International Journal of Cyber Warfare and Terrorism
Volume 6 • Issue 3 • July-September 2016

Semantic Technologies and Big Data Analytics for Cyber Defence

Louise Leenen, DPSS, CSIR and Cape Peninsula University of Technology, Cape Town, South Africa
Thomas Meyer, CAIR, CSIR Meraka, and Computer Science, University of Cape Town, 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.

KEYWORDS

Automated Systems, Big Data Analytics, Cyber Defence, Decision Making, Semantic Technologies

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,

DOI: 10.4018/IJCWT.2016070105

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.

1.1. Overview of Ontologies

An ontology consists of a shared domain vocabulary and a set of assumptions about the meaning of terms in the vocabulary. A formal definition of an ontology is given by Gruber (1993): a "formal, explicit specification of a shared conceptualisation". An ontology is a technology that enables a formal, shared representation of the key concepts of a specific domain and it provides a way to attach meaning to the terms and relations used in describing the domain.

The main benefits of ontologies are the ability to perform semantic search, provision of a common shared vocabulary and sharing of domain knowledge, and the facilitation of semantic integration and interoperability between heterogeneous knowledge sources. Any satisfactory solution to search and integration problems will have to involve ways of making information machine-processable, a task that is only possible if machines have better access to the semantics of the information.

The information in an ontology is expressed in an ontology language (which are frequently logic-based languages), and then progressively refined. The construction and maintenance of ontologies greatly depend on the availability of ontology languages equipped with a well-defined semantics and powerful reasoning tools. Fortunately there already exists a class of logics, called description logics, that provide for both, and are therefore ideal candidates for ontology languages. The web ontology language, OWL 2.0, which was accorded the status of a W3C (World Wide Web Consortium)

recommendation in 2009, is the official Semantic Web ontology language. OWL was designed to provide a common way to process the content of web information instead of just displaying it. There are a number of tools and environments available for building ontologies.

2. BIG DATA ANALYTICS IN CYBER DEFENCE

Big data analytics in cyber defence focuses on the ability to gather massive amounts of digital information to process, analyse, visualise and interpret results in order to predict and stop cyber-attacks. Advance warning of attacks and threat intelligence are becoming essential in security technologies.

According to the Gartner report (Litan, 2014), big data analytics will play a crucial role in detecting crime and security incidents. The Vice president of Gartner, Avivah Litan, said that big data analytics gives companies faster access to their own data than ever before. It also enables them to integrate different data sources to get an overall picture of threats against their institutions (The Cloud Times, 2014). A Trend Micro white paper on big data security challenges stated that (Trend Micro, 2012): "Successful protection relies on the right combination of methodologies, human insight, an expert understanding of the threat landscape, and the efficient processing of big data to create actionable intelligence".

Teradata sponsored the Ponemon Institute (Ponemon Institute, 2013) to perform an investigation on organisations' cybersecurity defences and their use of big data analytics in their study "Big Data Analytics in Cyber Defense", published in 2013. The report covers a wide range of pertinent issues and gives a good overview of the current awareness and the application of new data management and big data analytics of organisations in the fight against network attacks and other cyber threats. Although 61% of the respondents agreed that launching a strong defence against hackers and other cyber criminals requires their organisations to be able to detect and quickly contain anomalous and potentially malicious traffic in networks, at the time of the investigation, only 35% of respondents' organisations employed these tools. A positive result is that 51% of the organisations had knowledge of big data products that were available and regarded these tools as necessary in the fight against cyber threats. Only 23% of companies regularly applied big data analysis to counter threats but many organisations had plans to incorporate these tools in the future.

The Cloud Security Alliance (CSA, 2013) published a report, "Big Data Analytics for Security Intelligence ", on how the incorporation of big data is changing security analytics by providing new tools for leveraging data from both structured and unstructured sources. In this report, it is mentioned that people now create 2.5 quintillion bytes of data per day. The rate at which data is currently generated is creating a need for new technologies to analyse huge data sets. Big data analytics can be leveraged to correlate different sources in order to get a big picture. For example, financial transactions, log files and network traffic can be analysed to identify suspicious activities. The report points out that the urgency for collaborative research on big data topics is emphasised by the US Federal government's $200 million funding for big data research in 2012 (Lohr, 2012). The report points out that big data analytics can, for instance, advance security intelligence produced by Security Information and Event Management (SIEM) alerts by "reducing the time for correlating, consolidating, and contextualising diverse security event information, and also for correlating long-term historical data for forensic purposes". According to the report, big data tools provide an advantage in:

- More economical storage of large data sets;
- Much faster analysis;
- Being able to analyse and manage unstructured data; and
- Providing cluster computing infrastructures which are more reliable and available.

