XU ZHANG

zhangxu1115@gmail.com

(505) 415-8103

Objective

Software Engineer (Machine Learning)

Education

Ph.D. Computer Science, Adviser: Jedidiah R. Crandall *May 2018*
University of New Mexico, Albuquerque, NM

M.S. Computer Science, Adviser: Jedidiah R. Crandall *Dec 2013*
University of New Mexico, Albuquerque, NM

M.E. Computer Engineering, Adviser: Chang Hoon, Kim *July 2011*
Daegu University, South Korea

B.S. Mathematics *July 2009*
Dalian University of Technology, China

Highlighted Project Experience

Spam Filtering System

- Built a spam filter which is able to classify emails into spam and non-spam email with high accuracy
- *Preprocessing*: Converted letters into lower case Punctuation and non-words were removed; words were reduced to their stemmed form; numbers, dollar signs, email addresses, URLs are normalized
- Built a vocabulary list by selecting words that occur least 100 times in the spam corpus
- Given an email, converted it to a vector of features x_i. (x_i means if the i-th word in the vocabulary is present in the email)
- Used SVM with Gaussian kernels to perform the classification.Training set contained 4000 examples, test set contained 1000 examples. Achieved 98.9% test accuracy

Malware Classification System

- Developed a malware classification system which is able to classify different types of exploits: stack overflow, heap overflow, and format string attacks
- Run the exploits and benign using Vector-based dynamic information flow tracking system
- Utilized octave to extract features of each matrix. Extracted features includes: the average value of each cell in the matrix, the 2 norm, the infinite norm, the sum of the eigenvector of the matrix times the transpose of itself, the minimum, maximum, and sum values of the eigenvector of the matrix, the trace of the matrix, the rank of the matrix, the percent of nonzero elements
- Feed the features to a python program(utilized pybrain), in order to create a neural network to classify programs. We used some machine learning techniques such as neural networks, SVM, Logistic Regression, among them, neural network works best
- Created a recurrent neural network with LSTM in the middle layer (200 nodes) and softmax function in the final layer

Off-path Network Latency Measurement

Off-path round-trip time (RTT) measurement has many potential applications, including: improved geolocation capabilities, measuring the performance of parts of the Internet where there is not much measurement infrastructure (e.g., PlanetLab), and providing data plane measurements to better understand global Internet routing. Off-path means that the measurement machine is not on the path being measured. More specifically, the technique I developed can measure the RTT between essentially any two machines (A and B) on the Internet without having special access to A or B or having any presence in the path between A and B. My technique uses a TCP/IP side channel called TCP SYN backlog. It is more robust to packet loss and more accurate across different RTT ranges compared to previous off-path RTT measurement techniques. Overall, 91.18% of our RTT measurement results are within 20% of the actual RTT.

Work Experience

Internship, *International Computer Science Institute (ICSI), Berkeley* *May 2017 - Aug 2017*

- Worked on circumvention tool testing. Based on the types of the circumvention traffic, I implemented two approaches to extract possible features. I extracted possible features from unencrypted TLS handshake headers if the circumvention tool used TLS to hide its content. If the circumvention traffic was Non-TLS, I extracted raw payload from TCP layer and build data matrix by using the designed algorithm.

Research Assistant, *University of New Mexico, Albuquerque* *Fall 2012 - Spring 2018*

- Invented a technique to test off-path trust relationship in network layer 3
- Invented a technique to do IPv4 and IPv6 Alias Resolution
- Improved the fidelity of previous Off-Path latency measurement technique using TCP/IP side channels
- Invented a technique to find machines hidden behind firewalls
- Developed a new RST attack to evade IDS

Projects Developed

- Built a movie recommender system by using collaborative filtering learning algorithm
- Wrote K-means clustering and PCA (principal component analysis) to compress image
- Implemented one-vs-all logistic regression as well as neural networks to recognize hand-written digits
- Implemented anomaly detection to find failing servers on a network
- Wrote AI controller for game MS Pacman that took best actions
- Designed Genetic Algorithm to evolve compete & corporate ant colonies
- Wrote compiler in python and performed data-flow analysis
- Compared the performance of TCP congestion control algorithms in BGP connected virtual network
- Reversed engineering malware from the AURIGA malware family
- Reversed engineering windows mine sweeper and added backdoor to win easily
- Helped to develop and update multi-player, educational cybersecurity game, website: http://werewolves.cs.unm.edu/
- Wrote administrator control function for website http://www.siba.sg/

Leadership

- Volunteer lecturer for computer networking class
- Reviewer for journal *IEEE/ACM Transactions on Networking*
- Volunteer lecturer on GenCyber Summer Camp 2015 about Network RE and Password Cracking
- Participated in Tracer Fire IV Capture the Flag Competition

Skills

Programming: Python, Java, C, Bash, Octave, Matlab, Haskell

Environment and Tools: VirtualBox, VMWare, TensorFlow, IDA Pro, OllyDbg, Git, SVN, Vim

Networking Skills: Scapy, Bro, Wireshark, Tcpdump, Tcpflow, Hexdump, Traceroute, Nmap, P0f, PF_RING, Snort, Hping, Farpd, Netcat

Publications

Xu Zhang, Jeffrey Knockel, and Jedidiah R. Crandall. "**ONIS: Inferring TCP/IP-based Trust Relationships Completely Off-Path**," *Proceedings of IEEE INFOCOM* 2018.

Xu Zhang, Jeffrey Knockel, and Jedidiah R. Crandall. "**High Fidelity Off-Path Round-Trip Time Measurement via TCP/IP Side Channels with Duplicate SYNs**," in the *Proceedings of IEEE GLOBECOM* 2016.

Xu Zhang, Jeffrey Knockel, and Jedidiah R. Crandall. "**Original SYN: Finding Machines Hidden Behind Firewalls**," in the *Proceedings of IEEE INFOCOM* 2015.