Adaptive Autonomous Agent Responses to Targeted Malware Attacks

Steven Noel
The MITRE Corporation
McLean, Virginia
USA

snoel@mitre.org

Arun Lakhotia
Cythereal, LLC
Lafayette, Louisiana
USA

arun@cythereal.com

ABSTRACT

We describe the application of machine learning and data mining techniques for defensive autonomous agent responses to targeted cyberattacks. This approach clusters and classifies captured enterprise malware, and fuses the inferred classes with other relevant threat data to detect targeted attacks indicative of malware delivery attempts by advanced persistent adversaries. Defensive autonomous agents, whose behaviors are learned through specialized process modeling algorithms, are then improved through our enhanced situational knowledge of targeted attacks. The autonomous agents are guided by high-level process models, with which human operators interact for orchestrating lower-level autonomous agents. Agent responses are focused on protecting critical cyber assets, leveraging knowledge of potential paths of exploitation through the network. We employ agent-based simulation for rapid testing and refinement of process orchestration and agent behaviors.

1.0 RESEARCH CHALLENGES

Organizations undergo continual attacks from a range of threat actors with varying capabilities and intent. When defenders are better able to understand their adversaries, they are better able to respond. Indeed, it has been recognized that incident response would greatly benefit from improved capabilities for malware analysis [1].

Among the most serious threats are advanced adversaries who are targeting a specific organization. Attackers morph their malware to help evade detection by anti-malware systems, and target different individuals in the organization with different malware delivery methods. Since malware campaigns are tracked using hashes computed using byte code of malware, morphing of the malware also makes it difficult to recognize targeted campaigns. Security operations suffer from largely manual processes for correlating attack indicators, making it difficult to keep pace with adversary activities.

There are numerous open problems in recognizing targeted malware attacks and mounting adaptive autonomous responses against them. While advanced methods exist for computing malware semantic similarity, malware is readily shared through cybercrime industry, so that features derived from other malware artifacts and context are needed for distinguishing among adversaries. Knowledge about defensive responses by human operators needs to be captured by autonomous agents, and continually updated as defensive processes improve; operators also need to interrogate and orchestrate autonomous agents when needed.

Ideally, attack responses should ideally be guided by paths of potential adversary movement through the network, and focused on protecting critical cyber assets. Overall, this problem involves rich webs of interrelated data, which requires a flexible and manageable knowledge base to be maintained and shared among defensive agents. There are also scalability issues, since the space of malware is large and their correlations scale quadratically, as do other kinds of relationships such as potential adversary paths.

2.0 APPROACH

We propose an intelligent autonomous agent architecture for recognizing and responding to targeted malware attacks. Through unsupervised learning on malware features, this discover clusters of related malware indicative of targeted attacks, and then classifies those clusters according to known adversaries. For this, we can leverage our previous work in symbolic interpretation to extract generalized semantics from binaries, for fast similarity matching against large malware repositories [2]. These malware inferences can be fused with other threat information (delivery mechanisms, social engineering employed, known threat actor behaviors, etc.), for more accurate and fine-grained classification of adversary campaigns. This enhanced situational awareness can guide the responses of our defensive autonomous agents, e.g., applying similar responses for similar attacks.

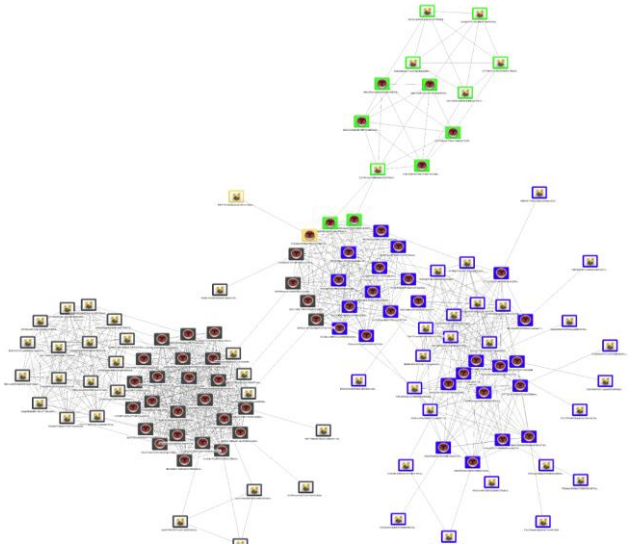

Figure 1: Clusters of Related Malware [10].

