Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2016, Article ID 9175418, 15 pages
http://dx.doi.org/10.1155/2016/9175418

Research Article

A Grey Theory Based Approach to Big Data Risk Management Using FMEA

Maisa Mendonça Silva,[1] Thiago Poleto,[2] Lúcio Camara e Silva,[1] Ana Paula Henriques de Gusmao,[2] and Ana Paula Cabral Seixas Costa[2]

[1]*Technology Centre, Department of Management Engineering, Universidade Federal de Pernambuco, Rodovia BR 104, Km 62, Nova Caruaru, 55002-960 Caruaru, PE, Brazil*
[2]*School of Engineering, Centre for Technology and Geosciences, Department of Management Engineering, Universidade Federal de Pernambuco, Caixa Postal 5125, 52.070-970 Recife, PE, Brazil*

Correspondence should be addressed to Maisa Mendonça Silva; maisa.ufpe@yahoo.com.br

Received 18 March 2016; Revised 3 July 2016; Accepted 26 July 2016

Academic Editor: Gang Kou

Big data is the term used to denote enormous sets of data that differ from other classic databases in four main ways: (huge) volume, (high) velocity, (much greater) variety, and (big) value. In general, data are stored in a distributed fashion and on computing nodes as a result of which big data may be more susceptible to attacks by hackers. This paper presents a risk model for big data, which comprises Failure Mode and Effects Analysis (FMEA) and Grey Theory, more precisely grey relational analysis. This approach has several advantages: it provides a structured approach in order to incorporate the impact of big data risk factors; it facilitates the assessment of risk by breaking down the overall risk to big data; and finally its efficient evaluation criteria can help enterprises reduce the risks associated with big data. In order to illustrate the applicability of our proposal in practice, a numerical example, with realistic data based on expert knowledge, was developed. The numerical example analyzes four dimensions, that is, managing identification and access, registering the device and application, managing the infrastructure, and data governance, and 20 failure modes concerning the vulnerabilities of big data. The results show that the most important aspect of risk to big data relates to data governance.

1. Introduction

In recent years, big data has rapidly developed into an important topic that has attracted great attention from industry and society in general [1]. The big data concept and its applications have emerged from the increasing volumes of external and internal data in organizations and it differs from other databases in four aspects: volume, velocity, variety, and value. Volume refers to the amount of data, velocity refers to the speed with which data can be analyzed and processed, variety describes the different kinds and sources of data that may be structured, and value refers to valuable discoveries hidden in large datasets [2]. The emphasis in big data analytics is on how data is stored in a distributed fashion that allows it to be processed in parallel on many computing nodes in distributed environments across clusters of machines [3].

Given the significance that big data has for business applications and the increasing interest in various fields, relevant works should be mentioned: [4] argued that consumer analytics lies at the junction of big data and consumer behavior and highlights the importance of the interpretation of the data generated from big data. Reference [5] examined the role of big data in facilitating access to financial products for economically active low-income families and microenterprises in China. Reference [6] investigated the roles of big data and business intelligence (BI) in the decision-making process. Reference [7] presented a novel active learning method based on extreme learning machines with inherent properties that make handling big data highly attractive. Reference [8] developed a selection algorithm based on evolutionary computation that uses the MapReduce paradigm to obtain subsets of features from big datasets. Reference [9] discussed the

advancement of big data technology, including the generation, management, and analysis of data. Finally, [10] described a brief overview of big data problems, including opportunities and challenges, current techniques, and technologies.

Big data processing begins with data being transmitted from different sources to storage devices and continues with the implementation of preprocessing, process mining and analysis, and decision-making [6]. Much of this processing takes place in parallel, which increases the risk of attack, and how best to guard against this is what big data management seeks to do [11].

Over the last few years, several researchers have proposed solutions for mitigating security threats. In [12], a taxonomy of events and scenarios was developed and the ranking of alternatives based on the criticality of the risk was provided by means of event tree analysis combined with fuzzy decision theory. Reference [13] developed a mathematical model to solve the problem according to the risk management paradigm and thereby provided managers with additional insights for making optimal decisions. There has also been research on the use of large network traces for mitigating security threats [14].

However, research analyzing the risks associated with big data is lacking. Moreover, from this perspective, information security measures are becoming more important due to the increasingly public nature of multiple sources. Hence, many issues related to big data applications can be addressed first by identifying the possible occurrences of failure and then by evaluating them. Consequently, this paper proposes the use of a specific Failure Mode and Effects Analysis (FMEA) method and Grey Theory, which allows for risk assessment at the crucial stages of the big data process. Both mathematical rigor, which is necessary to ensure the robustness of the model, and the judgments of those involved in the process, given the subjective characteristics of the types of assessments made, are considered in this model. This paper contributes to the literature in the following aspects. First, it offers new insights into how the different characteristics of big data are linked to risk in information security. Second, it provides a model risk analysis based on a multidimensional perspective of big data risk analysis.

The first section of the paper discusses big data and information security issues. Then, the discussion that follows relates to existing methodologies for information security and background information, which are necessary for developing the proposed approach. Next, we introduce the methodology and present a real case that illustrates how the methodology validates the proposed approach. Finally, the discussion presents the limitations of the research, suggested areas for further study, and concluding remarks.

2. Background

2.1. Big Data and Methodologies for Risk Management. As mentioned before, big data has different characteristics in terms of variety, velocity, value, and volume compared to classic databases. Consequently, big data risk management is more complex and is becoming one of the greatest concerns in the area of information security. Currently, another important point is that data availability and confidentiality are two top priorities regarding big data.

Recently, several works relating to big data and security have been published. Reference [15] proposed a new type of digital signature that is specifically designed for a graph-based big data system. To ensure the security of outsourced data, [16] developed an efficient ID-based auditing protocol for cloud data integrity using ID-based cryptography. In order to solve the problem of data integrity, [17] proposed a remote data-auditing technique based on algebraic signature properties for a cloud storage system that incurs minimal computational and communication costs. Reference [18] presented a risk assessment process that includes both risk arising from the interference of unauthorized information and issues related to failures in risk-aware access control systems.

There are many methods and techniques with respect to big data risk management. Table 1 lists and briefly describes qualitative methodologies for risk analysis.

Some approaches based on quantitative methods have also been proposed. Reference [19] presented an approach to the risk management of security information, encompassing FMEA and Fuzzy Theory. Reference [20] developed an analysis model to simultaneously define the risk factors and their causal relationships based on the knowledge from observed cases and domain experts. Reference [21] proposed a new method called the Information Security Risk Analysis Method (ISRAM) based on a quantitative approach.

As can be seen, the purpose of big data security mechanisms is to provide protection against malicious parties. Hence, researchers have also identified several forms of attacks and vulnerabilities regarding big data. Reference [22] investigated key threats that target VoIP hosts. Reference [23] analyzed the impact of malicious servers on different trust and reputation models in wireless sensor networks. Reference [24] examined a cloud architecture where different services are hosted on virtualized systems on the cloud by multiple cloud customers. Also, [25] outlined a discussion of the security and privacy challenges of cloud computing.

