Chapter 2
Research Trends for Malware and Intrusion Detection on Network Systems:
A Topic Modelling Approach

Santosh Kumar Smmarwar
National Institute of Technology, Raipur, India

Govind P. Gupta
https://orcid.org/0000-0002-0456-1572
National Institute of Technology, Raipur, India

Sanjay Kumar
National Institute of Technology, Raipur, India

ABSTRACT

With more uses of internet-based services, the risk of cyberattacks is growing continuously. To analyze these research trends for malware and intrusion detection, the authors applied the topic modeling approach in the study by using the LDA (latent dirichlet allocation) and calculating the maximum and minimum probability of the words, which appears in the large collection of text. The LDA technique is useful in finding the hidden topics for further research in the areas of network and cybersecurity. In this chapter, they collected the abstract of two thousand papers from the Scopus library from 2014 to 2021. These collected papers are from reputed publications such as Elsevier, Springer, and IEEE Transactions. The main aim of this study is to find research trends based on keywords that are untouched or on which less research work has been done. To the best of the authors' knowledge, this will be the first study done by using the LDA technique for topic modeling in the areas of network security to demonstrate the research gap and trends for malware and intrusion detection systems.

DOI: 10.4018/978-1-7998-7789-9.ch002

INTRODUCTION

According to a report of Equinix (Catalin Cimpanu et al. 2020), a 45% rise in cyber-attack may be seen till 2023 on the network communication system domain with the increasing use of computer communication networks and internet-based services. The threat of attack is ever-expanding over network infrastructure as the different techniques of attacks are being used by attackers. The traditional method of detecting new attacks is not effective as of as it should be to detect and respond before misuse of unauthorized users to the computer resources and private information of the organizations. There are various kind of attack exist known as intrusion, anomaly, malware, viruses, ransomware, adware, Trojan horses, DDoS, DoS, and many more. Malware is the malicious code that moves across the computer network to gain unauthorized access into the computer's critical resources, files, root directory, etc. Malware is designed to steal information, provide losses for the computer system, embedded into the user's code to scan sensitive information, and sending data on third parties servers. Different types of Malware has been classified based on their activity into computer systems such as adware, Trojan, bot, worm, virus, spyware, Ransomware, Rootkit, downloader, Launcher, Backdoor, etc. (Gibert et al., 2020), malware analysis follows two approaches for detecting malware as static analysis and dynamic analysis. The static approach follows the finding unwanted pattern without executing the codes, whereas the dynamic approach works on running code and monitoring the behavior of systems activity (Ren et al., 2019).

Intrusion is unauthorized access that tries to intrude into the privacy of a network. There are different types of intruders such as masquerade and clandestine users to detect intruders, an in-network system deploys a smart system known as an intrusion detection system. An intrusion detection system (IDS) is a security model of the network system used to trace the unauthorized activity of the network through the scanning process of traffic analysis of network packets. The intrusion model identifies the system activities behavior whether it is normal or abnormal and responds to it to the network administrator. The IDS works based on predefined records it contains and finds out the intrusion. It is a predictive model used in the cyber and network security domain that consists of various machine learning algorithms in identifying the intrusion very accurately (Belavagi et al., 2016).

There is two class of intrusion detection systems such as network based detection system and host based detection system (Farnaaz et al., 2016). Network-based detection systems work as passive mode means only monitor the activity of network traffic and send it to central authority while in the case of host-based systems, it is capable of detecting the internal activity of the system and also filter the packet data. The IDS can be classified as anomaly-based and misuse-based detection to detect new attacks and unknown patterns as well as known attacks from databases respectively (Stein et al., 2005, Belavagi et al., 2016). There is a lot of work is being done by researchers in the field of intrusion and anomaly detection and proposed various techniques of machine learning, deep learning, and data mining tools in the areas of the internet of things to developed efficient IDS (Da Costa et al., 2019, Hasan et al., 2019), However less work has done to find out research trend in network security. So our contribution in this paper is to find out the research gap from a large collection of abstract papers by using the topic modeling approach and LDA technique of text mining. In this study, the authors have collected the data of 2000 abstract papers from the Scopus library. These collected data having the eight Columns such as Author's name, Title, Year of publication, Links of paper, Abstract, Author Keywords, Indexed keywords, and References. After collecting authors are doing text cleaning process in which removing unnecessary information which has less meaning in deciding the topic to further research gap, then after applying