There are many examples of the use of big data analytics in cyber defence systems but we only mention two:

- **Intel's *Threat Intelligence Exchange* (TIE) system (Marko, 2014):** Multiple systems all share security information detected on one device or system in a centralised big data repository which then informs other devices and systems. Each security system then adapts their policies and controls to block a newly detected threat;
- **IBM's *QRadar Security Intelligence platform* and *IBM Big Data Platform* (IBM, n.d.):** Provide threat and risk detection via an integrated approach that combines real-time correlation analytics across structured and unstructured data, and forensic capabilities for evidence. With this approach it is possible to address advanced persistent threats as well as fraud and insider threats. A wide range of data is analysed over years of activity.

3. CURRENT CYBER DEFENCE APPLICATIONS USING SEMANTIC TECHNOLOGIES

In this section we give an overview of existing application areas of semantic technologies in cyber defence with an emphasis on the development of ontologies.

3.1. Cyber Attack Classification and Prediction

A quick reaction to a network attack is one of the most essential requirements in cyber defence. When a system can identify an ongoing attack and classify the attack, efficient counter-measures can be taken. Balepin et al. (2003) highlighted the need for quick responses with the increase in the speed of computer attacks. Various researchers have developed ontological applications to identify and classify network attacks. A few examples of such ontologies are listed below:

- Bhandari et al. (2014) developed an ontology to perceive the security status of a network;
- A peer-to-peer multi-agent distributed intrusion detection system (Ye et al., 2008);
- A network attack ontology intended to support the automated classification of attacks (van Heerden et al., 2013);
- Salahi & Ansarina (2011) developed an ontology-based system to predict potential network attacks.

3.2. Malware Classification

The classification of malware is a very complex discipline due to the fact that there does not exist clear boundaries for the different groups of malware; characteristics are often shared by different types of malware. Many attributes and state changes have to be considered to detect a piece of malware; this complexity also results in problems with the naming and the classification of malware. Good classification and naming schemes support the sharing of information across organisations, facilitate the detection of new threats, and assist with risk assessment in quarantine and clean-up (Bailey et al., 2007). Bailey et al. also highlight that the complexity of modern malware makes the classification process increasingly difficult, especially in terms of consistency and completeness. Another problem is the rapid increase in the number and diversity of Internet malware. There are a number malware naming schemes, for example the CARO scheme, but there does not exist a commonly accepted standard scheme.

Automated malware detection and classification systems currently classify malware inconsistently across products, and their results tend to be incomplete (Bailey et al., 2007). Some of the available analysis systems are Cuckoo Sandbox, Malwr, VirusTotal, and Yarae. One of the major problems with these classification systems is that they are inconsistent, incomplete and fail to be concise in their semantics (Bailey et al., 2007).

The first recommendation in the JASON cyber report (MITRE, 2010) is that the cybersecurity community should develop vocabularies and ontologies such that a common language and a set of

basic concepts can be developed for a shared understanding. This report was commissioned by the United States Department of Defense.

Mundie and McIntire (2013) state that: "Nowhere in the cybersecurity community is the lack of a common vocabulary, and the problems it causes, more apparent than in malware analysis." Mundie also stated on a podcast (Mundie & Allen): "And in my view, all the other aspects of a science – the statistics, hypothesis testing, etc. – all of that can only be built on top of that shared understanding that the report highlighted."

A growing number of researchers are investigating the use of semantic technologies to develop more efficient malware classification systems (Mundie & MacIntire, 2013; Tafazzoli & Sadjadi, 2008; Huang et al., 2010; Chiang & Tsuar, 2010).

3.3. Military Knowledge Management and Military Ontologies

The modern military environment is faced with an overwhelming amount of information from heterogeneous sources that has to be processed, integrated, interpreted, and exploited in order to gain situational awareness. The development and application of military related ontologies have grown tremendously the past 10 years. Curts and Campbell (2005), stated that "…the sorts of semantic interoperability provided by ontology technology are indispensable" in attempting to improve our understanding of Command and Control (C2).

