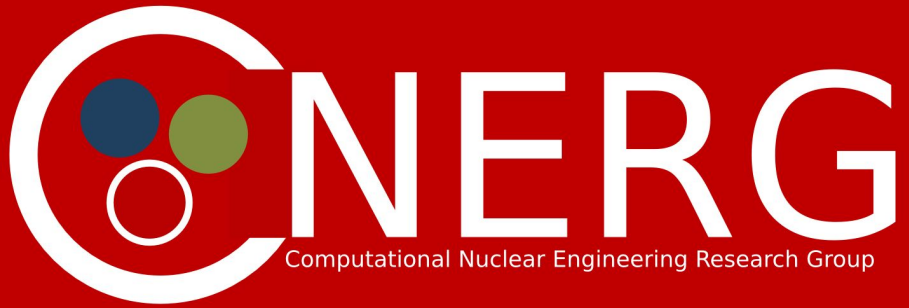
Leveraging HTC for nuclear security research

Arrielle C. Opotowsky
Advised by Paul P.H. Wilson
Dept of Engineering Physics, UW-Madison

HTCondor Week, 25 May 2022



Nuclear Security

The **prevention** and **detection** of and **response to**:

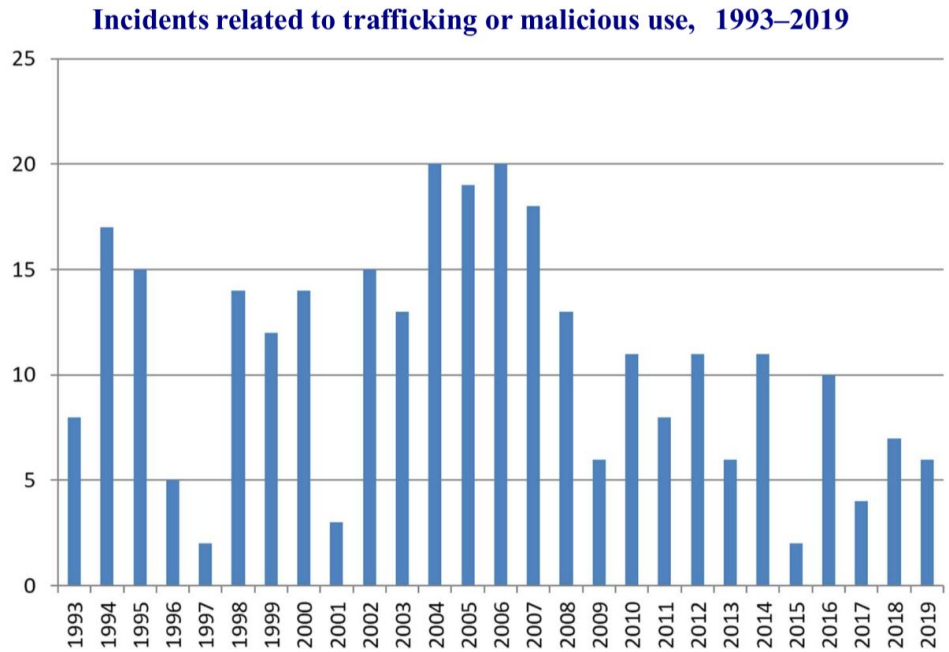
- Theft
- Sabotage
- Unauthorized access
- Illegal transfer
- Other malicious acts

involving:

