WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Systematic Analysis of Big Data Security Framework Established on Encryption

Sheikh Md Zubair Md Zahoor

Former Research Scholar, Computer Science, OPJS University, Churu, Rajasthan, India

A B S T R A C T

From a security standpoint, the largest problem for the big data age is the safeguarding of user privacy and data security. The analysis of big data is significantly complicated, particularly if the data is unstructured. If we want to query the information, we need to convert the unstructured data to structured data. Companies will need to figure out which types of information in their big data are associated to data security, as well as the related technologies, such as Key Management for Access Control, Attribute-Based Access Control, and Secure Search for Parties Involved, Searchable Encryption, and Secure Data Processing.

Keywords: Big data security, Security framework, Encryption

1. Introduction

Everyone's behavior was recorded in the big data age. Financial institutions, governmental institutions, and perhaps our neighbors have placed a tag on us. Banks will not finance to individuals with weak credit records; organizations will be wary of frequent job hopping when hiring; those who move partners will be stigmatized with their hearts and rejected by others around them. People with bad credit records, such as high-speed rails and airlines, are limited in their ability to construct a more reliable system. We categorize the data based on numerous traits to make everybody stand in amazement of ethics and dignity with due diligence. There are two types of data: public data and private data.

Big data processing has grown important to numerous government and commercial applications as the volume of data produced, gathered, and processed by computing systems continues to expand at an incredible rate.

Private data about individuals or business secrets are frequently found in the gathered data, which might inflict significant damage if they gets into the wrong hands. Criminal networks are establishing subterranean marketplaces where stolen private information may be bought and sold. For surveillance and competitive edge, government spy agencies are attacking individual, commercial, and rival government systems.

From a security standpoint, the largest problem for the big data age is the protection of user privacy and data security. Because big data often comprises large volumes of personal identifiable information (PII), user privacy is a major problem.

Massive amounts of data were once stored in relational databases that were highly structured. You understood precisely where to seek and how to get sensitive data, such as a patient's medical record or a bank user's credentials, if you needed it. In relational databases, it was also easy to remove any traceable information. Big data complicates the procedure, particularly if the data is unstructured. If we want to query the information, we need to convert the unstructured data to structured data. Companies will need to identify which parts of their big data are critical, and they will need to securely separate this data to maintain safety.

* *Corresponding author.*
E-mail address: contactsmz@gmail.com

2. Overview of the Literature

We will first go over the literature in this subject before moving on to the next portion. Custers I et al. compared data privacy laws and policies across the EU in 2017. Due to disparities in legal competencies, available funds and staff, policies, and cultural variables, they pointed out that significant differences exist in (the levels of) enforcement by various data protection authorities in this study.

Researchers have also offered a number of approaches for securing massive data. In 2018, Singh et al. presented a cybersecurity architecture for fog computing and cloud-of-things environments to recognize malevolent edge devices. The proposed cyber-security framework in this analysis employs three technologies to detect suspicious edge devices in a fog computing environment: Markov model, Intrusion Detection System (IDS), and Virtual Honeypot Device (VHD).

Tankard et al. proposed big data security in 2012. They asserted in this study that the centralized nature of big data stores generates threats for businesses to overcome, necessitating controls to be established around the data itself instead of the applications and systems that store the data.

Tankard et al. presented encryption as the foundation of big data security in 2017, and this study elaborated that data security is a necessary for any business to secure its operations. To guarantee that sensitive data is securely protected, encryption should be a critical component of every big data environment.

Lafuente emphasized the big data security challenge in 2015. Any organization adopting a big data project will benefit from anonymizing and encrypting data, as well as putting in place suitable access control and surveillance capabilities, impactful policies, and governance frameworks.

Privacy-aware big data analytics as a service for public health policies in smart cities was presented by Anisettia et al. [6] in 2018. We emphasize on balancing life quality and privacy protection in smart cities inside this research by proposing a novel Big Data-assisted public policy making approach. The presented strategy is built on Big Data Analytics as a service of privacy-by-design.

