

Applying Digital Library Technologies to Nuclear Forensics

Electra Sutton, Chloe Reynolds, Fredric C. Gey, Ray R. Larson

School of Information and UC Data
University of California, Berkeley
Berkeley, California, USA, 94720-4600
electra@berkeley.edu, chloe_reynolds@ischool.berkeley.edu,
gey@berkeley.edu, ray@ischool.berkeley.edu

Abstract. Digital Libraries will enhance the value of forensic endeavors if they provide tools that enable data mining capabilities. In fact, collecting data without such tools can result in investigators becoming overwhelmed. Currently, the quantity of highly dangerous radioactive materials is increasing with the advancement of civilizations' scientific inventions. This creates a demand for an equivalently sophisticated forensics capability that prevents misuse and brings malicious intent to justice. Our forensics approach applies digital library and data mining techniques. Specifically, the forensic investigator will utilize our digital library system which has been enhanced with advanced data mining query tools in order to determine attribution of material to their geographic sources and threat levels, enabling tracing and rating of smuggling activities.

1 Introduction

Radioactive materials can cause harm in many ways, from poisoning to bombs. Today these materials are abundantly applied in the medical, energy, weapons, and space exploration fields. Keeping populations safe from their malicious use is the job of forensic investigators worldwide. Recent records from the Illicit Trafficking Database (ITDB) at the International Atomic Energy Agency (IAEA) show that high-value HEU (Highly Enriched Uranium), Pu, Cs-137, Am-241, Ir-192, Cr-60 and Sr-90 have been traded illegally [1]. Therefore, tracking these activities is urgent.

A literature search [3] has revealed that neither the database research community nor the information retrieval research community are aware of the unique search challenges posed by the nuclear forensics identification problem. Each of these communities has significant resources which could be brought to bear in solving the fundamental problems of large-scale nuclear forensic discovery. Among other goals, a primary focus of this digital library is to interest these communities in the issues of nuclear forensic identification and to supply education. This is the first project of its kind to apply modern search technology to the nuclear forensics matching and identification problems. The methodologies explored (directed graph similarity, automatic classification, and rule-based

matching) are expected to break new ground in providing rigorously evaluated approaches to nuclear signature matching. Just as traditional forensics (fingerprint matching, DNA matching) benefited from algorithms developed by information search specialists in their area, we expect the area of nuclear forensics to be significantly improved by focused attention by researchers in the search area. In particular, the projects results should help support the work of the Nuclear Smuggling International Technical Working Group.

In the remaining sections of the paper we provide descriptions of the data collections involved. Finally we explain the data mining techniques used and our implementation. In the conclusion we discuss future plans for the system.

2 Data Collections

Since this digital library is about search and search evaluation, it is crucial to have some data upon which to evaluate the effectiveness of our methodologies. We expect to include, among others to be assembled, the following data sets:

1. Radioactive decay chain patterns. The unclassified Nuclear Structure and Decay Data (NuDat) data set [7] will be used to build graphs of radioactive decay. The data set contains over 165 thousand records of isotopes of elements, decay daughters, radiation type, half life and energy levels.
2. Location identification collections
 - (a) Uranium Oxide Quality Control data set. This proprietary dataset from Lawrence Livermore National Laboratory contains principal components analysis to infer location from indicator trace elements which are geospecific [6]. The data set has 3,436 sample measurements for 21 sites. The data set has licensing restrictions but we are attempting to obtain access to the data.
 - (b) Pit Manufacturing Measurements. Isotopic measurement records of pit manufacturing materials from the Plutonium Metal Standards Exchange Program at Los Alamos National Laboratory. The data consists of analytical measurements within and between worldwide laboratories.
 - (c) Actinide measurement test results. This unclassified dataset is a report of all measurements made on actinide samples and contains the assay actinide and its isotopic and trace element amounts. It is produced by researchers from six national laboratories and the Atomic Weapons Establishment.
3. The Nuclear Fuel Cycle and Weapon Development Cycle [8]. The Pacific Northwest National Laboratory created it on behalf of the United States Department of Energy. Nuclear materials are naturally-occurring and can also be manufactured in a lab. Once obtained, either sort goes through any of various multi-stage processes to either become fuel or nuclear weapons material. This figure depicts the life cycle stages of both outcomes: fuel or weapon development. At each stage, the figure also names the processing facility types that convert materials from one stage to the next in the life cycle.