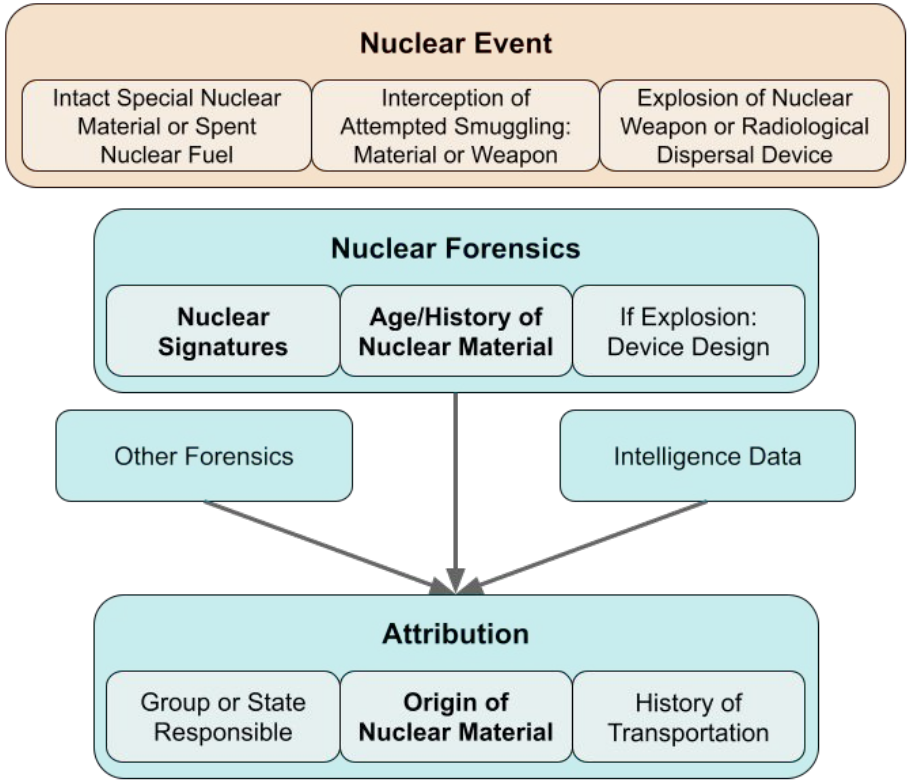
- Nuclear materials
- Other radioactive substances
- Their associated facilities

Definition from IAEA Safety Glossary and graphic obtained from this link

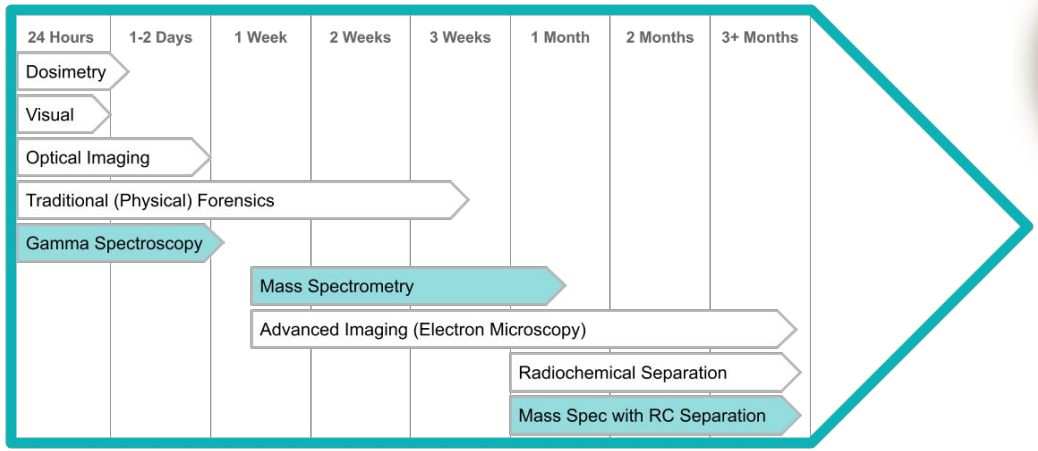
Nuclear Forensics



[IAEA Incident and Trafficking Database 2020 Factsheet: Incidents of nuclear and other radioactive material out of regulatory control]