3. Controls and Security of Big Data

3.1 Data Security Fundamentals

We'll start by going over some of the fundamental goals and methods of big data security. First, we'll look over some of the common adversary models that are utilized to identify the data threats which ought to be handled. Then we go over some of the most frequent security goals.
We have to identify the big data security goals now that we've established the attacker we want to protect against. Confidentiality, integrity, and availability, commonly known as the CIA trinity, are the three most essential security goals.

Confidentiality: The purpose of confidentiality is to keep all critical information hidden from an enemy. Traditional notions of secrecy, on the other hand, guarantee that an enemy will discover nothing about the sensitive material except its length. In big data applications, confidentiality is essential to ensure that sensitive data is not shared with the wrong people.
Integrity refers to the goal of being able to detect any illegal data alteration. In other words, a malevolent adversary should not be able to change such data without leaving a trace. This is critical in order to ensure the accuracy of data acquired in big data applications.
Availability: The purpose of availability is to be able to access one's data and computing resources at all times. An adversary should not be able to prevent important data or resources from being accessed. This is a critical security aim in big data processing since the sheer volume and velocity of the data makes ensuring continuous access challenging. Today's big data systems, on the other hand, often guarantee availability by non-cryptographic mechanisms such as replication and so on.

Governments will be better able to identify the numerous dangers they face, the possible avenues of attack, and the actors who might perpetrate them by analyzing big data.

3.2.1 Data Storing and Processing

Companies are increasingly seeking to aggregate data from a larger range of stores and applications in order to provide more context and so raise the worth of the data as they seek to generate value from it.

Another issue that could arise is regulatory compliance, particularly in the area of data protection. Some jurisdictions have more rigorous data storage and processing laws than others. To ensure that they remain in compliance with the restrictions that they face, organizations must carefully assess the legal implications of where they keep and process data.

3.2.2 Classify the Information

Big data projects also provide security benefits. When centralizing data repositories, businesses should first classify the data and implement relevant controls, such as establishing retention periods based on the legislation they are subject to. This will enable enterprises to weed out data that is of little value or no longer needs to be preserved so that it may be disposed of and is no longer vulnerable to theft or lawsuit requiring the presentation of records.

Another advantage in terms of security is that big swaths of data may be mined for security events like malware, spear phishing attempts, or fraud like account takeovers.

Data classification, on the other hand, may be a difficult, lengthy, and time-consuming process – one that many people have struggled with when trying to deploy data classification-based technologies. This is exacerbated by the fact that the data is frequently obtained from outside sources, making verification of its accuracy difficult. The first step for businesses is to determine what data is useful to them. They risk spending time and money analyzing data that will bring little or no value if they spend too much time obtaining all of the available data. Businesses must also conform to industry standards and government regulations, ensuring that records are preserved and archived for the time periods provided, and that data is protected in accordance with some standards' rules. Businesses should search for automated database and network discovery systems that can scan networks and identify all data assets to make the classification process easier.

Data classification is also crucial for identifying and protecting sensitive information. Data classification will aid in determining which information is the most sensitive and where it is kept. This should not be left to the IT department alone; line-of-business workers are more likely to comprehend the sensitivity of the data they work with. It should also contain compliance officers, who are responsible for staying up to date on new regulatory needs. This isn't a one-time event; it should be done on a regular basis.

3.3 Developing a Utility Program

The volume of big data generated and maintained by most enterprises and organizations can be a serious barrier, with traditional technologies needing weeks or more to search such vast amounts of data - most of which is unstructured.

Before beginning any large data management project, businesses and organizations must first locate and identify all data sources in their network, including where they came from, who developed them, and who has access to them. This should be a company-wide effort that entails locating and indexing data, with involvement from security and risk management, as well as legal and policy departments. This must also be a constant process in order to discover not only current data, but also new data as it is created across the network.

Organizations should aim to build or amend data handling rules as they move through the data classification process, such as defining what types of data must be stored and for how long, where they should be stored, and how data will be accessible when needed. Users will not be able to create their own data stores that are not under the jurisdiction of the IT department if such restrictions are enforced. For managing enormous amounts of data, data warehouses are a popular technology. Most, on the other hand, store data in a relational format, which works well for structured data, and as information is increasingly taken from a wide range of sources beyond typical enterprise applications, unstructured data makes up a large amount of data included in big data warehouses.

