

REFRAMING NUCLEAR FORENSICS DISCOVERY AS A DIGITAL LIBRARY SEARCH PROBLEM

Nuclear Engineering Colloquium Aug 27, 2012

Fredric Gey (gey at berkeley dot edu)

Institute for the Study of Societal Issues

University of California, Berkeley

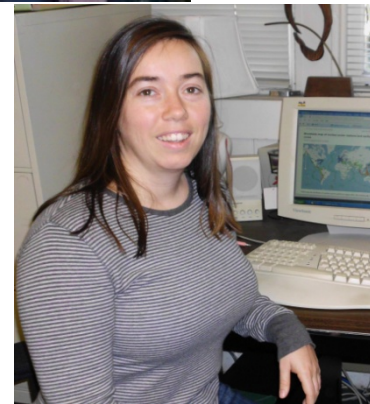
- First year funding source National Science Foundation Grant #1140073: *“ARI-MA Recasting Nuclear Forensics as a Digital Library Search Problem”*
- Thanks to Bethany Goldblum for helpful collaboration



Berkeley Nuclear Forensics Search Team



**left to right: Matthew Proveaux, Ray Larson, Fred Gey, Electra Sutton
and David Weisz. Inset: Chloe Reynolds**



Nuclear Forensics Attribution as a Digital Library Search Problem



- **Reframes the problem** of nuclear forensics discovery (identifying the source of smuggled nuclear material) as a digital library search problem against large libraries of analyzed nuclear materials, i.e.
 - Spent fuel from a nuclear reactor after fission
 - Enriched uranium or plutonium in the nuclear fuel
 - Refined uranium ore (yellow cake) from mines
- **Develops multiple models of the nuclear forensics search process** similar to how traditional forensics (fingerprint and DNA matching) benefited from specialized data representations and efficient search algorithms

Talk Overview



- **Nuclear forensics background**
- **One model of nuclear forensic search**
- **Prior work in nuclear forensics experimentation**
- **Our experiments with the SFCOMPO database**
- **Some conclusions and future work**

Search Scenarios – Nuclear Forensics



- **National Security Challenge:** terrorists wish to attack using a “dirty bomb” – a conventional explosive containing radioactive nuclear material which would cause widespread radiation poisoning, or **worse** **construct an actual bomb** from Special Nuclear Material
- If **smuggled nuclear material** is seized by authorities, how can you determine its origin?
- Nuclear isotopes decay according to well-known processes, creating a '**nuclear signature**' which identifies the time of creation
- **Chemical analysis** of a nuclear sample **can be matched** against digital libraries of existing samples collected from mines or nuclear processing plants worldwide

Nuclear Safeguards



- **National security challenge: As countries abandon their nuclear ambitions, what happens to their existing nuclear facilities?**
 - **Ukraine**
 - **South Africa**
 - **Congo**
- **These may be targets by criminals or terrorists to obtain illicit nuclear materials**
- **IAEA (International Atomic Energy Agency, Vienna) will have a role to play for decades to come**
- <http://www.iaea.org/NewsCenter/News/2008/forensicskit.html>

A Case of Nuclear Murder



- On November 1, 2006, **Alexander Litvinenko**, former Russian Federal Security officer was poisoned by **Polonium-210** isotope while having lunch with associates at a London sushi restaurant. He died of radiation poisoning three weeks later.
- According to doctors, "Litvinenko's murder represents an ominous landmark: the beginning of an era of nuclear terrorism"
- **Polonium-210 (^{210}Po)** is an isotope of Polonium with a significant half-life (138 days). It decays by emitting alpha particles which can be easily shielded by even pieces of paper or the human skin
- An alpha-emitting substance can cause significant damage only if digested or inhaled, acting on living cells like a short-range weapon.

A Case of Nuclear Murder (continued)



- Alpha (α) emitting isotopes like **Polonium-210 (^{210}Po)** can only be detected with special equipment which most hospitals don't have. Litvinenko was tested for α emissions only hours before his death.
- Estimates are that he was exposed to a radiation dosage of about 50 milliCuri (mCi) which corresponds to about 10 micrograms of ^{210}Po . That is 200 times the median lethal dose of around 238 μCi or 50 nanograms in the case of ingestion
- **Polonium-210 (^{210}Po)** is a decay product of Uranium and Plutonium and can only be isolated with special equipment found only in a nuclear establishments
- British authorities investigated the death and it was 'reported' that scientists had traced the source of the polonium to a nuclear power plant in Russia.

Nuclear Forensics Background



- Nuclear forensics has a lengthy history in national security and arms control verification
- In the 1950s through 1970s nuclear forensics included **seismology** (to detect underground nuclear tests) and **aerial sampling** (to detect atmospheric fallout -- in much the same way that UCB NE set up after the Fukushima accident)
- According to Michael May (chair of the 2008 APS/AAAS report): “the principal emphasis today is on the application of nuclear forensics techniques to help attribute either intercepted materials or an actual explosion to its originators.”

Nuclear Forensics Background



- “From January 1993 to December 2011, a total of **2164 incidents** were reported to IAEA’s Illicit Trafficking Data Base (ITDB) by participating States and some non-participating States.”
- “**399 involved unauthorized possession and related criminal activities.**“
- “Information reported to the ITDB demonstrates that:
 - The availability of unsecured nuclear and other radioactive material persists
 - Effective border control measures help to detect illicit trafficking
 - Individuals and groups are prepared to engage in trafficking this material”

Nuclear Forensics Background



- **399 incidents: unauthorized possession/criminal activities, 1993–2011**

[Confirmed incidents involving unauthorized possession and related criminal activities, 1993–2011

