Data-Driven Assessment of Cyber Risk: Challenges in Assessing and Mitigating Cyber Risk

Mustaque Ahamad, Saby Mitra and Paul Royal

Georgia Tech Information Security Center

Georgia Tech Research Institute

(In collaboration with the World Economic Forum)

Talking About Cyber Risk

- Risk = *Prob.[adverse event]**Impact[adverse event]*
- Attacks occur when threat sources exploit vulnerabilities
- Mean-time-to-compromise?
- Mean-time-to-recover? (assuming detection)
- Traditional dependability assumptions and solutions do not apply.

Why Even Try It?

- **Current cyber risk is anecdote and perception based** and we lack the ability to objectively assess the risk posed by ever evolving cyber threats.
- **Current cyber security threat data is fragmented** and collected by disparate entities such as security vendors, vendors serving different sectors and academic research centers.
- **Publicly available cyber security data is often delayed** and does not provide the ability to quickly respond to new threats that require coordinated effort within a short time.
- **A trusted data sharing and analysis platform** that brings data from multiple sources and provides novel analysis will increase our ability to respond to emerging threats quickly and effectively.

Approach

Explore partnerships to collect cyber risk relevant data from multiple sources and analyze it to create metrics that summarize current cyber security threats

- **Combine *public* and *proprietary* data sources** on cyber threats such as software vulnerabilities, drive-by downloads and malware from a variety of cyber security organizations.
- **Provide *threat analytics* and *visualization* tools** suitable for novice and advanced users, and that can be customized based on industry, technology platform, or geographic region