We built a utility program for converting unstructured data into structured data using Python in order to make processing and analysis of unstructured data easier, and a patent has been filed for this.

3.4 Big Data Security Controls

Controls should be shifted nearer to the data store and the data itself, instead of being placed at the network's edge, in order to provide better control over big data sets. The data management department must be precise enough to guarantee that only those authorized to access data can do so, restricting sensitive information against being infiltrated.

Controls should also be put up according to the concept of least privilege, particularly for individuals with higher access levels, such as administrators. As a result, organizations can pick who can examine the data or, in the case of an administrator, provide them physical access; but, if they tried to read the data, the process would have prevented decryption. Any multi-silo environment where any type of electronic data is stored benefits greatly from such an approach.

Access controls should be regularly checked and updated as employees change roles in the organization to ensure that they do not amass excessive rights and privileges that could be exploited. This can be accomplished utilizing already-in-use technology in many businesses, such as database activity monitoring systems, whose capabilities are being improved by numerous vendors to handle unstructured data in big data environments. Other valuable tools include SIEM (Security Information and Event Management) solutions, which collect log data from a wide range of network applications. Many manufacturers, like AlienVault, are expanding their SIEM solutions to include Network Analysis and Visibility (NAV) features, which monitor and analyze network data to check for potential attacks and malicious insider abuse, and are extremely scalable across big networks. Metadata analysis, packet capture analysis, and flow analysis are all available as NAV add-ons to SIEM solutions. Further efforts have been done in the instance of AlienVault in order to link the studied data and make proactive decisions in order to avoid or stop the intrusion. Another significant security aspect is ensuring that data is preserved as needed and disposed of when no longer needed, so that the business does not manage excessively huge volumes of data and the chance of sensitive data being breached is decreased. This can also be decreased by employing techniques like as encryption, tokenization, and data masking to render sensitive data unreadable, allowing only those with the keys to unlock the data to do so. Once the data has been appropriately categorised, this work becomes considerably easier. However, the legal department should be engaged in the formation of data retention and disposal policies to assure that they meet industry standards and government legislation.

3.4.1 Anonymized Data

It is critical to anonymize data in order to ensure that privacy concerns are handled. To maintain the data's security, it should be ensured that all sensitive information is deleted from the collection of records.

While generating information for big data, businesses and organizations must strike the correct balance between data utility and privacy. Before the data is kept, it should be sufficiently anonymized, erasing any user-specific identifiers. This can be a security issue in and of itself, as deleting unique identifiers may not be enough to ensure that the data remains anonymous. Following de-anonymization processes, the anonymized data could be compared to other publicly available data. As a result, data should be effectively encrypted in spite to being anonymous.

3.4.2 Encrypt Data

Enterprises and organizations will be required to encrypt their data when storing it. Since encryption has been the primary means of keeping data safe so far. At the same time, we need to identify external dangers using intrusion detection techniques. "It's critical to have a threat intelligence system in place so that increasingly complex assaults may be detected and organizations can respond to recognized security risks." One of the issues is that data cannot be sent encrypted by users if the cloud needs to execute operations on it. Fully Homomorphic Encryption (FHE), that enables data stored in the cloud to perform operations over encrypted data to create new encrypted data, has been proposed as a solution by certain experts. When the data is decrypted, the results will be identical to those obtained when the operations were performed on plain text data. As a result, the cloud will be able to perform operations on encrypted data without having access to the plain text data underneath.

3.4.3 Access Control and Monitoring

Data protection will rely heavily on adequate access control methods. Access control has generally been given by operating systems or apps that limit access to information, but if the system or application is hacked, all of the information is typically exposed. A better technique is to encrypt the data and allow decryption only if the entity attempting to access it has been granted access by an access control policy. Traditional Intrusion Detection Systems (IDS), firewalls, and application layer implementations are commonly used in big data systems to limit access to data. A big data project's security must also include real-time security monitoring. Organizations must keep a close eye on access to big data to guarantee that no unwanted access occurs. A threat intelligence system is also necessary to ensure that more complex assaults are recognized and that enterprises are able to respond to known security risks such as malware, vulnerabilities, flaws, and so on.