Nuclear Forensics Investigations



Author recreated from online article by Eric Feigl-Ding

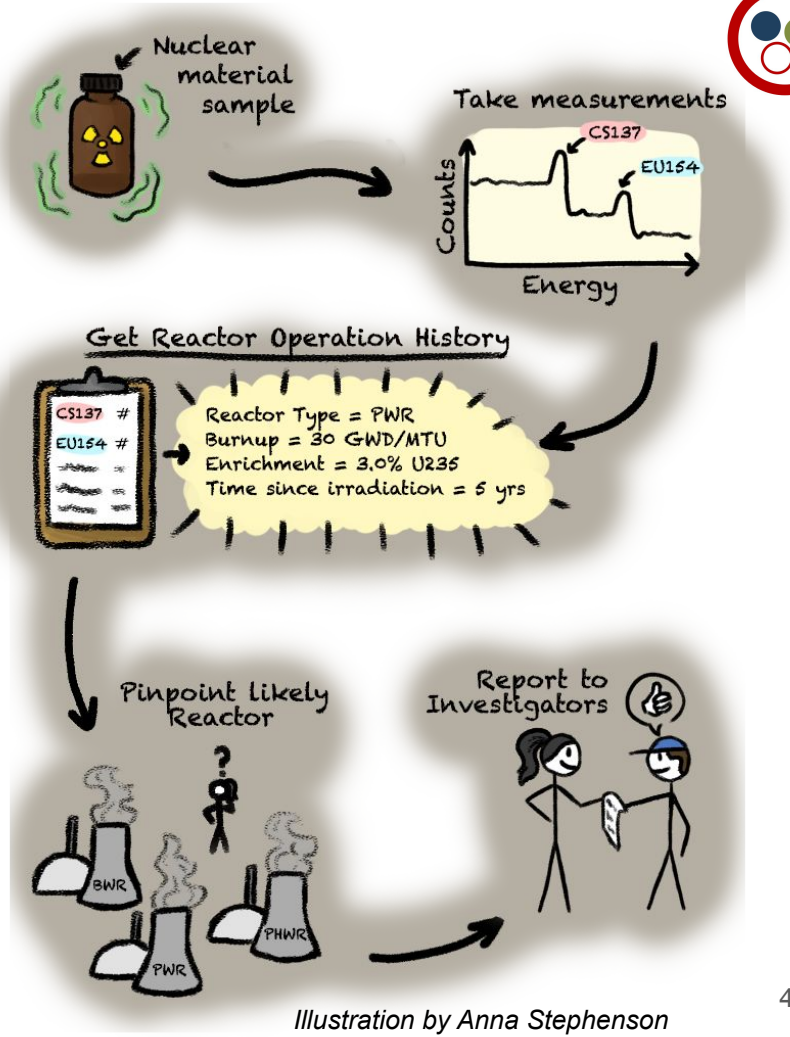


Illustration by Anna Stephenson

Can a computational approach expedite a nuclear forensics investigation?

Can a computational approach expedite a nuclear forensics investigation?

Learning About CHTC

- How I got there:
  - Step 1: Workflow no longer “works” on personal machine
  - Step 2: Complain dramatically to advisor
  - Step 3: CHTC to the rescue
- Phenomenal support:
  - One-on-one introduction/tutorial
  - Lots of email help + office hours
  - Once equipped with some experience, the manual helped a lot (not quite a power user …was cut off mid-montage)

Why CHTC?

- Machine learning-based workflow, ~10s of case variations
  - Algorithm & hyperparameter variations
  - Several labels to predict
- Size of training database(s)

