Identification and interpretation of anomalous system behavior through kernel event analysis

Sebastian Schrittwieser, Robert Luh, Stefan Marschalek

Josef Ressel Center for Unified Threat Intelligence
on Targeted Attacks (JRC TARGET), St. Pölten UAS

Josef Ressel Center TARGET

Funding Agency	Christian Doppler Forschungsgesellschaft
Funding Period:	April 2015 – March 2020
Industry Partner:	CyberTrap GmbH

Introduction

Driven by its primary goal of detecting unknown targeted attacks on IT infrastructures, the JRC TARGET has built a solid foundation in data science, malware analysis, and digital forensics. Since 2015, we have conducted comprehensive analyses of targeted cyber-attacks and existing detection as well as analysis methodologies, putting a strong focus on behavior-based concepts.

With a solid research infrastructure in place, we are able to collect bulk operating system event data. This data, which is collected on numerous computer workstations as well as in a dedicated malware lab, is processed and analyzed in a multitude of ways. We generally focus on capturing, formalizing, and analyzing operating system events subsequently used to classify and interpret (anomalous) application behavior. Research methodologies include text-based scoring of process tree structures, numeric and string-based clustering, grammar inference, sentiment mining, classification using Markov chains, and the extraction and matching of star-shaped graph structures.

Research Areas

The ultimate goal of the JRC is the research of APT-aware threat detection methodologies for IT infrastructures. We explored different methods for large-scale system event data collection, its preprocessing and reduction, clustering efforts based on sequential text corpora, as well as semantic event analysis based on sentiment mining and graph-based matching methods.

In the course of our work, we identified many data providers (collectors), correlation solutions, ontologies and languages, classification systems and other tools that can contribute to the semantics-aware recognition and mitigation of attacks.

We ultimately decided to focus on a few highly relevant topics, including the ones exemplified below.

Formalization, modeling and semantic web

Our APT ontology (TAON) combines actors, assets, specific attack objectives, as well as attack stages into a single, versatile model for planning an organization's defense against targeted attacks. TAON, an OWL-based ontology realized in Protégé, also facilitates the development of novel behavior-based detection systems by mapping attack specifics to individual abstracted events and anomalies that can be detected by today's monitoring tools. This helps analysts to understand how, why, and by whom certain resources are targeted. Populated by concrete data, the proposed ontology can become a smart correlation framework able to combine several data sources into a semantic assessment of any targeted attack – hypothetic and ongoing.

Data classification

For data classification, we developed a sentiment extraction and scoring system capable of learning the maliciousness inherent to n-grams of kernel events captured by a real-time monitoring agent. The approach is based on calculating the log likelihood ratio (LLR) of all identified n-grams, effectively determining neighboring sequences as well as assessing whether certain event combinations incline towards the benign or malicious. The extraction component automatically compiles a WordNet-like sentiment dictionary of events, which is subsequently used to score unknown traces of either individual processes, or a session in its entirety.

A lot of our work revolves around event propagation trees. These trees are representative for the behavior exhibited by computer programs. In an alternative classification approach, we use a moderately modified version of Markov chains to create a distance matrix based on the discretized behavioral profiles, which is subsequently used for clustering. Our evaluation results show that the Markov chain approach can be used to reliably classify arbitrary processes and helps identify potentially harmful outliers.

Anomaly detection and explication

For the automation of said event-to-activity mapping we included grammar inference. Methodologies based on inferring rules instead of relying on a manual definition of patterns are powerful knowledge generation tools. For the prototypical implementation of the system, we created a Sequitur-based inference and assessment mechanism that automatically extracts grammar rules describing potentially interesting patterns seen across several traces of a semantically comparable dataset.

One of TARGET's key goals is to explain anomalous behavior. We prototypically achieve this goal by considering anomalies identified through their deviation from a set of baseline process graphs, observed within the context of a user session. To minimize computational requirements, we have adapted star structures, a bipartite representation used to approximate the edit distance between two graphs using the Kuhn-Munkres algorithm. Baseline templates for benign process behavior are generated automatically and adapt to the nature of the respective process.

Conclusion

The research conducted by JRC TARGET will become the scientific foundation of a new, innovative line of security solutions by our industrial partner company CyberTrap, as well as the basis for further academic research at the UAS.

