Nuclear Forensics: A Scientific Search Problem

Fredric Gey†, Chloe Reynolds‡, Ray Larson‡, Electra Sutton†
†Institute for the Study of Social Issues and the ‡Information School
University of California, Berkeley
MS 5670 2420 Bowditch Street, Berkeley, California, USA
{gey, electra}@berkeley.edu, {chloe_reynolds, ray} @ischool.berkeley.edu
Keywords: Nuclear Forensics, Scientific Search

Abstract

In this paper, we introduce a unique sub-field of scientific search: nuclear forensics. Nuclear forensics plays an important technical role in international security. We describe a conceptual model of nuclear forensics matching as a particular form of directed graph matching. The unique characteristic of this match is that the attributes of the graph nodes (measurement of mass of nuclear isotopes in a decay chain) vary over time, so matching must include a time-varying computation at the heart of the match. Using a database of spent nuclear fuel samples we formulate a search experiment to try to identify the particular nuclear reactor from which a particular sample came. Preliminary results (Precision 0.34 at rank 10) are promising given the simplifying assumptions made.

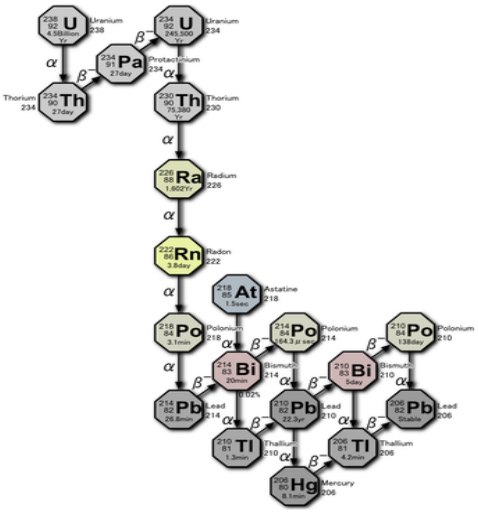
1 Introduction

One of the most frightening scenarios of individual terrorism has been the deployment of a conventional explosive device laden with nuclear material, a so-called "dirty bomb." This device (technically referred to as a radiological dispersal device or RDD [Scott 2005]) would release major amounts of radiation poisoning onto an urban population, causing untold human suffering. According to [Mayer, Wallenius and Fanghänel 2007] "Since the beginning of the 1990s, when the first seizures of nuclear material were reported, the IAEA recorded more than 800 cases of illicit trafficking of nuclear or other radioactive materials." Security agencies worldwide continue to work to prevent this scenario from happening. The two aspects of prevention are detection and forensics. Millions of dollars are being spent on improvement of devices to detect contraband radioactive material which might be hidden in shipping containers. The flip side of detection is forensics – if a significant amount of smuggled nuclear material is seized, can it's origin be traced to both track down the would-be terrorists and to prevent further smuggling activities [IAEA 2002, APS/AAAS 2008, GAO 2009]. To do this, a seized sample can be analyzed to ascertain its "nuclear signature" which can be compared to an archived digital library of nuclear signatures which have been abstracted by radio-chemical analysis of a large number (tens of thousands) of nuclear samples from uranium mines or nuclear reactors worldwide.

2 Nuclear Forensics as a Search Problem

Given a nuclear sample obtained from whatever process (interdiction, for example), the problem is to identify its source. Such identification requires clues to match against a dataset of samples for which sources and compositions have been identified. The process, abstractly, is not that different from matching fingerprints or DNA samples from a crime scene – both require a library against which the match will be made, and both require specialized matching technologies which execute the search. In the case of nuclear forensics, the library will consist of radioactive samples. Both plutonium and enriched uranium samples are characterized by their nuclear decay isotope production. Figure 1 displays the decay chain for Uranium 238U. The figure shows the parent and child isotope in each decay reaction, as well as the type of decay (alpha, beta).[1]

Figure 1: Uranium Decay Chain



If we examine this figure abstractly from a mathematical concept point-of-view, it resembles a

1 Source: http://en.wikipedia.org/wiki/Decay_chain

directed graph, in particular, a directed acyclic, weighted graph. The nodes of the graph are the particular isotopes of decay. The edges of the graph display the particular decay directions. The graph is 'directed' because decay proceeds from parent isotope to child (product) isotope. It is 'acyclic' because there are no cycles in the decay chain. In other words, no child isotope will subsequently decay to a parent isotope, producing infinite cycles. Finally the edges are 'weighted' in that the weights applied to the edges between parent and child consist of the decay rates (half-life). Thus, given the digital signature of an interdicted sample of nuclear material seized by authorities, we wish to identify its origin, and the conceptual search problem becomes one of a graph matching, in particular, of matching between weighted directed acyclic graphs. For background on graph matching, see [Bengoetxea 2003].