In this context, attacks themselves are becoming more and more sophisticated. Moreover, attackers also have easier access to ready-made tools that enable exploitation of platform vulnerabilities more effectively. For these reasons, the security risks of high volumes of data from multiple sources, complex data sharing, and accessibility-related issues arise in a big data environment. Therefore, there is an increasing need to develop and create new techniques for big data risk analysis.

2.2. Failure Mode and Effects Analysis (FMEA). FMEA was first proposed by NASA in 1963. The main objective of FMEA is to discover, prevent, and correct potential failure modes, failure causes, failure effects, and problem areas affecting a system [31]. According to FMEA, the risk priorities of failure modes are generally determined through the risk priority

Table 1: Qualitative methodologies for risk analysis.

Methods and techniques	Description and process	Author
CCTA risk analysis and management method (CRAMM)	Comprises three stages; the first two stages identify and analyze the risks to the system and the third stage recommends how these risks should be managed.	[26]
Expert system for security risk analysis and management (RAMeX)	Proposes examining the risk assessment portion of the risk management process in seven steps: define the problem, identify threats, determine the probability of occurrence, identify existing security, assess the business impact, assess security countermeasures, and generate report.	[27]
Facilitated risk analysis process (FRAP)	The process involves analyzing one system of the business operation at a time and convening a team of individuals who have business information needs and technical staff who have a detailed understanding of potential vulnerabilities of the system and related controls.	[28]
Information risk analysis methodologies (IRAM)	Provides three phases; first phase: conduct a comprehensive assessment of the business impact and determine the business security; second phase: assess threat and vulnerability of incidents occurring in a system; third phase: control selection.	[29]
Operationally critical threat, asset, and vulnerability evaluation (OCTAVE)	Organized into four phases: develop understanding of risk to the business, create a profile of each information asset that establishes clear boundaries and identify its security requirements, identify threats to each information asset, and mitigate this risk.	[30]

Table 2: Severity rating scale.

Rating	Effect	Severity of effect
10	Hazardous without warning	Failure is hazardous and occurs without warning; it suspends operation of the system and/or involves noncompliance with government regulations.
9	Serious	Failure involves hazardous outcomes and/or noncompliance with government regulations or standards.
8	Extreme	Big data is inoperable with loss of primary function; the system is inoperable.
7	High	The big data has severely affected performance but functions; the system may not operate.
6	Significant	The performance of big data is degraded; comfort or convenience functions may not operate.
5	Moderate	A moderate effect on the performance of big data; the product requires repair.
4	Very low	A small effect on the performance of big data; the product does not require repair.
3	Minor	A minor effect on the performance of the big data or system.
2	Very minor	A very minor effect on the performance of the big data or system.
1	None	No effect.

number (RPN), which assesses three factors of risk: occurrence (O), severity (S), and detection (D). Then, the RPN is defined by [32]

$$\text{RPN} = \text{O} \times \text{S} \times \text{D}. \quad (1)$$

Based on [33, 34], the classic proposal uses the 10-point linguistic scale for evaluating the O, S, and D factors. This scale is described in Tables 2, 3, and 4 for each risk factor. The failure modes with higher RPNs, which are viewed as more important, should be corrected with higher priorities than those with lower RPNs.

The FMEA method has been applied to many engineering areas. Reference [35] extended the application of FMEA to risk management in the construction industry using combined fuzzy FMEA and fuzzy Analytic Hierarchy Process (AHP). Reference [36] described failures of the fuel feeding system that frequently occur in the sugar and pharmaceutical industries [37]. Reference [38] proposed FMEA for electric power grids, such as solar photovoltaics. Reference [39] presented a basis for prioritizing health care problems.

According to [40], the traditional FMEA method cannot assign different weightings to the risk factors of O, S, and D and therefore may not be suitable for real-world situations. For these authors, introducing Grey Theory to the traditional FMEA enables engineers to allocate the relative importance of the risk factors O, S, and D based on the research and their

Table 3: Occurrence rating scale.

Rating	Description	Potential failure rate
10	Certain probability of occurrence	Failure occurs at least once a day or almost every time.
9	Failure is almost inevitable	Failure occurs predictably or every three or four days.
8 7	Very high probability of occurrence	Failure occurs frequently or about once per week.
6 5	Moderately high probability of occurrence	Failure occurs about once per month.
4 3	Moderate probability of occurrence	Failure occurs occasionally or once every three months.
2	Low probability of occurrence	Failure occurs rarely or about once per year.
1	Remote probability of occurrence	Failure almost never occurs; no one remembers the last failure.

Table 4: Detection rating scale.

Rating	Description	Definition
10	No chance of detection	There is no known mechanism for detecting the failure.
9 8	Very remote/unreliable	The failure can be detected only with thorough inspection and this is not feasible or cannot be readily done.
7 6	Remote	The error can be detected with manual inspection but no process is in place, so detection is left to chance.
5	Moderate chance of detection	There is a process for double checks or inspection but it is not automated and/or is applied only to a sample and/or relies on vigilance.
4 3	High	There is 100% inspection or review of the process but it is not automated.
2	Very high	There is 100% inspection of the process and it is automated.
1	Almost certain	There are automatic "shut-offs" or constraints that prevent failure.

experience. In general, the major advantages of applying the grey method to FMEA are the following capabilities: assigning different weightings to each factor and not requiring any type of utility function [41].

References [32, 33] pointed out that the use of Grey Theory within the FMEA framework is practicable and can be accomplished. Reference [42] examined the ability to predict tanker equipment failure. Reference [43] proposed an approach that is expected to help service managers manage service failures. Thus, Grey Theory is one approach employed to improve the evaluation of risk.

2.3. Grey Theory. Grey Theory, introduced by [44], is a methodology that is used to solve uncertainty problems; it allows one to deal with systems that have imperfect or incomplete information or that even lack information. Grey Theory comprises grey numbers, grey relations (which this paper uses in the form of Grey Relational Analysis, GRA), and grey elements. These three essential components are used to replace classical mathematics [45].

In grey system theory, a system with information that is certain is called a white system; a system with information that is totally unknown is called a black system; a system with partially known and partially unknown information is called a grey system [46]. Reference [47] argued that, in recent days, grey system theory is receiving increasing attention in the field of decision-making and has been successfully applied to many important problems featuring uncertainty such as supplier selection [48, 49], medical diagnosis [50], work safety [40], portfolio selection [51], and classification algorithms evaluation and selection [52].

According to [53], a grey system is defined as a system containing uncertain information presented by a grey number and grey variables. Another important definition is that of a grey set X (of a universal set U), which is defined by its two mappings $\underline{\mu_X}(x)$ and $\overline{\mu_X}(x)$ as follows:

$$\begin{aligned} \underline{\mu_X}(x) &: x \longrightarrow [0,1], \\ \overline{\mu_X}(x) &: x \longrightarrow [0,1], \end{aligned} \tag{2}$$

where $\overline{\mu_X}(x) \geq \underline{\mu_X}(x)$, $x \in X$, $X = R$, and $\overline{\mu_X}(x)$ and $\underline{\mu_X}(x)$ are the upper and lower membership functions in X, respectively.