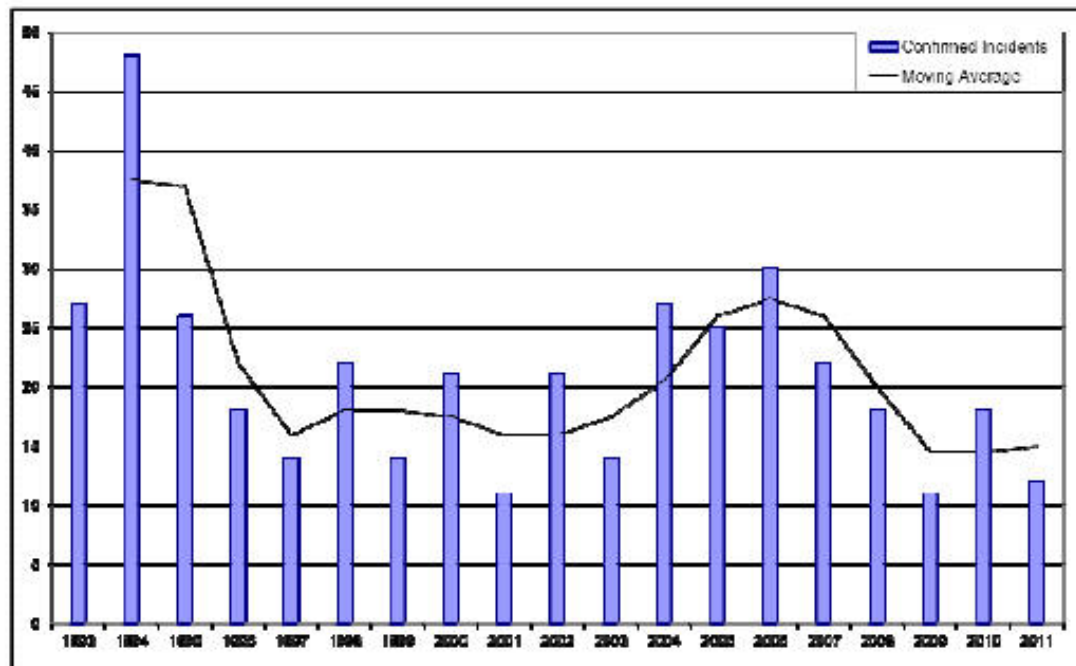


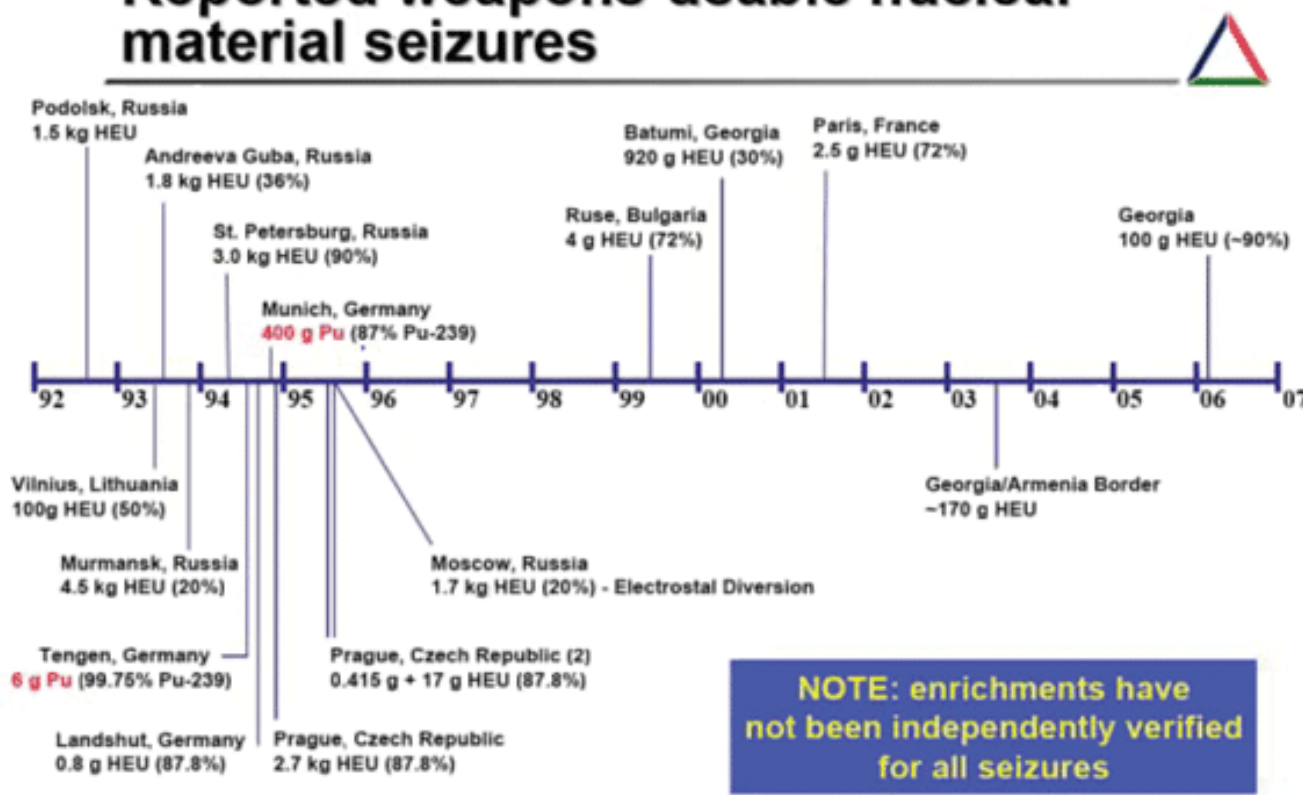
Figure 1. Incidents reported to the ITDB involving unauthorized possession and related criminal activities, 1993–2011.

Nuclear Forensics Background



- 16 reported incidents have been of weapons-usable material

Reported weapons-usable nuclear material seizures



NOTE: enrichments have not been independently verified for all seizures

As of September 21, 2007

Nuclear Forensics



- Dealing with terrorists nuclear intentions has two aspects – **detection** and **forensics**
- Large projects for improving **detection** (i.e. sensing radiation from outside shipping containers) are underway
- Equally large projects (>\$US100M) are underway for **forensics** in the USA and EU
- These projects are creating **digital libraries** of the composition of **existing nuclear material samples** collected from mines or nuclear processing plants worldwide
- The search aspect against these libraries has heretofore proceeded on an ad-hoc case-by-case basis