A number of efforts have been devoted to developing ontologies for military applications and we mention a few of these below. Lombard et al. (2012) developed an ontology for countermeasures against military aircraft. Belk & Noyes (2012) used an ontology to categorise all operations in cyber space. Smith et al. (2009) presented a process for constructing a concise, modular, and extensible core C2 ontology. Their aim was to support interoperability in a military environment by building a core ontology that can be extended with sub-domain ontologies. Nguyen et al. (2010) discussed the development of a set of ontologies for use in messaging systems within the military and first responder C2 applications.

Many military applications use an ontology in military as a supporting system for simulations. Haberlin et al. (2011) developed a simulation to evaluate a Hypothesis Management Engine.

3.4. Other Cyber Defence Applications

A few other application areas in cyber defence for which ontological approaches have been developed are discussed in this section.

Orbst et al. (2012) have done work in support of the development of an ontology of the cybersecurity domain that will enable data integration across disparate data sources. They propose a number of resources for the envisioned ontology that range from domain specific resources, languages, vocabularies, ontologies and schemas. Their ontology is currently focussed on malware but they propose the inclusion of actors, victims, infrastructure, and capabilities.

Ontologies have been developed to support cybersecurity policy implementation: Jansen van Vuuren et al. (2014) developed an ontology to support the implementation of the South African National Cybersecurity Policy. Due to the many role players, functions and relations that are involved in such an implementation, the authors present an ontology to represent the environment in which the policy implementation is to be done. Cuppens-Boulahia et al. (2008) proposed an ontology-based approach to instantiate new security policies to counteract network attacks.

Oltramari & Lebiere (2013) represent requirements for building a cognitive system for decision support with the capability of simulating defensive and offensive cyber operations by employing a semantic approach that includes the use of ontologies. Brinson et al. (2006) created an ontology for the purpose of finding the correct layers for specialisation, certification, and education within the cyber forensics domain.

4. EMERGING RESEARCH AREAS IN THE COMBINED DOMAINS

One of the necessary steps to obtain interoperability is to encourage research disciplines such as the big data and linked data communities to collaborate with the semantics research community. Linked data refers to a way of representing structured data so that it can be interlinked and become enhanced. There is a need for more research to be done in this area: Grobelnik et al. (2012) performed a quick test in 2012 by looking at the number of hits for key words such as "big data" (20 million), "semantic web" (9 million) and "big data & semantic web" (0.3 million). They also searched for the number of appearances of "semantic" in the four leading books published in 2011 on "big data" and found very few incidences.

Janssen & Grady (2013) explored the use of big data technologies augmented by ontologies to improve cybersecurity. They note that these technologies have the potential to revolutionise the handling of large volumes of cyber data. One way in which big data analytics will be effective in the cyber domain is to identify patterns rather than processing collections of pages. Janssen and Grady also maintain that semantic technologies are crucial for the handling of big data sets across multiple domains. Little inroads have yet been made to integrate big datasets. These researchers argue that integration ontologies will have to be developed to provide metadata for browsing and querying: the integrating ontology should automatically construct queries to the big data repository. A significant challenge in using ontologies for automated data analytics across data sets that requires attention is probabilistic reasoning. This is due to the fact that analysis will have to be done under some uncertainty.

Although there are a number of emerging trends in both the semantic research and big data communities, we focus on three main trends relevant for cyber defence: the creation of interoperability and platforms for the sharing of information, the development of global sources of information in specialised sub-domains, and the importance of intelligence-led approache. Scalable reasoning methods and stream reasoning are two emerging areas in the semantics community that should be noted by the cyber defence community. These methods can support the building of more efficient cyber defence systems.

