New tool could reduce security analysts' workloads by automating data triage

August 2 2019, by Jessica Hallman

Credit: CC0 Public Domain

During a cyberattack, security analysts focus on answering four key questions: what happened to the network, what was the impact, why did it happen, and what should be done? And while analysts utilize

advancements in software and hardware tools in their response, the tools are unable to answer these questions as well as humans can.

Now, researchers at Penn State and the U.S. Army Research Office have developed a technique that could significantly improve the performance of security analysts. Their tool, a finite state machine—a computational model that can be used to simulate sequential logic—was constructed to conduct automatic data triage of repetitive tasks that analysts regularly handle.

"Substantial amounts of analysis work are repeatedly done by human analysts," said Peng Liu, Raymond G. Tronzo, M.D. Professor of Cybersecurity in Penn State's College of Information Sciences and Technology and investigator on the project. "If an intelligent agent can help do the repeated work, then the analysts can spend more time dealing with previously unseen cyberattack situations."

"Cyber defense is always challenging due to the fact that adversaries always try their best to hide their acts among a huge amount of normal activities," added Cliff Wang, division chief, Computing Sciences, Army Research Office, an element of Combat Capabilities Development Command's Army Research Laboratory. "As cybersecurity becomes an increasingly important to Army operations, the ability to detect and analyze obscure and abnormal behavior is essential, especially during the early reconnaissance stage."

According to Liu and his collaborators, a time-consuming stage in cyber analytics is data triage, which involves an analyst examining the details of various data sources such as intrusion system alerts and firewall logs, weeding out false positives, and then grouping the related indicators so that different attack campaigns can be separated from one another. Their research aims to reduce the analysts' workloads by automating this process.

In their study, the researchers traced 394 data triage operations of 29 professional security analysts. Then, they utilized the finite state machines to recognize attack-path patterns in more than 23 million firewall log entries and more than 35,000 intrusion alerts collected from a 48-hour monitoring of a network with 5,000 hosts.

"Identifying attack paths in multiple heterogeneous data sources is a repetitive task for security analysts if the same type of attack path was analyzed before, and such repetitive tasks are often very time-consuming," said Liu. "Additionally, our interviews with security analysts revealed that they could be substantially affected by fatigue caused by analyzing a huge number of security-related events, and anxiety caused by time pressure."

The technique combines non-intrusive tracing of human data-triage operations, formal constraint graphs, and data mining of operation traces, and leverages principles in both computer science and cognitive science. The finite state machines are "mined" out of operation traces.

"Penn State researchers have been leading research efforts in applying statistical methods, artificial intelligence and machine learning to identify hard-to-find, low-level intrusion activities, and advanced the state in this field," said Wang.

The research, "Learning from Experts' Experience: Toward Automated Cyber Security Data Triage," was funded by the U.S. Army Research Office and published in the March 2019 issue of *IEEE Systems Journal*.

More information: Chen Zhong et al. Learning From Experts' Experience: Toward Automated Cyber Security Data Triage, *IEEE Systems Journal* (2018). DOI: 10.1109/JSYST.2018.2828832