Nuclear Forensics Search Models

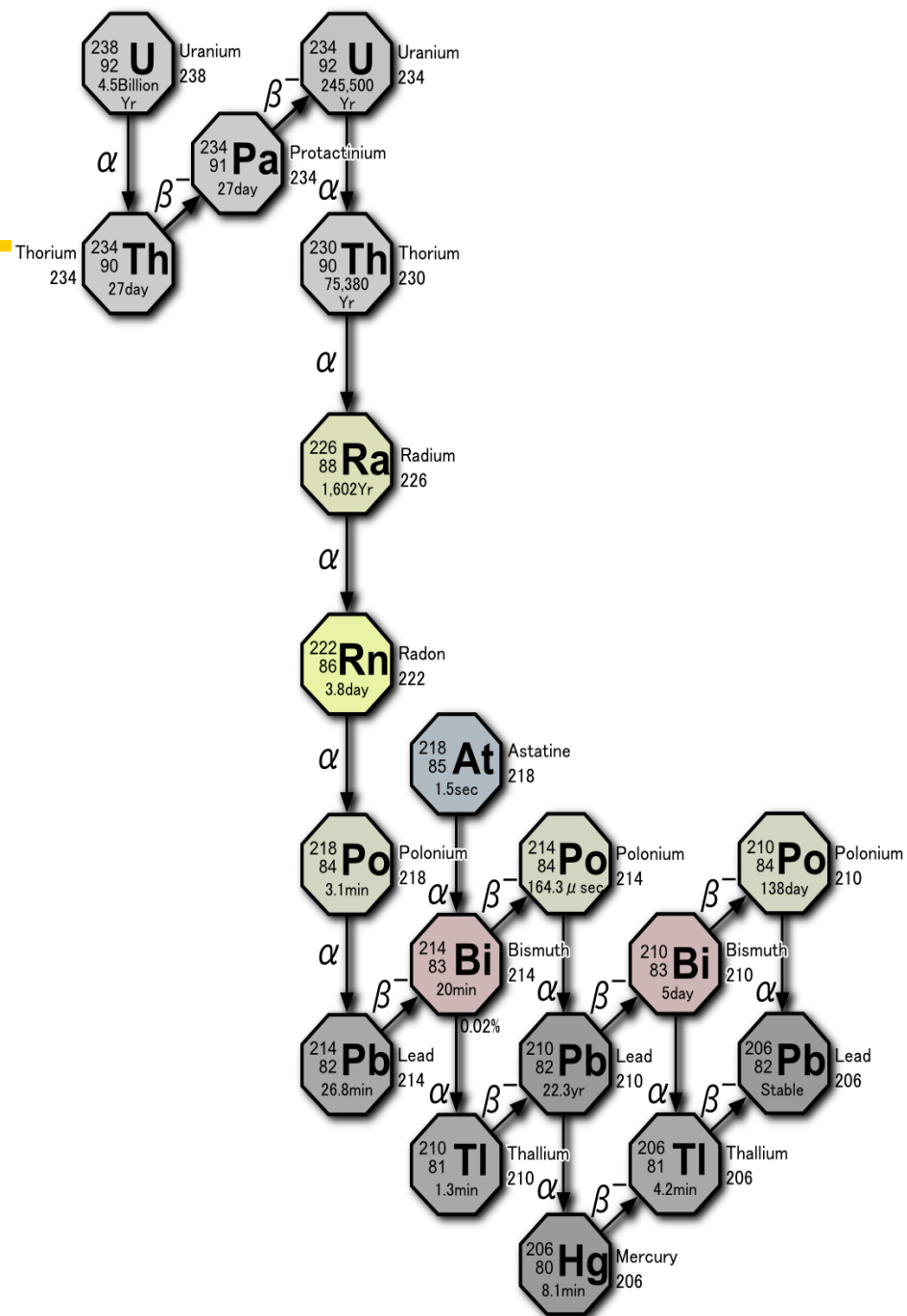


Nuclear forensics search can be framed as a:

- 1. Directed graph matching problem** (in particular a weighted, labeled directed graph matching problem)
- 2. Automatic classification problem** where machine learning is applied to classify a seized sample
- 3. Process logic problem**, whereby the forensic investigation capture the steps and logic which a human nuclear forensics expert would approach

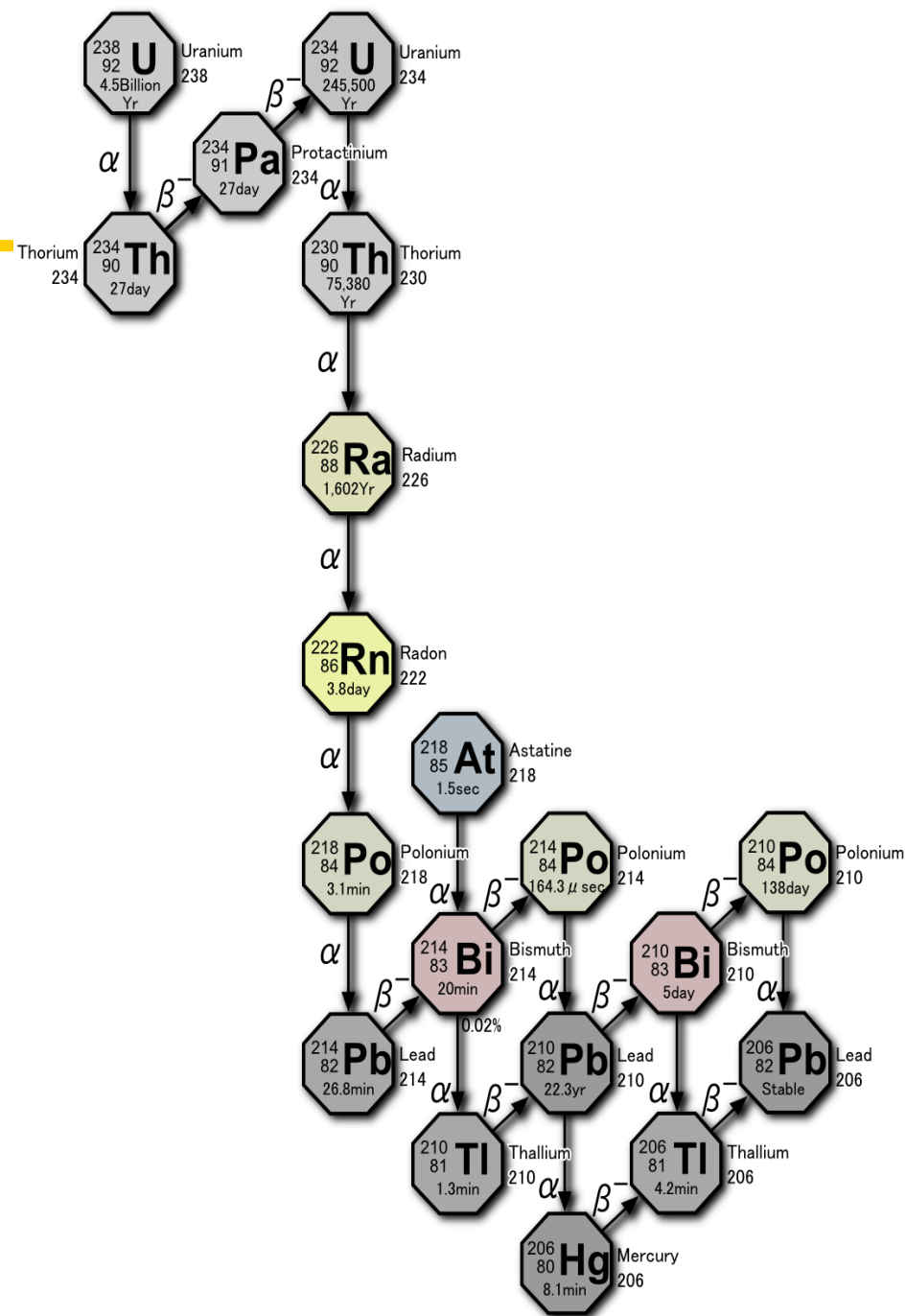
Nuclear Decay Chains

- Nuclear material search and matching depends upon two aspects – **decay chains** and **chemical impurities**
- When a **nuclear isotope decays to produce daughter isotopes** which in turn decay to other **child isotopes** until a **stable non-radioactive element, usually Lead (Pb)** is reached