4.1. Cybersecurity Information Sharing, Knowledge Representation and Interoperability

[The] Semantic Web in its most general aim is about interoperability being needed in almost all areas of research and business. (Grobelnik et al. 2012)

Formal models of cybersecurity information, vocabularies, standardised representations, data formats and exchange protocols are required to share cybersecurity information effectively in the cybersecurity community. Significant effort has been made to categorise cybersecurity information and standardise data formats and protocols (Hernandez-Ardieta & Tapiador, 2013). According to Dandurand and Serrano (2013) current practices and supporting technologies limit the ability of organisations to share information securely with trusted partners. These authors give an overview of a number of cybersecurity standards and initiatives that have been developed such as the European Information Sharing and Alert System (EISAS) (ENISA, 2011) and languages and structures developed by the MITRE corporation: Common Vulnerabilities and Exposures (CVE), Common Platform Enumeration (CPE) and others (Martin, 2008). Adoption of standards is improving but recently, subject-matter experts from the RSA organisation stated that:

Data standards for describing and transmitting threat information have advanced significantly, but much progress is needed to extend existing standards and drive wider adoption in vendor solutions. Threat information-sharing and collaboration programs help organizations augment their expertise

and capabilities in detecting and remediating advanced threats, but most sharing programs are hindered by a heavy reliance on manually intensive, non-scalable processes and workflows. (Hartman, 2012)

Janowics & Hitzler (2012) cite the usefulness in publication of own data as one of the examples of the added value of semantics: the creation of intelligent metadata enables researchers to support the discovery and reuse of their data. They also stress the shift from developing increasingly complex software to the creation of metadata, and that smart data will make all future applications more usable, flexible and robust. Ontologies should be used to restrict the interpretation of domain vocabularies towards their intended meaning and reduce the risk of combining unsuitable data and models, something which purely syntactic approaches or natural language representation often fail to do (Kuhn, 2005).

4.2. Global Cyber Attack Detection Systems and Automation

Numerous organisations across the globe detect and gather information regarding cyber-attacks, network intrusions and malware. Standard, shared systems should be developed to collate and encourage information sharing to enable improved protection against cyber events. However, due to the vast amount of information and the speed at which cyber-attacks take place, timely decision making and automated responses are required and the use of ontologies to accomplish this goal is important (Dandurand & Serrano, 2013). A 2008 review of existing security ontologies stated that the security community requires a complete security ontology that addresses insufficiencies in existing ontologies and provides reusability, communication and knowledge sharing (Blanco et al., 2008). Similarly, there should be a standard malware classification system and vocabulary.

Orbst et al. (2012) has made an attempt at creating an ontology for the cyber domain. They are using an initial ontology that is mainly focussed on malware but present a discussion of the development of an ontology for the whole domain. They give a description of the potential ontologies and standards that can be used in the global ontology. These resources include cyber and malware standards, schemas and technologies, foundational or upper ontologies, utility ontologies. An overview of the possible architecture is also given.

Janssen & Grady (2013) also proposed the development of a cyber domain ontology that will contain all knowledge necessary for assessment, decision, planning and response in this domain. They base their proposal on the fact that system awareness currently resides in the minds of large numbers of cyber professionals. This information should be gathered in a single repository. Although it is a daunting task, the researchers argue that the recent successes of ontology engineering and the high stakes in the cybersecurity domain makes it necessary to solve on a national level. This argument can also be applied on an international level in the view of the authors of this paper.

There are issues such as trust and willingness to share which will also have to be addressed.

4.3. Intelligence-Led approaches

Intelligence-led security is depicted by the *Information Age* as one of the 11 trends that will dominate cyber security ion 2016 (Rossi, 2015). Intelligence-led security approaches in cybersecurity will be able to produce better results in terms of tracking security incidents and analysing huge amounts of information. Traditional technologies cannot cope with the rate at which information is generated and are unable to tie together unlinked pieces of information in order to create situational awareness. Real-time monitoring, advance warning and speedy analysis time will support quick reaction to security incidents.

The research director of Gartner, Lawrence Pingree, said the lack of automated intelligence sharing prevents human and business processes from responding to breaches. Pingree also said security systems must become “adaptable based on contextual awareness, situational awareness and controls

themselves can inform each other and perform policy enforcement based on degrees or gradients of threat and trust levels" (Marko, 2014).

According to the Ponemon Institute's "2015 Global Megatrends in Cybersecurity" report (Ponemon, 2015), technology innovation will shift towards big data analytics, forensics and intelligence- based cyber solutions. They predict that the following technologies will gain the most in importance over the next 3 years: encryption for data at rest, big data analytics, SIEM and cybersecurity intelligence, automated forensics tools, encryption for data in motion, next generation firewalls, web application firewalls, threat intelligence feeds and sandboxing or isolation tools.