Another important challenge is how to create information ownership. If the data is kept in the cloud, a trust boundary between the data owners and the data storage owners should be established. This should not be presumed to be part of a contract, thus any agreements must include this as a factor.

3.4.4 Encryption as the Basis and Premise of Big Data Security

Big data refers to massive data sets that have resulted from enterprises' spectacular expansion in the volume of data they gather, produce, analyze, distribute, and store. By analyzing large data sets, new insights into how data patterns are linked can be discovered, allowing for better-informed decision-making that can boost competitiveness and promote innovation.

Databases, data warehouses, log and event files, security controls such as intrusion prevention systems, and user-generated data from sources such as emails and social media posts are all used in big data sets. The data collected can be structured, such as in a database's columns, or unstructured, such as in a word processing document.

3.4.4.1 The Information Stripping

All of this data is put into a centralized big data management system, which allows for data correlation and analysis. Much of the data will be very sensitive, including customer, employee, and supplier information, financial data, intellectual property, and a wide range of other data. Organizations face a variety of risks as a result of data breaches, including intellectual property theft, income loss, and reputational damage. Financial penalties and other sanctions for non-compliance with legislation requiring high levels of protection for sensitive data are among the potential dangers. "Finding all potentially sensitive information and understanding links across data sets is increasingly complex." It's also tough to keep track of who has access to sensitive data." Particularly where security policies are inconsistant, are applied differently un traditional and big data contexts, and access must be regulated across so many heterogeneous data sources. Security is a critical factor when creating large data analysis initiatives for reasons like these.

3.4.4.2. Secure Data through Encryption

Any data security policy, including big data projects, should include encryption and key management. Encryption reduces the dangers of data compromise substantially. The following are crucial for data protection in big data environments, according to ENISA (European Union Agency for Network and Information Security):

- To maintain data security and integrity, encrypt data in transit and at rest.
- Given the large number of devices that must be protected, make sure you have a proper encryption key management system.
- Think about how long you need to keep data - due to the nature of some data, data protection legislation may compel you to dispose of it after a particular amount of time.
- Create databases with confidentiality in mind — any sensitive information, for example, might be separated into distinct fields and easily filtered and/or encrypted.

Databases, spreadsheets, word documents, presentations, and archives should all have sensitive information encrypted. Data may leave the corporation at some time, either to be shared among employees and business partners or to be stored in the cloud and accessible via mobile devices. Whenever data is

transferred out of an organization, it is critical that the encryption keys remain within the business to prevent anyone from improperly accessing the keys and decrypting and reading the data. If the keys aren't kept safe, cloud service provider workers could gain access to data, or the data could be demanded by government authorities, typically without the knowledge of the data's owner. Encryption keys should not be stored with encrypted data, as this increases the risk of hackers gaining access to the data.

While encryption will aid secure data from misuse, fraud, and loss, it is also critical to maintain the capacity to undertake big data analysis. In order to ensure this, IBM has released certain best practice guidelines. It is suggested that data be masked both to protect the actual data from theft or loss and to provide a functioning alternative for times when the real data is not required, in order to improve privacy. Data that is sensitive can be hidden at the source or at the big data platform. Unstructured data found in textual, graphical, and form-based documents should be redacted to prevent misuse.

3.4.4.3. Integrated Security Platform

Encryption must be pervasive when working with large data environments that touch so many sections of the enterprise. This necessitates its provision as a platform with granular controls, powerful encryption, and centralized management that encompasses all data sources used in big data research. As all sensitive data sources will be included in the encryption program, this will help to improve efficiency and relieve security worries, as well as make compliance easier to achieve. It will ensure that policies are applied consistently, decreasing the amount of administrative work required for encryption.