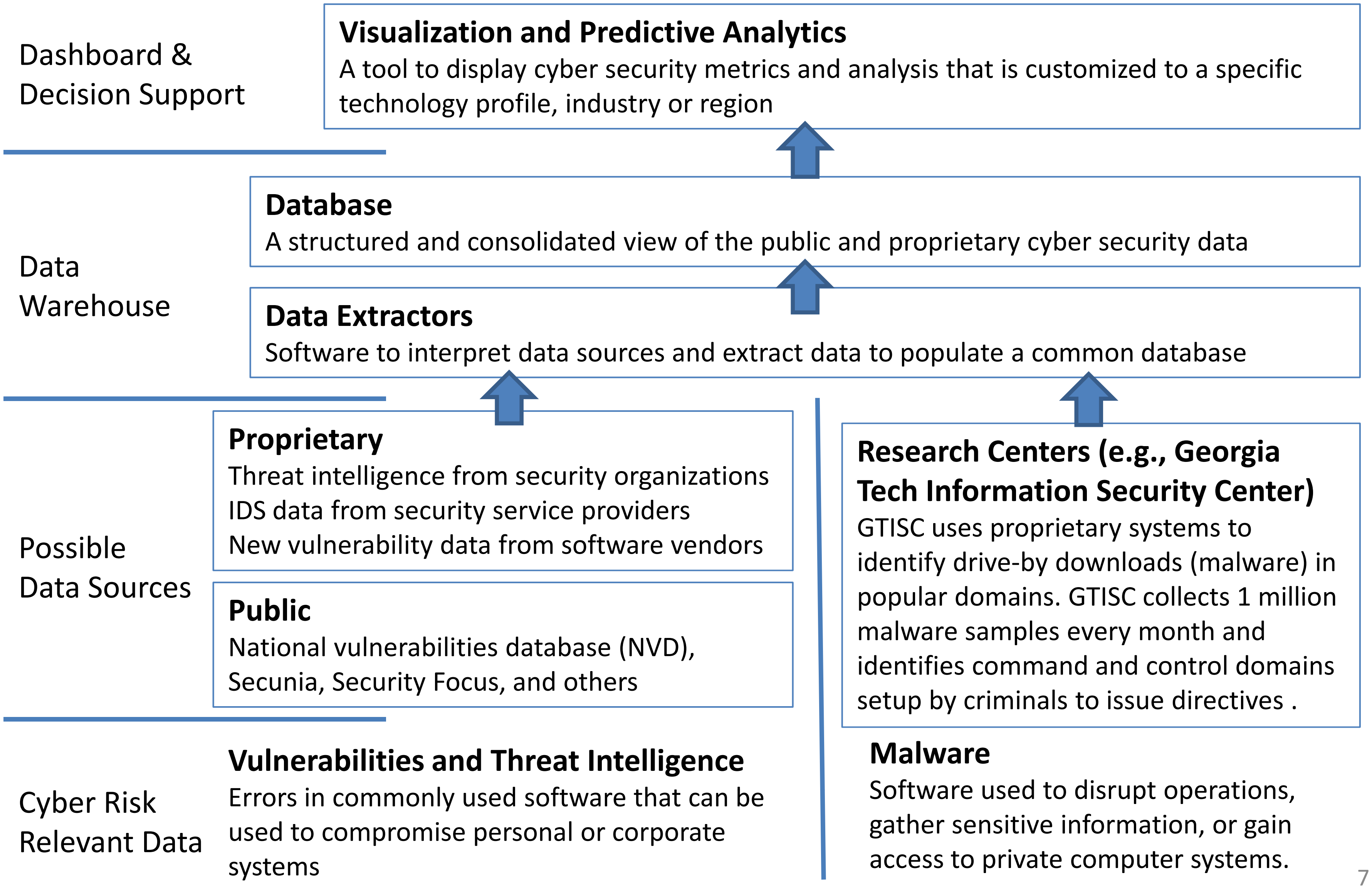
Key Questions

- What data is relevant?
  - Vulnerabilities, alerts from IDS system, compromised or malicious services?
- Where does the data come from?
  - Public, proprietary from security vendors or government or private entities?
- What can we do with such data for better understanding of cyber risk?
  - Analysis, visualization, prediction?
- What value does a cyber risk tool offer?
  - Actionable information?

Current Data Sources

- Public data
  - Vulnerabilities reported to NVD
- Summarized proprietary data
  - Drive-by-download risk data from a major security vendor
- Potentially malicious network traffic targeting an enterprise
  - IDS/IPS alert data captured from Georgia Tech networks

Overall System Architecture



Dashboard & Decision Support

Visualization and Predictive Analytics
A tool to display cyber security metrics and analysis that is customized to a specific technology profile, industry or region

Data Warehouse

Database
A structured and consolidated view of the public and proprietary cyber security data

Data Extractors
Software to interpret data sources and extract data to populate a common database

Possible Data Sources

Proprietary
Threat intelligence from security organizations
IDS data from security service providers
New vulnerability data from software vendors

Public
National vulnerabilities database (NVD), Secunia, Security Focus, and others

Research Centers (e.g., Georgia Tech Information Security Center)
GTISC uses proprietary systems to identify drive-by downloads (malware) in popular domains. GTISC collects 1 million malware samples every month and identifies command and control domains setup by criminals to issue directives .

Cyber Risk Relevant Data

Vulnerabilities and Threat Intelligence
Errors in commonly used software that can be used to compromise personal or corporate systems