4. Facilities. The International Atomic Energy Agency (IAEA) lists of world-wide reactors.

3 Data Mining Technologies

Our intent is to create a library of empirical radio chemical data with forensics capabilities in order to preserve public safety while accommodating the important utility of radioactive materials for medicine, energy and security. The nuclear materials addressed by our library collection are economically valuable commodities to criminals. As such, great resources are employed to smuggle nuclear materials to illicit customers who may have destructive or catastrophic objectives. To interfere with such aims, we utilize a digital library with advanced search capabilities.

3.1 Forensics Capacity

In order to furnish the forensics community with suitable protective force, the system's design not only incorporates advanced digital library and search capabilities, but also incorporates the practices and protocols of forensics. A forensics operation is distinguished by requirements for provability, traceability and credibility. The results of a query must stand up in court since the goal of a forensics operation is to make a claim during judicial proceedings. Therefore our design focuses on technologies that would survive the scrutiny of trial court requirements and other legal proceedings. Also, our data sources must constitute a reliable, authoritative base. We use collections from certified national laboratories. The following technologies were selected by virtue of their traceability to authoritative sources and for their quantifiable characteristics that will be credible in court.

3.2 Technologies

The data mining techniques employed are association relationships, a graph theory distance algorithm and rule-based classification. Each one was selected because it satisfies forensic practices and protocol requirement of reproducible results to which a subject matter expert witnesses can testify.

1. Relationship associations are utilized to make a conceptual link between the radiochemical properties of a nuclear material sample and where a sample lies in the Nuclear Fuel and Weapon Life Cycle corresponding to a threat level, as well as geographic origin, as discussed in the Data Collections section. These associations are drawn by using the transitive property of relationships using set operations and therefore can be demonstrated to have an irrefutable relationship.

2. The graph theory node distance algorithm identifies the threat level classification for nuclear materials. Node distance metrics determine the distance of a nuclear material's progress along the Nuclear Fuel and Weapon Life Cycle [8]. The lower the distance from current stage to end stage, the higher the threat level.
3. Rule-based data mining for classification of nuclear material signatures. This method is utilized to predict the radio chemical signatures of nuclear materials given temporal changes and the laws of radioactive decay.

4 Implementation

4.1 Attribution

Implementation of the attribution search involves association relationships that are found by joining data sets on common instances with known class labels and with the set operation of transitivity. Attribution will be further explained below.

Given chemical measurements of nuclear material samples, our digital library can return probabilistic geographic origin points and threat levels. Threat level refers to a material's position in the nuclear weapons life cycle, which begins at naturally-occurring Uranium and Thorium and progresses through various processes of enrichment - any of which has multiple stages - eventually becoming weapons grade material.

Chemical measurements of samples reveal trace element signatures which correspond to particular geographic sites or regions. Specific sites hold facilities that are capable of certain material processing steps (mining, enriching, etc.). These processes are associated with specific stages of the weapons life cycle. The closer a material is to the weapons grade end stage of the weapons development life cycle, the higher that material's threat level is. By multiple association relationships and the transitive property then, a user can input a smuggled material's chemical composition as a query. The query result will be (attribution of) the material's probable geographic origin and threat level.

4.2 Threat Level

Implementation of the search for a threat level of a nuclear material involves k-nearest neighbor distance computations [4] using the Nuclear Fuel Cycle and Weapons Development Process chart[8]. Considering each stage in this cycle as a node, node distance metrics can assign a numeric threat level to a given sample. The distance (in number of nodes) from a material's current state to the end state of weapons-grade material, determines its threat level; the smaller the distance, the more dangerous a material is. The distance attribute of this technique provides a compelling way to express viability of a material as a fuel or weapon and its threat level.

As mentioned in the previous section, the result of the relation association search is a point within the Nuclear Fuel Cycle and Weapons Development Process, e.g., chart position 5.10 “Spent Fuel or Special Irradiated U Target Containing Pu.” This point or position identifies the type of facility where a material with particular radio chemical properties would be found due to the manufacturing processes and development state of the material.