Encryption solutions should be combined with other security controls, such as endpoint security, to safeguard large data settings, which is especially important considering the amount of access provided by mobile devices in most enterprises. As more IoT devices are deployed, this will become increasingly significant, providing useful data sources for big data environments. Integration with other security controls, such as intrusion prevention systems and firewalls, can help to lower the risk of big data breaches and the identification of any network threats.

They were once regarded as valuable primarily for compliance reasons, but they have since proven their worth in terms of providing actionable intelligence for better decision-making and security readiness and protection.

3.4.5 Data Protection Policies and Legislation

Because big data is such a novel idea, there is currently no widely accepted list of best practices in the security field. The use of the internet, media attention, and law and regulation should all be focused on big data security.

3.4.5.1 Use of the Internet

Several key regulations on the use and perception of personal data were chosen to offer the overall backdrop and setting in this research while focusing more explicitly on informational privacy and personal data protection.

3.4.5.2 Attention from the media

Privacy and personal data protection should receive more media attention. The so-called Big Brother Awards, which are annual 'awards' for persons, companies, or government agencies whose actions violate privacy, may also be a good indicator of media attention. Obviously, media attention for privacy and privacy abuses may be the cause or outcome of the Big Brother Awards.

3.4.5.3 Legislation and Regulations

We need sectoral legislation to protect the processing of personal data even further. Health care, telecommunications, banking, criminal justice, and the public sector are examples of typical industries. In these sectors, there is sectoral legislation. In many cases, there are guidelines, codes of conduct, or other types of soft law controls on the handling of personal data even if there is no sectoral legislation. Many regulations that require impacted parties and authorities to be notified in the event of a breach provide a safe harbor if the stolen data has been securely encrypted, obviating the need for reporting. Even if a legislation does not give this safe harbor, the use of encryption will be taken into account when the precautions put in place by an entity are evaluated, potentially lowering the penalty that may be levied.

4. Big Data Security and Related Technologies

Encryption and access control are two fundamental technologies in large data security. The distribution and management of keys is a challenge. To safeguard data in storage, we now turn to cryptographic approaches. These approaches' main purpose is to control access to data held in potentially untrustworthy repositories. That is, we want to provide authorized parties access to the data they require while guaranteeing that unauthorized parties, such as outsiders attempting to acquire access or harmful insiders within the company administering the repository, are unable to access important information. In this section, we'll look at systems like file systems where data is stored in blocks that are stored and retrieved using a unique identification. We want authorized parties to be able to get data by its identifier in these systems, but we don't need to allow complicated search queries to retrieve subsets of the data.

4.1 Block Storage and Access Control in a Secure Environment

4.1.1 Access Control Key Management

Key management entails producing and disseminating cryptographic keys to system users in such a way that only authorized parties have access to sensitive data. Many commercially available, standardized ways for generating and managing keys are included in most modern systems for regulating access to data in this way. A trusted key management server is often used to manage all of the system's keys and deliver the appropriate keys to authorized parties. We mostly talked about a cryptographic approach called broadcast encryption or group keying, which allows a data owner to encrypt data and send it to a specific group of recipients without relying on a trusted key manager. This is especially significant in huge data applications, because storage may be managed by an untrustworthy repository without access to a trusted key manager.

4.1.2 Access Control Based on Attributes

Key management-based solutions, such as the ones mentioned above, have a built-in constraint. It is important to know the identities (and keys) of all authorized users in order to share data with them. This is a difficulty in large systems or systems with many organizational structures (as is frequent in big data architectures where data is collected, stored, and used across multiple contexts), because the data owner is unlikely to know the names of all authorized users. A technique known as attribute-based access control is an alternate way to access control (ABAC) in such situations. Data is encrypted in ABAC, together with a policy that describes the attributes of individuals who are authorized to view the data. Users are given keys for the traits they have, and they can only access data if and when those attributes are approved. This provides for data access control without knowing the whole list of people with the appropriate permissions.

4.2 Secure Search

Security and access control for block storage where data may only be retrieved by its unique identifier were discussed in the preceding section. Users rarely want to obtain all available data in a large data system; instead, they usually just want to retrieve a subset of the data based on some search criteria. We'll now shift gears and talk about cryptographic strategies for enabling secure search, which enables complicated, database-style queries on stored data.

