Universiti Teknologi MARA
Cawangan Perlis
Jurnal Intelek
Volume 14 Issue 1 (June 2019)

Towards the Big Data and Digital Evidences Integrity

Shekh Abdullah-Al-Musa Ahmed[1*], Nik Zulkarnaen Khidzir[2], Tan Tse Guan[3]

[1] Faculty of Creative Technology and Heritage, Universiti Malaysia Kelantan, Malaysia
[2] Global Entrepreneurship Research and Innovation Centre / Centre of Computing and Informatics Universiti Malaysia Kelantan

[3] Faculty of Creative Technology and Heritage, Universiti Malaysia Kelantan, Malaysia
*Corresponding author: * almusa.c17e002f@siswa.umk.edu.my*

Received Date: 18 January 2019
Accepted Date: 15 March 2019

ABSTRACT

Data plays an important role in the maintenance of collecting and processing huge amounts of information. The more data is stored, the more suitable and essential it is to ensure its safety. A lack of data security can result in great financial failures and reputational damage for any organization. Big data vulnerability comes when unauthorized person gains access to the data. Big data is a synthetic tools which gives us result. This paper displays various problems which are related to Big data problem and opportunity. Nevertheless this paper suggests a BDS model or Big Data Security diagram to provide the big data security. Just as companies start checking out e- Business activities such as sociable media for brand advertising and customer diamond, they have also started turning to data stats as well. Today, it is nearly impossible to find any brand that does not have sociable mass media presence. Looking at the growing of the data means the result of Big data issue will come. However, big data and digital evidences integrity require BDS model to provide the security of data in the organization.

Keywords: *Structured data; Unstructured data; Semi-structured data; Big data Security; Digital evidence; legal requirement; BDS Model*

INTRODUCTION

Actually technology that developed to control large and sophisticated data. Although big data means enormous data, it is a number of large datasets that cannot use traditional processing techniques. Big data is not just a data, alternatively it could be a complete subject matter that involves various tools, techniques and frameworks. In simple words, big data method resolves data conditions that are not controllable using traditional directories and tools. Big Data does not only mean large amount of info but also a sizable amount of data. So, this technology is a strategy to store, process, manage, evaluate and make statement. However, variety data, at the mandatory velocity, acceleration and within the mandatory time to permit real-time evaluation and response (Phillips ,2015).

In cases like this, the characteristics of big data are extremely large levels of data as well as extremely high acceleration of data and high variety of data. As a result, volume refers to the huge amount of information such as TB (Tera Bytes) to PB (Peta Bytes) to EB (Exa Byte) and even more. Whereas speed means how fast data is produced. It is indicating the quantity of data, speed is the velocity of data in and out, and variety selection of data types and resources.

ISSN: 2231-7716

Organized data means rows and columns of information which are very easy to store even in relational databases, for example relational data. Semi-structured data means that it is formatted for some reason and not formatted through rows and columns. One of these is XML data. Whereas un-structured data means the data is not formatted at all. It is not possible to store data in relational directories. For example, music tracks documents, videos, images, sensor data, web data, mobile data, GPS navigation data (Chaowei Yanget et al., 2017). Whereas small data is data that is small enough for individual's knowledge. It really is data in a volume level and format that means it is available, useful and actionable. However, big data means machines data and small data means data around people. Big data amount is big and visually-appealing items represent various areas of large collections of data such as histogram, graphs and scatter plots.

However, the Big data consists of the information made by different devices and applications such as black box data. The black box is an element of helicopter, airplanes, and aero planes etc. which catches the voices of the airline flight team, recordings of microphones and earphones, and the performance information of the aero planes. Sociable mass media data such as Facebook and forums keep information and the views uploaded by thousands of folks around the world. Also, the stock market data retains information about the buy and sell decisions made over the show of different companies created by the clients. From then on another is electric power main grid data keeps information employed by a particular consumer regarding a base train station along with transport data which includes model, capacity, distance and availableness to a car so search engines retrieve plenty of data from different databases.

LITERATURE REVIEW

Big data usually includes data sets with sizes beyond the capability of frequently used software tool to record, curate, manage, and process data within an endurable elapsed time. Big Info philosophy encompasses unstructured, semi-structured and structured data, however the key emphasis is on unstructured data. Generally Big data means large amount of data. Big data requires a pair of techniques and technologies with new kinds of integration to reveal insights from datasets that are diverse, complex along with a massive range (Seetha Raman et al,.2018).

