Information Analysis Technologies, Techniques and Methods for Safeguards, Nonproliferation and Arms Control Verification Workshop

Workshop Proceedings

May 12-14, 2014
Portland Marriott Downtown Waterfront Hotel
Portland, OR, USA

Co-Sponsored by:
Pacific Northwest Chapter of the Institute of Nuclear Materials Management
INMM Division of International Safeguards
INMM Division of Nonproliferation and Arms Control
Workshop Proceedings

Information Analysis Technologies, Techniques and Methods for Safeguards, Nonproliferation and Arms Control Verification Workshop

Portland, Oregon
May 12-14, 2014

Co-Sponsored by:
Pacific Northwest Chapter of the Institute of Nuclear Materials Management
INMM Division of International Safeguards
INMM Division of Nonproliferation and Arms Control
Thank You for the Generous Support of Our Sponsors:

NTI
BUILDING A SAFER WORLD

Pacific Northwest
NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

Tableau
SOFTWARE
Organizing Committee

**Workshop Chair**
Zoe Gastelum – PNNL

**Technical Committee Chair**
Seán Kreyling - PNNL

**Technical Committee:**
Giacomo Cojazzi – JRC
Courtney Corley – PNNL
Mathew Ferguson – IAEA
James Garner – ORNL
Mark Greaves – PNNL
Joao Goncalves – JRC
Karl Horak – SNL
Antonio Sanfillipo – Qatar Foundation
Karen Steinmaus – IAEA
Maikael Thomas – SNL
David Thurman – PNNL
Richard Wallace – MELE Associates

**Logistics Committee**
Anne Czeropski – INMM
Kim Elzie – PNNL
Travis Gitau – PNNL
Dorothy LaMastus – PNNL
Jodi Metzgar – INMM
Christine Ortega – PNNL

**Communications**
Ken Jarman – PNNL

And coordinated with:

[NNSA](https://www.nnsa.gov)
Contents

Technical Session 1: Open Source Information Analysis and Societal Verification

Examining the Role and Research Challenges of Social Media as a Tool for Nonproliferation and Arms Control Treaty Verification 1

The Efficacy of Social Media as a Research Tool and Information Source for Safeguards Verification 14

Societal Verification: Past and Present 27

The Loch Ness Monster, Bigfoot, and Safeguards Conclusions 38

Technical Session 2: Big Data and Machine Learning

Tessera: Open source software for accelerated data science 44

GPU Computing for Nonproliferation and Arms Control Applications 60

Machine Learning for Classification and Visualisation of Radioactive Substances for Nuclear Forensics 61

The Future of Intelligent Systems for Safeguards, Nonproliferation, and Arms Control Verification 76

Technical Session 3: Cyber Security, and Information Protection and Control

Enhancing Cyber Security through the Use of Testbeds 93

A Threat Assessment Methodology for Critical Digital Assets of Nonproliferation Agencies: Preliminary Results 101

Resolving the Information Barrier Dilemma: Next Steps Towards Trusted Zero-Knowledge Nuclear Warhead Verification 113

Technical Session 4: Data Visualization and Analysis

A Visual Analytics Approach to Structured Data Analysis to Enhance Nonproliferation and Arms Control Verification Activities 121

Generating Confidence from Heterogeneous Data in Safeguards and Arms Control Monitoring Concepts 134

Presenting Load Cell Data in 3D to Improve Situational Awareness during Gaseous Centrifuge Enrichment Plant Inspections 148
Insights from Information Analysis and Visualization in Support of the Global Nuclear Detection Architecture

Technical Session 5: Trade Analysis and Illicit Nuclear Trafficking

Signatures of Illicit Nuclear Procurement Networks: An Overview of Preliminary Approaches and Results

PNNL Strategic Goods Testbed: A Data Library for Illicit Nuclear Trafficking

Trade Analysis and Open Source Information Monitoring for Non Proliferation

Cost-Sensitive Classification Methods for the Detection of Smuggled Nuclear Material in Cargo Containers

Technical Session 7: Statistical Methods

Management and Analysis of RPM Data

Enhanced Spent Fuel Verification by Analysis of Fork Measurements Data Based on Nuclear Modelling and Simulation

Incorporation of Page’s Test in the Separation and Safeguards Performance Model

Developing a Validation Methodology for Expert-Informed Bayesian Network Models Supporting Nuclear Nonproliferation Analysis

Technical Session 8: Data Integration

VARO - The Euratom Toolbox for Safeguards Data Evaluation

Data Integration and Entity Resolution: Challenges and Opportunities for Nonproliferation and Arms Control

Integrating Video and Event Information to Improve Safeguards Verification Tasks Using Commercial Off-the-Shelf Software

Nuclear Forensics Driven by Geographic Information Systems and Big Data Analytics
Examining the Role and Research Challenges of Social Media as a Tool for Nonproliferation and Arms Control Treaty Verification

Michael J. Henry, Nick Cramer, Jacob Benz, Zoe Gastelum, Sean Kreyling, Curt West
Pacific Northwest National Laboratory

ABSTRACT
Traditional arms control treaty verification activities typically involve a combination of technical measurements via physical and chemical sensors, state declarations, political agreements, and on-site inspections involving international subject matter experts. However, the ubiquity of the internet, and the electronic sharing of data that it enables, has made available a wealth of open source information with the potential to benefit verification efforts. Open source information is already being used by organizations such as the International Atomic Energy Agency to support the verification of state-declared information, prepare inspectors for in-field activities, and to maintain situational awareness.

The recent explosion in social media use has opened new doors to exploring the attitudes, moods, and activities around a given topic. Social media platforms, such as Twitter, Facebook, and YouTube, offer an opportunity for individuals, as well as institutions, to participate in a global conversation at minimal cost. Social media data can also provide a more data-rich environment, with text data being augmented with images, videos, and location data.

The research described in this paper investigates the utility of applying social media signatures as potential arms control and nonproliferation treaty verification tools and technologies, as determined through a series of case studies. The treaty relevant events that these case studies touch upon include detection of undeclared facilities or activities, determination of unknown events recorded by the International Monitoring System (IMS), and the global media response to the occurrence of an Indian missile launch. The case studies examine how social media can be used to fill an information gap and provide additional confidence to a verification activity. The case studies represent, either directly or through a proxy, instances where social media information may be available that could potentially augment the evaluation of an event.

The goal of this paper is to instigate a discussion within the verification community as to where and how social media can be effectively utilized to complement and enhance traditional treaty verification efforts. In addition, this paper seeks to identify areas of future research and development necessary to adapt social media analytic tools and techniques, and to form the seed for social media analytics to aid and inform arms control and nonproliferation policymakers and analysts. While social media analysis (as well as open source analysis as a whole) will not ever be able to replace traditional arms control verification measures, they do supply unique signatures that can augment existing analysis.

INTRODUCTION
There has been increasing interest within the nonproliferation and arms control community to leverage the analysis of open source information to augment traditional methods in verification activities. As part of an increasingly available body of open source information via the internet, the rapidly evolving social

---

media landscape has led to new interest in taking advantage of these new data sources to support nonproliferation and arms control verification.

The preliminary research described in this paper investigates the feasibility of applying social media signatures as potential arms control and nonproliferation treaty verification tools and technologies. This research leverages the increased attention, and access to, publically open source information. The overarching goal of this research is to inform the nonproliferation community as to where and how social media can be effectively utilized to complement and enhance traditional treaty verification methods. The research and results presented in this paper also highlight the fact there is no “silver bullet” for leveraging social media analytics for treaty verification. It requires the right mix of tools, technologies, and techniques to create and maintain confidence in any verification regime.

This paper is organized as follows: first, existing relevant research around social and other open source data is presented. Next, a series of relevant research questions are proposed. This is followed by the presentation of a case study in which social and open source data were applied to a proxy scenario to understand how such data might support verification activities. Finally, a discussion of the challenges in social media research is presented as well as a summary of future research needs.

SOCIAL MEDIA RESEARCH
Open source information is defined as any information that is not proprietary or classified – it can include publicly available information from websites, but also other information such as that purchased through subscription, information specifically purchased from external sources, or information from internal sources, websites, or unpublished reports. In the context of open source information, social media consists of the tools, techniques, and technologies that use the Internet to facilitate communication between users. Common social media platforms include blogs and microblogs, social networking sites, photo-sharing sites, map-based sites, and online classifieds.

Current research in social media analytics focused on three main topic areas: 1) Content analysis, 2 Group and network analysis, and 3) Predictive analytics.

CONTENT ANALYSIS
Content analysis involves processing the information presented in media in order to derive an understanding of that data. Common examples of content analysis include topic identification (the categorization of common topics or themes within a text corpus) and sentiment analysis. Sentiment analysis involves labeling text with sentiment (generally positive, neutral, or negative) based on a pre-determined dictionary and word combinations. New research is looking at ways to make that analysis more accurate, and also to distinguish more complex human emotion from social media, such as happy, sad, surprised, or even distress. There is also research aimed at better analyzing irony, humor, and sarcasm within social media data.

New research is expanding the traditional approach of social media analysis of textual data to include “social multimedia” – photos, videos, maps and other “online sources of multimedia content posted in settings that foster significant individual participation and that promote community curation, discussion and re-use of content.” Social multimedia analysis sometimes focuses on content analysis, to identify multimedia related to a single event, location, or topic. Other research is designed to understand online communities surrounding social multimedia and their interactions, such as those who routinely post and comment on Flickr.
GROUP AND NETWORK ANALYSIS

The ability to identify a user as being a member of a group, or analyzing the emergence and growth of groups within social networks is a growing area of interest within the social media analytics community. Social media users form connections based on direct linkages (eg. a Twitter “follower” or a Facebook “friend”) or through a common activity (eg. commenting on or re-tweeting the same messages, or “like”-ing the same content on Facebook). Relationships between individuals can be surmised based on their communication events, such as which users have expertise, or those in a superior position to others. Some online community research analyzes the strength or weakness of relationships between group members, and the impact of those ties on the dissemination of information among social networks. Other researchers have examined communication styles of distinct online communities and found differences in communication styles, such as use of hash tags, timing of communications (when they occur), and language use (contractions, emotive language, use of punctuation, etc.) for online communities with different members, communication goals, and topics. Some research is using social media content to determine characteristics of an online community, such as the militancy of an online community. On a smaller scale, other research has been used to understand social media users’ personality from publicly available information from Facebook profiles, such as self-description, status updates, photos, and interests.

PREDICTIVE ANALYTICS

Prediction analytics seeks to utilize social media indicators to predict future, real-world events. Prediction from social media utilizes a combination of the scale of social media coverage, sentiment analysis, and influence analysis. Predictive analytics using social media have been recently used, for example, to predict movie revenues, outcomes of political elections, financial market activity, popularity of a song on the Billboard weekly chart, and the effects of product marketing.

RESEARCH QUESTIONS

Before any determination can be made as to the efficacy of using social media for treaty verification, it is imperative to understand what type of data can be gathered from social media, and whether or not it is relevant to treaty verification activities. To that end, a series of research questions were defined which helped guide the gathering, analysis, and interpretation of the research performed.

- Research Question #1: How can physical sensor data collected through social media be integrated and used to corroborate or enhance traditional sensor data such as that gathered by the International Monitoring System (IMS)? What signatures can be generated from social data that could support or augment arms control and nonproliferation verification?
- Research Question #2: How can citizen generated data be integrated and used to enhance confidence in a treaty regime?
- Research Question #3: What is the current state-of-the-art of social media analytics and how, if at all, can these tool and technologies be utilized to enhance traditional treaty verification methods?
- Research Question #4: What, where, how, and when are these different signatures useful, and to which nonproliferation or arms control treaties do they best apply?
- Research Question #5: What are the challenges and deficiencies of current social media analytics, and what additional analysis tools or techniques are required to allow for useful interpretation of social media data to be useful to arms control and nonproliferation analysts and negotiators?

ARMS CONTROL AND TREATY VERIFICATION CASE STUDY
This research focused on developing a case study appropriate for identifying key signatures in open source and social media data that would be applicable to nonproliferation and arms control scenarios. In order to conduct this investigation, this research focused on proxy events – events that are non-sensitive and display similar characteristics as what would be expected from a nonproliferation or arms control treaty-significant event.

The case study selected and described herein focus on a “loud boom” proxy event. This proxy event focuses on meteor events, as they have the potential to create infrasound signatures (indications of atmospheric nuclear tests). Additionally, meteors are a common occurrence and there was the potential to capture a significant number of events. This event serves as a proxy for events that would be applicable to the Comprehensive Nuclear Test Ban Treaty (CTBT) and to a lesser extent, the Nonproliferation Treaty (NPT) (since a nuclear test in a previously unknown nuclear state is a significant indicator of an undeclared nuclear program in that state).

A major goal of the Comprehensive Nuclear-Test-Ban Treaty (CTBT) is the detection of nuclear tests around the world. This is accomplished technically through a deployed worldwide array of seismic, infrasound, hydroacoustic, and radionuclide sensors, called the International Monitoring System (IMS). These sensors can reliably detect the occurrence of these events, however the determination of location or source of the event is more difficult due to the relatively limited number of sensors deployed globally. Social media has the potential to complement this sensor suite by providing a rapid indication of both the location and source of the event.

The objectives of this case study are to investigate the capability to utilize social media as a sensor node, and its ability to rapidly provide an indication and assessment of an event.

**METEOR EVENTS**

The phrase “loud boom” was selected as a key phrase for the description of meteors hitting Earth’s atmosphere and creating a sonic boom in social media data. One year’s worth of Twitter data (consisting of 10 Billion tweets) was searched for the term “loud boom”. Figure 1 shows the results of that query over time.

![Twitter Search for "Loud Boom"](image)

*Figure 1 Twitter search for keyword "loud boom"*
Examining the results of the social media searches, the research team identified several large data spikes. The first large spike shown in Figure 1 occurred the first week of October 2011. Utilizing topic clustering on the social media data, and conducting additional open source research for verification, the following information was uncovered:

- The Draconid meteor shower occurred on October 8-9, 2011
- SpaceWeather.com had photos of the event posted by observers from countries including Norway, Czech Republic, and Britain
- The International Meteor Organization’s website had user-generated data (spatial and temporal) from 128 observers located in 28 countries for this specific meteor event (8-9 October 2011).
This analysis began by identifying a spike in the simple Twitter search term “loud boom,” and determined this event occurred the first part of October 2011. Utilizing user-generated data, the team was able to determine the frequency of event and the locations of those events worldwide. Figure 2 presents this data, which was collected through an open source search on the Draconid meteor shower.

This same Twitter data was input into PNNL’s IN-SPIRE. IN-SPIRE is a “text analysis and visualization tool that statistically analyzes unstructured text within a collection of documents, identifies topics (i.e., terms with high frequency and non-uniform distributions), and visually clusters the documents based upon their topical similarity.” Figure 3 shows the result of importing the Twitter data into IN-SPIRE.

Other spikes identified in the “loud boom” Twitter search yielded results about non-meteor events that could be discerned easily through topic/theme analysis, such as posts about fireworks on the Fourth of July, and a highly re-Tweeted story regarding a large thunderclap at a baseball game in Texas which had players running for cover (as seen in Figure 4). Though non-meteor events were common for the “loud boom” search phrase, understanding other “loud boom” events can support researchers in crafting more sophisticated search strings, to decrease the number of events that an analyst might be presented with for additional investigation.
**IMPLICATIONS**
This case study highlights the ability of social media to leverage its dual use capability as both a physical and citizen sensor to provide corroborating and complementary evidence of the presence, or absence, of certain treaty relevant events. This research also highlights the usefulness of social media for characterization of the occurrence of rapid, single, high-visibility events such as a meteor shower or earthquake.

**CHALLENGES IN SOCIAL MEDIA RESEARCH**
Despite its potential to support nonproliferation and arms control verification efforts, social media faces many of the same challenges as other open source information analytics research, mainly: accessing, storing, and processing information; verification of sources and dealing with misinformation and deception; and fusing various types of data. In addition to those issues, social media research faces some challenges uniquely its own.

This set of challenges faced by social media researchers is broken into three main categories: 1.) data, 2.) analytics, and 3.) policy. A description of each category is presented below. This is meant as a high level overview of some of the major issues in each category; researchers interested in evaluating the utility of applying social media analytics to verification activities face the ongoing challenge of working with an ever-growing list of challenges.

**DATA CHALLENGES IN SOCIAL MEDIA RESEARCH**

**Data Volume**
The volume of data produced by social media data sources is staggering – as an example, roughly 200 million tweets are posted on Twitter a day. While technology exists to handle this large volume of data – such as cloud computing – providing access to those large-scale infrastructures is a challenge. Given these large volumes of data, performing analytics becomes a challenge. Even simple analytics are limited beyond a million tweets, and more comprehensive, deeper analytics, can only support data sets on the order of hundreds of thousands of tweets. This motivates further research into processing streaming...
social media, and the adaptation of traditional analytic methods to the special characteristics of social media data.

**Fusion of Multiple Platforms**
Despite the wide availability of social data, little has been done in the area of fusing multiple data types. While current social media analytic solutions can analyze micro-blogs, social networking information, and photo data, they are rarely incorporated into a single analytic environment.

**Supporting the Analysis of Multiple Languages**
Most content analysis tools are language-specific, in particular, oriented to English text. Social media is a worldwide phenomenon, so the ability to analyze non-English languages is critical. Enhanced translation capabilities are needed that will not only recognize words or sentences, but pick up on language nuances, such as idioms and sarcasm. While there is some work being conducted in non-English language social media, that research needs to continue to grow until capabilities are sufficient for languages in all areas of potential societal verification interest.

**Analytic Tools that go Beyond Twitter**
Most of the state of the art social data analytical tools are Twitter-centric, mostly due to fact that Twitter data is more readily available than other platforms. Globally, there are a myriad of social media platforms, but most of the state-of-the-art solutions that handle social media data beyond Twitter and Facebook are still in pilot deployment phases. As a generally U.S.- or English speaking-centric platform, Twitter’s breadth for societal verification is limited, except in cases where users are specifically targeting messages to the United States. As such, analytic tools that go beyond Twitter are needed to process those potentially relevant data sets that are not housed on Twitter.

**ANALYTIC CHALLENGES IN SOCIAL MEDIA RESEARCH**

**Finding Meaning in Low Signal-to-Noise Environments**
The vast majority of social media data will not have any relevance towards verification activities. Even when searches are targeted via an expert-defined keyword list, social data is often too noisy to evaluate without further processing. Therefore, methods such as looking for frequency spikes across multiple platforms, or deeper investigations within a detected event’s time range can help disambiguate the noise from the useful data.

**Timeliness of Monitoring**
Social media can be processed either after the fact, or in near real-time. The timeliness of process can affect the types of analytics performed, and is dependent on a number of factors:

- Data availability, which refers to the whether the raw social data is available in real-time and whether processing tools can bring that data into an analytical environment in close to real-time,
- Research strategy, which refers to focused or broad monitoring, and
- Analytical capacity, which refers to the analytical bandwidth, which can limit the timeliness of societal monitoring

**Disambiguation**
Disambiguation refers to determining the difference in meaning between two words that are the same, or essentially the same. In the social media context, this happens when a relevant issue is searched by keyword, or used in an analytical tool to determine “themes.” Presence of a keyword with the wrong meaning can lead to incorrect results.
**Geo-location**
Currently, only a small portion of social data is geo-tagged. If geo-tagging is present, social data can be analyzed within a geographic context. Some text, while not geo-tagged, refers to geographic location within content. Some vendors can attempt to geo-bound social data via IP address. Limiting one’s data collection to only those social data with geo-tags will be limiting. However, when social data is independent from geographic information, it may be difficult to determine whether the social data is relevant to the research question. Thus, enhanced capabilities for gaining access to, and understanding the implications of, geo-location will be key in future social data analytics.

**POLICY CHALLENGES IN SOCIAL MEDIA RESEARCH**
*Collection and Use of the Data*
The collection and use of social data is regulated by each social media platforms’ terms of service (TOS) or terms of use (TOU). Terms of service/use can vary by platform and by country. Many TOS have intellectual property rights clauses that explicitly forbid the unauthorized copying of material. Many go further to bar all forms of social media data collection. Subject to ‘fair use’ exceptions in certain countries, such TOS could prevent a researcher from even copying material to their computer for further analysis and forbid any form of sale of that information to their clients, without permission.

In addition, there are legal issues based on how the data will be used and by whom. If data from social media is to be used for verification of nonproliferation and arms control agreements, its collection will likely have to be legally and contractually authorized, not only in the country where it was collected, but also before an international body or court. As such, for social media monitoring, all terms of service/use should be consulted and complied with before data collection occurs.

**Disclosure of Information**
Another area of concern regarding disclosure of information. Depending on the information being collected or analyzed, disclosure of information could put individuals or groups at risk for their personal safety, or could cause serious operational security concerns for facilities.

The potential ethical issues surrounding inappropriate disclosure of information is not always obvious. This may be compounded when handling large data sets such as those found in social data research. For societal monitoring, concerns regarding information disclosure arise due to the potential risk posed to staff or infrastructure.

**Privacy and Personal Data**
Social media research must comply with national and international data privacy legislation and relevant requirements for notice, consent, accuracy, security, and access when personally identifiable data is collected and stored. Simply assuming “mirror compliance” with United States law (when applied to the data privacy rules of another country) can cause issues.

**DEFICIENCIES AND GAPS AS APPLIED TO TREATY VERIFICATION**
Over the course of conducting the research described above, the research team identified a series of deficiencies and gaps in currently available technologies. These deficiencies and gaps were evident in the analysis of our cases studies, but also apply to the larger concept of applying social media analytics to treaty verification objectives. A brief summary of those challenges includes:
• Scalability of current tools. Analytic capabilities are very capable, but many do not have the 
ability to handle the “big data” collected through social media.
• Real-time analytics. Very few tools have the ability to collect and analyze social media feeds in 
real-time. Social media has the potential to generate millions of data points a second. An 
analyst of that data must have tools to handle that constant volume of data and distill it into a 
usable form.
• Multimedia analytics. Social media goes beyong textual data. An assessment of what types of 
nonproliferation-relevant information can be mined and utilized from image and video sources.
• Language support. There needs to be a mechanism to work through and understand social 
media data and signatures in other languages. This includes the use of tone, slang, and sarcasm. 
Many of the tools utilized can accept and collect data in other languages, but it requires the user 
to manually input the search terms in the language of interest.
• Data integration. There is little data integration across multiple social media platforms. Having 
a fused or cohesive environment in which to collect, integrate, and analyze multiple information 
feeds/types would be extremely useful, as well as the integration of non-social information.
• User-centric analytics. Current tools were not developed for use in treaty verification, and so 
lack many features that would be of use of an arms control or nonproliferation analyst. Further 
discussions need to occur to clearly define who the end-user is in these cases and what their 
needs are.

Events of interest to the arms control and nonproliferation community pose unique challenges for social 
media analytics. There needs to be continued research toward identifying treaty-relevant events across 
all potential treaties of interest. This paper offers the first steps down this path, but there is significant 
work that still needs to be done to define, develop, and test treaty relevant signatures and apply them 
to social media analytics.

FUTURE RESEARCH NEEDS

Social media analytics will never replace traditional treaty verification measures, and there is no single 
solution or “silver bullet” in either area that can provide the required confidence on its own. Adequate 
confidence can only be acquired through the right mix of tools, technologies, and techniques and these 
will vary between treaty regimes. The research performed and presented in this paper have led the 
research team to identify a number of lessons learned and ongoing needs for utilizing social media tools 
and technologies for treaty verification.

To maximize the unique qualities of social media, further research needs to be performed to understand 
the potential signatures available for specific treaties. Not all treaty regimes will lend themselves 
equally to social media analytics, and it will be important to understand when and where social media 
signatures are most useful.

Future research will need to go further and take an in-depth look at potential signatures using treaty-
relevant events and data, and focus analysis on specific treaties to determine applicability. As an 
example, in the context of the CTBT, physical and citizen sensors can enhance and corroborate data 
collected through the IMS, and assist an analyst in quickly ruling out an event or flagging it for further 
investigation. On the other hand, in an arms reduction treaty where monitored dismantlement is 
monitored within a tight security environment, social media may prove less useful. Many of the 
processes, tools, material properties, and personnel are classified. The potential social media physical
sensor data related to dismantlement of nuclear weapons is therefore limited. Research must continue to adapt current tools and technologies to develop an integrated suite of tools that can work together to provide these signature discoveries of interest to the nonproliferation and arms control communities.

Many current social media technologies are just now beginning to collect, process, and analyze “big data.” Utilization for treaty verification will also require the collection of large amounts of data from disparate sources worldwide. Adding to this difficulty is the fact that treaty-relevant events are very infrequent, which means that tools must be able to identify the proverbial “needle in a haystack.” The varied needs of data collection and analysis of social media when applied to treaty verification, require continued research toward the development and refinement of collection and analysis algorithms.

**SUMMARY AND CONCLUSION**

Based on the preliminary work reported here, it is apparent that some areas of arms control and nonproliferation lend themselves more readily to social media analysis than others. High impact or highly publicized events are fairly easy to detect on social media, while less noticeable events, or those events without much publicity (especially those events which are by nature hidden or covert), are less easily detected in social data.

In addition, a clear understanding of the capabilities and limitations of each technique, and the applicability to each treaty, is required for success. Treaty relevant events are infrequent and maybe intentionally masked. While some preliminary exploration may be interesting, clearly defined research questions are necessary to navigate large quantities of data. Making sense of data by limiting timeframe or by keyword alone is generally insufficient for drawing meaningful conclusions from social media data. Iteration and evolution of the analysis processes allow an analyst to refine the research questions, change the direction of the research if necessary, and ultimately uncover the treaty relevant signatures.

Finally, social media analysis will never be able to replace traditional arms control verification measures. But, it could support them. The legal mechanisms to enable social data analysis to support verification are part of a larger endeavor to incorporate open source analysis in verification activities. While still in its infancy, the use of open source information to support treaty verification has significant potential and the ability to greatly enhance and further nonproliferation and arms control efforts in the future. Continued research and development into the tools, technologies, and techniques to apply social media to treaty verification is absolutely necessary if the international nonproliferation community hopes to utilize the growing power of social media to move beyond what has been done in the past and maximize scarce resources, while maintaining confidence in treaty obligations.

---


ix Brubaker, JR, F Kivran-Swaine, L Taber, and GR Hayes. 2012. “Grief-Stricken in a Crowd: The Language of Bereavement and Distress in Social Media.” Presented at the Sixth International Association for the Advancement of Artificial Intelligence Conference on Weblogs and Social Media, June 4-8, 2012, Dublin, Ireland.


Evolutions.” Presented at the Sixth International Association for the Advancement of Artificial Intelligence Conference on Weblogs and Social Media, June 4-8, 2012, Dublin, Ireland.


The Efficacy of Social Media as a Research Tool and Information Source for Safeguards Verification

Thomas Lorenz and Yana Feldman
International Atomic Energy Agency, Vienna, Austria

Abstract

The IAEA aims to provide credible assurances to the international community that States are fulfilling their safeguards obligations in that all nuclear material remains in peaceful use. In order to draw a soundly-based safeguards conclusion for a State that has a safeguards agreement in force with the IAEA, the IAEA Department of Safeguards establishes a knowledge base of the State’s nuclear-related infrastructure and activities against which a State’s declarations are evaluated for correctness and completeness. Open source information is one stream of data that is used in the evaluation of nuclear fuel cycle activities in the State. The Department is continuously working to ensure that it has access to the most up-to-date, accurate and relevant open source information available, and has begun to examine the use of social media as a new source of information.

The use of social networking sites has increased exponentially in the last decade. In fact, social media has emerged as the key vehicle for delivering and acquiring information in near real-time. Therefore, it has become necessary for the open source analyst to consider social media as an essential element in the broader concept of open source information. Characteristics, such as “immediacy”, “recency”, “interactiveness”, which set social networks apart from the “traditional media”, are also the same attributes that present a challenge for using social media as an efficient information-delivery platform and a credible source of information. New tools and technologies for social media analytics have begun to emerge to help systematically monitor and mine this large body of data.

This paper will survey the social media landscape in an effort to identify platforms that could be of value for safeguards verification purposes. It will explore how a number of social networking sites, such as Twitter, Facebook and LinkedIn, might be relevant in the context of overall State evaluation. The paper will further survey the tools available in the public domain that improve the monitoring of, searching for and extraction of safeguards-relevant information. The paper will conclude with an assessment of the value of social media and social media analytics as a component of the open source analyst’s safeguards verification toolbox.

Introduction

One objective of the IAEA is to provide credible assurances to the international community that States are fulfilling their safeguards obligations in that no nuclear material is diverted from peaceful uses to nuclear weapons or other nuclear explosive devices. The IAEA Department of Safeguards evaluates States’ compliance with their safeguards obligations by seeking to draw conclusions on the completeness and correctness of each State’s declarations and the Department’s knowledge of the State’s nuclear material inventory, nuclear activities and facilities. Examples of past known non-compliance have underscored the limits of verifying States’ declarations alone. The State evaluation process leverages multiple data streams, technical disciplines, and analytical methods, including state declared information, such as nuclear material accounting and additional protocol (AP) declarations; results of in-field verification activities, such as environmental samples, design information verification,
non-destructive assay and destructive analysis; as well as open sources, such as satellite imagery, trade data, scientific and technical (S&T) publications, etc. All safeguards-relevant information available to the IAEA is used not only to support the drawing of safeguards conclusions, but also to plan the optimal set of safeguards activities to be conducted for a State. The Department of Safeguards works continuously to ensure that it has access to the most up-to-date, accurate and relevant open source information available, and has begun to examine the use of social media as a new source of information. The use of social networking sites has increased exponentially in the last decade. In fact, social media has emerged as the key vehicle for delivering and acquiring information in near real-time. Therefore, it has become prudent for the open source analyst to consider social media as an element in the broader concept of open source information. Any assessment of social media must furthermore include careful consideration of ethical and privacy aspects of using such information for safeguards State evaluation purposes, along with credibility, bias, etc., as with all other sources of information.

This paper will survey the social media landscape in an effort to identify platforms that could be of value for safeguards verification purposes. It will explore how a number of social networking sites, such as Twitter, Facebook and LinkedIn, might be relevant in the context of overall state evaluation. The paper will further survey the tools available in the public domain that improve the monitoring of, searching for and extraction of safeguards-relevant information. The paper will conclude with an assessment of the value of social media and social media analytics as a component of the open source analyst’s safeguards verification toolbox.

Open Source Information for Safeguards

Collection and analysis of open source information is part of the Department’s efforts towards a holistic approach to safeguards implementation that ensures that the planning, conduct, and evaluation of safeguards for an individual State are based on the analysis of all relevant information available to the IAEA about that State.

Open source information is one of three information streams available for State evaluation. The utility of this category of information lies primarily in the comparison with other sources, especially State-declared information, and in the assessment of the consistency of all safeguards-relevant information regarding nuclear fuel cycle technologies and activities in a State. A wide spectrum of knowledge is derived from the open sources to

---

1 For the purposes of this paper, open source information is defined as information generally available to the public from various sources, such as government agencies, private entities, academic institutions, the media, commercial subscription databases, scientific associations and commercially available satellite imagery. This information may be available in print form, electronic database entries, web pages, audio-visual format.

complement the declared information and results of safeguards activities. The specialised S&T and other fuel-cycle related information about the infrastructure, research and development (R&D) and manufacturing activities are used, *inter alia*, to determine the current nuclear capabilities and future nuclear programme development plans of a State. The background legislative, regulatory, and security information, for example, provides an essential context for the evaluation process, and may also aid in understanding historical issues. As the declared information and safeguards activities are largely organised around a particular State, open sources may further offer insight into transnational activities, such as import/export of goods that may be subject to AP declarations. To ensure a comprehensive knowledge base, the open source information is drawn from a variety of sources – S&T publications, trade and patent data, company annual reports, government documents, news reports; different formats – textual, visual, data; and different languages.³

Open source information enhances the State evaluation process in several ways, including:

i) as part of continuous monitoring of safeguards-relevant activities within a State and to evaluate declared information;

ii) to maintain current awareness and serve as early warning for safeguards-significant events;

iii) focused research to support in-field verification activities such as complementary access; and

iv) special investigations related to a specific State, entity, fuel cycle step, or technology.

With over 20 years of experience, the Department possesses robust open source capabilities, both in acquisition and analysis, and continues to expand and improve its open source resources to ensure that it has access to the most up-to-date, accurate and relevant open source information available. The proliferation of information available on social media necessitates an examination of these platforms as sources of information, and a component of the open source analyst’s safeguards verification toolbox.

What is Social Media?

In their seminal 2010 paper “Users of the World Unite! The Challenges and Opportunities of Social Media”, Andreas M. Kaplan and Michael Haenlein define social media as “a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content”.⁴ More casually, social media is described as a landscape of internet platforms where users can interact easily and share ideas, opinions and information. In contrast, during the “Web 1.0” era, if one wanted to share or disseminate information, one had to set up a website, which meant getting a domain name, renting web space, designing and maintaining the site, usually with technical support from a webmaster. With social media, information dissemination – be it personal, commercial, or official government information – has become easier, faster, cheaper, and accessible to more people. One can target the audience in ways that wasn’t possible before.


Kaplan and Haenlein call social media a “revolutionary new trend” that companies and other organizations that want to reach a wide audience should not ignore. The authors group social media applications into six distinct categories:

- Collaborative projects (wikis, social bookmarking)
- Blogs (including Twitter)
- Content communities (YouTube, Flickr)
- Social networking sites (Facebook)
- Virtual game worlds (World of Warcraft)
- Virtual social worlds (Second Life)

The above categorization is useful to narrow down where the analyst is more or less likely to find safeguards-relevant information.

**Collaborative projects** such as wikis can contain basic and contextual information relevant to developing baseline knowledge about a State’s security and energy situation as well as its nuclear programme, but are unlikely to contain detailed up-to-date data to aid in verification of a State’s declaration.

**Blogs** are an excellent medium for inexpensive publishing of non-proliferation and related analysis. Arms control and non-proliferation blogs play an important role in collective or public monitoring and verification by providing analytical products based on crowd-sourced open source research. Personal blog posts may contain images of a location that could be of interest to the Safeguards analyst. The popular micro-blog Twitter provides a massive and continuous flow of information concentrated in one platform that can be evaluated as a source of information and also for network analysis purposes.

With their emphasis on the visual media type, and the high element of self-promotion, **content communities** and **social networking sites** are also likely places where information can be found relevant for safeguards State evaluation purposes.

While **virtual social and gaming worlds** can be useful tools for educational and training purposes within the arms control and nonproliferation realm, they are not likely to contain information relevant to evaluating a State’s nuclear programme.

**Why Pay Attention to Social Media?**

With the striking success and popularity of social media, more and more internet activity is taking place there. People are using Facebook not only for keeping in contact with friends and family. Networks like Facebook, LinkedIn and Twitter are used for asking questions, finding expert opinion, and researching commodities and services. In short the web is moving into social media.
Statistics suggest that about 100 hours of video are uploaded to YouTube every minute, 5 500 million tweets are sent daily, 6 1.6 million public photos are uploaded to Flickr per day, 7 and 30 billion pieces of content are shared on Facebook every month. 8 More information is made available faster and by a larger number of people. For news consumption, this means receiving news or updates in near real-time, from people who may be closer to the action or the source of information. The users of social media appear to have recognised this. A survey by the Washington-based Pew Research Center showed that roughly a third of the US adult population gets their news from Facebook. 9 The simplicity of Twitter makes it an attractive information-distribution platform, for both the established traditional media outlets and the “citizen journalist”.

The interactive nature of social media sites is added-value. Twitter’s reposts, LinkedIn’s connections and groups, and Facebook’s friends and sharing features, for example, have the potential to provide additional leads for information by offering insight into the information generator’s network.

**New content: business migration to Facebook**

Responding to the social media trend – possibly recognizing that it is easier and less expensive to maintain a Facebook profile than a website – companies and other organizations such as government agencies are finding social media platforms an effective way to promote and deliver their content, and to interact with their customers. In a 2011 survey of nearly 800 marketing executives, the consulting firm McKinsey & Company found that 39 percent of companies already use social media sites as their primary digital tool to reach customers, and that within the next four years that number is expected to reach 47 percent. 10 In some cases, organizations and institutions have more information on Facebook than elsewhere on the web.

---

**Social media and search**

Search engine optimization has for the past 4-5 years focused on either developing a social media platform or integrating with existing social media. Testament to this trend is Google’s development of the Google Plus social platform and the Bing-Facebook partnership in 2010. Because of the sheer volume of information available on social media sites, and the speed with which it is produced, internet users are conducting more of their general searches either within a social networking site or using a social media search engine that concentrates on providing access to that type of data. For example, the US-version of Facebook offers a powerful semantic search function that could be used as a complement to traditional search engines when conducting general internet searches for State evaluation purposes.

**Integrating Social Media in the Information Collection Toolbox**

A recent study by the James Martin Center for Nonproliferation Studies suggests that data mining of “New Media”, which includes social media, can be useful for Safeguards analysts. Specifically, the study says that such data mining can help analysts when: (A) trying to create a complete picture of a State’s nuclear programme; (B) improve efficiency of safeguards compliance assessments; and (C) create mechanisms for timely detection or prediction of States engaging in clandestine activities.

**News monitoring for current awareness/early warning**

The fast-moving fluidity of Twitter makes it a rapidly expanding medium for news distribution. According to one study, over 85% of Twitter activity (106 million tweets) over a specific period (two months) was news related. Twitter offers “immediate” news delivery, allowing for near real-time monitoring of global events. This may improve the Department of Safeguards’ information collection capabilities in two ways. First, getting the news in near real-time from a large and diverse body of sources advances the Department’s efforts to maintain awareness and ensure early warning of issues with potential implications for safeguards implementation. Case in point: some of the first indications about the effect of the 2011 Tohoku earthquake and tsunami on the Fukushima Daiichi nuclear power plant (NPP) appear to have come through a tweet. More importantly, against the backdrop of the limited information initially offered by the Japanese government regarding the NPP, people largely turned to social media – Twitter, Facebook, YouTube – not just to gather accurate information, but to share “news”, analysis, and peer review. While perhaps less useful for in-depth analysis of States’ nuclear programmes, Twitter can be valuable when monitoring developments that could have consequences for safeguards activities at a nuclear facility and the safety and security of visiting Agency staff.

---

The interactive nature of social media sites can help the user establish the credibility of the information provider by locating them within their network. For example, retweeting information by a known and trusted source from a previously unknown source adds legitimacy to the latter. This can help assess trends in trustworthy news reporting and recognize new sources to monitor. The LinkedIn connections of a particular entity of relevance to safeguards implementation can help flesh out social networks of other individuals with similar interests or activities.

Gathering contextual information and verifying absence of undeclared activities

As people move to social media for consuming and delivering information, new data becomes available. This information can be used to supplement State declared information and results of field activities for verifying the absence of undeclared activities. For example, the following items of information of potential relevance to safeguards implementation and State evaluation were identified in Facebook, LinkedIn and Twitter.

Example 1: A ministry responsible for energy development of one State with limited nuclear activities chose to distribute its press announcements about energy-related projects via a Facebook profile, while its website was kept static. Such information could be of direct relevance for assessing the State’s AP article 2.a.x declaration about future plans.

Example 2: The Facebook group dedicated to stopping uranium mining in another State provides an open platform dedicated to a campaign against developing uranium mining in the State. It allows group members – apart from discussing the issue – to post documents, photos and third party analysis related to an ongoing uranium prospecting project. While these advocacy groups may be biased, the information that is made available on such social media platforms contains useful knowledge as these local organizations are “on the ground” and have access to sites, people and documentation that a desk-bound analyst at headquarters may not. The details provided by these advocacy groups offer valuable insight and provide context into the progress and current status of nuclear fuel cycle activities in a State.

---

15 The information accessible from the LinkedIn and Twitter profiles was publicly available and did not require a login to the sites. The Facebook profile required a login.
Example 3: A newly established Medical Centre in a third State has no website. But it does have a Facebook page which contains external photos of the building during the construction period as well as internal photos of the finished laboratories, including equipment, such as a cyclotron. This would aid the Department in design information verification and planning of field activities.

Example 4: A LinkedIn profile of an international consultant includes information about his involvement in the construction of a phosphate chemical complex, including construction time and details about what type of plants and associated buildings the complex will house. The website of the mining company did not offer such details, only limited information announcing the project. Such information would be used to assess the completeness and correctness of the State’s AP declaration.

Example 5: A recent conference phenomena is participants’ tweeting live content such as pictures of slides of presentations and posters. This is a potentially unique monitoring tool for the analyst not present at the conference. Firstly, traditional media may not report on the happening or the content of the event. Secondly, the material may not become available online for non-participants. While the Department of Safeguards keeps track of upcoming safeguards-related conferences, it is not in the position to physically be present at all of them. The Twitter hashtag (conversation ID) for the conference is a handy way to focus on and monitor the specific event.
Relationship discovery and link analysis

The full power of social media information could be further realized through the application of network analysis and other “big data” analytical tools.

Because of the way social media platforms are organized, they are a unique source for relationship discovery and link analysis that can help develop a fuller picture of a State’s nuclear activities, such as mapping collaborative research and monitoring trade in nuclear-related commodities. It’s important to stress that the information mined via social media is just one source that must be used in combination with all other available data.

A project undertaken by the Pew Research Center demonstrates how Twitter can provide interesting network information from events such as conferences. They set out to map Twitter topic networks and observed that there are at least six types of structures to the conversations as people reply to and mention each other in tweets. In what the research calls “Tight Crowds”, discussions are characterized by highly interconnected people, such as professional or hobby groups.

The study notes that conferences could take this form (Figure 7). If one imagines a nuclear fuel cycle-related conference and the potential “Tight Crowd” Twitter conversations that could ensue from the attendees, it is possible to map out a network of links among the entities who are themselves interested in the conference topic, which would thus serve as a potential data source for State evaluation.

Figure 6: Pew Research Center showing a visualized example of “Tight Crowd” Twitter conversations. Source: http://www.pewinternet.org/files/2014/02/Examples-of-6-kinds-of-network-types-final.pdf.

Kristan J. Wheaton and Melonie K. Richey further demonstrate the analytical power of social network analysis in their article “The potential of Social Network Analysis in Intelligence”. By using the Twitter handle of the American Nuclear Society with its 200,000 connections, they are able to identify key actors and organizations in a network. This can have several uses for the Safeguards analyst, including using the technique to identify trustworthy and reliable sources in a specific topic area. In another example, more pertinent to evaluating a State’s nuclear fuel cycle activities, Wheaton and Richey use Twitter’s list feature and some cross-referencing techniques to identify 50 entities having a conversation about strategic mining and minerals (Figure 8). Techniques such as this can be used to gain a fuller understanding of the major players in a particular nuclear fuel cycle-related industry or commodity market.

Challenges in Using Social Media

As with information available elsewhere on the internet, there is a host of challenges to using social media as a data source. What sets social media apart from “traditional media” is simultaneously what presents a challenge for using it as an efficient information-delivery platform and a credible source of information. Some of these challenges include assuring data quality, evaluating the credibility of sources, dealing with the transitory and voluminous nature of social media information, data management, and ethics.

Data and source quality

The ease with which one can disseminate information via social media means data quality is of considerable concern. Anyone can post information to social media platforms and reach potentially millions of people very rapidly. There is an inevitable trade-off between immediacy and accuracy. The emphasis in social networking sites on getting information out, quickly, without moderation, runs the risk that an erroneous piece of information is released - intentionally or accidentally – and then propagated, despite it being possibly incorrect, out-of-date, or unsubstantiated.

Assessing the identity and credibility of the source of information acquired via social media can be more difficult than with “traditional media” – due to the tradeoff between anonymity and credibility. While a website usually has an “About” page and/or a traceable domain name, a user profile on Facebook or Twitter has little to no information associated with the author. Information presented on a website can also be wrong or misleading, but it’s easier to be less transparent via social media since it is more anonymous, and one is not expected to reveal a substantial amount of personal information. When

---

evaluating data stemming from social media, rigorous information and source evaluation becomes critical. A recent paper by Rob Thomson et al on the credibility of information shared on Twitter, found that “In general, profile anonymity proved to be correlated with a higher propensity to share information from low credibility sources”. However, the authors also noted that “Tweets with reference to third-party information made up the bulk of messages sent, and ... a majority of those sources were highly credible.” Information specialist Phil Bradley notes in his blog that the internet has recently seen a surge of accidental or deliberate hoax information on social media, and he warns of increased distrust. He calls for information professionals to step in and validate authority and content. If social media is to become a routine source of information, the Department would need to further improve its already robust methodology for evaluating source and information credibility to account for unique characteristics of social media.

**Transient and voluminous information**

To monitor and capture relevant information in a continuous stream of fast-moving data is a tall task. Firstly, the information is presented in multiple streams in a seemingly chaotic fashion. Secondly, the voluminous amount of information means there is a lot of “noise” around your specific search queries. How does one monitor and filter results?

**Monitoring Twitter:** Twitter has a search function and while it lets you save searches it does not let you set up alerts based on keywords, unless monitoring a specific Twitter account. Also, a search in Twitter will reportedly favour recent and favoured tweets, with not all tweets being indexed. Apart from Twitter’s inherent search and alert functions, there are a host of third-party applications to monitor conversations on Twitter. However, these tools are susceptible to Twitter’s API specifications, which change regularly, with the consequence that the tools need updating, or no longer work. Examples of third-party search tools that search Twitter are:

- Topsy (http://topsy.com/) offers comprehensive search of every tweet ever published, dating back to 2006.
- TwittStorm (http://www.twittstorm.com/)
- Twilert (http://www.twilert.com/) offers alerts (premium service)

There are other third-party applications that search not only Twitter, but other social media, acting as a meta- or federated search engine. Some examples are:

- SocialMention (http://socialmention.com/)
- Spezify (http://www.spezify.com)

---

22 An API (application programming interface) “specifies how some software components should interact with each other” ( “Application programming interface”, http://en.wikipedia.org/wiki/Application_programming_interface [last modified 2014-04-18]).
• Mention (https://en.mention.com/) Offers alerts and advanced analytics (premium service)

Monitoring Facebook: Monitoring and searching for information on Facebook carries with it a different set of challenges since the user in general is limited to seeing information posted within that user’s network of friends. Some information is publicly available and appears in search engine results. Facebook’s privacy pages notes that “People who aren’t logged into Facebook can still see things you’ve shared with the audience set to Public, as well as your public info such as your name, profile picture, cover photo, gender and networks”.23 However, the amount of information that is made public is small and is decreasing.24

Facebook’s Graph Search functionality lets users find information from within their network of friends. Graph Search, introduced in March 2013, is a semantic search engine allowing natural language queries. According to the Wikipedia entry for Graph Search, the feature “combines the big data acquired from its over one billion users and external data into a search engine providing user-specific search results”.25 It is worth reiterating here that Facebook search is powered by Bing, and as such, Facebook results also show up in searches conducted in Bing.

As mentioned above, general search engines are optimizing to integrate with social media, which means one can also use engines such as Google and Bing to search monitor social media. Setting up Google alerts would be one way to monitor social media.26

Information management

Related to the above challenge of identifying and capturing relevant data in live information streams on social media are the challenges of saving, storing and organizing the information, particularly before it disappears or is replaced with more up-to-date information. A number of questions remain to be answered:

• The nature and actual location of the data: Is the social media event pointing to the information that is residing on another website or is it the information itself?
• What format is most suitable for storing the data?
• How does one archive and organize information from social media?

Ethics – Developing social media Guidelines

Apart from dealing with technical challenges, there are ethical questions surrounding the use of social media for information collection purposes. The open source information collection efforts at the Department of Safeguards are conducted in an open and transparent manner. Analysts search only for publicly available information and without hiding or using false identities. However, since social media

24 “Facebook Users Take a Sharp Turn Toward Privacy”, Polytechnic Institute of New York University, 2012-02-21.
as a platform is anchored around networks, the extent to which information can be monitored from outside of that network may depend on several issues, such as how “public” the information and the user are, the level of permissiveness of the social media platform’s privacy policy, etc. It could be argued that if the information is available in open profiles on Facebook and LinkedIn (that is, indexed by search engines and accessible without a login), it should be treated as any other open source information on the internet. If searching social media is to become an integral part of the Safeguards analyst’s information collection toolbox, a thorough assessment of ethical and privacy aspects of the collection process is necessary as well as the development of guidelines and procedures.

Conclusions and Way Forward

The use of social networking sites has vastly increased in recent years, with social media becoming a key vehicle for delivering and acquiring information in near real-time, thus warranting a consideration as an element in the broader concept of open source information. This paper has taken an explorative look at the potential for social media to aid the Safeguards analyst in using open source information for State evaluation purposes. Content derived from social media platforms is one additional piece of open source information, which, once its credibility is established, is integrated with State-declared data and results of in-field verification activities for a consistency analysis of all information relevant to safeguards State evaluation. With concrete examples of where content on social media has offered such potentially relevant information, this paper has demonstrated that social media deserves further attention and assessment as a tool in the open source analyst’s tool box.

The collection and use of information contained in social media platforms present a set of technical and ethical issues. Some of these issues are: Can social media serve as an early-warning tool and make news monitoring of safeguards-relevant events more efficient? Can it be used as an additional information source for gathering contextual information for State evaluation? How can the credibility concerns arising out of the anonymity nature of social media be neutralized? Can deeper analysis of relationship discovery and network and link analysis be applied to social media for safeguards purposes? Any assessment of social media must furthermore include careful consideration of ethical and privacy aspects of using such information for safeguards State evaluation purposes.

The Department of Safeguards would benefit from an exploration – internal and jointly with its partners in the nonproliferation community – of ways in which social media can be mined for accurate and timely information of relevance and value for the State evaluation process, and to address the challenges related to the collection and use of social media as outlined above.
Societal Verification: Past and Present

Jessica Bufford
Pacific Northwest National Laboratory, Serving as a NGFP Fellow in NNSA

Abstract:
Verification is one of the greatest challenges to the vision of a world without nuclear weapons as laid out by President Obama at Prague in 2009. One potential solution is societal verification, or the participation of the general public in arms control verification. Though the concept was originally developed in the 1950s, societal verification remains vague and poorly understood. A survey of historical definitions reveals four key characteristics, namely that societal verification is society wide, treaty-based, systematic, and individually driven. Though a full societal verification regime does not exist, individual participation in treaty monitoring in the form of citizen monitoring is already evident. Thanks to advances in technology, non-governmental organizations and individual experts are able to provide insights into the implementation of certain international norms and treaties. Comparative analysis of citizen monitoring and societal verification demonstrates that there are significant gaps between the current monitoring capabilities and an official verification regime using socially generated information. Despite these gaps, there is significant potential for the public to support arms control in the future, and a better understanding of how to overcome some of these gaps could be an important contribution to efforts to establish a world without nuclear weapons.

Societal Verification: Past and Present

On April 9, 2009 President Barack Obama announced the U.S. commitment “to seek the peace and security of a world without nuclear weapons.” This landmark declaration elicited mixed reactions, welcomed by advocates of global nuclear disarmament and opposed by supporters of nuclear deterrence. Regardless of how it was received, President Obama’s statement provided new impetus and energy to discussions about global disarmament and the necessary steps to achieve such a vision. With President Obama’s statement came two questions: what would a world without nuclear weapons look like, and how do we get there?

Working towards a world without nuclear weapons will require overcoming significant challenges. President Obama recognized the enormity of this task, quickly acknowledging that “this goal will not be reached quickly – perhaps not in my lifetime.” One particularly difficult challenge will be verifying a treaty banning nuclear weapons. Today, the Nuclear Nonproliferation Treaty prohibits all but five member countries from possessing nuclear weapons and establishes a verification regime using safeguards implemented by the International Atomic Energy Agency. The safeguards regime, however, has proven to be imperfect, as a few states have been able to circumvent inspections to pursue nuclear weapons capabilities. Responding effectively to safeguards violations as an international community has also proven to be challenging, as demonstrated by the wide variation in state responses to Iran’s safeguards violations. In a world without nuclear weapons, treaty verification would need to be close to

1 Barack Obama, “Remarks at Hradcany Square, Prague, Czech Republic” (speech, Prague, Czech Republic, April 5, 2009), the White House Office of the Press Secretary, http://www.whitehouse.gov/the_press_office/Remarks-By-President-Barack-Obama-In-Prague-As-Delivered.
2 Ibid.
perfect. Any violation would be seriously destabilizing; therefore, the international community would need to be extremely confident in each state’s compliance with a treaty banning nuclear weapons. How to achieve that level of confidence remains unclear, but it will most certainly require new approaches to verification.

One potential approach is societal verification, or “...a system of monitoring compliance with treaties, and detecting attempts to violate them, by means...based on the involvement of the whole community, or broad groups of it...” Societal verification envisions the incorporation of the general public into arms control monitoring and verification. Though it was originally developed in the 1950s, it has gained renewed interest due to recent developments in technology. New communication capabilities like the Internet, smartphones, and public satellites have made the world more accessible to everyone. Some of these technologies could be used to support arms control treaty monitoring and verification, but how societal verification would be implemented remains unclear.

This paper attempts to shed some light on the ambiguous concept of societal verification. In order to understand the potential of societal verification, it is important to understand the history of the concept, which will be explored in the first section. Definitions of societal verification will then be analyzed to identify its characteristics. These characteristics will be compared against current civilian monitoring activities to assess the degree to which societal verification exists today and what gaps remain. The paper concludes with thoughts on the potential future of societal verification and its role in arms control.

History of Societal Verification

The concept of societal verification was initially articulated in the 1940s by Leo Szilard in response to the development of nuclear weapons. Szilard was the first physicist to conceive of the use of a nuclear chain reaction in a weapon. Though he was instrumental in the Manhattan Project, Szilard quickly became an arms control advocate, requesting that the U.S. government not drop a nuclear weapon on Japan. Szilard correctly predicted the nuclear arms race during the Cold War, noting that with the development of the nuclear weapon “the destructive power which can be accumulated by other countries as well as the United States can easily reach the level at which all the cities of the ‘enemy’ can be destroyed in one single sudden attack.”

To mitigate the risk of a “preventative” nuclear war, he advocated a system of controls on nuclear weapons production. Since the consequences of violating a ban on nuclear weapons would have significant global implications, Szilard encouraged engaging scientists and engineers as informants. He asserted:

...it would be desirable to create a situation which would permit us to appeal in various ways to physicists and engineers everywhere for information that would uncover violations of the controls...It would be quite essential that the people of this country and

---

the world be brought to understand from the start that any difficulties which any nation may place in the way of the established controls would have to be considered as tantamount to a “declaration of war”. Szilard’s ideas on arms control and verification were contained in a memo to President Franklin D. Roosevelt; however the unexpected death of the president prevented the memo from being delivered. These ideas did not end there, but inspired others in the 1950s and 60s during the advent of arms control.

Early critics of arms control argued that treaties could not effectively be verified. In response, a group of academics and leading thinkers on disarmament issues conducted a feasibility study to assess problems with implementing and enforcing disarmament agreements through inspections. In *Inspection for Disarmament*, Seymour Melman, among others, proposed that a disarmament verification regime include “inspection by the people.” Under a disarmament treaty, any citizen of any country would have an “explicit obligation...to report violations to [an] international inspectorate.” He and Lewis Bohn advocated mobilizing members of the public to report treaty violations within their own countries. They asserted that the general public supported nuclear disarmament and could feasibly participate in arms control out of a sense of moral obligation. Broad awareness of and participation in arms control agreements were central to the early articulations of societal verification, expanding the ideas of Szilard to encompass not just the expert population but also the general public.

Another important aspect of inspection by the people was the creation of reliable communication channels between governments, international governmental organizations, and the general public. Grenville Clark and Louis Sohn proposed the creation of a UN Inspection Service that would be accessible to the public for reporting of violations. They thus laid out the principles for citizen reporting in an annex as:

...Any person having any information concerning any violation of this Annex or any law or regulation enacted thereunder shall immediately report all such information to the United Nations Inspection Service. The General Assembly shall enact regulations governing the granting of rewards to persons supplying the Inspection Service with such information, and the provision of asylum to them and their families...No nation shall penalise directly or indirectly any person or public or private organisation supplying information to the United Nations with respect to any violation of this Annex...

While the individual was charged with reporting violations, the international Inspection Service provided the method for communicating those violations to the international community. Clark and Sohn recognized that such reporting would be dangerous, and recommended measures to protect whistleblowers.

As the Cold War progressed, however, support for societal verification declined. Mounting tensions between the United States and Russia rendered citizen reporting politically impossible. At the same time

---

government owned satellite technology improved, making it possible to verify treaty compliance using national technical means. As arms control developed using national technical means and on-site inspections for verification, societal verification faded from arms control and disarmament discourse for a time.

During the 1990s societal verification gained renewed attention. Joseph Rotblat picked up many of the same themes from early concepts, including the “deeply felt moral obligation” to report violations of a treaty to eliminate nuclear weapons, the need for broad participation, and the provision of legal protection for reporters.\textsuperscript{10} He also suggested that scientists and technologists in relevant industries could act as watchdogs of both their organizations and of their colleagues. A few years later Costa Rica brought the idea of an international disarmament verification body back into debate by circulating a Model Nuclear Weapons Convention at the UN in 1997. Their proposal included a Verification Agency to which individuals could report violations, similar to the Inspection Service suggested by Clark and Sohn.\textsuperscript{11}

Today, interest in the challenge of verifying a world without nuclear weapons has spurred renewed attention to societal verification. New technologies and media could make societal verification a meaningful addition to traditional monitoring and verification methods. The ability of the general public to observe, document, and share information has exploded in recent years. Smart phone technology alone has created a widespread network of potential sensors connected to social media that has enabled individuals to widely share information almost instantaneously. Though there appears to be great potential in new technology to support arms control, understanding the social and political context of societal verification to support an arms control treaty remains a significant challenge.

Definitions and Characteristics

Societal verification remains a poorly defined concept, in part because it has never been used in a legally binding agreement nor implemented in any country. Over time, a wide range of activities, behaviors, and principles have become associated with the term. A comparison of a few leading definitions provides valuable insight into the basic characteristics of societal verification.

One of the modern proponents of societal verification, Joseph Rotblat, defined societal verification as ...

...a system of monitoring compliance with treaties, and detecting attempts to violate them, by means other than technological verification...based on the involvement of the whole community, or broad groups of it...\textsuperscript{12}

His definition highlights some of the differences and similarities between traditional and societal verification. Like traditional verification, societal verification is intended to support formally negotiated treaties. All verification needs to be systematic to ensure it can effectively detect any violations. Unlike traditional verification methods, however, societal verification relies upon the observation and actions of the general populace to verify compliance and detect violations rather than relying upon national technical means and mutual inspections performed by government inspectors.

\textsuperscript{10} Rotblat, “Toward a Nuclear Weapon-Free World: Societal Verification,” 57.

\textsuperscript{11} Deiseroth, "Societal verification: wave of the future?” 270.

\textsuperscript{12} Rotblat, “Toward a Nuclear Weapon-Free World: Societal verification,” 52.
Kirk Bansak emphasizes similar points in his definition, describing societal verification as “the ways in which social actors and social activities can collectively contribute to the verification of arms control agreements.” Like Rotblat, Bansak identifies formal arms control agreements as the context for societal verification and looks to the broad participation of the general public to enact verification. Bansak diverges from Rotblat in how he describes the role of societal verification within treaty enforcement. While Rotblat defines societal verification as independent from national verification, Bansak places societal verification within and in support of broader verification goals, describing it as a “contribution” to arms control verification.

Finally, Dieter Deiseroth notes that the term “connotes the involvement of civil society in monitoring national compliance with, and overall implementation of, international treaties or agreements.” While the underlying assumptions about treaty-based, widespread participation in verification are the same, Deiseroth furnishes an important addition to the concept of societal verification by describing the activity of civil society as “monitoring.” While verification suggests judgment of the information collected, monitoring carries a connotation of simple observation.

From the above analysis, four common characteristics of societal verification shared among these definitions become clear. Societal verification can be characterized as:

1. **Society wide**: Under a treaty with societal verification provisions, all citizens or large portions of the population are interested, motivated, and able to participate in monitoring and verification of arms control agreements. This broad level of participation implies a society wide familiarity with treaty provisions, a strong interest in state compliance, and greater devotion to the enforcement of an international agreement than to one’s own state.

2. **Treaty-based**: Verification activity supports a formal, multilateral arms control treaty, which implies that societal verification is acceptable to all parties. For a state to accept any kind of verification, the capability of the other party or parties to look into the activities of that state must be matched by the ability of that state to look into the activities of the other party or parties. Consequently, the interest, incentive, and ability of citizens to report treaty violations needs to be comparable in all parties to such a treaty.

3. **Systematic**: With the call for reporting from a broad constituency comes the need to collect, organize, and analyze information coming from the general public. The quantity of information coming from citizen reporting could be enormous. Methods for organizing, assessing, and presenting information from citizens will be essential for the success of societal verification. Societal verification must also be equally applied to all states within a treaty, with standardized reporting processes and availability.

4. **Individually driven**: Societal verification shifts monitoring from the sole domain of governments to the responsibility of individuals as well as national technical means and verification organizations. Citizens become monitors of treaty implementation, and investigations of violations are driven by their reports.

In order for societal verification to realistically be implemented, all four of these characteristics will need to be present. One of the challenges for future efforts to implement societal verification will be to bring about those conditions in support of future arms control treaties.