A grey number is the most fundamental concept in grey system theory and can be defined as a number with uncertain information. Therefore, a white number is a real number $x \in \mathbb{R}$, and a grey number, written as $\otimes x$, refers to an indeterminate real number that takes its possible values from within an interval or a discrete set of numbers. In other words, a grey number, $\otimes x$, is then defined as an interval with a known lower limit and a known upper limit, that is, as $\otimes x\ [\underline{x}, \overline{x}]$. Supposing there are two different grey numbers

denoted by $\bigotimes x_1$ and $\bigotimes x_2$, the mathematical operation rules of general grey numbers are as follows:

$$\begin{aligned}
&\bigotimes x_1 + \bigotimes x_2 = \left[\underline{x_1} + \underline{x_2}, \overline{x_1} + \overline{x_2}\right],\\
&\bigotimes x_1 - \bigotimes x_2 = \left[\underline{x_1} - \overline{x_2}, \overline{x_1} + \underline{x_2}\right],\\
&\bigotimes x_1 \times \bigotimes x_2 = \left[\min\left(\underline{x_1}\,\underline{x_2}, \underline{x_1}\,\overline{x_2}, \overline{x_1}\,\underline{x_2}, \overline{x_1}\,\overline{x_2}\right),\right.\\
&\quad \left.\max\left(\underline{x_1}\,\underline{x_2}, \underline{x_1}\,\overline{x_2}, \overline{x_1}\,\underline{x_2}, \overline{x_1}\,\overline{x_2}\right)\right],\\
&\bigotimes x_1 \div \bigotimes x_2 = \left[\underline{x_1}, \overline{x_1}\right] \times \left[\frac{1}{\underline{x_2}}, \frac{1}{\overline{x_2}}\right],\\
&k \times \bigotimes x_1 = \left[k\underline{x}, k\overline{x}\right].
\end{aligned} \tag{3}$$

GRA is a part of Grey Theory and can be used together with various correlated indicators to evaluate and analyze the performance of complex systems [54, 55]. In fact, GRA has been successfully used in FMEA and its results have been proven to be satisfactory. Compared to other methods, GRA has competitive advantages in terms of having shown the ability to process uncertainty and to deal with multi-input systems, discrete data, and data incompleteness effectively [55]. In addition, [41] argues that results generated by the combination of Grey Theory and FMEA are more unbiased than those of traditional FMEA, and [42] claims that combining Fuzzy Theory and Grey Theory with FMEA leads to more useful and practical results.

GRA is an impact evaluation model that measures the degree of similarity or difference between two sequences based on the degree of their relationship. In GRA, a global comparison between two sets of data is undertaken instead of using a local comparison by measuring the distance between two points [56]. Its basic principle is that if a comparability sequence translated from an alternative has a higher grey relational degree between the reference sequence and itself, then the alternative will be the better choice. Therefore, the analytic procedure of GRA normally consists of four parts: generating the grey relational situation, defining the reference sequence, calculating the grey relational coefficient, and finally calculating the grey relational degree [55, 57]. The comparative sequence denotes the sequences that should be evaluated by GRA and the reference sequence is the original reference that is compared with the comparative sequence. Normally, the reference sequence is defined as a vector consisting of $(1, 1, \ldots, 1, \ldots, 1)$. GRA aims to find the alternative that has the comparability sequence that is the closest to the reference sequence [43].

2.4. Critical Analysis. Big data comprises complex data that is massively produced and managed in geographically dispersed repositories [63]. Such complexity motivates the development of advanced management techniques and technologies for dealing with the challenges of big data. Moreover, how best to assess the security of big data is an emerging research area that has attracted abundant attention in recent years. Existing security approaches carry out checking on data processing in diverse modes. The ultimate goal of these approaches is to preserve the integrity and privacy of data and to undertake computations in single and distributed storage environments irrespective of the underlying resource margins [11].

However, as discussed in [11], traditional data security technologies are no longer pertinent to solving big data security problems completely. These technologies are unable to provide dynamic monitoring of how data and security are protected. In fact, they were developed for static datasets, but data is now changing dynamically [64]. Thus, it has become hard to implement effective privacy and security protection mechanisms that can handle large amounts of data in complex circumstances.

In a general way, FMEA has been extensively used for examining potential failures in many industries. Moreover, FMEA together with Fuzzy Theory and/or Grey Theory has been widely and successfully used in the risk management of information systems [12], equipment failure [42], and failure in services [43].

Because the modeling of complex dynamic big data requires methods that combine human knowledge and experience as well as expert judgment, this paper uses GRA to evaluate the level of uncertainty associated with assessing big data in the presence or absence of threats. It also provides a structured approach in order to incorporate the impact of risk factors for big data into a more comprehensive definition of scenarios with negative outcomes and facilitates the assessment of risk by breaking down the overall risk to big data. Finally, its efficient evaluation criteria can help enterprises reduce the risks associated with big data.

Therefore, from a security and privacy perspective, big data is different from other traditional data and requires a different approach. Many of the existing methodologies and preferred practices cannot be extended to support the big data paradigm. Big data appears to have similar risks and exposures to traditional data. However, there are several key areas where they are dramatically different.

In this context, variety and volume translate into higher risks of exposure in the event of a breach due to variability in demand, which requires a versatile management platform for storing, processing, and managing complex data. In addition, the new paradigm for big data presents data characteristics at different levels of granularity and big data projects often encompass heterogeneous components. Another point of view states that new types of data are uncovering new privacy implications, with few privacy laws or guidelines to protect that information.

3. The Proposed Model

In this paper, an approach to big data risk management using GRA has been developed to analyze the dimensions that are critical to big data, as described by [65], based on FMEA and [31, 32]. The approach proposed is presented in Figure 1.