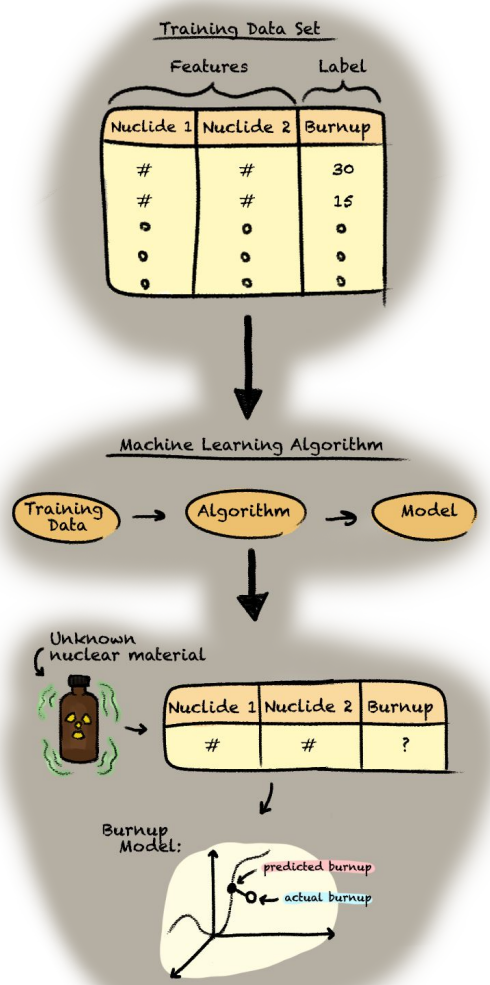


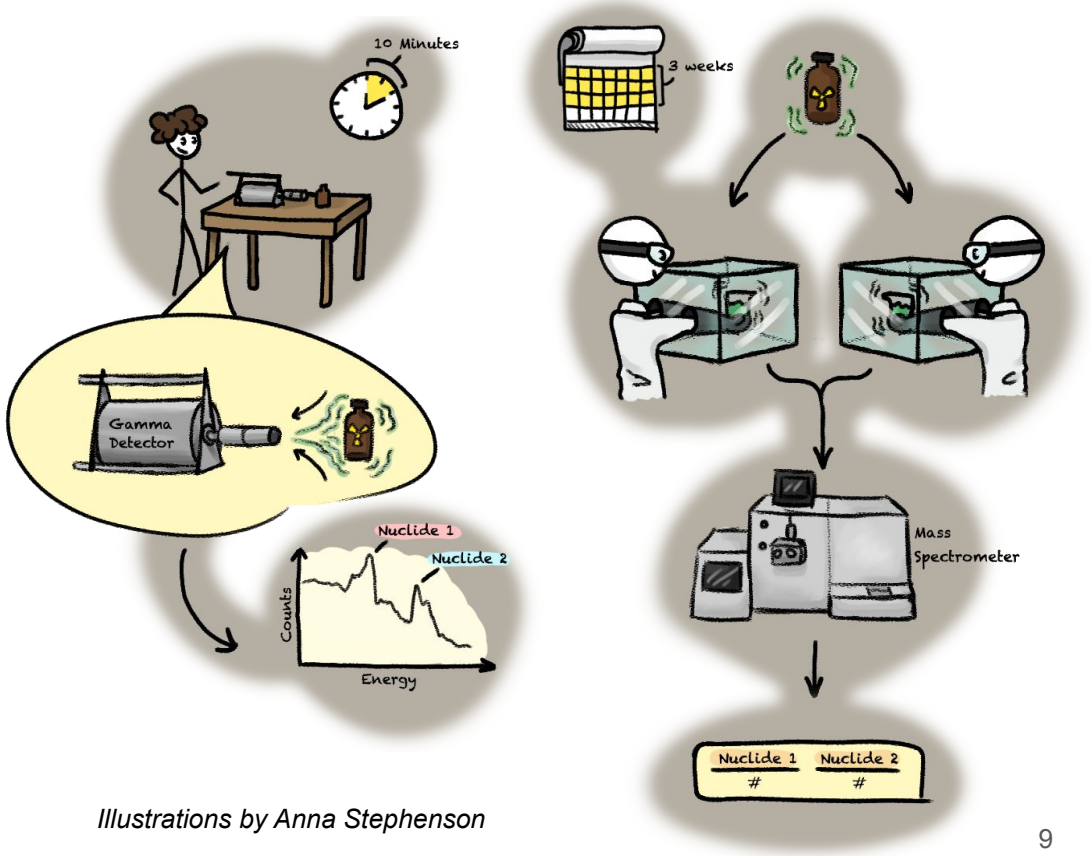
Illustration by Anna Stephenson

Database Dimensions

- Entries: ~500k
- Features (columns):
  - Slow: 29-32
  - Fast: 42-206

Fast Measurements

Slow Measurements



Illustrations by Anna Stephenson

2 Categories of Computation

- **Category 1**: Scikit-learn machine learning (2 algorithms)
  - One job per case submission
  - Some parallelism offered by scikit learn, typically requested ~4-8 CPUs
  - Large memory needs per job, many hours per job
- **Category 2**: 3rd algorithm based on other nuclear forensics work
  - Approx 500k calculations per case
  - 10k jobs per case submission (broke it into chunks ~50 calcs/job to not exceed 10k job limit)
  - Lower memory needs (1 CPU, <2GB) per job, minutes per job

2 Categories of Computation

- **Category 1**: Scikit-learn machine learning (2 algorithms)
  - One job per case submission
  - Some parallelism offered by scikit learn, typically requested ~4-8 CPUs
  - Large memory needs per job, many hours per job
- **Category 2**: 3rd algorithm based on other nuclear forensics work
  - Approx 500k calculations per case
  - 10k jobs per case submission (broke it into chunks ~50 calcs/job to not exceed 10k job limit)
  - Lower memory needs (1 CPU, <2GB) per job, minutes per job

