Chapter 76
Artificial Intelligence and Big Data Analytics in Support of Cyber Defense

Louise Leenen
University of the Western Cape, South Africa & CAIR, South Africa

Thomas Meyer
University of Cape Town, South Africa & CAIR, South Africa

ABSTRACT

Cybersecurity analysts rely on vast volumes of security event data to predict, identify, characterize, and deal with security threats. These analysts must understand and make sense of these huge datasets in order to discover patterns which lead to intelligent decision making and advance warnings of possible threats, and this ability requires automation. Big data analytics and artificial intelligence can improve cyber defense. Big data analytics methods are applied to large data sets that contain different data types. The purpose is to detect patterns, correlations, trends, and other useful information. Artificial intelligence provides algorithms that can reason or learn and improve their behavior, and includes semantic technologies. A large number of automated systems are currently based on syntactic rules which are generally not sophisticated enough to deal with the level of complexity in this domain. An overview of artificial intelligence and big data technologies in cyber defense is provided, and important areas for future research are identified and discussed.

INTRODUCTION

Governments, military forces and other organisations responsible for cyber defence deal with vast amounts of data that has to be understood in order to lead to intelligent decision making. The rapid increase in the number and variety of cyber threats, and in the volume of information that has to be processed, integrated and understood to provide efficient counter-measures provide challenges to the defence community. Integration of information requires an encoded common vocabulary and shared understanding

DOI: 10.4018/978-1-7998-7705-9.ch076

of the domain whilst the vast amounts of information pertinent to cybersecurity requires automation for processing and decision making. It is widely recognised that cyber defence requires the capabilities of artificial intelligence (AI) and big data processing (Vasudevan, 2018), (Graham, 2018) (Masimbuka, Grobler, & Watson, 2018). Big data provides the huge sets of data AI algorithms require to train data and to learn, i.e. to determine what normal behaviour is and thus to be able to detect abnormal events. These technologies are used for intrusion detection, malware classification and attribution, attack prediction and other applications. Artificial intelligence has made a resurgence in the past decade due to an underlying component, semantic technology. Semantic technologies represents a number of different technologies aiming to derive meaning from information. The combination of AI with big data capabilities go hand in hand to manage different data sets, to gain interoperability and insights and to make predictions[1]. One example of a limitation of current cybersecurity systems is that they tend to produce large numbers of false positives. Semantic technologies in conjunction with big data can improve this limitation due, in part, to recent advances in the scalability of techniques for managing semantic technologies.

Big data processing refers to data processing that is different from traditional processing technologies with respect to the volume of data, the rate at which data is data generated and rate at which data is transmitted, in addition to the fact that it includes both structured and unstructured data. Big data refers to volumes of data that are too large to handle by traditional data base systems. Big data analytics refers to advanced analytic techniques such as machine learning, predictive analysis, and other intelligent processing and mining techniques applied to big data sets. Big data analytics is required to combine different sources of information in order to recognise patterns for the detection of network attacks and other cyber threats. This must take place fast enough so that counter measures can be put in place. According to Saurabh, the CEO of a cybersecurity platform provider, cybersecurity in mature organisations depend on big data and in most cases it is big data that eliminates vulnerabilities and halt attacks (Saurabh, 2017).

Semantic technologies is a knowledge representation paradigm where the meaning of data is encoded separately from the data itself. The use of semantic technologies such as logic-based systems to support decision making and an ability to process large sets of data have become essential. Hernandez-Ardieta & Tapiador (2013) state that it is virtually impossible for any organisation to manage cyber threats without collaboration with partners and allies. Collaboration includes sharing of threat related and cybersecurity information on a near real-time basis and this requirement necessitates the development of infrastructure and mechanisms to facilitate the information sharing, specifically through standardisation of data formats and exchange protocols. It is not merely *how* to share information but also *what*, with *whom* and *when* to share, as well as reasoning about the repercussions of sharing sensitive data. This level of collaboration will be impossible without attaching meaning to data and the ability to reason over formal structures. The use of ontologies is the underlying semantic technology driving the Semantic Web initiative (Berners-Lee, Hendler, & Lassila, 2001). Blockchain technology is another emerging technology in the cyber defence domain.

Issues that arise from the use of AI and big data are the protection of privacy and the ethical use of these technologies. It should also be kept in mind that attackers can also use these technologies to their advantage.

Section 2 covers background on semantic technologies, including ontologies, and blockchain technology. Section 3 discusses current applications of AI and big data analysis in cyber defence. Section 4 focuses on emerging trends in the AI and big data communities that are relevant in the cyber domain. The cyber defence community should take note of the necessity to perform research in these identified areas. The paper is concluded in Section 5.