The new big data paradigm needs to work with far more than the traditional subsets of internal data. This paradigm incorporates a large volume of unstructured information, looks for nonobvious correlations that might drive new

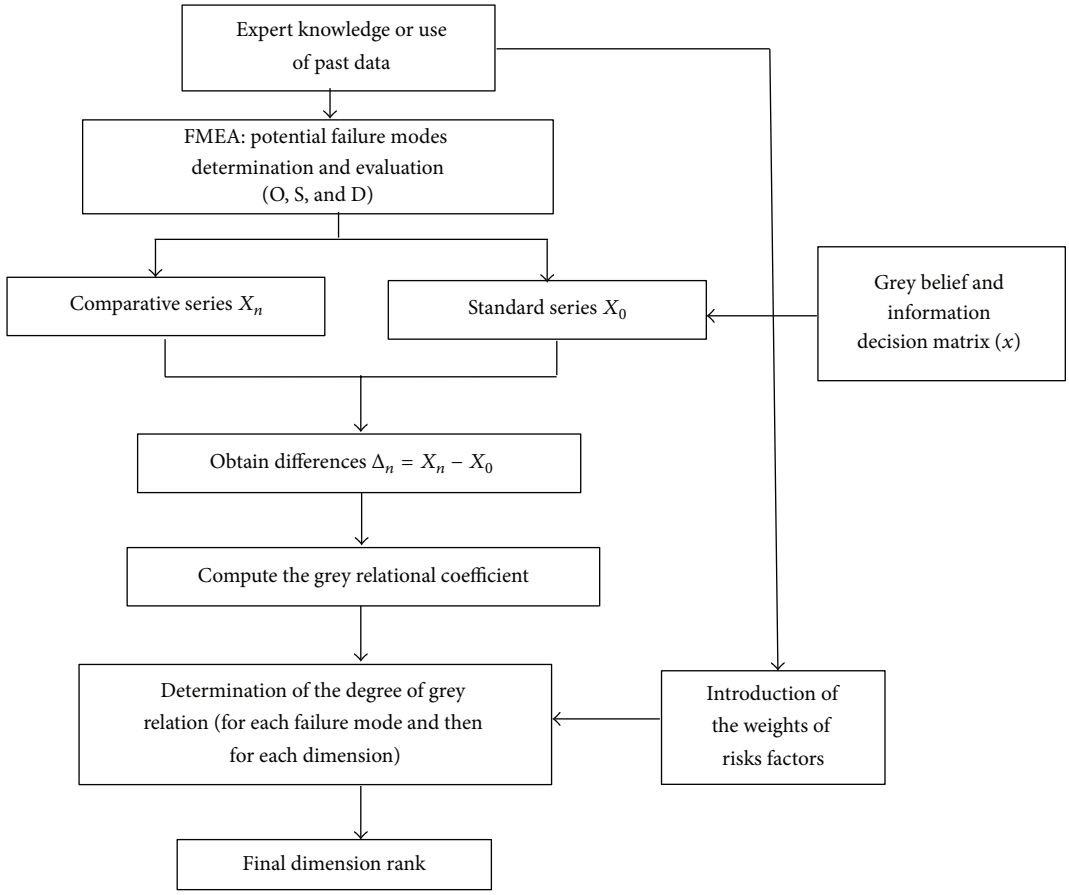

Figure 1: Flowchart of the proposed FMEA and Grey Theory based approach.

hypotheses, and must work with data that float into the organization in real time and that require real-time analysis and response. Therefore, in this paper, we analyzed the processing characteristics of the IBM Big Data Platform for illustrative purposes, but it is important to note that all big data platforms are vulnerable to both external and internal threats. Therefore, since our analysis model based on the probability of the occurrence of failure covers a wide view of the architecture of big data, it is eligible for analyzing other platforms, such as cloud computing infrastructures [66] and platforms from business scenarios [67]. Finally, our model considers the possible occurrence of failures in the distributed data and then we consider its implementation in a distributed way.

3.1. Expert Knowledge or Past Data regarding Previous Failures. The first step in the approach consists of expert identification or use of past data. The expert is the person who knows the enterprise systems and their vulnerability and is thus able to assess the information security risk of the organization in terms of the four dimensions [65]. One may also identify a group of experts in this step, and the analysis is accomplished by considering a composition of their judgments or the use of a dataset of past failures. The inclusion of an expert system in the model is also encouraged.

According to [68], an expert is someone with multiple skills who understands the working environment and has substantial training in and knowledge of the system being evaluated. Risk management models have widely used expert knowledge to provide value judgments that represent the expert's perceptions and/or preferences. For instance, [69] provides evidence obtained from two unbiased and independent experts regarding the risk of release of a highly flammable gas near a processing facility. References [70, 71] explore a risk measure of underground vaults that considers the consequences of arc faults using a single expert's a priori knowledge. Reference [19] proposes information security risk management using FMEA, Fuzzy Theory, and expert knowledge. Reference [72] analyzes the risk probability of an underwater tunnel excavation using the knowledge of four experts.

3.2. Determination and Evaluation of Potential Failure Modes (FMEA). In a general way, this step concerns the determination of the failure modes associated with the big data dimensions (Figure 2) in terms of their vulnerabilities. Each dimension is described in Table 5.

Furthermore, these dimensions can be damaged by various associated activities. Table 6 presents failure modes relating to the vulnerability of big data for each dimension.

Table 5: Description of dimensions.

Dimension	Description
Identification and access management	Given the opportunity to increase knowledge by accessing big data, it is necessary that only authorized persons can access it; thus, big data requires confidentiality and authenticity; to address this problem, [58] mentioned that sometimes both are needed simultaneously; this source recommended and proposed three different schemes: an encryption scheme, a signature scheme, and a sign-encryption scheme
Device and application registration	Data provenance refers to information about the history of a creation process; in other words, it refers to a mechanism that can be used to validate whether input data is coming from an authenticated source to guarantee a degree of information integrity [59]; then, provenance-related security and trustworthiness issues also arise in the system [60]; they include the registration of devices in machine-to-machine (M2M) and Internet-of-Things (IoT) networks, which can be considered one of the major issues in the area of security [61]
Infrastructure management	As big data physical infrastructures increase, difficulties associated with designing effective physical security also arise; thus, we use the term "system health" to describe the intersection of the information worker and the nominal conditions for infrastructure management monitoring of big data for security purposes, which include technical issues regarding the interoperability of services [62]
Data governance	Data governance can ensure appropriate controls without inhibiting the speed and flexibility of innovative big data approaches and technologies, which need to be established for different management levels with a clear security strategy

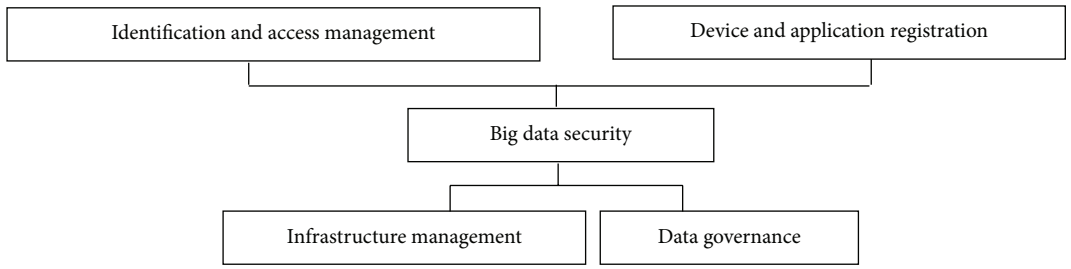

Figure 2: Big data dimensions.

In fact, the determination of the failure modes is achieved using the FMEA methodology and evaluated regarding its occurrence (O), severity (S), and detection (D).

3.3. Establish Comparative Series. An information series with n decision factors, such as chance of occurrence, severity of failure, or chance of lack of detection, can be expressed as follows:

$$X_i = \left(X_i(1), X_i(2), \ldots, X_i(k)\right). \tag{4}$$

These comparative series can be provided by an expert or any dataset of previous failures, based on the scales described in Tables 2–4.

