Chapter 74
Semantic Technologies and Big Data Analytics for Cyber Defence

Louise Leenen
Cape Peninsula University of Technology, South Africa

Thomas Meyer
University of Cape Town, South Africa

ABSTRACT

The Governments, military forces and other organisations responsible for cybersecurity deal with vast amounts of data that has to be understood in order to lead to intelligent decision making. Due to the vast amounts of information pertinent to cybersecurity, automation is required for processing and decision making, specifically to present advance warning of possible threats. The ability to detect patterns in vast data sets, and being able to understanding the significance of detected patterns are essential in the cyber defence domain. Big data technologies supported by semantic technologies can improve cybersecurity, and thus cyber defence by providing support for the processing and understanding of the huge amounts of information in the cyber environment. The term big data analytics refers to advanced analytic techniques such as machine learning, predictive analysis, and other intelligent processing techniques applied to large data sets that contain different data types. The purpose is to detect patterns, correlations, trends and other useful information. 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 is becoming increasingly popular. However, most automated systems are currently based on syntactic rules. These rules are generally not sophisticated enough to deal with the complexity of decisions required to be made. The incorporation of semantic information allows for increased understanding and sophistication in cyber defence systems. This paper argues that both big data analytics and semantic technologies are necessary to provide counter measures against cyber threats. An overview of the use of semantic technologies and big data technologies in cyber defence is provided, and important areas for future research in the combined domains are discussed.

DOI: 10.4018/978-1-5225-7501-6.ch074

1. INTRODUCTION

The rapid increase in the number and variety of cyber threats, and in the volume of information that has to be processed to provide efficient counter-measures require the ability to perform intelligent search and data integration. Integration of information requires an encoded common vocabulary and shared understanding of the domain. Due to the vast amounts of information pertinent to cybersecurity, automation is required for processing and decision making.

Big data is a term that is used to refer 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.

Semantic technologies is a term that represents a number of different technologies aiming to derive meaning from information. Some examples of such technologies are natural language processing, data mining, semantic search technologies, and ontologies. It should be noted that semantic technologies are not the same as Semantic Web technologies; the latter is a subset of the former. Semantic Web technologies are technology standards from the World Wide Web Consortium (WC3) that are aimed at the representation of data on the Web. Examples of Semantic Web technologies are RFD (Resource Description Framework) and OWL (Web Ontology Language). The Cambridge Semantics group (Bio, n.d.) defines semantic technologies as "…algorithms and solutions that bring structure and meaning to information" and Semantic Web technologies as "…those that adhere to a specific set of WC3 open technology standards that are designed to simplify the implementation of not only semantic technology solutions but other kind of solutions as well".

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 et al., 2001) and Section 1.1 thus provides an overview of ontologies.

This paper gives a brief overview of big data applications in cyber defence (Section 2), and a more thorough overview of application of semantic technologies in the cyber defence domain (Section 3). Section 4 takes a glance at the emerging trends in the semantics 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.