We are confident that our efforts will contribute to the timely detection and interpretation of current and future cyber-attacks on private, corporate, and public IT/ICT infrastructures and networks. This work is complemented by topically related research conducted by other UAS R&D teams active in the areas of industrial and smart grid security, security management, and data privacy.

Appendix A – Publications

[1] S. Marschalek, R. Luh, M. Kaiser, und S. Schrittwieser, „Classifying Malicious System Behavior using Event Propagation Trees“, in Proceedings of the 17th International Conference on Information Integration and Web-based Applications Services (iiWAS2015), 2015.

[2] M. Wagner u. a., „A Survey of Visualization Systems for Malware Analysis“, in Eurographics Conference on Visualization (EuroVis) - STARs, Cagliari, Italy, 2015, S. 105–125, doi: 10.2312/eurovisstar.20151114.

[3] D. Buhov, R. Thron, und S. Schrittwieser, „Catch Me If You Can! Transparent Detection Of Shellcode“, International Conference on Software Security and Assurance (ICSSA), 2016.

[4] P. Kieseberg, E. Weippl, und S. Schrittwieser, „Detection of Data Leaks in Collaborative Data Driven Research“, ERCIM News, Nr. 105, 2016.

[5] P. Kieseberg, E. Weippl, und S. Schrittwieser, „Forensics using Internal Database Structures“, ERCIM News, Nr. 108, 2016.

[6] R. Luh, S. Marschalek, M. Kaiser, H. Janicke, und S. Schrittwieser, „Semantics-aware detection of targeted attacks – A survey“, Journal of Computer Virology and Hacking Techniques, S. 1–39, 2016, doi: 10.1007/s11416-016-0273-3.

[7] R. Luh, S. Schrittwieser, und S. Marschalek, „TAON: An Ontology-based Approach to Mitigating Targeted Attacks“, International Conference on Information Integration and Web-based Applications & Services (iiWAS), 2016.

[8] B. Malle, P. Kieseberg, S. Schrittwieser, und A. Holzinger, „Privacy Aware Machine Learning and the Right to be Forgotten“, ERCIM News, Nr. 107, 2016.

[9] S. Marschalek, M. Kaiser, R. Luh, und S. Schrittwieser, „Empirical Malware Research through Observation of System Behaviour“, in First Workshop on Empirical Research Methods in Information Security, 2016, S. 467–469, doi: 10.1145/2872518.2888609.

[10] M. Pirker und A. Nusser, „A Work-Flow for Empirical Exploration of Security Events“, in 25th International Conference Companion on World Wide Web, 2016, doi: 10.1145/2872518.2888607.

[11] M. Pirker und A. Nusser, „Assessment of Server State via Inter-Clone Differences“, International Conference on Software Security and Assurance (ICSSA), 2016.

[12] S. Schrittwieser, S. Katzenbeisser, J. Kinder, G. Merzdovnik, und E. Weippl, „Protecting software through obfuscation: Can it keep pace with progress in code analysis“, CSUR, Bd. 49, Nr. 1, 2016.

[13] S. Eresheim, R. Luh, und S. Schrittwieser, „The Evolution of Process Hiding Techniques in Malware – Current Threats and Possible Countermeasures“, Journal of Information Processing, 2017, doi: 10.2197/ipsjjip.25.866.