---

Societal Verification Today: Citizen Monitoring

Societal verification has never been used in an international arms control agreement, but public participation in treaty implementation, or citizen monitoring, is already taking place. This paper defines citizen monitoring as the collection and dissemination of publically available information by civil society and individuals to monitor the implementation of treaty obligations by state parties. Like societal verification, citizen monitoring describes participation of segments of society in treaty implementation, but it is more limited in scope and capabilities than societal verification. Citizen monitoring emerged with the development of communication technologies. Today, organizations and individuals can access, analyze, and disseminate more information than ever before thanks to the Internet, enabling civil society to comment on treaty implementation with greater credibility than in the past. Some non-governmental organizations (NGOs) have taken on unofficial monitoring and verification roles in support of specific treaties or international norms. Loose networks of experts have formed that collect, analyze, and self-verify data to provide both historical and real-time snapshots of nuclear activities within certain states. Though there are many instances of citizen monitoring across the range of international issues, several prominent examples within international security suggest ways in which civil society could participate in future verification efforts.

The Landmine Monitor

One of the best examples of the potential contribution of NGO monitoring is the Landmine Monitor. Published by the International Campaign to Ban Landmines (ICBL), it is a compilation of open-source reports and assessments that tracks implementation of the Mine Ban Treaty.\(^\text{15}\) When the Mine Ban Treaty opened for signature in 1997, the only verification provision in the treaty was the regular submission of voluntary reports to the United Nations Secretary-General. Civil society rapidly identified this weak verification mechanism as a significant gap in the treaty regime. In response, ICBL started the Landmine Monitor to provide third party treaty verification. Using publicly available data, researchers analyze trends and developments to measure progress on treaty implementation. Though not officially recognized by governments, the Landmine Monitor has published reports every year since 1998, becoming the de facto monitoring organization of the treaty and providing important information on state implementation of the Mine Ban Treaty.\(^\text{16}\)

The Landmine Monitor demonstrates how civil society can effectively participate in treaty monitoring, even as \textit{ad-hoc} participants. Though they only have to accesses public information, the degree of connectivity afforded by the Internet has enabled the Landmine Monitor to provide significant insight into treaty implementation. As Michael Crowley and Andreas Persbo observe:

Despite the potential weaknesses sometimes inherent in a system which incorporates the collection and analysis of open source material, Landmine Monitor has over the years succeeded in collecting a large amount of information on state compliance with

\(^{15}\) Formally known as the Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on Their Destruction, the Mine Ban Treaty prohibits the use, stockpiling, production and transfer of antipersonnel mines. There are currently 161 States Parties.

the Mine Ban Convention. While this inevitably remains an imperfect monitoring system in some respects, it...has had an important impact on the Convention’s implementation." Information that was previously inaccessible outside a country can now be found online, facilitating efforts by NGOs to remotely monitor treaty implementation and making them more viable partners in future treaty verification efforts.

**The National Council of Resistance of Iran and the Institute for Science and International Security**

Citizen monitoring can also focus more narrowly on the implementation of treaty obligations by one country. In August 2002 the Natanz facility in Iran was disclosed by the National Council of Resistance of Iran (NCRI), an Iranian opposition group based in Paris. Following this revelation, the Institute for Science and International Security (ISIS) requested satellite imagery for the area around Natanz in September 2002 and February 2003. Based upon those images, ISIS developed a detailed assessment of the size of buildings within the nuclear complex, the likely nature and purpose of activities within those buildings, and the type of military strike that would be necessary to destroy certain facilities. Since then, ISIS has provided regular analysis of the Natanz facility through satellite imagery.

ISIS activities demonstrate two important aspects of citizen monitoring. First, individuals today are able to monitor a wide variety of activities using high resolution imagery that was previously available only to governments. During the Cold War, the only satellites in space were controlled by the U.S. and Soviet governments. Today, commercial satellites with imagery capabilities have become so common that an individual can buy images of almost anywhere in the world, facilitating the emergence of what could be considered open-source national technical means.

The second important characteristic of NCRI and ISIS activities is that citizen monitoring occurs remotely, taking place largely outside of the state being monitored. Outside observers are free to make controversial claims or revelations because they have few concerns about government repercussions. Internal reporters would likely be detected and stopped by government forces, but outsiders are able to report without fear of reprisals from the state they are monitoring. Such observation, however, is limited to what the researchers know to investigate. Major revelations like Natanz may always require assistance from individuals or entities within the state or dissenters from the government, as was the case with the NCRI revelations. Regardless of this potential limitation, the ISIS reports demonstrate that civil society has the capacity to provide some *ad-hoc* monitoring of treaty commitments, even if the analysts are working thousands of miles away.

**Expert Blogs**


Informal networks of experts have emerged that can also act as whistleblowers and monitors. Today, experts in the nonproliferation and arms control community share ideas and observations through blogs and social media, reaching a broader audience more quickly than traditional publishing avenues. More than just publishing information, members can pulse the community with a question and rapidly receive an informed answer. A researcher can ask their colleagues around the world to identify a picture or verify a piece of information and receive responses in minutes. Sharing ideas and analysis is also much quicker, as a post on a blog has the capability to start a lively debate within a few hours.

Blogs have become an important tool for experts to communicate their ideas rapidly. Some are hosted by organizations, such as the Nukes of Hazard\(^{20}\) and ANS Nuclear Café\(^{21}\), while others are maintained by individuals like Restricted Data and Arms Control Wonk.\(^{22}\) Arms Control Wonk in particular provides useful insight into how experts are able to collect, analyze, and discuss information from a wide variety of sources. Blog creator Jeffrey Lewis and other contributing authors analyze information collected from news organizations, government publications, and media such as photos and video and post it on their blog, distributing it to the community and generating discussion. This discussion can then assess the accuracy of information and identify problematic analysis. For example, in November 2011 Georgetown professor Phillip Karber asserted that China possessed 3,000 nuclear weapons hidden in a vast network of underground tunnels.\(^{23}\) One week later, Jeffrey Lewis responded with a post on Arms Control Wonk analyzing and critiquing Karber’s claims.\(^{24}\) It resulted in a lively debate on the blog that ultimately led the expert community to refute the results of the Karber study.

This self-reflection suggests that this community can self-correct. Though governments will need to treat information from citizen reporting judiciously, the expert network can evaluate and refine information being published and circulated. This will improve the quality of any potential contributions to arms control efforts. The expert community has also demonstrated a willingness to sift through large amounts of data generated by the public and have the background to identify potentially significant clues. Open-source conversations could serve as first notifications to governments of treaty violations, though official investigations would likely be needed before any information could be presented in official diplomatic channels. As well informed non-governmental observers, the extensive expert community network could provide important support to future arms control.

**Gap Analysis**

Though citizen monitoring today represents a potential fore-runner of societal verification, there remain significant gaps between ideal societal verification and modern citizen monitoring. These gaps can best be understood by comparing the four characteristics of societal verification with citizen monitoring.

---


Individually Driven: Insider versus Outsider Reporting

Individual participation is at the core of both societal verification and citizen monitoring, but the location of the participants differs significantly. Most citizen monitoring is conducted remotely by foreigners, whereas societal verification envisions citizens reporting suspicious activities within their own countries. Globalization and electronic communication technology has made countries more accessible to outsiders by facilitating both in-person and electronic visits. Outsiders are also more able to report on potential treaty violations because they do not face negative repercussions from the government being monitored. As Ronald Mitchell observes:

- Outsiders have stronger net incentives to monitor and provide information, although they may have more limited capacities, since the risk from the suspect government is far less. Indeed, most governments would consider any effort to retaliate against their citizens for helping to reveal clandestine nuclear activity as warranting severe sanctions.
- Thus, these actors face far less risk of retaliation, assuming they are outside the suspect country at the time the information becomes public.  

Outside reporting may bring attention to events that would not be reported from inside a country; however it is more susceptible to inaccuracies because the reporting is done from a distance. Behavior that a local would recognize as not in violation or irrelevant to an arms control treaty may appear suspicious to an outsider. Remote observation may also miss important activity that cannot be easily tracked remotely, namely human interactions. Modern technology has greatly facilitated citizen monitoring, and outsiders can now report with relatively high confidence about certain activities within a state, but it is still a far cry from the mobilization envisioned by societal verification.

Internal mobilization is one of the toughest challenges for societal verification. Early articulations of societal verification relied on the creation of a supranational loyalty to global disarmament that would supersede national loyalty. This loyalty has yet to develop, and today those leaking sensitive government information are still tried for treason and espionage. Developing loyalty to an abstract idea rather than a particular country could be difficult and could have unforeseen consequences. It also may not be possible without a significant change in the international system and the development of a governance system outside of states. Though citizen monitoring demonstrates that individuals can identify, analyze, and distribute treaty-relevant information even from afar, systematic voluntary reporting on one’s own state has not yet emerged.

Systematic and Society Wide: Informal Networks versus Global Organization

Societal verification was envisioned as an organized and systematic activity, coordinated through a global verification agency. Citizen monitoring is conducted informally, coordinated through a patchwork of individuals and organizations. Without a dedicated organization to manage data flow, citizen monitoring falls short of the societal verification standard. Citizen

---

monitoring may miss important information coming out of states violating treaty obligations. Only information deemed interesting by experts and nongovernmental organizations circulates through the citizen monitoring community. As a result, some indicators of treaty violations may not be identified and shared. A systematic approach to open-source data management would be able to mitigate that risk, though no system will be able to perfectly screen all information available. False information may also be picked up and circulated as credible. Though citizen monitoring has proven its ability to correct some errors, it is not clear that all false information is caught. Without careful data authentication by an over-arching organization, data coming out of citizen monitoring may not be credible, hampering its ability to contribute to arms control verification.

Citizen monitoring lacks the broad scope and participation of societal verification. Because citizen monitoring is informal in nature, only areas which are of interest to those participating in monitoring are covered. This is evident today in the number of blogs and discussions focused on Iran and North Korea, while other regions of the world are not discussed. Societal verification, however, would apply a similar level of attention to all parties of an arms control regime. Citizen monitoring is also limited participation. Many organizations and individuals who conduct citizen monitoring are from North America or Europe. There is a dearth of monitors reporting out of other areas of the world, potentially both biasing information and leaving dangerous gaps in monitoring activities. For societal verification to be truly systematic and society wide, the degree of scrutiny and reporting out of any state should be similar to all other states. Without systematic and society wide reporting, citizen monitoring has serious gaps. To be used in a treaty, societal verification will need to bridge those gaps and assure participating governments that its verification system is effective and reliable.

*Treaty-based: Ad-Hoc versus Legal Foundation*

Citizen monitoring is different from societal verification in part because it exists as an *ad-hoc* activity and lacks any legal basis which societal verification would possess. Because it is conducted by citizens outside of any formal treaty, information from citizen monitoring has no standing in inter-state relations. Without a legal basis, citizen monitoring is limited to commenting on observed behavior, activities, and events. While it is unclear to what degree citizen monitoring provides new information or simply duplicates information available through other national means, it is clear that in the future, societal verification would need to be recognized by all governments as a valid component of an arms control agreement. In order to be included in formal treaty negotiations, all parties would need to believe that societal verification provided meaningful and accurate information. Providing trusted information will be an important and necessary condition for societal verification, but citizen monitoring has clearly not yet crossed that threshold.

*Societal Verification, Citizen Monitoring, and Future Arms Control*

A world without nuclear weapons will require a strong verification regime, stronger than any negotiated to date. Societal verification has the potential to contribute to strengthened verification regimes. Initially developed to support the first arms control agreements and to prove that arms control was possible, today societal verification is being re-examined in discussions about the complete elimination of nuclear weapons. A form of societal verification has emerged in the development of citizen monitoring capabilities. There are still significant
gaps to overcome, however, before societal verification could be applied in the context of a treaty.

Some of these gaps may be impossible to bridge, in part because of the connotation of the term “verification.” Within an arms control context, verification implies not just collection and analysis of information but also judgments of state compliance with treaty obligations. Treaties are complex legal agreements that require extensive knowledge of treaty obligations to assess and evaluate implementation. It is unrealistic to expect that the general public will understand arms control treaties well enough to provide a meaningful compliance assessment that could be used to confront a state government. Additionally, any allegation of non-compliance will require very strong evidence, and it is extremely unlikely that states will accept the judgment of the general public.

Despite the semantic problems with “societal verification,” the term encapsulates useful ideas regarding public involvement in arms control efforts in the future. As demonstrated with the study of citizen monitoring, at least some portion of the public has the potential to provide valuable information about possible treaty violations. “Societal monitoring” could be a more appropriate term, recognizing that publicly available and collected information could be incorporated into formal verification regimes in the future to serve as supplemental information to already existing national verification activities.

There is significant potential for the public to contribute to arms control. Establishing a credible method for collecting, analyzing, and disseminating information from the public on treaty implementation and overcoming current limitations within citizen monitoring will require further study. Understanding how the public could contribute to future arms control verification could be an important component of efforts to establish a world without nuclear weapons.
The Loch Ness Monster, Bigfoot, and Safeguards Conclusions

What Crowdsourcing and Pervasive Mobile Sensors Tell us about Proving a Negative

Karl Horak
Sandia National Laboratories, Albuquerque, NM, USA

ABSTRACT
With over 5 billion cellphones in a world of 7 billion inhabitants, mobile phones are the most quickly adopted consumer technology in the history of the world. In advanced Western countries, over 91% of the population own cellphones and over 56% have smartphones. Miniaturized, power-efficient sensors, especially video-capable cameras, are becoming extremely widespread with the additional availability of wearable technology like GoPro, Google Glass, and lifelogging systems. The operators of these mobile systems form a large population of networked individuals that are spontaneously crowdsourcing the creation of enormous sets of visual and other data. Seemingly rare and nearly impossible to document events are now routinely captured, for example, the Chelyabinsk meteor, the “miracle on the Hudson” ditching, and the Boston Marathon explosions. By now one would have expected, say the skeptics, to have high quality, definitive video evidence of the Loch Ness Monster and Bigfoot. Have cellphone cameras proved the negative, that these crypto-zoological beasts do not exist? Can the same technology help us with safeguards conclusions? In this paper the author will explore recent trends in mobile and wearable technology, look at projected data volumes, examine some open-source tools to analyze these data streams, and discuss the implications for nuclear safeguards. Dr. Horak will recount his experience of using Google Glass for the past six months and consider what the future of pervasive mobile technology may hold for international treaties and agreements.

TRENDS IN MOBILE TECHNOLOGY
By the end of 2013 there were more mobile devices on earth than the 7 billion plus people. More than 2 billion mobile devices will ship this year. Those without cell phones have shrunk from 17% in May 2011 to 12% in February 2012 to 9% in May 2013, making mobile phones the most quickly adopted consumer technology in history [1].

This places small video-capable cameras in the hands of 91% or about 6.4 billion people worldwide. Combine that with the fact that all those cameras are each sitting atop a convenient Internet telecommunications device and one realizes that we are creating enormous web-accessible visual data sets.

From this vast network of embedded cameras, we are routinely seeing very rare events recorded and distributed within minutes.

1Sandia National Laboratories is a multi program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND 2014-3770C.
Chelyabinsk Meteor (still from video) | US Airways Flight 1549 | Boston Marathon Bombing
---|---|---
![Figure 1](image)
![Figure 2](image)
![Figure 3](image)
Within 65 hours over 100 million viewers had seen videos of the bolide [2].
Ditched at 3:27, this photo was uploaded to Flickr at 4:23 PM EST [3].
Cellphone photo taken within a few minutes of the explosions [4].

Table 1. Examples of rare events captured with mobile devices

With such a large and widely distributed array of camera sensors, we should have outstanding photographic and video imagery of such rare phenomena as the Loch Ness Monster, Bigfoot, and other cryptozoological beasts. The result has found its way into popular online humor in the form of the online comic XKCD [5].

![Figure 4](image)

Figure 4. XKCD online comic from December 2013

Besides smartphones and tablets (and even the ambiguously named “phablets”), there are now “wearable” computing devices. Whether watches like Pebble, rugged miniature video cameras like GoPro, or headsets like Google Glass, these small devices bring a new dimension to portable computing.

**DATA VOLUMES**

Image sites like Flickr, Instagram and Pinterest are adding billions of photographs every year. As of June 2012, every minute Flickr users were adding 3,125 new photos, Instagram users shared 3,600 new photos, and YouTube users uploaded 48 hours(!) of new video. Twitter users were sending over
100,000 tweets per minute while Facebook users were sharing 684,478 pieces of content. Mobile data traffic will reach the following milestones within the next five years according to statistics from Cisco [6].

- Monthly global mobile data traffic will surpass 15 exabytes by 2018.
- The number of mobile-connected devices will exceed the world’s population by 2014.
- The average mobile connection speed will surpass 2 Mbps by 2016.
- Due to increased usage on smartphones, smartphones will reach 66 percent of mobile data traffic by 2018.
- Monthly mobile tablet traffic will surpass 2.5 exabyte per month by 2018.
- Tablets will exceed 15 percent of global mobile data traffic by 2016.
- 4G traffic will be more than half of the total mobile traffic by 2018.
- There will be more traffic offloaded from cellular networks (on to Wi-Fi) than remain on cellular networks by 2018.

The number of mobile-connected devices currently exceeds the world’s population. Cisco, the network hardware company, estimates that by 2018 monthly global data traffic will surpass 15 exabytes. Network traffic from tablets alone will be greater than 2.5 exabytes by then. That is about 18% of the total.

The question becomes one of whether or not there is a needle of safeguards-relevant information in this enormous digital data haystack. With these enormous numbers of devices throughout the world, it seems likely that either intentionally or inadvertently, someone or some group somewhere will capture an image or share a message or collect data that at least indirectly has a bearing on safeguards.

Safeguards-relevant information may become published through the intentional actions of whistleblowers, WikiLeaks, hackers, disaffected employees, or as a result of analysis by NGOs or private individuals. Disclosure may be inadvertent in the case of citizen radiation sensors detecting an accidental release, of a facility incident being photographed from off-site, or of the unintended publication of what should have been controlled sensitive information. However, there could be misinformation on the Web as well, so determination of what information is trustworthy can be difficult.

Ultimately, we will not be able to prove the negative, that no proscribed nuclear activities are taking place, just because mobile technologies with increasingly powerful capabilities are becoming universally widespread. However, these technologies and the data that they are generating can help increase one’s confidence in safeguards conclusions.

**OPEN-SOURCE TOOLS**

SQL has long been the mainstay of database management. In recent years alternatives to RDBMS have begun appearing that are particularly well suited to handling the large volumes of public data. Being able to quickly and efficiently manage network information is useful for safeguards applications because many problems are in fact network analyses. Acquisition path analysis, proliferation networks, social network applications, transportation routing networks, process flow models, many geospatial problems, and some types of semantic networks all can benefit from the use of a graph database architecture.
A recent ranking of graph database engines places these five as the top contenders.

<table>
<thead>
<tr>
<th>Neo4j [7]</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Titan [8]</td>
<td></td>
</tr>
<tr>
<td>OrientDB [9]</td>
<td></td>
</tr>
<tr>
<td>Sparksee [10]</td>
<td></td>
</tr>
<tr>
<td>Giraph [11]</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2. Graph Database Management Systems**

There is considerable turnover in the popularity of these systems; only Neo4j and OrientDB have stayed in this Top 5 list for the past year.

Besides databases designed for storing the large data volumes involved, one needs specialized tools for analysis and visualization of these sorts of large and complex data sets. While an exhaustive listing is not possible, here are two suggestions.

Gephi [12] is an open-source graphing tool capable of handling very large data sets and generating 3D networks. DataMaps [13] is a jQuery plugin that provides data visualizations based on geographical data. Out of the box it can produce choropleths (color-coded areas) and bubble maps, among others.

**IMPLICATIONS FOR SAFEGUARDS**

IAEA Safeguards continues to monitor open-source channels of publicly generated information and publishes daily summaries. With the anticipated growth of these data streams, other strategies will be required. Some possibilities that suggest themselves include various kinds of crowdsourcing:

- **Microwork.** Small tasks, low payouts, for example, Mechanical Turk [14].
- **Macrowork.** Fixed tasks requiring specialized skills. A type of outsourcing where experts contribute as they have time. Examples include systems that pay experts to develop online training material.
- **Crowdsourced challenges.** Web-based inducement prize competitions, for example, InnoCentive [15] and DARPA’s Red Balloon Challenge [16].
- **Implicit crowdsourcing.** Other activities that have a useful side effect, for example, online games like FOLD-IT [17], voluntary systems like Zooniverse [18], and human-recognition systems like reCAPTCHA [19].

The difference between the xkcd view of the world and safeguards is that facilities under safeguards do not generally allow cellphone camera access to sensitive areas. This means that there is a significant gap in coverage and that these “ubiquitous” mobile devices aren’t getting 100% coverage.

While we have made a case for the use of public information generated from the rapid expansion of mobile Internet and communications technologies, these same devices have the capability to be used by Safeguards inspectors and analysts for their Safeguards-related tasks. Within Safeguards, there have been identified a number of needs [20], many of which can be addressed entirely or in part with mobile solutions:

- Miniaturization: bulky -> smart and portable
- High level of integration of sensors and instruments
- Robust devices with multipurpose detection functions
• Secure and wireless transmission
• Enhanced mobility for the inspector
• Ability to receive early feedback from HQ

However, one must keep in mind that most safeguarded facilities will not allow the entry and use of mobile computers with their cameras and wireless communications. Rethinking and renegotiating facility agreements with these technologies in mind will be necessary if they are to be used on site.

GOOGLE GLASS
Since November, 2013 the author has been what’s known as a Google Glass Explorer, basically a beta tester of Google’s entry into the wearable computer market. Reaction to Glass has ranged from “This is the coolest thing ever” to “It’s just a webcam you wear on your face.” In general, the technology has been enthusiastically regarded and people often are eager to try them on.

However, there are a number of concerns that wearable technology are bringing to the fore.

• Expense
• Usability
• Still in beta
• Apps aren’t there... yet
• Other wearable devices
• Cultural factors
• Privacy/sensitivity concerns
• Do wearables make sense?

The primary challenge is that no one has arrived at a truly new use case. Simply moving from the smartphone to smartglasses or smartwatches gives us “less of everything” except hands-free. That said, I have found in my own experience that the convenience of a wearable webcam allows me to take photographs of scenes where I never would have bothered to or had time to pull out a camera or smartphone. That in turn will lead to a further burgeoning of online photo collections and possibly result in the very occasional capture of a safeguards-relevant image or video.

These devices will become less obtrusive, more useful, and more common in the everyday world as they mature. Already wrist computers like Pebble blur the line between digital watch and smartphone. At the point where they become as innocuous and pervasive as Bluetooth-enabled hearing aids, one may well have to accept them as just another unremarkable, utilitarian item much as we deal with cellphones today.

CONCLUSIONS
In an increasingly connected world, Internet and communications technologies continue to exhibit extreme growth and with that come exponential growth in online data. At the same time, computing platforms are diversifying and miniaturizing, creating even more opportunities and challenges. But in the final analysis, pervasive mobile communications devices with high quality cameras and other built-in sensors will only add to our confidence in safeguards conclusions.

New software tools, many of them open-source, are becoming available to harvest, organize, and analyze this burgeoning volume of data. Novel methods like crowdsourcing may help reduce the burden...
on safeguards inspectors and analysts. Mobile devices may provide inspectors with useful capabilities if hurdles to implementation and other challenges can be overcome.

ACKNOWLEDGEMENTS
The author wishes to thank Andrew Steele for the original invitation to become a Google Glass Explorer.

REFERENCES
Tessera: Open source software for accelerated data science

Landon H. Sego, Ryan Hufen, Hannah Director, Ryan LaMothe
Pacific Northwest National Laboratory

Abstract
Extracting useful, actionable information from data can be a formidable challenge for the safeguards, nonproliferation, and arms control verification communities. Data scientists are often on the front lines of making sense of complex and large datasets. They require flexible tools that make it easy to rapidly reformat large datasets, interactively explore and visualize data, develop statistical algorithms, and validate their approaches—and they need to perform these activities with minimal lines of code. Existing commercial software solutions often lack extensibility and the flexibility required to address the nuances of the demanding and dynamic environments where data scientists work. To address this need, we developed Tessera, an open source software suite designed to enable data scientists to interactively perform their craft at the terabyte scale. Tessera automatically manages the complicated tasks of distributed storage and computation, empowering data scientists to do what they do best: tackling critical research and mission objectives by deriving insight from data. We illustrate the use of Tessera with an example analysis of computer network data.

Introduction
Obtaining useful, actionable information from the data explosion is a grand challenge for industry, academia, and government. The impetus to leverage the deluge of data has led to notable advances in hardware, networks, and software. Many data challenges, be they large or small, are sufficiently complex to require a data scientist—a technical expert trained in statistics, applied mathematics, and computer science.

Tessera is a software tool developed by data scientists, for data scientists, to enable them to work more efficiently, develop better algorithms, and produce accurate analyses of large data. Tessera is general-purpose and domain-agnostic. It can be applied to virtually any kind of data. This naturally includes, but is not limited to, datasets that arise in the safeguards, nonproliferation, and arms control verification communities.

Tessera consists of four open source software tools: DataDR (Divide and Recombine), Trelliscope, SQM (Signature Quality Metrics), and LIFT (Laboratory Integration Framework and Toolset). DataDR, Trelliscope, and SQM are written in the statistical programming language R (R Core Team 2014), while LIFT, written primarily in Java, is an enterprise service bus architecture that facilitates the integration of the three R packages with data sources, third-party software tools (e.g., SAS, Weka, Python), and user interfaces (e.g., Taverna, Tableau) into a unified workflow.

While Tessera is designed to run on parallel and distributed architectures that use the MapReduce framework, it is not constrained to a particular hardware platform, data architecture, or operating system. It can be used on a single desktop, a small inexpensive cluster, or a high-end data center.

Large, complex data require deep analysis. While descriptive statistics (e.g., mean, median, standard deviation) and tabulations are indispensable in exploring and summarizing data, deriving insight from
large datasets often requires extensive exploratory data analysis and visualization, coupled with the development of statistical or machine-learning models that explain and predict complex behavior. This process is iterative and interactive; it involves data explorations, visualization, hypothesizing and fitting statistical models, and validating those models.

**Divide and Recombine Methodology**

Tessera uses the *divide and recombine* technique to compute on large data, wherein a dataset is divided into subsets, a computation is applied to each subset, and the results of the computations are recombined in some fashion. Recombination can take a variety of forms: a statistical computation, an analytic recombination (i.e., staging for another round of computation), or visual displays for one or more of the subsets. The divide and recombine approach is illustrated in Figure 1.

A simple example of a statistical recombination would be the calculation of a global mean for a variable in a large dataset. This would be accomplished by first calculating the mean and the number of data points for each subset and then recombining these results by calculating the weighted average of the subset means to produce the overall mean.

To illustrate an analytic recombination, suppose five replicate measurements were taken for each of several million subjects and that each subject belonged to one of 500 possible groups. If the goal were to create a regression model for each group, we might wish to first divide the data by subjects and calculate the average for each subject and then recombine the subject averages according to groups. This process would constitute an analytic recombination in preparation for the final recombination step: fitting a separate regression model for each of the 500 groups.

Continuing the example, a visual recombination could involve creating a separate plot of the fitted regression models with the corresponding data points (the subject averages) for each group.

![Figure 1: Divide and Recombine methodology](image)
Many analyses of large datasets lend themselves to the divide and recombine approach, as there are often logical subgroupings that can be used to divide the data. These subgroupings may be defined by geographic region, units of time, entities, subjects, etc. If variables that define logical subgroups do not exist, or if it does not make sense to divide the data, then random subsets of the data may also be constructed. Random subgroups lend themselves naturally to large data bootstrapping approaches (Kleiner et al. 2012), which can be used to fit close approximations of statistical models to large datasets without digesting all the data at once.

In the R community, thousands of statistical, machine-learning, and data-processing methods have been developed for many types of problems and domain areas (CRAN 2014). Most of these methods were not developed for large-scale computation. They work quite well on small datasets but are difficult (or impossible) to apply to large datasets. However, the divide and recombine approach provides a natural way to employ these unscaled computing methods to large data, provided each subset is not excessively large. This ability to employ unscaled methods is one of the compelling, practical benefits of the divide and recombine methodology.

**Tessera Components**

We now discuss in greater detail the four components of Tessera: DataDR, Trelliscope, SQM, and LIFT.

**DataDR**

DataDR is an engine for parallelized, distributed computing with big datasets. It extends and simplifies the widely employed Hadoop Map-Reduce framework by providing an intuitive programming interface that requires simpler code (in R) rather than traditional Hadoop implementations (which use Java). It was designed specifically to facilitate principles of common statistical practice (Hafen 2014, Guha et al. 2012, Guha et al. 2009, Hafen et al. 2013 *Power*, Hafen and Critchlow 2013, Guha 2013). DataDR uses the R package RHIPe to interface with Hadoop (Guha 2013). While RHIPe requires that data scientists cast their algorithms in terms of map and reduce steps, DataDR allows them to use a simpler interface designed to facilitate data processing and analytic tasks commonly used by R programmers. Using the DataDR and RHIPe pairing, several lines of R code are translated into dozens, even hundreds of lines of Java MapReduce code required by Hadoop. The relative simplicity of DataDR provides an environment where data scientists can focus on the nuances of their analysis instead of wrestling with the complexities of computing on large, distributed datasets.

DataDR is not wedded to Hadoop: it can be integrated with any distributed data architecture that uses MapReduce. It also includes its own MapReduce engine that operates on single nodes or desktops, using multi-core R to distribute the computation while storing data on the native file system. This local MapReduce engine makes it easy to develop DataDR code on a small scale before deploying it on a larger system, and it allows new users to learn DataDR and enjoy its benefits, even if they do not have access to big-data infrastructure.

**Trelliscope**

While DataDR provides methods for data division, computational recombinations, and analytic recombinations, Trelliscope was developed to support visual recombinations and to rapidly create and view customized graphics for each subset. In many divide and recombine analyses, it is common for the division to result in many thousands of subsets (Hafen et al. 2013 *Trelliscope*). Trelliscope provides a way to rapidly render the plots and then filter the results to focus on examples that share common traits or characteristics (e.g., display only plots with correlation above 95%). Trelliscope can render any type
of plot that R or R packages are capable of producing. It derives its name from trellis displays that lay out individual plots, or panels, in a row-column format that facilitates visual comparisons between groups (Cleveland 1993, Sarkar 2008). Trellis displays guide scientists to focus on relatively small, rational subsets of data. This approach helps them develop statistical models that account for the scientific phenomenology that gives rise to the data—an essential requirement for rigorous analysis. An example of plots produced by Trelliscope is shown in Figure 2.

![Figure 2: Unusual events in electrical power grid data, visualized in Trelliscope](image)

A natural approach to viewing many panels is simply scrolling through all the panels that have been tiled across a screen. Although this alone can be powerful, it becomes unrealistic when the number of panels is large. In our own applications, the number of subsets range from thousands to millions. Our approach to viewing a large number of panels in a meaningful way is to compute metrics on the data being plotted in each panel that identify interesting qualitative or quantitative attributes, and then sample, filter, or sort the panels based on these metrics. Here, we loosely refer to any such metric as a cognostic, a term coined by John Tukey for “computer guiding diagnostics” (Cleveland 1998).

What constitutes an “interesting attribute” of a subset? In most cases, this is left up to the analyst to determine for the visualization at hand. A cognostic metric is simply the result of any computation that exploits or explains some behavior in the data being visualized. Examples are simple statistical metrics such as the mean, standard deviation, range, quantiles, and number of observations. When a model fit is being plotted, a measure of the goodness-of-fit for the model for each subset can also be a meaningful cognostic. Other examples are the percentage of values exceeding some limit, number of times a specific value is seen, difference between means of two groups, etc.

**Signature Quality Metrics**

Once exploratory data analysis and the development of statistical models have taken place, an essential next step is to validate the performance of the models. Often, the purpose of these models is to predict, characterize, or detect a phenomenon of interest, i.e., to discover or detect a signature of interest (Baker et al. 2013). However, signature discovery and detection is more than statistical model fitting; it requires developing and employing the encompassing signature system that includes the choice of
observables, measurement techniques, data processing, exploratory data analysis, feature extraction, and statistical model fitting. Naturally, every signature discovery effort requires a transparent approach for evaluating the quality of the resultant signature system.

SQM provides a mathematical framework for assessing the quality of signature systems, accounting for tradeoffs between system performance and other attributes of interest, like cost. To that end, we developed a formalism based on decision theory (Berger 1985) and multi-attribute decision science (Edwards et al. 2007, Keeney & Raiffa 1976) for measuring the quality of a signature system in terms of fidelity, risk, cost, other attributes, and ultimately, utility, which we define below:

- **Fidelity** refers to how well the signature system detects, predicts, or characterizes the phenomenon of interest. It includes metrics such as sensitivity, specificity, positive predictive value, and mean squared error, to name a few (Sokolova & Lapalme 2009, Hand 1997, Japkowicz & Shah 2011).
- **Risk** refers to the assessment of likelihoods and consequences associated with decision errors that may arise when employing the signature system.
- **Cost** refers to the resources expended to develop, deploy, and/or utilize the signature system. Examples include acquisition and operational costs of hardware, consumables, labor, etc.
- **Other attributes** include any other criteria or factors that distinguish one signature system from another and that are not already accounted for by fidelity, risk, or cost. Examples include human safety, ease of use, system portability, policy considerations, the risk associated with employing the signature system, etc.

These four components provide a heuristic for guiding investigators in their assessment of signature systems. Naturally, only a subset of them may be relevant for a particular assessment. The mathematical details of SQM are available in Sego et al. (2013), with additional examples provided by Watkins et al. (2013) and Nobles et al. (2013).

**Laboratory Integration Framework and Toolset**

Multidisciplinary research efforts increasingly involve complex data management and analysis tasks. These tasks and workflows are often developed using a variety of database types and programming languages. The influence and effectiveness of statistical modeling in these environments often depends on the extent to which statistical codes can be easily integrated into a larger analysis pipeline or signature system. LIFT (SDI 2014) addresses this challenge by encapsulating, orchestrating, and exposing analytics and data as cloud services.

LIFT enables the development and execution of complex data analysis and signature discovery workflows, including 1) loading and transforming heterogeneous data sources, 2) facilitating feature extraction, exploratory data analysis, and statistical modeling, and 3) validating algorithms and statistical models. All of this can be done using any combination of database and analytic tools, such as MySQL, NetCDF, NoSQL, Weka, R, SAS, OpenCV, Python, Java, etc. LIFT can also link storage and computation across heterogeneous hardware architectures, including desktops, commodity clusters, data appliances, and high-performance computing systems. LIFT contains the following high-level features:

- **Extensibility**: The framework allows the addition of analytical components, such as machine-learning and statistical algorithms, as services. The integration of data, signature discovery services, and workflows requires support for a composable architecture that can accommodate
a diversity of integration methods such as SOAP-based web services, RESTful APIs, as well as scripts and executables written in any language for any hardware platform.

- **Reusability**: Client applications must be able to re-use existing components and workflows developed by others. As a result, a managed user library exists that allows users to store and retrieve services and workflows that support a variety of tasks.

- **Scalability and Performance**: The framework supports the coordination of data-intensive and high-performance computational workflows as well as high-throughput event-based pipelines while maintaining performance and response time.

Virtually any user interface can be employed to control LIFT, including interpreted languages (e.g., R, Python), workflow tools (e.g., Taverna), or customized applications. A hypothetical workflow using LIFT is shown in Figure 3.

![Figure 3: Hypothetical example of a LIFT workflow](image)

**Example: Analysis of the Visual Analytics Science and Technology conference challenge data**

To illustrate the utility of Tessera, we applied the tools to analyze synthetic network security data publicly available from the IEEE Conference on Visual Analytics Science and Technology (VAST) Challenge (VAST Challenge 2013). The data consist of various types of computer network summaries over a two-week period for an enterprise of about 1800 computers. The data consist of network health metrics, intrusion protection system data, and NetFlow records. For most of our analysis, we focused on NetFlow data.

NetFlow data contain summaries of connections between computers on a network with information such as the source and destination IP addresses and ports, the number of packets sent and received in the transaction, the number of bytes that went back and forth, the duration of the connection, and the internet protocol used (such as TCP and UDP).

To better understand these data, we performed a preliminary exploratory analysis, making heavy use of visualization throughout.

**Extract, transform, and load (ETL) data and initial data exploration**

One of the inevitably tedious prerequisites of data analysis is transforming the data into the proper format. DataDR aspires to provide functionality to make this process as painless as possible. All of the VAST Challenge data were made available as delimited text files. Using DataDR, only a few lines of code
were required to go from the raw text data to more useful and efficient R data objects, creating distributed data frames from each of the input datasets.

While the analysis of large, complex data is best served by the ability and willingness to look at the data in detail, high-level summary statistics are often a good place to start. DataDR automatically provides summary information for the variables in our distributed data frame, so summary analysis is straightforward.

We began our analysis by studying the number of times the internal and external host IP addresses occur in the data, broken down by the type of the machine and whether the host was the first-seen source or destination of the connection. This approach helps us gain a high-level understanding of host behavior.

![Figure 4: Quantile plots, which plot the frequency of the ordered data for each category, provide a visual representation of the cumulative distribution of the frequency of connections](image)

In Figure 4, in the far right panel, we see that workstations are more often the first-seen source IP than they are the first-seen destination IP. This implies that, for these data, workstations typically initiate network connections rather than receiving them. In the same panel, we also observe a group of source workstations that initiate at least ten times more connections than the remainder. These may be remote desktop connections, a conjecture we investigate further below. Another notable set of outliers occurs in the fourth panel, where the four outlying hosts for HTTP source and destination have orders of magnitude more activity than the rest of the corresponding distribution of source and destination hosts. These summary plots help guide us to areas of the data we will want to look at in more detail.

Data quality, while perhaps mundane, is an essential part of drawing defensible conclusions from data. Our initial summaries and graphics led us to find a problem with the data; meta-data for some 100 internal hosts were not provided. We were able to verify and fix this mistake. The errors in the VAST
data are not unique: it has been our experience that Trelliscope visualizations often lead to identifying bad data values that, otherwise, go undetected.

There are many other interesting exploratory plots we made that provided additional insights and directions for further analysis, but for the sake of brevity, we omit these results.

**Investigating outlying HTTP hosts**

To further investigate the outlying HTTP hosts that we saw in Figure 4, we can use DataDR to create a subset of the data corresponding to these hosts. We are interested in how the activity of these outlying hosts varies over time. We can issue a simple command in DataDR to tabulate counts of activity of these hosts for each minute. Figure 5 shows a plot of the frequency of connections for each outlying host across time, where it is obvious that there are two time points where all hosts are experiencing an abnormally high amount of traffic. This is indicative of a coordinated denial of service attack, where one or more external hosts are flooding the web servers with traffic.

**Figure 5: Plots of frequency over time allow the scientist to quickly identify time periods with unusual activity**

We can now create a subset of the data corresponding to these significant time periods and investigate the traffic in detail for all hosts. Looking at a one-minute period of high-intensity traffic, using
Trelliscope, we quickly find nine external hosts responsible for nearly all of the traffic. Over 200,000 connections are attempted from these hosts during a one-minute period. Figure 6 shows the cumulative number of connections over the course of this one-minute period for each of these nine hosts.

![Graph showing cumulative number of connections over time for nine hosts.](image)

**Figure 6:** This panel display of the nine hosts makes it easy to see a consistent pattern of numerous connections taking place simultaneously for these nine hosts, suggesting a machine-coordinated denial of service attack.

We can use the knowledge we have gained about this type of behavior to build models and algorithms to detect such behavior in the future. This apparent attack was easily detected via initial looks at the data. Having accounted for it, we can now look for more subtle behavior.

**Investigating remote desktop connections**

Another way to assess the network activity is to examine the remote desktop connections (RDCs). With Tessera, we are not limited to working with a few variables at a time. Rather, we can easily analyze the relationships among groups of variables. In this case, we are interested in the connectivity relationships among the computers on the network and the information being transferred between them. Using Trelliscope, we can build a graph to visualize the RDCs among the computers by representing each computer as a node and the lines between them as the connections.
Figure 7 is a panel display of four such representations during four arbitrary one-minute intervals. Grey indicates a possible connection, yellow indicates an attempt to connect was made but no payload (i.e. data) was exchanged, and red indicates a payload was exchanged.

In Figure 7, we notice that some internal workstations are receiving a high number of RDCs initiated by external computers during some of the one-minute intervals. This may be indicative of an attack, so we would like to further investigate the time periods where this is occurring. However, there are hundreds of one-minute intervals, so it would be time consuming to manually inspect each graph.
Instead, we take advantage of the sorting functionality in Trelliscope, which enables us to quickly identify the hosts that are receiving the largest amount of data from external RDCs. Figure 8 shows the interactive screen scientists can use to sort and select a small subset of the thousands, or even millions, of panels.

Figure 8: Trelliscope graphical user interface for sorting panel displays using cognostics

In this case, we use the pictures of the distributions of the data for each variable on the bottom of the screen to identify variables of interest. Noting that nPayloads, a measure of the data transferred during the one-minute intervals, is very skewed, we elect to sort the data by this variable. This correspondingly changes the ordering of the graphs, so that now when we go to view the graphs, the first six visualizations in the panel display correspond to the time intervals with the highest payload, as seen in Figure 9.

We also see in Figure 8 that the variables, mostActive, which gives the IP address of the most active computer during the one-minute interval, and nConn, the number of connections made in the interval, also have non-uniform distributions. Therefore, these variables may also provide some information to help identify a security threat, so we do not want to ignore them, even though we have not elected to sort based on them. Fortunately, we can easily retain and display their information by selecting these variables as Visible Cognostics. This places a label below each graph that gives the value for mostActive and nConn. We also elect to display the time interval, time10, so that we know when these events are occurring. The resulting panel display is given in Figure 9.

Looking at the graphs for the top six payloads in Figure 9, we notice there is always one green node, indicating an external computer, with a large number of red arrows pointing out from it toward the other nodes in the network, indicating the delivery of a large number of payloads from that external computer. This shows that in all six intervals, the source of the extremely high number of payloads is a
single external source. Using the labels we retained, we also see that the IP address 10.138.235.111 is the most active IP address during two of these periods. This is unusual. It is unlikely that there would be a legitimate reason for an external computer to be making large numbers of payloads in a one-minute time period. It is particularly suspicious that some of these payloads were delivered from the same computer.

Figure 9: Visualization of the time intervals with the six highest payloads. During each period, an external computer, indicated by a green node, is making many RDCs that deliver payloads to internal host computers, indicated by red arrows

Discussion

Using Trelliscope, we quickly determined with high confidence the existence of an apparent coordinated denial of service attack. We identified the times of the attack and the IP addresses from which the attack originated. We also identified numerous possible RDC intrusion attempts, along with their originating IP addresses and when they occurred.

It is also worth noting that the visualization tools in Trelliscope facilitate multiple analysis paths by which a user may identify phenomena of interest, such as possible attacks. Stated differently, peculiarities in the data that are indicative of attacks may be discovered by a variety of means, and the “correct” starting point is not unique. Consider if we had instead started our analysis by comparing relationships between two variables. In Figure 10 we use Trelliscope’s bivariate filter tool to plot maxAttempt, the maximum number of RDC attempts made during the one-minute time intervals, and nConn, the number
of connections in the intervals. Here, we quickly notice strange behavior in that there are a large number of remote connections that make exactly 800 attempts. This behavior suggests that the connections have been automated with the intent to infiltrate the network. With Trelliscope, we can now further investigate these unusual events simply by drawing a box around the data of interest and clicking the Apply button, which immediately gives us specific information on the time periods and IP addresses from which these suspicious connections were made, enabling us to characterize the suspected RDC attack.

**Figure 10:** Scatter plot of the number of RDCs vs. the maximum number of RDC attempts during one-minute intervals. A cluster of suspicious outliers is indicated by the grey box

In short, Trelliscope allows us to quickly identify subtle patterns—patterns that would easily be overlooked—by combining numerical filtering with visual inspection. This capability enables scientists to quickly visualize and characterize data, providing a rigorous basis for the development and validation of statistical models, machine-learning algorithms, or predictive analytic models. These models, in turn, can be used in systems that serve broader missions, be that cyber security, arms control verification, radiation detection, etc.

**Conclusion**

The suite of R packages in Tessera (DataDR, Trelliscope, and SQM) enables data scientists to effectively perform the deep analysis necessary to address big-data challenges, while maintaining a programming environment that closely resembles familiar desktop computing. The LIFT component of Tessera provides a mechanism for managing multiple static or streaming data sources, incorporating algorithms
written in other programming languages, and deploying a unified workflow in a cloud-based architecture.

In our example, we used Tessera to identify potential cyber attacks in the 2013 VAST challenge data by exploratory data analysis and visualization. This example illustrates that large and complicated data cannot be fully understood through summary statistics alone; rather, an iterative process of exploratory analysis and visualization is required. Furthermore, exploratory analysis and visualization should be conducted before attempting to fit statistical models or applying analytic algorithms to the data. This approach is essential, because statistical models and algorithms are only as good as their fit to the data—and each dataset is unique and requires individual assessment.

While we have emphasized the data scientist as the principal user of Tessera, it is not our intent to dissuade others from using the Tessera tools. While moderate proficiency in R is admittedly a prerequisite, we advocate that Tessera be used in pairs, where domain scientist and data scientist examine the data together, drawing upon their separate expertise and expanding the knowledge they share in common. The requirement for R programming has one notable exception: the web-based interface of Trelliscope, which we designed to be intuitive and easy to learn. After providing assistance by writing the snippets of R code that govern the plots produced by Trelliscope, we have observed domain experts unfamiliar with R successfully derive unexpected insights from their data using the Trelliscope web interface.

The ability to extract and carefully analyze large volumes of complex data is a continual need in numerous domains, including the safeguards, nonproliferation, and arms control verification communities. To maximize the value of data and the information they contain, data scientists need to easily manipulate massive data sets, rapidly visualize data in a variety of ways, and confirm their results. All of these needs to be done with minimal coding, so that data scientists may focus on their area of expertise: identifying, analyzing, and interpreting subtle patterns in complicated data. Tessera, an open source software suite, addresses this critical need, providing a valuable tool for data scientists not only to make better sense of large and complex datasets but also to do it more rapidly.

Resources

Additional information, source code, tutorials, and documentation about DataDR and Trelliscope can be found at www.tesseradata.org. Additional details of the analysis (including code examples) of the VAST challenge using DataDR and Trelliscope are available at http://tesseradata.github.io/example-vast-challenge.

Acknowledgments

This work was supported by the Science and Technology Directorate of the U.S. Department of Homeland Security, the XDATA program of the U.S. Department of Defense Advanced Research Projects Agency (DARPA), and the Signature Discovery and Future Power Grid Initiatives at Pacific Northwest National Laboratory (PNNL), conducted under the Laboratory Directed Research and Development Program at PNNL, a multi-program national laboratory operated by Battelle for the U.S. Department of Energy.

We would like to recognize the contributions of our colleagues at PNNL who comprise the Tessera team: Anthony Bladek, Robert Brigantic, Wendy Cowley, Darren Curtis, Christine Noonan, Ralph Perko, and Amanda White. In addition, we thank Sharon Eaton for peer review and Ted Tannasse for the graphics.
shown in Figures 1 and 3. Likewise, we thank our collaborators William S. Cleveland and graduate students at Purdue University for their contributions to RHIPE, DataDR, and Trelliscope, and Saptarshi Guha for his original development and maintenance of RHIPE.

References


http://cran.case.edu.


GPU Computing for Nonproliferation and Arms Control Applications

Michael McDaniel, Heidi Smartt, Maikael Thomas
Sandia National Laboratories

Paper not received at time of publication.
Machine Learning for Classification and Visualisation of Radioactive Substances for Nuclear Forensics

Andrew Jones*
The University of Liverpool, Department of Electrical Engineering and Electronics,
Brownlow Hill, Liverpool, L69 3GJ
a.jones3@liverpool.ac.uk

Phillip Turner
AWE, Aldermaston, Reading, Berkshire, RG7 4PR

Colin Zimmerman
The UK National Nuclear Laboratory, The Central Laboratory,
Sellafield, Seascale, Cumbria, CA20 1PG

J. Y. Goulermas
The University of Liverpool, Department of Electrical Engineering and Electronics,
Brownlow Hill, Liverpool, L69 3GJ
j.y.goulermas@liverpool.ac.uk

Abstract – The IAEA (International Atomic Energy Agency) figures from 2012 show that there have been in excess of 2,331 reported cases of trafficking of radioactive materials worldwide[1]. The hazardous nature of this material means that trafficking of such substances is cause for concern and should be combatted appropriately. Nuclear forensics is a field that has emerged as an anti-proliferation solution to this problem by extracting forensic information from radioactive materials to answer questions about the source of such unknown materials. Traditionally the elemental and isotopic composition of the material can be analyzed to determine attributes of the materials history using known characteristics such as isotopic ratios and trace elements. Through novel implementation of machine learning and pattern recognition techniques it is envisaged that spectral analysis will become faster and more accurate. Rather than using the ratios of specific isotopes and specific elemental compositions for analysis we have shown that machine learning allows the user to make use of the full extent of elemental and isotopic data available. These techniques have further been able to provide an effective set of tools for finding valuable insight into the provenance of a particular sample utilizing the posterior probability distribution of samples from a given reference dataset. Our work has focused on finding novel and effective machine learning techniques to aid the nuclear forensics process by providing visualisation of the complex datasets that result from spectral analysis of radioactive samples as well as generating classification models for determining attributes such as reactor design or the age of a given sample. A number of classification and visualisation techniques have been successfully tested on a set of datasets that are representative of a number of different radioactive substances from the nuclear fuel lifecycle. While each of these substances have their own challenges to address in terms of source determination, our work has successfully addressed the need to provide a well-informed analysis of the nuances of these datasets to effectively and accurately utilise the characteristics of the materials to make an informed decision regarding the origin of the substance. For instance, a representative dataset of post-irradiation samples have been used to demonstrate an effective classification model for determining the reactor design from which a sample may have originated. This analysis has been further supported by dimensional reduction techniques to generate visualisation of the dataset such that further conclusions can be drawn from the spread of the data. Results show that these techniques are very much capable of classifying spent reactor fuel with a level of accuracy that would certainly aid the
decision making process with regards to determining the source of unknown substances in the fight against nuclear smuggling.

1 Introduction

The aim of this work will be to explore the use of machine learning techniques with an implementation towards nuclear forensics. Nuclear forensics attempts to determine the origin of unknown nuclear materials based on analysis of the substance composition itself. In doing so, this will provide a tool for quickly and efficiently determining details about such substances to help facilitate origin determination[2]-[4]. Through novel implementation of machine learning and pattern recognition techniques[5] it is envisaged that analysis will become faster and more accurate. Ultimately this will reduce the proliferation of nuclear materials by making it possible to trace the history of the material based on purely the composition of the substance.

There are a number of attributes that can be used to differentiate uranium compositions, notably these include the isotopic signatures of the plutonium and uranium isotopes, nuclear decay products, anionic and metallic impurities and the morphology of the material itself[3]. Measurement of each of these characteristics allows us to infer details about the history of the material and what its intended use may have been. Therefore it is very possible to determine the origin of such a material as well as answering other forensic questions. The focus of this work will be concerned with novel application of machine learning and pattern recognition techniques targeted at answering questions about a substance based on the isotopic composition. The most abundant form of nuclear material comes from the nuclear fuel cycle, which encompasses materials that are used for civil nuclear power plants throughout the world. Specifically we will be focusing the spent nuclear fuel resulting from the irradiation of fuel rods in civil nuclear reactors. The isotopic measurements from these materials have been used as the basis for this investigation. These measurements have been investigated using two techniques. Firstly, dimensional reduction methods have been employed to facilitate visualization of the high-dimensional dataset and aid qualitative analysis of the distribution of the data. Further to this, the data has been quantitatively assessed using a classification technique called random forest. Random forest has been used to classify the materials with respect to a number of different attributes of the fuel sample.

Here we present the results from using these techniques in a case-study scenario. The visualization and classification tools have been used to determine some useful forensic attributes of a selection of real-world measures taken from samples of the Spent Fuel Composition (SFCOMPO) database. These samples have been classified and visually compared with a synthetic dataset representing a number of different spent fuel materials. The synthetic samples have been generated at the National Nuclear Lab (NNL) using the FISPIN depletion code.

Initially an investigation has been performed to determine the reactor design from which a sample may have originated. Secondly the initial U-235 enrichment of the fuel has been determined and finally the burnup that the fuel has undergone in the reactor. Each of these attributes could be very helpful in determining a smaller sub-set of reactors from which an unknown sample may have originated and thus be very helpful in an investigation. Results show that the visualization and classification methods that have been employed here are well suited to answering these questions with good accuracy and would certainly for a good basis for forensic investigation of an unknown sample.
2 The Dataset

A synthetic dataset has been generated using the FISsion Product Inventory (FISPIN) depletion code. FISPIN has been used to generate the expected composition of the spent reactor fuel from 5 reactor types; Advanced Gas Reactors (AGR), Boling Water Reactors (BWR), CANada Deuterium Uranium (CANDU), Magnesium Non-Oxidising (Magnox) and Pressurised Water Reactors (PWR). As shown by Table 1 the samples included in this dataset represent spent reactor fuel from a number of different conditions including burnup, cooling time, rating and U-235 enrichment. The FISPIN package uses a WIMS (Winfrith Improved Multigroup Scheme) 2D reactor model with JEF (Joint Evaluated File) 2.2 nuclear data to calculate the representative data for our investigation. FISPIN has been well validate and can been seen as representing a good average composition of expected materials[6]. Calculations are based on decay, neutron capture and fission for the reactor fuel using simultaneous differential equations. The dataset does not represent the variance that would be expected within a fuel rod, but is representative of the average expected composition. There are 3,682 samples in the dataset, which represents spent fuel compositions from nuclear fuel with respect to different enrichment levels, ratings, irradiations and cooling periods for the 5 different reactor types. Each of the samples assumes the same input fuel composition with the exception of the U-235 and U-234 isotopes. As shown in Table 1 a variety of U-235 enrichments are represented. Furthermore, there is a slight variance in the U-234 isotope due to incidental enrichment during the U-235 enrichment process. This has been calculated using proprietary ratio information provided by Westinghouse.
Table 1. Breakdown of the different samples represented in the synthetic FISPIN dataset, which has been used for classification and visualization comparison with the measured SF COMPO samples.

<table>
<thead>
<tr>
<th>Reactor Type</th>
<th>U-235 Enrichment (%U-235)</th>
<th>Cooling Times</th>
<th>Burnup (MWD/te)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced Gas Reactor (AGR)</td>
<td>1, 1.5, 2, 2.5, 3, 3.5, 4</td>
<td>1, 5, 10, 15, 20, 30, 40 years</td>
<td>1000, 5000, 10000, 15000, 20000, 25000, 30000, 35000, 40000, 45000</td>
</tr>
<tr>
<td>Boiling Water Reactor (BWR)</td>
<td>2, 2.5, 3, 3.5, 4</td>
<td>1, 5, 10, 15, 20, 30, 40 years</td>
<td>5000, 10000, 15000, 20000, 25000, 30000, 35000, 40000, 45000, 50000, 55000</td>
</tr>
<tr>
<td>Canada Deuterium Uranium (CANDU)</td>
<td>0.711, 1.2</td>
<td>0, 1, 2, 10, 15, 20, 30, 40 years</td>
<td>5000, 10000, 15000, 20000, 25000, 30000</td>
</tr>
<tr>
<td>Magnesium Non-Oxidizing (Magnox)</td>
<td>0.711</td>
<td>0, 16, 90, 115, 280, 365 days</td>
<td>500, 1,000, 2,000, 3,000, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 1,000, 12,000, 13,000, 14,000, 15,000</td>
</tr>
<tr>
<td>Pressurized Water Reactor (PWR)</td>
<td>2, 2.5, 3, 3.5, 4, 4.5</td>
<td>1, 5, 10, 15, 20, 30, 40 years</td>
<td>5000, 10000, 15000, 20000, 25000, 30000, 35000, 40000, 45000, 50000, 55000</td>
</tr>
</tbody>
</table>

Our dataset includes samples representing 5 reactor types; Advance Gas Reactor (AGR), Boiling Water Reactor (BWR), Canadian Deuterium (CANDU), Pressurized Water Reactor (PWR) and Magnox. BWR and PWR reactors are similar in concept and they both use light water as a coolant and enriched uranium dioxide fuel. In a BWR reactor the water is heated until it boils and turns to steam, which in turn drives the turbines, whereas in a PWR the coolant is pressurized to prevent boiling and steam is produced via a heat exchanger which drives turbines external to the reactor safety containment. HWR reactors are characterized by the fact that they use heavy water as a moderator. Heavy water is less prone to neutron absorption than light water and therefore the fuel in these reactors does not, in general, require enrichment to sustain fission. Magnox reactor designs use a pressurized carbon dioxide...
coolant with graphite moderation together with natural enrichment metallic uranium fuel and derive their name from the fuel cladding, which is an alloy of magnesium and aluminium. AGR reactors are based on the gas cooled Magnox design. Instead of using Magnox cladding, stainless steel is used to allow the reactor to operate at higher temperatures and use enriched uranium dioxide fuel. As shown in Table 1, a selection of different samples from each of these reactor designs are represented in synthetic dataset, each representing the range of sample types that would typically be found in a real-world scenario.

In this case study we classify a number of spent fuel samples that have not been generated using the FISPIN depletion code. This exercise will help to determine whether the FISPIN data will be capable of classifying real measured materials using the machine learning techniques that have been employed here. For this exercise, the samples have been taken from the openly available SFCOMPO database[7], [8]. Nine samples have been taken from the SFCOMPO database all originating from the Calvert Cliffs-1 reactor, a PWR reactor originating in USA. The isotopic composition was taken in the Material Characterization Center (MCC) of Pacific Northwest Laboratory (PNL). Nine samples are included in the dataset, representing three fuel assemblies. Table 2 shows the actual U-235 enrichment and the burnup for each of the samples. The samples have been numbered in Table 2 and these numbers correspond to those represented in the visual assessment in Section 4.1 and the classification assessment in Section 4.2.

Table 2. Breakdown of the PWR Calvert Cliffs-1 samples showing initial U-235 enrichment and Burnup.

<table>
<thead>
<tr>
<th>Sample Number</th>
<th>Sample Name</th>
<th>U-235 Enrichment (%U-235)</th>
<th>Burnup (GWd/t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>USACC1PWR-1</td>
<td>3.038</td>
<td>27.35</td>
</tr>
<tr>
<td>2</td>
<td>USACC1PWR-2</td>
<td>3.038</td>
<td>37.12</td>
</tr>
<tr>
<td>3</td>
<td>USACC1PWR-3</td>
<td>3.038</td>
<td>44.34</td>
</tr>
<tr>
<td>4</td>
<td>USACC1PWR-4</td>
<td>2.72</td>
<td>18.68</td>
</tr>
<tr>
<td>5</td>
<td>USACC1PWR-5</td>
<td>2.72</td>
<td>26.62</td>
</tr>
<tr>
<td>6</td>
<td>USACC1PWR-6</td>
<td>2.72</td>
<td>33.17</td>
</tr>
<tr>
<td>7</td>
<td>USACC1PWR-7</td>
<td>2.453</td>
<td>31.4</td>
</tr>
<tr>
<td>8</td>
<td>USACC1PWR-8</td>
<td>2.453</td>
<td>37.27</td>
</tr>
<tr>
<td>9</td>
<td>USACC1PWR-9</td>
<td>2.453</td>
<td>46.46</td>
</tr>
</tbody>
</table>

For both the FISPIN and the SFCOMPO datasets the same measurement have been taken for each of the samples and these measurements have formed the basis of the following investigation. Measurements are represented in kg/MTU. Briefly, these comprise of Pu-238, Pu-239, Pu-241, Pu-242, Am-241, U-234, U-235, U-238, Tc-99, Cs-137 and Np-237.
3 Methods

3.1 Visualisation Methods

For visual assessment of the samples it is first necessary to reduce the original 11 features that represent the 11 isotopes to a lower dimensional space of 2 or 3 dimensions such that it can be visually represented. The majority of the dimensional reduction work that has been carried out for this investigation has been done using the Laplacian Eigenmaps[9] (LE) method. This is an unsupervised method that works by converting the \( n \times d \) data matrix (in the case \( n = 3682 \) and \( d = 11 \)) to an \( n \times n \) similarity matrix \( W \) where \( w_{ij} \) represents the similarity between the \( i^{th} \) and \( j^{th} \) rows of \( X \). The data is then projected non-linearly to a lower dimensional space of \( k \) dimensions. This is achieved by minimising the sum of the weighted pairwise distances between all \( n \) embeddings, expressed as:

\[
\min_{Z \in \mathbb{R}^{n \times k}} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \| z_i - z_j \|_2^2 .
\] (1)

The embedding matrix \( Z \) is an accurate representation of the original \( n \times d \) feature matrix \( X \). This can be solved with an eigen-decomposition of the Laplacian matrix of the graph defined as \( L = D - W \), where \( D \) is the diagonal matrix composed of the row sums of \( W \). LE optimizes Eq.(1) with the constraint that \( Z^T D Z = I_{k \times k} \), which imposes a further scaling of the embedding space to keep the dimensions separate.

