IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.39 ⁜ Vol. 11, Issue 8, August 2022

DOI: 10.17148/IJARCCE.2022.11805

Design and Development of Honeypot to prevent Phishing using Machine Learning Techniques

Pavan Gupta[1], Prathamesh Mishra[2], Mausam Bhunia[3]

Student, Electronics & Telecommunication, Thakur College of Engineering and Technology, Mumbai, India[1]

Student, Electronics & Telecommunication, Thakur College of Engineering and Technology, Mumbai, India[2]

Student, Electronics & Telecommunication, Thakur College of Engineering and Technology, Mumbai, India[3]

Abstract: Mechanized malware utilizes honeypot identifying instruments inside its code. When honeypot usefulness has been uncovered, malware, for example, botnets will stop the endeavoured split the difference. Ensuing malware variations utilize comparative methods to sidestep identification by known honeypots. This decreases the expected size of a caught dataset and the ensuing investigation. This paper includes many research done on honeypot with machine learning. And also include our methodology for detecting the attackers and learning the attacker's method for intrusions through reinforming learning and capturing different data about attackers.

Keywords: Cybersecurity, Machine Learning, Python, Hacking, Cyber-Crime

I. INTRODUCTION

Malware is perhaps the most expanded security danger in late years. Malware is explicitly intended to assault frameworks to take delicate and classified data. Likewise, malware can be intended to upset running frameworks to make advanced disarray [1][2]. A few different methodologies have been proposed by numerous analysts to identify malware, including utilizing machine learning. AI can be very compelling and proficient for malware location [4], and AI-based antimalware programming is a successful technique to utilize [5]. Furthermore, contrasted with the mark investigation-based approach, AI has better viability [6]. Machine learning will distinguish the presence of malware in view of the order of a class that has been characterized in a dataset. Characterization of a class in a dataset is regularly called a mark that is separated into two classes, specifically harmless and malware [7]. Simultaneously, the specialists proposed a honeypot as a gadget fit for social affair data on malware action [9][10][11]. Honeypot is a framework that is intentionally left as a trap to bait potential aggressors so they avoid basic frameworks. Honeypot can assist with machining learning update preparing information so it can give better precision. Notwithstanding, in its application, the honeypot doesn't have a reasonable plan, so it becomes troublesome and confounding when it is executed, particularly in acclimating to one's necessities in fostering an AI model. Hence, important writing is required to have been explored further to distinguish drifts and give guidance later on. This study led to a writing audit utilizing the precise writing survey (SLR) technique to distinguish honeypot patterns. By utilizing SLR, analysts can track down arrangements by looking into applicable SSH session that is carrying out malicious activities feature set is significant in determining SSH sorts of malware dealt with by honeypots. This study dissects articles with writing information base sources on the IEEE Xplore and ACM from 2010 to 2020

RESEARCH METHODOLOGY AND LITERATURE SURVEY

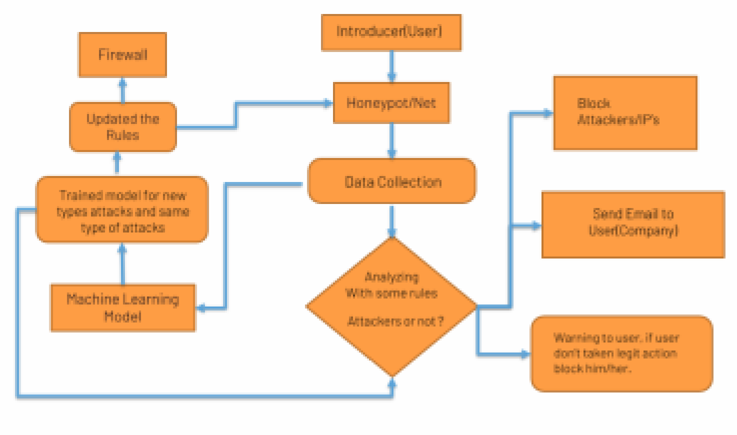
A) A Smart Agent Design for Cyber Security Based on Honeypot and Machine Learning [13] To manage 0-day and future assaults, the honeypot procedure can be utilized inactively as a data framework, yet in addition to support the conventional protection frameworks against future assaults. The particular arrangement in view of honeypot and the mix of AI calculations shapes a strong displaying and prescient framework for dubious profile acknowledgment and characterization. Subsequently, it addresses an incorporated productive framework for network safety to manage future and 0-day assaults. B) Automatic Identification of Honeypot Server Using Machine Learning Techniques [14] Many traditional security detection strategies such as firewall or intrusion detection systems (IDS) have been invented to protect the system's security, but there are still many critical issues which are reported every day. The situation has become more complex with the development of Internet technologies. Honeypots are defined as a new deception technology of cybersecurity defence in recent years, which can be used to strengthen the company's security detection level. In other words, honeypots are used to lure attackers into interacting with them and collect the in-formation which will be used to analyse and study the attack way of attackers. past exploration. This study means to get patterns, procedures, and Have proposed a machine learning model to classify SSH attacks based on the attack nature. It found that the high-speed networks demand better mechanisms to detect the compromises. In this paper, we detect a compromised compromises.