Represented as a Graph G = (V,E), a nuclear sample consists of a finite number of vertices (sometimes referred to as nodes) v_1 ... v_n representing elements in a decay chain, i.e. For Uranium above, n=19, v_1 = 238U v_2 = 234Th and v_{19} = 206Pb the terminal stable element of lead. Associated with each vertex at time t_m, is an amount $m(t_m)$, the measured mass of the element at the time of measurement. The edges (or arcs) between elements represent the decay direction: thus $e_{7,8}$ = (226Ra, 222Rn), representing the decay path from Radium to Radon. However when it comes to real samples, we will measure different amounts of each element at different times, and we have to use the basic equations of nuclear decay to compute from a measurement at one time to compare to a sample which was measured at a different time.

2.1 The Graph Similarity Function

We can simplify our representation by saying that a seized material sample at time t_m, is referred to as $G_s(t_m,)$. Let us further say that there exists a digital library of k samples each measured at different times LIB={$G_1(t_1)$ .... $G_k(t_k)$}. We wish to match the seized sample to appropriate library samples. But there are differences in times of measurement – to do the match we have to forwardly compute each of the library samples from $t_{k,}$ to time t_m (or backwardly compute the seized sample from time t_m to time $t_{k,}$). Thus we seek a similarity function:

$$SIM\ (G_s(t_m,),G_i(t_i)\ \varepsilon\ LIB) = SIM(G_s(t_i)=f_b(G_s(t_m,),G_i(t_i)\ \varepsilon\ LIB)$$

for the ith sample in the library and where f_b is the backward computation function. The backward computation will, for a multiple number of isotopes in the decay chain, be derivable from the general form of Bateman's equation [Ehman and Vance 1991].

Of course this is the simplest model – in reality, all samples will have additional geolocation clues L (manufacturing impurities or geologically specific elements) which may or may not have a time dependency. Thus G = (V,E,L) for a more complex model. It will be the first priority of the authors to develop a time-varying graph model which accurately captures the complexity of the nuclear forensic discovery (matching) problem.

3 DATA & AN EXPERIMENT

3.1 Spent Nuclear Fuel Rod Measurements

SFCOMPO is a database of spent nuclear fuel (fuel rods from a nuclear reactor after the energy has been extracted by the nuclear fission process) measurements. The data has been carefully vetted and deemed reliable by nuclear engineering experts and has been released to the public via the Organization for Economic Cooperation Nuclear Energy Agency (OECD-NEA) web site. [2] The process by which the samples are measured (the geometry of where the sample has been drilled and extracted from the fuel rod) is described on the web site. The data consists of 274 samples from 14 nuclear reactors (some no longer in operation) in four countries (Germany, Italy, Japan and the USA). There are a variable number of samples from each reactor, ranging from two for the Genkai-1 reactor in Japan to 39 for the Trino Vercellese reactor in Italy. Each sample has a variable number of isotope, isotope ratio and burn-up measurements, ranging from one measurement for Europium 155 (^{155}Eu) to 261 measurements (a measurement is found in almost all samples) for two Uranium ratios (^{235}U/^{238}U and ^{236}U /^{238}U). The total number of measurements is 10,339. Figure 2 is a frequency distribution of the count of samples for the top ten measurements:

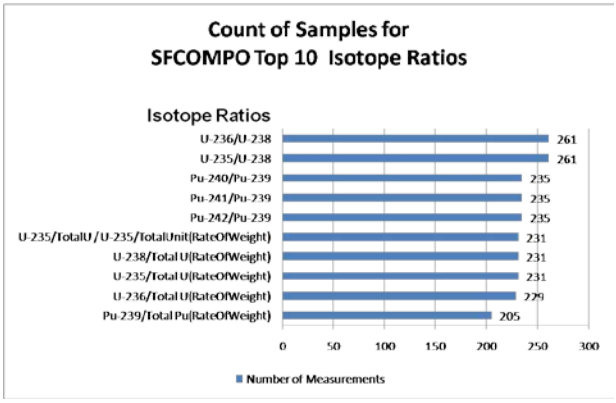


Figure 2: SFCOMPO top measurements

3.2 A Naive Search Experiment

In this section we develop a bare-bones search experiment using the SFCOMPO database. By bare-bones, we mean we ignore all temporal effects on measurements and measurement rations. We understand that this crude search method would be unrealistic for use in an applied setting, but we invoke it to get at core issues in the utility of the database for search. The goal of our experiment is to determine whether a sample and its constituent measurements can be used to identify which reactor the sample came from. The structure of the search experiment is thus:

1) A single sample is removed from the set of samples in the database. This sample becomes the "query sample" and all other samples are the "known samples", or "document samples" to invoke search terminology.
2) A similarity matching algorithm is developed which matches the measurements in the query sample with each of the measurements in each document sample. This match results in a number between zero and 1

[2] http://www.oecd-nea.org/sfcompo/

called a Retrieval Status Value, or RSV. (Ideally the RSV is an estimate of a matching probability).

3) Document samples are ranked by this similarity value.
4) Relevance of the document sample to the query sample is assessed as follows: If a document sample comes from the same reactor as the query sample, then the document sample is judged relevant. Samples not coming from the query reactor are judged irrelevant.
5) The usual information retrieval evaluation measure of precision can be calculated.

3.3 The Similarity Matching Algorithm

In order to explain the algorithm, we first briefly describe the dataset. To reiterate, the dataset contains 274 nuclear material samples taken from 14 nuclear reactors. Each of the 274 samples has up to 113 measurements associated with it. That is to say, the dataset has 274 rows and 113 columns. The "measurements" are a) the quantities present in the sample of isotopes and isotopic ratios, and b) burnup values. Because conducting such measurements is expensive, many samples' measurements are missing values (the measurements were not conducted). The population density for rows is low (mean = 38 out of 113 measurements present per sample, 34%). The population density for columns is slightly higher (mean = 91 out of 247 measurements per sample, 37%).

We extract one sample (at a time) from the database which becomes the **query sample**. The remaining 273 samples left in the database at that time are the **known samples**.

A sample **pair** is a pair of two samples, one of which is the query sample, the second of which is a known sample.

A **column_in_common** is a column for which both samples in a pair have values, i.e. neither column is missing values.

Our naïve algorithm compares the differences between columns of the query sample (input) and the columns of each known sample, ultimately creating a list of the top 10 most similar "known sample" results (output). This process works much like an internet search engine (i.e. Google). However, instead of a search term, a user enters (up to 113) isotopic measurements of an interdicted nuclear material. This set of isotopic measurements of this interdicted sample is the "query sample." In lieu of getting ranked website/documents as results, the user will receive a list of relevant "known samples." The results will also display the reactor from which each result/known sample originated. In this way, we aim to detect the probable reactors of origin for interdicted nuclear samples.

First, we looked at the **range** (maximum value over the entire 274 samples minus the minimum value) for each column. Then we calculated weighted column distances for each column of each pair.

Formula. weighted_column_distance (for a column x, for a pair) = absolute value (column_x_value of the query sample – column_x_value of the known sample) / (range of column x).

Next, we computed the retrieval status value for each of the 273 pairs.

Formula. retrieval_status_value , or **RSV**, (for a known sample) = 1 – (sqrt(sum of squares of weighted_column_distances of columns_in_common for a pair) / (number of weighted_column_distances for a pair))

We also evaluated a depth score for each pair. The **pair_depth** is equal to the number of columns_in_common for a pair. Pairs which have no columns in common, have a depth score of 0. Pair_depth scores ranged from 0 to 83 in this dataset. When a pair has depth score of 0, an RSV cannot be calculated because there are no points of comparison between the query sample and the known sample.

We call this type of pair a **ute pair**, short for "unable to evaluate." Ute pairs will negatively affect not only recall but also forensic investigations.

The **minimum_depth** (for a query) is a chosen minimum number of columns_in_common that a pair must share in order to be evaluated. The minimum_depth threshold in this experiment was 1. Each known sample must have had at least one column_in_common with the query sample (a depth score of 1) in order for that known sample to have been evaluated against each other. Future experimenters could opt to choose a minimum_depth setting higher than 1, which would likely result in more ute pairs. The topics of ute pairs and minimum depth thresholds are revisited in the *Analysis and Next Steps* section.

The **results** (for a query) are the known samples of all pairs for which the pair_depth is equal to or higher than the minimum_depth threshold chosen for the query.