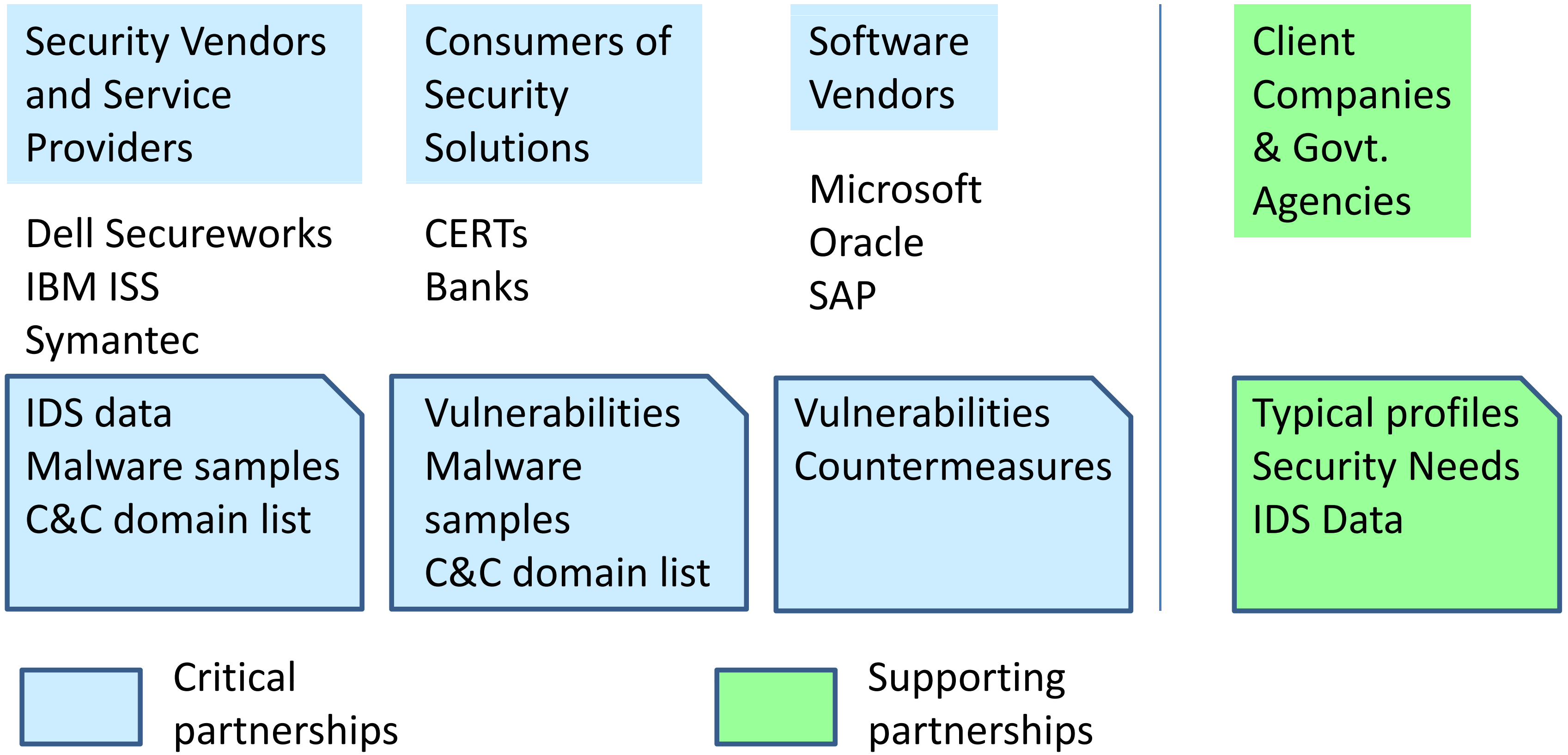
Malware
Software used to disrupt operations, gather sensitive information, or gain access to private computer systems.

The Why and What

	Vulnerabilities	Malware
Why we need	**Predictive Analysis** Expected volume/severity of attacks on a day Expected number of 0 day vulnerabilities on a day **Coordinated Response** Sharing of countermeasures / response to threats	**More Comprehensive Response** More malware samples and more C&C domains will provide for a more protected environment for everyone
What we need	**Threat Intelligence** Emerging threat intelligence from security organizations **Alert Data** *Intrusion Detection System* Data from security service providers like IBM and Dell **New Vulnerabilities** *New Vulnerability Data* from software vendors	**Malware samples and C&C Domains** Additional malware samples and C&C domains from security service providers and security vendors to be shared within a trusted group
What we have	**Public Vulnerability Data** National vulnerabilities database (NVD), Secunia, Security Focus, and others	**GT Information Security Center** GTISC collection of 1 million malware samples every month, as well as command and control (C&C) domains.

