Predicting and Fixing Vulnerabilities Before They Occur: A Big Data Approach

Hong-Mei Chen, Rick Kazman, Ira Monarch, Ping Wang
A Cautionary Tale

Secure Coding in C and C++

Robert C. Seacord

Foreword by Richard D. Peterson
CERT Director

May 2005

2007
A Cautionary Tale
66% CAGR for security incidents since 2009

Reactive, not Proactive
Hacker Communities

- Hackers form communities.
- They are learning communities
- Hacker communities, as with all innovation communities, need to share information
- By analyzing the topics in hackers' discussions, we will be able to get early indication as to which vulnerabilities are likely to be the focus of upcoming attacks.
- Early insight can lead to early quality assurance and mitigation strategies.
Our Approach

- Mining publicly available vulnerability, exploit, and attack databases.
Challenge: Tracking Concept Evolution

1) mining security data sources
2) mapping the relationships among these concepts and tracking their changes through time
3) building a security ontology based on the results of 1) and 2)
End Game

- The PCS rests on a big data infrastructure for:
  - extracting information from public data sources,
  - transforming (cleaning) and loading the data,
  - clustering and visualizing it, and
  - curating it for future use.

- This research, if successful, will help the security assurance community to be proactive rather than reactive.
Shameless Plug

- Submit a paper to the
- HICSS 50 Big Data Engineering Minitrack
- Submission Deadline: June 15, 2016
- Conference: Hawaii (Big Island)
  January 4-7, 2017