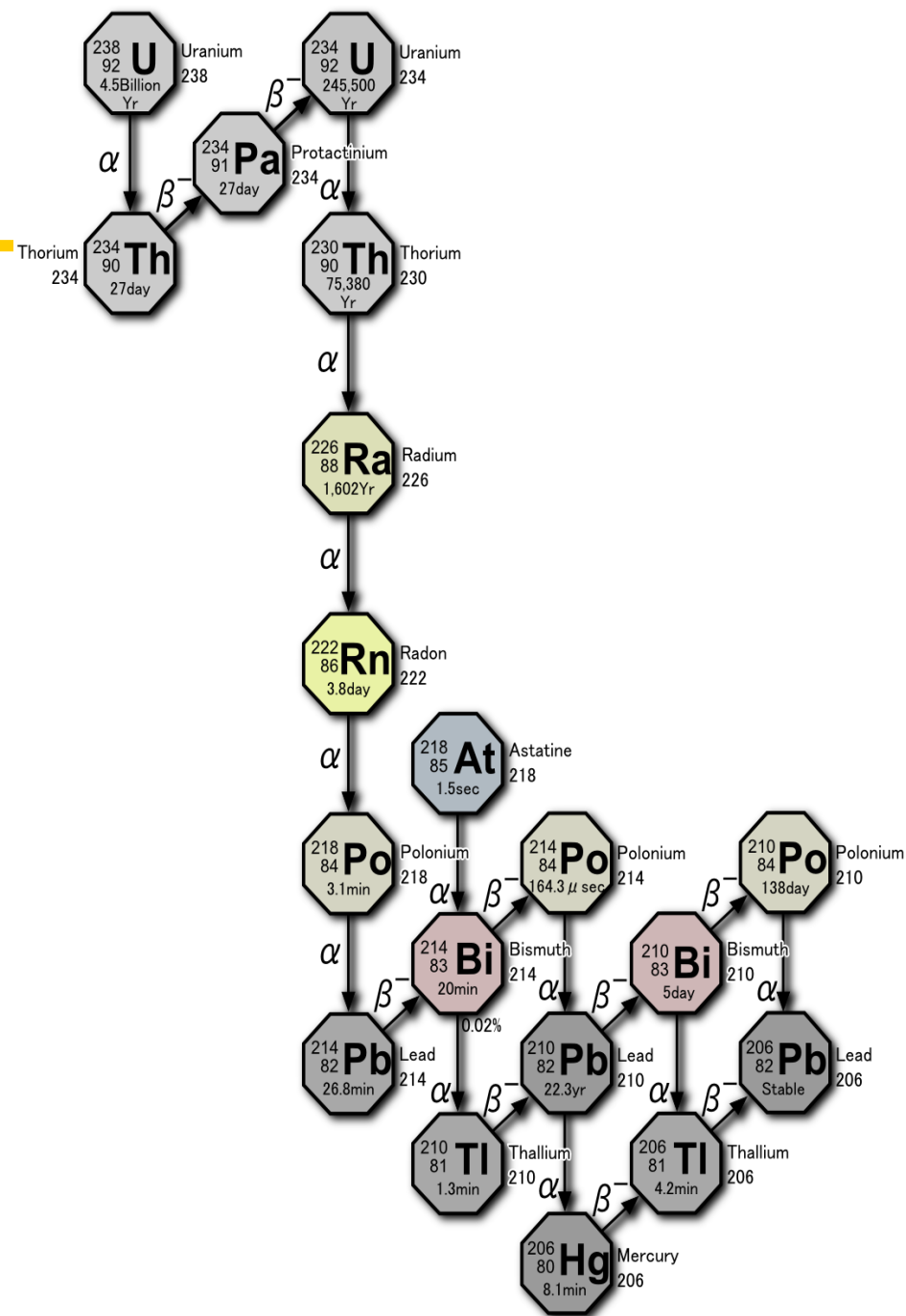
Nuclear Decay Chains (2)

- A nuclear sample will be a snapshot in time of a particular **decay chain**
- Analysis of the sample will yield the concentrations of each element and, using backward inference from the differential equations of decay, establish time since manufacture (if from a nuclear installation).
- Analysis yields a nuclear signature for the sample



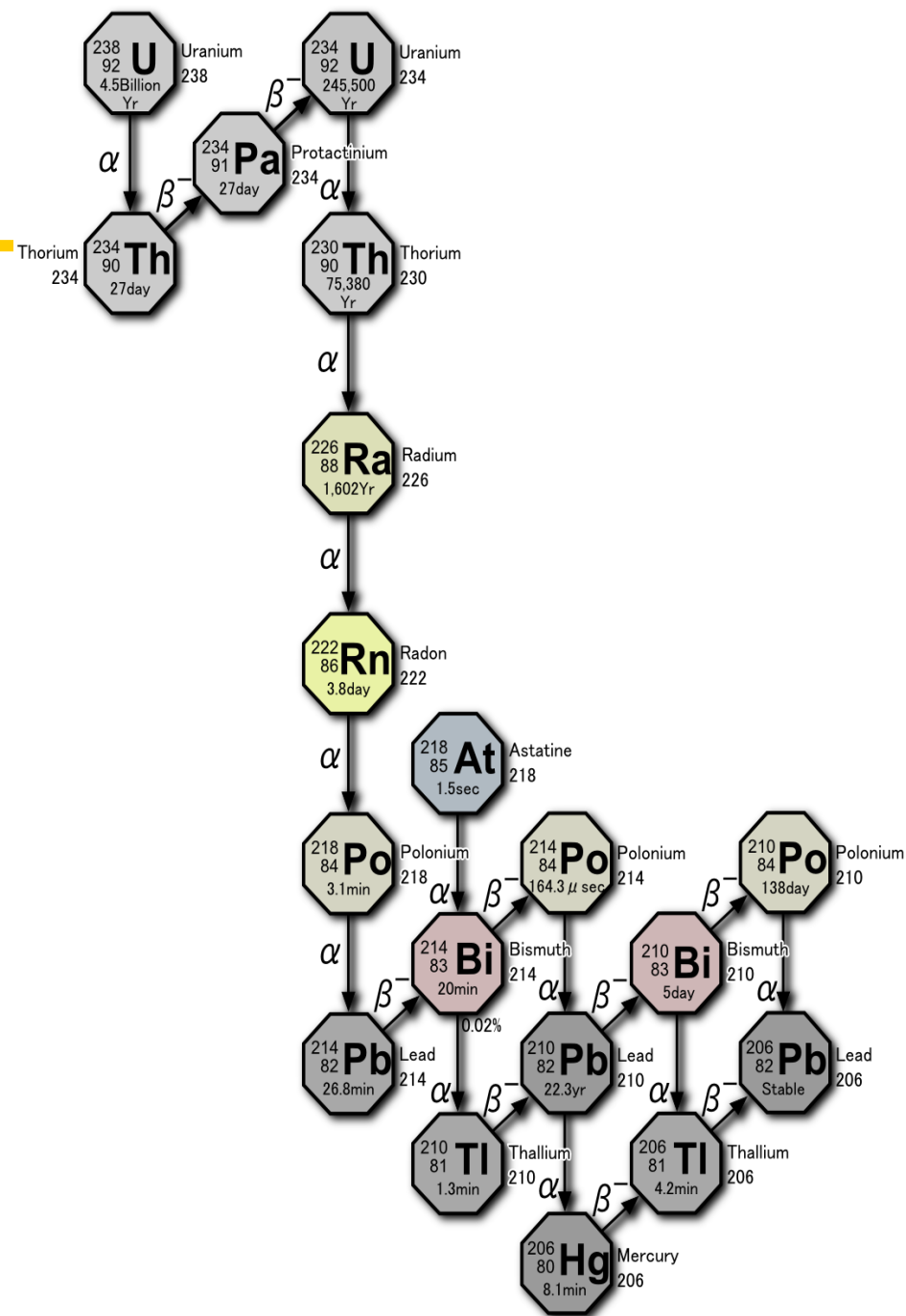
Nuclear Decay Chains (3)

- A nuclear sample will be a snapshot in time of a particular **decay chain**
- Analysis of the sample will yield the concentrations of each element and, using backward inference from the differential equations of decay, **establishes time since manufacture** (if from a nuclear installation).
- Analysis yields a **nuclear signature** for the sample



Nuclear Decay Chains (4)