3.4. Establish the Standard Series. According to [41], the degree of relation can describe the relationship of two series; thus, an objective series called the standard series is established and expressed as $X_0 = (X_0(1), X_0(2), \ldots, X_0(k))$, where k is the number of risk factors (for this work, $k = 3$, i.e., occurrence, severity, and detection). According to FMEA, as the score becomes smaller, the standard series can be denoted as $X_0 = (X_0(1), X_0(2), \ldots, X_0(k)) = (1, 1, \ldots, 1)$.

3.5. Obtain the Difference between the Comparative Series and the Standard Series. To discover the degree of the grey relationship, the difference between the score of the decision factors and the norm of the standard series must be determined and expressed by a matrix calculated by

$$\Delta_{0j}(k) = \left\| X_0(k) - X_j(k) \right\|, \tag{5}$$

where j is the number of failure modes in the analysis [31].

3.6. Compute the Grey Relational Coefficient. The grey relational coefficient is calculated by

$$\gamma\left(X_0(k), X_j(k)\right) = \frac{\Delta_{\min} - \zeta\Delta_{\max}}{\Delta_{0j}(k) - \zeta\Delta_{\max}}, \tag{6}$$

where ζ is an identifier, normally set to 0.5 [31]. It only affects the relative value of risk, not the priority.

Table 6: Failure modes associated with each dimension of big data.

Dimensions	Associated activities
A1: Identification and access management	A1.1: Loss of secret keys
	A1.2: Cryptanalysis of a ciphered signal
	A1.3: Secret password divulged to any other user
	A1.4: Intentional access to network services, for example, proxy servers
	A1.5: Spoofing: impersonation of a legitimate user
A2: Device and application registration	A2.1: Facility problems
	A2.2: Failure of encryption equipment
	A2.3: Unauthorized use of secure equipment
	A2.4: Ineffective infrastructure investment
	A2.5: Failure of application server
A3: Infrastructure management	A3.1: Cabling problems
	A3.2: Failure of radio platform transmission
	A3.3: Failure of cipher audio (telephone) and video
	A3.4: Failure of sensor networks
	A3.5: Failure of potential of energy
	A3.6: Unauthorized readout of data stored on a remote LAN
A4: Data governance	A4.1: Failure of interpretation and analysis of data
	A4.2: Failure of audit review of implemented policies and information security
	A4.3: Failure to maximize new business value
	A4.4: Failure of real-time demand forecasts

3.7. Determine the Degree of Relation. Before finding the degree of relation, the relative weight of the decision factors is first decided so that it can be used in the following formulation [31]. In a general way, it is calculated by

$$\Gamma\left(X_i, X_j\right) = \sum_{k=1}^{n} \beta_k \gamma\left(X_i(k), X_j(k)\right), \tag{7}$$

where β_k is the risk factors' weighting and, as a result, $\sum_{k=1}^{n} \beta_k = 1$.

3.8. Rank the Priority of Risk. This step consists of dimension ordering. Based on the degree of relation between the comparative series and the standard series, a relational series can be constructed. The greater the degree of relation, the smaller the effect of the cause [31].

4. An Illustrative Example

To demonstrate the applicability of our proposition based on FMEA and Grey Theory, an example based on a real context is presented in this section. The steps performed are the same as shown in Figure 1, explained in Section 3. Following these steps, the expert selected for this study is a senior academic with more than 20 years' experience. She holds a Ph.D. degree in information systems (IS), has published 12 papers in this field, and also has experience as a consultant in IS to companies in the private sector.

In the following step of the proposed model, the four dimensions associated with the potential failures of big data are represented according to Figure 2 and described in Table 5. Furthermore, Table 6 presents the failure modes relating to the vulnerability of big data for each dimension. Based on these potential failures, Tables 7 and 8 show the establishment of comparative and standard series for occurrence, severity, and detection, respectively.

To proceed to a grey relational analysis of potential accidents, it is necessary to obtain the difference between comparative series and standard series, according to (4). Table 9 shows the result of this difference.

In order to rank the priority of risk, it is necessary to compute both the grey relational coefficient (Table 10) and the degree of relation (Table 11) using (5), (6), and (7). Therefore, the greater the degree of relation, the smaller the effect of the cause. Assuming equal weights for risk factors, Table 11 also presents the degree of grey relation for each failure mode and dimension and final ranking.

From the analysis of failures using the proposed approach, we have shown that big data is mainly in need of structured policies for data governance. This result was expected because the veracity and provenance of data are fundamental to information security; otherwise, the vulnerabilities may be catastrophic or big data may have little value for the acquisition of knowledge. Data governance is also an aspect that requires more awareness because it deals with large amounts of data and directly influences operational costs.

Since the model works with a recommendation rather than a solution and compatible recommendations depend on expert knowledge, it is important to test the robustness of

Table 7: Comparative series.

Dimensions	Associated activities	O	S	D
A1: Identification and access management	A1.1: Loss of secret keys	5	7	4
	A1.2: Cryptanalysis of a ciphered signal	5	5	4
	A1.3: Secret password divulged to any other user	2	7	5
	A1.4: Intentional access to network services, for example, proxy servers	6	5	7
	A1.5: Spoofing: impersonation of a legitimate user	6	5	7
A2: Device and application registration	A2.1: Facility problems	8	7	5
	A2.2: Failure of encryption equipment	6	9	5
	A2.3: Unauthorized use of secure equipment	6	5	4
	A2.4: Ineffective infrastructure investment	8	5	4
	A2.5: Failure of application server	5	4	5
A3: Infrastructure management	A3.1: Cabling problems	6	5	4
	A3.2: Failure of radio platform transmission	2	9	4
	A3.3: Failure of cipher audio (telephone) and video	2	7	4
	A3.4: Failure of sensor networks	5	7	2
	A3.5: Failure of potential of energy	2	7	2
	A3.6: Unauthorized readout of data stored on a remote LAN	5	5	4
A4: Data governance	A4.1: Failure of interpretation and analysis of data	8	9	5
	A4.2: Failure of audit review of implemented policies and information security	8	9	4
	A4.3: Failure to maximize new business value	8	7	5
	A4.4: Failure of real-time demand forecasts	8	7	7

Table 8: Standard series.

