Visualization, Modeling and Predictive Analysis of Internet Attacks

Thermopylae Sciences + Technology, LLC

Administrative POC: Ms. Jeannine Feasel, jfeasel@t-sciences.com
Technical POC: George Romas, gromas@t-sciences.com

1400 14th Street North
Arlington, VA 22209
703-740-8768 (Main)
703-842-8599 (FAX)

Technical Content

Executive Summary

The current state of cyber security focuses a great deal on security information and event management (SIEM). Information gathering outside the firewall on potential attackers and attack vehicles is often not integrated with SIEM analysis. While SIEM analysis is valuable, it is at best only a portion of an overall integrated approach to cyber security that must include better tools to understand what is happening inside the firewall, combined and correlated with information from outside the firewall. This integration will provide far more effective models and response tools.

Thermopylae Sciences + Technology, LLC (TST) with subcontractor Recorded Future (RF) propose to use our proven, existing technologies as a basis to build predictive models to identify likely or ongoing attacks and coordinate appropriate response efforts. TST is a leading Google Enterprise Partner and has developed cutting edge geospatial and crowdsourcing applications that have saved lives and significantly enhanced the situational awareness and decision-making capabilities at the Department of State, US Southern Command (SOUTHCOM) and elsewhere. RF is an In-Q-Tel and Google Ventures funded company engaged in cutting edge research and development in predictive modeling. Both have robust management teams that can manage complex research and development efforts.

The team will use existing DHS data sets, as well as message traffic and web sites to build an integrated set of predictive models that help to predict likely sources and targets of cyber attacks. This model will then be validated using actual intrusion and attack data and web traffic. The goal is to develop the capability to correlate extensive structured and unstructured data and analyze it to identify emerging threats and predict likely attack venues. The second part of our effort will be to geo-tag the system and threat data and display it in a collaborative geospatial common operating picture. The goals are to clearly and intuitively show vulnerabilities, facilitate analyst collaboration and visualize attack parameters.

The significant problem areas this research and development proposal will address include:

- Prevention of cyber attacks by identifying emerging threats
- Rapid identification of imminent or ongoing attacks to support deterrence and rapid response
- Rapid analysis of infections and incursions to identify sources and attack vehicles in order to support rapid response and recovery

This application has direct commercial applications not only for cyber security, but also for fraud detection and other efforts. Given TST's and RF's commercial operations and capabilities, the design can readily extend to the civilian cyber security market space to help government and

civilian companies to collaborate and jointly work together, without compromising sensitive data, to defend the overall network.

Utility to Department of Homeland Security

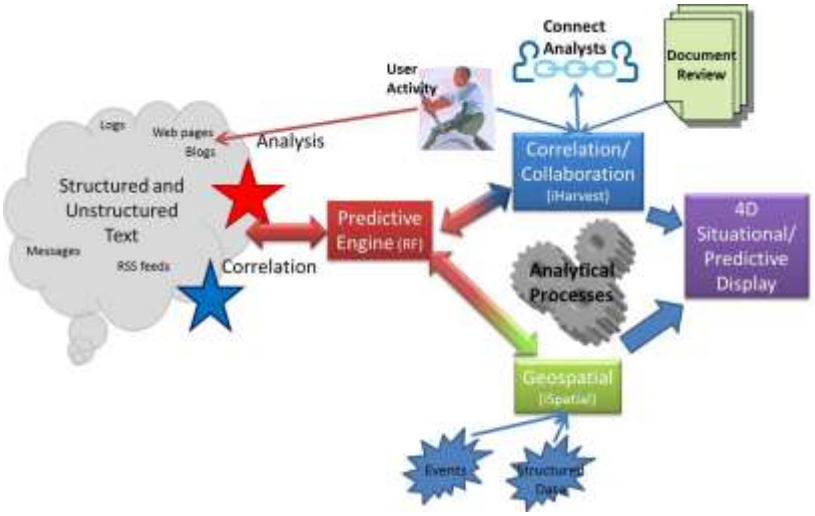
The integrated, predictive capabilities this research will produce will significantly enhance DHS's ability to defend against cyber attacks throughout both the government networks and civilian networks. The approach will significantly extend beyond the boundaries of the firewall and incorporate predictive analysis from websites, blogs, message traffic and other sources in cyberspace that may either indicate attacks in progress or pre-sage the development of attack vehicles and plans. The predictive models produced by the research will include data within the firewall and data outside the firewall to allow analysts to see what's happening within the network to identify potential targets and sources of attacks. The predictive results will be displayed in an integrated, configurable Common Operating Picture to give analysts and decision-makers the tools they need to make timely and effective decisions.