4.2.1 Involved Parties

The conventional model of searchable encryption considers three parties: the owner of the data, the querier who wants to learn more about the data, and the server who does the bulk of the storage and processing work. There may be numerous data owners or queriers in various contexts, such as the publish-subscribe and email scenarios described above.

Some cases just have two parties, which we refer to as the querier and the server. For example, in the cloud outsourcing scenario above, a client serves as both the initial data owner and the subsequent querier, whereas in the secure database scenario, the data owner handles all server processing herself. We may conceive of these circumstances as instances of collusion4 between two of the three parties in our three-party framework: in the cloud outsourcing scenario, the data owner colludes with the querier, and in the second case, the data owner colludes with the server.

4.2.2 Searchable Encryption

The seminal work on searchable encryption [14] was published in 2000 by Song, Wagner, and Perrig. They presented a cryptographic protocol for searching over encrypted data in a two-party environment (with a querier and a server) that had the following four properties:

- Encryption with provable security: The untrusted server holds ciphertext material that has been encrypted using a semantically safe encryption algorithm, preventing the server from learning anything about the plaintexts.
- Controlled searching: The untrusted server can only do searches that the querier has approved. The server is unable to conduct searches on her own.
- Song et al. support keyword searches with hidden queries. The keyword is not learned by the server during a query. (Goh later formalizes this concept as "selected keyword security," which is defined similarly to semantic security.)
- Query isolation: While the server does learn which records are returned to the querier, it does not learn anything else about plaintexts.

Furthermore, the Song et al. procedure achieves good results by sticking to a few simple rules that are followed by future works: The search techniques are straightforward, and they're similar to their unsecure equivalents. Instead of using slower public-key cryptography, the system uses faster symmetric-key encryption. In fact, Song et al. only use symmetric-key cryptography, and techniques that follow their lead are known as searchable symmetric encryption, or SSE. While public-key cryptography is used in some of the future tasks mentioned below to give additional functionality, it is used as infrequently as possible.

Song et al., like most unprotected database search solutions, improve query performance by pre-compiling an index that maps keywords to records that match the query. While Song et al research's focuses on non-indexed search, this observation foreshadows many of the future SSE studies.

4.3 Secure Data Processing

We explored doing secure searches on encrypted data in the last section. However, in many big-data applications, merely retrieving stored data is insufficient. Rather than returning the raw data, it is preferable to run analytic computations on the data and only deliver the results of these computations. We emphasize the importance of protecting data and computations even while they are being processed. Homomorphic encryption (HE), verifiable computation (VC), secure multi-party computing (MPC), and functional encryption are the four cryptographic techniques described in this section (FE).

All of these can be utilized to safely outsource data processing in various contexts. Homomorphic encryption enables for calculation on encrypted data while keeping confidentiality; it can be used to outsource sensitive data processing to another entity that can be trusted to conduct the computation correctly but not learn the data. Verifiable computation enables for data processing to be outsourced to another entity that is allowed to learn the data but not trusted to do the computation correctly. To provide both input and output confidentiality as well as computer integrity, homomorphic encryption and verifiable computation can be used together. Secure multiparty computing allows many parties to do a distributed computation on sensitive inputs while guaranteeing the confidentiality of each party's inputs from all other parties and assuring that the computation was accurately performed.

5. Conclusion

The fundamental problem posed by big data is determining how to locate sensitive information within the unstructured data stream. Organizations must ensure that sensitive data is isolated, and they must be able to demonstrate that they have suitable processes in place to do so. The data that organizations acquire should be subjected to a risk assessment. They should assess whether they are collecting any personal information about their customers that should be kept private, and implement procedures to preserve their data and their clients' right to privacy.

If the data is to be shared with other organizations, the method should be explored. Deliberately published data that turns out to be infringing on privacy can have a major reputational and financial impact on a company. Anyone storing or processing data on a third-party cloud provider must ensure that the service complies with legislation. Regional regulations governing the management of client data should also be carefully considered by businesses.