4.4. Scalable Reasoning Methods

Scalability is a feature of a system that enables it to accommodate growth. The primary purpose of providing meaning to data is to facilitate reasoning about the data, so as to be able to perform sophisticated tasks such as intelligent search and data integration. Reasoning is an expensive computational endeavour, however. One of the major challenges in this regard is the development of scalable reasoning methods. In recent years there have been a number of breakthroughs in the design of scalable ontology languages. The most important of these are the three profiles of the Web Ontology Language OWL 2: OWL 2 EL, OWL 2 DL, and OWL 2 RL (Motik et al., n.d.). All three profiles are sub-languages of OWL 2, each designed expressly for representing a particular class of ontologies. The focus on specific classes of ontologies makes it possible to design reasoning methods with very attractive computational properties. To get a sense of the difference between the three profiles, it is important to understand that there is a distinction to be drawn between data and an ontology, the latter being used to provide meaning to the data.

OWL 2 EL is designed for scenarios in which the ontology is large and complicated, but with fairly small amounts of data underlying it. A representative example of a large OWL 2 EL ontology is the medical ontology SNOMED CT (http://www.ihtsdo.org/snomed-ct/), with more than 300 000 active concepts and more than 1 000 000 relationships between the concepts. With SNOMED being represented as an ontology in OWL 2 EL, modern reasoning methods are able to classify all the concepts in SNOMED CT within a matter of milliseconds – a feat that was considered impossible about 15 years ago.

OWL 2 DL, on the other hand, is designed for cases in which an ontology is relatively small but spans large amounts of data. It is frequently used by employing the ontology as a semantic layer into which large database systems are being plugged. This enables users to query a database through the semantic layer, thereby obtaining truly intelligent responses from the system. The power of OWL 2 DL querying lies in the development of techniques where queries posed through the ontology are rewritten as standard database queries. This makes it possible to exploit existing efficient database querying methods, and has the potential for very fast and efficient querying.

Finally, OWL RL exploits the fact that many domains of interest can be represented using rule-like statements, and adopts existing techniques for reasoning efficiently with rule-based systems. OWL 2 RL is aimed at applications that require scalable reasoning without sacrificing too much expressive power. It is designed to accommodate OWL 2 applications that can trade the full expressivity of the language for efficiency, as well as RDF(S) applications that need some added expressivity. OWL 2 RL reasoning systems can be implemented using rule-based reasoning engines. The ontology consistency, class expression satisfiability, class expression subsumption, instance checking, and conjunctive query answering problems can be solved in time that is polynomial with respect to the size of the ontology. The RL acronym reflects the fact that reasoning in this profile can be implemented using a standard Rule Language.

4.5. Stream Reasoning

Most of the currently available semantic technologies are based on the assumption that information is static. This is, of course, not a realistic assumption, and one of the important trends in this area is

the development of tools able to deal with dynamic information that changes over time. A particularly useful scenario to consider is one where an incremental flow of data is available. Examples of this include data obtained from sensor network monitoring, traffic engineering, RFID tags applications, telecom call recording, medical record management, financial applications, and clickstreams, and are frequently referred to as streams of data. Clearly, information needed for ensuring cybersecurity falls in this category as well. Reasoning over such streams of data is referred to as stream reasoning (Della Valle et al., 2009). The goal of stream reasoning is to draw relevant conclusions and react to new situations with minimal delays. It is needed to support a variety of important functionalities in autonomous systems such as situation awareness, execution monitoring, and decision-making.

What is needed for efficient, intelligent stream reasoning is the provision of the abstractions, foundations, methods, and tools required to integrate data streams and existing reasoning systems, and there is broad consensus that the ability to reason about streaming data to cope with the increasing amount of dynamic data on the web is the next big step in semantic technologies. The research agenda for this challenge has been picked up by a number of research groups internationally (Stuckenschmidt et al., 2010). At its core is the goal to combine existing semantic technologies with data streams in order to perform stream reasoning. Work has been done on the foundations of real-time reasoning on data streams as they become available (Beck at al., 2014). It has also led to alternative abstractions for representing and querying semantic streams of data. Various forms of deductive and inductive stream reasoning have been investigated (Barbieri et al., 2013). In terms of improving the efficiency of stream reasoning methods, the exploitation of the temporal order of data streams has been recognised as a key optimisation method for stream reasoning. In a similar vein, parallelisation and distribution techniques for stream reasoning have been investigated (Albeladi, 2012).