Within just 2001 research survey and related lectures, META Group (now Gartner) determined data development challenges and opportunities as being 3D, such as increasing volume level (amount of data), acceleration (speed of information in and out), and variety (range of data types and sources). Gartner, and most of the industry, then continued to use this "3Vs" model for explaining big data. In 2012, Gartner updated their description the following: "Big data is high-volume, high velocity and high-variety information possessions that demand cost-effective, intensifying varieties of information refinement that permit increased of insight, decision making, and process automation. The Gartner's definition of the 3Vs is still broadly used, in addition to arrangement with a consensual explanation that states that "Big data represents the knowledge and assets indicated by such a higher volume, velocity and variety to require specific technology and analytical methods for their transformation into Value". In addition, a brand new Sixth, sixth v "Veracity" is added by some organizations to demonstrate it, revisionism challenged by some industry authorities. The 3Vs have been expanded to other complementary characteristics of big data.

Digital Evidence for Big Data

In Big data forensics involves the storage, identification, removal, documentation, and interpretation of computer mass media for evidentiary and origin cause research. Big data Digital evidence might be expected for a variety of computer crimes and misuses.

The term digital forensics was formerly used as a synonym for computer forensics but has expanded to cover investigation of most devices capable of keeping digital data. With origins in the personal processing revolution of the past due 1970s and early on 1980s, and it was not until the early 21st hundred years that national policies appeared. Before the 1980's, crimes involving computer systems were addressed using existing laws and regulations. Along with identifying direct evidence of up against the legislation, big data digital forensics can be used to feature evidence to specific potential foods, confirm alibis or transactions, determine intent identify resources or authenticate documents. Anna (Kostikova et al,.2017).

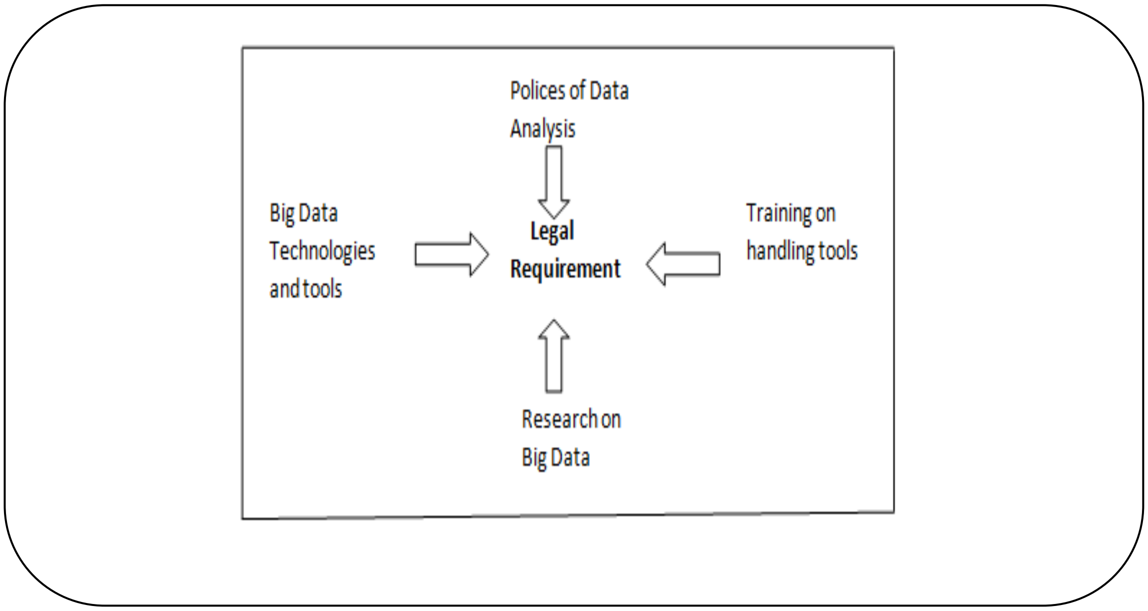


Figure 1 : Showing the process digital Evidence for big data

When it comes to digital evidence, examiner always gather evidence to confirm a suspect determined of a crime or broken a company policy. From the collected evidence which can be offered in court, the suspect's data or information is then checked out from the evidence maintained on a different computer or in the cloud hosting computer. This is usually followed by an accepted method to put together a good example. The server computer systems or cloud computer can contain information that helps legislation enforcement to look for the chain of incidents which could in the end trigger a crime. Nevertheless, to obtain the evidences that can lead to a conviction, Law observance officers should follow proper procedure when obtaining the Big Data digital proof. Since digital proof can be easily modified by any malicious person, there are cases where employees misused resources which could cost companies millions of dollars. Rather than working, these employees use them for Net surfing, sending personal emails, using company computers for personal tasks. Big data can be organized, unstructured and semi-structured data and also large in quantity, variety and velocity of information. Collecting the sensitive digital proof is a significant challenge (Gianluca Lax,. 2015).