Big data users will need to implement systems that will allow them to manage and safeguard their data successfully. Big data can benefit from traditional information life cycle management, which ensures that data is not stored after it is no longer required. Big data will also be subject to policies governing availability and recovery times. Organizations must, however, take into account the volume, velocity, and complexity of big data while managing and implementing their information life cycle management and policies.

At first glance, big data can appear to be a security nightmare. This does not have to be the case. Security must always be viewed as a method rather than a product. While there are numerous "big data solutions" available, it is critical for businesses to remember the process component. By all accounts, products can aid in the management of the organization's huge data collection. Companies will be considerably better positioned to handle data in a way that balances its value with the all-important element of consumer privacy if they consider the phases indicated above throughout the big data journey.

While encryption should be the foundation of any organization's data security, it is insufficient on its own. It should instead be strongly connected with other security systems.

Endpoint security, network security, application security, and physical security systems are all examples of controls that are increasingly being run across IP-based networks.

These techniques, we believe, will become an intrinsic element of the big data ecosystem as security becomes a crucial necessity for sensitive big data processing. We hope that this chapter's discussion will improve awareness of the most up-to-date technologies and protections for securing big data. We believe that a greater knowledge of data science and cryptography, as well as stronger collaboration between the two fields, will be vital to the future of big data processing.

REFERENCES

1. Sohal, A.S., Sandhu, R., Sood, S.K., Chang, V.: *A Cybersecurity framework to identify malicious edge device in fog computing and cloud-of-things environments. Comput*. Secur. 74, 340–351 (2018)
2. Tankard, C. (2012) : Digital Pathways. Big-data-security. Network-Security
3. The 2011 IDC digital universe. IDC (2011). www.emc.com/collateral/about/news/idc-emc-digital-universe-2011-infographic.pdf. Accessed 1 June 2012
4. Big data: the next frontier for innovation, competition and productivity. McKinsey Global Institute (2011). www.mckinsey.com/Insights/MGI/Research/Technology/Innovation/Big-data-The-next-frontier-for-innovation. Accessed 11 June 2012
5. Softserve Big Data Analytics Report. Softserve (2016). www.softserveinc.com/en-us-newsroom/knowledge-centre/softserve-big-data-analytics-report/. Accessed 11 May 2017
6. 2016 Global Encryption Trends Study. Thales e-Security/Ponemon Institute. www.thales-esecurity.com/knowl-edge-base/analyst-report-s/global-encryption-trends-study. Accessed 13 May 2017

7. Big data and infosecurity. Varonis (2012). http://blog.varonis.com/big-datasecurity/. Accessed 11 June 2012
8. Big data gap. MeriTalk (2012). www.meritalk.com/big-data-gap. Accessed 13 June 2013
9. Kajeepeta, S.: Strategy: Hadoop and big data. InformationWeek (2012). http://reports. informationweek.com/abstract/81/8670/Business-Intelligence-Information-Management/ strategy-Hadoop-big-data.html. Accessed 13 Apr 2014
10. Tankard, C.: Digital Pathways. Encryption as the cornerstone of big data security. Netw. Secur. (2017)
11. Lafuente, G.: The big data security challenge. Netw. Secur. 2015, 12–14 (2015)
12. Anisettia, M., Ardagnaa, C., Bellandia, V., et al.: Privacy-aware big data analytics as a service for public health policies in smart cities. Sustain. Cities Soc. 39, 68–77 (2018)
13. Custers, B., Dechesne, F., Sears, A.M., Tani, T., van der Hof, S.: A comparison of data protection legislation and policies across the EU. Comput. Law Secur. Rev. Int. J. Technol. Law Pract. (2017). https://doi.org/10.1016/j.clsr.2017.09.001
14. Mandiant, February 2013. http://intelreport.mandiant.com/Mandiant-APT1-Report.pdf
15. Song, D.X., Wagner, D., Perrig, A.: Practical techniques for searches on encrypted data. In: 2000 IEEE Symposium on Security and Privacy, Berkeley, California, USA, 14–17 May 2000, pp. 44–55. IEEE Computer Society (2000)