5. CONCLUSION

This paper considers the application of big data analytics and semantic technologies for cyber defence by giving an overview of the current state of affairs, and identifying emerging trends in the combination of these fields. Big data analytics provides the ability to deal with large sets of diverse data, structured and unstructured, in almost real-time while semantic technologies provide the ability to make sense of the resulting information. Semantic technologies allow one to tie together seemingly unrelated pieces of information. The emerging trends also serve as the authors' recommendations for future research areas in the domain.

REFERENCES

Albeladi, R. (2012). Distributed Reasoning on Semantic Data Streams. (2012). *Proceedings of the11th International Semantic Web Conference (ISWC)*, LNCS (*Vol 7650,* pp. 433-436). Springer Berlin Heidelberg.

Bailey, M., Oberheide, J., Andersen, J., Mao, Z. M., Jahanian, F., & Nazario, J. (2007). Automated Classification and Analysis of Internet Malware. *Proceedings of theInternational Symposium on Recent Advances in Intrusion Detection (RAID'07)* (pp. 187-197). Springer Berlin Heidelberg. doi:10.1007/978-3-540-74320-0_10

Balepin, I., Maltsev, S., Rowe, J., & Levitt, K. (2003). *Using specification-based intrusion detection for automated response. Recent Advances in Intrusion Detection* (pp. 136–154). Springer Berlin Heidelberg. doi:10.1007/978-3-540-45248-5_8

Barbieri, D. F., Braga, D., Ceri, S., Valle, E. D., Huang, Y., Tresp, V., & Wermser, H. et al. (2010). Deductive and Inductive Stream Reasoning for Semantic Social Media Analytics. *IEEE Intelligent Systems*, *25*(6), 32–41. doi:10.1109/MIS.2010.142

Beck, H., Dao-Tran, M., Eiter, T., & Fink, M. (2014). Towards a Logic-Based Framework for Analyzing Stream Reasoning. *Proceedings of the3rd International Workshop on Ordering and Reasoning* (*Vol 1303*).

Belk, R., & Noyes, M. (2012). *On the Use of Offensive Cyber Capabilities*. Master's thesis, Harvard Kennedy School. Retrieved from http://www.dtic.mil/docs/citations/ADA561817

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American*, *284*(5), 28–37. doi:10.1038/scientificamerican0501-34 PMID:11341160

Bhandari, P., & Guiral, M. S. (2014). *Ontology Based Approach for Perception of Network Security State. Recent Advances in Engineering and Computational Sciences* (pp. 1–6). RAECS.

Bio, L. F. (n. d.). Semantic Web vs. Semantic Technologies. *Cambridge Semantics*. Retrieved from http://www.cambridgesemantics.com/semantic-university/semantic-web-vs-semantic-technologies

Blanco, C., Lasheras, J., Valencia-García, R., Fernández-Medina, E., Toval, A., & Piattini, M. (2008). A Systematic Review and Comparison of Security Ontologies. *Proceedings of theThird International Conference on Availability, Reliability and Security* (ARES 08) (pp. 813-820). doi:10.1109/ARES.2008.33

Brinson, A., Robinson, A., & Rogers, M. (2006). A Cyber Forensics Ontology: Creating a New Approach to Studying Cyber Forensics. *Digital Investigation*, *35*, 37–43. doi:10.1016/j.diin.2006.06.008

Chiang, H.-S., & Tsuar, W.-J. (2010). Ontology-based Mobile Malware Behavioural Analysis. *Proceedings of theIEEE Second International Conference on Social Computing* (SocialCOM '10).

Cuppens-Boulahia, N., Cuppens, F., de Vergara, J. E. L., & Vazquez, E. (2008). An Ontology-based Approach to react to Network Attacks. *Proceedings of the3rd International Conference on Risks and Security of Internet and Systems(CRiSIS '08)* (pp. 280-305). doi:10.1109/CRISIS.2008.4757461

Curts, R. J., & Campbell, D. E. (2005). Building an Ontology for Command & Control. *Proceedings of the10th International Command and Control Research and Technology Symposium*. McLean, Virginia.