Challenge I – Access to Real-world Threat Data

Data Sources: Partnerships with various organizations to obtain cyber risk relevant data is critical for the success of the project



Security Vendors and Service Providers	Consumers of Security Solutions	Software Vendors	Client Companies & Govt. Agencies
Dell Secureworks IBM ISS Symantec	CERTs Banks	Microsoft Oracle SAP	
IDS data Malware samples C&C domain list	Vulnerabilities Malware samples C&C domain list	Vulnerabilities Countermeasures	Typical profiles Security Needs IDS Data

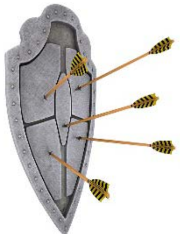
🟦 Critical partnerships

🟩 Supporting partnerships

Challenge II – Analytics

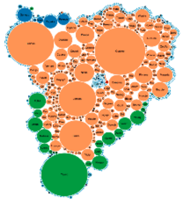
Analytics: While combining data sets provides new opportunities, developing customized tools will depend on the data feeds available

Drive-by Download Risk



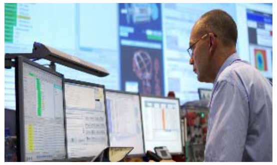
- Compromised websites infect user machines just because they visit
- Serious threats for everyday users
- Georgia Tech can detect likelihood of such infections

Behavior Fingerprints of Malware



- Rapidly changing malware means we must focus on execution behavior
- Georgia Tech processes about 100,000 samples each day
- Malware families and spread

What is My Cyber Risk Today?



- IT profile and security posture
- Value associated with target
- Observed malicious activity
- Mitigation options and ability

Predictive Analytics

- Epidemiological analysis
  - How far can an attack spread?How rapidly can it spread? Are certain sectors under higher risk?
- "What if" scenarios
  - How would these change with a specific mitigation plan?

Challenge III – Threat Visualization for Actionable Information

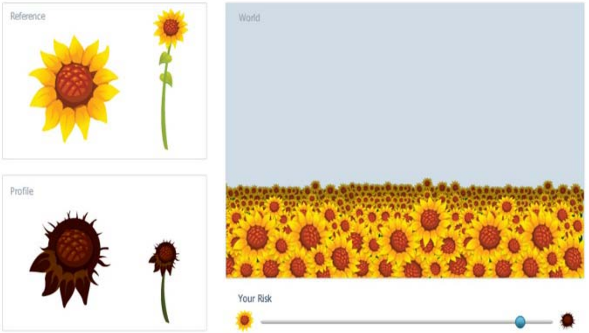
Visualization: Aggregating all the data feeds in a meaningful way to provide a cyber threat barometer is difficult.