Further to using LE, some supervised dimensional reduction has been carried out to determine how well the different reactor types in the dataset can be distinguished. This has been done using Fisher Discriminant Analysis[10], [11] (FDA). Unlike the LE method previously discussed, this is a supervised method that uses the class labels to aid visual representation. The aim of this transformation is not only dimensional reduction but also to maximize the distance between samples of different classes and to minimize the distance between samples of the same class. Thus the generated embedding is not only compacted into a lower dimensional space, but also represents the underlying discrimination task. This relies on using the class labels, in this case the reactor type of all the samples. FDA maximizes the following objective:

\[
J(W) = \frac{W^T S_B W}{W^T S_W W}
\] (2)

Where \( S_B \) is the between scatter matrix and \( S_W \) is the within class scatter matrix. FDA is the most commonly used linear supervised dimensional reduction technique for single label classification problems.

3.2 Random Forest Classification/ Regression

The random forest classifier is an ensemble learning methods that is based on the concept of constructing a number of prediction trees and aggregating the results form each tree to find a decision[12]. The dataset is bootstrapped into a number of random sample groups and each of these sample groups is used to train a prediction tree. In this way each of the prediction trees can be considered independent of one another, as they have been generated using independent sub-sets from the complete dataset. To find the class for a given sample, each of the prediction trees assess the sample individually and each will decide which class it belongs to. The class of a given sample will then ultimately be determined by aggregating the results from all of the prediction trees and the mode of the complete forest is taken as the final predicted class. This technique has been proven to perform well against other classifiers and is robust against overfitting[13]. For the purpose of this case study, the FIPSIN samples will be used to train to random forest ensemble and the Calvert Cliffs-1 samples will then
be classified using these models. Thus the synthetic FISPIN dataset will be used to classify the actual measured samples from the Calvert Cliffs-1 data.

When implementing the random forest classifier it is necessary to select the number of prediction trees that will be in the forest. Therefore a number of different tree numbers have been implemented and the out-of-the-bag error rate assessed for each to determine which number of trees has the best results for this classification problem. Figure 1 shows the error rates for the trees ranging from 1 to 100 for the classification for samples according the reactor design. As can be seen the error rate drops to a low level between 15 and 20 trees and remains at low from there. Therefore the implementation of random forest in this work has used 20 decision trees for the classification model. Similar results were observed for the initial U-235 and burnup regression models and therefore all implementations of random forest in this study have used 20 prediction trees.

![Figure 1. Classification error for different number of trees in the random forest when classifying reactor design.](image)

### 4 Results

#### 4.1 Visual Assessment

Figure 2 shows the visualisation results from using the FDA dimensional reduction method discussed in Section 3.1. Only the FISPIN samples have been included in this visualisation work as we rely on the class labels for visualisation and the Calvert Cliffs-1 samples are assumed to be unknown samples in the case study and therefore would not have class labels available. As can be seen FDA has been reasonably successful at separating the different reactor types in this 2-dimensional representation. However, there are some notable classes where separation has not been entirely successful. There is some significant overlapping between the BWR and PWR sample groups, suggesting that it would be difficult to distinguish between these two reactor types using the 11 isotopic measurements represented in this dataset. Overlapping between these reactor types has been noted in similar studies[3], there appears to be some intrinsic similarity between the spent fuel materials for these reactor types which makes them difficult to distinguish. Furthermore, there is some overlapping between the AGR and BWR samples, however this is not as significant and may not present as much of a problem during classification. The Magnox and AGR samples are well distinguished and it would be expected that these samples would be classified with a good degree of accuracy.
Figure 2. Fisher Discriminant Analysis (FDA) using a Gaussian W matrix to illustrate the distribution of the different sample groups with regard to the 5 different reactor types of the FISPIN dataset

Figure 3 shows the unsupervised LE visualisation of the FISPIN samples and the Calvert Cliffs-1 samples as discussed in Section 3.1. As can be seen, the samples from the different reactor types in the FISPIN dataset are well separated. As with Figure 2, some of the BWR and PWR samples are closely distributed within this 3D visualisation, but for the most part they are well separated. The CANDU and Magnox samples are closely grouped with one another in the far section of the X component of the plot. As can be seen in Table 1, the samples from these two reactor designs have only low enriched samples. Therefore it is hypothesised that the grouping of these two samples is caused by their low U-235 enrichment.

The Calvert Cliffs-1 samples in Figure 3 have been plotted with their associated numbers from Table 2. Clavert Cliff-1 is a PWR reactor type and therefore it would be expected that these samples would be close to the PWR samples represented in the FISPIN dataset. As can be seen, the samples do correlate well with the PWR samples of the FISPIN dataset. There are some notable samples that are very close to the BWR sample group, suggesting that there may be some difficulty to distinguish between these two reactor types. As noted in Figure 2, PWR and BWR are the most difficult reactor types to distinguish and this certainly seems to be the case for the Calvert Cliffs-1 samples shown here. Samples 9, 8 and 7 appear to be particularly difficult to distinguish from the BWR sample group.
The same LE embedding of the original 11 isotopes are shown in Figure 4 as are in Figures 2 and 5. In Figure 4, the FISPIN samples have been grouped according to the initial U-235 enrichment. Based on this visualisation it is clear that there is a distinct pattern in the distribution of the FISPIN samples with regards to the U-235 enrichment. The Calvert Cliffs-1 samples have been plotted into this embedding along with the FISPIN samples. It is clear to see that the samples are within an area of the plot that would be consistent with an initial U-235 enrichment of 2% to 3%. As can be seen in Table 2, there are three distinct U-235 enrichments represented by the Calvert Cliffs-1 samples. These three different enrichments are subtly visible in the visual representation of Figure 4. Samples 1, 2 and 3 are slightly further back in the plot and closer to the higher enriched samples in the FIPSIN reference dataset. These three samples are indeed the highest enriched samples from Calvert Cliffs-1 at 3.038% U-235. The next three samples (4, 5 and 6) are slightly lower enriched at 2.72. Finally samples 7, 8 and 9 are the lowest enriched at 2.453, this is consistent with these visual results. As can be seen in Figure 4 they are closer to the lower enriched samples from the FIPSIN reference dataset.
Figure 4. 3-Dimensional projection of the 11 feature dataset generated using the Lapasien Eigenmap (LE) showing the different enrichment values for each samples.

Figure 5 shows the same LE embedding of the samples as is shown in Figures 3 and 4, but this time the samples from the FISPIN dataset have been labelled to show burnup. The burnup is well represented by the X and Y components and therefore the Z component has been removed. More so than with the reactor type and the initial U-235 enrichment there is a very clear pattern in these samples with respect to the burnup of the reference FISPIN dataset. Therefore this visual embedding of the dataset should be very useful for determining the burnup of the Calvert Cliffs-1 samples. The Calvert Cliffs-1 samples are distributed throughout the different burnup values shown by the reference FISPIN dataset and as shown in Table 2 there are a range of different burnup values represented by the Calvert Cliffs-1 data. Based on this visual representation of the data it appears that samples 4 and 5 have the lowest burnup values while samples 3 and 9 would conversely have the highest. As shown by Table 2, this is in fact the case. The samples range between 46.36 GWd/t and 18.68 GWd/t and all he samples fall within this region in this visual representation.
Figure 5. 3-Dimensional projection of the 11 feature dataset generated using the Lapasien Eigenmap (LE) showing the different burnup values for each samples

4.2 Classification Assessment

Table 3 shows the results from using the random forest classification method for determining the reactor type that each of the samples originated from as detailed in Section 3.2. The Calvert Cliffs-1 reactor is a PWR reactor and therefore it is expected that the sample should be classified as PWR samples. However, as noted in Section 4.1 there is expected to be some miss-classification between the PWR and BWR samples based on the visual assessment of the FISPIN data. As can be seen in Table 3 the majority of the samples have been correctly classified as originating from a PWR reactor type. However samples 6, 8 and 9 have been incorrectly classified as originating from a BWR reactor. The miss-classification of samples 8 and 9 is consistent with the visual results shown in Figure 3, however sample 6 does not seem to be as clearly belonging to the BWR sample group in the visualisation results. While these results aren’t all correct, it has been noted that PWR samples would be difficult to classify based on the visual assessment in Section 4.1 and therefore other reactor types are expected to provide more reliable results.
Table 3. Results from using the random forest classification method for determining the reactor type of the Calvert Cliffs-1 samples.

<table>
<thead>
<tr>
<th>Sample Number</th>
<th>Expected Value</th>
<th>Random Forest Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PWR</td>
<td>PWR</td>
</tr>
<tr>
<td>2</td>
<td>PWR</td>
<td>PWR</td>
</tr>
<tr>
<td>3</td>
<td>PWR</td>
<td>PWR</td>
</tr>
<tr>
<td>4</td>
<td>PWR</td>
<td>PWR</td>
</tr>
<tr>
<td>5</td>
<td>PWR</td>
<td>PWR</td>
</tr>
<tr>
<td>6</td>
<td>PWR</td>
<td>BWR</td>
</tr>
<tr>
<td>7</td>
<td>PWR</td>
<td>PWR</td>
</tr>
<tr>
<td>8</td>
<td>PWR</td>
<td>BWR</td>
</tr>
<tr>
<td>9</td>
<td>PWR</td>
<td>BWR</td>
</tr>
</tbody>
</table>

Table 4. Results from using random forest regression to determine the initial U-235 enrichment of the 9 Calvert Cliffs-1 samples.

<table>
<thead>
<tr>
<th>Sample Number</th>
<th>Expected Value (%U-235)</th>
<th>Random Forest Result (%U-235)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.038</td>
<td>2.938</td>
</tr>
<tr>
<td>2</td>
<td>3.038</td>
<td>2.571</td>
</tr>
<tr>
<td>3</td>
<td>3.038</td>
<td>2.557</td>
</tr>
<tr>
<td>4</td>
<td>2.72</td>
<td>2.679</td>
</tr>
<tr>
<td>5</td>
<td>2.72</td>
<td>2.750</td>
</tr>
<tr>
<td>6</td>
<td>2.72</td>
<td>2.472</td>
</tr>
<tr>
<td>7</td>
<td>2.453</td>
<td>2.531</td>
</tr>
<tr>
<td>8</td>
<td>2.453</td>
<td>2.479</td>
</tr>
<tr>
<td>9</td>
<td>2.453</td>
<td>2.467</td>
</tr>
</tbody>
</table>

The results from using the random forest technique for determining the initial U-235 enrichment of the Calvert Cliffs-1 samples are shown in Table 4. There are three different enrichment levels represented and the regression results have had good success in determining these. Random
forest regression has been able to determine that the samples have been enriched between 2.938% U-235 and 2.467% U-235. The worst result seen in these results is sample 3 that is 0.481% off from the expected value. However, notably sample 9 has been determined within 0.014% U-235 enrichment of the actual value. Interestingly, sample 9 was incorrectly classified as a BWR reactor type in Table 4 but has been the most accurate sample when determining the initial U-235 enrichment. It is possible that determination of these two characteristics may be dependent on different features from the initial 11 isotopes, thus explaining this unusual result. The three different enrichment levels aren’t clearly represented in the results, for instance sample 5 has the second highest enrichment from the regression results and it does not have the highest expected enrichment. While not completely accurate, these results would certainly be a useful guide to forensic analysis and based on this study would be able to achieve a result that is within 0.481% U-235 enrichment of the sample in question.

**Table 5.** Results from random forest regression to determine the burnup of the 9 Calvert Cliffs-1 samples.

<table>
<thead>
<tr>
<th>Sample Number</th>
<th>Expected Value (GWd/t)</th>
<th>Random Forest Result (GWd/t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>27.35</td>
<td>26.45</td>
</tr>
<tr>
<td>2</td>
<td>37.12</td>
<td>37.00</td>
</tr>
<tr>
<td>3</td>
<td>44.34</td>
<td>38.85</td>
</tr>
<tr>
<td>4</td>
<td>18.68</td>
<td>20.17</td>
</tr>
<tr>
<td>5</td>
<td>26.62</td>
<td>26.70</td>
</tr>
<tr>
<td>6</td>
<td>33.17</td>
<td>31.50</td>
</tr>
<tr>
<td>7</td>
<td>31.4</td>
<td>24.50</td>
</tr>
<tr>
<td>8</td>
<td>37.27</td>
<td>29.25</td>
</tr>
<tr>
<td>9</td>
<td>46.46</td>
<td>37.50</td>
</tr>
</tbody>
</table>

Finally Table 5 shows the results from using random forest to determine the burnup of the 9 Calvert Cliffs-1 samples. Sample 5 shows a particularly accurate result with only a 0.08 GWd/t deviation from the expected value. On the other hand sample 9 has the least accurate results from with an 8.96 GWd/t deviation from the expected value. Interestingly sample 9 was also incorrectly classified as a BWR reactor type. Perhaps this sample has a somewhat unusual composition. Further study will be required to identify what is causing these unusual results. On the whole these results are well representative of the expected values and would provide a good guide for forensic investigation of an actual unknown sample.

5 Conclusion

The machine learning methods that have been demonstrated in this paper has shown that there is a good deal of promise in using these techniques for assessing the origin of radioactive material. We have been able to successfully use a synthetic reference dataset to determine details about real-world
measured samples with a good level of accuracy. The details that have been determined in this study would form a sound basis to forensic analysis of such materials. While the classification results have proven to have good accuracy, they have not been perfect. Therefore these methods are seen as an aid to the existing methods and not a replacement for them.

This approach has the advantage of utilising the full extent of the data available through dimensional reduction techniques and classification methods, as opposed to traditional techniques that focus on particular isotopes of interest[2]. While this particular investigation has used 11 isotopes, it is very possible that more features could be added to these classification and visualisation methods to improve results. As more isotopes and measurements are considered, it would be advantageous to use feature selection methods to help determine which of the features are particularly useful for particular classification problems[14]. Another advantage of this approach is that no prior expert knowledge has been required for analysis, for instance parent/daughter ratios. Because of this, it is envisaged that this approach could form the basis for a toolbox that would facilitate fast evaluation of an unknown sample. While the results are not completely accurate they would form a good basis for an initial investigation of intercepted materials.

The machine learning techniques that have been employed here are based on sound mathematical reasoning and therefore the results are well representative of the composition of the materials. However, it should be noted that these approaches are heavily reliant on the reference data that has been used for comparison in the visual assessment and to train the classification models. In the case we have had good success using samples generated using the FISPIN depletion code. The scope of this study has been limited to only spent reactor fuel. There are many other materials that may be encountered outside of regulatory control and it will be necessary to determine details about the source of such materials. Therefore future work will be required focusing on utilising these methods to determine attributes about other radioactive materials such as Uranium Ore Concentrate (UOC).

6 References


The Future of Intelligent Systems for Safeguards, Nonproliferation, and Arms Control Verification

Kari Sentz, Ph.D. François Hemez, Ph.D.
Los Alamos National Laboratory

Abstract

For any major safeguards, nonproliferation, and arms control verification effort, there is a wealth of textual, technical, and image-based information that is so extensive and voluminous that only a small fraction of it can be used. Unfortunately, this can result in lost knowledge, unrealized capabilities, and wasteful redundancy. This gap can be narrowed by taking advantage of the state-of-the-art of intelligent systems to preprocess and digest information that might otherwise be lost. As intelligent systems are machine-learning systems that are designed to emulate some aspect of human intelligence, these hold a particular appeal as a partner to the human analyst. We explore the value these human-inspired intelligent systems can have for high volume and high velocity unstructured data in a hard human-centric problem such as safeguards, nonproliferation, and arms control verification. This type of problem presents many challenges that resist a pure machine-learning based approach and the state-of-the-art is not sufficient to fully meet these challenges in its current form. This galvanizes innovation in the form of new technological combinations and the introduction of new algorithms and capabilities. More specifically, in this research supported by the Department of State Verification Fund, we investigated the state-of-the-art of text mining, deep question answering systems, and automated image and video analytics for human-centric data problems. This combination of tools provides substantive analytic capabilities for unstructured data in both text and image forms and allows for the exploitation of new data sources. We discuss how these technologies could mine emerging data sets from open social media sources in a new way. We also identify important limitations resulting from the uncertainty in the information itself as well as the uncertainty in the mechanism for fusion. We propose one untapped meta-analytic resource that can help to account for these uncertainties in the final result.

Introduction

The wealth, breadth, and depth of information relevant to nuclear safeguards, arms control, and treaty verification easily overwhelms any specialized human resources that can be brought to bear. While it is clear that there are a number of state-of-the-art technologies that can help to mine and organize high volume different types of data, how these can be synthesized and extended for hard human centric problems is the first step to realizing this promise. In this paper we explore recent advances in the deep question answering technology and image analytics and offer a mechanism for carrying this type of information forward to the decision-making environment.
Deep QA

It is well recognized that the most impactful innovations of recent times involve the ability to access and meaningfully navigate an ever-expanding corpus of information. This revolution has unfolded with the critical tools of the computer, the internet, the world wide web, and the search engine. As the technology evolves and the data accumulates, the need for a more intelligent way of obtaining and synthesizing useful information from an immense corpus becomes clear. In particular, what is needed is a system that can mine and synthesize vast amounts of available information to provide accurate and meaningful answers to questions posed by human users. Such systems are called deep Question Answering (QA) systems. These are distinct from rudimentary search engines in that search engines find key words and provide the most statistically significant matches from the web. It is left to the user to determine the relevance of such findings. A deep QA system seeks to understand the relevance of information to a query, provide an answer based on this information, and communicate a level of confidence in possible answers to the query.

Over the past decade, deep QA has been a central research focus in academic and industrial circles drawing on multiple disciplines, such as, machine learning, computational linguistics, data science, and uncertainty quantification. Deep QA launched into the mainstream in 2011 with a high profile televised Jeopardy! contest between the most elite human players and a deep QA system Watson specialized for the game, developed by IBM and academic partners. As is well known, the Watson system was the undisputed winner of the contest demonstrating superiority in disambiguation of clues as well as the retrieving and synthesizing relevant information in the shortest time horizon. This success reflected the willingness of the IBM team to overhaul everything from the basic technical approach and the underlying architecture to the metrics and evaluation protocols. Their 2010 journal publication [Ferrucci et al. (2010)] details the overarching principles critical to their success, namely: massive parallelism, many types of expert systems, machine learning techniques, pervasive confidence estimation, and the integration of both shallow and deep knowledge. This is realized in the development of over 100 different algorithms for analyzing natural language, identifying sources, finding and generating hypotheses, finding and scoring evidence, merging hypotheses, and ranking hypotheses. These techniques were combined in such a way to leverage any overlap in approach for improvements in accuracy, confidence, or speed. Following this game show success, the outstanding question remains is how can the state-of-the-art of deep QA can be leveraged for complex and multi-modal data problems with more ambiguous information.
Figure 1: High-level view of deep QA based on [Ferrucci et al (2010, 2012)]

At its core, deep QA seeks to structure knowledge from unstructured information and be able to extract the relevant information in response to a query, qualified by the strength of evidence. In the context of a specific question in a particular domain, we can think of this enterprise as involving four key transformations:

1. Raw Content Development
2. Content to Data
3. Data to Knowledge
4. Questions to Answers

The first phase involves the iterative development of raw content to a high quality corpus which requires acquisition, evaluation, integration, and expansion of resources. For a new domain associated with a query, collections of text documents are acquired in an iterative process in an attempt to cover salient topics that are considered missing from the current corpus. A subset of the information is transformed into a representation that the system can easily use. Sources are expanded by adding new
information as well as syntactic and lexical variants. [Ferrucci et al. (2013)] The transformation of content to data involves the extraction of relevant information from the cultivated corpus of unstructured information and leveraging structured data resources such as databases, knowledge bases, and ontologies. Ultimately, this phase involves the structuring of heterogeneous content. The third phase seeks to extract knowledge by identifying entities and relations between them based on the data. Tools used for knowledge extraction can be manually created rule-based expert systems and automatic statistical approaches.

The final transformation seeks to provide an ordered set of potential answers with quantified confidences. These potential answers (hypotheses) are generated from the content using as many interpretations of the question as possible. (Figure 2)

![Deep QA Hypothesis Generation](image)

**Figure 2: Deep QA Hypothesis Generation**

To achieve this requires extracting and structuring knowledge in the form of different pieces of evidence that support a hypothesis as well as the degree of confidence that supports the hypothesis. These confidences are obtained from scoring functions that pervade all levels of the system. In this way the final system can be said to reason from data to the hypotheses (abductive reasoning). This idea in shown in the notional Figure 3 with multiple types of evidence (called evidential dimensions) in support of a single hypothesis. In principle, these evidential dimensions can be fused into a single confidence measure in support of the hypothesis as is shown notionally in Figure 3.
**Figure 3: From Structured Knowledge to Confidence Measure**

In deep QA system, we know that both of the steps shown in Figure 3 were the results of a multi-phase machine learning process using at the highest level a supervised learning algorithm. How to do this in the absence of an appropriate training set is a current open research problem.

The type and amount of data required for meta-learning for a hierarchical machine learning system is not available in the context of safeguards, arms control, and treaty verification. This inspires another level of domain adaptation where we seek to leverage as many deep QA advancements as possible in the form of search, storage, speed, semantics, etc. and expand to include new data sets. The reformulation of deep QA for arms control verification and compliance is depicted in Figure 4. Here the question answering portion is taken out entirely. Despite the QA moniker, question answering is more about the user interface than the underlying machinery of evidential reasoning that is so useful. Interpreting questions is a risk for the system and unnecessary if the user is an experienced analyst. The analyst's experience is an important source of knowledge and designing a system that makes best use of that in terms of defining the question and assessing the results. To bring the user more into the loop, the input to the system is an initial set of variables that the user would like to explore in terms of establishing connections between existing variables and the evidence that supports it. The system can use concept expansion techniques to identify new salient variables to include in the set of variables of interest. This alleviates the burden of the user to be able to identify all of the potentially relevant variables a priori.
Another important advancement for the re-conceived system is expanding the data sets to include images, video, scientific models, and surveillance data as well as remote sensing through other available media. The extended data types require expansion of the capabilities to deal with it:

- Imagery and video---based data requires an automated analytic capability.
- To be relevant to a more ambiguous domain such as arms control verification and compliance that can be characterized by misinformation, denial, and deception, more capabilities need to be introduced for anomaly detection. This in the context of big data is a significant challenge. If a specialized system were to be developed for this domain, it would be advantaged by specialized systems developed in conjunction with detection experts and inspectors that would define compliance and non-compliance. Another related topic is that of mining and handling sparse data in big data. Both of these related topics are important features of re-conceiving system for this domain and constitute a significant research and development effort.
- In addition, would be the development of specialized systems that can take in scientific information and data from engineered systems. While this topic is beyond the scope of the current discussion, it is one of considerable interest for future work towards integrating the

Figure 4: Reconceiving deep QA for the Treaty Verification and Arms Control Domain
successes of sensor fusion with the individual successes in text and image classification and extraction.

- The final piece of the puzzle seeks to address the absence of the appropriate data for the multilevel machine learning algorithm. In this case we introduce another form of meta---analysis to qualify the results. Here we seek to understand the vulnerability of findings to uncertainty. This can be another opportunity to bring the expert into the loop. If the results are highly sensitive to uncertainty, the expert can be informed and potentially leveraged as a resource.

Automated image and video analytics

In addition to deep QA, another game changing technology involves automated image and video analytics. With access to larger and larger data sets, coupled with increasing computational resources, image classification and recognition algorithms drastically improved. This is well exemplified by the recent accomplishments of researchers from Facebook in face recognition systems that combine a three dimensional model-based alignment with a deep learning neural network yields impressive statistics of accuracy (>90% on most experiments). [Tiagman et al. (2014)] There are also a number of exemplary academic contributions to this renaissance of image---based object classification that have been made available to the open source community. One leader of the field, Jürgen Schmidhuber from the Swiss research institute for artificial intelligence, Istituto Dalle Molle di Studi sull'Intelligenza Artificiale (IDSIA) has won the past 9 first prizes in international competitions in pattern recognition in video and image analytics. [http://www.idsia.ch/~juergen/vision.html] IDSIA offers a modular machine- learning library specializing in deep learning neural networks in Python called PyBrain [http://pybrain.org/]. Another important open source contribution comes from the Berkeley Vision and Learning center (http://bvlc.eecs.berkeley.edu/) developed by Yangqing Jia that has implemented the fastest convolutional neural network publically available able to process more than 40 million images a day. [http://caffe.berkeleyvision.org/] Importantly, this resurgence in supervised and unsupervised object classification in images allows for the potential of including automatically generated textual annotations of images that would searchable and able to be incorporated in larger deep QA type technologies just like any other textual resource.

Robust Decision Making

Taking the output of an extended deep QA system to decision-making in safeguards, arms control, and treaty verification requires the integration of many different kinds of information ranging from soft data from textual sources such as policy documents and inspection reports to hard scientific data from engineered systems such as radiation detectors and seismic and acoustic sensors to image-based data available from overhead and ground sources. Each type of data from each of these sources has different type of uncertainty associated with it. This uncertainty can be lost in the decision process. As a result an outcome to a decision can be changed even if there is the slightest uncertainty over the inputs. In this section, we explore the value of an information gap analysis for robust decision-making with an eye to
the potential role in automated evidence-based reasoning systems that the next generation of the deep QA technology represents.

To better motivate this idea on a high-level for safeguards, arms control, and treaty verification, we notionally explore a classification decision whether a nation state is proliferating nuclear weapons in Figure 5. We define a classification decision space with “Yes” and “No” as potential outcomes with a clear boundary between them. We suppose that there is a decision process that yields a nominal value, which we designate with a white “x”. We show this for two cases: Case 1 on the left is comfortably in the “No” decision outcome space; Case 2 on the right is very close to the decision boundary. Uncertainty is introduced by turning a dial that dilates red uncertainty bounds around the nominal values. If the red bounds remain in a particular decision outcome space, that translates to confidence in a decision under uncertainty. If the red uncertainty bounds cross a decision boundary, the decision would have low confidence because the outcome could change as a result of uncertainty. In the example, both classification decision outcomes would be the same based solely on the nominal values: The nation state is not engaging in nuclear proliferation. However, they are clearly very different in their vulnerability to uncertainty. In Case 1, the decision is robust. In Case 2, the decision can easily change.

Figure 5: While the decisions based on the nominal values (represented by the white “x”) are the same in these two cases, they are very different in their vulnerability to uncertainty. In Case 1, the decision outcome does not change while in Case 2, the decision outcome can change as the result of the uncertainty. The more the red uncertainty bounds cross the decision boundary, the more compromised the confidence in the decision.
The reconciliation of such systems with the complexity and subtlety of decision support requires specialized capabilities. One useful capability is furnished through the gap in information between what is known and what needs to be known to justify a decision. The informational disparity described above is more commonly known as an information gap or info-gap. The best way to motivate the idea of an info-gap in the context of a decision is: If you had perfect information, what information would change the decision? By answering this question, we identify the cruxes of a decision. Then the current state of information can be viewed in terms of its distance from the perfect information case where the outcome of the decision changes. This distance between the current state of information and perfect information is the information gap. An info-gap provides a way to characterize the robustness of a model to the uncertainty in the decision space. In other words, this type of analysis can identify the outcomes that are most immune to failure due to uncertainty. [Ben-Haim (2006)]

This idea originated by Yakov Ben-Haim has been used in wide variety of applications such as finance, conservation management, infectious disease detection, and structural dynamics. [Ben-Haim (2006); Regan et al. (2005); Troffaes, Gosling,(2013); Hemez, Ben-Haim (2004)] There are three components to an information gap analysis:
1. The process model
2. The performance requirement
3. The uncertainty model

The process model is the mathematical representation of the system or decision-making process. The performance requirement is assessed through a performance measure computed using the process model. The uncertainty model characterizes what is unknown about the parameters of the process model. [Regan et al. (2005)] An info-gap model constitutes an unbounded family of consonant sets of uncertain events. The membership in the set is controlled by a horizon-of-uncertainty parameter, \( \alpha \). So info-gap theory proposes a structure for the uncertainty space and characterizes the clustering of uncertain events in sets. Importantly, info-gap does not assume a particular measure function for these uncertain events. As such, an info-gap model can admit any mathematical representation of uncertainty such as probability distributions, random sets, imprecise probabilities, etc. so long as it is consistent with the specified process model and the horizon-of-uncertainty parameter, \( \alpha \). [Hemez, Ben-Haim (2004); Ben-Haim, Hemez (2012)] Info-gap provides an indicator of robustness to uncertainty rather than a measure of the uncertainty itself.

Example

We would like to provide an example of the info-gap analysis. Suppose we have four pieces of evidence (which we call a evidential profile following [Ferrucci et al (2013)]) in favor of two possible decision outcomes summarized in Table 1.
Table 1: Strength of Evidence

<table>
<thead>
<tr>
<th></th>
<th>Alternative 1</th>
<th>Alternative 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evidence I</td>
<td>0.4</td>
<td>-0.8</td>
</tr>
<tr>
<td>Evidence II</td>
<td>0.53</td>
<td>-0.027</td>
</tr>
<tr>
<td>Evidence III</td>
<td>0.04</td>
<td>0.347</td>
</tr>
<tr>
<td>Evidence IV</td>
<td>-0.013</td>
<td>-0.173</td>
</tr>
</tbody>
</table>

The most significant evidential dimensions are Evidence I and Evidence III. The fact that these confidence scores are non-probabilistic uncertainty measures does not pose a problem for an info-gap analysis. As info-gap is a meta-analytic tool for quantifying robustness to uncertainty, it is agnostic as to the method used to quantify the uncertainty itself.

Process Model

It is easy to show that decisions can change as a function of different process models. Suppose the following notional weights for two different weighted averages as summarized in Table 2.

Table 2: Weighting Vectors for Notional Decision Algorithms

<table>
<thead>
<tr>
<th></th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Weighting Vector 1</td>
</tr>
<tr>
<td>Evidence I</td>
<td>2</td>
</tr>
<tr>
<td>Evidence II</td>
<td>1</td>
</tr>
<tr>
<td>Evidence III</td>
<td>5</td>
</tr>
<tr>
<td>Evidence IV</td>
<td>1</td>
</tr>
</tbody>
</table>

Utilizing these weights, we can construct decision algorithms based on these weighted averages. The relative confidence based on these decision algorithms change as a function of the weighting on the evidence dimension, Evidence I. The consequence of this is a change in the decision from Alternative 2 to Alternative 1 based on the weight of a single parameter. For the purposes of this simple example, we will explore the sensitivity of the decision algorithm based on Weighting Vector 1.

Performance Function

We define the second component of the information gap analysis the performance function $Y$ in terms of weights and the four evidential dimensions:

$$Y = w_1\text{Evidence I} + w_2\text{Evidence II} + w_3\text{Evidence III} + w_4\text{Evidence IV} \quad (1)$$

written symbolically as:
\[ Y = w_1 X_1 + w_2 X_2 + w_3 X_3 + w_4 X_4 \]  \hspace{1cm} (2)

where the weights \( w_k \) are those listed in Table 3 only scaled for convenience that the weights sum to unity:

\[ \sum_{i=1}^{4} w_i = 1 \]  \hspace{1cm} (3)

The performance function is evaluated for the two decision alternatives, labeled with a subscript 1 for the Alternative 1 and 2 for the Alternative 2. A large value of the performance function indicates the strength of the evidence that supports the corresponding hypothesis. Then, a ratio between the two values of the performance function \( (Y_r) \) is calculated:

\[ Y_r = \frac{Y(\text{Alternative 1})}{Y(\text{Alternative 2})} \]  \hspace{1cm} (4)

A ratio \( Y_r \geq 1 \) greater than one indicates that there is more evidence in support of Alternative 2. Likewise, a value \( Y_r \) indicates that the evidence in favor of selecting the Alternative 1 for decision Algorithm A is stronger. The info-gap analysis of robustness presented next is applied to the ratio metric (Eqn 4).

**Uncertainty Model**

The final requirement of an info-gap analysis is to identify the uncertainty model that we define relative to the values of the evidential dimensions describes in Table 2. We consider these to be “true-but-unknown” values that can be from different the values listed in the table. Further, we assume that no additional information is available to determine where the “true-but-unknown” values may be located relative to these values (or stated equivalently that no information is available to describe the uncertainty of the values of the evidential dimensions). The uncertainty model, therefore, is defined as a model of fractional change. The “true” value of one of the four evidence dimensions, denoted as a variable \( X_k \) in Eqn 2, deviates from the nominal value \( X_k^{\text{Nominal}} \) listed in Table 2 up to a percentage \( \alpha \):

\[ -\alpha \leq \frac{X_k^{\text{Truth}} - X_k^{\text{Nominal}}}{X_k^{\text{Nominal}}} \leq +\alpha \]  \hspace{1cm} (5)

Eqn 5 defines \( \alpha \) as a fractional change of variable \( X_k \). It has no physical unit. Equally important, the magnitude of \( \alpha \) is unknown, which expresses that the “true” value of the evidence dimension variable \( X_k \) is unknown.

**Information Gap Analysis of Decision Algorithm**
The info-gap analysis of robustness searches for the minimum and maximum values of the ratio metric, \( Y_r \) (Eqn 4) given that the four evidence dimensions are allowed to vary away, and up to \( \alpha \), from the nominal values defined in Table 2. At a given level of robustness \( \alpha \), two optimization problems are solved to search for the minimum and maximum values of the ratio metric (Eqn 4) given that the four evidence dimension variables \( X_k \) can vary within the bounds defined in Eqn 5.

In the numerical implementation illustrated next, the value \( \alpha = 1 \) corresponds to \( \pm 50\% \) variation of the nominal values of Table 2. This choice is arbitrary and requires justification. This formulation of the info-gap analysis is an eight-dimensional optimization problem since the four evidence dimension variables \( X_k \) are optimized for the Alternative 1 and the same four variables are optimized a second time, but independently, for the second alternative (Alternative 2). This implementation implies the assumption that the way the "true" values deviate from the nominal values for Alternative 1 is uncorrelated from the way the "true" values deviate from the nominal values for Alternative 2.

The analysis of info-gap robustness is summarized in Figure 18 that shows the minimum and maximum values of the ratio metric (Eqn 4) for increasing levels of robustness \( \alpha \). The dotted line at \( Y_r = 1 \) separates the region where the evidence supports Alternative 1 from the region where the evidence supports Alternative 2. From the region defined by the maximum and minimum performance curves, it is clear that the decision algorithm favors the Alternative 2 up to a level of robustness of \( \alpha = 0.2 \) as both curves lie to the right of the \( Y_r = 1 \) dotted line. (Recall that the level \( \alpha = 0.2 \) corresponds to \( \pm 10\% \) variation of the values for the evidential dimensions). This means that the decision in favor of Alternative 2 as the decision is best supported by the evidence profile, and this decision remains unchanged even if the evidence dimensions are incorrect up to \( \pm 10\% \) away from the nominal values listed in Table 2. If it is plausible that the true values of the evidence dimensions can deviate from the nominal values by more than \( \pm 10\% \), or \( \alpha = 0.2 \), then the evidential support in favor of Alternative 2 is more questionable because the minimum ratio (indicated by the solid line in Figure 5) decreases below one. In practical terms, this means that confidence in the diagnostic decision distills to the question of confidence in the evidence used to support it is correct to \( \pm 10\% \).
To illustrate how much evidence there is to support one decision alternative versus the other one, at a given level of robustness $\alpha$, this can be quantified in terms of the relative percentages of support. The difference between the maximum and minimum values of the ratio metric (Eqn 4), represents the entire range of possibilities supported by the evidence profiles and the associated uncertainty. These maximum and minimum values are shown as the robustness or performance curves in Figure 7 (same as Figure 6 with additional annotations). A simple approach to quantifying relative support is to measure how much of this range falls below the critical value for the Performance Ratio, $Y_r = 1$ and how much is above it. The values that fall below $Y_r = 1$ are in favor of Alternative 1 while the values greater than one are in favor of Alternative 2. These values on either side of $Y_r = 1$ are shown on Figure 6 and summarized in Table 3 as a percentage of the total range.
Figure 7: Strength of evidence interpretation of robustness curves

Table 3: Strength of evidence in support of the two decision alternatives

<table>
<thead>
<tr>
<th>Level of Robustness $\alpha$</th>
<th>Minimum Percent of Perturbation</th>
<th>Maximum Percent of Perturbation</th>
<th>Strength of Evidence in support of Alternative 1</th>
<th>Strength of Evidence in support of Alternative 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>-0.0%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>0.2</td>
<td>-10.0%</td>
<td>10.0%</td>
<td>0.0%</td>
<td>100.0%</td>
</tr>
<tr>
<td>0.4</td>
<td>-20.0%</td>
<td>20.0%</td>
<td>16.86%</td>
<td>83.14%</td>
</tr>
<tr>
<td>0.6</td>
<td>-30.0%</td>
<td>30.0%</td>
<td>20.20%</td>
<td>79.80%</td>
</tr>
<tr>
<td>0.8</td>
<td>-40.0%</td>
<td>40.0%</td>
<td>19.56%</td>
<td>80.44%</td>
</tr>
<tr>
<td>1.0</td>
<td>-50.0%</td>
<td>50.0%</td>
<td>17.33%</td>
<td>82.67%</td>
</tr>
</tbody>
</table>
Figure 8: Strength of Evidence in support of Decision Alternative 1

Figure 8 illustrates the “strength” of evidence in support of Alternative 1, which is the fourth column of Table 3, as a function of the level of robustness $\alpha$ on the vertical axis. Clearly, there is no evidence in support of Alternative 1 even if the evidence dimensions deviate from their nominal values by $\pm 10\%$, which is $\alpha = 0.2$. Beyond this value of $\alpha$, there is more uncertainty in the resulting decision. This uncertainty peaks where the nominal values of the evidence dimensions are incorrect by up to $\pm 30\%$ ($\alpha = 0.6$), then there is a 20.20\% “chance”, at most, that the correct decision is Alternative 1.

Conclusion

The accomplishments of these fields are many and perform very well under specific conditions. Extending these to a more general capability for application in more complex domains requires focused and intense research and development along a number of trajectories. On a high level, the central topics can be broken down into the following:

1 This “chance” is not “chance” in the sense of a probability; it may be more correct to call this metric a “strength” of the evidence collected in favor of one decision alternative or another.
1. Uncertainty quantification and decision support for information poor environments
2. Multi-modal data fusion: Scientific, Text-based, Image-based, Knowledge-based
3. Anomaly detection in big data: addressing misinformation and disinformation
4. Intelligent scientific data analysis

At the core of the information problem is an uncertainty problem that affects the estimation of source reliability, strength of evidence, and confidence in hypotheses based on evidence. This is already a challenge but in consideration of misinformation and disinformation, this becomes formidable. For this and other reasons, anomaly detection in large data sets becomes important, especially in the context of safeguards, arms control, and treaty verification-type problem where a small and seemingly insignificant departure from expectation can be the most important piece of information. Combining different sources of information in the context of high uncertainty is critical and the meta-learning approach of Watson is unsupportable in these domains domain because of the lack of appropriate training data. However the idea of many models attempting fusion and selecting what combination of models is relevant to a given problem is a very useful idea. Moreover, these tough problems inspire a more integrated approach that optimally leverages both expert and computational intelligence.

The potential of intelligent systems is vast in almost any conceivable discipline. This is especially true for the problems we face in the interest of national and global security. For any major verification and compliance effort, there is a wealth of textual, technical, and image-based information that is so extensive and voluminous that only a small fraction of it can be used. Unfortunately, this can result in lost knowledge, unrealized capabilities, and wasteful redundancy. With the research initiated here, supported by the Department of State, we have re-conceived the state-of-the-art of deep question answering systems, conceptualized the integration of this with visual analytic capabilities, and discussed the combined capabilities leveraging new data sets. Important algorithmic gaps have been identified that will inspire and inform the next phase of research. One particular meta-analytic capability, information gap analysis, was explored to account for the vulnerability of a decision model to uncertainty and can be integrated into any approach used to close the technological gaps.

Acknowledgements

The Los Alamos team is grateful for the support of the Department of State, Verification fund and in particular the enthusiastic support and feedback of Dan Wurmser and Astrid Lewis. We would like to thank Morag Smith and Kory Budlong- Sylvester of the LANL Program Office for their facilitation of this work.

References


Ferrucci, D. et al. (2012) This is Watson. IBM Journal of Research and Development vol. 56, no. 3-4.


Enhancing Cyber Security through the Use of Testbeds

A. David McKinnon, Mary Sue Hoxie, Kit Conklin
Pacific Northwest National Laboratory
Richland, Washington, USA

Abstract
We are in an increasingly data driven world. This trend has crossed over into safeguards, nonproliferation, and arms control verification as governments and international organizations automate data collection. Sensors deployed in the field typically send data back to a data collection site via a variety of transmission methods, including wireless networks. The need for security in these systems is recognized, but security efforts have often been focused only on hardening individual components. A holistic approach is needed, one that takes into account component security, network security, data integrity, and availability. Insecure sensors may collect and transmit bad data; insecure networks are vulnerable to replay attacks, data corruption, and denial of service; insecure databases are vulnerable to data corruption and the manipulation of data presented to analytical tools and analysts.

Testbeds are needed in order to assess the security posture of the end-to-end system. A testbed replicates a real world environment(s) by using the actual sensors, software, hardware, networks, and data streams used in the field. Bench top systems and stand-alone sensor tests are insufficient because many communication issues never arise until the systems are tested at scale in either a production environment or a well-designed testbed. Network bottlenecks, database I/O limits, and timing limitations must also be tested within realistic testbed environments. Many race conditions, which often lead to security issues, only occur when multiple systems/sensors are competing for the same resource. Testbeds can be configured to test a variety of use cases found in the field. Once potential vulnerabilities are identified within a given system cyber security mitigation strategies can be appropriately developed and then tested so as to determine their effectiveness.

Multiple users or researchers can share the resources within testbeds, which enables both increased access as well as reduced costs due to economies of scale. Industry can use a testbed to evaluate new sensor or communication technology and vendors can test interoperability with existing products and infrastructure. Testbeds can be used by academia to provide students with hands-on (or virtual) experience with real equipment.

PNNL has built and demonstrated testbeds in a variety of domains and our experience has shown that testbeds are feasible. For example a multi-lane simulator was developed to test different configurations of Radiation Portal Monitors. PNNL’s PowerNet testbed enables researchers to develop and test new ideas for the power grid. PNNL also is in the process of building sophisticated cyber testbeds and ranges that will allow a variety of equipment and data to be tested.

This paper will provide guidance for safeguards, nonproliferation, and arm control technology implementers as they seek to improve the cyber security of their deployed systems. In particular the pitfalls of isolated component testing will be addressed. Several key “use cases” for testbed based assessments will be presented. These include increasing the fidelity of modeling a network based attack upon (formerly stand-alone) sensors, identifying the impact of network security, and monitoring the integrity of data flowing through the complex systems of interconnected sensors, servers, and data stores.
The development and use of testbeds will enhance the understanding of current and future risks to the distributed sensor networks and distributed control systems used in the safeguards, nonproliferation, and arms control domain. These are large and complex systems of systems that are being deployed worldwide, sometimes in harsh and potentially vulnerable conditions. Future cyber security assessments of these critical systems must use a holistic, cradle-to-grave approach—an approach that will require the use of realistic testbeds.

1. Introduction

We are in an increasingly data driven world. This trend has crossed over into safeguards, nonproliferation, and arms control verification as governments and international organizations automate data collection and analysis. Sensors deployed in the field typically send data back to data collection sites via a variety of transmission methods, including wireless networks. Analysts at data centers will process this data, but often the first round of data analysis may be completely automated and occur without any direct human oversight. The need for security in these systems is recognized. This need will not go away because soon the “Internet of Things” will connect all of our devices into a ubiquitous always-on network intended for both computer and real-world devices.

Safeguards, nonproliferation, and arms control systems devices are used within larger systems. At a minimum, the system is composed of a trained inspector using these devices in the field. At the other extreme, unattended devices constantly monitor a facility or process and transmit their data via a communications network to a back-end data center where automated processes constantly analyze the data and alert human analysts when alarm conditions occur. Because both systems include devices, all too often our security focus has been only on the devices. But, a holistic approach is needed, one that takes into account component security, network security, data integrity, and availability. Insecure sensors may collect and transmit bad data; insecure networks are vulnerable to replay attacks, data corruption, and denial of service; insecure databases are vulnerable to data corruption and the manipulation of data presented to analytical tools and analysts. An adversary will often exploit the weakest link within the system of systems, which means that a holistic effort is needed to strengthen all components and processes with the system of system.

Testbeds are needed in order to assess the security posture of the end-to-end system. A testbed replicates a real world environment(s) by using the actual sensors, software, hardware, networks, and data streams used in the field. Bench top systems and stand-alone sensor tests are insufficient because many communication issues never arise until the systems are tested at scale in either a production environment or a well-designed testbed. Network bottlenecks, database I/O limits, and timing limitations must also be tested within realistic testbed environments. Many race conditions, which often lead to security issues, only occur when multiple systems/sensors are competing for the same resource. Testbeds can be configured to test a variety of use cases found in the field. Once potential vulnerabilities are identified within a given system cyber security mitigation strategies can be appropriately developed and then tested so as to determine their effectiveness.

The rest of this paper is organized as follows: Section 2 will present a cyber security overview, Section 3 will discuss the use of testbeds, requirements for testbeds will be given in Section 4, and we conclude in Section 5.
2. Cyber Security

Cyber security is a broad term that includes computer security, network security, and information security. Individually, securing just the devices (i.e., computers) or only the communication network, or just the back-end information analysis process is not sufficient. When discussing cyber security we often think of the triad of Confidentiality, Integrity, and Availability (CIA). Nonrepudiation is closely related to integrity, but sufficient different in focus that it deserves its own discussion as well. Policies that enable these four cyber security properties must be implemented with appropriate mechanisms. The three core mechanisms are: authentication, authorization, and audit. Collectively they are called the gold standard because each starts with “au” the chemical symbol for gold (Lampson 2004). The following subsections will discuss each of these terms in more detail.

2.1 Confidentiality

The United States government defines confidentiality as “preserving authorized restrictions on information access and disclosure, including means for protecting personal privacy and proprietary information.” (NIST 2004) Confidentiality is the property that is most commonly thought of cyber security property. However each cyber security property is important and, depending upon the context, their relative rank ordering will change. For example, while a newspaper reporter may want to protect the identity of an informant (high confidentiality), the same reporter would not want to restrict anyone from reading the resulting news story (high availability and low confidentiality).

Data encryption is a common mechanism used to ensure confidentiality. But simply rushing to encrypt all data, whether the data is at rest on a hard disk or in transit across a communications network is not the appropriate solution for all safeguards, nonproliferation, and arms control systems. For example, consider the scenarios where health and safety require real-time dosimetry. Encrypting dosimetry values would degrade the usefulness of these sensors because inspectors would not have ready access to their real-time radiation dose.

2.2 Integrity

Integrity is the cyber security property focused on “guarding against improper information modification or destruction, and includes ensuring information non-repudiation and authenticity.” (NIST 2004) Integrity is a critical to inspectors and analysts. If one can not guarantee that data has remained unchanged from its initial reading at the sensor until its final use, then one can not guarantee the validity of reports based upon that data. Even if data is transmitted securely from the field to the analysts’ back-end databases, one must still ensure that data integrity is maintained within the data center. Otherwise, any data corruption will cast doubt upon either a sensor’s reliability or upon the historical data analysis reports because of the mismatched data values.

2.3 Availability

Availability is defined to mean “ensuring timely and reliable access to and use of information.” (NIST 2004) However, “timely” is an intentionally ambiguous term. For example, on-site physical security personnel expect to receive perimeter alarms quickly so that they can apprehend intruders before they have a chance to complete their mission. On the other hand, inspectors may never receive real-time data feeds from facilities that do not allow their data to be transmitted off-site. In this later case, timely might be defined as “twice a year” and reliability would then be of paramount importance. It would be unacceptable if inspectors arrived on-site and found out that four-months of safeguards data was missing because of a hard disk crash. Availability is related to fault tolerance and in many instances fault tolerance mechanisms are used to achieve availability. But whereas fault tolerance is focused on
continuing normal operation despite the presence of hardware or software faults (IEEE 1990), availability is more focused on maintaining access to critical information (even while under attack). International treaties and agreements may specify availability constraints.

2.4 Non-repudiation
Non-repudiation provides “protection against an individual falsely denying having performed a particular action” (NIST 2013). As such it is a cyber security property that is essential, especially within the context of safeguards, nonproliferation, and arms control. Treaties and safeguards agreement must be enforceable, which means entities must not be able to repudiate data and information that has been collected and transmitted to others.

Integrity includes non-repudiation within its definition, but non-repudiation deserves special mention because of its implicit notion of having a trusted third party that can verify the authenticity of who the actors were for a given transaction. Without a strong non-repudiation framework, a malicious actor could falsify data or actions with little repercussion. At a minimum doubts could be raised as to which data set or sequence of actions was performed.

2.5 Authentication
Authentication is “verifying the identity of a user, process, or device, often as a prerequisite to allowing access to resources in an information system” (NIST 2013). Often we only think of authenticating people, but device authentication is equally important, especially within safeguards, nonproliferation, and arms control. For example, active seals must be uniquely associated with a given security enclosure. If not, then materials control systems may be readily spoofed. Protecting physical equipment with tags and seals can be done and this is solid step toward authenticating devices. But physical seals cannot protect data packets traversing a network, especially if facility-provided network infrastructure is used. Two-factor authentication improves authentication security by requiring users to use both something that users have (e.g., a token) and something that they know (e.g., a PIN) or something that they are (e.g., biometric authentication).

2.6 Authorization
Once individuals and devices are authenticated, their individual authorization levels must be established. Without solid authentication mechanisms in place, authorization can be quickly rendered moot. Consider for example the scenario where a host is able to forge the authentication tokens of an inspector. In this scenario, the host could then masquerade as an inspector and perform actions that only authorized inspectors are permitted to execute.

2.7 Audit
Audit is the third of the three golden (i.e., “au***”) properties. All significant actions and activities must be audited. Auditing serves a two-fold purpose. First, it enables post-mortem analysis of system events and user activities. Second, auditing acts as a deterrent because users know that their actions will be recorded and that they will be held accountable for their actions.

2.8 Policies
Cyber security is not a “one-size fits all” endeavor. Each organization must conduct a robust analysis of the threats to which it is exposed and the risks that it is willing to accept. As mentioned earlier, encryption is not always the solution. For example, newspapers achieve their purpose when their content is widely read (availability) and believed (integrity). On the other hand nuclear weapon designs are highly confidential and must be protected from disclosure with extreme diligence. Owners,
operators, and other stakeholders within safeguards, nonproliferation, and arms control enterprises must each set appropriate policies for their respective use.

3. Testbeds

Testbeds are needed in order to assess the security posture of the end-to-end system. A testbed replicates a real world environment(s) by using the actual sensors, software, hardware, networks, and data streams used in the field. Bench-top systems and stand-alone sensor tests are insufficient because many communication issues never arise until the systems are tested at scale in either a production environment or a well-designed testbed. Ideally testbeds use the same hardware and software as that which will be deployed in order to achieve high test fidelity. But either economic or laboratory constraints may require that some resources be virtualized through the use of software simulators or emulators.

Testbeds can be configured to test a variety of use cases found in the field. Network bottlenecks, database I/O limits, and timing limitations must be tested within realistic testbed environments. Many race conditions, which often lead to security issues, only occur when multiple systems/sensors are competing for the same resource. Virtualized systems can be run in real-time or even faster than real-time in order to stress computational systems. Once potential vulnerabilities are identified within a given system, cyber security mitigation strategies can be appropriately developed and then tested so as to determine their effectiveness.

Multiple users or researchers can share resources within testbeds, which enables both increased access as well as reduced costs due to economies of scale. For example, researchers could spread the baseline cost of extremely expensive cyber security tools across many projects and testing efforts. Similarly, the administrative and regulatory burden of storing special nuclear materials could be consolidated into a small core team of support staff. Industry can use a testbed to evaluate new sensors or communication technologies and vendors can test interoperability with existing products and infrastructure. Testbeds can be used by academia to provide students with hands-on (or virtual) experience with real equipment and analysis tools.

PNNL has built and demonstrated testbeds in a variety of domains and our experience has shown that testbeds are feasible. For example a multi-lane simulator was developed to test different configurations of Radiation Portal Monitors (McKinnon 2009). PNNL’s PowerNet testbed enables researchers to develop and test new ideas for the power grid (Edgar 2011). PNNL also is in the process of building sophisticated cyber testbeds and ranges that will allow a variety of equipment and data to be tested.

4. Testbed Requirements

The use of testbeds at PNNL has lead to general observations and lessons learned, which in turn have driven our proposed set of requirements for testbed developers.

4.1 Lessons Learned
Holistic, sensor to analyst testing is required. Over the years we have seen several systems that work during stand-alone testing, but when integrated into a larger testbed environments they simply do not function as expected. Safeguards, nonproliferation, and arms control systems are systems of systems and their individual components must be tested for interoperability with other components/subsystems within the larger system.
The network matters. Too often systems are implemented assuming that the network “just works.” Long network latencies may inadvertently impact system throughput. One system we tested worked well on a local area network, but failed to process the required number of events when operated in a wide-area network setting because the message round-trip times exceeded a second. Packet loss on a wide-area network will happen, but naïve software engineers might have assumed that everything just works and therefore the implemented error handling routine will be insufficient. Wireless systems may work within open spaces, but they may not work within nuclear facilities amidst all of the facility’s biological shielding that attenuates wireless signals. Bottom-line, real-world networks are anything but pristine and therefore testbeds should be augmented with the capabilities to simulate very noisy and error-prone networks.

Scale matters. Few vendors test their systems at world-wide scales or even state-level scales. Sensor networks generally work well when only one sensor is deployed. But as more and more sensors are interconnected, the aggregate data bandwidth to the backend (database) systems increases. If a poor network or data management algorithms were choose by naïve system designers, then at scale the systems may experience data loss because the system simply can’t keep up. This will of course have both integrity and availability impacts.

Undocumented features, unknown impacts. Vendor documentation covers all of the basics and the frequently asked questions. But with respect to cyber security, it is the “extra”, undocumented features that lead to increased vulnerabilities. Software bugs that impact core functionality are found because these bugs impact typical users. But typical users do not overflow buffers or commit any of the other atypical acts that system attackers perform. On the more benign side, consider whether the vendor has documented all of the used network ports and communication protocols? Can default passwords be changed? What happens if network addresses are misconfigured or if two sensors are configured with the same Internet address? Will the status messages of one sensor have priority over another vendor’s alarm message when both sensor systems are operated on the same network? Questions like these can often only be answered through robust experimentation and careful analysis of system logs.

4.2 Testbed Requirements
Four key requirements exist for testbeds in general, but especially for testbeds that will used to assess safeguards and arms control systems.

- Use actual systems
- Leverage virtualization
- Networking features must be configuration
- Isolate your testbed systems

Using actual systems significantly increases the fidelity of the testbed results. When simulated or emulated systems are used, the test results must be analyzed to determine if anomalous behaviors are due to the systems under test or if they are due to inaccuracies due to simplifying assumptions used by the simulation or emulation models.

Leveraging virtualization enables systems to be economically tested at scale. Many safeguards and arms control systems are expensive and therefore it is infeasible to conduct large-scale tests of how multiple systems at either state-wide or world-wide scales will interact. Using virtualization within the testbed, expensive systems can be replicated via software and scaled-up to ensure that real-world data volumes and data rates can be achieved.
Network features must be configurable. A well-designed system should be able to operate within a broad set of communications service levels. However, in order to verify this, the testbed must be able to replicate a broad range of real-world communication service levels. For example, latencies that result from moving data state-wide will be significantly different that if data were moved world-wide and therefore the testbed must be able to replicate latencies within both ranges. The networking features of the testbed must be able to replicate bandwidth constraints, network error rates (either bit errors or dropped packets), network latencies, and network jitter.

Testbed systems should be isolated from the production systems. This isolation helps ensure robust testing by limiting the impact of outside, non-controllable events on the systems under test. The isolation also enables robust cyber security testing because testers will have the freedom to conduct a wide variety of penetration tests and edge condition tests—tests that all too often cause systems to fail.

5. Conclusion

The development and use of testbeds will enhance the understanding of current and future risks to the distributed sensor networks and distributed control systems used in the safeguards, nonproliferation, and arms control domains. These are large and complex systems of systems that are being deployed worldwide, sometimes in harsh and potentially vulnerable conditions. These systems are at risk of failure, especially if performance testing and cyber security testing are not conducted before these systems are deployed to the field. Future cyber security assessments of critical safeguards, nonproliferation, and arms control systems must use a holistic, cradle-to-grave approach—an approach that will require the use of realistic testbeds.

Bibliography


Contact Information
A. David McKinnon, Ph.D., david.mckinnon@pnnl.gov, 509-375-3968, corresponding author
Mary Sue Hoxie, marysue.hoxie@pnnl.gov
Kit Conklin, christopher.conklin@pnnl.gov
All authors are affiliated with the Pacific Northwest National Laboratory, PO Box 999, Richland, Washington, USA.
A Threat Assessment Methodology for Critical Digital Assets of Nonproliferation Agencies: Preliminary Results

KM Whattam, S Godwin, ZN Gastelum, PA Craig Jr., GP Landine, K Conklin
Pacific Northwest National Laboratory

Abstract

As the global community becomes more reliant on interconnected tools, processes, and systems, they simultaneously become more vulnerable to cyber threats. Many nonproliferation agencies are ill-equipped to face the emerging cyber challenges confronting their organizations, either due to human capital, the scale and scope of cyber threats, or the international nature of their organization. While nonproliferation agencies are not unique in their cyber challenges, they find themselves in a precarious position because threats to an agency’s critical digital assets may also pose threats to the international nonproliferation regime. In this work, the authors will present a methodology developed to assess the cyber threats of a U.S. based nonproliferation agency’s critical digital assets, and a series of case studies developed for the nonproliferation agency with subject-matter experts based on known cyber threats and unique vulnerabilities of the organization. The authors will conclude by recommending areas of near-term research needs based on their findings.

Introduction

Digital infrastructure stakeholders live with a pervasive concern over potential and real cyber events. Various tools enable these stakeholders to assess the threats, vulnerabilities, and consequences that could potentially affect their digital infrastructure, as well as measure and assess cyber security effectiveness. Thorough assessments of digital infrastructure systems reveal a plethora of cyber security vulnerabilities.

As the global community becomes more reliant on interconnected tools, processes, and systems, they simultaneously become more vulnerable to cyber threats. Many nonproliferation agencies are ill-equipped to face the emerging cyber challenges confronting their organizations, either due to human capital, the scale and scope of cyber threats, or the international nature of their organization. While nonproliferation agencies are not unique in their cyber challenges, they find themselves in a precarious position because threats to an agency’s critical digital assets may also pose threats to the international nonproliferation regime. In this work, the authors will present a methodology developed to assess the cyber threats of a U.S. based nonproliferation agency’s critical digital assets, and a series of case studies based on known cyber threats and vulnerabilities developed for that agency with subject-matter experts. The authors will conclude by recommending areas of near-term research needs based on their findings.
Defining Cyber Security

For the purpose of this paper cyber security refers to the protection of data, systems, and services that are stored on, or controlled by, some type of computer, including systems with embedded microprocessors. Cyber-attacks target the confidentiality, integrity, and availability of data—these three are also known as the “CIA” triad (shown in Figure 1). Protecting the CIA triad means protecting data, systems, and services in storage, transit, and processing.

![Figure 1. The CIA Triad](image)

According to computer industry experts, confidentiality is defined as the prevention of “unauthorized disclosure of sensitive information. It is the capability to ensure that the necessary level of secrecy is enforced and that information is concealed from unauthorized users.” This could include, for example, information regarding sensitive nuclear operations, or technical specifications of controlled nuclear components and equipment. At the organizational level, confidentiality is achieved primarily through denial of access to unauthorized users such as through role-based access controls, “need to know” permissions, or classification of sensitive information. To protect this information from external threats, firewalls and other measures are used to provide layers of defense between internal operating systems and external adversaries. In addition, measures such as cryptography can be used to ensure that data are protected from unauthorized access, especially during transmission of information. Portable systems such as laptops or cell phones with potential access to sensitive information can also be protected via encryption.

Integrity is the prevention of “unauthorized modification of data, systems, and information, thereby providing assurance of the accuracy of information and systems.” Integrity ensures that data received from a system are the same, accurate (and not modified by another user) data that exist on that system. Integrity also ensures the results of analysis using the data have not been modified. An example is the 3rd Generation Attribute Measurement System, or 3G-AMS, which will integrate different measurement and analysis techniques, including determining the presence of plutonium, highly enriched uranium, and high explosives, into a system that can be certified and authenticated. Integrity also can ensure for

---


2 Ibid
example, that data from sensors (seals, flow monitors, radiation detectors, cameras, etc.) correspond exactly to the data that the sensor is collecting. While this is an issue for locally monitored systems, it becomes even more complex with the transmission of such data to remote locations (such as with remote monitoring systems).

Compromising the integrity of a control system could go beyond manipulation of data to insertion of malware, as the Stuxnet attack, which modified controller output and caused sensors to indicate normal operations even when that was not the case.

Availability refers to “the prevention of loss of access to resources and information to ensure that information is available for use when it is needed.” Essentially, availability is the ability to access information or services without disruption (for example from a denial of service attack that overloads a server’s capacity and makes a service or site unavailable for intended users). Disruption of availability can affect information services such as a website or a server, but can also refer to availability of a technology that is operated by those systems. For example, disruption of availability could refer to loss of access to data feeds from a variety of sensors, loss of availability of a particular system such as a proximity access reader, or disruption of a larger service such as the Comprehensive Nuclear Test Ban Treaty Organization’s (CTBTO) International Data Center.

Protecting information and deployed systems and materials

For nuclear nonproliferation, cyber security can be viewed from two distinct perspectives: protection of sensitive information and protection of deployed technologies and materials. The priority between these two areas varies significantly depending on the specific nuclear nonproliferation scenario at hand.

