ANALISIS DINAMIS DAN *DECISION TREE CLASSIFIER* UNTUK *MALICIOUS OFFICE* DAN *PDF*

Skripsi

diajukan untuk memenuhi bagian dari syarat memperoleh gelar Sarjana Komputer

Program Studi Ilmu Komputer

Dibuat oleh:

Jonathan Suara Patty 1804114

PROGRAM STUDI ILMU KOMPUTER

FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS PENDIDIKAN INDONESIA

2024

ANALISIS DINAMIS *MALWARE* DALAM SERANGAN *PHISHING* BERBENTUK DOKUMEN DENGAN *DECISION TREE*

Oleh

Jonathan Suara Patty

NIM 1804114

Sebuah skripsi yang diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Komputer Program Studi Ilmu Komputer di Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam

Universitas Pendidikan Indonesia

Agustus 2024

LEMBAR PENGESAHAN

ANALISIS DINAMIS MALWARE DALAM SERANGAN PHISHING BERBENTUK DOKUMEN DENGAN DECISION TREE

Oleh

Jonathan Suara Patty

1804114

Disetujui dan disahkan oleh:

Pembimbing I,

Rizky Rachman J., M.Kom.

NIP. 197711252006041002

Pembimbing II,

Yudi Ahmad Hambali, M.T.

NIP. 199005302019031013

Mengetahui,

Ketua Program Studi Ilmu Komputer

Dr. Muhammad Nursalman, M.T.

NIP. 197909292006041002

ANALISIS DINAMIS DAN *DECISION TREE CLASSIFIER* UNTUK *MALICIOUS OFFICE* DAN *PDF*

ABSTRAK

Penelitian ini menyoroti analisis dinamis terhadap *malware* yang digunakan dalam serangan *phishing* berbentuk dokumen, dengan memanfaatkan *decision tree classifier* untuk meningkatkan langkah-langkah keamanan siber. Sampel *malware* dikumpulkan dengan cermat dari honeypot suatu perusahaan, mewakili beragam potensi ancaman. Dari sampel yang dikumpulkan, beberapa akan ditetapkan untuk melatih *decision tree* dan beberapa sampel lainnya akan digunakan untuk mengevaluasi kinerjanya. Analisis dinamis dilakukan dalam lingkungan mesin virtual *Linux* untuk memastikan tempat pengujian yang terkontrol dan aman. *Decision tree* dibangun menggunakan *Python*, dengan mengintegrasikan pustaka *scikit-learn* yang kuat. Dengan menggunakan metode *classifier*, *decision tree* mampu membedakan secara efektif antara sampel *benign* dan sampel berbahaya, menunjukkan ketangguhannya dalam mengidentifikasi ancaman. Selain itu, *decision tree* mampu mengkategorikan *malware* yang teridentifikasi menjadi empat klasifikasi yang berbeda: *bot*, *trojan*, *ransomware*, dan *spyware*. Pendekatan komprehensif ini tidak hanya menyoroti efektivitas *decision tree classifier* dalam deteksi *malware* tetapi juga menegaskan potensinya dalam menyempurnakan proses klasifikasi *malware*. Temuan ini menunjukkan bahwa penerapan teknik semacam itu dapat secara signifikan memperkuat akurasi dan keandalan pertahanan keamanan siber terhadap serangan *phishing* yang canggih.

Kata kunci: Analisis Dinamis, *Decision Tree*, Dokumen, Keamanan Siber, *Malware*, *Phishing*.

DYNAMIC ANALYSIS AND DECISION TREE CLASSIFIER FOR MALICIOUS OFFICE DOCUMENTS AND PDFS

ABSTRACT

This research focuses on the dynamic analysis of malware used in document-based phishing attacks, leveraging a decision tree classifier to enhance cybersecurity measures. The malware samples were meticulously gathered from a company's honeypot, representing a wide array of potential threats. Among these, several samples were designated for training the decision tree, while several more were utilized to evaluate its performance. The dynamic analysis was executed within a Linux virtual machine environment to ensure a controlled and secure testing ground. The decision tree was constructed using Python, incorporating the powerful scikit-learn library. By employing the classifier method, the decision tree effectively distinguished between benign and malicious samples, showcasing its robustness in identifying threats. Additionally, the decision tree was capable of further categorizing the identified malware into four distinct classifications: bots, trojans, ransomware, and spyware. This comprehensive approach not only highlights the efficacy of decision tree classifiers in malware detection but also underscores their potential in refining malware classification processes. The findings suggest that employing such techniques can significantly bolster the accuracy and reliability of cybersecurity defenses against sophisticated phishing attacks.