Using Visualization for Navigating Large Amounts of Threat Data

Data overload is a serious problem

"Flower field" metaphor for presenting big picture

Threatened assets can be easily identified for additional analysis



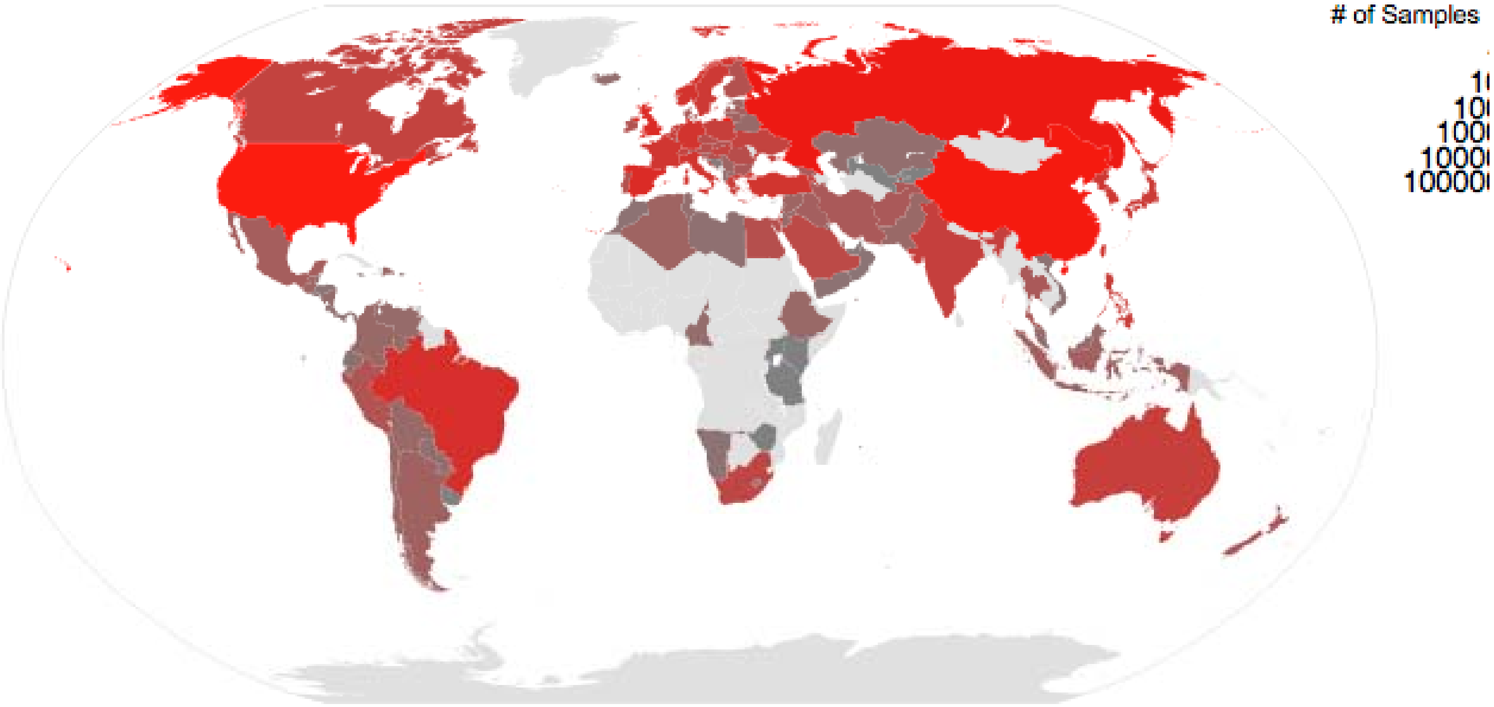
From Big Picture to Deeper Insights

An abnormal asset visualization points to increased risk

Click on it can provide details of vulnerabilities, exploits and attack information

