

A Mathematical and Data-Driven Approach to Intrusion Detection for High- Performance Computing

Brief Summary of Proposal Submitted to DOE

David H Bailey, LBNL

Participants

- David H. Bailey (LBNL)
- Juan Meza (LBNL)
- Deb Agarwal (LBNL)
- Scott Campbell (LBNL / NERSC)
- Sean Peisert (LBNL / U.C. Davis)
- Vern Paxson (LBNL / ICSI / U.C. Berkeley)
- Robin Sommer (LBNL / ICSI)
- Matt Bishop (U.C. Davis)

Approach

- Mathematics and statistics
 - Apply some known mathematical and statistical techniques to perform the equivalent of credit card fraud detection on the access and usage of high-performance computer systems.
 - Investigate some new promising techniques that potentially might be more effective.
- Data
 - Utilize the vast store of NERSC network access and user data.

Some Details on Approach

- Intrusion detection methods
- Forensic logging and analysis
- “Time Machine” – permits one to focus on “interesting” traffic
- Ensemble-based learning methods
- Spectral graph theory
- Outlier detection
- Modeling malicious behavior
- Provide “anonymous” data to others for research

Sample of Data to be Analyzed

- Network access patterns.
- Shell commands executed.
- Programs executed -- name and size of jobs.
- Collaboration network.