The **top 10 results** for a query are the results with the highest RSVs.

A **relevant_result** (for a query sample) is a known sample that came from the same reactor as the query sample.

The **precision_rank** (for a query) is the number of results to display in response to a query. A precision_rank of 10 means that 10 results will be displayed for any given query. Using a) the reactor as the condition of relevance, b) a precision rank of 10 and c) a minimum depth threshold of 1, we evaluated the query_precision of each query sample.

Formula. query_precision = (number of relevant results for a query among that query sample's top 10 results) / (minimum (precision_rank, number of relevant results))

Finally, we found the system precision score.

Formula. system_precision = (sum of 274 query_precision scores) / (count of query samples = 274)

3.4 Results

Utilizing all samples in the database results in 274 queries, where each query sample is matched against the other 273 samples. Since a premium is placed upon correct ranking at the top of the list, we report only precision in the top ten samples retrieved. Overall Precision at 10 of our naïve information system was 34%. Since several reactors have fewer than ten samples, we report for three different experiments, 1) all reactor samples used as queries, 2) all samples except for reactor Genkai-1 reactor who 2 samples are excluded from the experiment and 3) exclude reactors (and their samples) with fewer than ten samples, leaving 228 samples to utilize in the search experiment. Limiting the precision to rank 10 limited our maximum possible precision by definition. These 274 samples came from 14 reactors and the number of sample per reactor ranged from 1 to 39 samples per reactor. For example, the Trino Vercellese reactor had 39 samples. Since 39 is greater than 10 (the precision rank), so the maximum possible precision for a query sample taken

from this reactor would be 10 (the precision rank) divided by 38 (the number of remaining known samples from that same reactor), i.e. 10/(39-1) = 26%. Therefore, increasing the precision rank to 39 would improve the system's precision score. Table 1 summarizes these results by reactor.

Table 1: Precision-at-10 by Reactor

Overall Precision			**0.34**	
Reactor/country	**# of Samples**	**Max Possible Prec.**	**Precision (per Reactor)**	**Actual / Max Possible Prec.**
JPDR Japan	30	1.00	**1.00**	100%
Monticello USA	30	1.00	**0.85**	85%
Tsuruga-1 Japan	10	0.90	**0.53**	59%
Trino_Vercellese Italy	39	1.00	**0.24**	24%
Fukushima-Daini-2 JA	18	1.00	**0.21**	21%
Takahama-3 Japan	16	1.00	**0.16**	16%
Fukushima-Daiichi-3	36	1.00	**0.16**	16%
Obrigheim Germany	23	1.00	**0.15**	15%
Genkai-1 Japan	2	0.10	**0.10**	100%
H.B.Robinson-2 USA	7	0.60	**0.09**	14%
Cooper USA	6	0.50	**0.07**	13%
Gundremmingen DE	12	1.00	**0.06**	6%
Mihama-3 Japan	9	0.80	**0.06**	7%
Calvert_Cliffs-1 USA	9	0.80	**0.06**	7%

3.5 Analysis and Next Steps

There are a number of explanations for the unevenness of these results. Primary might be the naïve assumption that measurements are time-independent. However this is not the only explanation. Others are:

1) The number of measurements is not consistent across samples, nor is it always consistent between samples from the same reactor. Unlike the case of document retrieval, the absence of a particular isotope or isotope ratio only means that that particular value was not measured and it should have been – in document retrieval the absence of a matching term between a query and document is truly significant to the retrieval process.
2) Measurement of radioactive isotopes may not be a sufficient indicator of reactor location. Non-isotopic characteristics may be more important in nuclear forensics, in which case, in their absence, less can be deduced about which reactor the sample came from. Mayer et al [Mayer, Wallenius and Fangänel 2007] place heavy emphasis on visual and manufacturing forensic clues as well as isotopic measurements.
3) We may be studying the wrong problem. According to Table F.1 of the AAPS report [APS, AAAI 2008], ratios of various plutonium isotopes (238,240,241,242) with respect to Pu-239, definitively identify the reactor type (rather than the reactor origin). Since the current SFCOMPO database only covers two reactor types, Boiling Water Reactor (BWR) and Pressurized Water Reactor (PWR), and both are light water reactors, the search reduces to a binary classification problem.

