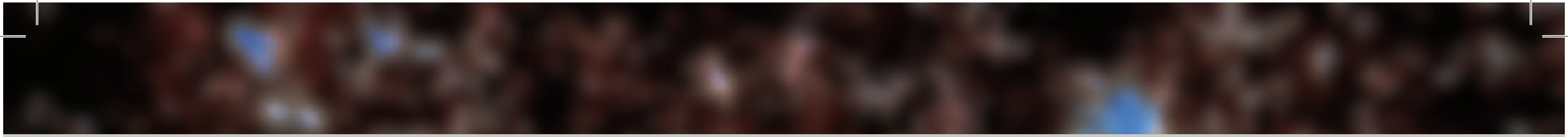


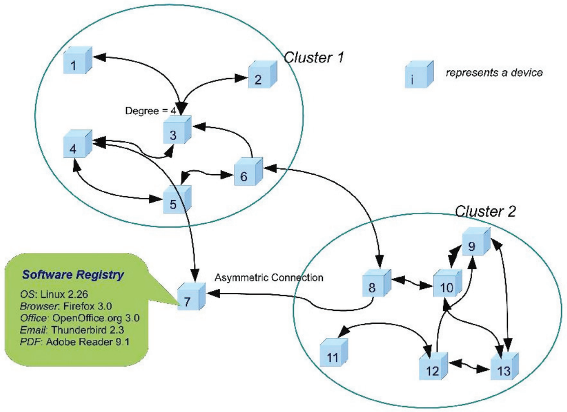
CyberSim: Geographic, Temporal, and Organizational Dynamics of Malware Propagation

Nandakishore Santhi, Guanhua Yan,
Stephan Eidenbenz, CCS-3

Cyber-infractions into a nation's strategic security envelope pose a constant and daunting challenge. We present the modular CyberSim tool, developed in response to the need for realistic simulations of software vulnerabilities and the resulting malware propagation in online social networks at a national level. CyberSim suite can (1) generate realistic scale-free networks from a database of geo-coordinated computers to closely model social networks arising from personal and business email contacts and online communities, (2) maintain, for each host, a list of installed software, along with the latest published vulnerabilities, (3) allow designation of initial nodes where malware gets introduced, (4) simulate the spread of malware exploiting a specific vulnerability, with packet delay and user online behavior models using distributed discrete event-driven technology, and (5) provide a graphical visualization of spread of infection, its severity, businesses affected, etc., to the analyst. We present sample simulations on a national-level network with millions of computers.

The combined real cost incurred by individuals, governments, and industries as a result of malware infecting otherwise productive computer networks is hard to estimate. The loss in productivity, perceived consumer lack of confidence due to explicit loss of privacy, irreplaceable loss of data, loss in hardware efficiency because of the malware running in the background, and costs associated with installing and maintaining anti-malware systems are all prominent factors. The direct cost of cyber crime to the US economy is in the tens of billions of dollars annually, and the federal government spends billions of dollars on cyber security each year [1]. Given the increasing intensity of cyber threats, the Obama administration has recently made cyber security a national security priority.

Fig. 1. A typical social network populated with software registries.



The extent and nature of malware infections on a large-scale computer network is decided by the specific software vulnerabilities that are exploited by malicious software. The malware spreads over a network by making effective use of the small diameter properties of social networks induced by email-contacts and various online communities: it is therefore crucial to include the specific nature of the social network in any serious study. A virtual controllable test bed is an indispensable tool to evaluate the effect of potential cyber threats and the effectiveness of proposed countermeasures.

When estimating the economic impact of an exploit, it is necessary to consider relatively large and complex computer networks with hundreds of millions of hardware hosts. As shown in Fig. 1, each host or device is in turn associated with myriad software applications which are installed on them and are routinely accessed by individual users. Further, an accurate model for the flow of packets in internet settings is called for. The rate at which a malware propagates over a network is further determined by the frequency and duration of online presence of a network user, as well as the usage rate of a particular vulnerable piece of software application.

The CyberSim suite was designed at LANL to address all these modeling and simulation challenges, while at the same time being flexible and scalable. CyberSim is designed to use the Common Vulnerabilities and Exposures (CVE) database maintained by the National Institute of Standards and Technology (NIST) National Vulnerabilities Database (NVD) [2], which is a reliable source of vulnerabilities and exploits of common software applications. The CyberSim tool is intended to be easy to use, providing an intuitive visualization interface. The toolkit includes a network generation module that takes various parameters as input; these are widely accepted to be representative of a social network model—scale-free nature induced through a fixed power-law degree distribution, and clustering based on geographic proximity of social contacts. The network generation module generates a social network according to the model parameters from a database of internet hosts with location information gathered from internet service provider surveys. Traditional tools for malware analysis typically lack the ability to model social networks in a realistic manner. The social network parameters taken as input by the network generation module are learned from real-world networks such as email contact lists and online

For more information contact Gary Sandine at gars@lanl.gov.