Big Data and Information Security

Big data digital evidence information is an important property. Big data information can be categorized into different categories. This is certainly typically done in order to control access to the information in several ways, depending on their importance, its level of sensitivity, and the vulnerability to theft or misuse. Companies typically choose to release more resources to manage information that has higher level of sensitivity.

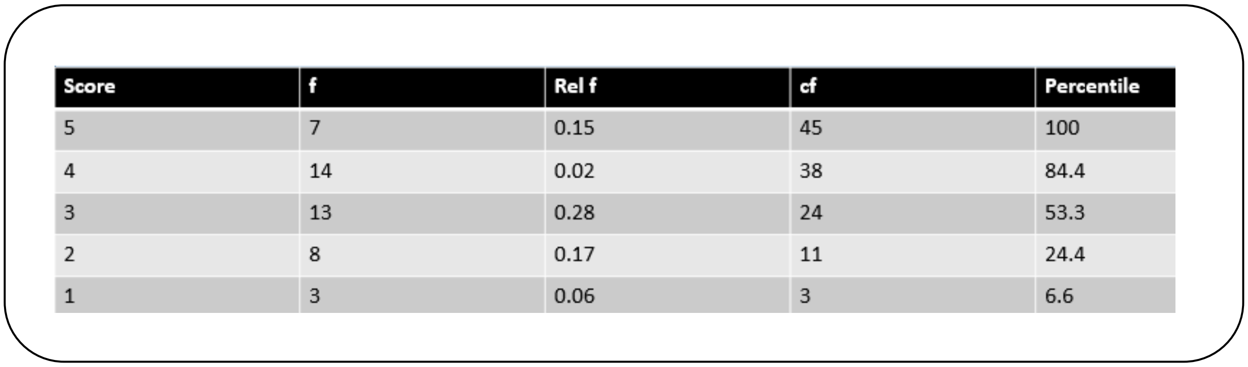
Big data organizations straighten out information in several ways to be able to differently deal with aspects of its handling, such as labelling (whether headers, footers, and watermarks specify how it should be handled), distribution (who gets to see it), duplication (how copies are manufactured and handled), release (how it is provided to outsiders), storage (where it is kept), encryption (if required) disposal (whether it is shredded or wiped) and methods of transmitting (such as e mailbox, fax, print, and mail). Big data information is suitable for internal only use and is usually intended to be seen by employees, contractors, and service providers, but not by the public.

Generally in information Security, the protection of information is throughout the life span of the data, from the initial creation of the information on before the final convenience of the information. The information must be protected while in motion and while sleeping. During its life span, information may be completed through many different information handling systems and through many different parts of information processing systems. There are many different ways the data and information systems can be threatened. In order to fully protect the information during its lifetime, each factor of the information handling system will need to have their own security mechanisms. The building up, layering as well as overlapping of security procedures is called security in depth. As opposed to a metal sequence, which is famously only as strong as it is most basic link, the defence-in-depth is targeted at a framework where, if one protective measure fail, other steps will continue to keep provide security.

METHODOLOGIES OF DIGITAL EVIDENCE

Computer system forensics involves obtaining and analysing digital information which will be the evidence in civil, criminal or management cases. Federal Bureau of investigation Computer Analysis and Response Team (CART) formed in 1984 to cope with the increasing quantity of cases involving digital facts. However, the Fourth Amendment to the U.S Constitution protects every person's rights to be secured in their personal information, house and property from search and seizure. Thus, when such information is needed, search warrants are required. Regarding to Seymour Bosworth and the book "Computer Security Handbook" described the laptop or computer forensics is the investigates data which can be retrieved from a pc's hard disk or other storage media (Venter, 2018). Whereas Network forensics produces information about how precisely or exactly a criminal or an attacker gained access to a network. And data recovery is the recovering of information that was deleted by oversight or lost during an ability surge or storage space crash. A small survey questionnaire was done for digital forensics examination analysis and the survey of data of frequency table as below:

Table 1 : Frequency of digital forensics examination analysis , (N=45)



Score	f	Rel f	cf	Percentile
5	7	0.15	45	100
4	14	0.02	38	84.4
3	13	0.28	24	53.3
2	8	0.17	11	24.4
1	3	0.06	3	6.6

In the year of 1970s, electronic crimes were increasing, especially in the financial sector where most police force authorities failed to know enough about computer systems and did not know where and when they should ask the right questions, as well as to protect the evidences for trial. Nevertheless, in 1980s PCs gained reputation and different OSs appeared and disk Operating-system (DOS) was available, and also Forensics tools were simple, and most were produced by Government departments. In the mid-1980s, Xtree Gold made an appearance on the market that recognized file types and retrieved lost or removed files. Then, Norton DiskEd soon followed to become the best tool for locating deleted record or record or data file. Within the 1987, Apple produced the Mac SE, a Macintosh, with an exterior EasyDrive hard disk with 60 MB of storage space. In the early nineties, tools for computer forensics were available and International Association of Computer Investigative Specialists (IACIS) were trained on software for forensics investigations such as IRS created search-warrant programs. the Professional Witness for Mac pc was the first commercial GUI software for computer forensics created by ASR Info. In the early 1990s, professional experienced the Mac computer that could recover deleted documents and fragments of removed files, had large hard storage posed problems for captivation. Other software such as iLook and Access Data Forensic Tool set (FTK) were also available for large data digital investigation.

BIG DATA DIGITAL EVIDENCE INTEGRITY

Big data vulnerably comes when unauthorized person gains access to the information. Big data tools produce an end result, thus, when used in a court, big data digital evidence integrity comes under the legal advice as other varieties of data whereas courts do not usually require more stringent rules. In the United States, the Federal Rules of Evidence are employed to examine the admissibility of digital evidence. The United Kingdom PACE and Civil Proof acts have similar tips and many more countries such as Malaysia, Bangladesh have their own laws and rules. US federal laws restrict seizures of any digital items where there with only clear digital certain value. Laws dealing with big data digital evidence are concerned with two issues: integrity and genuineness. Integrity is making sure the act of requisitioning and acquiring big data digital evidence does not improve the evidence (either the original or perhaps the copy). Genuineness refers to the capability to confirm the integrity info, to get example that the imaged media matches the original evidence (Laura Kinget al,.2016). The simplicity which big data digital proof can be altered means that documenting the sequence of custody from the crime scene, through evaluation and, ultimately, to the court, which is important to determine the authenticity of massive data digital data. Hence, the issue of big data digital evidence come up when different country has its own law for large data evidence collection.

BIG DATA DIGITAL EVIDENCE CHALLENGES

Computer technology Investigators Network (CTIN) meets monthly to discuss issues that law enforcement and corporations face regarding the big data digital evidence. High Technology Crime Investigation Association (HTCIA) exchanges information about techniques related to big data digital evidence investigations and security. Big data digital investigations and forensics falls into two distinct categories such as public investigations and private or corporate investigations. The public investigations of big data digital evidence requires government departments to be responsible for legal investigations and prosecution, organizations must observe legal guidelines, law of search and seizure, protects, and liberties under the law of all people, including potential proof. Private or corporate and business investigations of big data digital evidence are packaged with private companies, non-law-enforcement federal government agencies, and legal professionals. Non-public corporate and business investigations also require a lawsuit disputes investigation are usually conducted in municipal circumstances (Carrie Andersonet al,.2017).

Within a criminal arrest case, a suspect is attempted for a criminal offense such as burglary, homicide, or molestation. Computers and systems are tools which are often used to commit crimes. A large amount of states has added specific language to criminal arrest rules to define criminal offenses concerning computers. An unlawful circumstance commences when someone detects evidence of an illegitimate act. When the complainant makes an allegation, an accusations or supposition of truth, the law enforcement officer will then interview the complainant and makes a report about the crime. Police provides a record of indications to offences that have been determined recently investigators delegate, pick up, and process the info related to the complaint. Pursuing building a case, the info is turned over to the prosecutor. However, the challenging parts arise when someone maliciously uses the big data result set from different platform. So, the big data digital proof challenges are:

- Reliance on ICTs
- Number of users
- Missing mechanisms of control
- International dimensions
- Independence of location and presence at the crime site
- Automation Resources
- Anonymous communications
- Encryption technology