- Looking at the **decay chain** figure, what model seems to fit the structure is implied?
- **Directed graph**, in particular
- **Labeled directed graph** where the **nodes** are the **decay element isotopes** and the **edges** are the **types and direction of decay** (α or β)
- **Amounts imply a weighted, labeled directed graph**
- **So search can be recast as a graph matching problem**



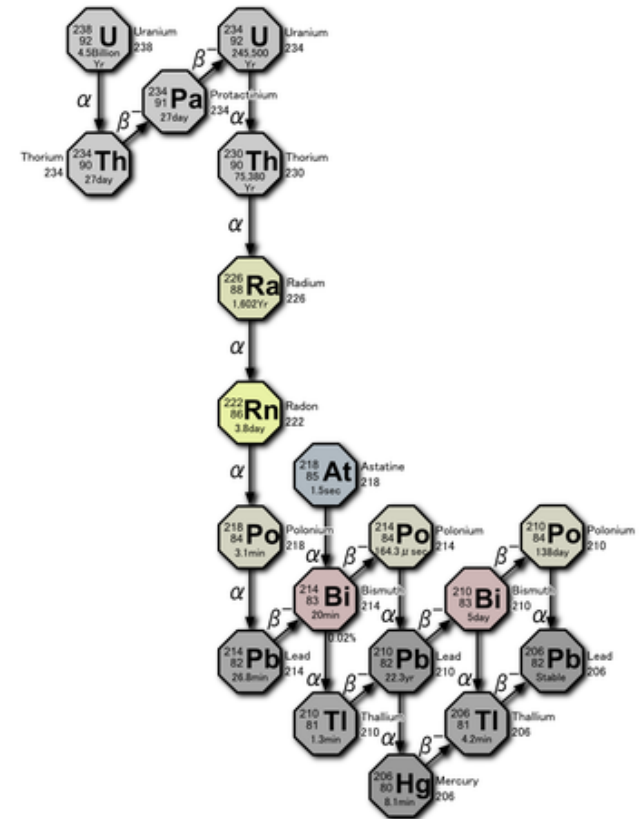
Search Model: Directed Graph Matching



Represented as a Graph $G = (V,E)$, a nuclear sample consists of a finite number of vertices (sometimes referred to as nodes) $v_1 \dots v_n$ representing elements in a decay chain.

For **Uranium 238**, $n=19$, $v_1 = {}^{238}\text{U}$ $v_2 = {}^{234}\text{Th}$ and $v_{19} = {}^{206}\text{Pb}$ 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 $e_{7,8} = ({}^{226}\text{Ra}, {}^{222}\text{Rn})$, represents the decay path from Radium to Radon.



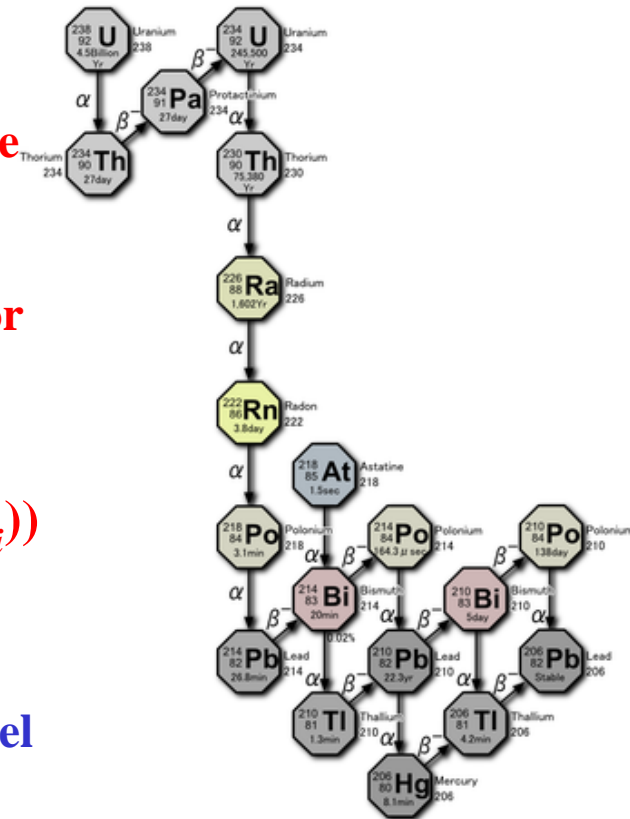
Search Model: Directed Graph Matching



A seized material sample at time t_m , is referred to as $G_s(t_m)$. Let us further say that there exist a digital library of k samples each measured at different times $LIB = \{G_1(t_1) \dots G_k(t_k)\}$. 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:

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

for the i th sample in the library and where f_b is the backward computation function. This is the simplest model – in reality, **all samples may have additional geolocation clues L** (manufacturing, irradiation period, operation history, etc) which may or may not have a time dependency. Thus $G = (V, E, L)$ for a more complex model.



Nuclear Reactor Database (Unifying Multiple Datasets)



We wanted a comprehensive detailed database about **worldwide nuclear reactors** including geographic coordinates

Searches for “nuclear dataset” and similar terms

- 200+ datasets found on web
- 80+ datasets downloaded (arbitrary subset)
 - Sorted into useful (65) / not useful (15) categories
 - Not useful example: Nuclear capacity by country
- Consolidation, done by choosing 5 reputable datasets (e.g. IAEA) and creating a unified database
- Unified dataset into a Google Earth viewer

Nuclear material could come from any of about 500 nuclear power plants worldwide



(Worldwide Nuclear Power Plants using Google Earth)

Original data source: <http://maptd.com/worldwide-map-of-nuclear-power-stations-and-earthquake-zones>
Supplemented with additional nuclear plant data from IAEA

Other Data Sets Assembled or Being Assembled in Support of the Project



The Nuclear Wallet Cards, J.K. Tuli, National Nuclear Data Center, Brookhaven National Laboratory.

Plutonium Metal Standards Exchange Program, Los Alamos National Laboratory (to benchmark code)

Reactor Isotopic composition data from Spent Fuel Isotopic Composition Database (SFCOMPO), OECD Nuclear Energy Agency (NEA)

Atomic Mass Data Center, CSNSM Orsay, France and hosted by National Nuclear Data Center (BNL, USA)

International Atomic Energy Agency (IAEA) nuclear material processing practices and telltale isotopic

Nuclear Fuel Cycle and Weapon Development Cycle, Prepared for DOE by the Pacific Northwest National Laboratory.