Keywords: Cybersecurity, Decision Tree, Document, Dynamic Analysis, Malware, Phishing

DAFTAR ISI

DAFTAR LAMPIRAN

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR PUSTAKA

Abdelghani, T. S. C. H. R. O. U. B. (2019, February). *Industrial control systems (ics) security in power transmission network. In 2019 Algerian Large Electrical Network Conference* (CAGRE) (pp. 1-4). IEEE.

Aonzo, S., Merlo, A., Tavella, G., & Fratantonio, Y. (2018, October). Phishing *attacks on modern android. In Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security* (pp. 1788-1801).

Arachchilage, N. A. G., Love, S., & Beznosov, K. (2016). *Phishing threat avoidance behaviour: An empirical investigation. Computers in Human Behavior*, 60, 185-197.

Årnes, A. (Ed.). (2017). *Digital forensics*. John Wiley & Sons.

Asamoah, H. (2020). *Antivirus software versus malware*. Архів кваліфікаційних робіт.

Atlas, L. G. (2019). *Anti Malware*.

Bendovschi, A. (2015). *Cyber-attacks–trends, patterns and security countermeasures*. *Procedia Economics and Finance*, *28*, 24-31.

Bhoopal, U. (2021). *Phishing Attempts in Cyber Crime. Supremo Amicus*, 24, 39.

Chakkaravarthy, S. S., Sangeetha, D., & Vaidehi, V. (2019). *A survey on malware analysis and mitigation techniques. Computer Science Review*, 32, 1-23.

Chaudhry, J. A., Chaudhry, S. A., & Rittenhouse, R. G. (2016). *Phishing attacks and defenses. International Journal of Security and Its Applications*, 10(1), 247-256.

SonicWall. (2022). *Cyber Threat Report. Available online*: https://theblockchaintest.com/uploads/resources/SonicWall%20-%20Cyber%20Threat%20Report%20-%202022%20Feb.pdf

Delaney, J. (2020). *The Effectiveness of Antivirus Software* (Doctoral dissertation, Utica College).

Díaz, R. M. (2020). *Cybersecurity in the time of COVID-19 and the transition to cyberimmunity.*

Djenna, A., Bouridane, A., Rubab, S., & Marou, I. M. (2023). *Artificial intelligence-based malware detection, analysis, and mitigation. Symmetry*, 15(3), 677.

Djenna, A., Saidouni, D. E., & Abada, W. (2020). A pragmatic cybersecurity strategy for combating IoT-cyberattacks. In *2020 International Symposium on Networks, Computers and Communications (ISNCC)* (pp. 1-6). IEEE. https://doi.org/10.1109/ISNCC49221.2020.9297251

Eng, E., & Caselden, D. (2015). *Operation clandestine wolf–Adobe flash zero-day in APT3 phishing campaign. FireEye. June, 23*.

Gupta, P., Srinivasan, B., Balasubramaniyan, V., and Ahamad, M. (2015). "*Phoneypot: data-driven understanding of telephony threats*," *in Proceedings 2015 network and distributed system security symposium*, (Reston, VA: Internet Society), 8–11. doi:10.14722/ndss.2015.23176

Han, Z. (2014). *Design and Implementation of Security Cloud Active Defense System Against Malicious Code. Journal of Chongqing University of Technology.*

Huang, L. S., Moshchuk, A., Wang, H. J., Schecter, S., & Jackson, C. (2012). *Clickjacking: Attacks and defenses. In 21st USENIX Security Symposium* (USENIX Security 12) (pp. 413-428).

Huang, Z., Wang, Q., Chen, Y., & Jiang, X. (2020). *A survey on machine learning against hardware trojan attacks: Recent advances and challenges*. IEEE Access, 8, 10796-10826.

Humayun, M., Jhanjhi, N. Z., Alsayat, A., & Ponnusamy, V. (2021). *Internet of things and ransomware: Evolution, mitigation and prevention. Egyptian Informatics Journal*, 22(1), 105-117.

Jakobsson, M., and Myers, S. (2006). *Phishing and countermeasures: understanding the increasing problems of electronic identity theft*. New Jersey: John Wiley and Sons.

Jamalpur, S., Navya, Y. S., Raja, P., Tagore, G., & Rao, G. R. K. (2018, April). *Dynamic malware analysis using cuckoo sandbox. In 2018 Second international conference*

on inventive communication and computational technologies (ICICCT) (pp. 1056-1060). IEEE.

Jansson, K., & von Solms, R. (2013). *Phishing for phishing awareness. Behaviour & information technology*, 32(6), 584-593.

Khan, S. K., Shiwakoti, N., Stasinopoulos, P., & Chen, Y. (2020). Cyb*er-attacks in the next-generation cars, mitigation techniques, anticipated readiness and future directions. Accident Analysis & Prevention*, 148, 105837.