THE SECURITY OF BIG DATA

The database security identifies a collection of set up procedure, standard, guidelines and tools which can be used to protect data from theft, improper use and unwanted intrusion, activities and disorders. It offers with the permission and access to the data structure and the data contained within it. Nevertheless, with the increased usage of web-affiliated, mobile and cloud-based applications, sensitive data has become accessible from different websites. These websites are highly vulnerable to hacking, especially if they are low-cost or free. The article Trust Management, by Blaze et al., shows a unified method to specifying and interpreting security policies, credentials, and human relationships that allows direct documentation of security-critical actions. . Credentials describe specific delegations of trust among public keys; unlike traditional certificates, which bind secrets to names, trust-management recommendations bind keys directly to authorizations to perform specific tasks. Whereas other articles such as - Introduction to Public key technology by Richard et al. recommended using characteristics apart from identity, attested to by known authorities in digital certificates, as a most basic for authorization on the World Wide Web (WWW). Blaze et al. introduced a complementary strategy to authorization based on delegation of privileges. Rivest et al. by Method for Obtaining Digital Signatures and Public-Key Cryptosystems introduced a scheme that provides a way to introduce names and bind them to public keys controlled by individuals and groups, which greatly facilitates identifying authorized principals electronically.

PROPOSED BDS MODEL

The overall problem of trust management, introduced by Blaze et al. Its application is actually specific so that identity-based documentation can be used in trust management security policy. However, trust management is completed by the application itself. Hence, in the case of BDS (Big Data Security) Model, we proposed there would be System Manager (SM) who will handle the entire system, registers data source users and providers, and issues keys. Then, the Big data Service agency manages the access to database resources. Finally, the users, who is a part of a group of authorized principals can access group resources. The good thing about BDS Model is that - this method supports the user confirmation so the integrity of database without retrieving original data from the storage space and probability detects data corruptions.

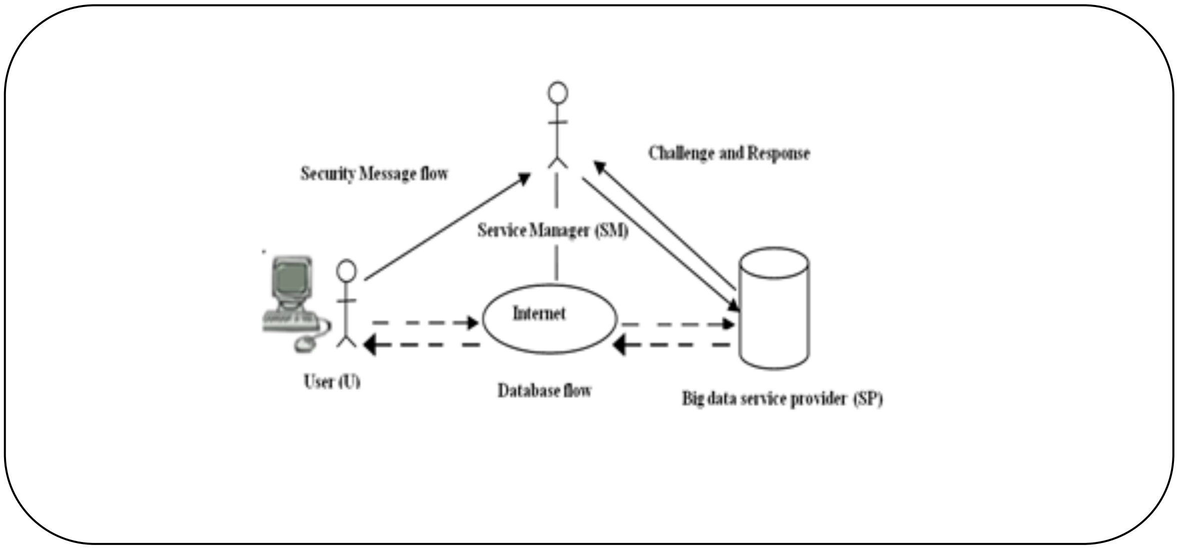


Figure 2 : Showing the BDS Model

In the BDS design, at first the user encrypts data to ensure the confidentiality, then, compute metadata over protected data. Later, the verifier can use remote data integrity checking protocol to verify the integrity. The verifier should able to discover any changes on data kept in database. Consequently, almost any data would be stored such as organized data means data that is in the form of rows and column. And it is very easy to store even in relational databases. By way of example: Relational data. Then can store Semi-Structured data means data that is set up in some way. Nevertheless, it is not set up in the form of rows and columns. For example: XML data. And Un-Structured Data means data is not organized in any respect. It is not possible to store big volume of data in Relational Databases. For example, audio files, videos, photographs, sensor data, web info, mobile data, GPS information.

