Performant Reasoning Systems for Cybersecurity in Data Center Computing

Nazario, Jose
jose.nazario@soartech.com

Sypniewski, Adam
adam.sypniewski@soartech.com

Schmorrow, Dylan
dylan.schmorrow@soartech.com

May 4, 2015

Abstract

Cybersecurity is largely driven by the "capture everything, and try to analyze it" approach, which has seen attempts at bringing big data to security. However, data center computing has traditionally been forced to choose between consuming overhead that could otherwise generate revenue and security. A new paradigm must be achieved to address this balance, which we feel can be met with abductive reasoning and agent-based security.

1 Discussion

SoarTech has dedicated work that focuses on long term research and development of intelligent systems for cybersecurity that highlight the capability to deliver trustworthy simulations and data analytics. We have ongoing research efforts that we believe have identified key scientific research opportunities relevant to the DOE Office of Science and would like to share recent findings and work with a coalition of academia, industry and government researchers to identify scientific research principles that (1) enable the trusted foundation for high performance supercomputing; (2) enable the application of extreme scale knowledge and analytics to address cybersecurity challenges; and (3) support a trustworthy data center ecosystem where reliably results are paramount.

Data center computing resources have made high performance, distributed computing accessible to a wide public. With this have come the risks of open computing, including abuse, theft of resources, malicious computation, and attacks on other tenants within the data center. Data center computing operators must therefore monitor their equipment to inspect for malicious activity, but this comes with a potentially high overhead that affects performance and fee-paying capacity. Modern cybersecurity monitoring has been built on the premise of recording as much as possible to facilitate any possible future investigation. This "capture everything, everywhere" approach is the root cause of the possible performance impact for secure data center operators.

One possible approach to address this tension between security and performance is to abandon this wholesale data collection effort and instead focus on applying reasoning to detect signs of abuse, specifically abductive reasoning. To accomplish this, an agent-based cybersecurity monitoring system would reason over selective data sets in an effort to prove candidate abuse models. In this scenario, it would be the responsibility of any specific agent that suspects an attack to gather evidence and support the hypothesis. SoarTech has been able to apply abductive reasoning to image recognition and in a past evidence-marshaling project, demonstrating the applicability of the Soar cognitive architecture the ab-

ductive reasoning. It is reasonable to expect that this approach could be applied to the cybersecurity domain. Past efforts at scaling Soar to massive data sets in cybersecurity include our Dynamic Understanding of DNS Events (DUDE) project for ODNI, where a community of diagnostic software agents inhabits a graph representing the URLs (which in general represent external sites) and IP addresses (usually corresponding to internal machines) in the DNS records. Then dynamics detection agents examine the time patterns of connections to these sites, seeking evidence of command and control activity and related download and exfiltration events. These agent activities are all localized in the graph representation, allowing the graph (and thus the agent processing) to be distributed across many processors.

We believe that lessons we are learning in these activities coupled with our long history in applying artificial intelligence technologies to related domains that we are uniquely positioned to support the design of built-in security mechanisms that will ensure trustworthy operation, usage, and results from extreme scale simulations and experiments that can be applied to a range of applications important to DOE such as securing high performance scientific data center computing at extreme scales.

2 About the Author

Dr. Jose Nazario is a Senior Scientist at Soar Technology leading cybersecurity research efforts. Prior to Soar Technology, Dr. Nazario was Chief Scientist at Invincea Labs, where he was PI for DARPA efforts including DARPA ICAS and PlanX, and led research into global cybersecurity metrics for JPCERT in the Cyber Green program. Prior to Invincea, Dr. Nazario managed the security research and analysis team ASERT at Arbor Networks. He has published research papers in forums including IEEE, USENIX, and ACM. This work has given Dr. Nazario a deep understanding and experience in developing approaches and technologies in a broad range of cybersecurity areas. Dr. Nazario also holds board member and/or technical advisory positions for the Honeynet Project and the Cyber Conflict Studies Association. Dr. Nazario earned his Ph.D. from Case Western Reserve University.