Dimensions	Associated activities	O	S	D
A1: Identification and access management	A1.1: Loss of secret keys	1	1	1
	A1.2: Cryptanalysis of a ciphered signal	1	1	1
	A1.3: Secret password divulged to any other user	1	1	1
	A1.4: Intentional access to network services, for example, proxy servers	1	1	1
	A1.5: Spoofing: impersonation of a legitimate user	1	1	1
A2: Device and application registration	A2.1: Facility problems	1	1	1
	A2.2: Failure of encryption equipment	1	1	1
	A2.3: Unauthorized use of secure equipment	1	1	1
	A2.4: Ineffective infrastructure investment	1	1	1
	A2.5: Failure of application server	1	1	1
A3: Infrastructure management	A3.1: Cabling problems	1	1	1
	A3.2: Failure of radio platform transmission	1	1	1
	A3.3: Failure of cipher audio (telephone) and video	1	1	1
	A3.4: Failure of sensor networks	1	1	1
	A3.5: Failure of potential of energy	1	1	1
	A3.6: Unauthorized readout of data stored on a remote LAN	1	1	1
A4: Data governance	A4.1: Failure of interpretation and analysis of data	1	1	1
	A4.2: Failure of audit review of implemented policies and information security	1	1	1
	A4.3: Failure to maximize new business value	1	1	1
	A4.4: Failure of real-time demand forecasts	1	1	1

this information and therefore to conduct sensitivity analysis. Thus, different weightings, based on the context, may also be used for different risk factors, as suggested by [33]. Table 12 presents a sensitivity analysis conducted in order to evaluate the performance and validity of the results of the model. As can be seen, the final ranking of risk is the same for all the different weightings tested (±10%).

5. Discussion and Conclusions

The main difficulties in big data security risk analysis involve the volume of data and the variety of data connected to different databases. From the perspective of security and privacy, traditional databases have governance controls and a consolidated auditing process, while big data is at an early

Table 9: Difference between comparative series and standard series.

Dimensions	Associated activities	O	S	D
A1: Identification and access management	A1.1: Loss of secret keys	4	6	3
	A1.2: Cryptanalysis of a ciphered signal	4	4	3
	A1.3: Secret password divulged to any other user	1	6	4
	A1.4: Intentional access to network services, for example, proxy servers	5	4	6
	A1.5: Spoofing: impersonation of a legitimate user	5	4	6
A2: Device and application registration	A2.1: Facility problems	7	6	4
	A2.2: Failure of encryption equipment	5	3	4
	A2.3: Unauthorized use of secure equipment	5	4	3
	A2.4: Ineffective infrastructure investment	7	4	3
	A2.5: Failure of application server	4	3	4
A3: Infrastructure management	A3.1: Cabling problems	5	4	3
	A3.2: Failure of radio platform transmission	1	8	3
	A3.3: Failure of cipher audio (telephone) and video	1	6	3
	A3.4: Failure of sensor networks	4	6	1
	A3.5: Failure of potential of energy	1	6	1
	A3.6: Unauthorized readout of data stored on a remote LAN	4	4	3
A4: Data governance	A4.1: Failure of interpretation and analysis of data	7	8	4
	A4.2: Failure of audit review of implemented policies and information security	7	8	3
	A4.3: Failure to maximize new business value	7	6	4
	A4.4: Failure of real-time demand forecasts	7	6	6

stage of development and hence continues to require structured analysis to address threats and vulnerabilities. Moreover, there is not yet enough research into risk analysis in the context of big data.

Thus, security is one of the most important issues for the stability and development of big data. Aiming to identify the risk factors and the uncertainty associated with the propagation of vulnerabilities, this paper proposed a systematic framework based on FMEA and Grey Theory, more precisely GRA. This systematic framework allows for an evaluation of risk factors and their relative weightings in a linguistic, as opposed to a precise, manner for evaluation of big data failure modes. This is in line with the uncertain nature of the context. In fact, according to [40], the traditional FMEA method cannot assign different weightings to the risk factors of O, S, and D and therefore may not be suitable for real-world situations. These authors pointed out that introducing Grey Theory into the traditional FMEA method enables engineers to allocate relative importance to the O, S, and D risk factors based on research and their own experience. In a general way, another advantage of this proposal is that it requires less effort on the part of experts using linguistic terms. Consequently, these experts can make accurate judgments using linguistic terms based on their experience or on datasets relating to previous failures.

Based on the above information, the use of our proposal is justified to identify and assess big data risk in a quantitative manner. Moreover, this study comprises various security characteristics of big data using FMEA: it analyzes four dimensions, identification and access management, device and application registration, infrastructure management, and data governance, as well as 20 subdimensions that represent failure modes. Therefore, this work can be expected to serve as a guideline for managing big data failures in practice.

It is worth stating that the results presented greater awareness of data governance for ensuring appropriate controls. In this context, a challenge to the process of governing big data is to categorize, model, and map data as it is captured and stored, mainly because of the unstructured nature of the volume of information. Then, one role of data governance in the information security context is to allow for the information that contributes to reporting to be defined consistently across the organization in order to guide and structure the most important activities and to help clarify decisions. Briefly, analyzing data from the distant past to decide on a current situation does not mean that the data has higher value. From another perspective, increasing volume does not guarantee confidence in decisions, and one may use tools such as data mining and knowledge discovery, proposed in [73], to improve the decision process.

Indeed, the concept of storage management is a critical point, especially when volumes of data that exceed the storage capacity are considered [11]. In fact, the emphasis of big data analytics is on how data is stored in a distributed fashion, for example, in traditional databases or in a cloud [66]. When a cloud is used, data can be processed in parallel on many computing nodes, in distributed environments across clusters of machines [3]. In conclusion, big data security must be seen as an important and challenging feature, capable of generating significant limitations. For instance, several electronic devices that enable communication via networks, especially via the Internet, and which place great emphasis on mobile trends allow for an increase in volume, variety, and even speed of data, which can thereby be defined as big

Table 10: Grey relational coefficient.

Dimensions	Associated activities	O	S	D
A1: Identification and access management	A1.1: Loss of secret keys	0.625	0.5	0.714286
	A1.2: Cryptanalysis of a ciphered signal	0.625	0.625	0.714286
	A1.3: Secret password divulged to any other user	1	0.5	0.625
	A1.4: Intentional access to network services, for example, proxy servers	0.555556	0.625	0.5
	A1.5: Spoofing: impersonation of a legitimate user	0.555556	0.625	0.5
A2: Device and application registration	A2.1: Facility problems	0.454545	0.5	0.625
	A2.2: Failure of encryption equipment	0.555556	0.416667	0.625
	A2.3: Unauthorized use of secure equipment	0.555556	0.625	0.714286
	A2.4: Ineffective infrastructure investment	0.454545	0.625	0.714286
	A2.5: Failure of application server	0.625	0.714286	0.625
A3: Infrastructure management	A3.1: Cabling problems	0.555556	0.625	0.714286
	A3.2: Failure of radio platform transmission	1	0.416667	0.714286
	A3.3: Failure of cipher audio (telephone) and video	1	0.5	0.714286
	A3.4: Failure of sensor networks	0.625	0.5	1
	A3.5: Failure of potential of energy	1	0.5	1
	A3.6: Unauthorized readout of data stored on a remote LAN	0.625	0.625	0.714286
A4: Data governance	A4.1: Failure of interpretation and analysis of data	0.454545	0.416667	0.625
	A4.2: Failure of audit review of implemented policies and information security	0.454545	0.416667	0.714286
	A4.3: Failure to maximize new business value	0.454545	0.5	0.625
	A4.4: Failure of real-time demand forecasts	0.454545	0.5	0.5

Table 11: The degree of grey relation for each failure mode and each dimension and the final rank.