CONCLUSION

This paper has presented the interrelated concepts, issues, difficulties and opportunities of big data and its structures as well as suggesting a model BDS (Big data security). Big data technology is an analytical tool which can create result set. One of the most challenging part is who would be assessing the database result set. Reliability issue appears when vulnerable user uses this information from different platform. In such case, we recommend BDS (Big data security) model. Consequently, the user here is anonymous to the big data service agency, in order to guarantee good security. Here, end user will get his key from the service manager (SM). When service administrator generates key for the user, the user will put structured, unstructured and semi structured data. Then, the service agency will use big data technology and get the result set. Following that, the user will get the database result arranged. While the case service agency and service manager all are trusted party, but the user can still act as an anonymous. Here we can consider two varieties threats - external and internal threats. Since the user acts here as an anonymous actor so the internal threat will be reduced. In addition to the case of external risk the service agency and service manager will be responsible to minimize the external threat.

REFERENCES

Abdullah, A., Yue, X., Taizan, C., 2017. An empirical study on the susceptibility to social engineering in social networking sites: the case of Facebook. Journal European Journal of Information Systems 26, p. 661.

Anna Kostikova, A. S., Galina Sorina & Sergey Spartak, 2017. Big Data: a loop or a challenge for human morality: mapping Russian tradition in philosophy and methodology. Russian Journal of Communication, 9, p. 252.

Alan, E., Mark, P., Carrie, M., 2006. The Evolution of Computer Forensic Best Practices: An Update on Programs and Publications. Journal of Digital Forensic Practice, Volume 1, p. 3.

Alistair, S., 2013. Social Engineering in the Information Age. An International Journal, 2, p. 67.

Carrie Anderson, G. C., Christopher Donaldson. 2017. Digital humanities and tourism history. Journal of Tourism History, 9, p. 246.

Chaowei Yang, Q. H., Zhenlong Li, Kai Liu, Fei Hu, 2017. Big Data and cloud computing: innovation opportunities and challenges. International Journal of Digital Earth, 10, p. 13.

Clay, P., Tom, L., Paul, B., 2015. The Impact of Organizational Commitment on Insiders' Motivation to Protect Organizational Information Assets. Journal of Management Information Systems, 32, p. 179.

Gavin, W., Elizabeth, D., 2010. What Security Professionals Need to Know About Digital Evidence. Information Security Journal: A Global Perspective, 19, p. 124.

Gianluca Lax, 2015. Digital Document Signing: Vulnerabilities and Solutions. Information Security Journal: A Global Perspective, 24, p. 1.

Hoskisson Phillips, W., 2015. An analytical journey towards big data. Journal of Decision Systems, 24, 87.

Jemal, A., 2014. User preference of cyber security awareness delivery methods. Behaviour & Information Technology, 33, p. 237.

John, D., Tejaswini, H., Mindy. K., 2014. Understanding Employee Responses to Stressful Information Security Requirements: A Coping Perspective. Journal of Management Information Systems, 31, p. 285.

Laura King, James F. S., Paul Cooke, 2016. Experiencing the Digital World: The Cultural Value of Digital Engagement with Heritage. Heritage & Society, 9, p. 76.

Seetha Raman, N. P., Indu Niranjan, Ujjwal Ranjan, Krishna Moorthy, Ami Mehta, 2018. Impact of big data on supply chain management. International Journal of Logistics Research and Applications, 21, p. 579.

Tommie, W., Singleton, D., Aaron J., 2008. The Potential for a Synergistic Relationship Between Information Security and a Financial Audit. Information Security Journal: A Global Perspective, 17, p. 80.

Venter, 2018. On digital forensic readiness in the cloud using a distributed agent-based solution: issues and challenges. Australian Journal of Forensic Sciences, 50, p. 209.

Victor, R., Kebande, I., Venter, 2018. Novel digital forensic readiness technique in the cloud environment. Australian Journal of Forensic Sciences, 50, p. 552.

Wiebke, A., 2009. Agents, Trojans and tags: The next generation of investigators. International Review of Law, Computers & Technology, 23, p. 99.