For example, the operation of certain components and sensors from a port of entry’s radiation portal monitors may be more critical than the protection of certain operational information. However, the protection of sensitive information at the IAEA would likely be deemed more critical than the operation of some components such as a server running the IAEA’s external website or email services.

Protection of information refers to protection from unauthorized access to view, download, modify, or transmit data. This includes the protection of sensitive information that is created and stored on a computer system, as well as information that is produced and transmitted by sensors and detectors. Examples of information that could be vital to protect in the nonproliferation field include information about sensitive nuclear technologies, information that will be used to support a verification objective, as well as operational data such as travel schedules and current locations of personnel in the field.

Threats to information can come from external users (hackers, terrorist groups, or rogue states) or from internal personnel with access to those systems. This requires an understanding of “normal” computer and network use—not just a snapshot in time, but also how network use changes over time. Abnormal use could be coming from a machine (virus/worm) or a human, and originate from internal or external sources. As such, a detailed understanding is needed of the different types of threats and how they might manifest themselves via network activity (e.g., off-hour access and printing/downloading, or using a virtual private network to access unusual locations).
Threats to information security require assurance of information confidentiality and integrity of data (both stored and transmitted). The protection of information is necessary to protect sensitive information from potential proliferators and to verify the integrity of, and provide access to, information needed by the international community to monitor or verify compliance with nonproliferation treaties and agreements.

Protection of technologies refers to the assured use and operability of equipment in the field. This can range from the safe and secure operation of a small system such as radiation portal monitors or larger systems such as a centrifuge or nuclear power reactor. Technologies can also include non-nuclear components such as a server on which data and services are stored. Protection of technologies spans disruption of service events in which key equipment is completely offline to deliberately cause the malfunction of equipment. For example, temporary disruption of service or corrupting data at the CTBTO’s International Data Center could disguise an event of interest. Frequent false alarms or malfunctions of equipment caused by unauthorized access to data systems could lead to loss of credibility of the system, so that less attention would be paid if there was a real event.

The protection of technologies also includes the physical protection of such technologies. One example of this is the use of cyber capabilities to override physical protection systems to access, steal, or sabotage nuclear material, equipment, or facilities. Cyber controls operate some physical protection systems, including proximity access, CCTVs, perimeter intrusion detection and assessment systems (PEDAS), and command and control stations, to name a few. Unauthorized cyber access to systems could cause frequent false alarms, drawing down the credibility of equipment and decreasing responsiveness of staff. Cyber threats to physical protection systems could also include distraction techniques, (i.e., setting off alarms in one area via unauthorized access to a physical protection system while breaching the security system in a different location).

Some systems need protection that cross-cuts technology and information

There are several areas that cross-cut both the protection of technology and the protection of information. One example is the use of information barriers in a measurement system for nuclear disarmament verification. Information barriers are used to assure nuclear weapons disarmament inspectors of the presence of certain treaty-accountable items without providing sensitive technical information about that item. Corruption of an information barrier via a cyber-attack could lead to the disclosure of that sensitive weapons design data.

Another example of cross-cutting information and technology protection for nuclear nonproliferation is a cyber-attack in which malware can make technologies provide erroneous information. Such would be the case (with a measurement system for nuclear disarmament verification) if the malware, which may be owner-installed, made the system appear to verify the presence of a treaty-accountable item when in fact there was not one present (or vice versa).

Conducting a cyber-security baseline survey of the nonproliferation domain

The objective of PNNL’s cyber security assessment methodology is to identify the technical areas within the nonproliferation mission space that require a more robust understanding of the potential
impacts of the cyber security environment, and to inform decision makers of risks, consequences, and potential solutions. This survey focused on identifying cyber risks of several critical nonproliferation activities and mapping those risks against relevant critical digital assets. The information that PNNL used in its survey was collected from PNNL subject matter experts who work in the nonproliferation and cyber security mission spaces. The methodology that PNNL incorporates into its cyber security assessment is a detailed and intensive process to identify vulnerabilities and flesh out potential mitigation strategies; the results of the survey, presented below, reflect only a cursory deployment of the full assessment methodology.

**PNNL’s methodology for assessing cyber security risks**

The methodology described in this paper is derived from industry-accepted practices. Similar frameworks have been implemented by the U.S. Department of Energy Office of Electricity Delivery and Energy Reliability. Other regulatory authorities including the North American Energy Reliability Corporation (NERC) and the U.S. Nuclear Regulatory Commission (NRC) have also used this framework to define cyber security programs, performance, governance of cyber effectiveness, and to facilitate further improvements in cyber defensive capabilities. Recognizing the benefits of using such a framework, the PNNL cyber-security assessment methodology sets the foundation for a full assessment that can be used to assist policy makers in addressing cyber vulnerabilities associated with deployed assets and to construct plausible mitigation strategies that meet mission needs and priorities.

To make informed decisions, policy makers must ensure that the cyber security assessment provides adequate vulnerability details and mitigation information to efficiently and effectively remediate each identified security weakness. Therefore the nonproliferation community must engage in the process of assessing, analyzing and determining threats associated with deployed assets. In order to accomplish these goals, our high-level framework for identifying mission-specific cyber resiliency strategies involves the following activities:

1. Analyze the system environment and assess threat information
2. Perform a detailed analysis to identify vulnerabilities and vectors of attack
3. Develop strategic objectives to mitigate attacks
4. Select security controls to meet the strategic objectives.

While each of these steps has unique differences, information found in each level is integral to the overall success of the assessment. For example, it is impossible to develop effective and efficient threat mitigation strategies without first identifying a project’s critical digital assets and their vulnerabilities. As with threats to physical security operations, cyber security threats do not exist within a vacuum. Vulnerabilities to one critical digital asset (CDA) could impact the operation of another CDA, network, or end-user analysis software. Recognizing this type of systems interconnectivity, it is critical that the assessment methodology be tailored for each stakeholder’s domain. In order to customize an

---

assessment approach for a particular stakeholder’s site or digital environment (e.g. deployed cyber assets), the methodology must also include the identification of a program’s:

- Mission
- Mission-critical functions
- System-critical infrastructure
- Support systems
- Devices

By establishing a simple framework that is mission-informed, the assessment is able to utilize established cyber security assessment practices for targeted nonproliferation missions. For instance, the first step of the methodology is to analyze system environments. As previously discussed, this process includes understanding the various computer components, sensors, and networks that make up a completed system. While this may seem relatively straight-forward, identifying cyber security vulnerabilities in the nonproliferation mission space requires a robust understanding of program objectives and how specific systems are built and fielded to support those objectives.

**Vulnerability Assessments**

A cyber vulnerability is a gap or weakness in a system’s security controls that a threat can exploit. The PNNL vulnerability assessment (VA) broadens and deepens awareness of threats, attacks, vulnerabilities, and the effectiveness of existing controls. From a cyber security perspective, it is not enough to only understand that the desktop computers analysts use to monitor access control systems in nuclear facilities should be password protected. Rather, the PNNL VA process seeks to understand holistic programmatic threats from mission-critical functions down to the lowest levels of implementation. The VA process also establishes baselines that future assessments can use to determine whether planned improvements have occurred.

There are three key phases in the PNNL VA process: threat vector analysis; attack trees; and consequences and susceptibility analysis. Once threat vectors (e.g. the areas of a system that are vulnerable to cyber-attacks) are identified, the framework utilizes cyber security subject matter expertise to build attack trees. Attack trees map out cyber-attack scenarios by identifying various nodes within a system that are susceptible to cyber-attacks. Utilizing attack trees within the PNNL methodology allows cyber security experts to better identify and understand both existing and potential systems vulnerabilities. Additionally, attack trees provide the roadmap that systems architecture experts can use to assess cyber-attack consequences and systems’ susceptibility. This process includes: assessing current system controls and how they protect against attacks; estimating potential for an attack (low, medium, high); estimating and ranking impact to CDA and Operations Mission from a potential attack; and developing risk ratings (e.g. the derived potential for occurrence and impact). Information gathered in this process is then used to develop the systemic risk mitigation plan for each CDA. Using this approach enables visibility of the threat within these priorities. “Visibility” is the careful articulation of the process to understand and organize the threat landscape with a better informed process that incorporates much more of the adversary view leading to a better visibility of the threat environment. The idea of the framework is enabled by the top-down approach in Figure 2.
Figure 2. Cyber Assessment Model

Figure 2 presents a typical model used by a cyber-security expert who has an in-depth understanding of the threat and an adversary. From information collected in each phase of the model, a cyber-security expert can ascertain objective, target, and adversary characteristics necessary to design the tools, techniques, and procedures (TTP) to thwart an attack. If the initial TTP is deemed credible, threat scenarios are designed that drive appropriate analysis to define the threat environment. Once the threat environment has been defined, a cyber-security expert can then begin to construct the programmatic and environment framework that will form the basis of a deployable cyber security posture. This simplified description of the assessment methodology downplays the significant level of complexity of each step in the process; a process that will be determined by the operating environment. To support system owners in understanding the environment, PNNL’s cyber assessment methodology incorporates a comprehensive list of questions meant to acquire threat information. The outcome from this exercise allows the system owner to fully appreciate the threats and vulnerabilities of deployed digital assets and begin the process of establishing mitigation strategies.

In order to demonstrate the framework, PNNL’s cyber security experts evaluated three systems now in place supporting the nonproliferation mission or being evaluated for future support. The assessment is presented in Table 1.
Conclusions

The purpose of this paper was to present a framework for assessing cyber vulnerabilities that could be associated with critical digital assets deployed and used by nonproliferation agencies, and to conduct a very preliminary baseline assessment using PNNL’s methodology. Moving forward, PNNL recommends that nonproliferation agencies conduct a comprehensive assessment of their digital assets using PNNL’s methodology (or similar cyber security methodologies). Understanding potential cyber threats and developing mitigation strategies will ensure that nonproliferation agencies are capable of protecting deployed assets from cyber-attacks.

As a first step, PNNL recommends establishing a multi-laboratory team of cyber experts to conduct a workshop that will ensure nonproliferation policy makers are cognizant of cyber security issues; appreciate the design complexity but relative simplicity of the cyber assessment methodology; and begin the process of mapping out cyber vulnerabilities unique to each office’s digital assets. Following the workshop, PNNL recommends that nonproliferation agencies develop guidelines to establish a consistent process for how offices should address cyber threats associated with deployed digital assets. PNNL and the multi-lab cyber team can provide the technical reachback for nonproliferation agencies to implement the cyber security guidelines.
Table 1. Nonproliferation Cyber Asset Vulnerability Map

<table>
<thead>
<tr>
<th>Mission Domain</th>
<th>Current State</th>
<th>Potential Vulnerabilities</th>
<th>Implication</th>
</tr>
</thead>
</table>
| Remote Monitoring System (RMS) | GTRI deploys the RMS capability in order to monitor the status of radioactive sources at facilities throughout the world.  
- The program provides propriety, self-contained remote monitoring systems designed to protect against the insider threat.  
- The RMS comprises multiple components within an individual facility, including seals, cameras, and a central RMS unit. Additionally, there is an active RFID-tagged steel wire that wraps around the facility’s irradiator. A primary objective of the RMS system is to monitor/verify that the source material in the irradiator has not been tampered with; this objective is accomplished in part by monitoring the RFID seals.  
- The central RMS unit monitors, via an antenna, the RFID seal.  
- The central RMS unit then provides facility data to an offsite monitoring station via a Russian-provided Ethernet connection and VPN. All of these capabilities operate on networks provided by the partner facility | GTRI has no control over the networking or information security measures associated with deployed systems.  
GTRI cannot confirm what, if any, cyber security measures are being taken.  
The site provides the VPN connection between the facility RMS unit and the offsite monitoring station. GTRI cannot verify the security of this connection. | Threatens security of entire system. Not only can someone alter and defeat the system, they can track through the facility network to the GTRI RFID system, which is connected to the central RMS unit, which is then connected to the offsite remote monitoring station. Thus, access to facility networks could give an attacker access to everything the program is trying to monitor.  
Cyber-enabled Physical Attacks. Some cameras are connected to the facility network. If someone were capable of gaining access to the facility network, you could, in theory, falsify or disrupt video feed.  
If you compromise the network, you can hack into a controller and change the access protocol password to 1234, which then would allow someone without the appropriate clearances to enter a closed and/or restricted area. |
<table>
<thead>
<tr>
<th>Mission Domain</th>
<th>Current State</th>
<th>Potential Vulnerabilities</th>
<th>Implication</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAUNA (Swedish Automated Unit for Noble Gas Analysis) System</td>
<td>Automatic radionuclide measurement system that analyzes air samples for xenon. After sample collection and analysis, the results are passed to a data center. However, not all of these capabilities have been established yet. Data are removed via a USB drive. (All USB drives used on the system must be cleaned/sanitized before being plugged into the system). This is an isolated system not connected to the internet. Physical security -- The system itself is protected by a locked gate as well as a locked door. Can download data off of the system via a USB drive.</td>
<td>The computer systems cannot be updated - There is no way to ensure that the updates would not disrupt systems operations. A virus/Malware could be installed on the USB drive.</td>
<td>Even though all data are currently manually interpreted and verified, it is unclear what threats exist should the system be more widely deployed. (e.g., data integrity through the remote monitoring processes)</td>
</tr>
<tr>
<td>LAARS-ES (Laser Ablation, Absorption Ration Spectrometry – Environmental Sampling)</td>
<td>Automated, unattended environmental aerosol sample collection that is being developed and combined with the uranium isotope ratio analysis system to detect enrichment facility misuse. After sample collection, inspectors collect the stand-alone environmental aerosol sampling system and then insert the system into the LAARS-ES instrument for rapid uranium particulate sample assay. The stand-alone system is enclosed in a tamper-indicated box. The LAARS-ES instrument can be designed for deployment within a standard IAEA instrument rack.</td>
<td>If deployed within a standard IAEA instrument rack, the system faces the same cyber threats that the IAEA does (e.g., use of a facility’s network, fiber optic lines, etc.</td>
<td>Digital data findings rely on the security/integrity of the IAEA instrument rack.</td>
</tr>
</tbody>
</table>
### Table 1. contd.

<table>
<thead>
<tr>
<th>Mission Domain</th>
<th>Current State</th>
<th>Potential Vulnerabilities</th>
<th>Implication</th>
</tr>
</thead>
</table>
| MPC&A Access Control Systems | Access control systems are installed at Russian nuclear facilities as part of the overall comprehensive physical protection security system. Access controls are applied in a variety of ways depending on the control requirements and the information, material, or location they are designed to control. Examples include:  
- doors and safes with mechanical locks and key controls  
- hardened doors with biometric controls, badge readers, and balance magnetic switches  
- entry control points that include turnstiles, metal and/or SNM detectors, badge readers, and guards. | Where access controls are employed that require more than just mechanical locks (i.e., at a locker or safe), the control system is typically linked, via fiber optic or similar system, to a Central Alarm Station (CAS) where sensor and alarm signals are evaluated, video is recorded through CCTV, and often capability is provided to CAS operators and/or security officers to lock-down facilities or to override alarms. Badging systems, PIN codes, and biometric data are often used in conjunction with access control systems, all of which are managed electronically. In some cases these systems are updated via sneaker-net using thumb-drives that communicate between the access control systems and the main security computer system that maintain these data. In some cases these systems are managed through an intranet system. | Unauthorized access to controlled areas, systems, and data.  
Data manipulation  
Increased risk of cyber-enabled physical attacks (e.g., override alarms and CCTV video feeds) |
Table 1. contd.

<table>
<thead>
<tr>
<th>Mission Domain</th>
<th>Current State</th>
<th>Potential Vulnerabilities</th>
<th>Implication</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLD 770 RPM Access Panel Issues</td>
<td>Model SC-770s, also known as 770 RPM Access Panels, are general purpose programmable controllers used in TSA Systems, Ltd. Nuclear Radiation Measurement Monitors. SC-770s are the main computer systems within a radiation monitor. The system is used to program and control the gamma and neutron radiation detectors as well as to control the radiation monitor's communication equipment. The unit formulates detection data and then supplies said data to the central alarm station system via an Ethernet connection.</td>
<td>Stationary Detectors -- In order to access the SC-770, the radiation portal monitor's pillar door must be unlocked with a key provided by the manufacturer. All detectors use the same lock and key combination. Mobile Detectors -- Unlike stationary detectors, access to SC-770s in mobile detectors does not require a key. The control panel is unlocked and accessed by sliding open a Plexiglas door. Common Vulnerabilities • Once physical access is obtained, one can use the SC-770 computer to program as well as turn on/off the gamma and neutron detectors. • All SC-770s are password protected. All of the passwords are “1234.” This information is readily accessible on the internet. All of the locks on the cabinets are the same. If you have a lock for one cabinet, you could unlock any deployed system. • Hardware and software in both systems are the same.</td>
<td>Possible to turn off detectors at will. Falsify detector data -- Central alarm system may not know if a detector is actually on. The publicly available manual reveals that every SC-770 password is “1234” and that the sequence cannot be changed. Hardware and Software • You can disconnect the port and connect a laptop to monitor/manipulate data. • The monitor itself would not be able to detect this connection. • Vendor has access to data stream via the Ethernet port.</td>
</tr>
</tbody>
</table>

Resolving the Information Barrier Dilemma:
Next Steps Towards Trusted Zero-Knowledge Nuclear Warhead Verification

Sébastien Philippe,1 Alexander Glaser,1 Mark Walker,1
Boaz Barak,2 and Robert J. Goldston3

1 Nuclear Futures Laboratory, Princeton University
2 Microsoft Research
3 Princeton Plasma Physics Laboratory and Princeton University

ABSTRACT
In the cryptography literature, zero-knowledge protocols are interactive proof systems where a prover can demonstrate the validity of an assertion to a verifier, who will accept the proof with a high probability while not gaining any knowledge beyond the validity of the prover's claim. Going at counter-current from previous historical development, we translated the concept of zero-knowledge proofs first developed for digital applications into the physical world by proposing an approach to nuclear warhead authentication envisioning an inspection protocol that a priori avoids detection of sensitive information. Under such physical zero-knowledge protocol, the host (prover) can prove to an inspector (verifier) that a warhead is authentic without revealing anything about its materials or design. The current developments focus on the practical implementation of such system using non-electronic detectors. After discussing the advantages of both the zero-knowledge and non-electronic properties of our authentication system compared to alternative proposals using so-called information barriers, we present some computational test-results and their impact on the design of our experimental proof-of-concept.

BACKGROUND
Existing nuclear arms-control agreements between the United States and Russia place limits on the number of deployed strategic nuclear weapons. Verification of these agreements can take advantage of the fact that deployed weapons are associated with unique and easily accountable delivery platforms, i.e., missile silos, submarines, and strategic bombers. The next round of nuclear arms-control negotiations, however, may begin also to include tactical weapons and non-deployed weapons. Both would require fundamentally new verification approaches, including authentication of nuclear warheads in storage and authentication of warheads entering the dismantlement queue. Dedicated inspection systems using radiation measurement techniques are likely to play a critical role in verifying such agreements, and different approaches have been proposed since the 1990s to accomplish this task.1 These are usually divided in two categories commonly referred as the template or attribute approaches. The so-called template method is generally considered the most robust verification approach. It envisions the comparison of a complex fingerprint of an inspected item against the fingerprint of a reference item, or template, to confirm that both items are substantially identical. In contrast, the attribute method verifies that the inspected item correctly features a limited set of previously agreed attributes with the consequences of providing a rigid framework where cheating may become easier as well as inherently revealing warhead design information.

What makes the authentication of nuclear weapons difficult in the framework of bi- or multilateral inspections is that any radiation measurements made on these objects, as well as any physical means (hardware) to process these measurements would themselves be highly classified. In the case of the United States, it is important to recall, “nuclear weapons knowledge is born secret”.2 This greatly limits
the ability of an honest host (prover) to convince the inspector party (verifier) that the object presented is in fact an authentic warhead, completeness of the proof, without compromising any classified information. On the other hand, it is equally difficult for the inspector to be certain that the verification procedure cannot be tricked into accepting false or tampered objects, soundness of the proof, without being able to see the measurement data and its associated processing equipment. Both completeness and soundness are in this case greatly affected by the classified nature of the object being inspected.

To address this problem, all proposed inspection systems have had, so far, to rely on engineered information barriers to protect classified information at the cost of introducing inherent limitation in the completeness and soundness of these protocols. Yet none of these have demonstrated the ability to achieve both perfect information security for the host and perfect information integrity for the inspector.

To free ourselves from this conundrum, we have proposed a fundamentally different approach based on the cryptographic concept of zero-knowledge proofs. Using a zero-knowledge protocol eliminates the need of an information barrier in the first place. Going further, we address the important problem of untrusted hardware/electronics and software in authentication system by using non-electronic (pre-loadable) detectors only.

This paper discusses the advantages of both the zero-knowledge and non-electronic properties of our warhead authentication system in the framework of information security and integrity. A series of computational test-results are presented and their impacts on the design of our experimental proof-of-concept are briefly described.

GOING NON-ELECTRONIC
Information barriers are often depicted as complex technological systems using electronic hardware components and potentially sophisticated software. They require both parties to trust that they have no trapdoors hidden from the inspector, which could be used to cause a system to declare invalid object as authentic, nor side channels unknown to the host, which could leak classified information. This raises the issue of the potential presence of untrusted hardware and malicious software within information barriers. Any piece of electronic or line of code can potentially be hacked and tampered.

Ideally, the development of such barriers is carried out jointly between the host and inspector parties to build confidence in the overall security architecture of the system. However, the host is still likely to supply all the building components as well as upload any software prior to measurement. There exists no enforcement mechanism to prevent the host from conducting secret parallel development with the goal of tampering equipment or exploiting any vulnerability he may identify, including after the development phase has been completed. All security checks during post-silicon validation and post-manufacturing testing would therefore be particularly difficult for the inspector party wishing to ensure that the hardware meets the design specifications and that it works as designed.

Hardware Trojans are one important example illustrating the inherent difficulty of hardware authentication. These Trojans are defined as malicious changes or additions to an integrated circuit that add or remove functionality or reduce reliability of the system. The host can potentially insert hardware Trojans at many stages in the product cycle: in the design stage, the manufacturing stage, the assembly stage, and the shipping (supply-chain) stage.
In the design stage, joint development on all aspects of the integrated circuit design from defining high-level functions down to wiring at the transistor level can prevent the host from including malicious side channels. The inspector can then check if the manufactured circuits correspond to the agreed design blueprints. If the information shared during the development phase is limited to high-level functions, while the low-level circuit design remains unspecified, then the inspector would need to reverse-engineer the hardware using recently developed techniques. If all building blocks of the information barriers can be selected by the inspectors at the time of the inspection, (e.g. by cut-and-choose one circuit out of ten), and if the remaining parts can later be (destructively) analyzed, then the swapping of circuits during the assembly and shipping stage should not be an issue.

 Trojan inclusion during the manufacturing stage probably remains the hardest host strategy to detect and deter. Material trojans, based for example on aging or environmental (i.e. temperature) effect on transistors, are extremely difficult to detect. The inspector may not be able to recreate all the variables including environmental ones necessary to trigger the hardware modification when testing the device independently. Recent work has shown that hardware can be even tampered at the sub-transistor level by modifying the dopant masks. Since only changes to the metal, polysilicon or active area can be reliably detected, these dopant Trojans remain immune to detection today.

Finally, once a measurement on a classified item is complete, the inspector would generally not be allowed to re-inspect the equipment. This constraint limits the ability to trust completely the “green light” given beyond the information barrier, even for a system that is assembled with components selected in a cut-and-choose manner because, using the example of the hardware trojan, the non-expected system behavior may only be triggered (temporarily or permanently) during the inspection. Despite these challenges, strong information barriers can be developed as the example of the UK-Norway Initiative has shown. Nevertheless, it may prove difficult or perhaps impossible to convince “truly skeptical” hosts and inspectors of the viability of an inspection system (used in a treaty context on real warheads or warhead components) given the inherent limits to security and integrity of the electronic components used in such a system.

For these reasons, we decided to examine the use of a non-electronic detection mechanism based on physical phenomena and permitting post-measurement inspection. To perform template authentication, when electronic real-time read-out cannot be used, non-electronic detectors must have the primary ability to store data, for example, the total number of neutron counts, in order to be read and analyzed at a later stage. Two detector technologies meet these criteria: superheated emulsions (“bubble detectors”) and neutron activation analysis detectors (e.g. cylinders or prisms of zirconium). These detector types have also important properties in line with the requirement of our zero-knowledge protocol. They have the capability to be preloaded with a desired neutron count prior to the inspection. This preload can persist for hours or days and its decay or aging rate, if present, is well characterized. Preload counts are indistinguishable from counts accumulated during irradiation of the test items. The detectors are energy selective so that the effect of low energy neutrons returning from room walls in an experiment can be minimized. They are insensitive to gammas, have relatively high efficiency, and permit total counts in the range of several thousands to tens of thousands. Other technologies may have similar properties but we will give these the highest priority for our experimental proof-of-concept.

In superheated emulsions, neutron recoil particle can trigger the formation of macroscopically observable bubbles from microscopic droplets that are dispersed in an inert matrix. If the detector vials are stored in adequate conditions to prevent aging, bubbles can be kept in steady state for years.
Gel formulation with high droplet density can be achieved. Temperature control is an important factor for these detectors as both the energy threshold and response of the emulsion are temperature sensitive. Adequate control of environment variables during experiment is therefore important.

Activation analysis is used in fusion research. The concept is based on using energy threshold reaction type such as (n,n’) and (n,2n) reactions with significant cross sections. Product isotopes of these reactions eventually decay emitting gamma radiations. These are consequently counted in adequate detectors after irradiation. Potential candidates for our experiment include \(^{90}_{\text{Z}}(\text{n},2\text{n})^{80}_{\text{Z}}\) with a 12 MeV threshold and \(^{115}_{\text{In}}(\text{n},\text{n}’)^{115}_{\text{mIn}}\) with a 1 MeV threshold. These types of detector prevent the host from visually inspecting any preload however they would require shielding. As opposed to bubble detectors, counting decay events may require hours or even days. However they have the ability to hold a much larger number of counts.

If no electronics is used in detectors, certain parts of the experiment apparatus will still require to be controlled electronically – the neutron generator and post processing equipment among others. These are believed to be much less sensitive since they do not hold, measure or store any classified information. Therefore the inspectors may access them for extensive study upon request. Finally it is important to stress that our non-electronic detectors are neither considered sensitive or classified after the measurements campaign. This is related to the zero-knowledge knowledge property of our authentication protocol and is an important difference with existing methods.

THE VIRTUES OF KNOWING NOTHING

Zero-knowledge protocols are interactive proof systems where a prover can demonstrate the validity of an assertion to a verifier, who will accept the proof with a high probability while not gaining any knowledge beyond the validity of the prover’s claim. In other words, after the proof protocol ends, the verifier gains no new knowledge about the object of the proof while being convinced of its validity. The notion of zero-knowledge remains a property of the prover and is not affected by the behavior of the verifier even when he intends to cheat.

This powerful concept, first developed in the digital domain, can have interesting non-trivial physical application. In the subject of our interest, we argue we can prove two nuclear warheads are identical or similarly close using a zero-knowledge template approach without revealing any information whatsoever on their design, composition or other sensitive data.

To illustrate the concept of physical zero-knowledge proof, we present a simple example related to our authentication protocol:

Alice (the host) has two small cups both containing X marbles where X is some number between 1 and 100. She wants to prove to Bob (the inspector) that both cups contain the same number of marbles, without revealing to him what this number X is. To do so, Alice prepares two buckets, which she claims each contain \((100 - X)\) marbles. Bob now randomly chooses into which bucket which cup is poured. Once this is done, Bob verifies that both buckets contain 100 marbles.

This simple protocol reveals no information on X since, regardless of its value, Bob should always counts 100 marbles in both buckets. If the cups did not have the same number of marbles, then no matter how the buckets are prepared, with probability of 50% after the pouring, one of the bucket will not contain
100 marbles. If Alice and Bob repeats the game five times then if Alice is consistently cheating, she will be caught with \((1 - 2^{-5}) > 95\%\) probability.

Similarly we design a zero-knowledge protocol for warhead authentication. We want to compare a candidate item with a template warhead by recording their direct transmission pattern of 14-MeV neutrons, as well as the intensity of neutrons emitted to the side of the items at a typical angle of 90 degree. In analogy to the marble example, the measurements are recorded using non-electronic detectors that are preloaded with the negative image of the radiograph. The protocol is as follow:

1. The host presents to the inspector the two items to be compared as well as two sets of preloaded non-electronic detectors. Preloaded values are not revealed to the inspector.

2. The inspector chooses which set of detectors will be used with which item.

3. Once the radiography is done. The inspector verifies that the final images are uniform.

If the items actually differ, and the preloads are chosen to complement the two items, then with significant probability of at least 50% the image will not be uniform.

As opposed to the marble example, neutron measurement is inherently statistically noisy. To avoid conveying information through the noise, the detectors are preloaded with noisy values. Since both the preload and the measurement distribution will follow a Poisson distribution and since the sum of two Poisson distribution with mean and standard deviation \((N_p, \sqrt{N_p})\) and \((N_m, \sqrt{N_m})\) is also Poisson with \((N_p + N_m, \sqrt{N_p + N_m})\), our protocol achieves the following:

*The neutron count obtained by any measurement on the template or on a valid submitted item is distributed according to the Poisson distribution with mean and variance equal to a previously agreed value \(N_{\text{max}}\).*

With \(N_{\text{max}}\) being chosen in advance by both sides, neither the measurement nor its noise reveals any new information. \(N_{\text{max}}\) could correspond for example to the maximum number of counts that is expected in the absence of a test item. If a submitted item varies from the template (or the submitted preloads are not identical) an image may be seen that could contain sensitive information. This will be an additional strong incentive for the host not to cheat. However, it also means that the system in its most simple form may be prone to human error. This might necessitate the addition of a fail-safe mechanism to prevent the host for reveling information when committing an unintentional mistake (i.e. wrong disposition of preloaded detectors). Finally, the steps following a measurement should be relatively straightforward. Since the information contained in the detectors is in principle unclassified, protocols can be devised that permit using both host-provided and inspector-provided measurement tools.

**COMPUTATIONAL ANALYSIS**

We now show our approach can be implemented in practice and that it can detect some notional diversions between the two objects. The current setup (see Figure 1) features an array of 367 superheated droplet neutron detectors, a deuterium-tritium (D-T) neutron generator, placed in a polyethylene collimator. Items to be authenticated are placed between the neutron generator and the
detector array. Bubble detectors have the particularity to be sensitive above a certain energy threshold. Transmission detectors are assumed to be sensitive to neutron energy above 10-MeV. Side detectors are currently getting developed and are likely to feature an energy threshold of 1-MeV in order to capture potential fission and other compound nucleus reaction neutrons.

![Figure 1. Experimental setup with neutron source in collimator, test item in container, and detector array. Large-angle detectors are not shown.](image)

Figure 1. Experimental setup with neutron source in collimator, test item in container, and detector array. Large-angle detectors are not shown.

Figure 2 illustrates the transmission results on a valid and invalid item. The sensitivity of the measurement scenarios increases with $N_{\text{max}}$ and the associated improvements of counting statistics.

![Figure 2. Results of MCNP5 simulations illustrating valid and invalid items.](image)

Figure 2. Results of MCNP5 simulations illustrating valid and invalid items. The radiograph of the test
item shown on the left is never measured, i.e., corresponds to a measurement without preloading the detectors. The other panels show total detector counts after measurements on a valid and an invalid item. Shades of gray and colors indicate absolute differences from $N_{\text{max}} = 1000$. The invalid item produces a larger number of suspicious data points, which are in this case spatially correlated.

Additional simulations were performed to determine the range of possible detection for the diversion scenarios of material removal and substitution in both large and small (spatially localized) quantities. We found that the required $N_{\text{max}}$ for large removal or substitution (for example tungsten versus lead) of materials can be as low as 1000. However local substitution of material, for example the replacement of 7% of a total mass of about 8 kg of tungsten in the British Test Object by lead may require a total count of 32,000 to achieve a detection probability of 95%.

These numbers are nevertheless conservative since they do not take into account the use of side detectors. The more realistic case of substitution of uranium-238 for uranium-235 in a nuclear weapon component results in a factor of about two reduction in the induced fission rate due to 14 MeV neutrons. Substitution of reactor-grade for weapon-grade plutonium has a small effect on the directly induced fission rate, but a large effect on the spontaneous fission rate, which could be detected passively by operating the side detectors in the absence of the neutron source.

CONCLUSION

Authenticating nuclear warheads without revealing classified information represents a qualitatively new challenge for international arms-control inspection. Here we have shown an example of a zero-knowledge protocol based on non-electronic differential measurements of transmitted and emitted neutrons that can detect small diversions of heavy metal from a representative test object. This technique will reveal no information about the composition or design of nuclear weapons when only true warheads are submitted for authentication. It therefore does not require an engineered information barrier. The use of non-electronic detectors prevents the most sensitive equipment of our authentication system to be vulnerable to untrusted hardware and software.

If the computational results are encouraging there is still a lot to achieve before a practical system can be readily implemented. Current research focuses on the design and construction of the first experimental proof-of-concept. Topics include shielding of the neutron source, development of the side detectors, effect of room returns on measurement. All have effect on the minimum achievable $N_{\text{max}}$ for the various diversion scenarios.

Timely demonstration of the viability of such an approach could be critical for the next round of arms-control negotiations, which will likely require verification of individual warheads, rather than whole delivery systems. Other such zero-knowledge protocols may be possible. The zero-knowledge approach has the potential to remove a major technical obstacle to new nuclear arms control agreements that include both deployed and non-deployed, strategic and tactical weapons, at substantially lower levels of armament than current agreements.

Acknowledgement. This project is supported by generous grants from Global Zero and the U.S. Department of State, funding from the Princeton Plasma Physics Laboratory, supported by DOE Contract
DE-AC02-09CH11466, and in-kind contributions from Microsoft Research. All simulations were run on Princeton University’s High Performance Cluster.


6 There are many low-level circuit designs that will produce the same high-level functionality.


8 Sharad Malik, personal communication, May 2014.


14 A general Monte Carlo N-particle (MCNP) transport code. Los Alamos National Laboratory, mcnp.lanl.gov.
A Visual Analytics Approach to Structured Data Analysis to Enhance Nonproliferation and Arms Control Verification Activities

David Gillen
Pacific Northwest National Laboratory

Abstract
Analysis activities for Nonproliferation and Arms Control verification require the use of many types of data. Tabular structured data, such as Excel spreadsheets and relational databases, have traditionally been used for data mining activities, where specific queries are issued against data to look for matching results. The application of visual analytics tools to structured data enables further exploration of datasets to promote discovery of previously unknown results. This paper discusses the application of a specific visual analytics tool to datasets related to the field of Arms Control and Nonproliferation to promote the use of visual analytics more broadly in this domain.

Visual analytics focuses on analytical reasoning facilitated by interactive visual interfaces (Wong and Thomas 2004). It promotes exploratory analysis of data, and complements data mining technologies where known patterns can be mined for. Also with a human in the loop, they can bring in domain knowledge and subject matter expertise.

Visual analytics has not widely been applied to this domain. In this paper, we will focus on one type of data: structured data, and show the results of applying a specific visual analytics tool to answer questions in the Arms Control and Nonproliferation domain. We chose to use the T.Rex tool, a visual analytics tool developed at PNNL, which uses a variety of visual exploration patterns to discover relationships in structured datasets, including a facet view, graph view, matrix view, and timeline view. The facet view enables discovery of relationships between categorical information, such as countries and locations. The graph tool visualizes node-link relationship patterns, such as the flow of materials being shipped between parties. The matrix visualization shows highly correlated categories of information. The timeline view shows temporal patterns in data.

In this paper, we will use T.Rex with two different datasets to demonstrate how interactive exploration of the data can aid an analyst with arms control and nonproliferation verification activities. Using a dataset from PIERS (PIERS 2014), we will show how container shipment imports and exports can aid an analyst in understanding the shipping patterns between two countries. We will also use T.Rex to examine a collection of research publications from the IAEA International Nuclear Information System (IAEA 2014) to discover collaborations of concern.

We hope this paper will encourage the use of visual analytics for structured data analysis in the field of nonproliferation and arms control verification. Our paper outlines some of the challenges that exist before broad adoption of these kinds of tools can occur and offers next steps to overcome these challenges.
Introduction
A primary focus of verification activities associated with Arms Control and Nonproliferation, such as those performed by the International Atomic Energy Agency (IAEA), is to verify the accuracy and completeness of declarations made by different states. The primary way this is accomplished is through onsite inspections of different facilities by subject matter experts. Before visiting a facility, inspectors need to obtain a complete understanding of the facility and state they are visiting. This understanding is achieved by analyzing many kinds of data, including data gathered from previous inspections, state-provided declarations about the facility, internal IAEA information and databases, third-party information, and open source material that pertains to the facility (Gyane 2010). The data that are analyzed come in many forms, with varied degrees of complexity and scale. Through their analysis of these data, analysts characterize the nature of the activities and intent of a facility so that onsite inspectors can verify both the completeness and correctness of the state declaration. This complete understanding, or situational awareness, of the facility provides comprehensive information to the inspectors to maximize their ability to cost effectively verify the completeness and the correctness of the state declarations (Ferguson 2010).

In studying a facility, organization, or state, analysts will look at many different aspects of its activities. One area of investigation might include analysis of commerce patterns, such as materials being shipped to and from the facility, economic patterns related to these commerce activities, and different commerce and economic networks that this facility might participate in. Another area of investigation might focus more on discovering what kind of research is being performed within a given state or organization: Is the organization or state engaged in any research collaborations of concern? These two areas are just examples of the kinds of data analysis that is performed and are the two areas that are the focus of this paper.

In both the case of analyzing commerce patterns and the case of analyzing co-authorship publication, the amount of data available makes it very difficult for analysts to simply read and digest all of the information available. Thus, analytic tools are needed to aid analysts in exploring these data sources, understanding of their contents, and capturing the knowledge gained from these data sources to form situational awareness of a facility.

The remainder of this paper explains how we applied the T.Rex application, a visual analytics tool developed at the Pacific Northwest National Laboratory (PNNL), to these problem domains. The paper starts with a description of visual analytics and what makes visual analytics tools different from other kinds of tools, followed by a brief introduction to T.Rex as a specific visual analytics tool. Next, we present two case studies: analysis of commerce patterns and analysis of research paper co-authorship. We then conclude with a section that describes the challenges for applying visual analytic tools, such as T.Rex, to the Arms Control and Nonproliferation domain and offers some potential ways to overcome these challenges.

Visual Analytics
Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces (Thomas and Cook 2005). Visual analytics expresses itself in the form of analysis tools that combine visual representations of data with capabilities for analytic reasoning, resulting in systems that enable an analytic discourse between analysts and their data. By visually exploring data, analysts can use their domain expertise to visually identify anomalies in the data based on previous observations or
constraints external to the data, help fill in gaps in incomplete data, and form hypotheses. But visual analytics is more than simply visualizing information. Visual analytics combines the best automated statistical or data mining algorithms with appropriate visualization and interaction techniques (Keim et al. 2008). Analysts are “in the loop” to guide the tool or system along a path specific for their analysis path.

Visual analytics complements traditional data mining techniques by adding user-guided exploration through visualization. Data mining, or knowledge discovery, is the science of extracting useful knowledge from such huge data repositories (Chakrabarti et al. 2006). It is typically based in statistical algorithms or machine learning and often displays results in the form of a visualization. When visual analytics is combined with such a visualization, the whole process becomes more collaborative between the user and the system, where the user’s interactions with the visualization can influence the outcome of statistical and machine-learning models. Visual analytics applied to data mining problems also helps an analyst understand previously unknown data to derive more insights through exploration (Keim 2002).

A number of visual analytic tools exist today and have been developed in academia, government organizations, the open source community, and commercial companies. Some are available to safeguards analysts, but many are not. In the Challenges section later in this paper, we speculate as to why many of these tools are not in use.

T.Rex

T.Rex is a PNNL-developed visual analytics tool to help analysts explore tabular data sources, to better understand their contents, and to capture the knowledge gained from these data sources. Analysts can quickly identify patterns of interest and the records and fields that capture those patterns. To discover these patterns, T.Rex uses a variety of visual exploration techniques, including a facet view, graph view, matrix view, and timeline view (see Figure 1). The facet view enables discovery of relationships between categorical information, such as countries and locations. The graph tool visualizes node-link relationship patterns, such as the flow of materials being shipped between parties. The matrix visualization shows highly correlated categories of information. The timeline view shows temporal patterns in data.
An original goal of T.Rex was to facilitate the analysis of transactional data, hence the name “T.Rex” which evolved from “Transaction Exploration.” Our definition of “transaction” is adopted from (Merriam-Webster 2014) to mean “an occurrence in which goods, services, or money are passed from one person, account, etc., to another” but is expanded to also include non-tangible goods or services. For example, in our definition, transactions could also include a record of phone call exchanges between parties, network traffic between computers, and general associations among individuals that don’t really involve “passing from one person to another,” such as co-authorship for research papers. We refer to the participants of a transaction as “actors,” and when we identify the actors of a transaction to T.Rex, we can perform advanced operations to leverage this information. Transactional data are difficult to analyze and understand because of the complex relationships among the actors participating in a set of transactions. Our view of a transaction is not limited to a single actor interacting with another actor, but can involve a complex sequence of interactions that together form a larger transaction. A design goal of T.Rex is to reveal relationships and patterns in these more complex styles of transactions. Besides focusing on transactions, T.Rex has become a generalized structured data analysis tool that has strengths for not only working with transactional data (by our definition) but also working with almost any structured dataset where all of the information about an individual transaction or relationship is contained within a single row of data.

A strength of T.Rex is its ability to support exploratory analysis of datasets. Exploratory analysis is used when little is known about the data being analyzed and no assumptions have been made about the data. The objective of exploratory analysis is to characterize the data and identify patterns and anomalies to provide better understanding of the dataset. Analysts can use their domain expertise in the subject matter at hand to supplement what they are seeing in the data to create a more comprehensive understanding. Analysts can also use exploratory analysis when testing hypotheses.
Rather than simply performing searches across a dataset, T.Rex provides a user interface that encourages exploration of facets of information related to those returned by user-initiated searches.

Another key feature of T.Rex is a rich annotation capability that helps the analyst capture knowledge gained from the data being analyzed. Often, analysis tools provide visual metaphors for interacting with and exploring data, but the user is responsible for taking notes and capturing knowledge in a separate tool. This disconnect makes it more challenging for analysts to keep track of their progress, link back to specific evidence that supports or refutes a developing hypothesis, and share information for presentation, dissemination to others, or training of junior staff. With T.Rex, analysts can annotate key parts of their analysis process—such as grouping records, filtering records, or for other ad-hoc points—within the tool itself.

**Case Studies**

**Characterizing Shipping Patterns with PIERS**

A case study for analyzing data in T.Rex for this domain is to characterize shipping patterns between different countries, looking for anomalies and patterns of concern. One such dataset to use for this analysis is from the Port Import Export Reporting Service (PIERS). PIERS collects more than 15 million bills of lading per year relative to over 20 million shipments for the United States, most of which are containerized commodities. PIERS processes these data into databases, facts, and figures, which others can then use to understand global trade. All commodity data used in this study were from calendar year 2010. Others have used these data for analysis (Sanfilippo and Chikkagoudar 2013) and have applied statistical models for anomaly detection. We build upon this work by adding user interaction to the statistical displays of T.Rex, which allows analysts to guide the exploration through analytic discourse.

For this case study, we start by loading the PIERS dataset into T.Rex. For this analysis, we use the PIERS 2010 dataset, comprising just over 200,000 records that describe container shipment imports to the United States. The dataset came to us as a CSV file, with 243,910 rows and 80 columns. Many of the columns are not relevant for our analysis, and we can instruct T.Rex to ignore them. We also instruct T.Rex to create facet columns for categorical data columns of interest, such as Importer Name, US Port, and Originating Country. These are the actors that participate in the transactions of this dataset: shipping containers around the world. Shipments have an originating country, are associated with an exporter company and an importer, depart from a shipping port, arrive at an arrival port, and are shipped by a shipping company on a named vessel registered to a country. Any of these values can be considered an actor in a shipping transaction, so we will mark them all as facet columns in T.Rex, to leverage some of the relationship discovery features of some of the visualizations.

For our case study, we will characterize the patterns of two different types of shipments: toys and scrap metal.

**Toy Shipments**

In recent years, toy-making has been an industry used to acquire dual-use materials (Warrick 2012). We can use T.Rex to characterize shipments related to toy-making and to look for anomalies that could warrant further investigation. When analysts begin their investigation, they may not have a hypothesis they are trying to test but instead are simply trying to characterize the shipping patterns related to the toy-making industry.
From the full dataset, we perform a search for “toy*” on the four cargo fields; we find there are just over 2,842 records containing some form of the word “toy.” If we look at the facet values, we can see the top cargo summary field is Automobiles. This is because “Toyota” is mentioned in quite a few of the cargo fields, so we need to filter out these records as well. After removing the Toyota records, we are left with 357 toy-related records to analyze.

One question we might want to explore is: What importer and exporter companies are used for shipments containing toys? We can use the graph tool to graph importer-to-exporter relationships. Upon constructing this graph, we can see that there are many one-to-one relationships, where one importer always uses the same exporter, and vice-versa. But there are some many-to-one and some many-to-many relationships among importers and exporters that can easily be seen in the graph, and these could be interesting patterns.

![Graph Visualization](image)

**Figure 2.** A graph visualization showing connections between Importers and Exporters. Singular relationship patterns of a single importer always using a single exporter have been collapsed to circles. Only non-singular relationships are expanded above, with importers shown as a square and exporters shown as a triangle.

Upon investigation of the individual shipping records that are behind some of these patterns, it appears that some of the shipments may contain multiple cargo categories or perhaps are mislabeled. One of the one-to-many relationships comprises shipping records for a Wind Energy company, and another is for an Energy Technology company (see Figure 2). These findings would likely be unexpected to an analyst searching for “toys.” Perhaps there are shipments that contain toys and other materials in the same shipment. One discovery is a majority of these records have a cargo summary of “toys nesoi” with
nesoi being an abbreviation for “Not Elsewhere Specified or Included.” Uncertain of the importance of this label, one might filter out any records containing nesoi, and analyze the remaining 20 records that do not contain the “nesoi” label. Using the facets tool and looking at the cargo summary facet, the values “Automobiles” and “Machinery Parts; Misc” continue to appear. The Automobiles value results from the use of the abbreviation “TOY” for “Toyota” in 6 of the records, and the Machinery value results from the Toyokoki manufacturer name being used in a cargo description field. Filtering out these remaining 7 records results in 13 records that appear to be representative of the shipments we are interested in. To answer the original question regarding characterization of the importers and exporters, these 13 shipping records use 12 individual pairs of exporters and importers, with two shipments sharing the same importer and exporter.

**Scrap Metal Shipments**

According to recently published information (IAEA 2014), one of the major categories of incidents involving unauthorized activities has been related to scrap metals. We use T.Rex to look at shipping records from PIERS containing scrap metals to discover anomalies that warrant further investigation. From the full dataset, we can perform a search for “scrap” on the four cargo-related fields, resulting in 3,249 records that were labeled as containing scrap. We then create a filter for these 3,249 records and continue our analysis on this filtered set. If we look at the U.S. dollar value of these shipments, we can see that the shipments range in value from $0 to over $800 million. The number of tons in each shipment ranges from 0 tons to more than 76,000 tons. We next use the Scatterplot view in T.Rex to look for patterns and anomalies in the relationship between these values (see Figure 3). There are many shipments of high tonnage but little value (the lower-right area of the visualization); there is a series of shipments that forms a direct correlation between tonnage and dollar value (the diagonal line from bottom left to upper right); and there are a small number of shipments that seem to have a higher than expected dollar value compared to their weight (to the left side of the visualization, above the diagonal line). There is also an anomaly record that seems to have a high dollar value and a fairly high weight as well—this record does not fit the pattern of the other shipments.
If we select the records that contribute to the diagonal pattern and also look at the categorical data shown in the facets tool, we can see that most of these records are in the Tantalum or the Titanium cargo category. The value of these shipments is directly related to the weight of the shipment. The records that have very little value (along the bottom of the visualization) are mostly found in the Scrap Metals or the Residues categories. The anomaly record at the top-center is for a shipment of ferrous scrap being sent from Rotterdam to Louisiana. Using a combination of the scatterplot, facets, and full-text searching, we find that most of the other ferrous scrap is along the bottom of the graph, with high weight and low dollar value. Maybe there is something special about this anomaly shipment that caused it to have high value.

Research Publication Co-Authorship
We can also use T.Rex to analyze patterns with co-authorship of research papers. Nonproliferation analysts keep track of researchers from different countries that collaborate with each other, especially collaborations that could result in knowledge sharing in certain domain areas. To discover these kinds of collaborations, we use T.Rex to analyze an export of some metadata records we downloaded from the IAEA International Nuclear Information System (INIS) website. Exporting these metadata from the INIS website was challenging. We used a combination of website crawlers, HTML parsing, and some data wrangling in Excel to produce a spreadsheet containing metadata for just over 4,000 publications. Our
spreadsheet contains columns for the publication title, date of publication, author names, author affiliations, the language used for the publication, and keywords/subject areas for each paper.

After loading the data into T.Rex, we use the graph tool to show connections between the countries of author affiliations. We can then explore the graph by either hovering our cursor over nodes with interesting patterns of connections or using the facets tool to select country names of interest and see what other countries are connected. For example, for no particular reason, we can click on France in the facets tool and see the other countries connected to France in the graph tool. These are the countries that researchers from France collaborated with for their publications. If we are interested in exploring France a little deeper, we can create a filter for only the publications that include France. With this filter active, all of the non-France publications are removed from all of the visualizations. Exploring the France publications a little more, we see that in 1983, there was a publication that was co-authored by an institution in France and an institution in Uruguay (see Figure 4). As we explore other countries that researchers in France have collaborated with, we may discover publications that warrant extra investigations simply because of the amount of developed technologies in each country.
Figure 4. T.Rex facets and graph view showing the relationship between publications co-authored by organizations in France and Uruguay.

A future version of T.Rex will allow the user to create different icons on the graph for different groups of entities. For example, more developed countries could be shown in one shape and one color, and less developed counties could be shown in another shape and a different color. These added visual indicators will reveal potential collaborations of concern that may warrant additional investigation.

Challenges
One of the largest challenges for organizations like IAEA to use visual analytics tools like T.Rex is the process by which organizations like IAEA must go through to acquire government-developed tools. Many high-quality analytic tools are available from the commercial marketplace and government
organizations, such as U.S. Department of Energy national laboratories. But the process for organizations like IAEA to bring new tools in house is very lengthy, because of the complex relationships IAEA has with different governments. For example, when IAEA is working with the U.S. government, requests for technology are channeled through the U.S. Department of State. Sometimes the requests for new capabilities are very specific, and sometimes the requests are less specific, requiring the Department of State to conduct a competitive award process. This process can often take up to five years from the time a new technology is available to the time it is deployed at IAEA. Early engagement with the IAEA that involves continued exposure of analysis tools may help streamline this process in the future.

Another challenge is getting data imported and ingested into analytic tools. For some of the analytic tools at IAEA, a certain amount of effort is required to map information sources into those tools, potentially limiting the tools’ usefulness for certain kinds of data. Often, tool developers and commercial companies address this problem through training, but the training is often not sufficient. Staff turnover at an organization leaves a knowledge gap between those who were trained and newly hired staff members who were not. Tool developers need to simplify the ability to import data into their analysis tools, eliminating the need for continual training.

Data providers also need to make their data available in formats more suitable for analysis. When trying to analyze publication information from the IAEA INIS system, we struggled with an efficient way to gather a collection of metadata for our analysis. There might have been other ways to get this data, but it was not obvious to our research team. Data providers have well-designed human-readable and searchable websites, but they also need to provide computer-friendly access to these same information sources, either through application programming interfaces (APIs), web services, or other more automated query interfaces.

A final challenge faced by Arms Control and Nonproliferation verification analysts is a common problem to many other domains: dealing with “big data.” IBM defines a big data problem by measuring it against four dimensions, known as the four Vs of big data: velocity, variety, volume, and veracity. “Velocity” refers to the speed at which data are produced and thus must be consumed and analyzed. “Variety” refers to all of the different formats that data can be in. “Volume” refers to the enormous sizes of today’s datasets and the sizes of datasets to come in the future. “Veracity” refers to the uncertainty that exists in data and whether the data can be trusted for critical decision making (IBM 2014). Today’s analysis tools are overcoming some of these challenges, but the problem space continues to grow as the solution space continues to evolve. Above, we mentioned that it can take up to five years for new tools to be deployed at IAEA. If big data continues to evolve in its four V dimensions at the rate it has been evolving, but analytical tools continually lag five years behind, organizations like IAEA will be at a tremendous disadvantage.

Conclusion
Researchers at PNNL have found that visual analytics tools such as T.Rex have applicability in the Arms Control and Nonproliferation domain with analysis of structured data sources. The analysis of these structured data sources provides onsite inspectors with additional situational awareness, so they can better verify the correctness and completeness of a state’s declaration of a facility. When analysts are unsure of exactly what they are looking for in their data, visual analytics tools help provide a way to interact with their data and learn about their data as they go through an analytic discourse. There is
simply too much data to analyze by reading each and every document, so organizations like IAEA need to continue to embrace the power that visual analytics tools can provide for analyzing large amounts of data, collecting evidence, testing hypotheses, and capturing the knowledge learned to provide the situational awareness needed for its verification activities.

Acknowledgements
The T.Rex software was supported by a government client and developed at the Pacific Northwest National Laboratory (PNNL). PNNL is managed for the US Department of Energy by Battelle under Contract DE-AC05-76RL01830.

A portion of this research was performed using PNNL Institutional Computing at Pacific Northwest National Laboratory.

References


Generating Confidence from Heterogeneous Data in Safeguards and Arms Control Monitoring Concepts

Jay Brotz, Michael Coram, Maikael Thomas, Sharon DeLand
Sandia National Laboratories¹, Albuquerque, New Mexico

Abstract
Safeguards and arms control monitoring regimes are designed to provide confidence that each State Party is fulfilling its agreement obligations. One mechanism to increase confidence with technical monitoring systems is to provide “evidence in depth” – an analogue to the “defense in depth” concept used in the physical security domain – with multiple sensors providing layers of information about the same actions. Such evidence allows for a single sensor to fail, have incomplete information, or simply be insufficient to provide adequate confidence. This paper describes the collection, organization, storage, and use of data generated by heterogeneous sensors and monitors as well as the data management architectural challenges of this approach. A system currently under research and development to address these challenges is described. This system provides evidence in depth, allowing continuity of knowledge of items of interest over a significant duration even in the presence of significant activity at the host facility. The system combines a set of facility sensors (door switches, cameras, and motion detectors) with a set of item monitors that can identify tamper attempts, movements, and other activities of interest to the containerized item being monitored. Sensors can trigger each other, providing more evidence than a single sensor can provide. In addition, the system is extensible, allowing additional technologies (sensors and item monitors) to be integrated. The system is currently fielded as part of a research and development testbed allowing sensor and item monitoring technologies to be evaluated.

Introduction
As international nuclear safeguards and nuclear arms control regimes become more sophisticated, so must the verification measures contained within them. On-site inspection verification was once limited to what an inspector could see and perhaps measure with simple equipment (Woolf 2011, IAEA 2007). However, as both types of regimes have expanded in both numbers of inspected entities and in scope of monitored items and materials, increasing the number of on-site inspections has diminishing returns and becomes prohibitively expensive. To address this gap between the ambitions of the monitoring regimes and the ability of inspectors, technology has been proposed to supplement inspectors and expand the regimes into areas that were difficult to reach previously. The International Atomic Energy Agency (IAEA) has already embraced the use of technology for unattended monitoring of safeguarded nuclear materials and technologies and continues to encourage research into technological means for enhancing international nuclear safeguards (Schanfein 2011). While current and past nuclear arms control treaties have been very limited in the use of technology for verification (not including national technical means), it has been suggested that the logical next step in the bilateral arms control regime between the United States and the Russian Federation will account for individual warheads instead of delivery systems (Pifer 2010). If this happens, either the number of on-site inspections must be increased or on-site unattended monitoring systems will be needed to supplement those inspections to

¹ Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy’s National Nuclear Security Administration under contract DE-AC04-94AL85000.
account for a greater number of smaller items and a greater number of sites (Brotz 2013). Given the expense and disruption to the host of on-site inspections, it is a distinct possibility that future arms control agreements rely on unattended, continuous monitoring systems for verification in addition to on-site inspections.

While safeguards monitoring systems and arms control monitoring system have different customers, users, and monitored items, there exists sufficient overlap in fundamental concepts of the two types of monitoring systems that we refer only to monitoring systems from this point on. Monitoring of accountable items or materials gives the monitoring party confidence that an agreement is being upheld by the host party – that the host is not diverting civilian materials for military use or producing, maintaining, or deploying more nuclear warheads than an agreement allows. This confidence is gained by inspecting these items and materials in person, but the confidence decreases over time between on-site inspections. A monitoring system consists of item or materials monitors that continuously verify the integrity and properties of a container, facility sensors that recognize and record facility access and the properties of an area or a perimeter, and a data management system to collect, store, and process the data generated by these sensors and monitors. The purpose of these components is to allow the monitor to have some knowledge of individual properties of individual items or attributes of a facility when inspectors are not present. The purpose of the system is to allow the monitor to gain sufficient confidence that the obligations of the host are being met in the absence of significant numbers of on-site inspections.

**Heterogeneous Data Synergy**

A single sensor can give the monitoring party a certain degree of confidence in the integrity of an accountable item or the use of a facility. For example, an active item monitor that seals a container surrounding an accountable item can provide the monitoring party confidence, once they have confirmed that the contents are as declared, that the item has not been accessed or moved since that confirmation. If there is access or if the container is moved to another area, the system will know and log the event, since the item monitor is active and transmits the “seal open” event as a message, or is no longer in communication with a local receiver. This is single-layer evidence. In order for this to be sufficient, the monitoring party has to accept that the host cannot defeat the item monitor in any way. An improvement is evidence in depth – an analogue to defense in depth. By placing door switches and motion detectors in the room with the accountable item, the monitoring party would have a record of access to the room. Adding cameras that are triggered by the door switch and motion detector provides the monitoring party with images of what happened in the room every time there was access. Triggering the camera images on more significant events, such as movements or tamper indications, can give inspectors information such as what was happening to an accountable item when such events occurred. Further layers can be added with portal monitors, radiation detectors, and perimeter monitors. Each layer gives the monitoring party more confidence that the host is not defeating the system, as the host would have to defeat each layer simultaneously. The ideal combination of sensors and monitors, as in the defense in depth principle used in the physical security realm, consists of multiple independent and greatly differing sensors. This drives diversity in the set of monitoring system components.

Another driver for technical diversity in monitoring systems is the difference between continuous, unattended sensors, and equipment used during on-site inspections. This equipment, such as portable radiation detectors, radioisotope air samples, or thermographic imagers, among many other types of equipment used only during an inspection, has different data requirements, but still may send data to the monitoring system.
As research into monitoring systems matures and expands, more and more types of monitoring systems are being investigated with different data types and uses. Heterogeneous data is likely to be the norm for future monitoring systems, especially within systems used in support of research into monitoring system, sensors, and item monitors.

A test bed for research and development of monitoring system concepts must support such diverse sensor and monitor nodes and their heterogeneous data. The primary requirements for the data management architecture are flexibility – the ability to interface with a variety of diverse sensors with different data types – and extensibility – the ability to expand the system whenever a new component is ready to be integrated. A derived requirement that allows the fulfillment of these requirements is a common interface definition. There are two ways to integrate discrete components into a data management system: 1) allow the components to have freedom of interface and create a new element in the data collection system that can translate that information for each component or 2) require that each component that integrates into the data management system use a common interface so that only one element is required in the data collection system to understand incoming information. This is shown figuratively in Figure 1. The former (the upper part of the figure) places more integration burden on the data management system, while the latter (the lower part of the figure) places the burden on the components. However, the sum of the integration burden is less in the latter case, especially if the interface definition is well conceived and even more so if the hardware and software necessary for that integration can be standardized and embedded into the components.
Components that have been considered for use in a monitoring system for safeguards or arms control include tags, seals, door switches, motion detectors, cameras, radiation monitors, and portal monitors. In some cases, the data from a particular sensor may stand alone. But in most cases, confidence is generated from data synergy: the combination of data from more than one source to strengthen the evidence of a given action or state. For example, a tag may be seen by a portal receiver and then by a receiver in a particular section of a storage bunker, at which time the system may check the seal. Simultaneously, door switches and motion detectors detect presence in that section of the bunker, and cameras collect images of that section during that time. All of these actions taken together give strong evidence of the location and integrity of that monitored item.

**Secure Sensor Protocol**

As the diversity of monitoring technologies increases in a safeguards or arms control monitoring test bed, the challenge of integration rises. As mentioned above, a common interface definition keeps this challenge from becoming untenable. The Secure Sensor Protocol (SSP), derived from the Secure Sensor Platform – a hardware and software framework for monitoring technologies – was developed by Sandia National Laboratories (SNL) with funding from the National Nuclear Security Administration (NNSA) Office of Nuclear Verification and is the standard interface definition for a multi-lab effort sponsored by...
the NNSA Office of Defense Nuclear Nonproliferation Research and Development. SSP consists of message definition for data to be transferred from monitoring nodes to the data collection system and command definitions for data flowing in the opposite direction. In addition, the SSP interface definition specifies message flow and expected message behavior. SSP messages and commands only define the application layer (of the OSI network model (ISO/IEC 7498-1 1994)), using the standards of TCP/IP and Ethernet for all lower layers, though other network and transport layers could be used if desired. Message and command requirements and behavior are dependent on node type. The types of nodes identified in the SSP interface definition are listed in Table 1 and examples of node types used in the Chain of Custody project are indicated in Table 2.