Progression of CHTC Use

1. Category 1 initially (~2019)
2. Category 2 later (~2020)
3. “Fast Measurements” database creation. One-off use case to run some simulations:
   docker-enabled for specific environment (April 2021)
4. Many Category 1 & 2 case variations:
   expanded to UW Grid and Open Science Grid (~June 2021)

Tangential Support

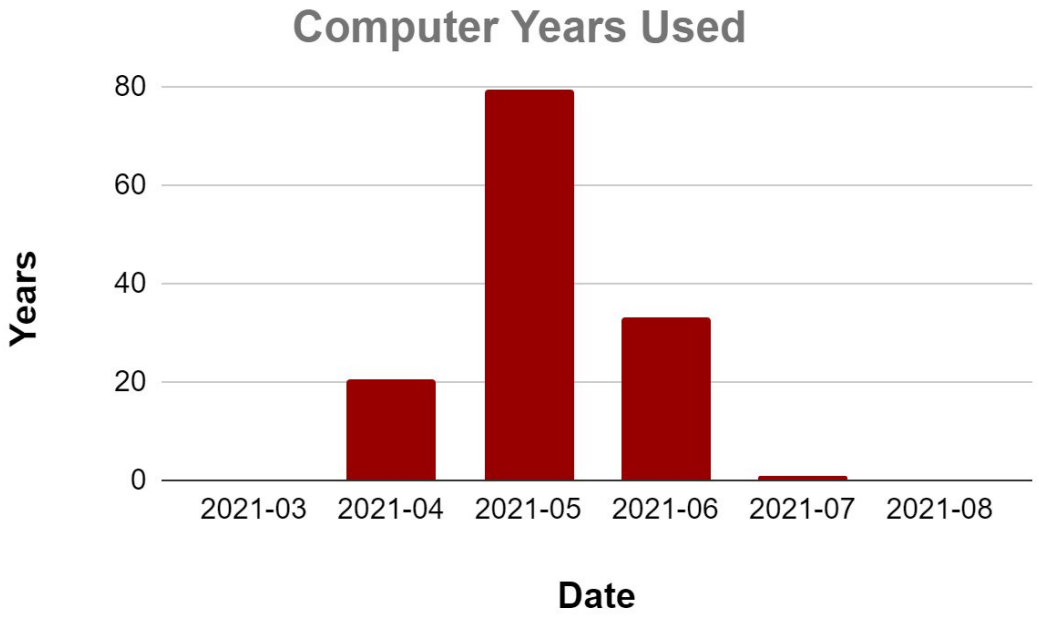
- Squid for large input databases
- Resolved stuck point 1: Migrating to OSG
  - For weeks, jobs kept getting booted!
  - Was misreading memory requirements (two different estimates existed) for awhile
- Resolved stuck point 2: Post-job large data transfers
  - Not feasible from off campus
  - Direct CHTC→UW Research Drive file transfer

Impacts of CHTC: Research

Computing power!

Without it, would have:

- much smaller database
- fewer parameter variations
- couldn't have done a "real world" scenario



Impacts of CHTC: Professional

- I would have gotten my PhD at ~ 171 years old
- Wrote a dissertation chapter for a general audience
- HTC experience helped me in job interviews
  - Knowledge of designing code
  - Knowledge of interacting with job submission systems
- Current job (nuclear methods software engineer) is computationally intensive

Acknowledgements

- Open Science Grid and CHTC!
- WI Institute for Science Literacy
  - http://scifun.org/Thesis_Awards/opotowsky.html
  - Illustrations are by Anna Stephenson
- DHS Nuclear Forensics Graduate Research Fellowship

This material is based upon work supported by the U.S. Department of Homeland Security under Grant Award Number, 2012- DN-130-NF0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

- NSF Graduate Research Fellowship
- Graduate Engineering Research Scholars
- Consortium for Monitoring, Technology, and Verification

Acknowledgements

- Open Science Grid and CHTC
- WI Institute for Science Literac
  - http://scifun.org/Thesis_Awards/opotowsky.ht
  - All illustrations are by Anna Stephenson
- DHS Nuclear Forensics Gradua

This material is based upon work supported by the U.S. Departm
Number, 2012- DN-130-NF0001. The views and conclusions con
should not be interpreted as representing the official policies, eith
Homeland Security.

- NSF Graduate Research Fello
- Graduate Engineering Resear
- Consortium for Monitoring, Te
  Verification