Dandurand, L., & Serrano, O. S. (2013). Towards Improved Cyber Security Information Sharing. *Proceedings of the5th International Conference on Cyber Conflict*. NATO CCD COE Publications, Tallinn.

Della Valle, E., Ceri, S., van Harmelen, F., & Fensel, D. (2009). It's a Streaming World! Reasoning Upon Rapidly Changing Information. *Intelligent Systems*, *9*(6), 83–89. doi:10.1109/MIS.2009.125

ENISA. (2011). EISAS (enhanced) report on implementation. Retrieved from https://www.enisa.europa.eu/activities/cert/other-work/eisas_folder/eisas-report-on-implementation-enhanced

Grobelnik, M., Mladenic, D., & Fortuna, B. (2012). Semantic Web in 10 years. *Proceedings of the11th International Semantic Web Conference (SWC2012) Workshop on "What will the Semantic Web look like in 10 years from now",* Boston, USA.

Gruber, T. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, *5*(2), 199–220. doi:10.1006/knac.1993.1008

Haberlin, R., Da Costa, P. C. G., & Laskey, K. B. (2011). An Ontology for Hypothesis Management in the Maritime Domain. *Proceedings of the16th International Command and Control Research and Technology Symposium*.

Hartman, B. M. (2012). Breaking Down Barriers to Collaboration in the Fight Against Advanced Threats. *RSA Security Brief February 2012*. Retrieved from http://www.emc.com/collateral/industry-overview/11652-h9084-aptbdb-brf-0212-online.pdf

Hernandez-Ardieta, J. L., & Tapiador, J. E. (2013). Information Sharing Models for Cooperative Cyber Defence. *Proceedings of the5th International Conference on Cyber Conflict*. NATO CCD COE Publications.

Huang, H.-D., Chuang, T.-Y., Tsai, Y.-L., & Lee, C.-S. (2010). Ontology-based Intelligent System for Malware Behaviour Analysis. *Proceedings of the2010 IEEE International ConferenceonFuzzy Systems (FUZZ)* (pp. 1-6).

IBM. (n. d.). IBM Security Intelligence with Big Data. *IMB website*. Retrieved from http://www-03.ibm.com/security/solution/intelligence-big-data/

Janowicz, K., & Hitzler, P. (2012). Key Ingredients for Your Next Semantics Elevator Talk. *Advances in Conceptual Modeling, LNCS* (Vol. *7518*, pp. 213–220). Springer Berlin Heidelberg.

Jansen van Vuuren, J., Leenen, L., & Zaaiman, J. (2014). Using an Ontology as a Model for the Implementation of the National Cybersecurity Policy Framework for South Africa. *9th International Conference on Cyber Warfare & Terrorism* (ICCWS 2014) (pp.107-115). Purdue University, Indiana, USA.

Janssen, T. and Grady, N. (2013). Big Data for Combating Cyber Attacks. In *Semantic Technology for Intelligence, Defense, and Security* (STIDS 2013).

Kuhn, W. (2005). Geospatial semantics: Why, of what, and how? Journal on Data Semantics III, LNCS (Vol. 3534, pp. 587–587).

Litan, A. (2014, January 14). Reality Check on Big Data Analytics for Cybersecurity and Fraud. *Gartner*. Retrieved from https://www.gartner.com/doc/2651118

Lohr, S. (2012). New U.S. Research will Aim at Flood of Digital Data. *The New York Times on 29 March, 2012*. Retrieved from http://www.nytimes.com/2012/03/29/technology/new-us-research-will-aim-at-flood-of-digital-data.html?_r=0

Lombard, N., Gerber, A., & van der Merwe, A. (2012).Using Formal Ontologies in the Development of Countermeasures for Military Aircraft. *Proceedings of theEighth Australasian Ontology Workshop*.

Marko, K. (2014, November 9). Big Data: Cyber Security's Silver Bullet? Intel makes the case. *Forbes*. Retrieved from http://www.forbes.com/sites/kurtmarko/2014/11/09/big-data-cyber-security/#5d833caa294e

Martin, R. (2008). Making Security Measurable and Manageable. *Proceedings of theIEEE Military Communications Conference* (MILCOM 2008) *IEEE* (pp. 1-9).