We propose to automate the learning of response behaviors through process mining [3]. This extracts hierarchical Markov models from cyber defender event logs, for learning patterns in defender operational processes (e.g., adversary hunters). We then map the discovered lower-level processes to autonomous agents, which communicate with high-level process models for organizational vetting and agent orchestration. The autonomous agents are informed by our knowledge of targeted attacks. Through Monte Carlo simulation and machine learning, the agent models are trained to adapt to different targeted attack situations.

Through machine learning of optimal processes, we thus adapt the autonomous agents to best respond to targeted attacks. We define orchestration processes that capture the high-level flow of an organization's security operations. An agent-based simulation framework then simulates attacker and defender agents, which generate simulation event logs for further iterations of process refinement. The agent responses are guided by an understanding of potentially exploitable paths through the enterprise network [4], as well as historical patterns of communication among mission-critical cyber assets [5]. For knowledge management and situational awareness within this complex web of interrelated information, we leverage our previous work in representing such interrelationships as a knowledge graph [6], with flexible schema-free design and ad hoc query-based analysis and visualization.

Figure 2 shows our architecture for responding to targeted malware attacks via autonomous agents. Agents deployed on enterprise endpoint nodes and gateways monitor and detect malware attempts (phishing emails, malicious web sites, etc.). Detected malware are sent to a "malware jail" for quarantine and dynamic behavioral observation and analysis. Malware are unpacked and subjected to deep semantic analysis of binary code [7], which extracts generalized semantics for machine learning features. From these features (along with semantic features from trusted partners sharing malware intelligence), an intelligent response orchestrator compares malware instances with historical archives, combines malware similarity inferences (e.g., cluster matches) with other relevant information (e.g., mined rules for response processes, enterprise security posture, mission dependencies, and external threat intelligence) for deciding optimal responses. The response orchestrator then pushes response decisions to agents deployed on network hosts.

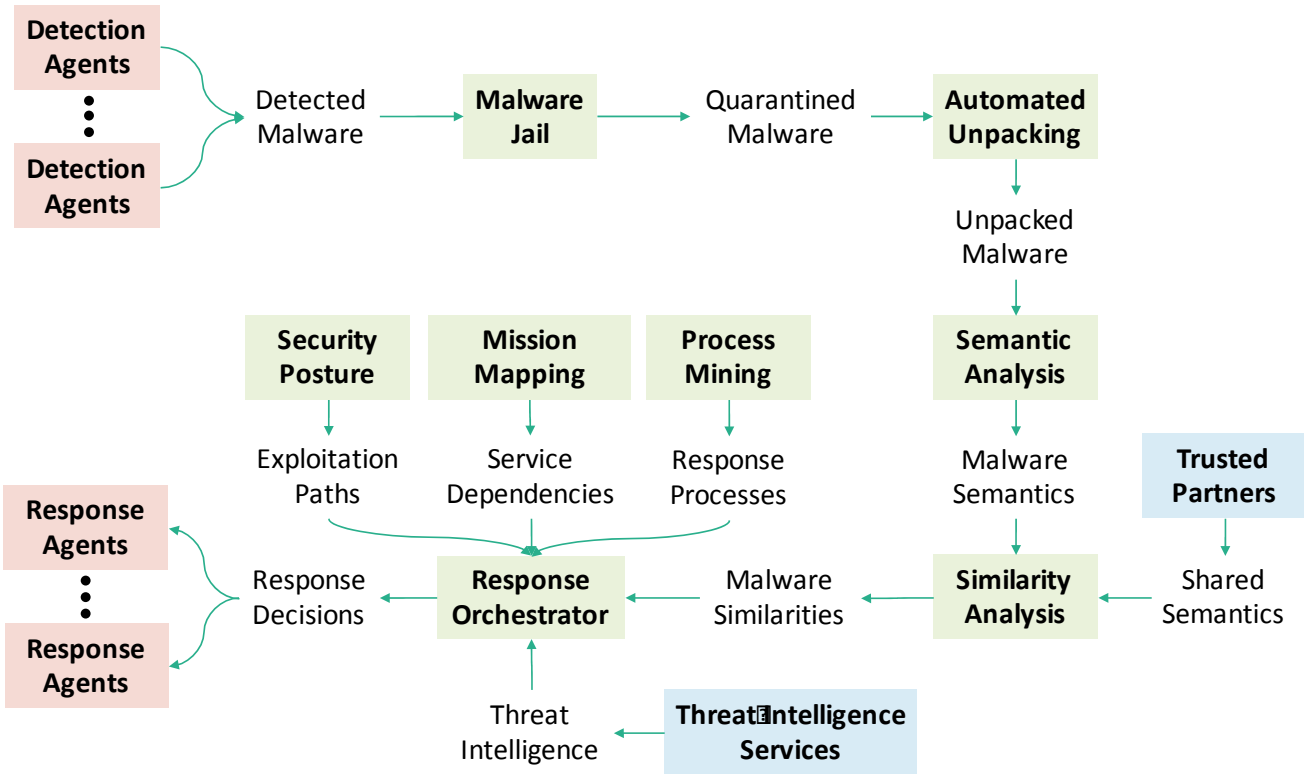

Figure 2: System Architecture for Adaptive Autonomous Agent Responses to Targeted Malware Attacks.

