An Interview of Daphne and Naren by *Comm. of ACM*
on Program Anomaly Detection

1-Your method only seems to work on programs whose behavior with which you have become very familiar. Is that true?

Response: Yes, that is correct. But this is a reasonable assumption since we are seeking to detect anomalies. Virtually all program anomaly detection approaches need to build models to represent *expected* program behaviors, by observing a program's execution traces and/or analyzing executables. In our ACM CCS '15 work, we construct such a behavioral model through data mining and learning methods on function and system call traces.

If so, how could it be used to foil hackers in general?

Response: The advantage of program anomaly detection techniques is their potential to detect new attacks and anomalies, compared to conventional signature-based intrusion detection (such as an anti-virus scan). The capability of detection is not limited to recognizing known attacks. For hackers to accomplish their attack goals, they inevitably need to alter aspects of the program execution. One needs to compare the observed runtime execution traces with the behavioral models inferred about the program.

Accuracy and efficiency are the two main factors in determining the quality of the detection. Accuracy refers to how well the model can distinguish between normal and attack traces. It is usually measured by the false positive rate (i.e., false alerts) and the false negative rate (i.e., missed detections). False negatives are obviously troublesome but false alerts also lead to costly human effort, as each alert has to be manually inspected and investigated by security analysts. Efficiency refers to all performance aspects of the detection, including space and runtime complexities of the model, training and detection algorithms.

Program anomaly detection offers flexibility in enforcing dynamic program behaviors, thus providing necessary complementary security protection against zero-day exploits and advanced persistent threats. Like all other security solutions, program anomaly detection is not perfect. The success of the detection is with respect to an attack model, which describes the adversary's capabilities.

2-How would you compare your techniques to others which also analyze anomalous program behaviors?

Response: Since Dorothy Denning's 1987 seminal paper on anomaly detection, many research solutions have been proposed to secure complex programs. Most of them build precise models to represent normal control flows and data flows of a program. In comparison, our work considers a much stronger adversarial model. Modern stealth attacks do not necessarily alter the control-flow or data-dependence properties of a program. Our attack model assumes that the adversary does not introduce conflicts with control-flow graph, anomalous call arguments, or unknown calls. One may wonder what hackers may accomplish under this restricted attack model. Surprisingly, many real-world attacks belong to this category, including service abuse attacks, denial-of-service attack, memory exploits such as heap feng shui, and workflow violations, e.g., bypassing access control.

Our method characterizes numerical and semantic properties associated with call events, specifically correlation patterns in long call sequences, intervals between calls and call frequencies. We

immediately faced a few technical challenges. Modern programs have diverse normal behaviors. How do we recognize all or most normal correlation patterns? Straightforward application of novelty detection or outlier detection techniques such as 1-class SVM results in unacceptably high false positives. For some programs, we found that normal call sequences are more different from each other than from anomalous traces! In addition, long traces may consist of tens of thousands of system calls. How do we efficiently represent them in training and detection?

For the diversity challenge, our approach categorizes normal behaviors into clusters. Through clustering, our model recognizes and supports many different types of normal execution patterns. Anomaly detection is performed within the boundary of the cluster. We found that this two-stage detection technique brings orders of magnitude reduction in false positive rates. In addition, we designed boolean-matrix and integer-matrix based data structures to compactly represent call correlation and frequency properties.

Our experimental evaluation showed that our prototype is effective against three types of real-world attacks and four categories of synthetic anomalies with less than 0.01% false positive rates and 0.1~1.3ms analysis overhead per behavior instance. (Each behavior instance has 1K to 50K function calls or system calls).

Over the years through our interactions with many security researchers in both academia and industry, we see great needs in designing and developing practical program and network anomaly detection solutions. In our other work that appeared in ACM ASIACCS '14 (H. Zhang, D. Yao and N. Ramakrishnan. Detection of Stealthy Malware Activities with Traffic Causality and Scalable Triggering Relation Discovery), we demonstrated a causality-based network anomaly detection approach. We hope the principles and lessons learned from our work can inspire and encourage others in their design.

3-What is the next step in your computer security work?

Response: **(Daphne)** I am passionate about designing and developing security methodologies that are innovative, useful, and most importantly have long-lasting and far-reaching impact in the cyberspace. I look up to the innovative work of female computer scientists, Stephanie Forrest, Elisa Bertino, Barbara Ryder, Klara Nahrstedt, Rebecca Wright and Anna Lysyanskaya, just to name a few.

For my next step, I am collaborating with ECE researchers on securing Internet of Things and smart city infrastructures, specifically on designing lightweight and tamper-resistant program anomaly detection systems. With the recent White House's Smart Cities Initiative, I would very much like to see that the past three decades of research and development into program anomaly detection finally benefits society.

(Naren) We are interested in developing data analytics approaches to not just detect cyber-attacks but also to forecast them. There are many signals “in the wild” that precede real attacks and our goal is to harness them with a view toward forecasting attacks against specific organizations, sectors, or websites.