Table 1. SSP Node Types

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed</td>
<td>Nodes that are intended to be stationary for their lifetime (e.g., attached to a wall in a bunker)</td>
</tr>
<tr>
<td>Transient</td>
<td>Nodes that are intended to move (e.g., attached to a treaty-accountable item)</td>
</tr>
<tr>
<td>Direct</td>
<td>Nodes that communicate to the data management system without an intervening proxy (i.e., they create and send formatted SSP messages and receive SSP commands directly); Note: proxies themselves are direct nodes</td>
</tr>
<tr>
<td>Proxied</td>
<td>Node that communicate with the data management system through a proxy</td>
</tr>
<tr>
<td>Volatile</td>
<td>Nodes that can be queried as to their identification or state, but cannot initiate actions and do not have their own message numbers</td>
</tr>
<tr>
<td>Non-volatile</td>
<td>Nodes that have their own message numbers and can initiate actions</td>
</tr>
<tr>
<td>RF</td>
<td>Direct nodes that use a TCP Transceiver as a bridge between the node and the data management system</td>
</tr>
<tr>
<td>TCP</td>
<td>Direct nodes that use an Ethernet interface on the local network to communicate with the data management system</td>
</tr>
</tbody>
</table>

Table 2. SSP Node Examples

<table>
<thead>
<tr>
<th>Type</th>
<th>Examples from the Chain of Custody Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Proxied Non-volatile</td>
<td>None</td>
</tr>
<tr>
<td>Fixed Proxied Volatile</td>
<td>None</td>
</tr>
<tr>
<td>Fixed Direct Non-volatile</td>
<td>SSP Camera (TCP), Integrated RF Seal Reader (TCP Proxy), Advanced RFID Tag Reader (TCP Proxy), SSP Door Switch (RF), SSP Motion Detector (RF)</td>
</tr>
<tr>
<td>Fixed Direct Volatile</td>
<td>Not possible</td>
</tr>
<tr>
<td>Transient Proxied Non-volatile</td>
<td>Integrated RF Seal</td>
</tr>
<tr>
<td>Transient Proxied Volatile</td>
<td>Advanced RFID Tag/Seal</td>
</tr>
<tr>
<td>Transient Direct Non-volatile</td>
<td>Chain of Custody Item Monitor, Tiny Gamma Spectrometer</td>
</tr>
<tr>
<td>Transient Direct Volatile</td>
<td>Not possible</td>
</tr>
</tbody>
</table>
The messages and commands of the SSP protocol are composed of Type Length Value packets, or TLVs in short form (see Figure 2). These packets are intended to be a compact container for a single datum. Multiple data are contained in multiple TLVs, sometimes in the same message. Four metadata TLVs are a part of every message: message timestamp, message number, node identification, and message reason. The message timestamp indicates when the message was created, not when it was received, so that network latency does not create misrepresentations of event times. The message number increases monotonically for each message sent, allowing the data management system to identify missing messages and avoid vulnerability due to node isolation. The message reason indicates, in part, what the data management system should do with the data sent. The metadata TLVs are followed by node-specific TLVs; for example, a temperature sensor could have a Temperature TLV, with the value encoded as a four-byte floating point number indicating degrees Celsius. In addition to the metadata and node-specific TLVs, digital signature TLVs can be added to the message if authentication is used and the entire payload can be encrypted prior to being sent. An ID and length header is appended to the TLVs to ensure proper transmission of messages and commands.

![Figure 2. SSP Message Structure](image)

**Data Management Architecture**

In order to support evidence in depth, the system must be capable of integrating data sets from disparate sources and correlating them. This support requires three capabilities: 1) the ability to receive data from disparate sensors, 2) the ability to perform actions based on input from one sensor, including triggering another sensor, and 3) the ability to present the combined data to the user such that it is actionable. This section describes a data management architecture that supports these three capabilities.

Figure 3 shows the current data management architecture implemented in the Chain of Custody (CoC) Data Management System (DMS). The CoC DMS was built as part of a test bed for evaluating technologies that may be useful in future arms control agreements. While the CoC DMS was built for arms control, the architecture is also believed to be appropriate for safeguards applications. The CoC DMS would be deployed to each site that may contain accountable items. Each site can be made up of multiple monitored areas where such items might exist, such as storage bunkers or maintenance areas.
Figure 3. Data Management Architecture

On the left side of Figure 3, a set of sensors and other infrastructure are shown in the monitored area that could support a given arms controls or safeguards agreement. These components could include item monitors attached to accountable items, facility sensors such as cameras, motion detectors, and door switches, and other components such as the tablet in the figure below that support operator interaction with the system. On the right side is the Data Management System. There are two major elements – the Data Collection and Inventory Management (DCIM) Software, which is responsible for receiving data and performing actions in response to this data, and a set of user interfaces for data review and analysis. During an inspection, the Joint Storage and Review (JSR) Software is used by the monitoring party and a host representative to jointly review data gathered during the inspection visit to ensure the system is working correctly. Even without an inspector present, the host is able to monitor the system and analyze all data for compliance with the terms of the governing agreement via the Host Monitoring, Analysis, and Validation (HMAV) Software. The HMAV Software contains a Data Storage Component (DSC) that stores messages from the system into a relational database and a traditional web application providing the user with access to the monitoring data. The data from each site is then transmitted to a central collection location (the National-Level Reporting System or NLRS) where regular data exchanges can occur with the monitoring party. The monitoring party would presumably have their own HMAV Software equivalent for analysis of the combined data. Analysis and visualization of data is discussed in detail in the next section.

It should be noted that this architecture could be reconfigured to support additional needs. For example, an application similar to the HMAV Software could exist for use by the monitoring party during
an inspection. The JSR Software is a simplistic verification tool with no analytic capabilities, so something more capable may be required. This may, in turn, require a host review component to support a classification review before data is accessible by the inspector. In a safeguards deployment, the JSR Software may not be needed and the NLRS might be replaced by a system provided by the monitoring party.

Internally, all components of the CoC DMS are multi-threaded, event-driven applications except for the HMAV Software, which is a traditional database-connected web application. The internals of the DCIM Software are shown below. The DCIM user interface is intended to be used by researchers as a diagnostic tool and would not be part of a deployed system.

![Figure 4. DCIM Internal Architecture](image)

Following data through the system, an SSP Message is generated by a sensor and sent to the DCIM Software. It is received by the SSP TCP Bridge. A Bridge is how the DCIM communicates with the outside world. The architecture allows new Bridges to be easily added, addressing the need to integrate disparate data sets. While the DMS currently only supports SSP over TCP, it was designed with the need to support SSP over other protocols, such as UDP or a serial interface. The architecture could also support protocols other than SSP, though the integration effort is greater for non-SSP protocols because more translation of node data would be required. A Bridge converts data received from a sensor (or any
other component) to an Event. It also converts outgoing requests like SSP Commands from Events to data to be sent to the sensor.

Events are central to the DMS architecture. As an event-driven architecture, all internal logic is performed in response to an Event. Many events can be received at the same time, so the DMS architecture is multi-threaded. Each Event is assigned to a thread obtained from a pool. The event is then processed by a sequence of Event Handlers that perform the required logic. This allows new Event Handlers to be added or the sequence to be reconfigured based on the needs of the system. The flexibility in the data management architecture is necessary because of its use in a research and development test bed. In a deployed system, the Event Handler configuration would be locked down to aid in authentication and certification of the software.

Event Handlers can perform a wide variety of actions. Most critical to this discussion is that an Event Handler can generate new Events in response to an incoming Event. For example, an Image Request Handler exists that will request imagery from cameras in response to an Event of interest. For example, a “door open” event might cause a request to capture imagery for a timespan that would show activity in a monitored area when the door was opened.

Analysis and Visualization of Data
While the DMS offers different data analysis capabilities and user interfaces at different points in the system, the discussion in this section focuses on a user who is assessing the data from a monitoring standpoint – either a monitoring party verifying accuracy of declarations or a host reviewing data prior to sharing it with the treaty partner.

A user will want to analyze the data with respect to four categories of questions:
- Is the data complete and authentic? If data is missing, what time periods need to be explored more closely? If the data cannot be authenticated, which events and time periods are affected?
- Do the monitoring system elements appear to be intact and functioning correctly? If not, when did the disruption occur and what can be determined about its cause? What action, if any, needs to be taken?
- Are all declared items accounted for? Has chain of custody been maintained? Is there evidence of noncompliance including diversion or undeclared items? What evidence is available to assess discrepancies?
- Are observed activities and accountable items consistent with declarations? What evidence is available to assess discrepancies?

Addressing the questions in the first two bullets ensures the data is complete and correct or that issues with the integrity of the data are identified before addressing the monitoring questions in the second two bullets. In this system, the sensor messages include a message number; the DCIM includes logic to check for the completeness and, if it has the authentication keys, validates the authenticity of messages. These checks could also occur downstream by the monitoring party and the host party independently.

Given the number and variety of messages that may arise in heterogeneous sensor systems, the user needs some way to sort through the messages and prioritize them. The DMS allows for the specification of sensor-specific business logic to categorize and prioritize messages with conditions of interest (i.e., alert levels). Table 3 lists alert levels used by the DMS.
Table 3. Alert Levels

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Normal events that do not require user notification</td>
</tr>
<tr>
<td>1</td>
<td>Normal events that require user notification, e.g., a door opening or seal open</td>
</tr>
<tr>
<td>2</td>
<td>A non-normal event that may not need immediate attention by the user, e.g., when a battery voltage drops but is still in operational range</td>
</tr>
<tr>
<td>3</td>
<td>A non-normal event that may require immediate attention by the user, e.g., an error or tamper event</td>
</tr>
</tbody>
</table>

To address the monitoring questions as well as many of the system integrity questions, the HMAV software provides a user interface with high-level summary reporting, the ability to drill down to event data in several ways, a timeline of events, and a search tool. The interface begins with a “dashboard” view of four classes of monitoring results – inventory changes, key events, monitoring system status, and item monitor status – as shown in Figure 5.

The figure shows data for a single site with two monitoring locations and a repository for spare item monitors (in this particular concept of operations, item monitors need to be continuously monitored to assure their integrity for future use). The Inventory Changes section captures all of the items that the monitoring system has detected either entering or exiting a monitored area. The table can be sorted by any of the column headings. The Key Events section shows monitoring system messages that have been identified as having a level of 1 or higher. Clicking on an event allows the user to see the full message. The Current Site Status section shows the state of the monitoring system components in each of the storage areas. Each monitoring system element is colored according to its alert level and the alerts roll up to indicate the status at each site. Clicking on the monitoring system element takes the user to more detailed information about the monitoring system element and its history. The Current Item Monitor Status section lists the status for each item monitor at the site.

The simple sorting, prioritization and visual displays in the dashboard allow the user to 1) quickly compare accountable item and item shipment declarations to the monitoring data and 2) identify events or time periods of concern.
The HMAV offers more detail on items, monitors, and events through alternative tabs in the interface. The search tab also allows the user to construct queries to look at specific types of messages, sensors, monitors, etc. Once the user identifies time periods of concern, the Timeline tab provides a way to look at all of the events, including images, occurring within a specific time period. This can be very helpful in assessing the cause of anomalies or discrepancies in the data. An example of a timeline display is shown in Figure 6 below. The time scale is shown on the black line near the top of the display. Alerts from item monitors, such as the seal tamper alert, are shown as with lines to the appropriate time on the time scale. Facility sensors events have been aggregated into periods of activity, shown as bars below the time scale. Icons indicate which type of sensor it was and when the sensor was active. Active alerts are also indicated with a bar. In the figure, the top bar is a door switch, the next two are motion sensors, the fourth bar is an active alert, and the last four bars are cameras. Available images are shown below the timeline display. The user can use this capability to assess alerts by scrolling through the data near the time of the alert and using the item monitors, facility sensors, and cameras for evidence.
The DMS addresses collecting and analyzing data at the site level. A site is likely to be a part of a larger enterprise. Further analysis and visualization tools are needed to ensure consistency across multiple sites.

Future Concepts
The previous sections identified four system functions that are necessary to create confidence from heterogeneous data:

1. Acquisition – acquiring heterogeneous data from multiple sources
2. Fusion – integration of heterogeneous data into a consistent and useful representation
3. Analytics – processing the data to come to useful conclusions to support decision making
4. Visualization – facilitating the study of data in a visual form so the analyst can intuitively draw conclusions and make decisions

The CoC DMS performs each of these system functions to varying degrees. Future research and development efforts will augment and improve the DMS in each area.

Data Acquisition and Fusion
As discussed in previous sections, the DMS uses SSP to standardize the data format from multiple heterogeneous sensors. As the arms control and safeguards requirements for this system continue to expand, there will undoubtedly be new sensors that must be integrated. Some will be new sensors using
SSP for integration, but with new data types and representations never before seen by the DMS, which will necessitate the creation and standardization of new TLVs. New representations of these data will need to be created in order to properly fuse the data for further analysis. Another challenge will be the integration of the DMS with legacy sensors and systems that do not natively understand and use SSP. Novel DMS Bridges will need to be implemented that handle the legacy data format and convert the data to SSP in real time. A final issue to consider is the need to combine data from sensor arrays to create an integrated representation of physical reality. For example, the data from an array of radiation sensors could be combined to create a volumetric map, which could provide the analyst information on movements of sources.

**Advanced Analytics**

The current DMS has some capability to correlate events to construct a narrative over time. The system, however, does not have enough analytical capability to automatically find trends, patterns, and anomalies buried in massive amounts of data. SNL previously developed a Knowledge Generation (KG) software package that can automatically identify anomalous sensor behavior and deviations from treaty declared activity (DeLand 2000). KG uses state machines to model sensor behavior and compares sensor input streams to the logic of the state machine, flagging any deviations from expected behavior.

Current research is investigating the integration of KG into the DMS to build upon its capabilities. A limitation of the current KG system is that anomalous behavior can only be discovered via the logic programmed into the state machine. Thus, if the state machine designer cannot anticipate an anomaly, pattern, or trend, KG will be unable to find it. By integrating machine learning algorithms into KG, it is possible to train KG on data generated from scenarios. An analyst will, thus, be able to accept or reject flagged events (trends, patterns, and anomalies) proposed by KG, and via the machine learning algorithms, it will be possible to improve performance over time.

**Advanced Visualization**

The DMS HMAV provides a comprehensive web-based interface to allow an analyst different ways to visually analyze textual and symbolic data. As the DMS faces increasing volumes of data, it will be necessary to discover novel methods to display data in an intuitive way for analysts. Multidimensional data presents a special challenge, especially as the number of dimensions increases past three. Research is needed so that data may be presented such that trends, patterns, and anomalies can be intuitively presented to the analyst so that important conclusions may be drawn. A future concept worth considering is novel ways to combine and visualize sensor data to create additional useful information for an analyst. For example, camera data from a calibrated camera array could be combined to create a 3D volume which could be interactively manipulated by an analyst, presenting an intuitive way to see behind occlusions in traditional 2D imagery.

**Conclusion**

Flexible, extensible sensor systems that enable evidence in depth provide significant capabilities for combining the strongest technical capabilities into monitoring systems capable of providing confidence that agreement obligations are being met. The data system architecture, including data semantics and communication, is critical for success. Future endeavors will add significant and novel capabilities into data acquisition, fusion, analytics, and visualization.
Works Cited


Presenting Load Cell Data in 3D to Improve Situational Awareness during Gas Centrifuge Enrichment Plant Inspections

Jim Garner, Ross Snow, and Michael Whitaker  
Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831

International Safeguards authorities face the challenge of achieving independent safeguards conclusions at many complex facilities. These facilities may have tens or hundreds of tanks, miles of pipes, and numerous feed and withdrawal stations. Facility staff working at these plants every day inherently have detailed, direct knowledge of interconnections between the equipment. Conversely, inspectors—who visit many different facilities and may visit each individual facility only a few times a year—have a more challenging task to understand how material is transferred in and between material balance areas. Currently, safeguards authorities are using two-dimensional (2D) representations to assist in the understanding of these complex facilities by using software packages like Data Analysis and Interpretation and Solution Monitoring System. Highly interactive three-dimensional (3D) environments of the complex facilities are being investigated for future use. The authors believe coupling 3D models to the process-related data further improves an inspector’s situational awareness and effectiveness. The authors have coupled these immersive environments with real data using a commercial off-the-shelf (COTS) software solution. The authors will describe how the COTS Iconics Genesis package has been used on Oak Ridge National Laboratory’s (ORNL’s) feed and withdrawal system and how similar systems could be used to enhance training, pre-trip inspection preparation, and data review—whether on-site at a review station, on the plant floor, or at headquarters from data retrieved by remote data transmission.

Oak Ridge Feed and Withdrawal System
ORNL has installed a feed and withdrawal facility to test process-monitoring technologies (Fig. 1). The facility uses water as a surrogate for UF₆ to simulate operations at a gas centrifuge enrichment plant (GCEP) feed and withdrawal (F/W) area. Water is pumped from three feed tanks, through a mock process, and then into three product tanks and two tails tanks. Each tank rests on a load cell platform that measures its weight. This facility has been described in previous literature [1-2].
Fig. 1. The mock feed and withdrawal area at ORNL pumps water from feed tanks through a mock process and into product and tails tanks. It is used to evaluate process-monitoring technologies.

**OPC-UA Based Process-Monitoring Architecture**

The data acquisition and data handling capability of the Oak Ridge feed and withdrawal system was implemented with the standards-based OPC-UA\(^1\) services and information exchange model. Each F/W station consists of a load cell connected to a weigh indicator that continuously measures weight placed at that station. An OPC-UA communications server, KEPServerEX, is used to communicate with each weigh indicator and source that weight information to other software or systems by publishing OPC and OPC-UA tags. Other software applications can subscribe to these tags to obtain the associated data. Live weight data are stored using the Canary Labs Enterprise Historian. The Iconics Genesis64 software package can retrieve live data directly from KEPServerEX or display historical data from the Canary Labs Enterprise Historian (all via the OPC-UA data interconnect mechanisms). Fig. 2 summarizes this data flow and shows how MATLAB, SQL Server, and SQL Server Reporting Services have been integrated to analyze data, store results, and present information to users. The automated analysis of load cell data using this architecture has been discussed in previous literature [3-6].

\(^1\)Object Linking and Embedding (OLE) for Process Control Unified Architecture
Fig. 2. Commercial off-the-shelf (COTS) software using OPC-UA and OPC communication protocols was implemented at the ORNL feed and withdrawal facility. This system demonstrates the power of COTS solutions using industrial communication standards to provide interoperable architecture that is flexible and scalable.

ICONICS GENESIS64

Iconics, Inc., packages several products for the human machine interface (HMI) and supervisory control and data acquisition (SCADA) markets under the name Genesis64. These products are built on top of the Microsoft .NET framework and can be deployed using Microsoft Windows Presentation Framework or Silverlight or can be integrated into SharePoint solutions. GraphWorX64 from Iconics, Inc., allows users to build scalable, vector-based graphics that do not lose details when enlarged. The authors have used GraphWorX64 to build intuitive graphics that depict real world locations like the ORNL F/W system as shown in Fig. 3. These 3D environments can also be integrated with other Iconics products like TrendWorX64 viewers to allow users to get a more detailed view of historical information through rich charting capabilities.
Fig. 3. Iconics Genesis64 enables users to create 3D models that can be animated with the latest\(^2\) process data. This display can be published using standard web browsers like Internet Explorer and can be rotated or zoomed into using the tip of a finger on a touch screen-enabled computer. Here ORNL’s feed and withdrawal system is represented. The weight of each tank is displayed by the green level. When a user hovers over a tank the current weight is displayed. A user can navigate to a plot of historical data by clicking on a cylinder.

By clicking on a tank in the 3D view, a user can navigate to charts displaying historical data, as shown in Fig. 4. Iconics offers other products as part of this suite of tools for predictive fault detection and diagnosis, alarm and event management, and integration with geographic information systems.

\(^2\)In the case of GCEPs, process weight data may be updated only once per hour to an on-site data store and may never be transmitted off-site.
Fig. 4. From the 3D environment, a user can click one tank to drill down for more detailed evaluation of the data. Here the Iconics TrendworX package is used to plot the historical data. A cursor can be used to show the exact weight in each cylinder at a given time. At the bottom of the plot, a summary view allows a user to quickly navigate in time to periods of interest.
The standards based approach described above using OPC and OPC-UA has enabled the authors to extend the live display of 3D data to allow users to play back historical data. As shown in Fig. 5, a user interface similar to that of a video cassette recorder (VCR) was built that allows historical data to be visualized in the 3D model. As described in Fig. 6, the user controls write to OPC tags that are monitored by a MATLAB script. The MATLAB script reads historical data from Canary Labs Enterprise Historian and plays this back, writing to tags in KEPServerEX. The weight and historical time is read by GraphWorX and displayed.

Fig. 5. The 3D environment can be modified to display historical data and can be played back to visually show how material was transferred between tanks.
SAFEGUARDS IMPLEMENTATION
This initial implementation was designed as a proof of concept and demonstration of integrating data with a 3D model. This architecture could be extended for more complex environments that require authenticated data from the source or remote monitoring. Although OPC-UA provides robust security mechanisms, the OPC-UA protocol does not directly align with the emerging international safeguards standards for data authentication. Custom software could be written to generate data files that are then digitally signed for remote monitoring transport. Once these 3D models are populated with real data, they can be used as a tool to show inspectors normal operating practices. ORNL also has developed 3D models and coupled them to simulated data. These types of tools could demonstrate what off-normal practices might look like.

For the application at GCEPs, the flows in the feed and withdrawal area may be relatively simple and better displayed as a plot, but a 3D representation could be tied to mailbox declarations to show declared occupancy status and carried with the inspector during short-notice random inspections of the feed and withdrawal area.

Fig. 6. Dataflow related to VCR functionality.
For other facilities, if data can be transmitted remotely, the 3D representation populated with real data could help an experienced inspector describe to someone less familiar with a particular site what normal operations are and actually show—not just describe—what has recently occurred.

CONCLUSION
The raw data by itself can tell a story in very technical details if the operator has the tools to decipher it. Iconics Genesis64 is just one example of how COTS software can be used to create 3D models to assist inspectors in better understanding the raw data. These types of tools would allow inspectors to generate “snap shots” of the current process conditions that would provide valuable information in conducting Short Notice Random Inspections (SNRIs) or other unannounced inspections. Three dimensional models showing past operations could provide inspectors with new capabilities for verifying operations as declared between inspections. The automatic “pre-analysis” of process data coupled with the remote transmission of agreed upon key performance indicators would also enable the IAEA to optimize its SNRI schedule.

REFERENCES


Insights from Information Analysis and Visualization in Support of the Global Nuclear Detection Architecture

Robert T. Brigantic and Daniel C. Skorski
Pacific Northwest National Laboratory

Paper not received at time of publication.
SIGNATURES OF ILLICIT NUCLEAR PROCUREMENT NETWORKS: AN OVERVIEW OF PRELIMINARY APPROACHES AND RESULTS


jennifer.webster@pnnl.gov
Pacific Northwest National Laboratory
Richland, WA 99352

Abstract
The illicit trafficking of strategic nuclear commodities (defined here as the goods needed for a covert nuclear program excluding special nuclear materials) poses a significant challenge to the international nuclear nonproliferation community. Export control regulations, both domestically and internationally, seek to inhibit the spread of strategic nuclear commodities by restricting their sale to parties that may use them for nefarious purposes. However, export controls alone are not sufficient for preventing the illicit transfer of strategic nuclear goods.

There are two major pitfalls to relying solely on export control regulations for the deterrence of proliferation of strategic goods. First, export control enforcement today relies heavily on the honesty and willingness of participants to adhere to the legal framework already in place. Secondly, current practices focus on the evaluation of single records which allow for the necessary goods to be purchased separately and hidden within the thousands of legitimate commerce transactions that occur each day, disregarding strategic information regarding several purchases. Our research presents two preliminary data-centric approaches for investigating procurement networks of strategic nuclear commodities.

Pacific Northwest National Laboratory (PNNL) has been putting significant effort into nonproliferation activities as an institution, both in terms of the classical nuclear material focused approach and in the examination of other strategic goods necessary to implement a nuclear program. In particular, the PNNL Signature Discovery Initiative (SDI) has codified several scientific methodologies for the detection, characterization, and prediction of signatures that are indicative of a phenomenon of interest. The methodologies and tools developed under SDI have already been applied successfully to problems in bio-forensics, cyber security and power grid balancing efforts and they have now made the nonproliferation of strategic goods into a challenge problem for testing their methodology and tools.

As a first step towards the detection and characterization of illicit procurement networks, our research examines procurement networks as defined by a system of entities (people or companies) that enter into transactions of specific items with one another. Once we have defined such networks, we are interested in answering questions about the behavior and characterization of such networks.

The questions we wish to answer regarding procurement networks are, first, “Can we detect networks within large, noisy datasets?” and second, “To what extent can we compare multiple networks and identify their similarities?” As procurement networks can be naturally viewed as a graph, we have employed several graph analytic tools to aid in these tasks. In particular, Graphscape, an SDI tool, uses a novel method to approximate edit distance, a graph distance measure based on the number of changes needed to transform one graph into another, in order to measure how similar two given graphs are to each other. Given a set of graphs where vertices represent companies and edges represent a shipment from company A to company B, we can calculate an all-for-all comparison of graphs. In this way, we are
able to determine which graphs are most similar, and which require more changes to transform one into the other. The set of graphs to be compared can be further specialized to provide more insight, e.g., using different time periods to explore events in a company life cycle.

The research team is still exploring the best methods for network characterization. For instance, GRADIENT, an SDI graph analysis tool that examines the ease of movement throughout a network, may be useful for understanding relationships between companies, addresses or commodities. Similarly, MLSTONES, another SDI product, may allow us to better characterize the purchase patterns of companies by treating shipment information as a combination of network flow analysis and fuzzy string matching. The team will examine the potential applications of MLSTONES and other techniques in the coming months.

1 Introduction

Strategic nuclear commodities are defined as those goods required for the development of a nuclear program, ranging from nuclear facilities and special nuclear materials, to dual-use equipment that can be used for nuclear or non-nuclear purposes (which are usually subject to export control regulations), to solely non-nuclear goods. The trafficking of strategic nuclear commodities poses a significant challenge to the international nuclear nonproliferation community. Export control regulations, both domestically and internationally, seek to inhibit the spread of strategic nuclear commodities by restricting their sale to parties that may use them for nefarious purposes. However, export controls alone are not sufficient for preventing the illicit transfer of strategic nuclear goods.

This project attempts to address one of the challenges that is unmet by current regulatory controls – namely, how to identify, characterize, and predict the intent of procurement networks that are involved in transactions of strategic commodities. For our purposes, a procurement network is defined by all of the entities, commodities, and other related data, which can be gathered from import and export transactions between such buyers, sellers, shippers, and other brokers. Ultimately, our research will study these procurement networks in the context of the trafficking of strategic commodities. However, to answer initial questions about the identification and characterization of networks, we examine legitimate international commerce. Legitimate commercial information provides a rich data environment for algorithm development and any base line for normal commerce discovered along the way can then be compared to the signatures produced by illicit transactions. Thus, for the purposes of this study, we will focus on building out our methodologies by examining commercial procurement networks in the automotive industry.

In this paper, we will discuss the current state of export control in the United States and then investigate two major questions from a data-centric perspective: (1) How can we identify and characterize procurement networks from within the greater set of all maritime shipping data? and (2) How can we compare such networks to one another while respecting the context and roles of companies within a larger network? To answer these research questions, we will utilize maritime shipping data [18] and explore methods for using existing data in order to identify and compare such procurement networks using graph analytic techniques developed under Pacific Northwest National Laboratory’s (PNNL) Signature Discovery Initiative (SDI).

2 Background on Current Export Control Practices

The U.S. export control system is based on the issuance of licenses from the Department of Commerce, the Department of State, and the Department of Treasury. When a license request is submitted,
relevant government agencies are under pressure to make a quick and accurate decision on whether to issue a license. Lengthy reviews have the potential to interfere with trade and shipping schedules and negatively impact international commerce. However, making a quick decision could result in the export of sensitive information or technology and threaten national security. In addition to these pressures, expert analysts in export control are inundated with an enormous amount of license requests and transaction data. Thus, nefarious actors have an advantage in that they not only have a geographically large region in which to hide, but they can obscure their purchases within the even larger data space. In assessing the legitimacy of an export license application or assisting customs agencies with border level enforcement of trade regulations, export control analysts are concerned with two key areas: 1) identifying the commodity and discerning the end-use of the product and 2) identifying the ultimate end-user. Dual-use strategic nuclear goods can be used for either nuclear or non-nuclear applications, and are often export controlled. As such, dual-use commodities present significant challenges to export control enforcement agencies when examining transactions individually. Furthermore, nuclear proliferation efforts require a combination of these dual-use items and other more common commodities. By considering the strategic view of a series of all such commodities being consolidated in a single location by a single entity, this study hopes to aid experts in the identification and classification of procurement networks that deal in export controlled strategic nuclear commodities, while utilizing information from the larger body of trade to better inform such analysis.

The United States has developed several mechnaisms to support the export license application review process, including creating lists of export controlled commodities and commodity identification through manual database queries. Additionally, each government agency involved in export control regulation and enforcement contributes to a consolidated export control list, which aides export control analysts in restricting exports to certain individuals or organizations. To develop these lists and enforce regulations, export control practitioners spend significant time and effort reviewing large and diverse data. Proliferation networks that obfuscate the end-use or user, the combination of legitimate and nefarious trade in dual-use goods, and intentional or unintentional bypassing of the export licensing system all undermine efforts by the United States to enforce its export controls. As the world continues to evolve, the United States will have to become more efficient and effective in its detection of procurement networks of strategic nuclear commodities.

Many studies have been conducted in an attempt to strengthen export control regimes. Several authors analyze export controls from the international perspective, they identify weaknesses in current regulatory and monitoring approaches and make policy recommendations to correct those weaknesses [6, 16], or forecast future export control challenges [8]. Some research is more tailored to improving the current U.S. system [13, 17, 19], however, the bulk of the literature is conducted on improving the enforcement of export controls by defining and identifying illicit procurement networks.

One of the key debates among authors in this field is determining if illicit networks actually “look” different than legitimate networks [8]. A general conclusion among the literature is that there isn’t a way to tell illicit and legitimate networks apart when focused on a single commodity [7]; there are limited studies available focused exclusively in this area [14]. Instead, authors argue it is more important and impactful to identify the procurement network and actors involved in it [20]. Additionally, several publications maintain that globalization and technology have effectively changed the way we need to look at trafficking, and identifying a few links in a proliferation network will not be enough to shut down a network entirely [4, 11]. Several authors have identified actions that could be considered “red flags” in identifying illicit procurement behavior [3, 15], but these articles fall short in identifying networks as a whole. This study takes a datacentric approach to examining several of the key aspects in this debate,
particularly those regarding the nature of procurement networks, as will be discussed in the following sections.

3 Identification of Networks within Shipping Information

The first question this study considers in the pursuit of locating illicit trafficking networks is “How can one identify any procurement network within large quantities of shipping transactions?” To answer this question, one must know what is meant by a “procurement network.” For this study, a procurement network is defined by all entities, commodities, and transaction information related to the manufacture/processing of an item of interest. However, the expression of this information requires some careful thought. This project has chosen to consider procurement networks by formulating the shipping transaction data as a graph of nodes and edges, for which there are many mathematical techniques that one can use to characterize and compare different networks.

To date, this study has used two different views of transaction data – company-based and commodity-based. The company-based view, in which the nodes of the graph are given by the companies involved in each transaction (shipper, consignee, and notification parties) and edge weights are given by the number of transactions between the two parties, has been explored by the research team and will be used as the model for the remainder of the paper. For a single transaction, three edges will be created/weighted higher based on the relations between shipper-consignee, consignee-notify party, and shipper-notify party.

In our experiment to discover legitimate procurement networks, we began by examining maritime trade into and out of the United States. As an example of a legitimate procurement network, the team selected the automobile manufacturing industry to facilitate development of proof-of-concept methodologies. An initial large graph was created by searching the shipping record database for companies containing the string “MOTOR,” and then used transactions where such companies were either the shipper, consignee or notification party to create a company-based graph view of the data. This included companies like “FORD MOTOR CO” as well as “BAYERISCHE MOTOREN WERKE AG” (more commonly known as BMW). A fundamental assumption is that this “MOTOR” graph, which we will refer to as \( M \), consists of many procurement networks, some of which may occasionally overlap or connect to one another. The automotive manufacturing industry example will be utilized throughout the rest of the paper as methodologies for identification, extraction, and comparison of procurement network signatures are developed.

By using this natural model for transaction data, one can concentrate on extracting information regarding individual procurement networks from the larger body of trade data. In the following sections, we will discuss two methods of separating the larger graph into the individual networks - search by known network and edge betweenness community detection.

3.1 Extraction via Search by Known Network

One method for extracting the procurement networks from this larger graph is for a subject matter expert to search out specific, known companies and extract all transactions where this company is involved. This involves choosing a specific string identifier that one would find in the company name. For instance, in extracting a graph of Toyota’s procurement network, we use the string “Toyota” to encompass all records that may be listed in alternative notations, such as “Toyota Inc.” or “Toyota Company.” This methodology for extracting graphs is incomplete in that it:
• Requires foreknowledge of at least one company involved in the transactions and thus in the network,
• Can only provide one level of company interactions as it forces one of the endpoints to be the known company, and
• Does not necessarily show interactions between suppliers of the known company.

In order to overcome some of the barriers listed above, the research team turned to edge betweenness community detection.

3.2 Edge Betweenness Community Detection

The central tenant of community detection algorithms in graph theory is that characteristics of the graph (irrespective of the data it represents) can be used to fracture a large graph into smaller communities. One such characteristic is betweenness, which is a mathematical measure of the strength with which a single element in a graph is related to the rest. Colloquially, if one considers a graph as a network for roads (edges) and villages (nodes), a betweenness score allows one to distinguish between highways or cities (those elements with a high betweenness score) and country roads or villages (low betweenness score).

Thus, given a graph, $G$, one can ask the question: which edges are more or less traveled? In other words, if one were to walk along all of the shortest paths between every pair of nodes in a graph which edges would one travel more often or less often? This concept is formalized in edge betweenness. For each edge, one calculates the number of shortest paths the edges falls on as a fraction of the total number of shortest paths. Given an edge $e \in G$, one then defines the edge betweenness, $b(e)$, as

$$b(e) = \sum_{v,w \in V} \frac{\sigma_{vw}(e)}{\sigma_{vw}}$$

(1)

where $\sigma_{vw}(e)$ is the number of shortest paths from $v$ to $w$ that pass through edge $e$, $\sigma_{vw}$ is the total number of shortest paths from $v$ to $w$, and the sum is taken over all of the pairs of vertices in $V$.

Edge betweenness community detection posits that edges with high betweenness are more likely to be bridges between different communities [10]. Therefore, if all such edges above a certain threshold are removed, then the graph will be disconnected into its constituent communities. This team has used a library called igraph within R to implement edge betweenness community detection in the graph of the automotive industry, $M$, which we will consider further in our Preliminary Results section. In our study of maritime shipping data, edge betweenness community detection eliminates ties between communities that are created by single, rare shipments, while identifying communities by the stronger trends of frequent shipments. This breaks the larger transaction graph into smaller networks of entities that are involved in trade frequently with one another. This approach allows researchers to process large graphs with an unknown quantity of procurements networks within them and attempt to automate the detection of procurement networks through the isolation of communities.

4 Network Comparisons

Now that methods have been developed to identify individual procurement networks, one might naturally consider to what extent one can compare multiple networks and identify their similarities. This general question can be broken down into several smaller tasks, i.e. comparing how closely related two procurement networks are in terms of manufacturing goal or company structure; identifying unusual
transactions in one network based upon normal trade patterns for the industry as a whole; or examining how much a network changes during various stages in a company’s life cycle (such as startup of a new facility). Once shipping transactions have been mapped into a graph and individual procurement networks have been identified, one can use a well known measure of graph edit distance (GED) to calculate the similarities between two graphs \([1, 2]\) and attempt to answer these questions.

4.1 Graph Comparison Approach

True GED cannot easily calculated for graphs with any significant size (>12 nodes) due to its time complexity. Therefore an approximation method must be identified if the aim is to leverage GED as a similarity measure. Many techniques have been developed to approximate edit distance \([9, 12, 21]\) which aim to speed up the calculation and to define upper or lower bounds for the GED. A major drawback to these techniques is that they generally focus on the structure of the graph, and do not inform the distance with attributes. For example, given two fully connected graphs with 4 nodes each, one containing social network information and another containing molecular information, a typical GED approximation will estimate them as equal due to their structure. This type of similarity in our problem space is not as informative as a distance measure that does utilize graph attributes for comparison. To approximate GED while utilizing graph attributes, we leverage a PNNL software called GraphScape which allows node comparison to be weighted based on domain-specific attributes. This provides a means to differentiate graphs not only on their structure, but their content as well. In the case of international maritime shipping data, this team has chosen to include information about companies and their respective transaction commodities as attributes for each node.

When comparing two nodes from separate graphs, it is important to decide what similarity means for the problem space. This comparison is largely defined by the metadata attached to each node, which in our case is the company name and their commodities. In some industrial domains, the node/company name may be of high importance, while in others the job family for the node/company is more important. We evaluated the available attributes in the data set and determined two features that should be used for initial evaluation, the name of the company and the commodities that company sends or receives. Node names have some bearing on the similarity problem, but there are obvious situations where the name of the company is less important than its function. For example, if two companies send commodities to the same company as an intermediary, then that similarity should be of more importance than if the two companies shipped similar commodities to two separate companies. Company names were compared using Levenshtein distance, allowing a similarity for the names as well as the graph \([5]\). Having such a “fuzzy” measure of similarity between company names is important for the data used in this analysis since many of the values are hand-entered and misspellings and various forms of abbreviation are commonplace. This makes the data unsuitable for many string matching comparison algorithms and can obscure relations in the data which drastically affect the graph structure.

The other measure used in this study for determining graph similarity is the list of commodities sent or received. For each node, we constructed a set of commodities (including harmonized tariff codes and free-text commodity names) that were part of a shipment. That set was then added as an attribute to a company. When doing a node comparison between \(v_1\) with commodity set \(S_1\) and \(v_2\) with commodity set \(S_2\), commodity similarity was given by \(|S_1 \cap S_2|/|S_1 \cup S_2|\), i.e. the number of commodities common to both lists divided by the total number of unique commodities in both lists. This component of similarity allows comparison of companies in regard to what types of materials and commodities they
deal in. The hypothesis is that this mechanism will allow comparison of companies that are similar in nature (such as car companies), but are geographically dissimilar or have different names.

5 Preliminary Results

Our initial efforts in the examination of the automotive industry have involved both the identification of individual networks, and comparisons between them. In the results shown below, we begin with our graph of the automotive industry, $M$, containing all transactions where one company name contains the string “MOTOR.” This graph was then fractured into subcommunities via edge betweenness community detection. The result was a large set of graphs, many of which were very small, but some were moderately sized graphs which this team hypothesizes could be procurement networks. In each of these resulting networks, the two vertices with highest degree were noted and then used to create “ground truth procurement networks,” $T$, using the search by known network methodology, corresponding to that set of high degree vertices. This effort resulted in two sets of graphs: the set of all ground truth procurement networks created by searching our database for specific companies ($T$), and the set of all pieces of $M$ achieved by using edge betweenness community detection ($D$). A subset of these graphs are shown for two automotive companies in Figures 1 and 2.

To illustrate our results, we will examine graphs from two particular companies – Harley Davidson and Interamerican Motor. In each case, we have a single graph that is generated by edge betweenness community detection and two that are generated via a search by known network. In the case of the search by known network graphs, the few highest degree nodes from the edge betweenness community detection graph were identified and used to search in the database for all other transactions in which it was involved.

![Harley Davidson Comparison](image1.png)

Figure 1: Harley Davidson Comparison: Left most graph created via edge betweenness community detection, central and right most graphs generated via search by known network utilizing nodes of high degree from the edge betweenness community detection as initial search terms.
As can be seen in Table 1, the results of graph edit distance comparisons provide a high confidence in identifying graphs of similar structure and content as the edit distances for the search by known network graphs are low for the edge betweenness community detection graph which led to its creation as well as to the other search by known network graph created from that same graph. Graph21 and graph22 (central and right images from Fig. 2) were based off of information from the Interamerican Motor graph (on the left of Fig. 2) where as graph15 and graph16 (central and right images from Fig. 1) were based on information from Harley Davidson (on the left of Fig. 1). Visually, one can recognize the similar structures in each of the graphs created for a single company as well. When a graph is compared to a similar graph created to depict another company (i.e. “Harley Davidson compared to Interamerican Motors), the edit distance increases to reflect the difference in companies involved in the network as well as the different procurement structures. However, more work is needed to investigate node comparison methods that do not rely on such strict string comparison qualities and reducing the role of the size of the graph in the comparison metric. As mentioned previously, we are continuing to work on the role of the shipping record data in the similarity metric. Upon expanding our graphs beyond initial samples, we came to realize that it is difficult to compare distances between the various graphs. If two large graphs are very similar, they may have an edit distance of 130 (as below) whereas two small graphs (under 130 nodes & edges) will have a lower edit distance, but may be completely dissimilar. The graphs shown here are all roughly the same size and, therefore, allow us to compare edit distance without dealing too sharply with graph size. This challenge will need to be addressed before moving forward.
6 Conclusions and Future Research Directions

Initial studies into procurement networks and their distinguishing features using methodologies generated under SDI have proved fruitful. With the community detection algorithms, we have successfully begun the process of fracturing the large shipment flow into smaller procurement networks. GraphScape and its edit-distance comparison allow us to compare business practices in terms of the companies involved and the commodities that they trade. There are still areas for improvement – in particular, the exact calculations of similarity metrics with respect to company comparisons and graph size.

There are several other avenues for examining procurement networks and their similarities that we intend to pursue in the coming months. First, we are intend to expand upon the graph based approaches by using this formulation of the problem to answer questions like “How can we examine changes in a procurement pattern and correlate them with influential factors that will not be explicitly listed as a transaction, e.g. the opening of a new factory or the beginning of illicit activity?” SDI’s Gradient tool may allow us to examine this question by finding paths of frequent and close communication among such companies. We also are looking into protein comparison based methods for analyzing procurement patterns of a single company with the SDI tool MLLSTONES – instead of treating the procurement network as a graph that is affected by multiple companies, this approach will treat the commodities as a string of transactions with time as the only free parameter. By looking at the procurements in this fashion, we can examine the motivation of a company and detect changes in the combination and frequency of purchases.

Following the development phase of the algorithms described here, we hope to extend this work to the identification and characterization of procurement networks of strategic nuclear goods. The competitive intelligence community has already demonstrated that one can gain key insights into a business’ motivation and strategy through analysis of that company's purchasing patterns. We believe that understanding trends and patterns in such procurement networks, and thereby being able to reliably detect anomalies, could support the export control community in its assessment of license applications.
as well as its analysis of trade patterns for potential interdiction. Thus, we think the application of our methodology to strategic nuclear commodity data sets and problem spaces will provide enhanced understanding to this domain, and advance PNNL’s mission to support national security.

Acknowledgements
The research described in this paper is part of the Signature Discovery Initiative at Pacific Northwest National Laboratory. It was conducted under the Laboratory Directed Research and Development Program at PNNL, a multi-program national laboratory operated by Battelle for the U.S. Department of Energy.

References


**PNNL Strategic Goods Testbed: A Data Library for Illicit Nuclear Trafficking**

Jennifer B. Webster, Luke E. Erikson, Christopher Toomey, Valerie A. Lewis  
Pacific Northwest National Laboratory  
Richland, WA 99352

**Abstract**

Pacific Northwest National Laboratory (PNNL) has put significant effort into nonproliferation activities as an institution, both in terms of the classical nuclear material focused approach and in the examination of other strategic goods necessary to implement a nuclear program. To assist in these efforts, several projects in the Analysis in Motion (AIM) and Signature Discovery (SDI) Initiatives at PNNL are developing machine learning methodology for human-computer interaction in real time environments to assist analysts in this domain. All of these technical projects require access to data – whether it is in terms of detector data, shipping records, financial information, company relations, or other communications. The first question that mathematical and computational researchers come up with when asked to build analyst assist or automated tools is “What does the data look like?” They become frustrated when basic questions like this can not be easily answered and this can have the effect of pushing researchers away from the nuclear trafficking domain, especially in strategic commodity and export control areas where data sets can not easily be generated through standard experimental techniques.

For small projects that are building a proof of concept for their methodology, obtaining this data can be arduous and expensive. To relieve the burden of data collection from these projects and grow a lab-wide capability, the Strategic Goods Testbed Team has taken over data collection and placed subscriptions and access to flat data files in a centralized location so that all projects can benefit from these items. We have collected shipping data in the form of PIERS records, judicial information about export control cases, NAC data on the nuclear fuel industry, and financial data from Dun and Bradstreet and our data sets are continuing to expand. With a single access agreement, researchers in data-mining and other fields can utilize all of the records that have been downloaded, make requests through subscription services, and interact with other researchers through our interface. Our testbed team provides more than a simply static repository by working with researchers to refine their data needs and insure data quality as well as quantity. We are currently working with laboratory and initiative specific management to examine effective ways for continuing data growth and sustainability.

**1 Introduction and Motivation**

With the continued advances in nuclear science, the disruption of illicit nuclear trafficking has become an area of international interest. Individual nations and industries have invested millions of dollars in the development of both hardware for the detection of radiological sources and analytical capabilities to investigate proliferation activities. Additionally, international organizations have emerged to prevent nuclear trafficking. Pacific Northwest National Laboratory (PNNL) has played a variable role in the enforcement of non-proliferation policies over the years including developments in the radiation portal monitoring program and technical reachback capabilities for export control licensing. Recently, PNNL has invested considerable effort into this area and named the disruption of nuclear trafficking as one of its laboratory level objectives.
As a result of the increased interest, several projects have begun investigating the trafficking of strategic goods and proliferation networks. To clarify, the strategic good problem involves the detection, monitoring and control of items that could be used to start a nuclear program including items on the Nuclear Suppliers Group (NSG) trigger list and dual-use goods. NSG trigger list items are those that are clearly meant for nuclear or radiological applications, such as fuel cells for reactors or medical isotopes. These items are regularly monitored by international agencies as well as country specific entities like the US Department of Commerce. Dual-use items present a much more significant challenge for monitoring agencies as these items have non-nuclear related applications for which they are regularly purchased.

There have been many policy level explorations of the strategic goods problem that have made recommendations on new export control legislation and procedures [1, 3, 4, 5] as well as monitoring of the risks presented by individual countries and their intent to proliferate [2]. In contrast, many new projects at PNNL are geared to take advantage of the laboratory’s strong history of data analysis, machine learning, and visual analytics capabilities. In order to apply these techniques to improving proliferation monitoring and detection, researchers need to locate or create adequate data sources that provide information about dual-use commodities and their potential applications to proliferation activities. Several staff members with interest in machine learning recognized this need and created the Testbed in concert with subject matter experts in the export control and nonproliferation domains. Their initial task was to identify and collect sufficient data to begin applying data mining techniques to this domain and then support projects who chose to do so by refining and adding to the available data.

In one year of Testbed operation, the team has identified several data sources that may provide insight into the proliferation of strategic goods and has begun to support several new projects around the laboratory. In the remainder of this paper, we will discuss the currently gathered information provided by the Testbed, a sampling of the supported projects, and our plans for future growth in this area.

2 Data Sources and Operational Setup

The Strategic Goods Testbed Team has identified a number of data sources including shipping records, financial profiles of companies, judicial information, and subject matter expert experiences. There are three classes of data storage/access that the Testbed makes use of: flat data files, query services, and policy/SME study summaries. We will discuss each of these in the context of our major data sources in the remainder of this section.

These data sources have been gathered in a single repository that can be accessed by any PNNL staff member. By storing everything in a central location, we have mitigated two common problems associated with gathering and using data – first, we have paid for data that would be impossible to afford on a small project and prevented duplication of data as it can be purchased by multiple projects; and second, we have consolidated all of the various user agreement and data restrictions into a single, concise document that outlines all of the responsibilities of data users. This allows smaller projects to utilize these rich data sources without waisting their valuable time and money.

2.1 Port Import Export Reporting Service (PIERS)

PIERS is a comprehensive database including all U.S. commercial maritime exports and imports. PNNL currently maintains a PIERS subscription service and has also purchased several large flat files. The subscription service provides researchers access to all PIERS data from the last five years (rolling start based upon the date of the query). Queries may be submitted to the test Testbed team and Testbed staff will work with researchers to refine their query and provide access to the requested data. Once records are downloaded they are added to the communal record storage for all other researchers to access. The
flat files include several large purchases, including a selection of export records from 2012 and an entire year’s worth of imports from 2013.

This is one of the most frequently accessed data sources for the current Testbed iteration. The PIERS records provide information on commodities, companies, and other pertinent details regarding shipping transactions as specified in shipping manifests. These manifests are generally used for tariff collection purposes, so they provide only high level information, i.e. a general category of commodity rather than a detailed part number or a cost estimate based upon shipment averages as opposed to the exact value.

2.2 NAC International

NAC International is a leading nuclear fuel cycle consulting and technology company. Currently, the Testbed contains three flat files from this company that contain information about worldwide fuel cycle transfers in the enrichment and fuel fabrication sectors. This allows researchers to examine trends in reactor types as broken down by fuel supplier, country and commercial reactor.

2.3 Dun & Bradstreet (D&B)

D&B maintains an industry standard database of corporate information. All information is company based and public domain, however D&B also provides a score for individual companies, which is a proprietary metric. Through the Testbed subscription service, access is provided to both Hoovers and the Global Database. Hoovers is a traditional database system that can be queried for multiple records. The Global Database is similar to PACER in that it is a one request/one result system and gives information on a companies credit history, management personnel, leading competitors, and financial patterns. Any selection of provided information about a specific company can be downloaded via the Global Database. As these data sources are both subscription services, requests can be submitted to the Testbed Team for implementation.

2.4 Public Access to Court Electronic Records (PACER)

PACER is a database of electronic court records that can be used to provide information about specific cases in federal court. Every Federal court and many State courts are represented, including: Supreme, Appeals, National, District and Bankruptcy. To find a specific case, you must have the case number or Party Name in question. For each request on PACER, a $0.10 fee per page is charged, although charges under $15 per quarter will be waved.

Currently, the Testbed has been using this resource to provide more detail on several export control cases mentioned in Department of Commerce Press Releases. By using the limited information in a press release, the Testbed Team has found information regarding time frame and companies involved that may be used to find specific shipping records of interest that correspond to export control violations. These few examples and a documented procedure have been made available to researchers as well as a documented procedure for requesting access.

2.5 Subject Matter Expert Problem Area Documentation

PNNL has several experts working in the realm of export control. Working in tandem with these personnel, the Testbed team has identified a collection of publically available regulations regarding export control restrictions pertaining to commodities and companies. These experts have graciously explained their role in export control and lent their expertise to the preparation of several documents aimed at providing data analysts with necessary context for their efforts. These documents include the commerce control list (regulatory information on commodities), an export control primer, a survey of
press releases regarding export control, and a survey of commonly used tools and techniques for working with the data sources available.

3 PNNL Projects and User Base

Currently, the Testbed team supports 20 individual users in several projects across the laboratory, including teams in two of the major initiatives at PNNL – the Analysis in Motion (AIM) and Signature Discovery (SDI) Initiatives. The AIM Initiative focuses on advancing the science of analytics to enable continuous synthesis of new knowledge in a streaming fashion and dynamic control of measurement systems in real time. Specifically, AIM members are focusing on semi-automated learning tasks that operate on high velocity and high volume data. The disruption of nuclear trafficking – specifically the identification and control of strategic goods as measured in shipping records – was identified as a challenge data set for the projects in visual analytics, incremental machine learning, and inductive reasoning. The Testbed team has provided both shipping data and subject matter expert anecdotal experience to these computational and mathematical analysts in order to familiarize them with the realm of export control.

The mission of the Signature Discovery Initiative is to use robust, repeatable statistical and mathematical methods for the development of signatures that one may use to predict, detect or characterize phenomena of interest. This team has also made the strategic goods problem a priority and has a large scale project working on the identification of illicit procurement networks intending to begin a nuclear programming. The SDI project has made use of the subscription and bulk purchases of the shipping records for both US imports and exports as well as the financial information from D&B databases and judicial information resulting from the Department of Commerce’s export control investigations. The SDI team has provided exceedingly useful feedback to the Testbed members that has guided efforts in maintaining the current data and in procuring new and more specific data to meet the needs of their project.

There are two smaller efforts (in terms of the Testbed team involvement) that have also been supported. The first task is to provide information to those subject matter experts based here at PNNL as they provide technical reachback expertise in the export control licensing and customs and border enforcement realms. Secondly, the large graphs that can be constructed from the shipping data can provide a valuable test set for scalable-parallel graph analysis and machine learning algorithms. We have provided information to such a group based at PNNL to further improvements in high performance computing with graphs.

4 Conclusions and Future Work

Moving forward, the Strategic Goods Testbed Team has a number of tasks, both in terms of administrative needs and data exploration. We are currently integrating with a number of teams and increasing our customer base across the laboratory by regular meetings with initiative leadership, as well as open brown bag presentations to the staff at large. This allows us to familiarize all levels of personnel with the mission and available data resources. In terms of data exploration, we have found several other potential data resources that expand upon the currently available data in shipping and finances. Testbed staff members are currently working to procure them and set up appropriate access control procedures to provide the same ease with which the current data sets are provided.
Acknowledgements

The research described in this paper was conducted under the Laboratory Directed Research and Development Program at PNNL, a multi-program national laboratory operated by Battelle for the U.S. Department of Energy.

References

Trade Analysis and Open Source Information Monitoring for Non Proliferation

Giacomo G.M. Cojazzi, Cristina Versino, Erik Wolfart, Guido Renda, Willem Janssens
European Commission, EC, Joint Research Centre, Institute for Transuranium Elements, Ispra, Italy
Contact: giacomo.cojazzi@jrc.ec.europa.eu

ABSTRACT

With the Additional Protocol in force, the IAEA has strengthened tools to verify the absence of undeclared nuclear activities. The new state level approach being proposed by IAEA envisions an objective based and information driven safeguards approach further utilizing relevant information to improve the effectiveness and efficiency of safeguards. To this goal the IAEA makes extensive use of open source information. Here open source is broadly defined as any information that is neither classified nor proprietary. It includes, but is not limited to: media sources, government and non-governmental reports and analyses, commercial data, and scientific/technical literature.

New sources taken into account include trade-related information. JRC has surveyed and catalogued open sources on import-export customs trade data and developed tools for their use in safeguards. Tests on the use of these data by the IAEA suggest safeguards relevance along the following lines: (i) support the IAEA State evaluation process and improve understanding of a State’s nuclear programme; (ii) verify import and export declarations made by States under Additional Protocols; and (iii) identify indicators of activities to be safeguarded or to be declared under Additional Protocols.

In the field of open source monitoring, JRC is developing and operating a “Nuclear Security Media Monitor” (NSMM), which is a web-based multilingual news aggregation system that automatically collects news articles from pre-defined web sites. NSMM is a domain specific version of the general Europe Media Monitor (EMM) and monitors – additionally to the 4000 general web news sources targeted by EMM- more than 150 nuclear specific sites (including NGOs, academic, (inter-)governmental and scientific/technical sites). Filters remove articles not relevant to the nuclear domain and group them into various areas of interest, e.g. related to different steps of the nuclear fuel cycle or to relevant countries. NSMM has been established in a joint project with IAEA with the aim to streamline IAEA’s acquisition and analysis of open source information and develop the current information collection/newsletter production process to a more efficient system.

The paper illustrates what are the sources of trade data relevant for nonproliferation and the possible uses to inform verification activities and presents the main aspects of the NSMM also by illustrating some of the uses done internally at JRC.

1. INTRODUCTION

As reported in [1], the IAEA “…collects and processes safeguards relevant information about a State from a wide range of sources: information provided by the State itself (e.g. declarations and reports); safeguards activities conducted by the Agency in the field and at Headquarters (e.g. inspections, design information verification, material balance evaluations); and other relevant information (e.g. from open sources and third parties)”. In states with an additional protocol [2] in force the agency has additional mandate and means to verify the absence of undeclared activities and, to this aim can make an even larger use of open source information. Open source is here loosely defined as any type of non-classified of proprietary information and includes, but it is not limited to media sources, government and non-governmental reports and analyses, commercial data, and scientific/technical literature, including trade data.
JRC is investigating information analysis for supporting nuclear nonproliferation since 2007 and actively supports the IAEA through the development of tools and approaches in the context of the EC support programme to the Agency. The paper presents two main research areas in this domain and is structured in two parts. In the first part the possible use of trade analysis to support IAEA verification activities is presented with reference to the type of data available, the approach to deal with them and the tools developed by JRC. The second part presents selected tools developed by JRC for open source information monitoring, their tailoring to the nuclear safeguards and security needs and some of their uses within both JRC and IAEA.

2. TRADE ANALYSIS: SOURCES, METHODOLOGY AND TOOLS

Data services on global trade referred to in this paper are open source, as described in a data catalogue [3] originally compiled by JRC for the IAEA. The data are available either for free or by a subscription fee. The information provided by trade data services has a regulatory origin as it stems from declarations made by traders to national customs authorities. Customs data are collected at national level and, by decision of individual States, published in transactional or statistical formats. 

**Transactional data** are close to declarations made by importers/exporters to customs. Declared data fields subject to disclosure may include:

- A code classifying the commodity traded; e.g., according to the Harmonized System (HS) [4] product nomenclature developed by the World Customs Organization.
- Free text description of the commodity.
- Quantity, expressed in weight or number of items.
- Value.
- Date of shipment.
- Country/port of import/export.
- Party names.

**Statistical data** on trade are derived by aggregating transactions data by country, trade flow (import or export), reference period of time (months, years) and product categories as specified by the HS. As an example, COMTRADE [5], by the United Nations Organisation, provides annual series of trade data for 150 reporting countries. COMEXT [6], by the Statistical Office of the European Union (EU), is a second example. Focused on European reporting countries, COMEXT provides monthly records on EU trade.

Global trade data can be used in the conduct of trade case studies. A case study may start with a piece of information or a hypothesis about trade activities on items subject to safeguards, export controls, or other proliferation-sensitive items. The goal of a case study is to consult web data sources on global trade to:

- Confirm or deny the information/hypothesis;
- Find new, related information.

**Example.** A case of trade in high-precision machine tools with unclear end-use has been reported by the press. Articles report that several high-precision machine tools have been exported from Country A to Country B over a period of a few months. Is the alleged trade confirmed by global trade data sources?
A preliminary step is to identify clearly items of interest to the case study at hand. This is supported by expertise on items’ and reference documents where materials and equipment subject to controls are listed, defined and described.

**Example.** Machine tools subject to export controls are listed in the Nuclear Suppliers Group ‘Guidelines for Transfers of Nuclear-Related Dual-Use Equipment, Materials, Software and Related Technology’ [7]. Machine tools of interest here include turning machines (item 1.B.2.a. in [5]), milling (1.B.2.b.) and grinding machines (1.B.2.c.) for removing and cutting metals, ceramics or composites. The main parameters for a machine tool to meet the specifications given in [7] are (i) the number of angular or rotary axes that can be moved simultaneously with the perpendicular or linear axes and (ii) the overall positioning accuracy of the machine.

With items of interest identified, it is necessary to map these items to HS descriptors in order to retrieve relevant trade data\(^1\). Mapping items of interest to nuclear-related trade to HS is not trivial, because:

- HS codes only approximate targeted items.
- Exporters may (intentionally or unintentionally) declare trade to customs authorities in HS categories aside from the correct ones.

Trade analysts must determine on a case-by-case basis which HS codes to use to best address the case study at hand. A starting point for the selection of HS codes is given by existing correspondence tables. Developed by experts of the HS, these tables map items listed for export controls to HS codes. One such a table is the EU ‘Correlation Table’ [8] CT) developed and maintained by the European Commission, DG TAXUD: CT maps to HS items listed for export controls [9] by EU Member States.

**Example.** Machine tools (turning, milling, grinding machines) are associated to a number of HS codes by the EU Correlation Table. Some are reported in Table 1. None of these codes is specific to export-controlled machine tools. For example, no precision parameter is indicated in the HS explanatory notes. Indeed HS is a taxonomy designed to describe generic categories and the most common from the trade point of view. This limitation implies that trade data retrieved by HS codes need then to be analysed by trade-related criteria that can, in favourable cases, reduce the inherent data ambiguity. See follow-up point in the Example.

With HS codes selected, a plan of queries on trade data services is designed taking into account the questions addressed by the case study and the range of services available on trade data. For example, querying for statistical data requires the specification of the following dimensions:

- A reporting country;
- HS codes related to the items of interest;
- A trade flow (import, export);
- A time period.

**Example.** We query trade data services for Country A exports to Country B. HS codes for queries are listed in Table 1. The time period is a few consecutive months in a given year.