Nuclear Forensics: Experimentation



- Most data is closely held in **secret laboratories** (Los Alamos, Livermore, etc). Unsure about the EU-JRC Institute for Trans-Uranium (ITU) Elements at U. Karlsruhe
- How to do experiments in search?
- Utilize the **equations of decay** to create a **simulated synthetic database** through reactor codes like ORIGIN-ARP
- This approach has been taken by others (Nicolaou 2006 and Robel & Kristo 2008 J Environmental Radioactivity), described next

Nuclear Forensics Experimentation

Nicolaou 2006 J Environmental Radioactivity “Determination of the origin of unknown Irradiated nuclear fuel”



- Nicolaou used the ORIGIN code to create a library of data simulating input fuel and burnup for a variety of different input fuels for four different reactor types (PWR, BWR, CANDU and FBR), U and Pu concentrations
- He used 4 different actual known fuel samples as if they were unknowns (PWR reactor type only)
- He reduced the 9 dimensional measurement space ^{234}U , ^{235}U , ^{236}U , ^{238}U , ^{238}Pu , ^{239}Pu , ^{240}Pu , ^{241}Pu and ^{242}Pu , to three dimensions using Principal Component Analysis
- He determined that results were robust ignoring cooling decay

Nuclear Forensics Experimentation

Nicolaou 2006 J Environmental Radioactivity “Determination of the origin of unknown Irradiated nuclear fuel”

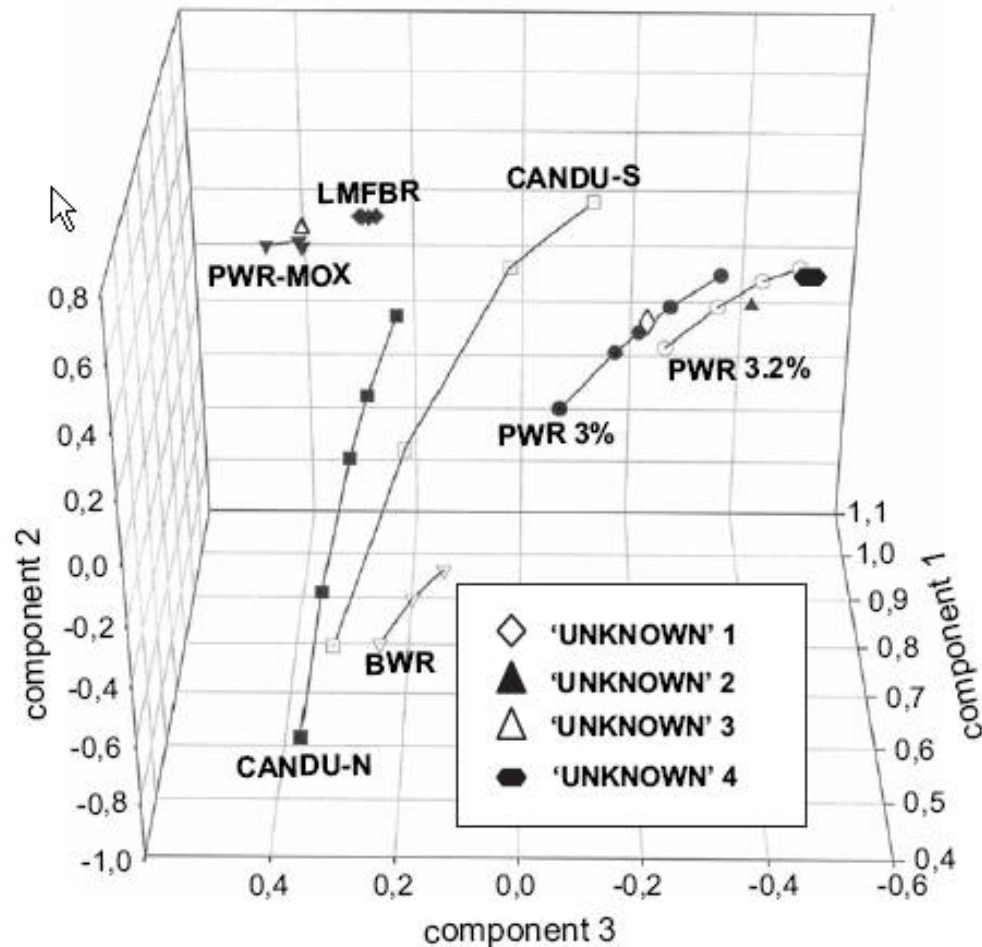


Fig. 1. A 3D plot of the factor analysis showing the similarities between the simulated and 'unknown' fuels.

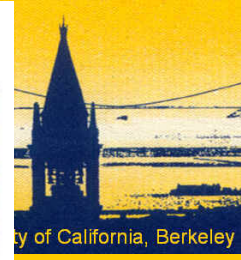


ELSEVIER

Contents lists available at ScienceDirect

Journal of Environmental Radioactivity

journal homepage: www.elsevier.com/locate/jenvrad



University of California, Berkeley

Discrimination of source reactor type by multivariate statistical analysis of uranium and plutonium isotopic concentrations in unknown irradiated nuclear fuel material

Martin Robel*, Michael J. Kristo

Chemical Sciences Division, Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 9455, United States

- **Roble & Kristo extended Nicolaou's results for more reactors using the ORIGIN-ARP code and accounting for different fresh fuel compositions.**
- **They also reduced the measurement space to three dimensions using Principal Component Analysis**
- **They compared results to Partial Least Squares Discriminant Analysis**

Nuclear Forensics Experimentation

Robel, Kristo, Heller 2009 Uranium Ores Samples from Mines



- **Robel, Kristo and Heller applied Principal Component Analysis to the problem of identification of uranium ore concentrate samples from 21 mines in 7 different countries, with a variable number of samples (from 1 to 397) for each mine. They again compared results to Partial Least Squares Discriminant Analysis and KNN and Classification and Regression Tree algorithms.**
- **Their iterative statistical method outperformed the traditional classification methods, at least for country identification. It is possible that the imbalance of the dataset affected the outcomes.**

Nuclear Forensics Experimentation

Robel, Kristo, Heller: Imbalance of the Uranium Ore dataset



- The imbalance of the Uranium Ore dataset**

Table 2. Uranium ore concentrate vendor QC data sources and numbers of samples.

Source ID (Class)	Source Name	Samples	Class	Source Name	Samples
1	USA 1	6	12	Australia 3	233
2	USA 2	3	13	Kazakhstan 1	37
3	USA 3	8	14	Kazakhstan 2	7
4	USA 4	85	15	Kazakhstan 3	4
5	USA 5	21	16	Kazakhstan 4	1
6	Canada 1	301	17	Kazakhstan 5	17
7	Canada 2	94	18	Uzbekistan 1	177
8	Canada 3	5	19	South Africa 1	56
9	Canada 4	11	20	Namibia 1	169
10	Australia 1	397	21	Czech Republic 1	27
11	Australia 2	59			