The metric aspect is suitable for forensics practices since it makes the classification of a material’s threat level measurable and illustratable in a courtroom setting. The threat level will also serve as a metric for the material’s value as a commodity (to smugglers).

4.3 Rule-based Matching

Any given nuclear material has a nuclide signature. This signature is determined by the properties of radioactive decay and fission. The training data set for our signature classifier is a set of measured quantities of isotopes and trace elements for particular nuclear material samples from well-established laboratories. This training set has created class labels for signatures, e.g., Highly Enriched Uranium (HEU) - more than 20 % of Pu-240 - and Low Enriched Uranium (LEU) - less than 4 % of Pu-240.

Laboratory data is usually incomplete and contains anomalies such as disagreeing production date isotopic ratios. For those reasons, we will generate our own set of simulated measurements. In the paper ‘Characterization of nuclear fuel using multivariate statistical analysis’ [5], the authors state, “It is impractical to obtain such data from real nuclear fuel sample analyses. Instead, the ORIGEN-ARP code package (Bowman, 2000) was used to predict U and Pu isotopic compositions at various burn up values.” In the AAAS/APS report on nuclear forensics [2], simulated data was also used to characterize different types of nuclear reactors. This simulated data set was generated by software called MCODE. MCODE is similar in function to ORIGEN-ARP but uses more computing resources because it implements a Monte Carlo method. For this project we expect to generate simulated data using one of the code packages mentioned above. We will build simulated decay chain graphs using the radioactive decay rules inferred from the NuDat database [7]. Our implementation will supplement the incomplete and anomalous laboratory measurements with our own simulated data.

One note in composing classifications for forensic use of raw data is that the modeling process must be sensitive to unexpected values. Unexpected values are important as unique attributes which could be evidence to prove a case for particular events rather than for generalized scientific knowledge. Investigative practices that search for malicious intent are focused on finding singular activities that can be traced to particular perpetrators rather than finding general observations about a phenomenon.

5 Conclusion

In this paper we have described the preliminary work being carried out on this project, and discussed some of design and search issues involved in nuclear forensics search. Our work in data collection, data mining and forensic matching algorithm creation is ongoing. We are in the process of collecting and indexing the various databases discussed in the Data Collections section and are investigating algorithms for graph-based search with time components.

6 Acknowledgements

The work described in the paper is part of the grant “ARI-MA: Recasting Nuclear Forensics Discovery as a Digital Library Search Problem” which was funded by the U.S. National Science Foundation and Department of Homeland Security (award #ECCS-1140073).

References

1. International Atomic Energy Agency. Illicit Trafficking Database (ITDB) Factsheet, 2009. <http://www.pnl.gov/publications/abstracts.asp?report=228018>.
2. American Physical Society (APS) / American Association for the Advancement of Science (AAAI) Joint Working Group: Nuclear Forensics: Role, State of the Art, Program Needs, 2008.
3. F. Gey, R. Larson, and E. Sutton. Ari-ma recasting nuclear forensics discovery as a digital library search problem, nsf/domestic nuclear detection office academic research initiative grant proposal from uc berkeley. 2012.
4. Y. Lifshits. Branch and bound algorithms for nearest neighbor search: Lecture 1. Steklov Institute of Mathematics at St. Petersburg, California Institute of Technology, RuSSIR, Russian Summer in Information Retrieval, 2007.
5. M. Robel and M. J. Kristo. Characterization of nuclear fuel using multivariate statistical analysis. Lawrence Livermore National Laboratory report, 2007.
6. M. Robel, M. J. Kristo, and M. Heller. Nuclear forensic inferences using iterative multidimensional statistics. Institute of Nuclear Materials Management 50th Annual Meeting Tucson, AZ., July 2007.
7. A. Sonzogni. Nudat (nuclear structure and decay data) database. 2010.
8. C. E. Willingham. Nuclear fuel cycle and weapons development process. Prepared for the United States of America Department of Energy by the Pacific Northwest National Laboratory. <http://www.pnl.gov/publications/abstracts.asp?report=228018>, 2004.