Technical Approach

Technical Concepts

TST will use a solid foundation of proven and rapidly emerging technology as the basis for a research and development endeavor that will build a truly integrated predictive capability to assist in deterring, preventing and mitigating the effects of cyber attacks.

The figure and charts below shows the overall technical approach.



Component	Status	Capabilities
Predictive engine	Recorded Future is used in security and commercial applications	• Predictive analysis • Modeling and temporal analysis of textual data • Identify emerging cyber threats and targets
Correlation/Collaboration	iHarvest® is used at INSCOM, SOUTHCOM	• Develop profiles of users' specific concepts and interests

	and other locations	• Match user profiles with data repositories to bring pertinent information back to users • Integrates search applications for minimal disruption • Alerts users of matching profiles, facilitating collaboration and awareness between interest groups • Identify potential suspicious user activity
Geospatial and Situational/Predictive Display	iSpatial® is in use in Department of State, SOUTHCOM and other locations	• Visualize and track data on a 3-D globe • Managing real-time information, intelligence feeds and mobile messaging devices directly from a Web-based user interface • View customized data layers of 3-D terrain, models and user graphics • Conduct geospatial analysis to determine cyber attack vectors and targets • Collaborate and share data instantly in real-time

Based on this framework of proven capabilities, TST and RF will conduct the following research:

Research	Capabilities Provided
Predictive attack modeling	• Research the basic building blocks of predictive modeling and analysis capabilities • Research malware and botnet identification on the web • Develop predictive model to identify likely attack vehicles, sources of attacks and targets
Predictive detection modeling	• Research data from attacked sites (logs, user activity, application activity) • Develop predictive prevention model to help identify attacks that are ongoing
Social networking analysis	• Research activity on system administrator forums, blog, and websites prior to attacks to see what the sys admins are discussing to see if they ask about issues and problems • Research activity on black hat forums, blog, websites and message intercepts prior to attacks to see what hackers are discussing • Build predictive model to emerging threats and new attack vehicles and targets
Analytical processes and collaboration	• Develop process that enhances situational awareness of malware and botnet activity • Develop effective capabilities to leverage existing tools to improve correlation and collaboration between both automated tools and human analysts • Develop capabilities to allow analysts to model and wargame potential attacks in order to test prevention and deterrence measures, as well as responses • Develop collaboration techniques to allow government and civilian researchers, analysts and security managers to share data and to collaborate without compromising sensitive/classified information.
IP address geospatial tagging and proxy piercing	• Research and develop mechanisms to geotag IP addresses • Research and develop mechanisms to pierce proxy addresses to trace back to original IP address • Develop geospatial displays that show attack vector origins
Predictive Situational awareness and collaboration	• Develop the feeds from predictive models into the situational display • Develop the ability to show the evolution of an attack temporally and geospatially (showing both attack sources and attack locations) • Develop the capability to coordinate and display prevention, deterrence, and response activities

Figure 2 Research

TST and RF both have strong R&D teams and reach back into academic institutions, such as MIT Computer Science and Artificial Intelligence Laboratory (CSAIL) and George Mason University Volgenau School of Engineering, as well as other partners, such as Draper Labs, that will facilitate the research. The team's combined capabilities will provide unique and effective research and development. TST devotes 10% or more of its budget to R&D and has clearly demonstrated its R&D capabilities through rapid development of capabilities for the Army G2 and the SOUTHCOM. RF has twenty personnel on staff, most of them PhD's devoted to R&D and is an In-Q-Tel and Google Ventures funded company. Both companies maintain strong R&D programs and have proven program management teams.