Queries return the list of partner countries for which trade on the query dimensions exist, specifying value and quantity of the trade over time. Results of queries are presented in tabular form for analysis and interpretation. The criteria adopted for the analysis are, in general, specific to the commodities at hand. One informative generic criterion is the ‘trade unit value’, defined as the ratio between value and

---

1 By definition, trade data referred to in this article do not generally include trade undeclared to customs. Some countries include such trade when discovered.
quantity traded for a given HS, in a temporal point (e.g. month or year), between traders (countries or companies). For limited quantities traded, the trade unit value approximates the declared price at which commodities have been traded. Knowing the market price for items targeted in a case study, one can compare it with the trade unit values of retrieved data records. Data points compatible with the market price of targeted items are highlighted as points of interest. These may confirm alleged information or provide new insights in the case study.

Table 1 – Some HS codes for declaring trade of machine tools to customs.

**Example.** Table 2 shows value (USD) and quantity (No of items) of exports declared by Country A to B on HS codes of interest in the time frame of the case study (M1, ..., M5). The data could refer to the alleged trade of machine tools in that the trade unit value is compatible with market prices for export-controlled machine tools. Data were extracted from Global Trade Atlas® accessible at http://www.gtis.com/gta/.

Table 2 – Country A exports to Country B on selected commodities.
**Tests** on the use of global trade data by the IAEA suggest safeguards relevance along the following lines [10]:

- Support the IAEA State evaluation process and improve understanding of a *State’s nuclear programme* – Trade information on exports can support the assessment of a State’s nuclear related industrial capabilities. Data on trade flows between States can be used to understand their international cooperation. Understanding mining-related activities can be improved by using data on the exports of raw materials and semi-finished products. Data on imports and exports of nuclear materials and equipment may also provide information on the development of the nuclear fuel cycle in general.

- **Verify import and export declarations made by States under Additional Protocols (APs),** article 2.a.(ix) [2] – Trade data can prove useful to identify flows of raw material subject to safeguards. Trade categories (of the Harmonized System) appear to be less specific than safeguards categories, but precise enough to be determined as safeguards-relevant. The identification of shipments of some AP Annex II equipment may represent a greater analytical challenge.

- **Identifying indicators of activities to be safeguarded or to be declared under APs,** article 2.a.(iv) [2] – In this context it is foreseen that trade data can be used to verify hypotheses about the absence of undeclared activities. Commodities to serve as indicators and methodologies then need to be identified on a case by case basis and in a hypothesis-specific way.

In this context a **software tool** has been developed by JRC in support to trade analysis. The tool, called The Big Table (TBT) [11], allows analysts to search control lists, identify items of interest to trade-related case studies, and link these to technical documentation and descriptors needed to retrieve global trade data. More specifically, TBT is designed to:

- Browse and search a collection of reference documents listing and describing items relevant to IAEA safeguards, export controls and case studies on nuclear-related trade. Reference documents include: regulatory documents, technical handbooks and the Harmonized System.

- By the above process, select items of interest to specific case studies related to nuclear trade.

- Map selected items to Harmonized System codes. HS codes are instrumental to the retrieval of trade data records.

- Export HS codes in formats suitable to query web trade data services and retrieve records about trade of interest for analysis.

In short, TBT offers functionalities for analysts to perform steps preparatory to the retrieval and analysis of data records pertinent to case studies on nuclear trade. TBT does not provide access to trade data services, nor tools to analyze data records. Data retrieval and analysis are performed outside TBT.

TBT, originally developed for IAEA, can serve a variety of tasks and communities underpinning export controls, including the rating of items by licensing authorities and supporting commodity identification for customs controls. It is made available to a number of EU Member States licensing and customs authorities. A special version of TBT has been recently developed exclusively for the IAEA that includes the Physical Model [12] in the document collection [13].

3. OPEN SOURCE INFORMATION MONITORING, METHODOLOGY AND TOOLS

*Europe Media Monitor Platform and Tools.* As summarised in [1], “the Europe Media Monitor (EMM), developed by JRC [14], is a web-based multilingual news aggregation system that collects over 200000
news articles per day in about 50 languages from more than 4000 web news sources. The sources are mainly general news sites with a world-wide coverage, but also include some specialist websites and twenty commercial news providers. The system employs text mining techniques to provide a picture of the present situation in the world as conveyed in the media. These techniques include automatic multilingual categorization, entity extraction, geo-location, quote extraction and sentiment analysis. In addition, an algorithm for detecting breaking news automatically clusters all collected news articles every ten minutes and displays the ten largest clusters per language by plotting them on a time-by-size graph. It also provides all the necessary hyperlinks to navigate through the clusters and to go to the source for a detailed exploration. Furthermore it applies some deeper semantic information analysis techniques, for example, to automatically detect violent events, derive reported social networks and analyze media impact [15]. EMM creates a searchable full-text index of all articles that flow through the system. For each article, it stores meta-information including title, description, source, category, language, and original URL (Uniform Resource Locator). However, it does not store the original article itself...Several EMM installations have been set-up providing varying thematic focus and user accessibility. The main EMM installation monitors generic news media with little coverage of specialised thematic areas and serves as a general media monitor and demonstrator of EMM capabilities. Its front page – the EMM Newsbrief [16] - provides a user interface to all this information and is visited on a regular basis by some 25 000 users, and gets some 1.5 million hits per day. Other EMM installations are targeted at specific thematic areas and/or customers. For example, MediSys is specialized in medical and health-related topics” [17].

“EMM is a powerful tool for automatically aggregating and analysing open source information from the Internet. However, if the information needs to be disseminated further or fed into an existing information analysis workflow, the generated news stream has to go through a review and selection process carried out by a domain expert. For this purpose, the JRC developed the NewsDesk application.

NewsDesk allows for manual review and selection of the most relevant articles with an easy-to-use drag-and-drop interface. It also supports the rapid production of newsletters which can be disseminated to interested user groups. The articles selected in NewsDesk including the meta-information extracted by EMM can be posted to existing third-party IT systems, for example, to ingest the information into a backend archive or to publish the selected news on a web portal.

NewsDesk is conceived as a collaborative environment, i.e. users are organised in virtual groups where they can work as a team on the news articles review and newsletter production. It is a web application, which uses RSS feeds as information input – typically generated by EMM. However, it is also possible to ingest feeds from third-party applications. NewsDesk allows users to send notifications via SMS or e-mail; it also integrates an automated notification system to alert personnel on duty during holidays or out of office hours.”

**Nuclear Security Media Monitor (NSMM): Extending Source Coverage.** The public version of EMM monitors pre-selected websites targeting mainly general news media with little coverage of specialised thematic areas. Typically, important news stories are duplicated across many media sites, thus generating some redundancy in the system. Consequently, the system tolerates a certain degree of undetected articles, meaning that not all relevant articles of all targeted sites need to be retrieved.

The IAEA, however, also needs to monitor a set of specific nuclear safeguards-related websites, including NGOs, blogs and sites of national and international authorities. Hence, IAEA provided a list of
more than 140 nuclear-specific websites, which it would like to be added to the EMM source list (hereafter referred to as nuclear websites).

In order to fine-tune EMM to Nuclear Security/Safeguards needs without interfering with the public EMM website, it was decided to set-up a separate EMM installation dedicated to the nuclear safeguards and security domain, hereafter referred to as Nuclear Security Media Monitor (NSMM) [18]. The types of websites of potential interest and therefore monitored by NSMM include: nuclear-focused news agencies and aggregators; regional, national and local government and intergovernmental organisations whose domain covers nuclear issues; NGOs, academic sources and blogs providing analyses on safeguards-relevant topics; general news sources and aggregators; technical publications on the nuclear fuel cycle, etc. A classification of the sources monitored by NSMM is given in Table 3.

<table>
<thead>
<tr>
<th>Category Name</th>
<th>Content</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>General News and Aggregators</td>
<td>Directly-accessed news sources, news aggregators, and &quot;fee-based&quot; comprehensive news archive with collection of newspapers, periodicals, and news wires, filtered by user defined keywords.</td>
<td>Very High</td>
</tr>
<tr>
<td>Nuclear News Aggregators</td>
<td>Articles from news agencies and news aggregators that customarily or primarily report on issues related to nuclear industry and safeguards.</td>
<td>High</td>
</tr>
<tr>
<td>NGO &amp; Academic</td>
<td>Non-governmental organization or university providing detailed reports and added value assessments concerning State's nuclear programmes and activities, and general nuclear nonproliferation issues.</td>
<td>Medium</td>
</tr>
<tr>
<td>Blogs</td>
<td>Interactive websites with commentaries on nuclear issues.</td>
<td>Medium</td>
</tr>
<tr>
<td>Government &amp; Intergovernmental</td>
<td>Information from relevant intergovernmental organizations and competent authorities at national level are a unique source of authoritative information on nuclear safeguards and nuclear industry issues.</td>
<td>Low / Medium</td>
</tr>
<tr>
<td>Nuclear Industry</td>
<td>Information on companies including location(s), products, capabilities, activities, number of employees, main customers, exports of nuclear related items.</td>
<td>Low / Medium</td>
</tr>
</tbody>
</table>

Table 3 - Classification of sources monitored by NSMM.

Some technical challenges still needed to be resolved as the nuclear websites differ from the general news media typically targeted by EMM in several aspects: i) the nuclear websites often publish unique information, thus a high reliability in the detection of new articles is required, ii) the nuclear websites are more static, i.e. the frequency of new articles/reports is much lower and iii) some of the nuclear sites require authentication.

Considerable effort was made to configure and test the monitoring of the nuclear websites: for each site, the relevant RSS feeds and/or HTML pages were selected and inserted into an EMM configuration file. The monitoring results are validated to ensure that all relevant articles are collected successfully. The maintenance of the nuclear sources is a continuous task: some of the source definitions have been optimized or updated following changes on the site or the source URL and additional sources have been added to the system.

Using NSMM. NSMM has initially been established in a joint project with IAEA’s Division of Information Management (SGIM) with the aim to streamline IAEA’s acquisition and analysis of open source information and develop the current information collection production process for the internal news review (SGIM Highlights) to a more efficient system. SGIM has started using NSMM/NewsDesk for the daily monitoring of its ‘web sources’ (over 150 nuclear-related web sites that were previously monitored manually) in the beginning of 2013. NSMM/NewsDesk is used for selected parts of the production workflow of the SGIM Highlights (see [1]). As well as the daily news monitoring and newsletter production, SGIM analysts also carry out country-specific monitoring and searches, which typically
involves monitoring additional national sources and generates more state-specific information. Following the positive experience in the general daily monitoring, SGIM analysts are considering using NSMM/NewsDesk for country-specific monitoring.

NSMM is also used operationally at the JRC for the daily creation of a Nuclear Security News review, which is distributed to selected recipients in the EU institutions.

Figure 1 illustrates the information flow for NSMM and the related NewsDesk application.

![Figure 1: Schematic overview of information flow in a possible setup for open source information collection and analysis based on NSMM and NewsDesk.](image)

The two main use cases are:

1. **Real-time Information Awareness.** The end user (e.g. an information analyst or nuclear inspector) can directly consult the NSMM Newsbrief page, which is always updated in near-real-time. Additionally, the nuclear-specific categories which automatically filter the incoming information can be further refined and thus allow the users to access the articles according to more specific geographic or thematic areas of interest, which might not be covered by a more generic newsletter. The NSMM could allow an inspector or analyst to set up filters for their specific area of interest to ensure access to relevant open source information on a near real-time continuous basis.
2. *Daily monitoring* for the generation of a domain-specific newsletter: The information analyst manually reviews all media articles that pass the NSMM filter; selects the articles that are most relevant to the subject (e.g. Nuclear Security); and arranges them in the resulting newsletter according to thematic and geographic area of interest. The newsletter is also archived in a repository for later reference. In the case of the JRC Nuclear Security News the figures on a typical day are as follows: NSMM collects 170000 articles on the monitored sites and filters about 500 articles relevant to nuclear security/safeguards issues. The information analyst selects about 25 articles to be included in the newsletter which is then distributed to selected recipients in the EU institutions.

4. **CONCLUSIONS AND OUTLOOK**

As acknowledged by [19]: “The collection and analysis of open source information is an essential element in the State evaluation process. Open source information, including trade and procurement data, can provide early indications of potentially undeclared nuclear activities. Analysis of open-source information is used by the Department to support the State evaluation process, in particular to help verify the completeness”.

Open source information analysis can thus play a role in the IAEA by contributing to the establishment of a global state picture within the context of nuclear safeguards. In this paper a particular focus has been put on the possible role and use of trade data for nonproliferation activities, on the open source information monitoring and collection process and on the tools and approaches developed by JRC to support them.

The importance of open source information analysis has the potential of being further increased in the future following the evolution of the state level concept by contributing to the establishment of a global state picture. This might inform the shaping of state specific factors and of the related technical objectives to improve safeguards effectiveness and efficiency.

**ACKNOWLEDGEMENTS**

5The work here presented is being carried out within the projects Open Source Information for Nuclear Security: applications and tools and Strategic Trade Analysis for Non Proliferation, both funded within the European Commission (EC) Euratom Horizon2020 Research and Training Programme, as contribution to the EC Support Programme to the IAEA. The Nuclear Security Media Monitor is developed by EC-JRC as collaboration between the Institute for Transuranium Elements and the Institute for the Protection and Security of the Citizen. The authors would like to thanks all the IAEA officers and JRC colleagues that contributed to the activities here presented.

**REFERENCES**


[18] https://nsmm.emm4u.eu/nucsec

Cost-Sensitive Classification Methods for the Detection of Smuggled Nuclear Material in Cargo Containers

Jennifer B. Webster
Pacific Northwest National Laboratory, Work Performed as a student in the Department of Mathematics, Texas A&M University

Abstract

Radiation portal monitors (RPMs) have become ubiquitous in the international nonproliferation effort. They provide a non-destructive and minimally invasive method for interrogating cargo containers, but, like all indirect testing methods, there are limitations based on the background environment and detector efficiency that will impact the assay performance. In a traditional RPM configuration, each detector is individually tuned to meet a desired performance criterion by setting a threshold which the observed counts must exceed in order for the detector to register an alarm. This threshold will impact both the false alarms and false negatives produced by the system if one were to rely on a single measurement for all decision making purposes. The RPM is created from an array of such tuned detectors, which will have systemic error rates that are based upon these individual detector rates. If one wishes to evaluate and adjust the performance of the entire system, then the thresholds for the individual detectors will need to be modified in order to reflect that objective. Usually, a single detector threshold is determined independently of the rest of detectors and without knowledge of the spatial relationship between detectors as reflected by the RPM design.

This adjustment of detector thresholds can be thought of in terms of a classic classification problem – we are determining an optimal threshold or decision rule that separates the observations corresponding to cargo containers without a hidden source from those which do. In this paradigm, it is easy to discuss the balance of the cost associated with scanning multiple completely harmless containers against the risk associated with allowing smuggled sources in terms of false positives and false negatives for the entire system rather than for an individual detector. Our methodology uses a support vector machine (SVM) framework with F-score feature selection to provide nearly optimal classification while simultaneously determining thresholds that take into account correlations between detectors as a result of spatial position and the properties of radiation. In some cases, this procedure can decrease the fraction of false positives by an order of magnitude over current methods.

1 Introduction

Classification problems occur in many different aspects of life, be they in the human decision process or in a computational evaluation. Classifying objects involves a complex interplay between the available information, the ultimate goals of the process, and the various consequences of the resulting actions. For instance, when examining information in preparation to making a decision, we analyze the information as a collection and not just as individual measurements, allowing us to utilize correlations between data and improve the accuracy of our decision. Furthermore, we as humans analyze the costs and risks involved in the situation and weight various pieces of information in order to compensate for the varying importance of goals in our decision making process. This could include restricting the likelihood of occurrences of one of the outcomes or weighting the different outcomes to compensate for differences in costs.
This study examines the balance of these ideas in the context of detecting smuggled nuclear material inside cargo containers. Every year, approximately 40 million shipping containers pass through American ports [17] – this is more than 100,000 crates each day. After the terrorist attacks on September 11, 2001, the border security problem has drawn more attention. This has led to increasing the number of checkpoints at border crossings and more stringent security requirements when flying. It has also facilitated the enactment of several border security laws, including the mandatory screening of maritime cargo as specified in Public Law 110-53 – August 3, 2007: Implementing Recommendations of the 9/11 Commission Act of 2007 [18]. Thus, there is a need for effective, efficient sorting (or classification) algorithms that locate smuggled nuclear material inside of cargo containers.

In this paper, we will discuss three decision rules for distinguishing cargo containers containing nuclear material from those that do not by using a gamma detector array to measure radiation from multiple positions around the container. The first method is most similar to that currently practiced in the detection community and finds the optimal threshold for each detector independently of the others. The second method will describe uses analytic knowledge of the distribution of measurements to determine the absolutely optimal decision rule for separating containers. The third method we propose is a cost-sensitive support vector machine implementation that can achieve nearly optimal performance while balancing the false alarm rate against the risk of allowing nuclear material to escape detection. It also has the advantage that it requires no assumptions about the underlying distribution of the measurements. All three of these methods rely on statistical theory to facilitate the cost-sensitivity of the algorithm. Thus, we begin with a section describing the theoretical underpinnings before moving on to the details of the three algorithms and finally a comparison between them.

2 The Theory Behind Bayes’ Optimal Decision Rule

There are many ways to go about classifying populations, as mentioned above. If exact information about the conditional distributions of these populations, the probability of each object type in the overall population and the costs of various classification actions (misclassification costs, for example) are known or can be approximated with some accuracy, then an optimal classification rule can be determined. This classification method is called the Bayes’ Rule for Cost Minimization or the Bayes’ Minimal Risk Decision.

This method is based on the fact that it is usually easiest in an experimental situation to control the type of an object rather than the exact measurement, \( x \). Consider an experiment to determine the gender of a person based on height measurements. It is relatively easy to find many people that are clearly male or clearly female and then measure each of those people’s height. It is much harder to find a sufficient sample of people of any gender that are exactly the same height for every possible height measurement.

These distributions, where we control either the class or the measurement and allow the other event to vary, are called conditional probabilities and can be transformed using Bayes’ Rule from functions that have the class controlled by the researcher into functions that presume that the measurement was the controlled quantity. As we can see in the following theorem, we can construct a decision rule that minimizes the expected cost of misclassification using analytic information about the probability distributions.

**Theorem 1** (Bayes’ Rule for Minimization of Cost of Misclassification [10]).

Let us consider a set of features \( x^N \) obtained from two populations, \( S \) and \( D \), with population conditional distributions \( p(x|S) \) and \( p(x|D) \), respectively, for \( x \in S \subseteq R^N \). Also, suppose \( P(S) \) and \( P(D) \) are the prior probabilities of encountering each respective population and \( P(S)+P(D)=1 \), i.e., all items that produce readings in the feature space must be classified as belonging to exactly one of the two populations. Further, let the cost of misclassifying a point from the \( S \) population as a member of population \( D \) be given
by $c_{D|S}$ and $c_{S|D}$ be the cost of classifying a point from the D population as coming from the S population. Then, for a given point in the feature space, $x$, the classification rule that minimizes the expected cost of misclassification is:

$$x \text{ is classified as belonging to } S \iff x \in \left\{ x \in S^n \mid c_{D|S}P(x|D)P(D) \leq c_{S|D}P(x|S)P(S) \right\} \quad (2.1)$$

**Remark 2.1.** Under the further assumption that $P(S)$, $c_{D|S}$ and $p(x|D)$ are strictly greater than zero for all points $x$ in the feature space, then

$$A = \left\{ x \in S^n \mid \frac{c_{S|D}P(D)}{c_{D|S}P(S)} \leq \frac{p(x|S)}{p(x|D)} \right\} \quad (2.2)$$

This demonstrates that the absolute cost of each individual error need not be known nor do the probability distributions need to be normalized for this analysis to be valid. Only the relative costs of the two error types and the proportionality of the two cargo container types are required. Further, by treating the cost ratio or indeed the whole left side of the inequality as an unknown, we can use this framework to impose constraints on the expected total misclassification rate or even the individual expected rates of false negatives and false positives through the Neyman-Pearson Lemma:

**Theorem 2** (Neyman-Pearson Lemma [14]).

*Suppose we are labeling objects as belonging to one of two populations: S and D. Given knowledge of the conditional distributions of measurements for each population and the proportion of each population in the overall context, then the likelihood-ratio test which gives the label S to an object having measurement $x$ with a specified false negative rate $\alpha$ is*

$$\Lambda(x) = \frac{L(D|x)P(x|D)P(D)}{L(S|x)P(x|S)P(S)} \leq \eta$$

*where $\eta$ is a constant chosen such that $P(\Lambda(x) \leq \eta | D = \alpha$*

### 3 Three Methods of Detector Threshold Design

Using the theoretical principles developed in the previous section, we will be comparing three different methods for defining detector thresholds which balance the possibility of allowing nuclear material to escape detection against the costs of invasively screening large numbers of containers with completely harmless cargo. The first two methods allow us to benchmark the performance of our proposed algorithm – one that approximates current methods by treating the detector array as a set of independent threshold problems and the next the analytically optimal bounding function as defined by the Bayes’ Optimal Decision Rule. Then we will move on to discuss our proposed algorithm – a cost-sensitive SVM method that can be trained using real world samples rather than making assumptions about the distribution of measurements.

For each of these tests, we will be considering the same physical setup of an array of gamma detectors surrounding two sides of a standard cargo container, as shown in Figure 1. Each realization of a detector can take on positive integer values denoting the count rate observed at that detector position and is represented by a separate dimension in an $n$-dimensional feature space, $S^n$, where $n$ is the
number of detectors in the array. Using MCNP [20], one can calculate the expected average count rate of each detector for a \( t \)-second measurement interval given a specific cargo distribution, source size and location. This was done for a variety of such combinations in the course of the study, although we will concentrate here on a single arbitrary theoretical model. For information about the effect of varying cargo and background on the algorithms, please refer to [19].

(a) Container with no source  
(b) Container with source

Figure 1: Average detector readings for a cargo container with an HEU source in a \( t \)-second interval. In both figures, the containers are filled with light density materials. The left figure shows only the photon counts produced by background radiation and the right shows the same background radiation in addition to a 1kg HEU source placed inside the container.

3.1 Detector Arrays as a series of 1-dimensional Neyman-Pearson Problems

The first algorithm we will discuss is what we will call the Box Threshold Method. The end goal of this method is, for a given set of \( n \) measurements, to define an \( n \)-dimensional box in the feature space, \( S^n \), such that the total expected misclassification error is minimal and the false negative rate is equal to the specified rate \( \alpha \). Any container that produces a measurement \( x \) that lies within this \( n \)-dimensional box will be classified as safe. To the best of the author’s knowledge, this method, where every detector has an alarm threshold that is independent of the other detectors, is similar to that currently in practice in border security checks. In this section, we make use of the fact that we are given exact probability distributions for the readings from each detector for each type of container. The standard statistical model for radiation detection is the Poisson distribution. The basic algorithm is as follows:

1. Specify an individual detector false negative rate, \( \alpha_{ind} \). This requires that every single detector has the same false alarm rate when taken separately.

2. Using the Neyman-Pearson Criterion, we enforce the constraint on the false negative rate by finding a value \( c_i \) such that, for the \( i^{th} \) detector of a container with an HEU source having mean \( \lambda_{i,D} \),

\[
P(x \leq c_i | D) = \sum_{x=0}^{c_i} \frac{\lambda_{i,D}^x}{x!} e^{-\lambda_{i,D}} = \alpha_{ind} \quad (3.1)
\]
3. Determine individual thresholds, \( t_i \leq c_i \), by minimizing either the expected cost of misclassification or the expected false positive rate, as chosen by the researcher and specified in the respective equations below:

\[
ECM(t_i) = P(x > t_i | S) + P(x \leq t_i | D)
\]

\[
= 1 - \sum_{x=0}^{t_i} \frac{x^0}{x!} e^{-\lambda_i} S^x + \sum_{x=0}^{t_i} \frac{x^0}{x!} e^{-\lambda_i} D^x \quad (3.2)
\]

\[
P(x > t_i | S) = 1 - \sum_{x=0}^{t_i} \frac{x^0}{x!} e^{-\lambda_i} S^x \quad (3.3)
\]

4. Arrange these values \( t_i \) into the vector \( \mathbf{t} \), then calculate the total expected false negative rate \( \alpha \), which in this case can be found by computing:

\[
\alpha := P(S | D) = P(x \leq t | D) = \prod_{i} \sum_{x=0}^{t_i} \frac{x^0}{x!} e^{-\lambda_i} \quad (3.4)
\]

5. Repeat steps 1-4 adjusting \( \alpha_{ind} \) until \( \alpha \) is as close as possible to \( \alpha \), the desired expected false negative rate as given by the researcher. In this particular instance, adjustment of \( \alpha_{ind} \) will be made using the bisection method where the function of which we want to find the root is

\[
f(\alpha_{ind}) = \alpha \left( \alpha_{ind} \right)^{-\alpha}.
\]

Then, \( \mathbf{xst} \) is the decision rule that classifies an object which produces a set of measurements \( \mathbf{x} \) as safe (without a source). As a result, we have guaranteed that the false negative rate, \( P(S | D) \), is less than or equal to the specified level \( \alpha \). A simple 2 dimensional representation of this method is given in Figure 2.

**Remark 3.1.** Current methods of radiation detection usually use a constant threshold for every detector in the system, but as we are interested in how correlations in measurements affect the classification process, we will allow each detector threshold to vary individually. Further investigation exploring this difference in approaches is warranted.

**Remark 3.2.** If one chooses to minimize the total expected cost of misclassification as in (3.2), it may not be possible to exactly achieve the specified false negative rate \( \alpha \). Even though the false positive rate will continue to decrease as the threshold increases, it may not offset the increase in the overall misclassification rate since the false negative rate will increase as the threshold increases.

If we choose to minimize the false alarm rate as in (3.3), we can further utilize the known structure of the distribution functions to simplify our computations. Using the fact that the cumulative distribution function of any one dimensional distribution increases as the argument \( x \) goes to infinity, we know that the false positive rate given by \( P(x > t_i | S) \) will decrease as the threshold value increases. More simply, the false positive rate for a one dimensional function is a monotonically decreasing function. Therefore, with this objective function, we can develop the thresholds using only information about the distribution of measurements from containers containing HEU, \( P(x | D) \).
Figure 2: Illustration of the Box Method implementation for a 2 detector system with minimization of the false positive rate. Assume that we are given many samples of two detector measurements that we might observe when scanning a specific kind of container that contains a HEU source (the scatter plot at the top right). First, we project these samples into the 1D spaces for the set of all detector readings, essentially looking at measurements from one detector at a time while completely ignoring the other measurements, as given by the frequency plots along each access of the scatter plot. Next, we use the Neyman-Pearson criterion to determine a threshold with a specified false negative percentage, $\alpha_{ind}$, which are denoted by the red line in each frequency plot. Finally, we use these individual thresholds to create a rectangular area in the original feature space, which corresponds to the box created by the axes and the two red lines in the scatter plot. The interior of this rectangle is designated as the region $A$ and points that fall inside this region will be classified as belonging to safe cargo containers. Given this defined region we can count the total fraction of false negatives that fall inside this region to get the total false negative rate, $\bar{\alpha}$ and iteration can be used to adjust the region until $\bar{\alpha} = \alpha$.

3.2 Analytic Bayes’ Optimal Decision Method

The other option for a classification method using analytic distribution information that will be discussed in this study is the Bayes’ Optimal Decision Method. Unlike in the previous method, all the detectors that will ultimately be used to make the classification will also be used simultaneously to arrive at the decision rule. This will allow correlations between the measurements to be exploited in order to produce
more accurate classifications. In fact, this method will achieve the optimal decision rule. However, it may not be possible to achieve this in practice as we are not usually provided with complete analytic knowledge of the distributions of measurements.

Following the procedure of Sec. 2, we would like to find the region of feature space $A$ which minimizes the total expected cost of misclassification (ECM)

$$ECM = c_S | D \sum_{x \in A} p(x|D)p(D) + c_D | S \sum_{x \in A} p(x|S)p(S)$$

(3.5)

subject to the constraint that the false negative rate $P(S|D)$ must be no more than $\alpha$, i.e.,

$$\sum p(x|D)p(D) \leq \alpha.$$ 

This will be done utilizing the likelihood ratio formulation where a cargo container is classified as safe if its measurement, $x$, satisfies the inequality

$$\frac{c_S | D p(D)}{c_D | S p(S)} \leq \frac{p(x|S)}{p(x|D)}$$

(3.6)

where $c_S | D$ is the cost of a false negative, $c_D | S$ is the cost of a false positive, $P(\cdot)$ is the fraction of containers of the specified class, and $p(x|\cdot)$ is the conditional distribution of measurements of the given class.

Determining the exact costs of each type of misclassification and even the correct proportions of containers in the overall population can be exceedingly difficult. Therefore, we will treat the left hand side of (3.6) as a constant and utilize the Neyman-Pearson Lemma (Theorem ) to enforce our constraint. Thus, our problem becomes to determine a value for the cost ratio $\eta$ such that a container is labeled safe if

$$\eta \leq \Lambda(x) = \frac{p(x|S)}{p(x|D)}$$

(3.7)

where $\eta$ is chosen so that $P(\Lambda(x) \geq \eta | D) \leq \alpha$. In this fashion, we will trade estimation of the misclassification costs in the original cost minimization formulation for control of the global false negative rate.

**Remark 3.3.** It is important to note that we could have specified the overall false positive rate that we would find acceptable instead of the false negative rate and this would not substantially change the overall algorithm. The general statement of the Neyman-Pearson Lemma makes this possible by changing the labeling system and a few inequalities.

### 3.2.1 Determination of Cost Ratio

One way to determine the cost ratio is through a root finding approach. We begin by assuming that there is a value $\eta_0$ such that the safe region is defined by $\frac{p(x|S)}{p(x|D)} \geq \eta_0$ and has the desired false negative rate $\alpha = P(\Lambda(x) \geq \eta_0 | D)$. Then for any other choice of $\eta$, we can determine the associated false negative rate, $\alpha_\eta = P(\Lambda(x) \geq \eta | D)$. One can then adjust $\eta$ until $\alpha_\eta$ is as close as possible to the specified level $\alpha$ by finding the roots of

$$f(\eta) = \alpha_\eta - \alpha = P(\Lambda(x) \geq \eta | D) - \alpha.$$ 

(8)
This turns the problem of finding thresholds for multiple detectors into an effectively one dimensional problem.

Furthermore, we can notice that \( f(\eta) \) is a monotonically decreasing function as a result of the properties of probability distributions. Combining this monotonicity with the normalization of probability measures, we can see that, as \( \eta \to 0 \), \( P(\Lambda(\eta) \geq \eta | D) \to 1 \) and for \( \eta \) large, \( P(\Lambda(\eta) \geq \eta | D) \) will be near 0, so \( f(\eta) \) will range from \( 1-\alpha \) to \( -\alpha \) as \( \eta \) grows. The discrete nature of the Poisson distribution means that we may not be able to choose one particular value of \( \eta \) as the root of the function \( f(\eta) \) since, by definition, neither function varies smoothly. However, we can choose the closest possible choice of \( \eta \) to \( \eta_0 \) via a standard bisection method for root finding.

Because we are generally working in high dimensions, we will use stochastic integration techniques to approximate \( P(\Lambda(\eta) \geq \eta | D) \) instead of actually performing the summation:

\[
P(\Lambda(\eta) \geq \eta | D) = \sum_{\{x | \Lambda(x) \geq \eta\}} p(x | D).
\]

(9)

This means that we also avoid having to parameterize the boundary of the region and avoid some of the computational problems that can occur in high dimensional integration.

### 3.2.2 Initial Tests of the Bayes’ Optimal Method

The actual shape of the region defined by specifying \( \Lambda(x) \geq \eta \) is highly dependent on both the character of the distributions used to determine the value \( \eta \) and the global false negative rate \( \alpha \). For example, tests were completed using two detectors where the distribution of measurements for containers with a source, \( p(x | D) \), changed from a single Poisson distribution to a bimodal distribution that is the sum of two Poisson distributions. Both tests had the same distribution for safe measurements, \( p(x | S) \), which is a two-dimensional Poisson distribution with mean (13.0,17.0). In the first test (Fig. Error! Reference source not found.), the mean of the dangerous distribution was placed at (30.0,23.0). In the second test (Fig. Error! Reference source not found.), the means of the two dangerous distributions are (30.0,23.0) and (25.0,35.0). Given the same false negative rate \( \alpha = 0.05 \), we can see that the two curves generated differ substantially as the bimodal distribution has forced the boundary of the Bayes’ Optimal Decision Region to bend. As a result, accurate and complete characterization of the distribution of measurements is necessary in order to develop the most accurate classification algorithms.
Figure 3: Given here are the Bayes’ Optimal Decision Boundaries for two different distributions as determined by a root finding method. This simple example shows how the character of the distributions influences the shape of the region A and emphasizes the need for an accurate characterization of the entire feature space for the most complete classification.

![Bayes' Optimal Decision Boundaries](image)

Figure 4: Effects of varying the allowable false negative rate on the shape of the Bayes’ Optimal Decision Region. As the level $\alpha$ increases, the boundary moves further away from the origin, but the overall shape remains constant, suggesting that the shape is controlled by the character of the distributions alone.

Alternatively, we can study the region described by the Bayes Optimal Decision rule for a fixed distribution as the percentage of false negatives is varied. Using the bimodal test distribution as in the previous discussion, tests were completed allowing only the desired false negative rate $\alpha$ to vary. In this particular case, it appears that the boundaries vary along two vectors that are linked to the difference in means of the two distributions and the parameter that changes is the distance from each major boundary portion to the origin, as seen in Figure 4. More generally, the overall shape of the region, A, appears to be controlled by the character of the distributions used to generate the decision boundary and the false negative level $\alpha$ controls the size of the region. This could prove useful in updating algorithms in practice as one large detailed study could be made to understand the character of the region, which is then adjusted using more naive methods to obtain specific false negative rates as requested.

It is interesting to note the trade off in error types caused by adjusting the decision boundary. Table 1 shows the computed values of $\eta$ and the false negative and false positive rates for each of the optimal decision regions in Figure 4. As the percentage of false negatives doubles, the false positive rate decreases by an order of magnitude. This may not be true for the realistic distribution of measurements, but it will allow us to discuss the effects of the disparate numbers of safe and dangerous containers in our population. In particular, with the nearly 40 million safe containers each year passing through ports, decreasing the false positive rate corresponds to a significant decrease in the number of highly scrutinized containers. This translates to a reduction in cost for scanning systems and hence must be evaluated when determining the acceptable risk of undetected smuggled nuclear material.
Table 1: Variations in the cost ratio \( \eta \) and false positive rate as the false negative constraint is changed for a fixed distribution, corresponding to the curves in Figure 4.

<table>
<thead>
<tr>
<th>False Negatives</th>
<th>( \eta )</th>
<th>False Positives</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.025</td>
<td>1.3580</td>
<td>0.0307</td>
</tr>
<tr>
<td>0.05</td>
<td>0.3629</td>
<td>0.0121</td>
</tr>
<tr>
<td>0.1</td>
<td>0.0736</td>
<td>0.0039</td>
</tr>
<tr>
<td>0.2</td>
<td>0.0115</td>
<td>0.0006</td>
</tr>
</tbody>
</table>

4 Support Vector Machines

Support Vector Machines (SVM) are a specialized optimization method for kernel based classification rule development. The goal of SVM algorithms is to find the hyperplane separating two populations which minimizes misclassifications. Studies have shown that decision rules produced in this manner approach the Bayes’ Optimal Rule in the space of such restricted classifiers as the sample size increases [13, 15, 7, 3]. In the linear case, SVM methods optimize over both the normal and distance to the origin of a plane separating the two populations, i.e., SVM algorithms seek to find an optimal separating hyperplane to partition the feature space, \( S^\eta \), according to the labels \( S \) and \( D \). There are algorithms that use a kernel such as an \( n^{th} \) degree polynomial or radial basis function in order to transform the high dimensional feature space into a lower dimensional space where optimization algorithms can work more effectively. The hyperplane is found in this transformed space and then pushed back to the original feature space to perform classification. The choice of kernel is initially provided by the researcher, so it requires expert judgment and experimentation to choose the appropriate function.

Consider a set of \( n \) samples \( y_i x_i \), where \( x_i \) is the vector of measurements on which the decision is based and \( y_i \) is either 1 or \( -1 \) and corresponds to the labeling of the two classes we are trying to distinguish. There are two steps to any support vector formulation – transformation of the vectors of measurements, \( x_i \), to the Hilbert space on which the decision function will operate and then utilization of quadratic programming methods to find the optimally separating hyperplane between the two classes of samples in this space. Using the kernel function chosen by the researcher, \( \Phi(x) \), each sample point can be transformed into a space that is more conducive to finding a linear separation of the sample points from each class.

After completing the kernel application step, a quadratic programming method chooses the optimal separating hyperplane by determining the normal vector \( w \) and associated distance \( b \) that maximizes the margin, the distance between the decision boundary and the nearest \( x_i \) from each class, on either side of the separating hyperplane, as shown in Fig. **Error! Reference source not found.**. The \( |b|/|w|_2 \) is the distance from the hyperplane to the origin along the normal \( w \).
Figure 5: SVM methods choose the separating hyperplane that maximizes the margins between the two classes of points. In the original formulation, the two margins have equal size $1/|w|$. The ν-SVM formulation allows some of the points from the training method to be inside of these margins or to be misclassified. This can be necessary, especially when the two sets of points $S$ and $D$ are not easily separable.

In order to make this method cost-sensitive as in the other methods in this study, we will utilize the 2ν-SVM formulation proposed in [2, 5], which relies on the introduction of a single parameter $γ \in [0,1]$ that gives the trade off in the two error types:

$$\min_{w, b, q} \frac{1}{2} |w|^2 + \frac{ν}{n} \sum_{i \in I^+} \xi_i + \frac{1-ν}{n} \sum_{i \in I^-} \xi_i - νq \quad (4.1)$$

s.t. $y_i w \cdot \Phi(x_i) + b \geq ϱ - \xi_i$, for $i = 1, ..., n$

$\xi_i \geq 0$ for $i = 1, ..., n$

$q \geq 0$

where $I^+ = \{ i : y_i = 1 \}$ ($n_+ = |I^+|$) and $I^- = \{ i : y_i = -1 \}$ ($n_- = |I^-|$). As further shown in [Error! Reference source not found., Error! Reference source not found.], the choice of ν and γ can provide bounds on the fraction of margin errors of each type, $ν_+$ and $ν_-$:

$$ν_+ = \frac{vn}{2γn_+} \quad ν_- = \frac{vn}{2(1-γ)n_-} \quad (11)$$

There are many pieces of software that need to work together for the support vector machine framework to operate effectively. Therefore, we used the LIBSVM software package [1] with the modifications for the 2ν-SVM formulation as given in [4]. In order to bound the false negative rate, a simple search over the feasible region is performed to gain a basic understanding of how the false negative rate changes and then localized to find a specific false negative rate in a similar fashion to the coordinate descent method described in [6]. The largest difference is that our method varies ν and γ where Davenport’s coordinate descent method varies $ν_+$ and $ν_-$. 
5 Comparison of Methods

As one can see, in Figs. 6 and 7, for the two detector case, the SVM method is an improvement on the Box Threshold Method and matches well with the Bayes’ Optimal solution in the space of restricted classifiers [12]. By this, we mean that, in the space of hyperplane decision boundaries, the SVM method will choose the Bayes’ Optimal solution, which may differ from that described in Sec. 3.2 because the likelihood ratio test does not necessarily generate a planar boundary. In this case, this may provide a skewed sense of the efficacy of the SVM as compared with the Box Method since the Bayes’ Optimal solution is linear here. The kernel choice can effect the outcome as much as the sample points provided for training and there is a danger of overfitting in higher dimensions for all of the methods, as can be seen with the Box Method in Figure 8.

Table 2: Misclassification Rates produced by the three algorithms.

<table>
<thead>
<tr>
<th>Method</th>
<th>False Neg.</th>
<th>False Pos.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box</td>
<td>0.0663</td>
<td>0.0567</td>
</tr>
<tr>
<td>Bayes’</td>
<td>0.0669</td>
<td>0.0066</td>
</tr>
<tr>
<td>SVM</td>
<td>0.0652</td>
<td>0.0070</td>
</tr>
</tbody>
</table>

Figure 6: Examples of decision boundaries created using the three methods.
Figure 7: As before, we can use the ROC and Accuracy curves for comparing all three classification methods (Box, Bayes’ and SVM) irrespective of a specific desired false negative constraint, $\alpha$ for a two dimensional feature space. In this case as well, the SVM method which takes into account the correlations in measurements improves upon the classification provided by the Box Method. The means of the distributions are (17,13) and (25,30), as before.

(a) ROC Plot  (b) Accuracy v. Precision Plot

Figure 8: Adding features until we are working with a 4 dimensional feature space, we can utilize the ROC and Accuracy curves for comparing classification methods irrespective of a specific desired false negative constraint, $\alpha$. The SVM method performs almost as well as the optimal choice. However, we are starting to add information about features that do not differ between safe and dangerous containers. Therefore, the overall classification performance has begun to decline (as seen by the Box curve) and we are in danger of overfitting the SVM method to our sample points since they do not provide information about the greater volume of the feature space.

Furthermore, some of the efficacy of using multiple detectors is lost in the Box Method because correlations are not taken into account and the number of detectors containing minimal signal from the source far outweighs the number of detectors with a noticeable signal increase. Choosing the
appropriate detectors with which to make the best decision is just as important as choosing the correct decision rule.

5.1 Effects of Enforcing Misclassification Percentage Constraints

As we saw in Sec. 3.2.1, one can translate a constraint on the false negative rate into a single cost ratio \( \eta \), which is a function of the probability of the percentage of dangerous containers in the population and the costs of misclassification errors. Thus, we can make use of our calculated values for \( \eta \) and known information to estimate the dollar cost of a false negative \( c_{S|D} \) for a particular implementation of the algorithm. For example, we know that there are roughly 11 million containers entering US ports each year [16] and according to the International Atomic Energy Agency (IAEA), there have been 2331 confirmed incidents involving illicit trafficking and other such unauthorized activities involving nuclear material in the period from 1993 to 2012, 16 of which have involved “unauthorized possession” of HEU or Plutonium [8]. Supposing that all of these events were to occur by using cargo containers to smuggle such material into American ports, we have an average of 0.8 events per year, giving us the proportions of each container type in our population as \( P(D)=7.3\times10^{-8} \) and \( P(S)=0.99999993 \). We can further estimate the cost of physically searching the cargo container needlessly, \( c_{D|S} \), by assuming that it will take 8 man hours to perform the search and the average dock worker is paid \$14 per hour. Giving these workers a hazard pay of \$25 per hour, we can assume that the cost of such a needless search is \( c_{D|S}=\$200 \). Thus, the cost of allowing nuclear material to escape detection by our algorithm is

\[
c_{S|D}=\eta c_{D|S} \frac{P(S)}{P(D)}=2.7\times10^9 \eta
\]

(5.1)

Therefore, if we found that a 5% false negative rate gave a value of \( \eta=0.1 \), then our algorithm has assigned a dollar value of around 274 million to the destruction of a city. In comparison, Hurricane Katrina cost roughly 108 billion dollars in property damage and destruction [11]. It should be noted that the estimate of \( P(D) \) used here is different from the true average value as a result of our assumptions about the IAEA statistics. The events recorded by the IAEA are international statistics, not just those events occurring at American ports. Furthermore, 25 kg of HEU are required by the IAEA before a “significant quantity” of material is obtained. The IAEA defines a significant quantity to be “the approximate amount of nuclear material for which the possibility of manufacturing a nuclear explosive device cannot be excluded” [9]. Based on this fact, 25 crates, each containing 1 kg of HEU, would need to be smuggled into the country in a relatively short time period. If we assume that all of the material must make it into the country within a single year, the proportions of each container type become: \( P(D)=2.3\times10^{-6} \) and \( P(S)=0.999997 \). Performing the calculation in the same manner as before, \( c_{S|D}=8.8\times10^7 \eta \), which decreases the cost of allowing a single container to escape detection significantly (=\$9\times10^6). Since the total smuggled source requires 25 crates, this translates to a cost of approximately \$220 million for the entire 25 kg of HEU.

We can also reverse our procedure and determine the equivalent cost of searching every container, \( c_{D|S} \), if we suppose that the destruction of a city through nuclear material is equivalent to the impact of this hurricane. In this case,

\[
c_{D|S}=\frac{c_{S|D}}{\eta} \frac{P(D)}{P(S)}=7.8\times10^3 \eta
\]

(5.2)
With the same 5% false negative rate and $\eta=0.1$ as in the previous analysis, this computation suggests that hand searching a single container is worth on the order of 78 thousand dollars. In reality, the cost of searching containers is far less than this and leads one to conclude that searching containers is worth the cost in return for a large decrease in risk. Furthermore, these calculations suggest that it might be more appropriate to control the overall false alarm rates of the system if one wants to control the cost of the entire system since the general population contains many more safe containers than those with a source. However, we will continue to constrain the false negative rate in this study for consistency.

6 Conclusions

Using information from an entire array of detectors can be more effective than individually evaluating the various measurements in the course of developing detector measurements. In some cases, we can even achieve nearly optimal separation. The methods discussed in this paper use cost-sensitive statistical and machine learning methods to translate a restriction on the global error rates of the system into an impact ratio of the costs of false positives and false negatives on the system. By using this type of approach, we can reduce the instances of false alarms by a measurable amount, which translates to direct monetary savings in fielding such monitoring devices. Further information regarding the details of these methods and an expansion on the effects of physical cargo variation on each algorithm may be found in [19].

7 Acknowledgments

I would like to thank Dr. Chirayath, Texas A&M University, and Los Alamos National Lab, especially Dr. C.J. Solomon and Dr. Avneet Sood for their help with MCNP simulations and computing resources. Funding for this work was provided by ARI Grant #2008-DN-077-ARI018-05. Most importantly, I would like to thank Dr. Wolfgang Bangerth for his guidance throughout this project.

References


Management and Analysis of RPM Data
Nathan C. Rowe, Scott Alcala, Jason Crye, and Angela Lousteau
Oak Ridge National Laboratory
Oak Ridge, TN, USA

ABSTRACT

Oak Ridge National Laboratory (ORNL) receives, archives, and analyzes data from radiation portal monitors (RPMs). Over time the amount of data submitted for analysis has grown significantly, and in fiscal year 2013, ORNL received 545 gigabytes of data representing more than 230,000 RPM operating days. These data come from more than 900 RPMs. ORNL extracts these data into a relational database, which is accessed through a custom software solution called the Desktop Analysis and Reporting Tool (DART). DART is used by data analysts to complete a monthly lane-by-lane review of RPM status. Recently ORNL has begun to extend its data analysis based on program-wide data processing in addition to the lane-by-lane review. Program-wide data processing includes the use of classification algorithms designed to identify RPMs with specific known issues and clustering algorithms intended to identify as-yet-unknown issues or new methods and measures for use in future classification algorithms. This paper provides an overview of the architecture used in the management of these data, performance aspects of the system, and additional requirements and methods used in moving toward an increased program-wide analysis paradigm.

I. INTRODUCTION

Oak Ridge National Laboratory (ORNL) receives, analyzes, and archives radiation portal monitor (RPM) data to provide insight into a wide variety of factors that affect site operation and RPM maintenance. This allows better recommendations on how to effectively manage and maintain RPM operations. The data are also reviewed on a lane-by-lane basis to monitor the RPM state-of-health to provide information on needed maintenance, configuration problems or optimizations, and potential improvements for operations. ORNL has been receiving and analyzing RPM data since 2004, and in that time, the amount of data submitted for analysis has grown significantly. In fiscal year 2013, ORNL received data from over 900 operating RPMs; the 545 gigabytes of data represented more than 230,000 RPM operating days. The data are extracted and stored in several formats, including the raw text files (commonly referred to as Daily Files), a Microsoft SQL database, and Matlab or HDF5 formatted flat files. The data are typically analyzed lane-by-lane on a monthly basis, RPM availability is calculated and reported quarterly, and an overall program analysis is conducted and reported annually. Data also are used on a case-by-case basis to identify, locate, or better understand specific problems or conditions. Recently, ORNL started using data analysis to evaluate programwide data to extract higher level program statistics and to develop automated measures of RPM state-of-health.

Notice: This manuscript has been authored by UT-Battelle, LLC, under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.
II. Data Management

There are a number of tasks involved in receiving, storing, and extracting the data that are necessary to make it available and accessible for analysis. This includes steps that are commonly referred to as data processing and data cleaning. The automation of these data processing steps has been an ongoing process since early batch mode analysis processes were introduced. (Ronald J. Livesay 2009) ORNL stores data in several formats that are useful for different analysis tasks. This section provides a brief description of the raw RPM data, the steps by which it is saved and transmitted to ORNL, the process of cleaning and extracting the data, and the various forms in which it is stored and finally used for analysis.

RPM data are received at ORNL in the form of “Daily Files,” which are ASCII-based text files containing a single day’s worth of data. An example of a daily file can be seen in Fig. 1. The file contains various types of data including background measurements, occupancy measurements, vehicle speed, tamper sensor state changes, fault conditions, and RPM settings information. It is organized sequentially by time with each line containing a single type of data recorded at a specific time. The type of data is identified by a two-letter code at the beginning of each line, followed by the applicable data. A timestamp is applied at the end of each line by the central alarm system (CAS) software that records the data. The data types and formats are fairly consistent for all of the RPM data received, with some exceptions due to errors and format differences introduced in the writing of the daily files by various CAS software. In 2013 daily files averaged 2.3 MB but can be over 20 MB under some conditions.

```
GS, 000036, 000033, 000030, 000030, 09:20:37.3
NS, 000000, 000002, 000002, 000003, 09:20:37.5
GS, 000034, 000020, 000031, 000023, 09:20:37.5
GS, 000039, 000030, 000035, 000027, 09:20:37.7
GX, 000121, 003850, 000000, 000000, 09:20:37.9
```

**Fig. 1. Example of a Daily File**

The data originate as ASCII-based text sent over a TCP/IP socket connection from the RPM to the CAS computer. The CAS software creates the daily files by writing the received data to individual files by RPM and date. The files are collected manually from each site and sent to ORNL via an FTP server. The received data are then processed automatically each night by custom data extraction software called AutoDART, which inserts the data into a Microsoft SQL server database. The processed daily files are then archived for later use and for long-term storage. Periodically daily files are re-extracted into a flat file format for additional analysis. Fig. 2 shows the data flow from generation at the RPM to analysis and finally to long-term storage.
The SQL database is structured to facilitate metric generation, long-term trending, and information mining. The general database structure is shown in Fig. 3. The database has two main segments: the location hierarchy and the extracted data. The location hierarchy defines the RPMs from which data may be received along with the expected base file name used to identify the appropriate files during extraction. The extracted data segment contains a list of extracted daily files as well as all of the data taken from these files. To save space and improve data recall speeds, only a subset of the overall data is stored in the database. For example, background measurements are stored at a rate of one every 10 minutes in the database but are captured at a rate of one every 5 seconds in the daily files. Also for non-alarming occupancies, only overall statistics like maximum, minimum, mean, and occupancy length are stored, rather than the full 200 ms gamma data and 1 s neutron data. Data are split among databases by year to improve performance. The most recent 3 months of data are stored in a separate database called “RPMcurrent,” which is used for most monthly reporting needs.
Fig. 3. The DART database structure is broken down into two segments: the location hierarchy, which is predefined, and the extracted data. All tables of extracted data are referenced to the location hierarchy by country, site, and lane identification.

Daily files are extracted periodically into a flat file format for additional analysis. This extraction combines an entire quarter (three months) of data for a single lane into one file using a Matlab version 7 file format. This format is convenient for the case-by-case type of analysis that is typically done using Matlab, and the format provides for some compression to make the storage requirements more manageable. Keeping a single lane per file is advantageous for parallel processing because file reads are independent when the analysis is broken down into tasks by lane. Data also have been extracted into an HDF-5 format for comparison, and this format may be advantageous for analysis using other software packages.

A rough comparison of the storage size for the various formats is provided in Fig. 4. Storage size was calculated as the average size per daily file over about 3 months of data (43,837 daily files) from the beginning of 2014. All of the flat file formats contain essentially all of the data and cover an identical set of daily files. The SQL database format contains the reduced data set described previously and consists of the same set of daily files, plus an additional 374 that were extracted after the flat file extraction. The Matlab v7 is the smallest format due to compression that is applied to files, followed by the HDF-5 formatted files. Despite the reduced data set, the SQL database is similar in size to the original daily files because of the additional lane information that is stored with the records and multiple indexes that are stored for access speed.
III. Data Analysis

Data analysis covers a variety of tasks, both automated and manual, where RPM data are used to provide information that is useful for operational planning, state-of-health monitoring, trouble-shooting, and overall program reporting in consistent formats. The majority of ORNL data analysis is conducted through a custom software solution called the Desktop Analysis and Reporting Tool (DART). In addition to providing a tool for accessing and visualizing the data (allowing for manual analysis of operational conditions, maintenance problems, and general troubleshooting), DART also provides some automated analysis to generate a green, yellow, or red state-of-health (SOH) and availability measure. On a case by case basis, specialized analysis routines can be run as queries directly to the SQL database or on the flat file data using a 32 core high performance computer platform.

The RPM SOH is determined through an automated process based on criteria designed to identify and highlight common hardware problems. The SOH is categorized as green, yellow, or red indicating, respectively, that it operates effectively, needs improvement, or has significant weaknesses. The determination is made based on 11 criteria including problems with daily files (empty or oversized), percent of occupancies with speed messages, comparison of background in adjacent detectors for both gamma and neutron, incorrect settings, fault messages, and excessive number of tamper messages. Thresholds for yellow and red status are set for each criterion. The overall SOH is the lowest of the status from each individual criterion. The SOH determination is made nightly as a part of the AutoDART process. It is based on all available data from the past month. The automatic determination is reviewed monthly by a data analyst and can be overridden.

The RPM availability calculation is another automated analysis process used to measure the proportion of time the RPM is in a functioning condition. (Tyler Guzzardo 2012) The availability metric is determined based on the received data only. Because there are many instances where the data received at ORNL are not continuous, the absence of data is not considered evidence.
that the RPM is unavailable. As seen in Eq. (1), availability is reported as the fraction of data received without evidence of a problem. Problems are defined as empty daily files, hours of missing data, hours without a background update, gamma and neutron faults, and two neutron detector signals measuring zero for 24 hours. For each problem condition a period of detection is defined, and a similar period is considered unavailable if the problem is detected. For example, “hours without a background update” is specifically monitored on an hourly period. If no background updates are recorded in a given hour of data, the monitor is considered unavailable. Similarly “gamma and neutron faults” are expected to be recorded in the file on a 5 second basis, so for each fault recorded, the monitor is considered unavailable for that 5 second period.

\[
\text{Availability} = \frac{\text{Daily File Data Received} - \text{Daily File Data with Evidence of RPM Failure}}{\text{Daily File Data Received}}
\]

DART also provides a tool for data visualization, manual analysis, and reporting. The “data dashboard” provides a quick, easy interface for accessing data on a selected lane over a selected data range. Lane metrics are automatically calculated for the selected date range, including occupancy count, gamma and neutron alarm counts, and median vehicle speed. Bar charts are provided to compare background mean and standard deviation between detectors as well as the total number of faults by type. The dashboard also provides the latest SOH reporting and RPM settings. A variety of plots can be generated, including detector backgrounds, individual alarms, and histograms of occupancy length, vehicle speed, calculated container length, background suppression, and traffic throughput. These data visualization capabilities allow an analyst to manually review RPM SOH and operational conditions as well as to troubleshoot problems to report issues and make recommendations for project teams and local maintenance providers.

Beyond the automated and manual analysis tools provided by DART, additional specialized analysis can be done on a case-by-case basis using either custom database queries or flat file analysis. This type of analysis can be conducted for a variety of reasons but typically will answer a specific question about programwide conditions, attempt to isolate and identify a particular problem, or be conducted in an exploratory manner to identify or test new analysis routines or metrics. Examples of this type of analysis include: (1) database queries run to identify typical neutron backgrounds that were used to set thresholds for the availability metrics and (2) a search of flat files for occupancies with timestamp errors indicating a possible CAS problem.

The first example comes from a programwide analysis of neutron detector backgrounds conducted to develop the availability metrics. (Tyler Guzzardo 2012) The results were used to set the threshold for determining that a neutron detector assembly was non-functioning. This analysis was performed using direct queries of the database to produce a histogram of average daily neutron measurements taken from occupancies for all available 2011 data. A histogram of the daily averages can be seen in Fig. 5. There appear to be three separate peaks in the histogram that likely result from the different number of $^3$He tubes installed in various monitors. Pedestrian monitors and later generation vehicle monitors typically use 4 $^3$He tubes, older vehicle monitors have 8 $^3$He tubes, and rail monitors have 16 $^3$He tubes. Overall the data were used to determine appropriate thresholds for identifying non-functioning neutron detector assemblies.
Fig. 5. Histogram of daily average neutron signal used in defining a threshold for minimum number of neutron signal lines required to assess state-of-health.

The second example is based on the discovery of incorrect time stamping of occupancy data, which prompted a general analysis of programwide data to attempt to isolate specific lanes and sites with this problem. The problem results in a gap in timestamps within an occupancy, followed by multiple measurements being assigned identical timestamps. An analysis of the flat file data using Matlab was conducted to identify occupancies with either a gap of greater than 1 second between timestamps or more than 4 duplicate timestamps. A significant amount of processing is required to run this analysis for all lanes. Parallel processing of the Matlab v7 flat files on a 32 core HPC machine accelerated the processing, allowing results to be generated for the fourth quarter of 2013 in only 8 minutes.

IV. Automated Analysis

The current ORNL data analysis tools and methods provide an effective means of gaining insight into a wide variety of factors that affect program goals, but these analyses are still labor intensive and rely heavily on the judgment of trained data analysts to identify problems and interpret results. The development of pattern recognition methods and use of knowledge discovery algorithms promises great potential for improving the accuracy and consistency of data analysis while also improving efficiency. Near-term results likely will be applied in a “data analyst in the loop” model, where algorithm results will provide additional tools and information for the data analysts’ review. Long-term, it is possible that algorithms of this type could be deployed in the field with the capability to automatically identify and diagnose problems or even predict them in advance.
While there may be significant promise to this approach, a number of challenges exist in implementing any such system. There are two general approaches (See Fig. 6) that could be considered: option 1 would try to isolate lanes or detectors that are behaving abnormally and then attempt to explain, or classify, each abnormality; option 2 would begin with known problem categories and attempt to design classifiers to isolate individual occurrences. The first approach has the advantage of potentially detecting and isolating problems that are currently unknown. In this approach, unsupervised learning techniques might be applied to look at natural groupings in the data. The second approach has the advantage of basing development on the existing understanding of RPM problems and failures and could make use of supervised learning techniques. With either approach, segmentation and feature extraction are likely to be necessary steps in the development of a working solution.

**Fig. 6. Comparison of two options for developing an automated data analysis system.**

Option 1 uses a method of anomaly, or change, detection to identify potential problems; option 2 is based on the classification of data into known categories.

Initial work into automated analysis routines has looked at neutron alarms programwide in an attempt to classify them into the broad categories of alarms resulting from normal operation, those resulting from detector problems, and those resulting from system testing. Neutron alarms are an advantageous place to begin work because the data are naturally segmented into individual occupancies and because the most important features of each alarm are typically contained in the measurements done at the peak of the alarm, which provides for a natural feature extraction. For this reason the initial work has been done using the ratio of 1 second neutron count rate and ratio of 200 ms gamma count rates from each detector at the point of
maximum summed neutron count rate. This produces a compact data set for which various classification algorithms can be applied and evaluated.

This feature extraction may not be optimal for all purposes because some shape information may be significant. This is particularly true for alarms that are caused by a known interaction between the gamma alarm lights and the neutron detection system in some monitors. These alarms now are rare because a programwide fix was adopted. They also appear relatively normal to operators because their occurrence is random and caused by only minor changes in the overall neutron count rate. For these reasons, in the initial automation these alarms will be grouped with other alarms that occur during normal operation. In the future classification algorithms operating on the full neutron alarm profile may be able to take advantage of shape information to separate alarms caused by this specific condition.

The classification was done using a k-Nearest Neighbor (k-NN) classifier using the three nearest neighbors and a Euclidean distance measure. A small “training” set of 91 neutron alarms was used in the development. The “training” set was drawn randomly and assigned appropriate classes by manual inspection of the alarm profiles and associated lane data. “Problem alarms” are typically those where either a single detector assembly is failing and generating an increased number of counts or where two detectors in a single side of the monitor have completely failed and read all zeroes. “Testing alarms” are identified by an increased count rate primarily in adjacent gamma and neutron detectors caused by placing a $^{252}\text{Cf}$ source on the face of the neutron detector. “Normal alarms” are everything else, including statistical false alarms and real alarms. Sometimes the appropriate category for a neutron alarm is difficult to determine based on the available data and because some subjectivity exists in the manual classification.

This initial KNN-classifier has been assessed against an independent set of 109 neutron alarms—the “test” set—that were manually categorized by data analysts. The results for this testing are given in Table 1. This classifier gave the correct response to an alarm category 90.6% of the time. The goals of this work with neutron alarms are to provide an indication for data analysts on potential problems with specific RPMs and to allow for programwide trends to be tracked. The KNN-classifier was selected for this task as an initial starting point, and other options will be explored. There also is potential for improving the results using other distance measures or by modifying the feature extraction to provide more data to the algorithm.