Nuclear Forensics Experimentation

Robel, Kristo, Heller: Imbalance of the Uranium Ore dataset

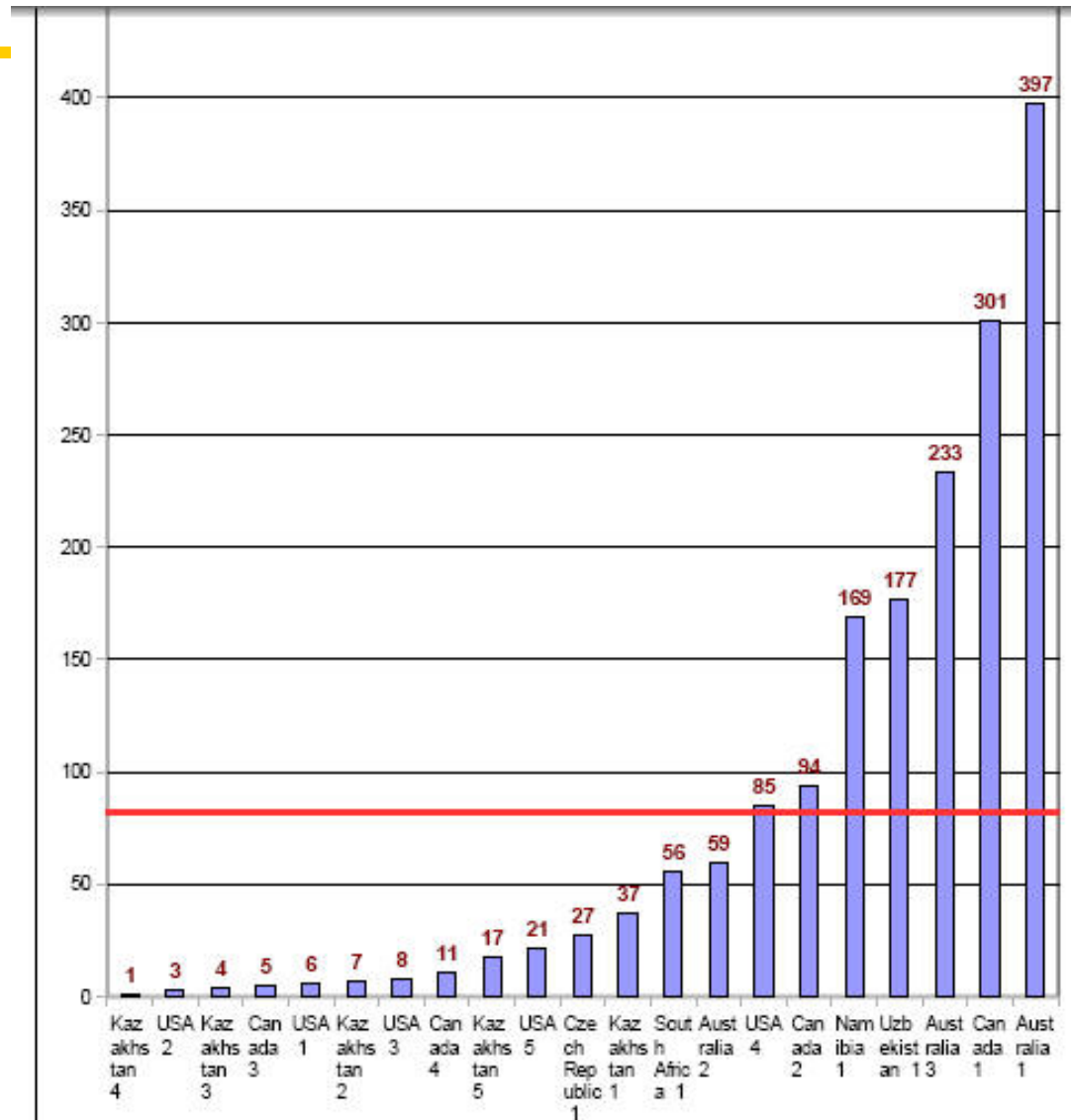


Figure 3: Bar Chart of Uranium sites from LLNL OC data with mean sample size

Spent Nuclear Fuel Database SFCOMPO (source: OECD Nuclear Energy Agency)



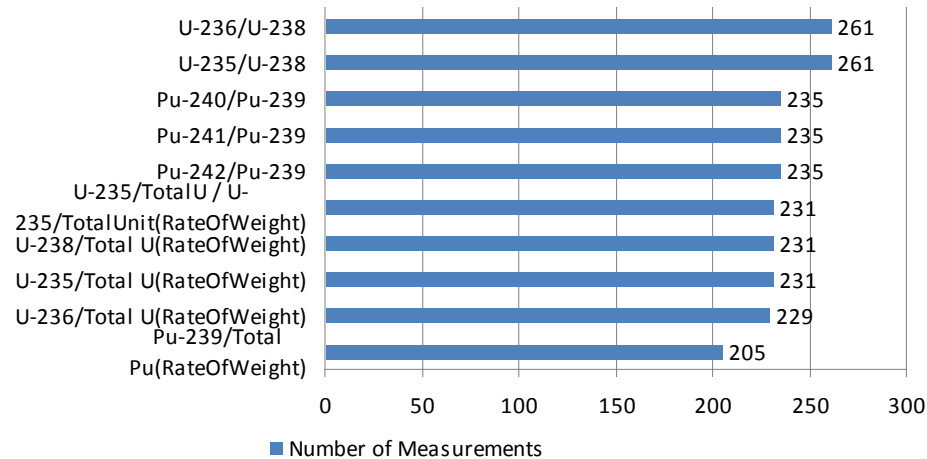
To experiment, we downloaded this spent fuel measurement database (html tables) from the web :

- **14 reactors from 4 countries** (light water. BWR,PWR)
Germany, Italy, Japan, USA
- **273 Samples** (variable number per reactor)
 - Maximum samples (Trino Vercellese, IT): 51
 - Minimum samples (Genkai-1, JA): 2
- **10,339 Measurements** of Isotopes, Isotope Ratios and Burnup, (variable number for each sample)

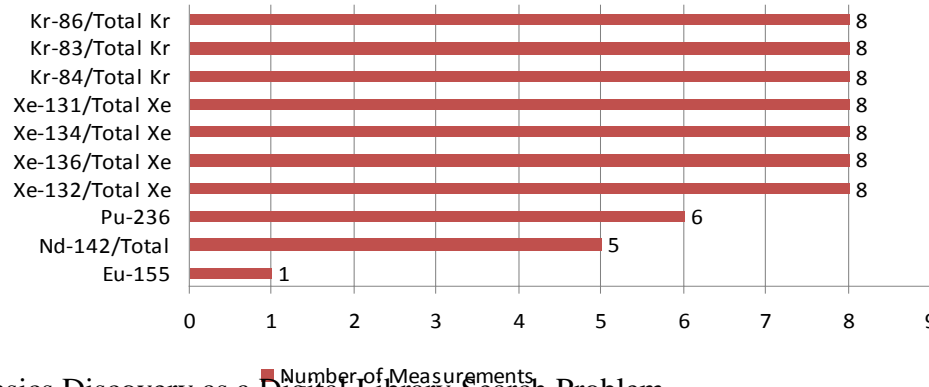
SFCOMPO Spent Nuclear Fuel Variable Measurement Characteristics



Top 10 Isotopes and Ratios Measurement Counts



Bottom 10 Isotopes and Ratios Counts



Nuclear Murder and Attribution



- On November 1, 2006, **Alexander Litvinenko**, former Russian Federal Security officer was poisoned by **Polonium-210** isotope while having lunch at a London sushi restaurant. He died of radiation poisoning three weeks later.
- According to doctors, "Litvinenko's murder represents an ominous landmark: the beginning of an era of nuclear terrorism."
- **Polonium-210** (^{210}Po) is an isotope of Polonium with a significant half-life (138 days). It decays by emitting alpha particles which can be easily shielded by even pieces of paper or the human skin
- UK authorities were reported to have traced the material to a nuclear reactor in Russia **HOW DID THEY DO THIS?**