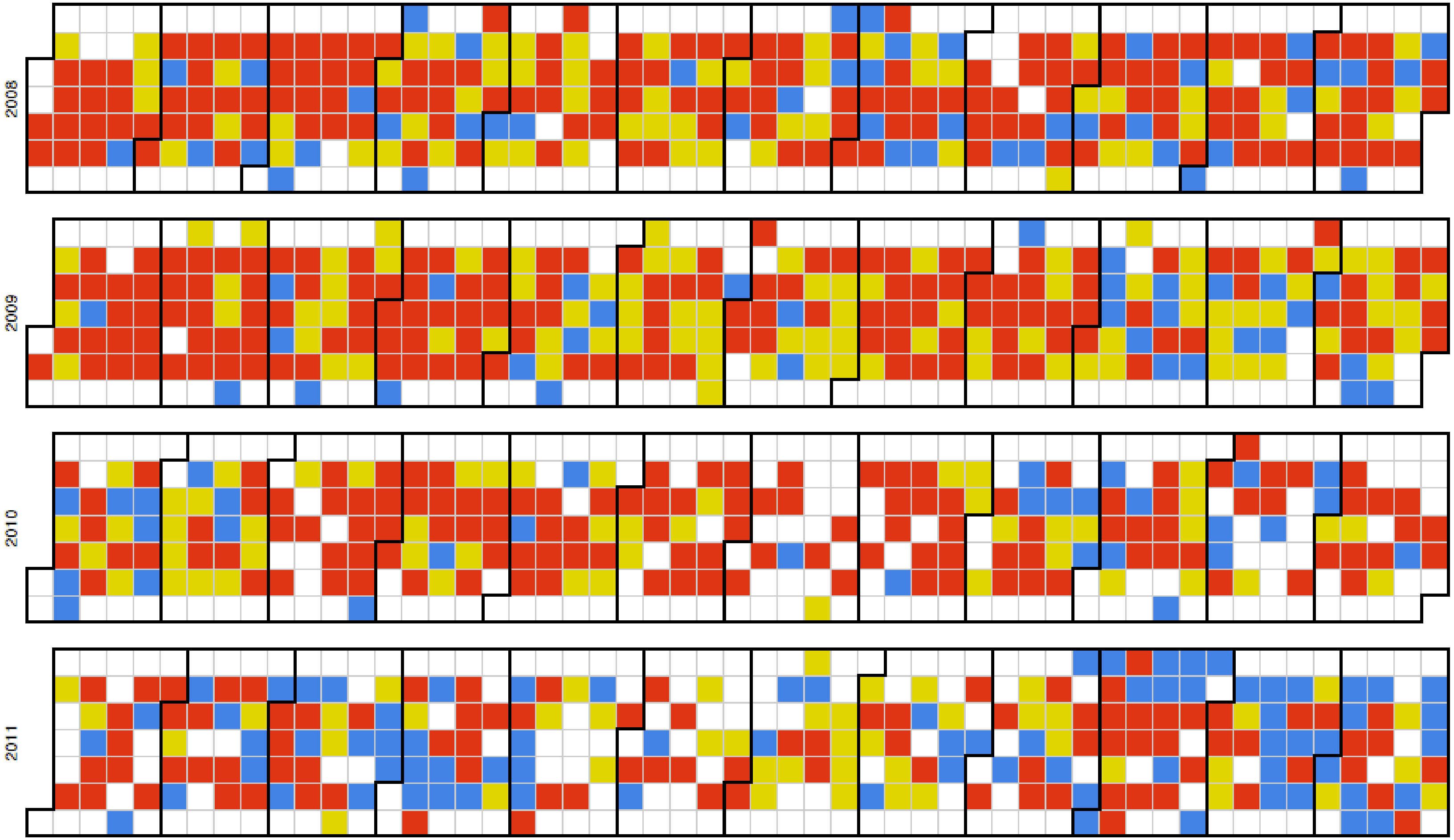
Better situation awareness and response strategy