**Table 1. Results of KNN-classifier applied to a set of 109 neutron alarms independent from the original training set**

<table>
<thead>
<tr>
<th>Classifier Result</th>
<th>Actual</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal</td>
<td>58</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>Problem</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>Testing</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>

**VI. CONCLUSION**

Data analysis provides significant benefits by providing insight into a wide variety of factors that affect site operation and RPM maintenance. The collection, processing, and storage of RPM data is a significant undertaking, and a great deal of development has been completed to streamline this process. A number of automated processes currently provide input for data analysis and for
reporting of program level information, including the SOH and availability reports. Despite all of
the available tools, data analysis remains fairly labor intensive and relies heavily on the expertise
of trained data analysts. Automated analysis methods could provide more consistent and more
accurate analysis while also improving the efficiency of the analysis process. Initial results for
automated classification routines are promising, but significant work remains to be done.
REFERENCES


Enhanced Spent Fuel Verification by Analysis of Fork Measurements Data Based on Nuclear Modelling and Simulation.

Stefano Vaccaro\textsuperscript{1}, Ian C. Gauld\textsuperscript{2}, Jianwei Hu\textsuperscript{2}, Peter Schwalbach\textsuperscript{1}, Paul De Baere\textsuperscript{1}, Willi Koehne\textsuperscript{1}

\textsuperscript{1}European Commission, Directorate General for Energy, Luxembourg
\textsuperscript{2}Oak Ridge National Laboratory, Oak Ridge, TN, USA

Email Address of Main Author stefano.vaccaro@ec.europa.eu

Abstract:

Currently, in the EU, activities related to storage of spent fuel are constantly increasing. In fact, while in Finland and Sweden final geological repository sites are planned to be operational in 2023 and 2026 respectively, in several other EU Member States fuel is going to be moved from wet ponds to dry storage (Germany, Belgium, Spain, Czech Republic, Bulgaria and Lithuania). The required verification activities present a considerable challenge to the EURATOM Safeguards authority. Safeguards inspectorates use frequently the Fork detector for gross gamma and neutron counting, taking advantage of the relationship between those radiation fluxes and the fuel's burn up and cooling time to verify operator's declarations.

Under the framework of the U.S. DOE-EURATOM agreement on nuclear safeguards and security, a module for automated enhanced analysis of spent fuel measurement data, using the ORIGEN (Oak Ridge Isotope GENeration) code, part of the SCALE nuclear systems modelling and simulation package, has been integrated into the EURATOM-developed automated review package CRISP (Central RADAR Inspection Support Package). Measurement data are thus collected in an unattended mode by the data acquisition application RADAR (Remote Acquisition of Data and Review) and then processed by CRISP, which outputs, for each fuel assembly, the measured gamma and neutron count rates. In parallel, ORIGEN performs burn-up calculations based on operator declarations previously entered into CRISP and calculates the expected neutron and gamma count rates for each assembly. These calculations use detector response functions, developed using Monte Carlo transport modelling, to account for the detection probabilities of both neutron and photon particles that originate in each fuel pin. Finally, CRISP correlates and compares the expected (calculated) gamma and neutron signals with the measured values. The comparison is annexed to the inspection report to support drawing safeguards conclusions.

While previous works were more focused on the software aspects and on early case studies, this paper will show in-field applications of the CRISP-ORIGEN approach for safeguards inspection activities during the verification of spent fuel from different design, discussing fuel-type-specific issues in the measurement evaluation.

Index Terms— Depletion Modelling, Nuclear Safeguards, Spent Fuel Verification, Fork Detector Measurements
INTRODUCTION

IRRADIATED spent nuclear fuel is included in the range of materials concerned by international safeguards as implemented in the framework of EURATOM and Non-proliferation Treaties because of its fissile material content. Therefore, continuity of knowledge regarding the irradiated fuel assemblies needs to be maintained to ensure the integrity of the assemblies. Verifications are also needed to detect total or partial defect to ensure that no material diversion has taken place from these assemblies.

Spent fuel assemblies are kept in wet storage, usually between 5 and 10 years, in some cases up to 40-50 years. Underwater storage allows continuous cooling, decay of short-lived isotopes and provides shielding from their radioactivity. Afterwards, the assemblies are then sent to an interim dry storage or encapsulation facility for final disposal. In many European Union member States, the institutions in charge of nuclear Safeguards (namely EURATOM and IAEA) are presently facing new challenges. One aspect is related to the fact that wet storage pools are reaching their design capacity limits, therefore an increasing number of cask loading activities are planned for the years to come, requiring an important inspection effort for data collection and evaluation. In addition, the encapsulation and final disposal facilities in Sweden and Finland are intended to commence operation in the near future, also posing new challenges in spent fuel characterization for nuclear safeguards, since new and enhanced existing methods are therefore needed for the verification of irradiated nuclear material and for the re-verification of same in the future.

This work describes the implementation of the improvements to the Fork detector, one of the most widely used instruments for Spent Fuel verification. The enhanced Fork would tackle some of the above-mentioned challenges that safeguards inspectorates are facing at the moment.

In fact, the change in the measurement approach should improve the partial defect detection capability of the Fork detector, by integrating modern depletion and neutron/photon transport models for enhanced analysis of the measurement data within a fully automated data acquisition and analysis software system. This method is easy to implement and it can be applied to 100% of cask loadings, even in the case of intense loading activity. Finally, it can also be combined with other less frequently-used, more time-consuming verification techniques (e.g., gamma-ray tomography). In this paper, we discuss the development of the data analysis software and preliminary results obtained from spent fuel cask loading campaigns. This work illustrates the improvements to the Fork detector technique which can be achieved with advanced computational methods, and some of the challenges being addressed by current studies. This work was carried out by EURATOM and Oak Ridge National Laboratory under the framework of the U.S. DOE-EURATOM agreement for nuclear safeguards and Nuclear Security.

MATERIALS AND METHODS

A. Spent fuel gamma and neutron signatures

Passive neutron or gamma-ray assay of irradiated fuel cannot directly measure fissile content because the neutron or gamma is not primarily emitted by fissile material. Nonetheless, empirical relationships have been used in the past to interpret measurement data to relate burn-up, cooling time and the neutron and gamma count rates [1]. The European Commission (EURATOM) and International Atomic Energy Agency (IAEA) have been using this approach for the Fork measurements in their joint inspections all over the European Union, considering this instrument as one of the best available techniques for spent fuel partial defect verification (i.e. defect of less than a full item, which
is in this case a fuel assembly), despite the limitation that it may not be able to detect pin removal of less than 50% in an assembly [2].

In order to cope with the full complexity of spent fuel features (very short to very long cooling times, very low burnups, intermediate cooling between irradiation cycles), it appears more and more necessary to overcome the boundaries imposed by the simplified assumptions on which the empirical method is based, towards a more physically-founded approach. Several works have been made in this direction [3, 4], but a very promising approach, as already shown by recent work of Borella et al. [5], is to use modern nuclide transmutation and decay codes combined such as ORIGEN with up-to-date nuclear data to evaluate measurement data. The fuel evolution during irradiation is thus modelled in detail and the total passive neutron and gamma emission for each assembly can be calculated.

Additional calculations by suitable transport codes are needed to simulate detector signals. Detailed MCNP [6] models have been developed to simulate the transport of neutron and gamma ray, and thus the response of the detectors.

**B. Fork detector and associated electronics**

The Fork detector is a fork-shaped watertight polyethylene enclosure containing two sets of sensors in order to measure two opposite sides of a fuel assembly. In each of the prongs, there is an ion chamber for gamma flux measurement (in current mode) and two fission chambers for neutron measurement. One of the neutron detectors is covered with a thin layer of cadmium coating in order to filter out thermal neutrons and it is designed to primarily detect fast neutrons; the other neutron detector is not covered with cadmium and it is designed to detect primarily thermal neutrons. In this paper, the former is referred to as “Neutron-B” and the latter is referred to as “Neutron-A”. The ratio between Neutron A and Neutron B can be used as a correction for Boron content in water [1]. The rear part of the enclosure hosts the detectors’ preamplifiers (Fig. 1).

![Fork detector head](image)

**Fig. 1. A dismantled Fork detector head**
The detector head described above is mounted under water onto a stainless steel pipe, through which the connecting cables are fed to the electronic unit at pond side. The electronic unit is an all-in-one box, including a very compact, low power-consumption, fan-less computer, an uninterruptible power supply and a signal processing unit (Fig.2). The neutron and gamma signals are processed by the modular unit SMC 2100 by Freiberger Sensortechnik [7], which can receive up to four single-channel neutron signals and up to two current signals from ionization chambers. The same unit supplies the high and low voltage needed by the detectors and preamplifiers. The electronic unit can be sealed, enabling the system for fully unattended operation. An intuitive indication of the status of hardware operation is visible on the front panel, as shown in Fig. 2.

Fig. 2 A portable Fork measurement system indicating the status of hardware operations.

C. A standardised data acquisition and review platform: RADAR-CRISP

The Fork data are collected in unattended mode using the software package RADAR (Remote Acquisition of Data and Review), that is a modular and standardised software platform for data acquisition from different sensors. The development of RADAR has been financed by the European Commission starting in 1997 [8, 9].

The review module associated to RADAR is CRISP (Central RADAR Inspection Support package) and, like RADAR, has been based since the start of its development in 2001 on the principles of modularity, standardization and openness to 3rd party suppliers.

CRISP is a software package with several tools to configure or to access a database system for a full data analysis. The database contains all the necessary information to perform an automated data analysis; it contains plant information, item movements and operator declarations. Furthermore the database contains a number of algorithms to extract information (events) out of raw data files and process these events.

In a recent development, functionalities of the SCALE-ORIGEN package can be called by CRISP using a newly-developed module to allow automated evaluation of Fork measurement campaigns. This module, called ORELLA, takes basic information on the assembly design and declared operating history information from RADAR, creates input files for the ORIGEN code, executes that code, and returns calculated Fork detector signals based on ORIGEN depletion modelling routines and Fork detector response models [10] [11].
D. SCALE and CRISP Spent Fuel Analysis Module

1) SCALE Nuclear Systems Modelling and Simulation

SCALE is a comprehensive Modelling and simulation code suite for nuclear safety analysis and design developed and maintained by Oak Ridge National Laboratory (ORNL) [12]. SCALE is a modular system that integrates a diverse set of analysis capabilities into user-friendly calculation sequences that are designed to solve complex problems in reactor lattice physics, criticality safety, radiation shielding, and spent fuel characterization for nuclear facilities. For analysis of systems involving irradiated nuclear material, accurate spent nuclear fuel characterization is a key requirement.

SCALE includes the ORIGEN code (Oak Ridge Isotope GENeration) as the burn-up transmutation and decay analysis module to calculate the complete isotopic content, associated neutron and gamma ray emissions, and decay energy release from irradiated fuel. This is the only version of ORIGEN that is currently maintained and supported with state-of-the-art nuclear data.

2) ORIGEN Depletion Module

ORIGEN solves the complete transmutation and decay matrix currently for about 2200 individual isotopes necessary for the analysis of irradiated nuclear fuels on times scales from seconds to millions of years after fission. An assembly depletion and decay analysis typically requires 2-5 seconds, depending on the number of irradiation cycles and the complexity of the operating history. Accurate spent fuel calculations require accurate nuclear cross section data developed for the different assembly designs and reactor operating conditions. SCALE includes ORIGEN cross section libraries for most of the worlds’ commercial reactor types and assembly design classes; including PWR $14\times 14$, $15\times 15$ $16\times 16$, and $17\times 17$ designs, and BWR $7\times 7$, $8\times 8$, $9\times 9$ and $10\times 10$, including different vendor designs. Libraries for uranium-oxide (UOX) and mixed-oxide (MOX) are included. Other reactor types include VVER 440, VVER 1000, RBMK, AGR, Magnox, and CANDU. These libraries contain cross sections tabulated over a range of initial fuel compositions, burnup, and operating conditions that are rapidly interpolated for any problem-specific fuel assembly by the ARP (Automatic rapid processing) utility code in SCALE. The fuel assembly libraries are developed from detailed lattice physics models (e.g., Fig. 3) using the SCALE system.

The accuracy of these nuclear data libraries has been extensively validated using data from destructive radiochemical assay measurements of nuclide compositions of spent fuel from multiple experimental programs [13] and measurements of assembly decay heat [14].

In addition to requiring complete nuclide concentrations and activities in spent fuel, accurate calculations of the gamma ray and neutron source intensity and energy distribution are necessary. ORIGEN currently uses gamma emission libraries based on the latest Evaluated Nuclear Data Files ENDF/B-VII.1. The neutron methods in ORIGEN are based on the SOURCES code and nuclear data that include rigorous spontaneous fission and (alpha, n) capability [15].
Fig. 3. Assembly lattice model used to generate ORIGEN library for the GE14 10x10 design showing the configuration of fuel rods, gadolinia-bearing fuel rods, water holes, etc. (Different colors represent discrete material zones as simulated in the model.)

3) Fork Detector Module

As part of the spent fuel verification capability, a Fork detector analysis module, FDET, was developed by ORNL to calculate expected neutron and gamma ray signals measured by the Fork detector using the neutron and gamma ray emission rates in the spent fuel as evaluated by ORIGEN. The energy-dependent detector response functions were calculated using MCNP using detailed models of the assembly and detector configuration in the storage pond [16], as illustrated in Fig. 4.

The system calculates the predicted signals from the neutron-A, neutron-B and gamma detectors and returns these values to CRISP. In addition CRISP automatically executes checks on the data for completeness, extracts detected events from neutron and gamma measurements, creates a combined event for each assembly and links this combined (or correlated) event to the declaration provided by the operator. In a final step, extracted events from measurements and returned values from the SCALE calculations are compared and displayed in a report to the inspector [17].

Fig. 4 Fork detector response model for the 17x17 assembly design (the dimensions are also marked on the X and Y axis in units of centimeters).
E. Performance and Fuel-specific issues

The knowledge about system performance has to be established by building-up experience on actual measurement data. In fact, as shown in Fig. 5, the number and structure of uncertainty sources can be very complex, therefore a big number of assemblies have to be measured, in order to account for all the possible variations of these factors. This is of high importance, since the accurate quantification of uncertainties is needed to evaluate the total system uncertainty and to establish target tolerance values and “go/no go” acceptance criteria that can be used by inspectors.

At this development stage, variations caused by the assembly design, enrichment, burn-up, cooling time, and exposure history can be accounted for, to a certain extent, in the burn-up simulation. However, within a given assembly design class there are many variations including enrichment zoning (especially for BWR and MOX fuels), number and location of burnable poison rods for reactivity control and reactor’s control rods and absorber blades that represent additional sources of uncertainty. These details are generally not available and are not provided in operator declarations. Even assemblies with identical designs and burn-up can have different characteristics due to different irradiation exposures within the core and different local coolant and operating conditions.

![Ishikawa (fishbone) diagram of the uncertainty sources for an enhanced Fork measurement](image)

**1) PWR Fuel**

The statement above is confirmed by the first analysed measurement campaigns on PWR (LEU) fuel. For campaigns on very similar fuel, in terms of geometry, burn-up, operating history, and cooling time the discrepancies are much less scattered, resulting in a measured-calculated difference typically within ±5%. These discrepancies become larger (up to 10-15%), when less homogeneous sets are measured. For this type of fuel the uncertainties in these simulations generally consist of two main sources: (a) the neutron and gamma ray source terms and distribution in the irradiated
fuel, and (b) the neutron/gamma fluxes that reach the detectors based on the detector configuration and on the factors influencing the radiation transport. The total system uncertainties are currently being assessed using measurement data from a large set of diverse spent fuel assemblies, in a bottom-up approach that overcomes the limitations in quantifying the different single uncertainty sources. These uncertainties will be used to establish threshold for partial defect detection.

2) BWR Fuel

Although analysed on a single limited of relatively similar assemblies, the data from BWR (LEU) spent fuel exhibit, if compared with the PWR, relatively greater scattering of measured-calculated count rate deviations (within ±15%), despite the similarity of enrichment, cooling time and cycle history for these assemblies. These assemblies were, in fact, irradiated in the same three consecutive cycles and they all had the same cooling time (~4 years), same initial enrichment of 2.6% 235U, and similar burnups (from 28.8 to 32.4 GWd/tU).

The explanation of this higher results scattering resides in the fact that BWR assembly designs and operating conditions are usually much more complex than those for PWR assemblies. BWR assemblies routinely use variations in enrichment zoning, gadolinium rod loading, and are exposed to control blades during operation. In addition, the reactors operate with variable and time-dependent moderator void conditions which can have significant impacts on the actinide production rates in the fuel, and therefore the neutron multiplication rate. This also results in more variable axial burn-up distributions that have an impact on the local burn-up at the vertical position of the Fork detector measurement. In general, this level of information is not available to the inspector and the simulations must rely on typical assembly design configurations and average operating conditions. Given all these uncertainties, and the expected accuracy of the operator declarations themselves, the calculated results are considered to be generally in good agreement with measurements. Finally, it is important to ascribe to the uncertainty sources the fact that the response function used in this analysis was based on a single neutron multiplication factor of a representative spent fuel assembly, whereas the actual multiplication factors in each assembly will vary.

3) MOX fuel assemblies

MOX fuel reflects most of the features affecting calculation scattering in BWR and PWR LEU fuel. In fact it has specific designs with enrichment zoning that can heavily influence the performance of the simulation when not known or not taken into account. Another factor influencing the quality of the calculations is the level of detailed of the initial Pu isotopic vector. In fact, it can happen that the inspector is provided only with the total Pu and Pu-fissile, rather than the detailed Pu isotopic vectors with the appropriate reference dates to compensate for nuclear transmutation. In these cases a representative vector (as found in typical spent fuel) has to be used, with the introduction of a very important uncertainty source. The information about the initial Pu isotopics is indeed important, because the generation of 242Cm and 244Cm are much more sensitive to the initial amount of 242Pu than other Pu isotopes.

F. Ongoing work and future outlook

The enhanced approach Spent Fuel by analysis of Fork Measurements data based on nuclear modelling and simulation is on the way to become the usual spent fuel verification method at EURATOM. Nonetheless an intense work is ongoing to improve the usability of the tool, the accuracy of the models and to establish well defined target tolerance values and “go/no go” acceptance criteria for inspection use.
The main improvement effort to the user-friendliness of the software application is the development of automated data import functionality. At the moment, automated import of Operator Declarations is already possible using a XML file and a standard format for data is being defined in order to enable the conversion of the declarations to the XML format. This effort is not limited to the technical aspects, but involves the agreement with the operators and to a certain extent the State authorities of an acceptable data sharing protocol.

To improve the modelling calculation, a better characterisation of some factors contributing to the uncertainty has to be developed. Studies are underway to estimate the impact in the modelling of variations in radial and axial distributions of burnup, nuclide concentrations and associated neutron/gamma source terms. Future improvements to account for variation of neutron multiplication factors in the fuel are also planned.

In parallel, a measurement campaign on a set of 50 assemblies (25 PWR and 25 BWR), selected to ensure the maximum representativeness of spent fuel features, is planned for summer 2014 in order to improve the uncertainty estimate and the partial defect detection capability assessment. This study, together with an extensive pool of other measurement data, including other fuel types like VVER-440, is expected to be also a good benchmarking of the depletion and transport model.

In this perspective, the integration with the RADAR-CRISP measurement data acquisition and evaluation software would enable the safeguards inspectors from international organizations like IAEA and EURATOM to use the Fork detector as an automated spent fuel verification system.

REFERENCES


INCORPORATION OF PAGE’S TEST IN THE SEPARATION AND SAFEGUARDS PERFORMANCE MODEL*

Benjamin B. Cipiti & Michael McDaniel
Sandia National Laboratories
P.O. Box 5800, MS 0747, Albuquerque, NM 87185-0747

Tom Burr
Los Alamos National Laboratory

ABSTRACT
The Separation and Safeguards Performance Model (SSPM) uses MATLAB/Simulink to provide a tool for safeguards analysis of reprocessing facilities. The SSPM has been used for designing the overall safeguards system architecture for reprocessing, testing new measurement instrumentation, and analyzing diversion scenarios. The key use of this modeling capability is to determine the overall detection probability for various safeguards system designs. Currently PUREX, UREX+, and electrochemical versions of the model exist. The SSPM simulates materials accountancy and process monitoring measurements throughout the plant and calculates inventory differences as time progresses. Diversion scenarios can be chosen to examine the safeguards system response to material loss. The statistical test used to set alarm conditions for detecting material loss is Page’s test, which is a well-known sequential test that is effective for detecting abrupt or protracted loss. This paper describes the incorporation of the Page’s test in the MATLAB/Simulink environment, with attention to how the Simulink environment provides a visualization capability for monitoring fuel cycle facilities.

INTRODUCTION
Safeguards modeling for fuel cycle facilities requires a systems level view. The Separation and Safeguards Performance Model (SSPM) [1,2,3] was developed to provide a platform for designing new safeguards systems, evaluating the performance of improved measurement instrumentation, and performing diversion scenario analysis. This modeling effort is focused on reprocessing because these facilities face the largest challenges for materials accountancy, but other fuel cycle facilities have been considered. The SSPM simulates material flow through a facility and measurements that would or could be used for materials accountancy or process monitoring.

Inventory difference calculations across material balance areas are calculated in the model, and Page’s test of SiTMUF (Standardized Independent Transformed Material Unaccounted For) is used to set alarm conditions for material loss. This paper describes the model with attention to how Page’s test is incorporated in the SSPM. Examples of how the model is are shown to demonstrate how the Simulink environment can be used.

SEPARATION AND SAFEGUARDS PERFORMANCE MODEL
The SSPM is built in MATLAB/Simulink. The reprocessing plant unit operations are modeled as subsystems that simulate tanks filling and emptying or other processes. The signals connecting the blocks track the material flow through the plant. Elemental mass flow rates are tracked along with bulk

* Sandia is a multiprogram laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Company, for the United States Department of Energy’s National Nuclear Security Administration under Contract DE-AC04-94AL85000.
Figure 1 shows the PUREX version of the SSPM. The key unit operations are indicated as black boxes. The blue boxes represent measurement points, although this view only shows a small subset of the measurements simulated in the model. Red boxes are diversion points which can be turned on to simulate a diversion scenario.

The measurement blocks are designed to measure the quantity or concentration of interest. These can be designed to measure continuously (as in the case of a flow meter), or measure at specific times (as in the case of an accountability tank sample). The user can enter the random and systematic relative error, and the simulation uses a random number generator to assign errors to the actual measurements. The random error varies around a zero mean with a specified standard deviation and is different for each material (such as bulk liquid flow rates for aqueous plants or salt flow rates for electrochemical plants). Separations are mostly modeled through assumptions, but past work has also linked more detailed chemistry models if higher fidelity modeling is required.
measurement. The systematic error also varies around a zero mean with a specified standard deviation but is determined randomly only at the beginning of the run, and then is held constant for the entire run to simulate bias. In this manner, all measurement points are simulated.

**INVENTORY DIFFERENCE CALCULATION**
For each material balance area (MBA), an inventory difference (ID, also known as material balance or material unaccounted for) is calculated using all of the measurements within that MBA. This usually involves one input measurement, several inventory measurements, and multiple output measurements. The SSPM tracks all this data over a given material balance period. Timing is an important consideration because measurements are taken at different times. Delay blocks are used to synchronize the measurements for a particular point in time. The input and output accountability batches are summed over the material balance period, and the difference of the inventories from the previous measurement point are used. Error propagation is also programmed in the model.

This data is all fed into an embedded function block that uses MATLAB script to calculate the inventory difference. The measurement errors are also fed into the script because they are used in the SITMUF calculation. The following section describes Page’s test applied to the SITMUF sequence in more detail.

**PAGE’S TEST OF SITMUF**
The Page’s Test used in the SSPM is a somewhat complex calculation described in references [4,5,6]. This test is performed over both the U and Pu IDs, and the purpose of the test is to detect material loss while accounting for covariances that arise due to the systematic errors.

The first step is to calculate the ID for each material balance period:

\[ ID = \sum inputs - \Delta inventory - \sum outputs \]

The standard Page’s test assumes statistical independence of each value in a series of ID measurements. However, all ID measurements are correlated since the ending inventory of one balance period is equal to the beginning inventory of the subsequent balance period. There is also a correlation since subsequent measurements are using the same measurement technology. Therefore the ID series (or MUF, material unaccounted for) series is transformed into a Standardized Independent Transformed MUF, or SITMUF (see references 12 and 13).

An ID series:

\[ \overline{ID} = [ID_1 ID_2 ID_3 \ldots] \]

Has a variance/covariance matrix:

\[
\begin{bmatrix}
V_{11} & C_{12} & C_{13} \\
C_{21} & V_{22} & C_{23} \\
C_{31} & C_{32} & V_{33}
\end{bmatrix}
\]

There exists a lower triangular matrix [T] and diagonal matrix [U] such that:

\[ [T][V][T]^T = [U] \]
The ITMUF \([i]\) is calculated as:

\[
[T] \times \overline{\mathbf{D}} = \bar{I}
\]

The SITMUF is calculated as:

\[
\overline{SITMUF} = \frac{\bar{I}^T}{|U|}
\]

The Page’s test is then applied to this SITMUF series. Page’s test uses chosen \(h\) and \(k\) values that are modified to achieve the desired sensitivity and false alarm probability. The \(k\) variable changes the sensitivity, and the \(h\) value is the threshold condition to signal an alarm. For this test, \(k=0.5\) and \(h=5\) are a good starting point. Page’s test is calculated at each balance period as:

\[
S^+_1 = SITMUF_1
S^+_n = \max(S^+_n + SITMUF_t - k, 0)
\]

An alarm condition is reached when:

\[
S^+_n > h
\]

The calculation of the variance/covariance matrix is the most complex step in the process. In the following equations, \(in\) is the input, \(out\) is one of the three outputs, and \(inv\) is one of the 17 inventory measurements. The subscript \(i-1\) or \(j-1\) refers to the value from the previous time step. The subscripts \(R\) and \(S\) refer to random or systematic errors. The diagonal terms of the variance/covariance matrix are calculated as:

\[
V_{ii} = in_i^2(\sigma_{iin,R}^2 + \sigma_{iin,S}^2) + \sum_{k=1}^{3} out_k^2(\sigma_{out,R}^2 + \sigma_{out,S}^2)
\]

\[
+ \sum_{k=1}^{17} \{inv_i^2(\sigma_{inv,R}^2 + \sigma_{inv,S}^2) + inv_{i-1}^2(\sigma_{inv,R}^2 + \sigma_{inv,S}^2)\} - 2 \sum_{k=1}^{17} \{inv_i inv_{i-1} \sigma_{inv,S}^2\}
\]

The off-diagonal terms in the variance/covariance matrix are calculated as:

\[
C_{ij} = in_i in_j \sigma_{iin,S}^2 + \sum_{k=1}^{3} (out_i out_j \sigma_{out,S}^2) + \sum_{k=1}^{17} \{(inv_i inv_j + inv_{i-1} inv_{j-1}) \sigma_{inv,S}^2\}
\]

\[
- \sum_{k=1}^{17} \{inv_i inv_{j-1} (\sigma_{inv,S}^2 + \sigma_{inv,R}^2 [if j-i = 1])\}
\]

\[
- \sum_{k=1}^{17} \{inv_{i-1} inv_j (\sigma_{inv,S}^2 + \sigma_{inv,R}^2 [if i-j = 1])\}
\]

In the last two terms, the random error of the inventory term is only applied if the condition is true.
Within the Simulink embedded function, after the variance/covariance matrix is calculated, the Cholesky decomposition is used to calculate the [U] and [T] matrices. This is a MATLAB function that can be directly applied to the matrix. Extracting only the diagonal terms of the Cholesky decomposition of the variance/covariance matrix gives the [U] matrix. The transpose of the Cholesky decomposition matrix divided by the [U] matrix is equal to the [T] matrix. From there, ITMUF and SITMUF can be calculated as shown above.

**JOINT PAGE’S TEST**

Previous work by Jones [5] has examined the use of a joint Page’s test to monitor simultaneously for both abrupt and protracted material diversion. This prior work suggests choosing a large \( k \) and \( h=0 \) for an abrupt test and a smaller \( h \) with larger \( h \) for the protracted test. The abrupt test is essentially just setting a threshold condition for the SITMUF values. Past work has examined effective settings for \( h,k \) values. With a joint test, it is important to note that the combined false alarm probability (FAP) for both tests must be below 5% due to regulatory requirements. A 1% FAP for the abrupt test and 4% FAP for the protracted test will approximately satisfy that requirement.

The SSPM was used to run multiple simulations of a no diversion case to determine the SITMUF threshold for the abrupt test for a 1% FAP per year using a ten day material balance period. That value was estimated to be 2.8, so the conditions for the abrupt test are \( k=2.8, h=0 \). Tests of both diversion and no diversion cases found that \( k=0.5 \) is an effective condition for the protracted test. To keep the FAP at or below below 4%, \( h \) was estimated to be approximately 3.8. Thus the protracted test conditions are \( k=0.5, h=3.8 \).

The SSPM contains both tests which are calculated as the model runs. Depending on the diversion scenario, both tests will yield different results, but the highest detection probability will be used.

**MODELING EXAMPLE**

The SSPM has been used to examine how a particular safeguards system design will respond to material loss. Figure 2 shows an example of an abrupt removal of Pu and how the Page’s test responds to the loss. The top figure shows the abrupt Page’s test which indicated an alarm at hour 960 because the Pu SITMUF value surpassed the threshold \( k=2.8 \). The bottom figure shows the protracted Page’s Test which also indicated an alarm at hour 960.
Every model simulation provides slightly different results due to the random nature of the errors. The above results show just one simulation. Multiple simulations are run to determine overall detection probabilities. As just described, Page’s test is tuned to achieve a false alarm probability of <5% per year, so a detection probability of 5% or less would be expected in a no diversion case. If a diversion scenario is protracted and near the limit of detection, the detection probability may be near 50%. An abrupt diversion is expected to have a detection probability >95%.

While the Page’s test provides definitive results during diversion scenario analyses, other metrics can be plotted using Simulink to provide more visualization into the process. Overall measurement error, the individual inventory differences, and the cumulative sum of the inventory difference all provide useful information for plant monitoring. Process monitoring data may also be used to monitor plant conditions. Figure 3 shows an example of tank inventories for some of the key unit operations in an aqueous plant as a function of time. The material loss from Figure 2 can be seen as a reduction of inventory from the Pu Separation Buffer Tank.
Figure 3. Visualization of bulk tank inventories in Simulink

The key use of this modeling capability is to determine the overall detection probability for various safeguards system designs. New measurement technology that can decrease measurement uncertainty may be able to decrease the overall error, and thus increase detection probability for diversion scenarios. However, the results are not always intuitive since changing one measurement may only
have a minor impact on the overall system. This type of analysis can be used to guide future research on what measurement uncertainties are required to meet regulation or improve overall safeguards.

CONCLUSION
The integration of Page’s test of SITMUF has provided a key metric in using safeguards modeling to make conclusions about safeguards system designs. The overall impact on detection probability has been used to analyze new measurement technologies and propose new monitoring strategies. The specific applications have also allowed for testing of the Page’s test to work out bugs in the programming, giving confidence that the current model is performing as expected. The modeling in Simulink provides a visualization capability that allows researchers to see material loss as it is occurring.

ACKNOWLEDGEMENTS
The modeling development was funded through the Fuel Cycle Technologies program under the U.S. Department of Energy Office of Nuclear Energy. Additional funding has been received from the National Nuclear Security Administration through the Next Generation Safeguards Initiative.

REFERENCES
Developing a Validation Methodology for Expert-Informed Bayesian Network Models Supporting Nuclear Nonproliferation Analysis

Amanda M. White, Zoe N. Gastelum, Paul D. Whitney
Pacific Northwest National Laboratory
Amanda.White@pnnl.gov

Abstract
Under the auspices of Pacific Northwest National Laboratory’s Signature Discovery Initiative (SDI), the research team developed a series of Bayesian Network models to assess multi-source signatures of nuclear programs. A Bayesian network is a mathematical model that can be used to marshal evidence to assess competing hypotheses. The purpose of the models was to allow non-expert analysts to benefit from the use of expert-informed mathematical models to assess nuclear programs, because such assessments require significant technical expertise ranging from the nuclear fuel cycle, construction and engineering, imagery analysis, and so forth. One such model developed under this research was aimed at assessing the consistency of open-source information about a nuclear facility with the facility’s declared use. The model incorporates factors such as location, security and safety features among others identified by subject matter experts as crucial to their assessments. The model includes key features, observables and their relationships. The model also provides documentation, which serves as training materials for the non-experts.

The research team developed a validation study to determine if the model encapsulates sufficient expertise such that a non-expert using the model can arrive at similar assessments as an expert. If this were true, greater efficiency could be achieved by using model-assisted assessments of non-experts to augment ongoing work or triage cases needing expert attention. To support a comparison of expert versus model-assisted non-expert assessments, we fabricated five data sets for five mock facilities, including geospatial imagery, news reports, press releases, and scientific publications.

Four experts were asked to review the documentation and assess the consistency of these five facilities with their declared use, by specifying ‘High consistency’, ‘Medium consistency’ or ‘Low consistency’. Likewise, four non-experts were trained to use the Bayesian network model with evidence, and asked to attach evidence from the facility documentation to the model. The model incorporates an expert-informed logic that transforms the attached evidence to an assessment of consistency with the declared purpose (also measured in ‘High consistency’, ‘Medium consistency’ or ‘Low consistency’). A comparison of the expert and non-expert responses shows a good correspondence between the two.

This paper will describe the validation study, compare the expert and non-expert assessments, and evaluate the similarities and identify areas where further model development is needed to improve the model-assisted assessments.
Introduction

Nonproliferation assessments require the integration of a great deal of diverse knowledge, including the nuclear fuel cycle, the implications of scientific research and development, international commerce patterns, construction and engineering, and the interpretation of satellite imagery, among others. While highly trained experts may be able to assimilate some of the information, very few analysts have expertise in all of the areas required to integrate all nonproliferation-relevant information about a state, site, or event of interest. This research team has developed a variety of Bayesian network models focused on integrating diverse information to compare competing hypotheses [1-6]. This includes several models focused on multi-source signatures of nuclear programs, developed for the purpose of allowing non-expert analysts to benefit from the use of expert-informed models. One such model developed was aimed at assessing the consistency of open-source information about a nuclear facility with the facility’s declared use. The model includes several indicators of consistency and relationships between them, and provides documentation to serve as training materials for non-experts.

A validation study was performed to assess whether the model captures sufficient expertise such that a non-expert using the model can arrive at similar assessments as an expert. If this were true, greater efficiency could be achieved by using model-assisted assessments of non-experts to augment ongoing work or triage cases needing expert attention.

Bayesian Network Models

A Bayesian network (BN) is a statistical model consisting of a set of random variables and the probabilistic relationships between some of them. Each variable can take on one of a set of discrete states. A BN is typically represented as a set of nodes and edges, where the nodes are the variables and the edges represent conditional dependencies. When knowledge about the state of one variable is entered into the model, updated probabilities of the states of the other variables can be obtained. BNs can be used to compare the likelihood of competing hypotheses in light of available information.

PNNL has a history of developing Bayesian networks for assessing a variety of national security related problems, including assessing nonproliferation problems [2,3] and safeguards [4]. Analysts populate these models with available evidence in order to compare hypothesis likelihoods, or to see trends in likelihoods over time. A sensitivity analysis of a BN model can yield information about the most critical pieces of evidence used to drive changes in likelihood, or can guide collection requirements for additional information that would be most influential to the model.
Developing Expert-Informed Models

Since the goal of these BN models are to capture subject matter expertise such that it can be used by non-experts, the process of building a model depends on incorporating expert knowledge at several steps (see Figure 1) [5]. Model building begins with consulting the literature and identifying and interviewing experts in order to identify important indicators and states. When estimating the probabilistic relationships between indicators, both expert elicitation and historical data, where available, are used. In order to estimate probabilities from expert elicitation, we use an elicitation strategy called conjoint analysis, in which people compare the relative likelihoods of two scenarios [6]. The scenarios consist of combinations of indicators and states in the BN model, and the users rate the relative likelihoods of each pair on either a sliding scale or a 5-point Likert scale (Figure 2).

**Figure 2: Conjoint Analysis Elicitation Example**

1. **Indicate in the blue boxes which of the following scenarios is more likely to occur.**

   **Scenario A:**
   - Limited points of entry.
   - Personnel appear to be primarily academics and students.
   - Located in a remote region

   **Scenario B:**
   - Multiple points of entry.
   - Personnel appear to be primarily academics and students.
   - Located in or near a population center

   Scenario A is much more likely to occur than Scenario B

   Scenario A and Scenario B are equally likely to occur

   Scenario B is much more likely to occur than Scenario A
Before deploying a model for use for analytic judgments, another validation step is performed using expert knowledge; an experiment conducted for this validation step is the primary subject of this paper. The purpose of this stage of the model development is to compare the analytic assessments resulting from the model to those made by subject matter experts without the model.

Site Characteristics Consistency Model
The model under study here was developed to compare site characteristics to the declared use of a facility, to assess their consistency (Figure 3). There are five indicators in the model that affect the overall consistency evaluation:

- Security Features
- Safety Features
- Construction/Maintenance Activities
- Operational Activities
- Location

![Figure 3: Facility Features Model](image)

Each indicator has a variety of potential observables associated with it; Table 1 lists a few.

Table 1: Selected Model Observables

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Observables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Security Features</td>
<td>Limited points of entry</td>
</tr>
<tr>
<td></td>
<td>Guard shack</td>
</tr>
<tr>
<td></td>
<td>Degree of information about the facility in open sources</td>
</tr>
<tr>
<td></td>
<td>Facility visibility from public spaces</td>
</tr>
<tr>
<td>Safety Features</td>
<td>Evident shielding and transfer systems</td>
</tr>
<tr>
<td></td>
<td>Decontamination vehicles</td>
</tr>
<tr>
<td></td>
<td>Buildings’ structural support</td>
</tr>
<tr>
<td>Construction/Maintenance Activities</td>
<td>Use of machines and vehicles</td>
</tr>
<tr>
<td></td>
<td>Maintenance activities</td>
</tr>
<tr>
<td>Operational Activities</td>
<td>Personnel: military or academics?</td>
</tr>
<tr>
<td></td>
<td>Scale of operations</td>
</tr>
<tr>
<td></td>
<td>Disposal of waste</td>
</tr>
</tbody>
</table>
The purpose of the site characteristics model is to formalize expert knowledge regarding site analysis to use for training of junior analysts, or to potentially automate some analytic processes so that analysts can focus their time on the most critical analytic tasks. To use this model to assess a single facility, an analyst would collect information about the model observables (i.e. collect evidence that model observables were either true or false at this particular facility). It’s important to note that it is not expected that all observables will have information for a single facility. Upon entering this information into the model, the BN calculates posterior probability assessments of High/Medium/Low consistency with the declared use.

**Validation Study Design**

A validation study was developed to determine if the model encapsulates sufficient expertise such that a non-expert using the model can arrive at similar assessments as an expert. To support a comparison of expert versus model-assisted non-expert assessments, we fabricated scenarios for five mock facilities (four research reactors and one commercial nuclear power reactor), including geospatial imagery, news reports, press releases, and scientific publications. Participants in the study were asked to determine the consistency of the mock scenarios with a statement about the declared use of the facility.

Four experts were chosen based on their expertise in satellite imagery and other information analysis in support of nonproliferation verification. Each expert was asked to review the scenarios and assess the consistency of these five facilities with their declared use, by specifying ‘High consistency’, ‘Medium consistency’ or ‘Low consistency’. The experts were also asked to briefly summarize which information they felt was most crucial to their assessments.

Likewise, four individuals with no expertise in nonproliferation or nuclear technologies also agreed to participate. The non-experts were trained to use the BN model software. Each non-expert was given a copy of the model for each scenario, and was asked to attach evidence from the facility documentation to the model. The state of the indicator ‘Facility Features’ Consistency with Declared Use’ with the highest posterior probability (‘High consistency’, ‘Medium consistency’ or ‘Low consistency’) was taken to be the non-expert response for each scenario.

The five case studies were constructed to represent a wide spectrum of potential consistency states (high, medium, or low), as well as a variety of information availability situations, ranging from little available information (but not necessarily inconsistent) to lots of available information. For each scenario, a collection of fabricated open source information was compiled including:

- A description of the declared use of the facility
- Satellite images of the facility and surrounding area. In most cases, we included one or more close-up images of the facility, an image of the surrounding site in order to show site safety and security features, and an image of nearby population centers, if any.
- Scientific and technical literature being published by the facility. Scientific and technical literature collections were fabricated based largely on the International Atomic Energy Agency’s (IAEA) International Nuclear Information Service (INIS)\(^1\).
- Information from IAEA databases regarding the design, construction, operational characteristics, and other technical details about the facility. Templates for the fabricated data from these

\(^1\) http://www.iaea.org/inis/
databases were borrowed from the IAEA’s Research Reactor Database² and the IAEA’s Power Reactor Information Service.³

- News reports, press releases, and other open source data regarding the facility. Open source information collections were generally based on real news events combined from a variety of research and power reactors in the United States. News reports and press releases were fabricated to appear as if they came from university press departments, commercial industry, and specialized news agencies such as World Nuclear News.⁴

Many of the scenarios, most notably those intended to have high consistency, were inspired by research reactors operating at U.S. universities and other research institutes, with much of the data about the scenarios coming from the actual institutes with falsified names, images, or other identifying characteristics. Other facilities, intended to be less consistent, included an amalgamation of fabricated data with images of non-reactor sites such as the Washington State Penitentiary, located in Walla Walla, Washington.

**Study Results**

The scenario assessments by the four experts and four non-experts are shown in Table 2.

<table>
<thead>
<tr>
<th>Person</th>
<th>Scenario 1</th>
<th>Scenario 2</th>
<th>Scenario 3</th>
<th>Scenario 4</th>
<th>Scenario 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert 1</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>Expert 2</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>Expert 3</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Expert 4</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
</tr>
<tr>
<td>Non-expert 1</td>
<td>Low</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Non-expert 2</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Non-expert 3</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Non-expert 4</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>

By eye, the consistency of assessments between the experts and non-experts appears quite strong. A statistical hypothesis test was performed to formalize that assessment. The test compares hypotheses

- $H_0 =$ the distribution of assessments is the same for both experts and non-experts
- $H_1 =$ the distributions are distinct

Each column of Table 2 can, in principle, be compared (i.e. experts versus non-experts) using a chi-square test of the similarity of distributions. Two problems with this are:

1) The chi-square statistic for columns with identical assessments (e.g. Scenario 2) cannot be calculated

2) The correlation of information between the columns of the table (due to the same individuals assessing each scenario) implies there is no closed form (nor asymptotic approximation) for the sum of the 5 chi-square statistics.

---

² http://nucleus.iaea.org/RRDB/RR/ReactorSearch.aspx
³ http://www.iaea.org/pris/home.aspx
⁴ http://www.world-nuclear-news.org/
To address these issues the data was pooled across the five scenarios for the experts and also for the non-experts, giving the values in Table 3.

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert</td>
<td>5</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Non-expert</td>
<td>3</td>
<td>7</td>
<td>10</td>
</tr>
</tbody>
</table>

The chi-square statistic comparing the rows in Table 3 is 1.1; however, this isn’t interpretable using chi-square probability tables due to the point 2) above.

Therefore, a randomization test is employed. A randomization test (strongly related with permutation tests) is a test of statistical significance in which the distribution of a test statistic is estimated by creating a series of random permutations of the class labels. This gives a set of test statistic values under the null hypothesis assumption (i.e. no difference between the classes). The actual test statistic is compared to the estimated distribution to give an approximate p-value [7].

For this purpose, the randomization test works as follows. Using the data in Table 2, the row labels ‘expert’ and ‘non-expert’ are randomly permuted 1000 times (always with four of each). For each randomization, the pooled assessment values are computed (e.g. Table 3) along with the chi-square statistic on the pooled values. The proportion of these 1000 statistics greater than 1.1 is 0.45; thus, this is the estimated p-value of the alternative hypothesis $H_1$. That is, so far as the data (and form of comparison) can discern, the null hypothesis cannot be rejected and there is no statistical difference between the experts’ assessments and those of the non-experts.

If the comparison of the expert and non-expert assessments did not agree, these assessments could be used for further model calibration. In [6], the research team formulated an optimization problem using expert elicitation as an input to an optimization problem to estimate model parameters. A similar optimization problem could be formulated with the assessments of this validation study. The new objective function would be to find the model parameters that minimize the difference between the model consistency level and the experts’ assessments, using the evidence items identified by the non-experts in the model. The optimization constraints for the new problem would be the same as those used in [6]—probabilities between 0 and 1, and some probabilities summing to 1.

Another consideration in model use is the time factor. To reach these assessments, the experts each spent up to 1 hour of time, whereas the non-experts spent a minimum of 8 hours each. Follow-up interviews with the non-experts revealed difficulties in using the Bayesian network software to attach evidence to the models. They also found the satellite images difficult to interpret, and there were a few observables that required more nuclear facility knowledge than they possessed. More training would offset these difficulties, but would negate the difference between experts and non-experts.

**Conclusion**

This study demonstrates that expert-informed Bayesian network models provide value in the nonproliferation domain. Such models can assist individuals with less training to make similar assessments to those of experts. The study did demonstrate that non-experts using the model will need more time per assessment (at least initially) but one hopes that would decrease with subsequent assessments. The non-experts could also benefit from an easier interface to attach information to the BN model.
Model-based assessments could be used to support ongoing assessments and to triage cases needing expert attention. BN models can also provide a good training tool, since they encapsulate and make clear the reasoning of the experts who were involved in their development. For example, the Bayesian network demonstrates what characteristics an expert looks for, and which factors they consider most influential in the overall assessment. Overall, the development of expert-informed BN models seems to be a worthwhile investment for enhancing the kind of human assessment driven work of nonproliferation assessments.

Acknowledgments
The research described in this paper is part of the Signature Discovery Initiative at Pacific Northwest National Laboratory. It was conducted under the Laboratory Directed Research and Development Program at PNNL, a multi-program national laboratory operated by Battelle for the U.S. Department of Energy under contract DE-AC05-76RL01830. The authors would like to acknowledge the contributions of Alex Stephan, Andy Luksic, Cindy Henderson and Carrie Varley to this study.

References
VARO - The Euratom toolbox for safeguards data evaluation

W. Koehne, P. Dossogne, J. Lupo, J. Garijo, V. Canadell, D. Ancius, M. Lahogue, M. Vlietinck, M. Vassart
European Commission, Directorate General for Energy, Directorate E Nuclear Safeguards, Luxembourg
Email address of main author: wilhelm.koehne@ec.europa.eu

Abstract:
The accountancy data submitted by nuclear operators to the EURATOM Headquarters as part of the reporting requirements of the EURATOM Regulation 302/2005 need to be checked against on-site data and operational records and compared with verification measurements. In modern, complex installations of the nuclear fuel cycle, these data can easily amount to several thousand data sets per safeguards inspection, which need to be compared and evaluated for consistency.

Since manual data treatment is not an option for such an amount of data and considering the given time constraints during inspections, it was decided to start with the development of a new software application for the evaluation and verification of accountancy data of operators, and their comparison with operating records and related safeguards verification measurements.

The VARO (Verification of Accountancy Records of Operators) application aims to make safeguards evaluations and verifications coherent without limiting inspectors in their inspection scope. A generic application, centrally managed, reduces the resources needed for development, maintenance and training, thus leading to a more consistent evaluation of safeguards data.

The application uses for the accountancy data the XML data format as defined in EURATOM Regulation 302/2005, but uses for operating and measurement data predefined loading modules which can be adapted to a wide range of data formats. All inspectors using the application work with a predefined set of selection and acceptance criteria for verifications, thus ensuring that verifications in the different installations are coherent. Measurement data, resulting from hand-held, fixed or unattended instruments are fed through predefined interfaces and used to verify the operating data.

The application is part of the standard inspection software package used by EURATOM inspectors, which is available in a so called Mobile Kit, thus allowing inspectors to work almost seamlessly while being on inspection and allowing them to synchronise their inspection results upon return to HQ. The centralisation of inspection data allows for a direct comparison of their performance regarding safeguards between installations and helps focussing inspection resources.
Introduction:

The EURATOM Treaty is law in the European Union (EU) and implemented by the European Commission (EC). One of the major challenges for the Commission over the next few years will be to find the right balance between available resources, in particular personnel, on the one hand and safeguards verifications on the other.

A number of industrial operations will generate extra workload, such as the defueling of nuclear power reactors, the production and export of large numbers of nuclear fuel elements for overseas markets, the consolidation of strategic material in the UK, as well as the construction of encapsulation plants and final repositories in Finland, Sweden and Spain.

Due to the known staff reduction targets of the Commission, increased effectiveness and efficiency of safeguards tools are required to cope with the resulting constraints at a time when the nuclear material under control and activities continue to grow steadily.

The use of standardised software tools is one way to save resources in several aspects such as development, harmonisation, training, documentation, etc.

All nuclear operators in the European Union are required to maintain a system of accountancy and control for nuclear materials in their possession. According to EURATOM Regulation 302/2005, these systems shall include accounting and operating records with information on quantities, category, form, composition and location of these materials, together with information on recipients or shippers, when nuclear materials are transferred.

The amount of these data might be limited for small installations and by consequence the related safeguards verifications could be done manually. However, this has proven to be very resource demanding or even impossible in bigger or more complex installations. It is therefore essential to provide inspectors with clear guidance and efficient software tools to allow them to carry out their verifications in an efficient and coherent manner.

Generic tools required

Almost immediately when computing power became more widely available, nuclear safeguards inspectors started to use the then new technology to evaluate nuclear accountancy data and related operational records. From simple spreadsheet type evaluations these developments evolved in a number of cases into very specific and complex applications that are essential for the successful execution of inspections.

However, the constant development of the hard- and software platforms, combined with staff mobility and the evolution of nuclear installations, have made it very difficult to maintain a large number of often installation specific applications, to keep them in working order and adapted to the inspection needs.

Apart from the difficult maintenance and long-term development, these different applications created a constant need for inspectors training. When being on inspection, inspectors are facing all sorts of difficulties and the number of different software applications in use for similar tasks is a further complicating factor that needs to be avoided wherever possible.
In addition to these difficulties, the relevant reporting format was changed from 80 character lines to XML records with EURATOM Regulation 302/2005 coming into force, thus making most of the existing installation specific applications obsolete, due to the incompatibility of the input format.

The EURATOM VARO software project was started to address these difficulties and to develop a common verification tool for safeguards data that nuclear safeguards inspectors can use at a broad range of installation types as well as at headquarters to verify accountancy data against operational records and safeguards measurements performed.

**Concept**

EURATOM Safeguards uses database systems at their Luxembourg Headquarters to process, store and evaluate the accountancy declarations that nuclear operators are required to send according to Regulation 302/2005 for most of the installations on a monthly basis. These records are first checked for compatibility with the required reporting format, checked for internal consistency, and then loaded into the accountancy database. The records are then checked against the reporting rules of the Regulation and facility specific criteria stored in a so called Rulebook database.

---

**Euratom Safeguards IT concept for inspection data**

For installations under IAEA safeguards, these declarations are also translated into the IAEA format and forwarded to them.

Nuclear safeguards inspectors carry out inspections to verify these declarations by comparing the data with the physical reality on-site at the installations. In preparation for their inspections, they download the last accountancy data sent by the operator from the HQ accountancy database system (CMF) for the relevant Material Balance Areas (MBA). The deadline for transmission of these Inventory Change Reports (ICR) is the fifteenth of the following month, so there is at least a gap of 2 weeks between EURATOM's HQ and local accountancy data that needs to be updated locally. On arrival at the
installations the HQ data are complemented with accountancy data from the local accountancy records of the operator to update the VARO data to the day before the start of the inspection.

This accountancy update is straightforward if the operator is prepared to provide the required inventory change data in the XML format of the regulation. In some installations the accountancy systems do not allow interim ICRs without creating the monthly file for EURATOM, thus creating a new report number. To avoid any issues with the report numbering according to EURATOM Regulation 302/2005, all operators are asked to implement a functionality to have an interim report without an increment of the report number.

However, many operators at reactors and smaller facilities, who only declare a few lines per month, do not have their own accountancy system and use the EURATOM ENMAS (EURATOM Nuclear Material Accountancy System) application to prepare their monthly declarations. Since the accountancy systems in these installations are often paper based or use incompatible data systems, there is no other way for the inspectors to transcribe these data into the required XML format than to use the ENMAS application. During inspections this can be relatively time consuming and some inspectors prefer to do this on return in HQ. So far there is no tool existing in VARO to load data in another format than XML, but mapping tools like for the loading of operating data, as described later, might be considered later in the project to save inspection resources.

When loaded, these local records are checked for consistency and compliance with the reporting requirements to allow the inspector to establish an updated book stock. These data are then compared with operating data for consistency.

Physical verifications are normally carried out based on a stratification of the inventory to be verified. Statistical models with detection goals are used to calculate a sample size and select items to be chosen for physical verification activities. Data are loaded using mapping files to translate the operator and measurement data into the standardised internal data base format. A powerful, but easy to use, query engine has been developed to display and compare the different data sources in a coherent, common and user-defined way.

**Software development**

To be compatible with other EURATOM software projects, the VARO application is designed using web technology with an underlying Oracle database system, which is running either on a HQ server or as a mobile version using a so-called Mobile Management System (MMS), which allows the creation of Mobile Kits containing the application and the relevant data for the inspection that can be copied to a storage device. Inspectors therefore always leave HQ with the latest software version and data which are synchronised with HQ. The look and feel of the mobile version for VARO is exactly the same as with the HQ version. With the development of the MMS it is possible to control the software development centrally, thus ensuring that the infield version is fully compatible with HQ, that there is no conflict with software versions and that data are handled in a coherent way both in HQ and on inspections.
**VARO Project Phases**

Similar to other large, complex software projects with a number of different uses running over a longer period of time, it is difficult to define all functionalities in sufficient detail at the conception phase to allow for a writing of the full specifications. Since the VARO development team was relatively small in relation to the size of the project, a step approach was chosen, splitting the project in different phases. The definition of these phases was mainly based on the functionalities in question. Since the loading and verification function was the most generic of the functions required, the project was started with an accountancy tool for the loading and checking of these data. This function is used in all facility types and was the one in most urgent need. Further project phases incorporated the treatment of operating and measurement data.

However, since some of the functionalities needed in these phases are common, the start-up development effort was significant. Functionalities like the storage and sharing of user defined queries were brought forward to avoid the repeated loss of definition work, which could have caused a user acceptance problem.

The development is done in iterations, with four versions managed in parallel. The developers are working on version N, while version N-1 is being tested by professional software engineers, while version N-2 is in user acceptance testing (UAT) and version N-3 is operational. At the moment iteration 28 is operational. This sequence is ensuring a close feedback circle for the development team and allows the analyst to define the way forward in close cooperation with the project committee, which is consisting of representatives of all units of the safeguards directorate. Whereas the project committee meets at different intervals to discuss strategic issues and to review functionalities, a smaller project team meets at weekly intervals for status meetings to review user comments and more imminent development issues.

A user feedback function "Problems? Questions? Remarks?" at the bottom of the application interface allows users to send messages to the application helpdesk at any time by e-mail. Attachments can be used to explain the issue using screenshots or other documents and with the e-mail a list of the last commands used is sent to allow the development team to establish the operational background for
comments. All user comments and bug reports are followed up in a database, thus allowing the project team to prioritise and monitor progress.

This development environment is ensuring a short iteration cycle and clear guidance for the development team.

**Mobile Kit**

All new EC safeguards applications are now based on web interfaces, using Oracle databases. With the start of the development of VARO it was decided to develop a mobile working environment for essential safeguards applications to run on, called "Mobile Kit". Besides VARO, applications for seal management (ESAM) and reporting (IMIS) have or will have mobile versions, thus allowing the inspectors to use their HQ tools when being out of the office. Since the sensitivity of safeguards data does not allow to use fully web based applications, these mobile versions work within an inspection context and are run either on the inspectors' notebooks, which have encrypted hard drives and are cleared to be used for the storage, treatment and transfer of these data or on local PCs in the on-site safeguards offices. Before leaving for a mission the inspectors define the content of their specific Mobile Kits by selecting the required applications, the inspections to be covered, and the data to be copied from HQ systems. Once the selections are made, the Mobile Kit is prepared as a self-extracting jar-file serving as a data container. These files are then either copied to the encrypted hard drives of the inspectors' notebooks or any other approved storage medium for sensitive safeguards data, like encrypted USB sticks with biometric (fingerprint) user identification. The jar-files can be opened directly on the inspectors' notebooks or at bigger installations, with safeguards IT infrastructure, on a server, thus allowing multi user access. In this respect the web based applications are versatile as they are multi-user capable by design.

At the end of an inspection the preparation of an inspection outcome kit is necessary for the return of data to HQ. This outcome kit is used to upload the inspection data to the HQ database. It is to be noted that there is no synchronisation of accountancy data with HQ, but all data are stored in the inspection context for later reference.

**Leaving for an inspection**

Inspectors leaving for missions can carry in the Mobile Kits data for several different inspections. After having selected the relevant inspections, the inspector can change the predefined periods for HQ data copied to the Mobile Kit. Once the inspection data are defined the inspector can start with the Mobile Kit preparation. To protect the data a mobile kit password needs to be entered for the preparation and later installation of the Mobile Kit.
**VARO data handling**

**Working at the installation**

On arrival at the installation the inspector either uses VARO running on his inspector’s notebook or the local safeguards PC network to install the Mobile Kit and run the applications.

**Accountancy data**

The accountancy data sent to EURATOM have to be in the appropriate format, as defined in Regulation 302/2005. Many operators use the ENMAS application as provided by EURATOM, whereas others use plant, site or operator specific applications to prepare these declarations. Since the format of accountancy data is defined in detail, there is no need for plant specific data loading routines.

One of the first activities is to check whether inventory changes occurred since the last declaration was sent to HQ in order to establish the book inventory at the time of inspection. To do so, the operator is asked to provide an update of the inventory change reports (ICR) to the day before the inspection starts. These ICRs are loaded into VARO, checked for errors and can be merged with the HQ data.

It is essential that inspectors use coherent verification rules when loading accountancy data. VARO uses the same rules as the HQ application (CMF). So there is only one reference used for verification rules and the inspectors are sure that they are using the correct, up to date syntax for these checks as used in HQ and by the accountancy unit. After loading of the accountancy declarations, errors and warnings can be checked and the inspector has the option to compare and combine the accountancy declarations.
Different browse and filter options allow adapting the views on the data and help the inspectors with consistency and coherence checks.

Since the application is based on a query editor, predefined queries can be created by experienced inspectors being in charge of the installation, thus sharing the knowledge and allowing for facility adapted views and filters. All queries can be saved and selected to be shared amongst the inspectors.

Operating data

The consistency of accountancy and operating data is one of the essential checks inspectors carry out during their verifications. Operating data are very plant specific and their detail and structure depends on the plant type, operational concept and size, but also on the inspection type.

Whereas in small installations the material tracking is done mainly manually using paper documentation, more complex installations with higher throughputs use IT systems that allow for online tracking of material flows and quantities in the different areas. This serves many purposes, like process and criticality control, quality assurance, product documentation and is also important for nuclear material accountancy and control. Plants using fully integrated material tracking systems normally have a very good documentation of their operations and are able to provide this information in electronic format. However, also smaller installations use mostly electronic tools to keep their records up to date and these data are often in the form of spreadsheets or databases, which can relatively easily be converted into a format suitable for evaluation by software tools.

The VARO application is supposed to help inspectors with the evaluation of these operating data and their comparison with related accountancy declarations. However, since operational data and their structure are plant specific, it is essential to normalise them. Mapping files are used to relate the different formats of operating data to the internal operating data structure of VARO. This is another area where the use of this common software tool is of help to get to more coherent data structures. New installations providing operational data in electronic formats can use already existing data structures for which mapping files already exist.

Operating data are needed for flow and inventory verifications. Depending on the type of inspection, interim or inventory verification, and the inspection context, the required data are different. For inventory verifications the operating data consist of inventory listings and related documentation. Flow verifications are normally based on documentation related to material movements, like shipments in automated plants where fuel assemblies pass through an unattended measurement station and these results are compared with the flow information, solvent flows in reprocessing plants, weighing data in bulk handling facilities etc.

Measurement data

Physical verifications are an essential part of inspection activities. After the accountancy data have been verified against operating records, these are used to divide the nuclear material into different strata depending on the material type, category, location and available instrumentation for safeguards verifications. Depending on the inspection type, the List of Inventory Items (LII), store inventory listings or movements at strategic points of the plant are used to select items for measurements. The number of items to be selected for verifications is made either by using the Jaech sampling algorithm, depending on the required detection probability, the number of items, the goal quantity and the average nuclear material quantity of the strata or predefined percentages / fixed numbers, like 10% or 5 items out of a
Stratification and Sampling

Once the item selection is done, the inspectors prepare their paper print-outs of working papers for use in the plant. These working papers provide information specific to the verifications to be done, like area listings for tag checks sorted by location or measurement data sheets for specific instruments. There is also the option to use electronic spreadsheets as an output for the working papers, that can be filled in by the different inspection teams whilst carrying out their verifications.

After having carried out the physical verifications the inspectors have to enter the measurement data into the application. This can be done either by manual transcription from the working papers, by re-importing the electronic spreadsheets, or by importing instrument specific data files. The measurement data are linked with the operating data using item IDs or time related information.