SFCOMPO Spent Nuclear Fuel Data

A Naive Search Experiment: Structure



1. **Assume the temporal effects are negligible** on measurements and measurement ratios[†]
2. **A single sample is removed from the set of samples in the database. This sample becomes the “query sample” (the seized sample of unknown origin) and all other 260 samples are the “document samples” (to invoke search terminology).**
3. **A similarity matching algorithm is developed which matches each measurement in the query sample with its equivalent measurement in each document sample. This match results in a number between zero and 1 called a Retrieval Status Value (RSV) (ideally it is an estimate of a matching probability).**
4. **Document samples are ranked by this RSV similarity value.**
5. **Relevance of the document sample to the query sample is assessed as follows:**
 1. **If a document sample comes from the same reactor as the query sample, then the document sample is judged relevant.**
 2. **Otherwise it is Irrelevant**
6. **Standard web retrieval performance measure (precision at rank 10) is used**

[†] assumption used by Robel and Kristo (2008)

Search Performance Measure: Precision (standard web retrieval evaluation)



1. The standard measure of performance for web retrieval is the computation of **precision at rank ten**.
2. **Precision** for each ranked document (web page) is **the fraction of relevant documents divided by the rank number, i.e.**
 1. If the first document is relevant, precision at 1 is 1.0
 2. If the second document is irrelevant, precision at 2 is 0.5
 3. If the third document is relevant, precision at 3 is .667
 4. If the fourth document is irrelevant, precision at 4 is again 0.5
3. **Only the first ten ranked web pages are judged** for relevance or irrelevance

SFCOMPO Search: Performance by Reactor



Reactor Name	Reactor Country	Number of Measurement Sets	Max Possible Precision	Random Expected Precision	Actual Precision (per reactor)	Actual / Max Possible Precision	Actual / Random Precision
JPDR	Japan	30	1	0.11	1.00	1.00	8.96
Monticello	USA	30	1	0.11	0.85	0.85	7.62
Tsuruga-1	Japan	10	0.90	0.04	0.53	0.59	14.25
Trino_Vercellese	Italy	52	1	0.19	0.24	0.24	1.27
Fukushima-Daini-2	Japan	18	1	0.07	0.21	0.21	3.14
Takahama-3	Japan	16	1	0.06	0.16	0.16	2.69
Fukushima-Daiichi-3	Japan	36	1	0.13	0.16	0.16	1.20
Obrigheim	Germany	33	1	0.11	0.15	0.15	1.40
Genkai-1	Japan	2	0.10	0.01	0.10	1.00	13.32
H.B.Robinson-2	USA	7	0.60	0.03	0.09	0.15	3.47
Cooper	USA	6	0.50	0.02	0.07	0.14	3.14
Gundremmingen	Germany	15	1	0.06	0.06	0.06	1.00
Mihama-3	Japan	9	0.80	0.03	0.06	0.08	1.76
Calvert_Cliffs-1	USA	9	0.80	0.03	0.06	0.08	1.79
Overall		273	0.84	0.07		0.37	4.86
Overall Precision					0.34		

Search Experiment Implications and Next Steps



- 1. Performance seems promising considering the crudeness of the assumptions (however we may only be correlating burn-up -- needs further investigation)**
- 2. What might happen if the following improvements were made?**
 - 1. All measurements are available instead of selected ones**
 - 2. All measurements are normalized to a uniform precise time**
- 3. Collaborators at PNNL (funded by DNDO/NTNFC) are doing just that, by computationally:**
 - 1. Filling in (imputing) the missing values**
 - 2. Normalizing the actual /imputed measurements to a precise time**
- 4. Our group (with help from Bethany) is independently doing this.**
- 5. We will then re-run our search experiment on the “improved” database**
- 6. PNNL is expanding the database to other reactor types (e.g. graphite moderated)**

Future Directions and Activities



- 1. Expand collaboration to forensics groups at LLNL and ORNL. Martin Robel has suggested that the newest version of ORIGIN expands reactor types, so a more extensive analysis might be possible**
- 2. Attend the SFCOMPO meeting at OECD/NEA 9/19-20 in Paris**
- 3. Look into joining the Round Robin Exercises of the Nuclear Smuggling International Technical Working Group.**
- 4. Seek to access data about Uranium mines/ores for equivalent search experiments**
- 5. Professor Sunil Sunny Chirayath at TAMU may have access to real measurement data from Indian Nuclear Reactors (CANDU and FBR types).**
- 6. Begin to create nuclear forensics educational materials in collaboration with the UCB Nuclear Engineering Department**

Collaborators/Subject Matter Experts



Department of Nuclear Engineering, University of California, Berkeley
(**Bethany Goldblum**, Prof. Jasmina Vujic)

Nuclear Systems Design, Engineering and Analysis, Pacific Northwest National
Laboratory, Richland, WA (**Michaele (Mikey) Brady Raap**, Jon Schwantes)

Nuclear Science Division Isotopes Project, Lawrence Berkeley National
Laboratory, (**Richard Firestone**)

Chemistry & Materials Science Division, Los Alamos National Laboratory, Los
Alamos, NM (Lav Tandon. **Kevin Kuhn**)

Students



Chloe Reynolds, Masters of Information Management and Systems, School of Information, June 2012

Matthew Proveaux, incoming Masters student, Nuclear Engineering (MS 2014) (BS Physics, UC Davis)

David Weisz, incoming PhD student, Nuclear Engineering (MS Health Physics, nuclear non-proliferation track, Georgetown University), summer only.

Charles Wang, incoming Masters student, School of Information (MIMS 2014) (B.S. computer science)

Actively recruiting for fall 2012

Planning a steady state of 2+ graduate students until the FY 2013 budget situation is clarified.

Seeking NSF REU Undergraduate funding for summer 2013

Publications and Presentations



“Database Heterogeneity in a Scientific Application,” poster presentation at the IASSIST 2012 conference, June 6, Washington DC

“Applying Digital Library Technologies to Nuclear Forensics” to be published at the *International Conference on Theory and Practice of Digital Libraries (TPDL)*, Cypress September 23-27, 2012

“Nuclear Forensics: A Scientific Search Problem” to be presented at LWA 2102: *Lernen, Wissen, Adaption* Dortmund, Germany, September 12-14, 2012

Nuclear Forensics Search

Grant home page



<http://metadata.berkeley.edu/nuclear-forensics>

- **Contacts**
 - **Fred Gey (gey at berkeley dot edu)**
 - **Ray Larson (ray at ischool.berkeley dot edu)**
 - **Electra Sutton (electra at berkeley dot edu)**

Nuclear Forensics: Further Study



- **AAAS/APS 2008 report: Nuclear Forensics: Role, State of the Art, Program Needs**
- **ITU 2009 Presentation at IAEA: Data Interpretation in Nuclear Forensics**
- **Kristo, Tunny 2012, The State of Nuclear Forensics**
- **Robel, Kristo and Heller, LLNL report 2009: Nuclear Forensics Inferences Using Multidimensional Statistics**
- **Moody, Hutcheon and Grant, Nuclear Forensics Analysis, CRC Press 2005**

Finis (終わられる, Das Ende)



Quick Summary – Nuclear forensics discovery (attribution) can be approached as a search problem against libraries of nuclear signatures

We are developing various models of the search process

We have performed some experiments with at spent fuel database that look promising for our approach.

- **Thank you very much**
- **本当にありがとう (I hope Google translate is correct)**
- **Vielen Dank für Ihre Aufmerksamkeit**