Chapter 15
Threat Detection in Cyber Security Using Data Mining and Machine Learning Techniques

Daniel Kobla Gasu
https://orcid.org/0000-0002-6208-6095
Department of Computer Science, University of Ghana, Ghana

ABSTRACT

The internet has become an indispensable resource for exchanging information among users, devices, and organizations. However, the use of the internet also exposes these entities to myriad cyber-attacks that may result in devastating outcomes if appropriate measures are not implemented to mitigate the risks. Currently, intrusion detection and threat detection schemes still face a number of challenges including low detection rates, high rates of false alarms, adversarial resilience, and big data issues. This chapter describes a focused literature survey of machine learning (ML) and data mining (DM) methods for cyber analytics in support of intrusion detection and cyber-attack detection. Key literature on ML and DM methods for intrusion detection is described. ML and DM methods and approaches such as support vector machine, random forest, and artificial neural networks, among others, with their variations, are surveyed, compared, and contrasted. Selected papers were indexed, read, and summarized in a tabular format.

INTRODUCTION

Cyber security requirements in organizations have evolved in the last several decades as a consequence of communication networks and information systems having become an essential factor in economic, social development and almost every facet of our daily lives (Singh & Nene, 2013). Security challenges such as intrusion, malware, phishing, misuse of the system, unauthorized modification of information (Vani & Krishnamurthy, 2018) and denial of service attacks pose threats to cyber infrastructure. Moreover, attackers constantly adapt to detection schemes and actively seek to exploit new vulnerabilities. Threats are becoming more advanced with the emergence of Advanced Persistent Threats (APTs), social engi-

DOI: 10.4018/978-1-7998-3149-5.ch015

neering, ransomware, and fraud committed through digital identity theft (Suraj, Kumar Singh, & Tomar, 2018). Hence, for detection schemes to remain relevant they must necessarily deal with the distribution of data changes over time (non-stationarity) (Verma, 2018).

This survey paper focuses on Machine Learning (ML) and Data Mining (DM) techniques for cyber security, particularly intrusion detection. Papers that had more citations were preferred because these described popular techniques. However, it was also recognized that this emphasis might overlook significant new and emerging techniques, so some of these papers were chosen also. Four research questions were posed. These questions were then used to collect the necessary information from papers in the review process. The section below enumerates the review questions.

SRQ1: Which journal is the dominant cyber threat detection journal?
SRQ2. What kind of data mining and machine learning algorithms were used in detecting threats in cyber space?
SRQ3. What kind of datasets were used for training algorithms to detect threats?
SRQ4. What methodology was adopted in conducting the research?

The aforementioned review questions were motivated by the following objectives. They are arranged in the order the review questions are stated.

1. To identify the most important cyber threat detection journal
2. To identify the effectiveness of using data mining and machine learning in cyber security analytics to detect threats to cyber infrastructure
3. To identify whether predictive models are repeatable or not by examining the usage of public datasets.
4. To identify the appropriateness of methodologies used.

This systematic literature review (SLR) is being undertaken to:

- Systematically review literature on various data mining and machine learning techniques in support of cyber security analytics to detect threats and predict cyber-attacks.
- Conduct an examination of papers in data mining and machine learning in relation to the various algorithms implemented.
- Present a clear picture of the current state of research in the field of data mining and machine learning in support of threat detection and intrusion detection.
- Present a summary of research results and provide pointers to areas and ideas that may be identified as candidates for future research.

This paper is divided into 6 sections. Section two describes the main steps in conducting this review. Background to study and overview of Data Mining and Machine Learning methods for attack/Intrusion detection is presented in Section three. Section four presents the results of the review. Sections 5 discusses the results and section six concludes the paper by providing an outlook on future research.