Driving Data in the Cybersecurity Economy

Erin Kenneally
U.S. Dept of Homeland Security
Cyber Security Division
IMPACT Motivation: The ‘Open Secret’ of Effective R&D

- **Data are critical to R&D capabilities**
  - Exactly 0% of R&D (quality) possible sans data
  - Cybersecurity needs real-world data to develop, test, evaluate knowledge & tech solutions to counter cyber threats
  - “Big Data” may grow on trees but still has to be picked, sorted, trucked

- **Decision analytics are critical to HSE capabilities**
  - Cybersecurity needs integrated, holistic understanding of risk environment
  - Gap between Data <---> Decisions: multi-dimensional, complex association and fusion, high-context presentation elements

- **Data sharing + Analytics |= Easy**
  - High value data = High legal risk + $$
  - Data rich vs. data poor
  - Expensive to abstract away low level knowledge- and labor- intensive tasks
  - Technologists optimize for Efficiency, Lawyers optimize for Certainty

2018 Kenneally
- Data profiles and central indexing
- Connect distributed providers and seekers of data and analytics
- Build collective knowledge
- Optimize data use and analysis

Business Rules Broker, Legal, Ethics

Metadata Discovery (FIND)
Data Matchmaking (GET)
Social Networking (FEEDBACK)
Tool Matchmaking (USE)

2018 Kenneally
IMPACT ROI

- **Parity**: lower barrier to entry for data impoverished viz federation of data Supply & Demand (academic, industry, govt)
- **Scale**: beyond interpersonal relationships, ad hoc acquisitions
- **Sustainable**: Uniform, repeatable process
- **Utility**: responsible innovation over risk-aversion
- **Trust**
  - Vetted data, researchers, providers
  - Balance efficiency and certainty
  - Legal and ethical accountability
Shop til You Drop
IMPACT Portal <www.ImpactCyberTrust.org>

This is a central metadata index of all the data available in IMPACT from our federation of Providers. Browse our data catalog using the Text Search box or the Filter Search feature on the left side of the page. Note: You must log in as a Researcher to request data.

Filter:
- Year: 2015
- Cat: DNS Data
- Cat: Internet Topology Data

Result Count: 12
(results sorted by search relevance)

<table>
<thead>
<tr>
<th>Cart</th>
<th>Name</th>
<th>Provider</th>
<th>Collection Dates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GT Malware Passive DNS Data Daily Feed</td>
<td>Georgia Tech</td>
<td>2015-07-01 to Ongoing</td>
</tr>
<tr>
<td></td>
<td>IPv4 Prefix-Probing Current</td>
<td>UCSD - Center for Applied Internet Data Analysis</td>
<td>2015-12-09 to Ongoing</td>
</tr>
<tr>
<td></td>
<td>IPv4 Route/24 DNS Names Current</td>
<td>UCSD - Center for Applied Internet Data Analysis</td>
<td>2008-03-01 to Ongoing</td>
</tr>
<tr>
<td></td>
<td>IPv4 Route/24 Topology</td>
<td>UCSD - Center for Applied Internet Data Analysis</td>
<td>2007-09-13 to Ongoing</td>
</tr>
<tr>
<td></td>
<td>IPv4 Route/24 Topology Current</td>
<td>UCSD - Center for Applied Internet Data Analysis</td>
<td>2007-09-13 to Ongoing</td>
</tr>
</tbody>
</table>
Data Trends

Source: DHS IMPACT program; SRI analysis, Apr. 17

No Data in 2007
Global, Multi-Sector "Impact" (as of Jul 2017)

Total Users (1,987)

- Government: 228, 11%
- Commercial: 762, 39%
- Foreign: 223, 11%
- Academic: 709, 36%
- Private: 36, 3%

Research papers, journals, tech reports (>300 “known”)

Dataset Provisioned (>3,500)

Approved Foreign Users (236 Total)

- UK: 25%
- CAN: 20%
- ISRAEL: 14%
- JP: 7%
- NL: 5%
- SG: 3%
- AUS: 25%
- NL: 5%
- SG: 3%
- AUS: 25%

Source: DHS IMPACT program; SRI analysis, Apr ’17
Success Elements

- Findable
  Centralized Mediation

- Responsible
  Legal & Ethical framework integrated

- New, high-value datasets

- FREE

- Tools to USE the data

- Diverse
  Real-world Problem-driven Data

- Distributed Provisioning

- Engage International data and researchers

2018 Kenneally
Evolved IMPACT R&D Approach

Market need:

- Existing capabilities do not provide cyber risk decision analytic support needed by HSE
  - Security, Integrity, Stability, Resilience of networks
  - Sensitive data sharing and controlled data disclosure
  - Interdependencies, cascading, and aggregate effects of cyber-vulnerabilities and attacks across platforms and enterprises
  - Changing risk environment demands dynamic cyber security R&D
  - < time & effort to find, curate, normalize, understand high volume, velocity, variety, value
  > time extracting insight and meaningful decisions from data

Product:

- 1st-gen R&D-enabling infrastructure democratized **data raw materials** (Data Providers)
- New BAA fosters evolved R&D infrastructure adds **derivative data products and tools** for HSE: Decision Analytics-as-a-Service Providers (DASP)

2018 Kenneally
NGI Recap

Recap

Business Layer

Application Layer

Data Layer

Analytic Requirements

Data Requirements

Decision-Analytics-As-a-Service Provider Network

PERFORMERS

(1) Cyber Security Challenge Problem(s) (What)

(2) Data or Analytic/Tool Capability (How)

Mediator Infrastructure

Mediator Infrastructure

Decision Analytics

Data

Data Provider Network

External Data Providers

Researchers & Developers

DHS Components

DHS (Cyber) Critical Infrastructure Providers

Decision-Analytics-As-a-Service Provider Network

DaaSP

DaaSP

DaaSP

DaaSP

DaaSP

Inputs

Outputs

RFI

DHS Components

Authoritative Govt Docs

Other

Census

Legacy & Social Networking Mechanisms

CSD ‘Data Mgmt Plan’
Socialization

IMPACT Survey

IMPACT Data Request and User Experience

This survey is sent in conjunction with the evolved data sharing program, IMPACT (Information Marketplace for Policy and Analysis of Cyber-risk & Trust), supported by DHS S&T Cyber Security Division. This survey will help us in our efforts to enhance data sharing to support cyber security research & development for industry, academia and the government. Please fill out the survey as completely as you can to the best of your knowledge. The survey will remain open for two weeks.

https://www.ImpactCyberTrust.org/#knowledgebase
Why Engage IMPACT

- How do companies address risks associated with data sharing for academic research?*
  - Engage in a rigorous internal review of proposed academic research projects.
  - Close to half of the companies retain custody and control over the research data at all times.
  - Companies employ rigorous data use agreements to limit access to and use of shared data.

How IMPACT addresses risks

- Vet Researchers, Providers, Data
- Provider can host and provision own data
- Provider can engage Disclosure Control-as-a-Service for very sensitive data that allows analysis without Researcher seeing data
- Provider leverages standardized Researcher data use agreements with customized additional restrictions by Provider