Some of the mobile instruments in use have an internal memory for measurement results that can be exported as instrument data files. Specific mapping files will be used to import these data and feed them into the VARO application for evaluation and comparison. The use of these routines is essential to avoid transcription errors, gain traceability, and save inspection resources.

For unattended measurement systems EURATOM uses its own RADAR (Remote Acquisition of Data and Review) software, which allows for remote control of and data gathering from unattended instruments. So called DAMs (Data Acquisition Modules) are used as interfaces for the different instruments, like ID readers, Neutron Coincident Counters (NCC), High Resolution Gamma Systems (HRGS), transducers etc. The recorded data of the DAMs are collected on local PCs and at bigger installations are automatically copied to central servers in the inspectors' offices. The evaluation software CRISP (Central Radar Inspection Support Package) evaluates the recorded data and detects safeguards relevant events based
on triggering devices like limit switches, ID readers or signal patterns, like weighing processes or level readings of vessels being filled or emptied. If more than one signal is used for the event definition, like neutron and gamma measurements with an ID reader or level and density measurements for tank transfers, these are combined to establish safeguards relevant information, like element or isotopic data or material transfer quantities.

If operating data are needed to evaluate the measurement data, like isotopic data for a NCC without a HRGS, or tare weights of containers for weighing, VARO will provide these data sets to CRISP using inspector defined queries and CRISP specific protocols. After CRISP has evaluated the data and formed the measurement events, these are exported to VARO for comparison with the operating data.

Therefore, a comparison of safeguards verifications with accountancy declarations through operating records and related statistical analysis is possible. With related measurement uncertainties the inspectors get immediate feedback if and when measurement results are outside of predefined acceptance criteria and have the option to re-check or re-measure and/or seek for background information. This avoids late surprises and ensures that re-verifications are done as early as possible to avoid unnecessary interference with plant operations later. At the end of the physical verifications specific routines and listings will allow the inspectors to calculate overall statistics on measurement uncertainties.

**Query engine**

One of the main challenges of the project is to be generic enough to cover all installation types under EURATOM Safeguards and being at the same time specific enough to allow inspectors to carry out their verifications in an efficient and effective way.

One of the main features of the application is the query engine, allowing inspectors to filter, organise and display the available information in a very flexible, but specific way to support the inspectors with the detection of inconsistencies and errors. These queries are using a SQL type syntax, which is relatively easy to use with pull-down menus for the available commands and data fields. Since the queries are displayed in a window at the top of the result list, immediate feedback is given to the user when defining the criteria. Complemented by sorting and display options the user has the full flexibility to display and check the data available.
Comparison Accountancy Date with Operating Data

This query engine is used for all data sets, accountancy, operating data (moves, inventory listings) and measurement data.

The queries, sorting and display settings can be saved by each user under his/her preferences and can be made available to others by sharing them. This allows the management and facility officers to define standard verification routines without limiting the inspectors in any way to carry out the verifications under their responsibility.

Document checks and versioning

During the course of an inspection it is sometimes necessary for the operator to update the documentation provided. Since inspectors often immediately process information made available, each updated set of data causes a problem for the inspectors to know which data were changed and if these changes have an impact on the verifications already performed or the conclusions already drawn. A typical example is the List of Inventory Items (LII), which as an operational document, is sometimes updated by the operator during a Physical Inventory Verification. It is therefore essential to have a routine to check updated data files against the original and to highlight changes found. This also allows for a statistical evaluation on the quality of original documents, e.g. how precise they were, the number of errors corrected etc. However, verifications already carried out and their established relation to operating records should not be compromised. Therefore, the application has to be able to detect changes in updated data sets, whilst at the same time maintaining record relations between operating and measurement data based on the item ID.
Data transfer to HQ

Normally the inspectors enter all relevant data into the VARO application and have them available for later use, evaluation, documentation and statistical analysis. However, some Member States have very specific legal restrictions on the transfer of data related to nuclear material quantities and locations. To address these restrictions it is possible to limit the data travelling back with the inspectors to HQ by taking only pre-defined data back, i.e. without location information.

Return to HQ

On return to HQ the inspectors are required to upload their inspection data using the outcome kit to the HQ database. After this upload, it is possible to continue with the data evaluation seamlessly. The use of the mobile version allows for the inspectors' independence from on-site data and takes away the pressure to finish evaluations on-site.

Dedicated data output formats and print-outs allow the inspectors to document their inspection activities and verification results. These documents are used as annexes to the inspection reports, which is the overall documentation and evaluation of the inspection activities.

After the evaluations are finished it is the responsibility of the head of the inspection to close the inspection. This is the moment when the inspection data are fixed and considered as reference.

Data storage

Whereas the HQ database for accountancy data is to be seen as reference, similar central repositories for operational and measurement data do not exist yet. These data are normally stored at the installations and taken, if possible, back to HQ as separate files. These files have different structures, depending on their origin and are not stored in a single, compatible data format.

With the start of the VARO project it was decided to normalise operating and measurement data using an internal data structure that allows for a central storage.

However, depending on the installation type, it might be necessary for the inspectors to reference back to data received at earlier inspections or to have older measurement data available for comparison and to complement data. This is for example necessary at storage locations used at fabrication and reprocessing facilities, where isotopic composition details are needed for later in-process measurements or reverifications. Since the VARO concept is relying on HQ data as reference, it is essential that inspectors are able to take relevant operating and measurement data for predefined areas and periods with them when leaving for an inspection.

Statistical evaluation

The centralised HQ storage of relevant inspection data allows not only for a statistical evaluation in the inspection context, but also for data analysis over several inspections and years. It is foreseen to have statistical routines allowing for periodical evaluations on MBA and installation level. This will help the inspectorate to better focus the inspection effort and to adapt it regularly to requirements.
Material Balance Evaluation

At bulk handling facilities, the evaluation of the difference between the book stock and the physical inventory, called Material Unaccounted For (MUF), is an important indicator for the quality of the Nuclear Material Accountancy and Control (NMAC) system of the operator and essential to have an indication for a possible protracted diversion. Since all the data required for a MUF evaluation are available in VARO, if the operator provides the relevant operating data in electronic format, automatic routines are foreseen to be developed. An essential element of the MUF evaluation is to have the related operator's measurement uncertainties, which are needed for the sigma MUF calculation.

The operators are required pursuant to EURATOM Regulation 302/2005 to describe in the Basis Technical Characteristics (BTC) of their installation the methods for measurements, sampling and analysis and are also required to provide derived estimates of random and systematic errors with the operating records. If these uncertainties are not available, the International Target Values for Measurement Uncertainties in Safeguarding Nuclear Materials are used as an estimate. It is foreseen to implement a MUF and sigma MUF calculation using these uncertainties to support the Material Balance Evaluation (MBE) at the end of the Material Balance Period (MBP) after the Physical Inventory Taking (PIT) of the operator and its verification at the Physical Inventory Verification (PIV) by the inspectorate.

Long-term trending of the MUF and the Cumulative MUF (CUMUF) help to detect possible bias effects, operational changes, measurement deficiencies etc. With the storage and evaluation of related data it is foreseen to automate this process, as far as possible, allowing for a direct comparison of the different installations and a better focusing of inspection resources.

Co-operation with the IAEA

The functionalities of VARO are tailored to the needs of the EURATOM Safeguards Directorate, however, all of its functions are needed by IAEA inspectors as well. The IAEA is aware of the development and with the common development of the RADAR/CRIP instrument interface software the use, exchange and comparison of operating data is an essential VARO feature to encompass the different operating data formats.

The exchange of inspection data is an essential element of the EURATOM cooperation with the IAEA. The exchange of EURATOM’s inspection results is foreseen in the Safeguards Agreement and could be done by making the relevant Mobile Kit for an inspection available through the existing VPN channel used for the transmission of surveillance data.

Present status of the development

The development is split into different phases to allow for a stepwise implementation and to avoid developments that prove to be inadequate for prolonged periods of time. Phase one of the project was focused on the accountancy evaluation. This part of the application is needed at all installations and addresses the most urgent requirement of the inspectors. Phase 2 of the project is focused on the loading and evaluation of operating data and their comparison with the accountancy data. Phase 3 is now developed and in testing. It addresses the stratification, sampling and comparison of measurements with the operating data. Additional phases are foreseen for the statistical evaluations, automatic routines, and a data repository allowing data warehouse functionalities together with the other HQ applications.
**Difficulties encountered**

Due to the complexity of the project and the required start-up time until the first version came into common use, it proved to be difficult to keep stability in the project team. Because of limited contracts and existing mobility of staff within the organisation, the project management changed several times. However, resilience on the user side has ensured that the project did not lose track. Due to the differences between the installations, ranging from small laboratories handling milligram quantities to complex reprocessing plants handling tonnes of sensitive nuclear material, the use of data tools varies strongly between the different inspection groups. The harmonisation and centralisation of inspection concepts and documentation is still a matter of continued discussion with all stakeholders.

**Co-operation of all stakeholders required**

The scope of the VARO application includes, besides other elements, the aim to increase the efficiency of the limited inspection resources available. It is in the interest of all stakeholders to carry out inspections as smooth as possible, without any unnecessary delays. A fundamental requirement is an efficient and effective data processing. Whereas in almost all installations the use of computerised material tracking systems is an established practice, the data exchange with the safeguards authorities is often done in a patchwork fashion. To avoid cumbersome and time consuming manual data treatment during inspections, it is essential that inspectors receive predefined datasets, which can be loaded and evaluated automatically. To support the use of VARO, inspectors are required to agree common data interfaces with the operators. This will save time and effort at all sides, but requires an initial investment and co-operation to get these arrangements into routine operation.

**References:**

P. Schwalbach, A. Smejkal, D. Ancius, M. Boella, W Koehne, P. Chare
*Data acquisition systems in large nuclear facilities - challenges, experiences, solutions*,
IAEA Safeguards Symposium 2010, Wien

P. Schwalbach, A. Smejkal, E. Roesgen, T. Girard
*RADAR and CRISP – Standard Tools of the European Commission for remote and unattended data acquisition and analysis for nuclear safeguards*
Proceedings Symposium on International Safeguards, Wien 2006, Austria

W. Koehne, P. Dossogne, P. Meylemans, V. Canadell, M. Lahogue, S. Jurkonis, S. Ciccarello
*VARO - A Euratom software project for safeguards data evaluation*
Proceedings of the 33rd ESARDA Symposium on Safeguards and Nuclear Material Management, Budapest, Hungary 16-20 May 2011
Data Integration and Entity Resolution:
Challenges and Opportunities for Nonproliferation and Arms Control

Reid Porter\textsuperscript{1}, Christy Ruggiero\textsuperscript{2}

rporter@lanl.gov, ruggiero@lanl.gov

\textsuperscript{1}Intelligence and Space Research Division,
\textsuperscript{2}Nuclear Engineering and Nonproliferation Division,
Los Alamos National Laboratory, NM, USA, 87545

ABSTRACT

Combining multiple information sources for integrated safeguards could potentially provide a more complete picture of a State’s nuclear program than we currently have. However data integration and entity resolution are extremely challenging problems, and the effectiveness of automated tools on real data has been extremely limited. More often than not, this means analysts are left manually collating and cross-referencing data sources, and this quickly overwhels analyst time and resources. In this paper we outline some of the key information science advances that have emerged over recent years, and discuss how they might be brought together to help analysts with this problem. One potential opportunity is enabled by Statistical Relational Learning (SRL), a relatively new sub-field of machine learning that extends Graphical Models to the relational domain. The utility of SRL can be further advanced by Interactive Machine Learning (IML), a sub-field of machine learning that moves learning into the analyst environment, and collects training examples from analysts as they interact with their data. In this paper we place these advances into context with the data integration and entity resolution challenges. We also walk through the data integration and entity resolution steps that would be part of an IML facility monitoring architecture in order to illustrate how these information science advances could be applied to the specialized data integration challenges found in safeguards, non-proliferation and arms control applications.

1. INTRODUCTION

Integrated safeguards requires a continuous evaluation process in which information from multiple disparate sources are assessed in order to determine that all activities associated with a State’s nuclear program are consistently and completely declared. Multiple information sources are critical for this assessment. One of the most powerful and cost-effective sources of information is from the persistent unattended monitoring of facilities. Low cost commercial sensors, such as image and video cameras, as well as more specialized radiation detectors, are in wide spread use, and can give a continuous state of health of facilities. This data, when combined with state-supplied information, data from in-field verification activities, as well as the growing amount of open source data, could potentially provide a more complete picture of a State’s nuclear program than we currently have.

However, as the data volume and complexity increases, it also becomes increasingly difficult for analysts to synthesize and evaluate the data with a robust, consistent and defensible methodology. It is too time consuming for analysts to inspect all of the data available, so typically only a small fraction is ever
examined. In this paper we outline some of the key information science advances that have emerged over recent years, and discuss how together, they can potentially help analysts integrate data from a variety of sources into a comprehensive view of facility and State activities. To provide a concrete example of the potential opportunities, and the associated challenges, we imagine an analyst asking increasingly ambitious questions of the data:

**Do I see it?**

In any single data stream, objects, signals, or other features of interest need to be identified. In some cases this is straight forward and in other cases it is less clear. For example, a crane in a video monitoring system, or a gamma spectrometer reading indicating activity and isotope, compared to a seismic or radiation sensor signal significantly above background or a supplier manifest indicating a shipment was delivered.

**Do I see it in other data streams?**

Multiple data streams enable analysts to detect abnormalities or inconsistencies in activity by cross-referencing observables. This can also be straight forward or ambiguous. For example, overlaying a facility site map with a satellite image of the same site compared to cross-referencing shipped objects visible on unattended monitoring cameras with the supplier manifest.

**Do related data streams tell me more about it?**

A nuclear program is a related set of activities with critical interrelations and interdependences that occur at many spatial/temporal scales. The ultimate goal of data integration would be to enable analysts to efficiently explore, formulate, and evaluate questions in this space. For example, do the operational indicators and known centrifuge capacity correspond to the material accounting enrichment data, and material shipment data? Is the radiation sensor data at a reprocessing facility consistent with the activities observed through video monitoring?
2. ENTITY RESOLUTION

The questions in Section 1 require vastly different levels of data integration: from relatively straightforward object detection on a single data stream to coordinated analysis of diverse data streams from across the facility, complex and State. Information science provides a very general framework in which many of these questions can be represented in computation systems. We call this framework Entity Resolution. In our definition an Entity is an abstraction related to how an analyst would like to interpret the data. An entity often corresponds to a real world object, such as a crane in Question 1. But it could also correspond to more abstract information of interest such as facility states, processes, or activities, as in Question 3. For the purposes of our discussion it is helpful to think of entity resolution as two problems:

1. Data association: we need to identify the data sources, and the subsets of data from those sources, that are most relevant to the entity and particular question at hand.
2. Data combination: we need to use the data that was identified in data association to answer the particular question at hand.

For single data sources and simple questions such as Question 1, data association is fairly straightforward and the most significant challenge is data combination. However, as the diversity and number of data sources increases, and/or as the complexity of the entity of interest increases, the data association problem quickly becomes the most significant challenge. Note that this is true for both human analysts and for automated methods [1].

A simple example of the data association problem is shown in Figure 1, and in this case the entities of interest are vehicle shipments between facilities. We collect two pieces of data from two different sensors which could be associated with the same vehicle, and therefore a possible shipment. This is potentially very important information and this association must be maintained. However, when a third piece of data is collected, the entity resolution must be revised.

![Diagram](image)

Fig. 1: The challenge of data association when integrating multiple data sources (adapted from [1]). Two pieces of data are obtained from two different sensors that suggest that a shipment has occurred between facilities A and B. Recognizing that these two pieces of data are associated with the same real-world object is potentially very important. Raw camera data had to be processed to give a location and an object. In one case, a ‘truck,’ perhaps by size of the vehicle in the images, and in the other a license plate, which gave an (fortunately) unambiguous text translation. The third camera data was processed to give both a vehicle type and a license plate text, which shows that they are not the same vehicle through the ‘domain’ knowledge, that a sedan is not a truck. If any of these data were less clear, the entity resolution would not have been possible. For example, if the third camera only had coupled seismic or vehicle tire spacing data that could only indicate it was larger than a compact car, or if the text translation of the license was unclear on one letter, we could not be sure it wasn’t the same vehicle in all cases.
Note that even in the idealized noise free example (Figure 1) data association required subtle contextual domain knowledge of the data sources and data types to prioritize the potential associations. The biggest challenge facing automated entity resolution, and therefore data integration, is the fact that most computational tools have a complete lack of domain knowledge.

In the subsequent sections we will describe recent advances in information science that together lead to a data integration and analysis architecture for capturing and accumulating domain knowledge to improve entity resolution performance. In Sections 3 and 4 we will set the stage with a brief historical account of where machine learning and semantic representations have traditionally focused and then in Section 5 we will describe how in more recent years these technologies have found common ground in the field of Statistical Relational Learning.

In section 6 we will turn out attention back to the analyst and discuss the field of Interactive Machine Learning, which aims to exploit large-scale multi-user environments and the increased sophistication of human-computer interaction to capture domain knowledge from analysts as they analyze their data. We describe a relatively new set of human-computer interactions in Section 7, and describe how advances in statistical relational learning enable us to learn from these interactions over time. In Section 8 we walk through the data integration and entity resolution steps involved in a hypothetical data integration and analysis architecture for facility monitoring.

3. MACHINE LEARNING

Machine Learning (ML) provides tools that transform raw data into higher-level information that can facilitate search, summarization, cross-referencing and many other analysis tasks. One of the most widely studied problems in machine learning has been object detection and classification¹. The performance of ML on this problem is most impressive for very specific and well defined objects of interest where the data association problem is well defined. For example, in almost all cases, the only data we need to determine if an image contains a vehicle, is the image itself.

There has been some success in applying ML to data from multiple sources. For example, the vehicle detection in our example (Figure 1) could be improved by combining the data from the video camera with data from a seismic sensor. In cases like this the data association problem is often avoided by carefully designing the data collection system so that data streams are co-collected with respect to a common reference frame. Given this, the data combination proceeds as it does in face detection except the object detector has more input data to work with. This approach requires coordinated data collection and therefore it does not scale to the rapidly growing number of data sources and data types that are available to today’s analysts.

4. SEMANTIC REPRESENTATIONS

Technologies to support interoperability and data integration have seen rapid growth as web-standards and service orientated architectures have developed. Historically, data integration focused on databases and were developed as an alternative to data warehousing that could provide lower-cost integration of legacy systems. The key idea is to define a global interface through which analysts (and tools) access data [2]. This approach hides the complexities associated with data sources and tool chains from the user. However the interface must be specified up-front and must be updated as new data becomes available and analyst’s priorities change.

¹ ML is also applied to events, activities and other types of entities, but we will use the term object detection for clarity.
In more recent years, data integration efforts have expanded beyond database centric views of data to more abstract application-level views of data, which we call semantic representations. As an illustration of semantic representations we use the simple graphical representation of a nonproliferation domain shown in Figure 2. There are three entity types. A Facility is associated with a geographic location such as a place, factory or building. A Shipment is associated with transportation of people, information, or materials between Facilities. A Shipment also typically involves people or a vehicle. An Enterprise is associated with known commercial, industrial or academic activities that might involve multiple Shipments and/or multiple Facilities i.e. an organization.

Fig. 2: Non-proliferation applications involve data associated with multiple complex objects and relationships that must be integrated to solve complex analysis problems.

Each object has a set of attributes, as well as a set of relationships to other objects. There are many different data sources that can be used to populate attributes for each object. Data related to Shipments and Facilities can be obtained from local or remotely sensed imagery, reports, inspections or sensors. Data related to Enterprises might come from historical records, declarations, or even newspapers and other open media sources.

The promise of semantic representations has fueled much activity in semantic web technologies. Semantic representations try to represent data in terms of information that the analyst cares about, they succinctly describe different types of entities and their relationships at various scales (objects, activities, scenarios), and they improve scalability and interoperability through an encapsulated data representation. Similar to traditional data integration efforts, the interface (also called an Ontology) must be specified up front and maintained over time. But the biggest limitation of semantic representations is due to the fact that semantic queries, like database queries, operate on the assumption that data is correct, noise free and complete. This means queries are usually only applicable to carefully curated datasets, and generally not that useful for resolving entity resolution problems.

5. STATISTICAL RELATIONAL LEARNING

At a high level, machine learning and semantic representations provide complementary approaches to entity resolution. Machine learning tools can robustly populate specific attributes of specific entities in the presence of errors, noise and missing data. Semantic representations help analysts connect and query attributes of different entities that could have been populated from significantly different data sources. This can enable more complex activities to be uncovered.

These complementary approaches inspire the popular metaphor of combining bottom-up (machine learning) and top-down (semantic representations) components. The performance of systems that directly link these components is typically poor, due to the mismatch in representations that are used. Machine learning tools make mistakes but can tolerate certain amounts of noisy missing data. Semantic representations compactly represent complex relational dependencies but require perfect data.

One way to improve the match is to use a Bayesian Network for the top-down component [3]. Bayesian Networks have had a long history in knowledge representation, and eliciting domain knowledge from analysts is an ongoing activity in many applications. It is the analyst who typically knows what data is relevant to what questions, and Bayesian Networks encode this in a model with inputs that have the
same probabilistic format as the output from machine learning tools. The combination has been successfully applied in a number of applications, but Bayesian Networks lack the compactness, encapsulation and scalability of semantic representations.

Statistical relational learning (SRL) is a relatively new thrust in machine learning that tries to addresses these shortcomings. SRL generalizes Bayesian Networks (and other graphical models) to the relational domain, and thereby unifies the probabilistic and logical foundations of machine learning and semantic representations. A detailed discussion of SRL, the formal math behind it, and the methods of implementation are beyond the scope of this paper, but have been thoroughly reviewed [4]. Here, we will simply list the advantages of SRL:

1. SRL enables machine learning tools to incorporate domain knowledge encoded in ontologies [5].
2. SRL ontologies can be fine-tuned, adapted and matched over time by learning from data [6], [7].
3. SRL generalizes proven machine learning methods to tackle more difficult entity resolution problems [8, 9].

These advantages have led researchers to develop a wide range of frameworks that offer various tradeoffs between SRL model expressiveness and computational complexity. These frameworks provide new, more general ways of approaching problems, but the large variety of frameworks is an indication of just how difficult it is to match complex relational models to real data. Hand-coding ontologies is critical to encoding domain knowledge, and improving machine learning tools. However ontologies are abstract and subjective and often do not align with real data. SRL provides theory and methods for bringing the ontology and the data into closer alignment, but in many applications, the gap between initial representations is simply too large.

6. INTERACTIVE MACHINE LEARNING

Interactive Machine Learning (IML) can help narrow the gap between semantic representations and data. The key difference between IML and SRL is when and where learning occurs. SRL methods, like traditional machine learning, is formulated as an offline batch training process, where large amounts of data are collected, classifiers are designed by training on this data over long periods of time (often days or weeks), and then deployed to analysts to be used. In contrast, IML moves learning into the analyst environment, and collects training examples from analysts as they interact with their data [10]. Figure 3 illustrates the main components we have discussed so far, and indicates in orange and red, the iterative IML process that learns from an analyst to close the gap between semantic representations and real data.

The accumulation of analyst interactions (provenance data) is, of course, required to support IML. Fortunately, accumulating provenance data is already part of many data integration and analysis architectures, since it serves a variety of purposes. In some applications, provenance data is required for traceability and validation. In other applications, having a record of past work empowers analysts
to backtrack and explore alternatives. Modern architectures are also beginning to support the organization, sharing and playback of these interactions as user workflows which is useful for worker training and collaborative work. However, in IML this data is also used as training data to improve the machine learning tools and semantic knowledge base.

Machine learning has had the ability to learn from analyst interactions in the end-user environment for some time, and in fact, this ability is being used to varying degrees in a number of applications. A commonly experienced and very effective example is internet search, where user responses to search results are used to improve the search over time and profile users to better target advertising. However, in most cases the interactions are limited to labeling, and the objective is to minimize interaction.

We suggest that there are classes of applications, particularly in highly specialized domains like safeguards, arms control and nonproliferation, where experienced analysts are able and willing to interact with the analysis environments in a more sophisticated way than current systems allow. Work in support of this argument appears in the Visual Analytics community. Here researchers have been developing a set of more complex interactions that they call Semantic Interactions [11] (a name we gratefully borrow for our own purposes!)

7. SEMANTIC INTERACTIONS

The advances in machine learning methods that have brought us to statistical relational learning, have also provided a foundation for learning from a number of more complex (but still intuitive) interactions that we call semantic interactions. Semantic interactions provide increased bandwidth between analysts and the analysis architecture, which helps analysts help the machine learning tools to narrow the gap between the semantic knowledge base and the real data. In the following sections we describe the semantic interactions that we believe are particularly important to entity resolution and data integration: grouping, ungrouping and matching.
7.1 Grouping and Ungrouping

Two of the most important semantic interactions are grouping and ungrouping, or merging and splitting. These interactions are intuitive for most analysts to understand and are important for entity resolution because they enable analysts to directly manipulate what data is associated with what real-world entity. A simple example is illustrated in Figure 5. Once an analyst is satisfied with the correct data association, they can provide labels or relations. Learning methods for these interactions build on ‘learning to cluster’ methods [12]. In previous work we developed a particularly simple approach for learning these interactions for geospatial mapping applications [13].

7.2. Matching

Another important semantic interaction is matching, and it enables analysts to identify equivalent instances of the same real-world object. This interaction is often required when analysts analyze multiple images in time (the same real-world object is observed at different points in time, as in tracking) or from different views. Note, an example of the latter is image registration, and a common special case of the matching interaction is the identification of ground control point pairs. In Figure 6 we show an application example related to facility monitoring. Learning methods for these interactions are related to learning costs for the assignment problem (also called the matching problem) [14].

---

2 In some literature this, and only this, is referred to as entity resolution.
Fig. 5: On the left is the initial prediction made by the machine learning tools. Some objects appear to be correctly delineated, but the cooling tower is incorrect. Some parts of the tower have been associated with the shadow, and other parts have been assigned to separate objects. The former is corrected with a Split interaction (Top) and the latter is corrected with the Merge interaction (Bottom).

Fig. 6: Identifying equivalent instances over time enables the movement and change in objects to be monitored. In this example, linking the red gate between the two images informs the machine learning tools that the gate is often in one of two states. Also in this scenario, we see that the analyst has matched a large blue truck. In this example we imagine the analyst could not make this determination from the satellite imagery alone, and so matched these instances via data from the ground based security camera. This informs the machine learning tools that integrating security camera and satellite imagery can be accomplished at the level of vehicles, and that in this case the security camera data is more useful than the satellite imagery for matching vehicle instances.
8. FACILITY MONITORING WALKTHROUGH

We will walk through some of the interactions that help an analyst answer the question “What operational state is the reactor facility in?” using a hypothetical IML architecture for facility monitoring shown in Figure 7. In this example, three types of data systems are integrated: satellite image data, video surveillance data, and various other sensor data. We will describe the model system using predominately satellite image data since it defines the underlying geospatial reference frame.

The initial machine learning tools predict objects (object labels and edges) in a first set of satellite images. Images are further annotated and labeled by the analyst through grouping/ungrouping and labeling. The interactions are used as training data to tailor the tools to this particular image stream leading to better predictions on subsequent images. Eventually, only images with high uncertainty in predictions need to be annotated and reviewed by the analysts.

With the basic objects defined, the analyst can specify new relationships, define custom higher-level groupings, and efficiently quantify content such as object area and counts (for example, only count cars within a parking lot). Some common relationships, such as adjacency, can be learned from data, however the real value of interactive machine learning comes when the analyst identifies longer range dependencies based on their domain expertise (e.g. parking lots at 3 distant locations are known to be related to facility operations).

The analyst uses the matching interaction to associate specific objects over time. These examples are used to automate the matching of additional objects that have similar changes in position, which may due to an actual coordinated movement of equipment, or due to imperfect image registration.

The analyst flags certain image features for review, such as movement or change in an image, and provides annotations based on these features (e.g. no change, door opened) using the organizations shared vocabulary. The analyst requests vehicle counts from the surveillance cameras, and annotates a subset of the data with high counts using an activity-level attribute (high, low, or, given sufficient data, normal versus maintenance type vehicle pattern).

Finally, sensor data is linked to objects in the image data, allowing even more relations to become apparent, e.g., reactor power monitors, electromagnetic sensors in the switch yard and turbine hall, and thermal sensors in the cooling tower. More distant data is also added by the analysts based on their expertise, e.g., hotel and RV park data indicating increased contract worker presence; social media sources indicating movement of large equipment or shielded materials.

Over time, analysts build up a tailored semantic hierarchy for the domain, but they are always working through mappings to real data to do this. Defining the semantic hierarchy in this way requires more work than defining it up-front (as has been traditionally done with ontologies), but it can produce more accurate entity resolution throughout the hierarchy, and provide analysts with an understanding of when, and when not, to trust automated predictions. In addition, the ability to drive development of a tailored semantic hierarchy that is unique to each dataset while still in the same information science architecture is ideal for assessing the activities of each State’s unique nuclear program with the data available through their particular safeguards agreements.
Fig. 7: Example of an interactive machine learning architecture: Labeling and relating geospatial data with hypothetical sensors and video surveillance data helps to build a semantic representation of the facility in a normal operation state (top images). In the lower images data gathered at a different time clearly indicates the reactor is not in a normal operational state, but it takes further data integration to assess that this is not an Emergency (for example: training lot is occupied, worker lot is full, the Emergency operations facility lot (not shown) is normal), or a refueling shut down (No increased activities at warehouse, no movement of major equipment, no preparative activities/changed pattern before this state). It was, in fact, a hot shutdown that was captured in the satellite image.
9. SUMMARY

The data integration and analysis architecture described in the previous section is not as far off as it may seem. In fact, with the exception of interactive machine learning, the main components in Figure 3 appear in many commercial systems already. Palantir Technologies is one example with many of these components. In addition Palantir has developed visual analytics tools that empower analysts to explore, annotate and search complex relational representations of structured and unstructured data [15]. Our immediate future work is focused on empowering analysts to accomplish similar tasks with specialized image data sets related to facility monitoring and nuclear material forensics. Our initial experiments with analysts merging, splitting and correcting low level image segmentation have been encouraging. But it is still unclear how scalable the image domain will be in terms of knowledge base complexity and the level of abstraction in which analysts will be able to work effectively.

It is interesting to note how similar the objectives and directions of interactive machine learning are to visual analytics. Both fields try to leverage the complementary strengths of humans and machines during analysis, but from different directions. Visual analytics has emerged from the visualization science community which focuses on the user and develops tools and techniques that tailor user interfaces to the data analysis problem, with the objective of maximizing user productivity [16]. Interactive machine learning has emerged from the machine learning community which focuses on the machine and develops tools and techniques that tailor the data analysis tools to the problem at hand, with the objective of maximizing automation and accuracy.

It is also interesting to note that both fields of research appear to have converged independently on semantic interactions as a key step in optimizing interaction. The visual analytics work in this area found semantic interactions enhanced the analysts ability to explore and make sense of diverse collections of documents [11]. In our work, the motivation for semantic interactions is to capture and encode greater domain knowledge from analysts, and thereby provide analysts with greater control over their tools. The semantic interactions presented here are not identical to those used in the visual analytics work, but there is significant overlap, and this defines a rich area for future work.

References


Integrating Video and Event Information to Improve Safeguards Verification Tasks Using Commercial Off-the-Shelf Software

Ben Dabbs*, Jim Garner, Jim Younkin, and Michael Whitaker

Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831

Email: dabbsbt@ornl.gov, garnerjr@ornl.gov, younkinjr@ornl.gov, whitakerjm@ornl.gov

*This research was performed under appointment to the Nuclear Nonproliferation International Safeguards Graduate Fellowship Program sponsored by the National Nuclear Security Administration’s Next Generation Safeguards Initiative (NGSI).

ABSTRACT
In a world where sensors and cameras are ubiquitous, users can easily drown in data from the sea of sensors. The authors will discuss a commercial off-the-shelf (COTS) solution they have used to integrate COTS network video cameras with radio-frequency identification (RFID) readers to improve simulated inspections at the ORNL feed and withdrawal system. Integrating heterogeneous data has enabled the researchers to make more effective and efficient data-driven decisions. This paper will primarily discuss integrating sensor data to more efficiently verify declared operations, but it will also provide examples of how this can improve the effectiveness of detecting undeclared activities. The authors will discuss some of the features and benefits of the Genetec Security Center software package that are applicable for use by the safeguards community. Typically, international safeguards authorities have leveraged bespoke software solutions like RADAR/CRISP, DAI, or NDAR, and future implementations may also need to be bespoke due to very specific security requirements. However, the gap between commercially available security products and international safeguards requirements is narrowing, and COTS software solutions can offer suggestions and provide a vision for better ways to integrate heterogeneous data. The authors will also discuss future work to integrate camera systems currently implemented in international safeguards environments (NGSS, DCM-14, and Gemini), EOSS seals, incorporate bar code scanners, and triggering based on sensor data or analyzed sensor data.

Keywords: COTS, Data Integration, Monitoring, Surveillance

INTRODUCTION
As the use of unattended and remotely monitored systems for the facilitation of safeguards verification activities increases, the need for integrating the abundance of data collected from multiple independent sources and sensors is important enough to ensure that the analysis and review of such data can be done in an efficient manner. While specifically designed software solutions (e.g., GARS, RADAR/CRISP, DAI, and NDAR) have been developed in the past for use by international safeguards authorities [1-2], the gap between capabilities in commercial off-the-shelf (COTS) software solutions and bespoke solutions for international safeguards environments is closing quickly. While COTS solutions may not be drop-in replacements for international safeguards review tools, they provide certain advantages, including the benefit of having a large user base, which can provide increased support as well as a “think tank” for designing and developing better ways to integrate the heterogeneous data [3]. The Genetec Security Center internet protocol (IP) surveillance system provides, for example, simultaneous review of multiple video streams, motion detection, data authentication, access control capabilities, and the ability to bookmark events.
This paper describes the use of the Genetec Security Center software solution as used with the Feed and Withdrawal (F&W) Facility within the Containment and Surveillance (C&S) Lab at Oak Ridge National Laboratory (ORNL). The purpose is to use the F&W system as a test bed to illustrate the features of this software as it relates to integrating data from the monitoring of nuclear fuel cycle processes (e.g., the movement of UF₆ cylinders) and how this could improve both the efficiency and effectiveness of the inspection activities of international safeguards authorities. While this paper focuses solely on the integration of video and radio-frequency identification (RFID) events, additional types of sensors (e.g., remotely monitoring capable seals like the EOSS, barcode scanners, etc.) could be added and integrated in the same manner to provide an archive of every cylinder event for streamlined analysis and review.

**GENETEC SECURITY CENTER OVERVIEW**

Genetec Security Center is a security platform that combines IP security surveillance systems (e.g., Omnicast and Synergis) within a single user interface (Security Desk) for the efficient monitoring and review of activities. Omnicast provides the IP video management software (VMS), and Synergis provides the IP access control system that allows real-time monitoring of access events and personnel activities. Additionally, a major advantage of this platform is that both systems are capable of incorporating and supporting a wide range of third-party hardware (e.g., RFID readers, RF Code Asset Tracking, and Honeywell and Bosch intrusion detection systems) [4]. When these systems are combined, an access control event can be used to initiate a video recording, which is then archived for later review. An additional module of Security Center (Plan Manager) allows for the real-time visualization of the security environment through an interactive map.

Key features of this software include, but are not limited to, Monitoring, Area Presence, Cardholder Activities, and People Counting. The Monitoring feature allows for real-time viewing of activities occurring within the different areas through the network cameras, as well as the status of the cameras and access control readers (i.e., online or offline) and the “people” currently within an area on the map using the Plan Manager module. The Area Presence feature provides the ability to see a snapshot of the cardholders within a certain area, the Cardholder Activities feature allows every event that a “person” was involved in to be reviewed, and the People Counting feature allows a real-time view of the people within all defined areas.

**LAYOUT OF THE F&W FACILITY IN GENETEC SECURITY CENTER**

While typically used for monitoring the activities of personnel within a facility, the Security Center platform has been easily adapted to monitor the movement of mock UF₆ cylinders within the C&S lab at ORNL. In this sense, the mock cylinders (i.e., water tanks) are the “people” described above. Two sizes of water tanks (10 liters for product and 100 liters [25 gallons] for feed/tails, later referred to as 10L and 25G, respectively) are used to represent the different sizes of UF₆ cylinders, each equipped with its own unique RFID tag. Within the software, each tank represents a “Cardholder” that is assigned to a Cardholder Group. Access rules defined for these groups determine which tanks will be granted or denied access to certain areas. Within the C&S laboratory, 10 cardholders exist (four 10L tanks and six 25G tanks), as shown in Fig. 1.
As defined in Security Center, the F&W facility consists of an accountancy area, a process area, and a storage area. The accountancy area consists of large and small accountancy stations, and the process area consists of three feed stations, three product stations, and two tails stations. Each station can accommodate only one cylinder at any given time. The storage area represents a holding area for cylinders when not involved in the process or accountancy activities and allows a multi-cylinder occupancy. Each station is defined as its own area within its larger parent area. An RFID reader (defined as a door within the software) is placed at the entrance to each station to either grant or deny access. When the RFID tag is read, the system registers an event, and if the tank’s credentials comply with the reader’s access rules, the system records the entry of the tank into that area. Additionally, six COTS network cameras have been positioned to record the area events at the feed, tails, product, large accountancy, small accountancy, and storage areas.

The layout of this facility within the Plan Manager module is shown in Fig. 2. Also included in the definitions are areas for Water Conversion and Water Fabrication to simulate the transfer of UF₆ cylinders to and from conversion and fabrication facilities.
CYLINDER MOVEMENT SCENARIOS
To illustrate the many features and capabilities of this software solution, including how it seamlessly integrates video and RFID events, cylinder movement scenarios were simulated by moving certain tanks from one station (i.e., area) to another. An RFID event triggers a specific camera to record and archive a video clip of this event, and the features of Security Desk allow for analysis and review of these events. These features would allow an international safeguards inspectorate to more efficiently verify the operator’s declarations as well as more effectively detect an undeclared diversion scenario.

INITIAL TANK LOCATIONS
To perform the following scenarios, the cylinders were given an initial location within the F&W facility. Five 25G tanks (three in feed stations and two in tail stations) and two 10L tanks (in the product stations) were placed within the process area. One 25G tank and two 10L tanks were placed in the storage area. The accountancy areas were left unoccupied. Fig. 3 shows a screenshot from the archived video of tank 25G06 being loaded into the Tails 2 station. Upon an RFID event, the data archive stores a two-minute video, including one minute before and one minute after the event. When replaying this video, a baseline picture of the cardholder (tank 25G06) and an “Access granted” message are displayed in the top right of the video window. The nameplate visible in the event video can be compared against the baseline image to ensure that the item being placed in the station is indeed tank 25G06 and not a substitute tank.

Fig. 3. Screenshot of archived video showing tank 25G06 being granted access and loaded into the Tails 2 station. The recording was initiated upon the RFID event.
After initial placement of all tanks, the Area Presence feature was used to generate a report of all tanks within an area. The report for the Process Area (see Fig. 4) displays the sub-areas (e.g., Feed 1, Tails 1, etc.), the tank IDs, the time at which each tank was granted access, and a picture of each tank.

Fig. 4. Area presence of the Process Area showing which tanks are within each of the sub-areas.

LARGE TANK TRANSFER
To simulate a typical cycle of a feed/tail UF₆ cylinder, tank 25G04 was moved throughout the lab, where it was registered at the following stations/areas: Feed, Accountancy, Storage, Conversion, Storage, Accountancy, Storage, Feed, Accountancy, Storage, Tails, Accountancy, and Storage. This simulates the steps in which a feed cylinder is emptied, sent to a conversion facility for additional material, emptied again, filled with tails, and then left in storage at the enrichment facility. As seen in Fig. 5, this entire cycle can be viewed (and exported) showing the times this tank was granted access to the individual areas. The People Counting feature shows that tank 25G04 ultimately joined the other three tanks already in storage (see Fig. 6).
Fig. 5. Exported report displaying the movement of tank 25G04 throughout the different areas to simulate a feed/tails UF₆ cylinder cycle.

Fig. 6. People Counting showing tanks within the Storage Area, including tank 25G04. The list on the left also displays the number of tanks within the other areas at that time.
SMALL TANK TRANSFERS
To simulate a product UF₆ cylinder within an enrichment facility, small tank 10L03 was transferred from the storage area to the Product 1 station to be filled. During this transfer, it was weighed at the small accountancy station. Fig. 7 below shows how the Cardholder Activities feature can be used to track actions of a specific cylinder and also displays a screenshot of when the tank was granted access and registered at the Product 1 station.

![Image of small tank transfer](image)

**Fig. 7.** A Cardholder Activities report of the 10L03 tank showing when it was granted access to the Small Accountancy and the Product 1 station as well as a screenshot of the video.

ACCESS DENIED EVENT
The authors tried to illustrate a simple facility misuse scenario by moving a mock product cylinder (tank 10L03) to the Feed 1 station. This scenario represents the misuse of an enrichment plant to feed enriched material to produce material of a higher-than-declared enrichment.

By assigning access rules in the software, the small product tank was denied access to this station because it is reserved for big feed/tail tanks only. Fig. 8 shows a screenshot of the archived video where it is superimposed with an “Access denied: Denied by access rule” message. Additionally, within the software, it is possible to signal an alarm on these events and then generate a report of all alarms, thus providing a historical record of every situation in which a cylinder was registered in an unauthorized area.
SAFEGUARDS RELEVANCE
During a routine inspection, the inventory and operations declared by the operator must be verified. For enrichment plants and other facilities that process UF₆ cylinders, this includes verifying the number and type of cylinders present at the facility. This is performed at the end of each material balance period during a physical inventory verification (PIV) and currently is conducted annually to meet the International Atomic Energy Agency’s (IAEA’s) goals for detecting a diversion. In today’s world, inspectors must individually count and verify the identification of these cylinders, which is a time-consuming process. Software tools such as Genetec Security Center could significantly reduce the time required to perform item counting, thereby improving the efficiency of these verification activities. While inspectors must be careful that they continue to verify the nuclear material items and not just an RFID tag that is supposed to be associated with the item, the inspectors could easily verify the number and identification of the cylinders that reside in each area using the features described above.

In the same manner, the software’s many features provide the means to potentially improve the effectiveness of detecting undeclared activities. Within an enrichment plant, an indicator for potential production of higher-than-declared enrichment is the presence of undeclared or extra UF₆ cylinders within the process area. If such a cylinder is found during an inspection, the history of all events associated with this cylinder could be generated and viewed, including the associated video footage, to determine why it is there. This would allow the inspectors to effectively conclude whether or not illicit activities have transpired. The system described does not try to authenticate the RFID events to international safeguards standards or try to ensure all station access attempts are recorded as events. If our system included sensors capable of capturing all attempts to place a cylinder into a station, a report of all “Access denied” events or alarms could be generated to determine if the operator attempted to place a cylinder in an unauthorized location.

CONCLUSION
The features of the Genetec Security Center software solution with regards to integrating video and event data have been illustrated. Using COTS network video cameras and RFID tags and readers, the software was used to analyze different cylinder movement scenarios in the F&W facility. The many...
features of the software were shown to provide information that could assist international safeguards authorities in increasing the effectiveness and efficiency of their inspection activities. In the future, integration of additional sensors (e.g., EOSS seals, barcode scanners, etc.) and currently utilized camera systems (i.e., NGSS and DCM-14) could be incorporated. Software programs such as Genetec Security Center offer cutting edge user interfaces and data integration solutions that should influence the development of future international safeguards software solutions.

REFERENCES
Nuclear Forensics Driven by Geographic Information Systems and Big Data Analytics

Clair J. Sullivan
Department of Nuclear, Plasma, and Radiological Engineering
University of Illinois at Urbana--Champaign

ABSTRACT

Numerous entities have recently established themselves to enable citizen scientists to collect radiation data that is temporally and geographically tagged. The available detectors include commercially-available systems, self-assembly kits, and even smart phone applications that can be purchased to turn the camera of the phone into a crude radiation detector. These developments, fueled by events like the Fukushima nuclear disaster, are resulting in the creation of large, GIS-based radiation sensor networks – big data. This paper focuses on how such networks can be leveraged for nuclear forensics purposes. We shall present data from some of these networks and describe how the coupling in of other data sets is necessary for nuisance alarm adjudication. Emphasis will be placed on the benefits of GIS techniques and big data analytics for improving the statistics provided in the receiver operating characteristic (ROC) curve.

Introduction

Radiological standoff detection of special nuclear material (SNM), shielded or otherwise, is challenging because the signal at large distances is small, the background and noise are large, and signal acquisition times need to be short (thus further hampering the statistics of detection). Current practice involves searching for SNM either through passive detection of the characteristic gamma rays and/or neutrons emitted from the material or through an active approach where a target or region is interrogated with probing radiation (x-rays, gamma rays, neutrons, or cosmic muons) and measuring reactions products from the interaction of the probe with SNM.1 Both of these approaches are limited by a fundamental problem: the amplitude of the radiation signal of interest decreases as the square of the distance to the source, known as the “1/r² problem.”

The majority of current research into both passive and active detection has focused on the development of new materials for radiation detection. However, despite the advances in detection materials, the majority of fielded gamma-ray detectors still rely on sodium iodide (NaI, developed in the early 1950’s), high-purity germanium (HPGe, developed in the 1970’s), and a variety of plastic scintillators (developed in the 1950’s). Deployed neutron detectors are further limited to ³He proportional counters (developed in the 1960’s), although there is much work to identify replacements for these detectors as the world-wide supply of ³He dwindles.1,2
The reason these relatively-old technologies are still used as the deployable detectors of choice is because of one key quality – their detection efficiency. It is not possible to solve the $1/r^2$ problem using a single detector. Therefore, organizations like the International Atomic Energy Agency, Department of Homeland Security, and Department of Defense are extremely limited in their detector selection because having detectors with high efficiency is the only way to decrease the impact of having fewer incident quanta of radiation at a large distance when only a single detector is used.

To combat these problems, there has been research focused on using networks of radiation sensors to localize sources that are either fixed or moving. These networks have relied on small, stationary detectors or detectors with limited mobility. Additionally, the density of these sensor networks was reasonably small, implying that each sensor was still individually limited by $1/r^2$. Further, inter-network data exchange and analysis was not explored in depth.

In this paper, we propose thinking about radiation sensor networks in a new way. Most people view the standoff problem as a problem of trying to find a source using a single, stationary (or minimally mobile) detector or a limited number of sensors in a statically-configured network. Using this approach, it is clear the efficiency is the most important attribute of any single detector since it is assumed to be most likely that the source will not be in close proximity to the detector. In this paper, we will explore ways to cut down on the distance by taking advantage of a potential myriad of existing radiation detectors whose network density can be significantly larger than currently-deployed detectors. Further, we will discuss the benefits of treating radiation data as one data layer among many other layers in a “big data” environment. Emphasis will be placed on key attributes of the overall data picture that can potentially contribute to nuclear forensics solutions. As motivation for future research, we will present some existing networks and potential networks of radiation sensors that can be evaluated immediately in such scenarios.

**Relevant Existing Technologies**

The current network of radiation detectors deployed worldwide for nuclear forensics through programs like the Department of Homeland Security’s Global Nuclear Detection Architecture largely consist of stationary portal monitors and other similar counters. Due to the low geographic density of detectors, there is no overlap of the response of each detector to source activities of practical concern. Rather, each detector functions individually since there would be no correlation between measurements of neighboring detectors.

However, it is possible to create a very dense network of sensors based on commercially-available technology. Several companies are selling smartphone apps to turn the camera of the phone into a radiation detector. Some have extended this same approach to web and security cameras. Consider if every cell phone and security camera within an environment like a city were a radiation detector. Some of these devices are at fixed positions whereas others are mobile, creating a dynamic network of sensors. It is clear that such a network could potentially generate a great deal of data, which could include the time-tagged GPS coordinates of each individual measurement.
Additionally, as a result of the Fukushima accident, many commercial entities have emerged to provide citizens with the tools necessary to measure and monitor radiation in their own environment. The result of these commercial solutions in an increasingly-rich network of inefficient detectors, coupled together to form highly-informative, real-time pictures of the radiation environment. Essentially, these companies and the citizen scientists obtaining measurements for them are creating a layer of big data that represents a radiation map of the world. Some have offered their data for public download.

**Smartphone Apps**

There are several commercially-available apps that have been created that can turn a smartphone into a radiation detector. Some use the microphone of the phone to listen to the audible clicks made from a detector like a Geiger-Muller counter while other use the onboard camera(s) of the device. The later is based on the physics of the semiconductor within the phone. When ionizing radiation (gamma rays in this case) interacts within the silicon of the CCD or CMOS device, it creates a small burst of charge. In an image, a single radiation interaction event appears as a single bright pixel for a brief period of time. Thus it is possible to count the number of bright pixels per time interval and correlate that to radiation exposure or dose.

For this work, we explored a variety of apps available on the iTunes store and shall present the results of a few of these apps below. (Similar apps exist for Android phones, but were not considered in this study.) Data was collected using an iPhone 4. The apps that were selected for study were chosen based on their ability to collect data using the onboard camera(s) over a short time interval (less than 1 minute) and output their numeric measurement in appropriate units. In the discussion below, we shall consider two apps that met these requirements: “GammaDetector,” which provides measurements of counts per minute (cpm), and “RadMonitor,” which provides a dose measurement in μSv/hr.

Each app was used to measure a 29.1 μCi $^{137}$Cs source for the pre-programmed integration time at a variety of distances. Results of these measurements are shown in Figure 1. For GammaDetector, the average count rate on contact was measured to be 1699±392 cpm with an average background of 15±6 cpm. For RadMonitor, the average dose on contact was measured to be 630±101μSv/hr and the average background was measured to be 18±4 μSv/hr. Based on this data, the intrinsic efficiency of each app was measured to be $2.599 \times 10^{-5}$ and $7.423 \times 10^{-3}$ respectively.
It is instructive to consider what these efficiencies imply for source detection in a very basic, assumed network of smartphones. To illustrate this, consider if every smartphone on the so-called Magnificent Mile in Chicago, IL was operated using one of these apps. It is estimated that 22 million people visit the Magnificent Mile per year.\footnote{23} (Note that this does not include the population of people who work or live there.) If we assume the average daily rate of visitors and then further assume that 10% of them are on the street at any given time, the density of smartphones (assuming a one-dimensional problem) would be 4.9 phones per linear meter.

It is important then to compare this to the efficiency numbers measured above to understand whether each detector is operating individually within the $1/r^2$ problem or whether the network is dense enough to overcome this limitation. To do so, we must explore what the detection range would be for an assumed source activity. For illustrative purposes, let us assume that a detection was the result of a deviation from background of greater than 3σ and use the measured values of background above for each app. For a 1 Ci $^{137}$Cs source, GammaDetector would measure greater than 3σ above background at 1.32 m whereas RadMonitor would measure it at 0.45 m. Given the assumed sensor density above, several detectors would likely simultaneously detect the source, implying that the $1/r^2$ problem is eliminated with this sensor density.

**SafeCast**

It is possible to study this problem from the opposite direction, assuming a single, mobile detector or series of uncorrelated individual detectors. After the Fukushima Daiichi accident, a group of citizen scientists started a movement called SafeCast to collect time-- and geo-tagged radiation data as a global sensor network and make that data publicly available.\footnote{21} This non-profit organization provides kits for building Geiger-Muller counters and an open-source API for uploading data taken from commercial detectors assuming they are time- and geo-tagged. Additionally, they provide a smartphone app for making maps of their data. Some screen shots of that app are provided in Figure 2. As can be seen from the figure, not only do they have a great deal of members providing data, but they also have the ability to incorporate other data sets, such as the fly-over measurements conducted by the Department of Energy shortly after the accident.
Figure 2. Data plotted from SafeCast app for all of Japan (left), the Fukushima region (middle), and the DOE/NNSA-collected measurements (right).

While the original intent of SafeCast was to collect data in Japan, many other users around the world have been uploading data to make it a truly global sensor network. To date, the database contains approximately 15 million individual measurements that have yet to be statistically analyzed or leveraged.

**Relevant Big Data Concepts: “The Four V’s” and GIS**

The traditional definition of big data characterizes the data based on "The Four V's:" volume, velocity, variety, and veracity. In the example of smartphone radiation detectors, it is clear that the potential volume exists and the data and analysis from each device is moving at tremendous velocity. The veracity is also clear -- if a sensor network of smartphones were to be made, there would clearly be a great deal of uncertainty associated with fluctuations in background, nuisance alarms from the presence of legitimate radioactive sources such as certain commercial goods and people who have recently had a medical uptake of radiation, etc.

The true opportunity presents itself when considering the variety of data available for analysis. This potential is easily illustrated through an example analysis from the open source data set available through SafeCast. For this example, we have selected a single series of measurements taking by a mobile sensor that drove from Tennessee to Michigan one day and then returned by the same route a day later. While this data was obtained with a single moving detector, it could also be explored by thinking of the data as obtained through a network of detectors along the same route that simultaneously took one measurement.

The count rate as a function of time obtained by this detector is shown in Figure 3. It is difficult to make sense of this data just as a single detector's count rate as a function of time. During this time interval, the average number of counts measured was $29.14 \pm 9.59$ counts/5 seconds. Many fielded detectors such as radiation portal monitors can use this information to set alarm values, although frequently this is done by obtaining a moving-window average to better quantify background. Several different averages are shown in Figure 4 based on simple window averaging of a variety of different window widths. As can be seen from the longer window data in this figure, the overall background values increase slowly towards the end of the route and then decrease as the detector returns to it's starting location.
Figure 3. Counts measured with a 5 second integration time. In this figure, measurements are not shown when the detector was not moving.

It is instructive to be able to plot this data on a map. One key feature of the SafeCast data set is that all of the data is time- and geo-tagged. Figure 5 shows how the count rate changes with position based on the available GPS coordinates. Based on this map, it is clear that the radiation counts steadily increased as the vehicle drove towards Michigan and then symmetrically decreased when driven away.
Figure 4. Count rate as a function of time showing smoothing with a simple, moving average approach and a variety of different window widths.

Figure 5. Map showing variation in count rate as a function of GPS coordinates.
While interesting, if radiation counts are the only data used for analysis, there is not enough information to determine the complete picture. One could, for instance, plot the radiation count rate as a function of velocity to see if longer dwell time had impact on the counts, as shown in Figure 6. Another hypothesis is that the natural background due to the presence of naturally occurring radioactive material (NORM) could potentially be higher in the regions with higher count rates. However, as shown in Figure 7, this is not the case. As is evident from this map, the radiation counts in Ohio should be greater, but this was not seen in the SafeCast data set.

**Figure 6.** Analysis of correlation between count rate (green) with velocity (blue) for the window widths used in Figure 4. The time centered around a normalized index of 0.3 indicates when the system was turned off over night.

**Figure 7.** Distribution of NORM radiation across the U.S.\textsuperscript{24}
A good clue to the potential cause of the higher background can be found if one considers the weather along the route for those days. A radar map of the United States is shown in Figure 8. As can be seen from the map, the northern portions of this route had precipitation (most likely rain) during this time. Rain is known to cause an increase in background due to the leaching of gaseous NORM from the soil as a result of the increased soil moisture. While it is not possible to 100% determine the exact cause of the elevated background, this scenario is plausible.

Figure 8. Radar map showing precipitation at northern portion of the measurement route.

Another example of a complicated radiation environment can be seen when considering measurements in the region surrounding the Fukushima Daiichi reactor. The SafeCast data is shown in Figure 9 for a single day in the region of the reactor and neighboring communities. As can be seen from the figure, there is significant variability in the values of the measured radiation. To elucidate this information further, it is helpful to consider how the radiation count rates have varied in time in this region. A similar region of data is shown in Figure 2 showing how the regions of higher count rates have remained consistent. In particular, it is useful to note the geographic regions of lower dose rates after the accident, such as the region to the southwest of Sukugawa. It is obvious in this example that if lower concentrations of isotopes like $^{137}$Cs were deposited in a region from the accident, then we would expect to see a lower dose measurement there several months after the fact. But this also points to an important concept that can be used by a time-evolving, dynamic sensor networks. There are regions of the world known to have higher background, whether it be from variations in NORM in the soil, the presence of large marble structures, or the regular production of radioactive materials such as for a medical isotope production facility. As is immediately visible in the SafeCast data, these regions can be quantified a priori within the dynamic sensor network and used as the basis for future background observations.
Figure 9. Radiation count rate information for a single day near Fukushima taken from the SafeCast data set. Note the area southwest of Sukagawa with lower count rates corresponding to the regions of lower $^{137}$Cs concentration as seen in Figure 2.

Needs for Future Research

While the conclusions that can be drawn from the Fukushima example are seemingly obvious, it is instructive to consider their implications on the overall operation of sensor networks of the future. Should it be possible to create a network of sensors with the density of every smartphone in a city, it will be very important to be able to accurately determine background in order to assess potential alarms. One very practical way to do this is to use the network itself to determine the background. For example, if one sensor passed by a certain point at one time, we would expect a similar sensor to measure similar background within a short time later. This requires that the sensor network must be able to communicate data from sensor to sensor in real time. This is shown again using the SafeCast data in Washington, DC in Figure 10. There are regions in this figure illustrating higher background such as the World War II Memorial and steps of the Capitol Building, both of which have significant amounts of marble. Any sensor who GPS coordinates correspond to one of these locations should automatically adjust its alarm settings to a higher threshold. In this way, the network can inform itself on background variations through regular measurement. Additionally, as measurements are collected over a long period of time, a map could be made of the natural background variations that can further inform individual measurements. For example, if a sensor passed by a marble statue, it would know from prior measurements of that region that the background was always higher in this region.
As was demonstrated in Figures 3-8, without additional data, it can be difficult to adjudicate whether and alarm was coming from a normal variation in background associated with either weather or geography or rather an actual source. This example took into account both weather conditions as well as the known maps of terrestrial NORM in an attempt to understand the background fluctuations within the data. However, it would also be helpful if the time evolution of the radiation profile within the network could be analyzed in the context of available GIS data. This concept is graphically shown in Figure 11. In this example, the GPS positions of a series of sensors (which are of fixed position in a rectilinear grid for this example for simplicity) are shown superimposed on a notional map. If sensors 1 through 11 in Figure 11a detect an event above background sequentially, then it might be suspected that there is a source moving along the road. However, if the same network detects an event that is not sequential (i.e. the positions of points 1 through 11 are randomized), the cause could be attributed to background fluctuations. Assuming the detections are sequential, the velocity of the potential source can be inferred. This velocity could then be compared to the local traffic conditions to further adjudicate whether the source was moving at the speed of vehicular traffic or if the speed was more consistent with a pedestrian’s walking speed.

Additionally, a map of the source movement should be considered. If, as in the case above, sensors 1 through 11 detect sequential hits, it can be observed that they all lie along a road. However, further information would be required if, for example, the scenario of Figure 11b occurred where several detectors simultaneously detected an event. In this case, the movement of the source does not correspond to the known positions of the roads. However, based on the fact that there is time correlation of a movement along a path, one might suspect a background variation from weather similar to that shown in Figure 8. In this case, it would be necessary to correlate the direction of movement of the potential source with the wind direction and other weather data. Additionally, in this example a potentially-broad area of sensors detected the event. If the detection range of the smartphone-based sensors is truly as small as was shown in Figure 1, then this would suggest a very distributed source of radiation, which would further motivate the conclusion of a weather-related background fluctuation.
A key problem in big data as applied to sensor networks is where the data should be processed. On one hand, any big data system is limited by its velocity -- there is only so much bandwidth that can be consumed by the problem. This would motivate allowing the smartphones to process the majority of information. However, as shown from the previous examples, it is also necessary to evaluate in real time the state of the network as a whole. This would suggest the opposite approach, where a central server or cloud network should handle the majority of data analysis. (Bandwidth limitations suggest the use of a distributed cloud computing environment is much more scalable than the use of an individual server.) Finding the optimum division of computing resources will require significant research.

Lastly, it is important to mathematically quantify whether the proposed networks and data layers truly make solutions that are better than what exists today. Since any proposed approach is rooted in statistics, there are potentially clear mathematical metrics for success. The Receiver Operating Characteristic (ROC) curve provides a direct method for measuring the probability of detection versus the probability of false alarm. Through careful simulation and measurement of ROC curves for the complete system, overall improvement in the detection of SNM using this approach can be quantified. Any future research done in this subject should, therefore, demonstrate improvements in the ROC curve as a function of algorithm.

Conclusions

With the growth of concepts such as geospatial information systems, big data analytics, and sensor networks, a great deal of potential exists for taking a leap forward in improving the problem of standoff detection of radioactive materials that is presently limited by the $1/r^2$ problem. Data sets already exist that can be analyzed for these purposes and more data can easily be obtained using commercially-available apps and smartphone technology. The availability of these tools provide the ability to instantly create a network of potentially billions of sensors. However, this cannot be done without a concerted effort to analyze the statistics of the problem and how additional data layers beyond radiation measurements can be used to adjudicate alarms. Fundamental to this problem is developing an understanding of how to measure background and assess deviations from it for a potential threat. We have shown some examples using orthogonal data sets in an attempt to understand the available radiation data with some suggestions for studying the time evolution of both the network and the individual sensor. The importance of inter-network communication was also demonstrated.
Acknowledgements

The author would like to recognize the valuable assistance provided by Karl Roth, Dr. Shaowen Wang, Dr. Yan Liu, and Dr. Jing Gao who all helped in the GIS analysis of the Safecast data.

References