Dimensions	Associated activities	Degree of grey relation	Degree of grey relation (dimension)	Risk ranking
A1: Identification and access management	A1.1: Loss of secret keys	0.613095	0.619312	3
	A1.2: Cryptanalysis of a ciphered signal	0.654762		
	A1.3: Secret password divulged to any other user	0.708333		
	A1.4: Intentional access to network services, for example, proxy servers	0.560185		
	A1.5: Spoofing: impersonation of a legitimate user	0.560185		
A2: Device and application registration	A2.1: Facility problems	0.526515	0.588648	2
	A2.2: Failure of encryption equipment	0.532407		
	A2.3: Unauthorized use of secure equipment	0.631614		
	A2.4: Ineffective infrastructure investment	0.597944		
	A2.5: Failure of application server	0.654762		
A3: Infrastructure management	A3.1: Cabling problems	0.631614	0.712743	4
	A3.2: Failure of radio platform transmission	0.710317		
	A3.3: Failure of cipher audio (telephone) and video	0.738095		
	A3.4: Failure of sensor networks	0.708333		
	A3.5: Failure of potential of energy	0.833333		
	A3.6: Unauthorized readout of data stored on a remote LAN	0.654762		
A4: Data governance	A4.1: Failure of interpretation and analysis of data	0.498737	0.50965	1
	A4.2: Failure of audit review of implemented policies and information security	0.528499		
	A4.3: Failure to maximize new business value	0.526515		
	A4.4: Failure of real-time demand forecasts	0.484848		

Table 12: Sensitivity analysis.

Weights of risk factors	Degree of grey relation (dimension) and risk ranking
Occurrence: 0.30 Severity: 0.35 Detection: 0.35	D1: 0.616667 (3) D2: 0.591629 (2) D3: 0.645833 (4) D4: 0.512405 (1)
Occurrence: 0.36 Severity: 0.32 Detection: 0.32	D1: 0.621429 (3) D2: 0.586264 (2) D3: 0.641071 (4) D4: 0.507446 (1)
Occurrence: 0.35 Severity: 0.30 Detection: 0.35	D1: 0.621528 (3) D2: 0.589271 (2) D3: 0.644097 (4) D4: 0.512216 (1)
Occurrence: 0.32 Severity: 0.36 Detection: 0.32	D1: 0.61754 (3) D2: 0.58815 (2) D3: 0.64246 (4) D4: 0.507597 (1)
Occurrence: 0.35 Severity: 0.35 Detection: 0.30	D1: 0.619742 (3) D2: 0.585045 (2) D3: 0.639633 (4) D4: 0.504329 (1)
Occurrence: 0.35 Severity: 0.35 Detection: 0.30	D1: 0.618968 (3) D2: 0.591531 (2) D3: 0.646032 (4) D4: 0.513907 (1)

data content. This fact adds more value to large volumes of data and allows for the support of organizational activities, bequeathing even more importance to the area of data processing, which now tends to work in a connected way that goes beyond the boundaries of companies.

This research contributes as a guide for researchers in the analysis of suitable big data risk techniques and in the development of response to the insufficiency of existing solutions. This risk model can ensure the identification of failure and attacks and help the victim decide how to react when this type of attack occurs. However, this study has limitations. For instance, it does not measure the consequences of a disaster occurring in the field of big data. This measurement could be carried out based on [74]. Future work should focus on developing a model to ensure the working of data governance and should recommend specific actions to ensure the safety of big data and to help managers choose the best safeguards to reduce risks. Further studies may also consider security-related issues in the fields of enterprise architecture, information infrastructure, and cloud-based computing.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

This research was partially supported by Universidade Federal de Pernambuco and GPSID, Decision and Information Systems Research Group.

References

[1] R. Tinati, S. Halford, L. Carr, and C. Pope, "Big data: methodological challenges and approaches for sociological analysis," *Sociology*, vol. 48, no. 4, pp. 663–681, 2014.

[2] M. Chen, S. Mao, and Y. Liu, "Big data: a survey," *Mobile Networks and Applications*, vol. 19, no. 2, pp. 171–209, 2014.

[3] H. Hu, Y. Wen, T.-S. Chua, and X. Li, "Toward scalable systems for big data analytics: a technology tutorial," *IEEE Access*, vol. 2, pp. 652–687, 2014.

[4] S. Erevelles, N. Fukawa, and L. Swayne, "Big Data consumer analytics and the transformation of marketing," *Journal of Business Research*, vol. 69, no. 2, pp. 897–904, 2016.

[5] N. Kshetri, "Big data's role in expanding access to financial services in China," *International Journal of Information Management*, vol. 36, no. 3, pp. 297–308, 2016.

[6] T. Poleto, V. D. H. de Carvalho, and A. P. C. S. Costa, "The roles of big data in the decision-support process: an empirical investigation," in *Decision Support Systems V—Big Data Analytics for Decision Making: First International Conference, ICDSST 2015, Belgrade, Serbia, May 27–29, 2015, Proceedings*, vol. 216 of *Lecture Notes in Business Information Processing*, pp. 10–21, Springer, Berlin, Germany, 2015.

[7] E. G. Horta, C. L. de Castro, and A. P. Braga, "Stream-based extreme learning machine approach for big data problems," *Mathematical Problems in Engineering*, vol. 2015, Article ID 126452, 17 pages, 2015.

[8] D. Peralta, S. del Río, S. Ramírez-Gallego, I. Triguero, J. M. Benitez, and F. Herrera, "Evolutionary feature selection for big data classification: a MapReduce approach," *Mathematical Problems in Engineering*, vol. 2015, Article ID 246139, 11 pages, 2015.

[9] X. Song, Y. Wu, Y. Ma, Y. Cui, and G. Gong, "Military simulation big data: background, state of the art, and challenges," *Mathematical Problems in Engineering*, vol. 2015, Article ID 298356, 20 pages, 2015.

[10] C. L. Philip Chen and C.-Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: a survey on Big Data," *Information Sciences*, vol. 275, pp. 314–347, 2014.

[11] A. Siddiqa, I. A. T. Hashem, I. Yaqoob et al., "A survey of big data management: taxonomy and state-of-the-art," *Journal of Network and Computer Applications*, vol. 71, pp. 151–166, 2016.

[12] A. P. H. De Gusmão, L. C. E Silva, M. M. Silva, T. Poleto, and A. P. C. S. Costa, "Information security risk analysis model using fuzzy decision theory," *International Journal of Information Management*, vol. 36, no. 1, pp. 25–34, 2016.

[13] W. T. Yue, M. Çakanyildirim, Y. U. Ryu, and D. Liu, "Network externalities, layered protection and IT security risk management," *Decision Support Systems*, vol. 44, no. 1, pp. 1–16, 2007.

[14] K. Singh, S. C. Guntuku, A. Thakur, and C. Hota, "Big Data Analytics framework for Peer-to-Peer Botnet detection using Random Forests," *Information Sciences*, vol. 278, pp. 488–497, 2014.