[14] P. Kieseberg, P. Frühwirt, und S. Schrittwieser, „Security Testing for Mobile Applications“, ERCIM News, Bd. 109, S. 52–53, 2017.
[15] P. Kieseberg, S. Neuner, S. Schrittwieser, und M. Schmiedecker, „Real-time Forensics through Endpoint Visibility“, International Conference on Digital Forensics & Cyber Crime (ICDF2C), 2017.
[16] P. Kieseberg, S. Schrittwieser, B. Malle, und E. Weippl, „Das Testen von Algorithmen in sensibler datengetriebener Forschung“, Rundbrief des Fachausschusses Management der Anwendungsentwicklung und -wartung (WI-MAW), 2017.
[17] P. Kieseberg, E. Weippl, und S. Schrittwieser, „Forensics using Internal Database Structures“, ERCIM News, Nr. 108, 2017.
[18] J. Kim, K. Kim, J. Cho, H. Kim, und S. Schrittwieser, „Hello, Facebook! Here is the stalkers’ paradise!: Design and analysis of enumeration attack using phone numbers on Facebook“, 13th International Conference on Information Security Practice and Experience (ISPEC 2017), 2017.
[19] R. Luh, G. Schramm, M. Wagner, und S. Schrittwieser, „Sequitur-based Inference and Analysis Framework for Malicious System Behavior“, First International Workshop on Formal Methods for Security Engineering, 2017.
[20] R. Luh, S. Schrittwieser, H. Janicke, und S. Marschalek, „Design of an Anomaly-based Threat Detection & Explication System“, Third International Conference on Information Systems Security and Privacy, Madeira, Portugal, 2017.
[21] R. Luh, S. Schrittwieser, S. Marschalek, H. Janicke, und E. Weippl, „Design of an Anomaly-based Threat Detection & Explication System“, 22nd ACM Symposium on Access Control Models and Technologies (SACMAT), 2017, doi: 10.1145/3078861.3084162.
[22] R. Luh, S. Schrittwieser, und S. Marschalek, „LLR-based Sentiment Analysis for Kernel Event Sequences“, 31th International Conference on Advanced Information Networking and Applications, 2017.
[23] J. Rauchberger, R. Luh, und S. Schrittwieser, „Longkit - A Universal Framework for BIOS/UEFI Rootkits in System Management Mode“, Third International Conference on Information Systems Security and Privacy, Madeira, Portugal, 2017.
[24] M. Geko und S. Tjoa, „An Ontology Capturing the Interdependence of the General Data Protection Regulation (GDPR) and Information Security“, in CECC 2018: Proceedings of the Central European Cybersecurity Conference 2018, Ljubljana, Slovenia, 2018, doi: 10.1145/3277570.3277590.
[25] R. Luh, G. Schramm, M. Wagner, H. Janicke, und S. Schrittwieser, „SEQUIN: a grammar inference framework for analyzing malicious system behavior“, Journal of Computer Virology and Hacking Techniques, S. 01–21, 2018, doi: 10.1007/s11416-018-0318-x.
[26] Luh, Robert, M. Temper, S. Tjoa, und S. Schrittwieser, „APT RPG: Design of a Gamified Attacker/Defender Meta Model“, in International Workshop on FORmal methods for Security Engineering, 2018.
[27] M. Pirker, P. Kochberger, und S. Schwandter, „Behavioural Comparison of Systems for Anomaly Detection“, in Proceedings of the 13th International Conference on Availability, Reliability and Security, Hamburg, Deutschland, 2018.
[28] J. Rauchberger u. a., „The Other Side of the Coin: A Framework for Detecting and Analyzing Web-based Cryptocurrency Mining Campaigns“, in Proceedings of the 13th International Conference on Availability, Reliability and Security, Hamburg, Deutschland, 2018.
[29] D. Rieger und S. Tjoa, „A Readiness Model for Measuring the Maturity of Cyber Security Incident Management“, International Conference on Intelligent Networking and Collaborative Systems (INCoS-2018), 2018.
[30] S. Tjoa, „Big Challenges – Future cyber-security challenges and the role of software security and assurance in the era of IoT, industry 4.0 and big data“, International Conference on Software Security and Assurance (ICSSA), Seoul, South Korea, 2018.
[31] T. Dam, L. D. Klausner, D. Buhov, und S. Schrittwieser, „Large-Scale Analysis of Pop-Up Scam on Typosquatting URLs“, in Proceedings of the 14th International Conference on Availability, Reliability and Security, Canterbury, United Kingdom, 2019, S. 53:1–53:9.
[32] R. Luh, „Advanced Threat Intelligence: Interpretation of Anomalous Behavior in Ubiquitous Kernel Processes“, De Monfort University Leicester, Dissertation, Juli 2019.
[33] R. Luh, H. Janicke, und S. Schrittwieser, „AIDIS: Detecting and classifying anomalous behavior in ubiquitous kernel processes“, Computers & Security, Nr. 84, S. 120–147, Juli 2019, doi: https://doi.org/10.1016/j.cose.2019.03.015.
[34] R. Luh, M. Temper, S. Tjoa, S. Schrittwieser, und H. Janicke, „PenQuest: a gamified attacker/defender meta model for cyber security assessment and education“, Journal of Computer Virology and Hacking Techniques, Nov. 2019, doi: 10.1007/s11416-019-00342-x.
[35] Luh, Robert und S. Schrittwieser, „Advanced threat intelligence: detection and classification of anomalous behavior in system processes“, e & i Elektrotechnik und Informationstechnik, Bd. Springer, S. 1–7, Dez. 2019.