* “UNDERSTANDING CORPORATE DATA SHARING DECISIONS: PRACTICES, CHALLENGES, AND OPPORTUNITIES FOR SHARING CORPORATE DATA WITH RESEARCHERS” Future of Privacy Forum (2017)
<table>
<thead>
<tr>
<th>Name</th>
<th>Data Provider</th>
</tr>
</thead>
<tbody>
<tr>
<td>GT Malware Passive DNS Data Daily Feed</td>
<td>Georgia Tech</td>
</tr>
<tr>
<td>Historical GT Malware Passive DNS Data 2011-2013</td>
<td>Georgia Tech</td>
</tr>
<tr>
<td>US Long-haul Infrastructure Topology</td>
<td>University of Wisconsin</td>
</tr>
<tr>
<td>DARPA Scalable Network Monitoring (SNM) Program Traffic</td>
<td>DARPA</td>
</tr>
<tr>
<td>Skaion 2006 IARPA Dataset</td>
<td>SKAION</td>
</tr>
<tr>
<td>GT Malware Unsolicited Email Daily Feed</td>
<td>Georgia Tech</td>
</tr>
<tr>
<td>DSHIELD Logs</td>
<td>University of Wisconsin</td>
</tr>
<tr>
<td>syn-flood-attack</td>
<td>Merit Network, Inc.</td>
</tr>
<tr>
<td>Netflow-1</td>
<td>Merit Network, Inc.</td>
</tr>
<tr>
<td>DoS_traces-20020629</td>
<td>University of Southern California-Information Sciences Institute</td>
</tr>
<tr>
<td>NCCDC 2013</td>
<td>Center for Infrastructure Assurance and Security (UTSA/CIAS)</td>
</tr>
<tr>
<td>NCCDC 2014</td>
<td>Center for Infrastructure Assurance and Security (UTSA/CIAS)</td>
</tr>
<tr>
<td>DoS_80_timeseries-20020629</td>
<td>University of Southern California-Information Sciences Institute</td>
</tr>
<tr>
<td>CAIDA DDoS 2007 Attack Dataset</td>
<td>UCSD - Center for Applied Internet Data Analysis</td>
</tr>
<tr>
<td>Netflow-2</td>
<td>Merit Network, Inc.</td>
</tr>
<tr>
<td>Netflow-3</td>
<td>Merit Network, Inc.</td>
</tr>
<tr>
<td>NCCDC 2011</td>
<td>Center for Infrastructure Assurance and Security (UTSA/CIAS)</td>
</tr>
<tr>
<td>NTP DDoS 2014</td>
<td>Merit Network, Inc.</td>
</tr>
<tr>
<td>NCCDC 2015</td>
<td>Center for Infrastructure Assurance and Security (UTSA/CIAS)</td>
</tr>
<tr>
<td>UCSD Real-time Network Telescope Data</td>
<td>UCSD - Center for Applied Internet Data Analysis</td>
</tr>
<tr>
<td>Resource Provider</td>
<td>Overview</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Massachusetts General Hospital</td>
<td>- Network attack detection and analysis of medical device honeypots. - Continuous monitoring of medical device traffic. - Implementation of medical device forensic analysis tools. - Symantec: SymantecEndpoint Threat Visibility</td>
</tr>
<tr>
<td>Partners</td>
<td>- DARPA: DARPA's DARPA Cybersecurity Applications and Tools program supports the development of new tools and methodologies for medical device security. - DARPA: DARPA's DARPA Cybersecurity Applications and Tools program provides funding for research and development of medical device security tools. - DARPA: DARPA's DARPA Cybersecurity Applications and Tools program supports the development of new tools and methodologies for medical device security.</td>
</tr>
<tr>
<td>RI</td>
<td>- Medical device honeypots: Multiple honeypots are deployed to attract and analyze threats to medical devices. - Real-time monitoring of medical device traffic: Continuous monitoring of medical device traffic to detect potential threats. - Case studies: Multiple case studies of medical device cyberattacks are analyzed to provide insights into threats and vulnerabilities.</td>
</tr>
<tr>
<td>Foundation</td>
<td>- Medical device data: Multiple medical device data sources are analyzed to provide insights into threats and vulnerabilities. - Real-time monitoring: Continuous monitoring of medical device traffic to detect potential threats. - Case studies: Multiple case studies of medical device cyberattacks are analyzed to provide insights into threats and vulnerabilities.</td>
</tr>
<tr>
<td>IJ</td>
<td>- Medical device honeypots: Multiple honeypots are deployed to attract and analyze threats to medical devices. - Real-time monitoring of medical device traffic: Continuous monitoring of medical device traffic to detect potential threats. - Case studies: Multiple case studies of medical device cyberattacks are analyzed to provide insights into threats and vulnerabilities.</td>
</tr>
<tr>
<td>GTSC</td>
<td>- Medical device honeypots: Multiple honeypots are deployed to attract and analyze threats to medical devices. - Real-time monitoring of medical device traffic: Continuous monitoring of medical device traffic to detect potential threats. - Case studies: Multiple case studies of medical device cyberattacks are analyzed to provide insights into threats and vulnerabilities.</td>
</tr>
<tr>
<td>U. Wisconsin</td>
<td>- Medical device honeypots: Multiple honeypots are deployed to attract and analyze threats to medical devices. - Real-time monitoring of medical device traffic: Continuous monitoring of medical device traffic to detect potential threats. - Case studies: Multiple case studies of medical device cyberattacks are analyzed to provide insights into threats and vulnerabilities.</td>
</tr>
<tr>
<td>CALDA</td>
<td>- Medical device honeypots: Multiple honeypots are deployed to attract and analyze threats to medical devices. - Real-time monitoring of medical device traffic: Continuous monitoring of medical device traffic to detect potential threats. - Case studies: Multiple case studies of medical device cyberattacks are analyzed to provide insights into threats and vulnerabilities.</td>
</tr>
</tbody>
</table>