Micro, T. (2012). Addressing Big Data Security Challenges: The Right Tools for Samrt Protection. *Trend Micro.* Retrieved from http://www.trendmicro.com/cloud-content/us/pdfs/business/white-papers/wp_addressing-big-data-security-challenges.pdf

MITRE. (2010). Science of Cybersecurity. *(JSR-10-102) MITRE Corporation*. Retrieved from http://fas.org/irp/agency/dod/jason/cyber.pdf

Motik, B., Cuenca Grau, B., Horrocks, I., Whu, Z., Fokoue, A., & Lutz, C. OWL 2 Web Ontology Language Profiles (2nd ed.). Retrieved from http://www.w3.org/TR/owl2-profiles

Mundie, D., & Allen, J. (n. d.). Using a Malware Ontology to Make Progress Towards a Science of Cybersecurity. Retrieved from http://www.cert.org/podcast/show/20130509mundie.html

Mundie, D., & McIntire, D. M. (2013). An Ontology for Malware Analysis. *Proceedings of theEighth International Conference on Availability, Reliability and Security*. IEEE (pp. 556-558).

Nguyen, D. N., Kopena, J. B., Loo, B. T., & Regli, W. C. (2010). Ontologies for Distributed Command and Control Messaging. *Proceedings of theSixth International Conference on Formal Ontology in Information Systems*.

Oltramari, A., & Lebiere, C. (2013). Towards a Cognitive System for Decision Support in Cyber Operations", Semantic Technology for Intelligence, Defense, and Security (STIDS 2013) (pp.49-56).

Orbst, L. Chase, P. and Markeloff, R. (2012). Developing an Ontology of the Cyber Security Domain. In *Semantic Technology for Intelligence, Defense, and Security* (STIDS 2012).

Ponemon Institute. (2013). Big Data Analytics in Cyber Defense. Retrieved from http://www.ponemon.org/local/upload/file/Big_Data_Analytics_in_Cyber_Defense_V12.pdf

Ponemon Institute. (2015). 2015 Global Megatrends in Cybersecurity. Retrieved from http://www.raytheon.com/news/rtnwcm/groups/gallery/documents/content/rtn_233811.pdf

Rossi, B. (2015, December 4). 11 Trends that will Dominate Cyber Security in 2016. *Information Age*. Retrieved from http://www.information-age.com/technology/security/123460617/11-trends-will-dominate-cyber-security-2016

Salahi, A., & Ansarinia, M. (2011). Predicting Network Attacks Using Ontology-Driven Inference, arXiv prepring arXiv:1304.0913. Retrieved from http://arxiv.org/ftp/arxiv/papers/1304/1304.0913.pdf

Smith, B., Miettinen, K., & Mandrivk, W. (2009). The Ontology of Command and Control. *Proceedings of the14th International Command and Control Research and Technology Symposium*.

Stuckenschmidt, H., Ceri, S., Della Valle, E., & van Harmelen, F. (2010). Towards Expressive Stream Reasoning. *Semantic Challenges in Sensor Networks, Proceedings of Dagstuhl Seminar*.

Tafazzoli, T., & Sadjadi, S. H. (2008). Malware Fuzzy Ontology for Semantic Web. *International Journal of Computer Science and Network Security*, *8*(7), 153–161.

The Cloud Security Alliance (CSA). (2013). Big Data Analytics for Security Intelligence. Retrieved February 10, 2016 from, https://downloads.cloudsecurityalliance.org/initiatives/bdwg/Big_Data_Analytics_for_Security_Intelligence.pdf

The Cloud Times. (2014). By 2016, 15 Percent of Large Global Companies will have Adopted Big Data Analytics for at least one Security or Fraud Detection Use Case. Retrieved from http://www.gartner.com/newsroom/id/2663015

van Heerden, R., Leenen, L., & Irwin, B. (2013). Automated classification of computer network attacks. *Proceedings of theInternational Conference on Adaptive Science and Technology* (ICAST 2013) (pp. 157-163).

Ye, D., Bai, Q., Zhang, M., & Ye, Z. (2008). P2P distributed intrusion detections by using mobile agents. *Proceedings of theSeventh IEEE/ACIS International Conference on Computer and Information Science* (ICIS 08) (pp. 259-265).