DOI: 10.17148/IJARCCE.2022.11805

After analyses on the performance of several machine learning algorithms, the performance of J48 and PART seemed to be promising and produced better result. C) Using Reinforcement Learning to Conceal Honeypot Functionality [15] They deployed two Cowrie honeypots at the same time; one adaptive as detailed the adaptive honeypot was developed to generate rewards in 75 states and is freely available. on the adaptive honeypot EC2 instance. and brute-force attempts. commands executed on the honeypots. Dictionary, and brute-force attacks are excluded as they represent pre-compromise These commands all represent interactions post-compromise. attacks were recorded on the honeypot. Other SSH malware interacted with the honeypot. D) Detection of Severe SSH Attacks Using Honeypot Servers and Machine Learning Techniques [16] There are attacks on or using an SSH server – SSH port scanning, SSH brute-force attack, and attack using a compromised server. Attacks using a server could be DoS attack, Phishing attack, E-mail spamming and so on. Sometimes an attacker breaks into a public SSH server and uses it for the above activities. Mostly, it is hard to detect the compromised SSH servers that were used by the attackers. However, by analysing the system logs an organisation can know about the compromises. For an organisation holding several SSH servers, it would be tedious to analyse the log files manually.

Proposed idea

We are aiming to build open source or a tool for users (companies). Through this user (company) can: 1. User can download the tool in his/her server and configure it to according to its requirement. 2. It will setup multiple private ports that can be attacked by attackers. If someone connect to these ports it will blacklist them 3. It will monitor all the users and make a log and looks for brute force attempts 4. Have a profile of user who are attacking a system. So, it will be beneficial for system to protect useful and sensitive information 5. System can classify the attackers and can learn new way of attacking with help of machine learning 6. It will email users when attack occurs and let you know what the attack was.


Introducer(User)
Firewall
Honeypot/Net
Updated the Rules
Block Attackers/IP's
Data Collection
Trained model for new types attacks and same type of attacks
Send Email to User(Company)
Analyzing With some rules Attackers or not ?
Machine Learning Model
Warning to user. if user don't taken legit action block him/her.

RESULT AND DISCUSSION

The improvement of Internet and web-based media adds to increasing the information delivered on the Internet and the associated hubs, yet the default establishment and the design of assortment of programming frameworks address some security openings and weaknesses, while most of Internet clients have not exactly set up wellbeing mindfulness, prompting tremendous security chances. With the improvement of organization assault procedures, each host on the Internet has turned into the objective of assaults. In this manner, the organization data security can't be overlooked as an issue. From different research we can get that, 1. With the help of machine learning and deep learning the power of honeypot can be enhanced. 2. We can different classification algorithm to determine the types of attack and types of users.The improvement of Internet and web-based media adds to increasing the information delivered on the Internet and the associated hubs, yet the default establishment and the design of assortment of programming frameworks address some security openings and weaknesses, while most of Internet clients have not exactly set up wellbeing mindfulness, prompting tremendous security chances. With the improvement of organization assault procedures, each host on the Internet has turned into the objective of assaults. In this manner, the organization data security can't be overlooked as an issue. From different research we can get that,

1. With the help of machine learning and deep learning the power of honeypot can be enhanced.
2. We can different classification algorithm to determine the types of attack and types of users
3. With Reinforcement learning we can train our model to develop a logic to identify attackers and block them.
4. If we improve existence technology, we may secure many data and resource.

DOI: 10.17148/IJARCCE.2022.11805

CONCLUSION

As an active cybersecurity defence strategy, the emergence of honeypots further reduces the attacker's activity space. Although honeypot technology is constantly improving, attackers are constantly searching for weaknesses in honeypots to identify them. we have discussed different methodology used by security researcher. With the help of machine learning we may get a better security system.

REFERENCES

[1] C. Vatamanu, D. Cosovan, D. Gavrilu, and H. Luchian, "A Comparative Study of Malware Detection Techniques Using Machine Learning Methods," Int. J. Comput. Electr. Autom. Control Inf. Eng., vol. 9, no. 5, pp. 1115–1122, 2015.