Data Next Steps

However, the first next step to improve this research is to adjust the SFCOMPO test dataset by using nuclear decay reaction software to compute data to a uniform time for all measurements and also fill in all the unmeasured items by imputation and simulation. Efforts to do this are underway at the Pacific Northwest National Laboratory in Washington USA by a collaborating research group.

Algorithmic Next Steps

Numerous steps could be taken to improve the algorithm. One is weighting. In this experiment, we averaged the distance score for the pair across all 113 columns evenly, giving each column the same weight. Future experiments could weight the distance of certain columns more heavily than others, based on real-world rules and/or learnings from clustering.

Another improvement would be determining and setting an ideal minimum depth. In this experiment, the minimum depth was set to 1, the lowest feasible setting, meaning that pairs only needed to have one point of comparison in order to be evaluated against each other. Like DNA analysis, fingerprint analysis and other forensics analyses, our accuracy would be improved by increasing the required minimum number of acceptable points of comparison (columns_in_common). In our case this would mean increasing the minimum depth threshold. Of course, demanding more comparison points will increase the number of ute pairs, and therefore decrease recall. Further investigation is needed to determine the best balance for applied use cases.

A different algorithmic enhancement would be to set the rank level dynamically for each query. (This would adjust the denominator in the query_precision formula, thereby affecting the system precision as well.) We would determine a minimum distance threshold such that matches with too low of an RSV score would not retrieved. Similarly, the minimum depth threshold could be allowed to reset dynamically either (up or down, or up only) per query; this is an alternative to the previous suggestion of determining a single across-the-board minimum depth threshold.

Deciding how to handle and report other metrics (relevancy, pair_depth, and a UTE score) would also become important before this system is implemented in an applied setting. We may opt to list the known samples of each ute pair as a supplement to the top results, which could be beneficial in the course of a forensic investigation.

4 Summary

The purpose of this paper is to present a new application in the area of scientific search, *nuclear forensics*, which has not only international security importance, but also presents interesting challenges to develop new search methodologies. The approach described, temporal directed graph matching, is not the only approach. Another interesting approach is to cast the nuclear forensics matching problem as an automatic classification problem [Robel, Kristo and Heller 2009]. The appeal to the Information

Retrieval research community is to see if additional models can be founded to apply to nuclear forensics discovery.

Acknowledgments

This research was sponsored by the USA National Science Foundation under Grant # 1140073, entitled "*ARI-MA: Recasting Nuclear Forensics Discovery as a Digital Library Search Problem*." We thank Bethany Goldblum (UC Berkeley Nuclear Engineering) and M.C. (Mikey) Brady (Pacific Northwest National Laboratory, Richland Washington USA) for helpful discussions.

References

[APS, AAAI 2008] 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.

[Bengoetxea 2003] Endika Bengoetxea, **Inexact Graph Matching Using Estimation of Distribution Algorithms, PhD Thesis,** available at

http://www.sc.ehu.es/acwbecae/ikerkuntza/these/Ch2.pdf) Ecole Nationale Supérieure des Télécommunications (Paris), 2003.

[Ehmann and Vance 1991] W. D. Ehmann and D. E. Vance, **Radiochemistry and Nuclear Methods of Analysis**, John Wiley and. Sons, New York, 1991

[GAO 2009] General Accountability Office (GAO) report, **Nuclear Forensics: Comprehensive Interagency Plan Needed to Address Human Capital Issues,** *available* at http://www.gao.gov/new.items/d09527r.pdf

[IAEA 2002] International Atomic Energy Agency (IAEA) Staff Report, "Tracing the Source: Nuclear Forensics and Illicit Nuclear Trafficking," October 2002.

[Mayer, Wallenius and Fangänel 2007] K. Mayer, M. Wallenius, T. Fanghänel, "Nuclear forensic science — From cradle to maturity," Journal of Alloys and Compounds 444–445 (2007) 50–56.

[Robel, Kristo and Heller 2009] M. Robel, M. J. Kristo, and M. A. Heller, "Nuclear Forensic Inferences Using Iterative Multidimensional Statistics," Institute of Nuclear Materials Management 50th Annual Meeting Tucson, AZ. July, 2009, 12pp.

[Scott 2005] Mark Robert Scott, **Nuclear forensics: attributing the source of spent fuel used in an RDD event,** Masters Thesis, Department of Nuclear Engineering, Texas A & M University, 2005, available at: http://txspace.di.tamu.edu/bitstream/handle/1969.1/2368/etd-tamu-2005A-NUEN-Scott.pdf?sequence=1