3.0 PREVIOUS WORK

Previous work under the DARPA Cyber Genome Program has yielded a fast and robust capability for malware analysis and attribution based on "genomic correlation" [7] [8] [9]. That capability has subsequently been extended for mining relationships over numerous malware artifact types, including code, code semantics, dynamic behaviors, malware metadata, distribution sites, and e-mails [2], and is available commercially as Cythereal MAGIC [10]. We leverage this mature capability for unpacking malware, extracting semantic "juice" (generalized semantics), and for performing similarity analysis over large malware corpuses.

Some cybersecurity tool vendors have begun incorporating machine learning techniques for malware detection. Such tools perform binary classification, in which malware-related events are classified as either "good" or "bad," e.g., based on surface-level (bytecode and file structure) analysis or behavioural analysis from log data. These capabilities are inadequate for our purposes. Beyond classification (supervised learning), we require unsupervised clustering for detecting patterns of coordinated malware attacks, through using features that characterize semantic invariants of related malware. This is exactly what Cythereal MAGIC provides, through a combination of dynamic analysis to unpack malware and static analysis to extract features and compute malware similarities.

There are many sources available (both commercial and open source) for shared threat intelligence that can enhance the discriminatory power of detecting and classifying targeted attacks. The area of process modeling is well established, including tools for process mining algorithms and standardized languages for expressing process mining inputs (event logs). Agent-based simulation frameworks are also available, as well as process modeling and simulation tools.

4.0 IMPACT

This line of research addresses many challenges in cyber defensive operations, in terms of enhanced situational awareness of targeted attacks by advanced adversaries, autonomous agents for rapid adaptive attack response, and high-level orchestration of low-level autonomous agents. In combining these aspects, the value of this synergistic solution is greater than the sum of its parts.

5.0 REFERENCES

[1] L. Pingree, MacDonald and Neil, "Best Practices for Mitigating Advanced Persistent Threats," Gartner Report G00224682, 2012.

[2] C. Miles, A. Lakhotia, C. LeDoux, A. Newsom and V. Notani, "VirusBattle: State-of-the-Art Malware Analysis for Better Cyber Threat Intelligence," in *7th IEEE International Symposium on Resilient Control Systems*, 2014.

[3] F. Szimanski, C. Ralha, G. Wagner and D. Ferreira, "Improving Business Process Models with Agent-based Simulation and Process Mining," in *Enterprise, Business-Process and Information Systems Modeling*, Springer, 2013, pp. 124-138.

[4] S. Noel, E. Harley, K. H. Tam and G. Gyor, "Big-Data Architecture for Cyber Attack Graphs: Representing Security Relationships in NoSQL Graph Databases," in *IEEE Symposium on Technologies for Homeland Security (HST)*, Boston, Massachusetts, 2015.

[5] S. Musman, "Automagical Cyber Dependency Mapping," The MITRE Corporation.

[6] S. Noel, E. Harley, K. H. Tam, M. Limiero and M. Share, "CyGraph: Graph-Based Analytics and Visualization for Cybersecurity," in *Cognitive Computing: Theory and Applications, Handbook of Statistics 35*, Elsevier, 2016.

[7] A. Lakhotia, M. Preda and R. Giacobazzi, "Fast Location of Similar Code Fragments using Semantic 'Juice'," in *2nd ACM SIGPLAN Program Protection and Reverse Engineering Workshop*, 2013.

[8] M. Preda, R. Giacobazzi, A. Lakhotia and I. Mastroeni, "Abstract Symbolic Automata: Mixed Syntactic/Semantic Similarity Analysis of Executables," *ACM SIGPLAN Notices*, vol. 50, no. 1, pp. 329-341, 2015.

[9] A. Pfeffer, C. Call, J. Chamberlain, L. Kellogg, J. Ouellette, T. Patten, G. Zacharias, A. Lakhotia, S. Golconda, J. Bay, R. Hall and D. Scofield, "Malware Analysis and Attribution using Genetic Information," in *7th IEEE International Conference on Malicious and Unwanted Software*, 2012.

[10] Cythereal, "Changing the Rules of Cyber Engagement," [Online]. Available: http://www.cythereal.com. [Accessed 5 June 2017].