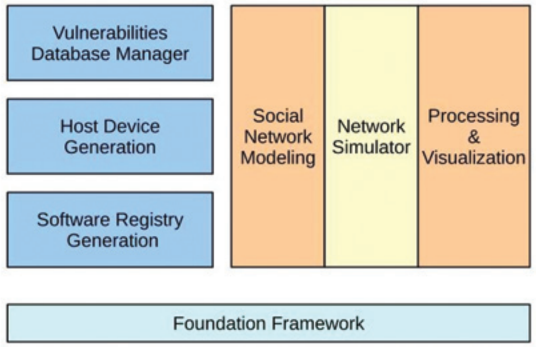

Vulnerabilities Database Manager
Host Device Generation
Software Registry Generation
Social Network Modeling
Network Simulator
Processing & Visualization
Foundation Framework

Fig. 2. The CyberSim modular architecture.

community graphs as a result of surveys. CyberSim has the ability to model various types of social networks such as personal email contact lists, business contact lists, and online communities. The toolkit has the ability to capture the inter-connections between several types of social networks at the same time.

CyberSim includes a highly distributed, scalable, discrete-event network simulation module built upon the LANL-developed SimCore framework [3]. This module allows the simulation of social networks with millions of networked devices on heterogeneous distributed platforms consisting of a grid of multi-core computers, each with limited physical memory. SimCore in turn relies on a simulation engine such as PrimeSSF/MPICH [4]. SimCore has been previously used in building complex simulation tools such as Multiscale Integrated Information and Telecommunications System (MIITS) [5] and ActivitySim [6].

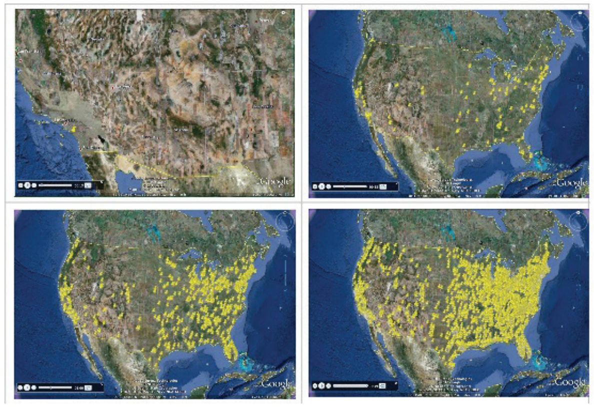


Fig. 3. Visualization of the spread of malware on a large-scale network.

The main design objective of CyberSim is to estimate the impact of software vulnerabilities on various business sectors, looking at its severity as well as rapidity. Realistic estimates of malware dynamics can be used to develop remedial actions. Figure 2 shows the layered architecture of CyberSim. The Foundation Framework is the core layer providing the support libraries for the other layers and modules. The Host Device Generation module populates a database of internet-enabled hardware hosts from data gathered through surveys. The Software Registry Generation module populates each hardware device with a software registry of installed operating systems, and internet applications such as media players, browsers, email-clients and office documentation software.

The Vulnerabilities Database Manager module is responsible for periodically updating the database of known vulnerabilities for the installed software on the hardware nodes. The Social Network Modeling module takes as input the databases maintained by the Host Device Generation, Software Registry Generation, and Vulnerabilities Manager modules in order to generate a realistic social network. The Network Simulation module uses a distributed message-passing simulator to simulate the spread of a malware over the social network model. The Processing and Visualization module generates detailed true-scale time-lapse location information on the malware spread and its impact on various businesses in a commonly supported markup language such as KML. The Social Network Modeling module of CyberSim is capable of modeling a wide variety of social networks by capturing the salient features of realistic social networks such as their scale-free nature and geographical proximity-based clustering.

Figure 3 shows the time-lapse distribution of infected devices during a CyberSim simulation case study involving millions of computer nodes with the real-world vulnerability in a popular office suite, which was exploited in the wild during August 2009 [7]. The CyberSim tool provides [8] the network security analyst a wide array of flexible options in the selection of social network types, distribution of social contacts and software, network parameters, and, finally, visualization.

[1] Pulliam, D. *Government Executive,* http://www.govexec.com/dailyfed/0305/031705p1.htm (2005).
[2] NIST 2010. National Vulnerabilities Database maintained by NIST, htttp://web.nvd.nist.gov (2010).
[3] Kroc, L., et al., "Simcore: Los Alamos National Laboratory Unclassified Technical Report," LA-UR: 07-0590 (2007).
[4] PRIME 2010. "PRIME: Parallel Real-time Immersive network Modeling Environment," https://www.primessf.net/bin/view/Public (2010).
[5] Waupotitsch, R., et al., "Multi-scale Integrated Information and Telecommunications System (MIITS): First Results from a Large-scale End-to-End Network Simulator," *in Proc 38th Winter Simulation Conf* (2006).
[6] Galli, E., et al., "ActivitySim: Large-scale Agent-based Activity Generation for Infrastructure Simulation," *in Proc 2009 Spring Simulation Multiconf* (2009).
[7] CVE 2009. Entry in Common Vulnerabilities and Exposures Candidate List, http://cve.mitre.org/cgi-bin/cvename.cgi?name=cve-2009-1136 (2009).
[8] Santhi, N., et al., "CyberSim: Geographic, Temporal, and Organizational Dynamics of Malware Propagation," *in Proc 2010 Winter Simulations Conf* (2010).

Funding Acknowledgments
Department of Homeland Security (DHS); National Infrastructure Simulation and Analysis Center (NISAC) Program