Example of System Provided Intelligence: Malware Source

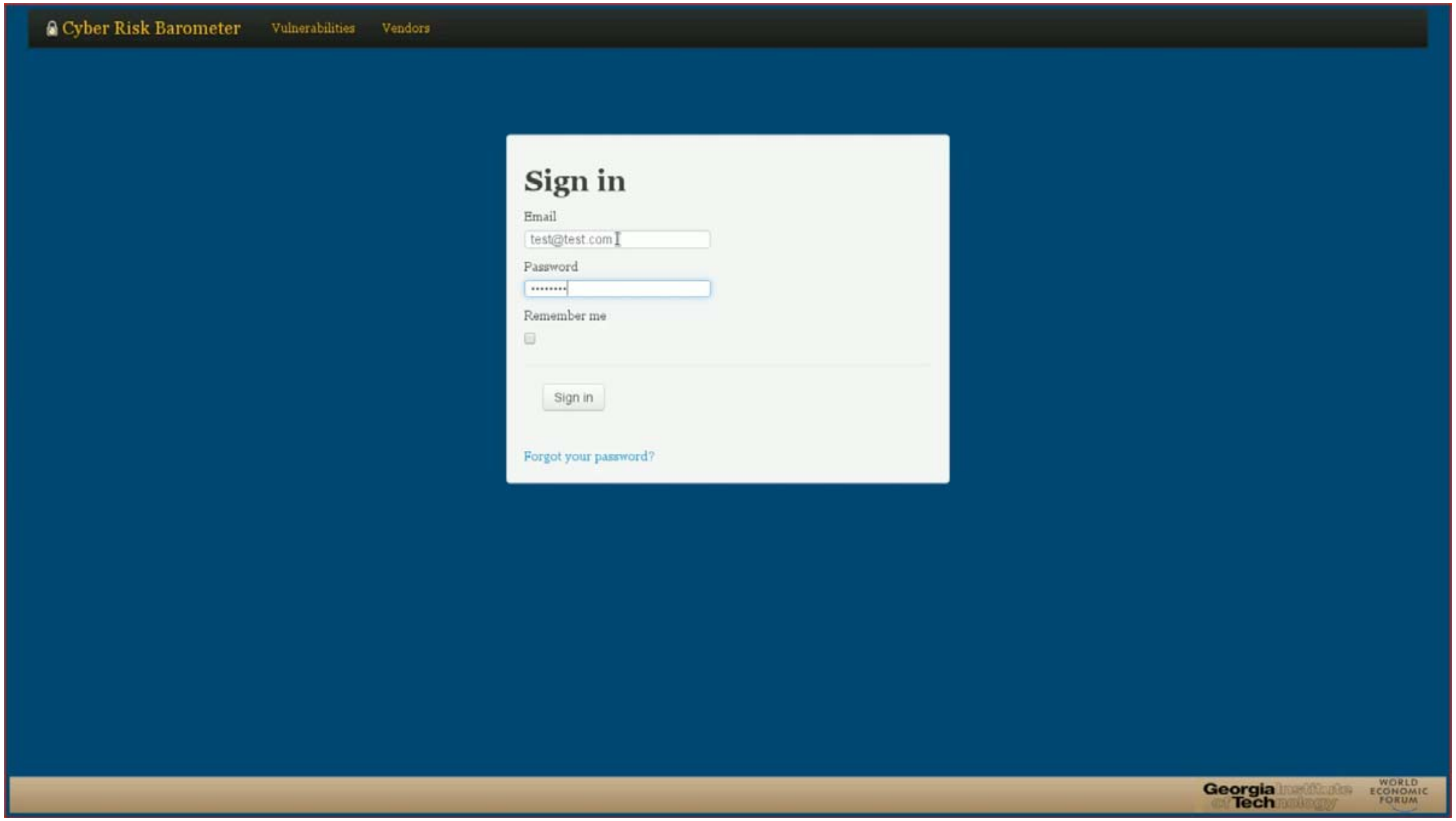


Source Country by Compilation Language (Hover over countries to see exact statistics)

Vulnerability Disclosure Calendar



Vulnerability Data Visualization Demo



Sign in

Email

test@test.com

Password

••••••••

Remember me

□

Potential Benefits

- Data-driven cyber risk assessment can enhance cyber resilience
  - Modeling attacks: Will we ever have be MTTA and MTTR for cyber attacks?
  - Predictive value: early attack warning & proactive response
  - Better intelligence about emerging threats and vulnerabilities
  - More effective human-in-the-loop decision making with analytics and visualization
- “CERT 2.0”
  - Real-time access to threat information

Planned Work: Threat Weather Reports

- Public vulnerability data collection and analysis
  - Calendar style visualization shows high level trends and allows drill down for deeper insights
  - Customization for given information technology profile (sector or organization specific)
- Malware Threat Intelligence
  - Drive-by-download risk by daily analysis of popular websites
- “Attempted attack” data visualization and and time-based trends