Kravchik, M., & Shabtai, A. (2018, January). *Detecting cyber attacks in industrial control systems using convolutional neural networks. In Proceedings of the 2018 Workshop on Cyber-Physical Systems Security and PrivaCy* (pp. 72-83).

Kruse II, W. G., & Heiser, J. G. (2001). *Computer forensics: incident response essentials. Pearson Education*.

Lin, T., Capecci, D. E., Ellis, D. M., Rocha, H. A., Dommaraju, S., Oliveira, D. S., & Ebner, N. C. (2019). *Susceptibility to spear-phishing emails: Effects of internet user demographics and email content*. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *26*(5), 1-28.

Martín, A., Lara-Cabrera, R., & Camacho, D. (2018). *A new tool for static and dynamic Android malware analysis. In Data Science and Knowledge Engineering for Sensing Decision Support: Proceedings of the 13th International FLINS Conference* (FLINS 2018) (pp. 509-516).

Möller, D.P., Haas, R.E., 2019. *Guide to Automotive Connectivity and Cybersecurity. Springer*.

Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). *An introduction to decision tree modeling. Journal of Chemometrics: A Journal of the Chemometrics Society*, 18(6), 275-285.

Nurhayati, A., & Frencius, F. (2019). *Mapping perception of consumer antivirus software with multidimensional scaling method. Aptikom Journal on Computer Science and Information Technologies*, 4(3), 91-95.

Or-Meir, O., Nissim, N., Elovici, Y., & Rokach, L. (2019). *Dynamic malware analysis in the modern era—A state of the art survey. ACM Computing Surveys* (CSUR), 52(5), 1-48.

Ramzan, Z. (2010). *Phishing attacks and countermeasures*. *Handbook of information and communication security*, 433-448.

Pachhala, N., Jothilakshmi, S., & Battula, B. P. (2021, October). *A Comprehensive Survey on Identification of Malware Types and Malware Classification Using Machine Learning Techniques. In 2021 2nd International Conference on Smart Electronics and Communication* (ICOSEC) (pp. 1207-1214). IEEE.

Pierazzi, F., Mezzour, G., Han, Q., Colajanni, M., & Subrahmanian, V. S. (2020). *A data-driven characterization of modern Android spyware. ACM Transactions on Management Information Systems* (TMIS), 11(1), 1-38.

Proofpoint. (2022). S*tate of the Phish 2022 Report. Proofpoint.*

Qamar, A., Karim, A., & Chang, V. (2019). *Mobile malware attacks: Review, taxonomy & future directions. Future Generation Computer Systems*, 97, 887-909.

Shalaginov, A., Banin, S., Dehghantanha, A., & Franke, K. (2018). *Machine learning aided static malware analysis: A survey and tutorial. In Cyber threat intelligence* (pp. 7-45). Springer, Cham.

Shankar, A., Shetty, R., & Nath, B. (2019). *A review on phishing attacks. International Journal of Applied Engineering Research*, 14(9), 2171-2175.

Sihwail, R., Omar, K., & Ariffin, K. Z. (2018). *A survey on malware analysis techniques: Static, dynamic, hybrid and memory analysis*. Int. J. Adv. Sci. Eng. Inf. Technol, 8(4-2), 1662-1671.

Sudhakar, & Kumar, S. (2020). *An emerging threat Fileless malware: a survey and research challenges. Cybersecurity*, 3(1), 1.

Usman, N., Usman, S., Khan, F., Jan, M. A., Sajid, A., Alazab, M., & Watters, P. (2021). *Intelligent dynamic malware detection using machine learning in IP reputation for forensics data analytics. Future Generation Computer Systems*, 118, 124-141.

Utku, A., Doğru, İ. A., & Akcayol, M. A. (2018, *May*). *Decision tree based android malware detection system. In 2018 26th Signal Processing and Communications Applications Conference* (SIU) (pp. 1-4). IEEE.

Vayansky, I., & Kumar, S. (2018). *Phishing–challenges and solutions. Computer Fraud & Security*, 2018(1), 15-20.

Vidyarthi, D., Kumar, C. R. S., Rakshit, S., & Chansarkar, S. (2019). *Static malware analysis to identify ransomware properties. International Journal of Computer Science Issues* (IJCSI), 16(3), 10-17.

Von Neumann, J. (2012). *The computer and the brain. Yale university press*.

Whittaker, C., Ryner, B., & Nazif, M. (2010). *Large-scale automatic classification of phishing pages.*

Yeboah-Ofori, A. (2020). *Classification of malware attacks using machine learning in decision tree. International Journal of Security,* 11(2), 10-25.