[2] T. Mithal, K. Shah, and D. K. Singh, "Case Studies on Intelligent Approaches for Static Malware Analysis," Emerg. Res. Comput. Information, Commun. Appl., pp. 555–567, 2016, DOI: 10.1007/978- 981-10-0287-8.

[3] Kaspersky Lab, "Kaspersky Lab′s Cyber Security report," 2016.

[4] I. Firdausi, C. Lim, A. Erwin, and A. S. Nugroho, "Analysis of machine learning techniques used in behavior-based malware detection," Proc. - 2010 2nd Int. Conf. Adv. Comput. Control Telecommun. Technol. ACT 2010, pp. 201–203, 2010, DOI: 10.1109/ACT.2010.33.

[5] Z. Markel and M. Bilzor, "Building a machine learning classifier for malware detection," WATeR 2014 - Proc. 2014 2nd Work. Anti-Malware Test. Res., 2015, DOI: 10.1109/WATeR.2014.7015757.

[6] M. G. Schultz, E. Eskin, E. Zadok, and S. J. Stolfo, "Data mining methods for detection of new malicious executables," Proc. IEEE Comput. Soc. Symp. Res. Secur. Priv., pp. 38–49, 2001, DOI: 10.1109/secpri.2001.924286.

[7] U. Pehlivan, N. Baltaci, C. Acarturk, and N. Baykal, "The analysis of feature selection methods and classification algorithms in permission-based Android malware detection," IEEE SSCI 2014 2014 IEEE Symp. Ser. Comput. Intell. - CICS 2014, 2014. IEEE Symp. Comput. Intell. Cyber Secur. Proc., pp. 1–8, 2014, DOI: 10.1109/CICYBS.2014.7013371.

[8] Y. Roh, G. Heo, and S. E. Whang, "A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective," IEEE Trans. Knowl. Data Eng., vol. 4347, no. c, pp. 1–1, 2019, DOI: 10.1109/tkde.2019.2946162.

[9] K. Saikawa and V. Klyuev, "Detection and Classification of Malicious Access using a Dionaea Honeypot," Proc.2019 10th IEEE Int. Conf. Intell. Data Acquis. Adv. Comput. Syst. Technol. Appl. IDAACS 2019, vol. 2, pp. 844–848, 2019, DOI: 10.1109/IDAACS.2019.8924340.

[10] P. D. Ali and T. Gireesh Kumar, "Malware capturing and detection in dionaea honeypot," 2017 Innov. Power Adv. Comput. Technol. i-PACT 2017, vol. 2017-Janua, pp. 1–5, 2017, DOI: 10.1109/IPACT.2017.8245158.

[11] V. Sethia and A. Jeyasekar, "Malware capturing and analysis using dionaea honeypot," Proc. - Int. Carnahan Conf. Secur. Technol., vol. 2019-October, pp. 0–3, 2019, DOI: 10.1109/CCST.2019.8888409.

[12] W. Stallings, Computer Security Third Edition, 3rd Editio. United States of America: Pearson Education, 2015

[13] A Smart Agent Design for Cyber Security Based on Honeypot and Machine Learning, Mohamed Eddabbah,2 Youssef Lmoumen,1 and Raja Touahni1, Nadiya El Kamel

[14] Cheng Huang, Jiaxuan Han, Xing Zhang, Jiayong Liu, "Automatic Identification of Honeypot Server Using Machine Learning Techniques", Security and Communication Networks, vol. 2019, Article ID 2627608, 8 pages, 2019. https://doi.org/10.1155/2019/2627608

[15] Using Reinforcement Learning to Conceal Honeypot Functionality, Seamus Dowling,1[0000−0001−8722−2009] Michael Schukat2 and Enda Barrett2 1 Galway Mayo Institute of Technology, Castlebar, Mayo, Ireland 2 National University of Ireland Galway, Galway, Ireland

[16] Detection of Severe SSH Attacks Using Honeypot Servers and Machine Learning Techniques Gokul Kannan Sadasivam1,*, Chittaranjan Hota1, Bhojan Anand2 •Department of Computer Science and Information Systems, BITS, Pilani – Hyderabad Campus, Hyderabad, Telangana, India-500078 •School of Computing, National University of Singapore, Computing 1, 13 Computing Drive, Singapore - 117417.

BIOGRAPHY

Pavan Gupta Student at Thakur College of Engineering and Technology in the branch Electronics & Telecommunication City- Mumbai, India