Uniqueness

- First, the approach includes the standard cyber security information that applications like CSET, SNORT and ArcSight consume and can incorporate these tools in a comprehensive framework. It incorporates data from websites, blogs and other message traffic to help identify trends, emerging attack vectors and monitor message boards for security professionals and hackers. This approach will provide both a comprehensive view of threats as well as new data feeds to build more effective predictive models.
- Integrated predictive models will use data from blogs, message traffic and other sources to predict potential attack mechanisms, targets and perpetrators.
  - Mining of social media will identify malware/botnet activity threads (i.e., it monitors and analyzes relevant cyber security activity on the Internet).
    - looks for posts/chatter on infection, propagation, destructive mechanisms, etc. of known and new malware/botnets
    - records originator and participants of posts/chatter
    - provides a new approach in malware and botnet detection and identification through natural language algorithms
  - Collaborative creation of a "signature library" of malware and botnet activity
    - aggregates posts/chatter of activity from Internet social media sources can potentially spot internal security threats through unintentional leaks on Internet our process can be applied to multiple data sources resulting in a comprehensive repository of data and metadata on specific malware and botnet activity (i.e., the signature or profile)
- The research will develop dynamic collaboration and correlation tools to synthesize data from disparate sources and to connect analysts together to help them share information and approaches. This value added information will dynamically feed the predictive models to help refine and validate them and make them more robust.

Required Material

TST will base the development on TST's existing iHarvest, iSpatial, and RF's Recorded Future platforms. The effort will also employ DHS S&T's DETER testbed facility and the PREDICT repository. The initial R&D efforts will be on TST and RF networks in their facilities. Once the

models are ready for more extensive testing and evaluation, the effort will move to the DETER test facility.

Personnel Qualifications and Experience

TST and RF both have a strong team with a wealth of cyber security, analytical, geospatial, and predictive modeling and managerial experience. Representative personnel include:

George Romas

George Romas is the Chief Technology Officer for Cyber at TST and has over 28 years' experience within the Intelligence Community, providing his strategic and tactical expertise across a wide range of technologies. George holds BS degrees in Computer Science and Economics from Union College and has done postgraduate work in the MS program for Computer Science at Virginia Tech. His experience includes modeling and simulation, systems programming, database development, systems engineering, cyber security, and enterprise architecture - with a strong focus on intelligence analysis solutions. Taking hiatus from the federal sector, Mr. Romas co-founded an Internet startup in July 1999 to develop a network-based multilevel security (MLS) appliance built on a trusted operating system. Prior to that, he worked at another startup that was developing control solutions for robotic manufacturing assembly lines. George holds an active TS/SCI security clearance.

Dr. William Ladd

Dr. Bill Ladd is Chief Analytic Officer at Recorded Future. Bill holds a PhD in Statistics from the University of Wisconsin-Madison and a BSE in Chemical Engineering from Princeton University. Before joining Recorded Future in 2010, Bill was Senior Vice President of Technology and Operations at Genstruct, the leading systems biology modeling company. At Genstruct, Bill led the development of an extensive knowledgebase derived from scientific literature and the application of artificial intelligence methods to this knowledgebase to address pharmaceutical research challenges. Prior to Genstruct, Bill was responsible for analytic strategy for Spotfire (now TIBCO) and Gene Logic, and was a statistical analyst at Merck. Bill consults with several high tech companies and has published numerous peer-reviewed papers. At Recorded Future, Bill provides the cornerstone for Recorded Future's analytic approaches in converting unstructured text data to actionable findings. Bill holds an active TS/SCI security clearance with full-scope polygraph.

Commercial Capabilities and Plan

All capabilities developed in this program will be immediately exportable for commercial application in not only the cyber security realm, but in many other realms as well. TST and RF both have commercial lines of business as well as our governmental lines of business that will be excellent vehicles to commercialize this research. Initial commercial applications include fraud detection, brand awareness and protection, and corporate security.

Costs, Work and Schedule

The R&D effort is planned as a Type II effort with a 24 month schedule. TST will develop a detailed milestone based schedule and cost breakout as part of a proposal. Milestones and costs will be tied to specific deliverables of predictive models and analytical tools.