[15] S. Hou, X. Huang, J. K. Liu, J. Li, and L. Xu, "Universal designated verifier transitive signatures for graph-based big data," *Information Sciences*, vol. 318, pp. 144–156, 2015.

[16] J. Zhang and Q. Dong, "Efficient ID-based public auditing for the outsourced data in cloud storage," *Information Sciences*, vol. 343-344, pp. 1–14, 2016.

[17] M. Sookhak, A. Gani, M. K. Khan, and R. Buyya, "Dynamic remote data auditing for securing big data storage in cloud computing," *Information Sciences*, 2015.

[18] N. Baracaldo and J. Joshi, "An adaptive risk management and access control framework to mitigate insider threats," *Computers and Security*, vol. 39, pp. 237–254, 2013.

[19] M. M. Silva, A. P. H. de Gusmão, T. Poleto, L. C. E. Silva, and A. P. C. S. Costa, "A multidimensional approach to information security risk management using FMEA and fuzzy theory," *International Journal of Information Management*, vol. 34, no. 6, pp. 733–740, 2014.

[20] N. Feng, H. J. Wang, and M. Li, "A security risk analysis model for information systems: causal relationships of risk factors and vulnerability propagation analysis," *Information Sciences*, vol. 256, no. 20, pp. 57–73, 2014.

[21] B. Karabacak and I. Sogukpinar, "ISRAM: information security risk analysis method," *Computers and Security*, vol. 24, no. 2, pp. 147–159, 2005.

[22] R. Farley and X. Wang, "Exploiting VoIP softphone vulnerabilities to disable host computers: attacks and mitigation," *International Journal of Critical Infrastructure Protection*, vol. 7, no. 3, pp. 141–154, 2014.

[23] V. K. Verma, S. Singh, and N. P. Pathak, "Impact of malicious servers over trust and reputation models in wireless sensor networks," *International Journal of Electronics*, vol. 103, no. 3, pp. 530–540, 2016.

[24] V. Varadharajan and U. Tupakula, "Counteracting security attacks in virtual machines in the cloud using property based attestation," *Journal of Network and Computer Applications*, vol. 40, no. 1, pp. 31–45, 2014.

[25] H. Takabi, J. B. D. Joshi, and G.-J. Ahn, "Security and privacy challenges in cloud computing environments," *IEEE Security and Privacy*, vol. 8, no. 6, pp. 24–31, 2010.

[26] SANS, "A Qualitative Risk Analysis and Management Tool-CRAMM," 2002.

[27] M. P. Kailay and P. Jarratt, "RAMeX: a prototype expert system for computer security risk analysis and management," *Computers & Security*, vol. 14, no. 5, pp. 449–463, 1995.

[28] T. R. Peltier, *Facilitated Risk Analysis Process (FRAP)*, Auerbach Publications, 2000.

[29] J. Creasey, "A complete information risk management solution For ISF members using IRAM and STREAM," in *Managing Information Risk*, pp. 1–7, 2013.

[30] C. Alberts and A. Dorofee, *Managing Information Security Risks: The OCTAVE Approach*, Addison-Wesley, 2002.

[31] R. J. Mikulak, R. McDermott, and M. Beauregard, *The Basics of FMEA*, vol. 2, CRC Press, Boca Raton, Fla, USA, 2009.

[32] A. Pillay and J. Wang, "Modified failure mode and effects analysis using approximate reasoning," *Reliability Engineering and System Safety*, vol. 79, no. 1, pp. 69–85, 2003.

[33] M. Ben Daya and Abdul Raouf, "A revised failure mode and effects analysis model," *International Journal of Quality & Reliability Management*, vol. 13, no. 1, pp. 43–47, 1996.

[34] J. B. Bowles and C. E. Peláez, "Fuzzy logic prioritization of failures in a system failure mode, effects and criticality analysis," *Reliability Engineering and System Safety*, vol. 50, no. 2, pp. 203–213, 1995.

[35] M. Abdelgawad and A. R. Fayek, "Risk management in the construction industry using combined fuzzy FMEA and fuzzy AHP," *Journal of Construction Engineering and Management*, vol. 136, no. 9, pp. 1028–1036, 2010.

[36] A. Mariajayaprakash and T. Senthilvelan, "Failure detection and optimization of sugar mill boiler using FMEA and Taguchi method," *Engineering Failure Analysis*, vol. 30, pp. 17–26, 2013.

[37] O. Kaljević, J. Djuriš, Z. Djurić, and S. Ibrić, "Application of failure mode and effects analysis in quality by design approach for formulation of carvedilol compression coated tablets," *Journal of Drug Delivery Science and Technology*, vol. 32, pp. 56–63, 2016.

[38] A. Colli, "Failure mode and effect analysis for photovoltaic systems," *Renewable and Sustainable Energy Reviews*, vol. 50, pp. 804–809, 2015.

[39] C. Kahraman, I. Kaya, and Ö. Şenvar, "Healthcare failure mode and effects analysis under fuzziness," *Human and Ecological Risk Assessment*, vol. 19, no. 2, pp. 538–552, 2013.

[40] J. Wei, L. Zhou, F. Wang, and D. Wu, "Work safety evaluation in Mainland China using grey theory," *Applied Mathematical Modelling*, vol. 39, no. 2, pp. 924–933, 2015.

[41] C.-L. Chang, P.-H. Liu, and C.-C. Wei, "Failure mode and effects analysis using grey theory," *Integrated Manufacturing Systems*, vol. 12, no. 3, pp. 211–216, 2001.

[42] Q. Zhou and V. V. Thai, "Fuzzy and grey theories in failure mode and effect analysis for tanker equipment failure prediction," *Safety Science*, vol. 83, pp. 74–79, 2016.

[43] Y. Geum, Y. Cho, and Y. Park, "A systematic approach for diagnosing service failure: service-specific FMEA and grey relational analysis approach," *Mathematical and Computer Modelling*, vol. 54, no. 11-12, pp. 3126–3142, 2011.

[44] J.-L. Deng, "Control problems of grey systems," *Systems & Control Letters*, vol. 1, no. 5, pp. 288–294, 1982.

[45] J. L. Deng, "Introduction to grey system theory," *The Journal of Grey System*, vol. 1, no. 1, pp. 1–24, 1989.

[46] H. Kuang, M. A. Bashar, K. W. Hipel, and D. M. Kilgour, "Grey-based preference in a graph model for conflict resolution with multiple decision makers," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 9, pp. 1254–1267, 2015.

[47] H. Kuang, D. M. Kilgour, and K. W. Hipel, "Grey-based PROMETHEE II with application to evaluation of source water protection strategies," *Information Sciences*, vol. 294, pp. 376–389, 2015.

[48] M. S. Memon, Y. H. Lee, and S. I. Mari, "Group multi-criteria supplier selection using combined grey systems theory and uncertainty theory," *Expert Systems with Applications*, vol. 42, no. 21, pp. 7951–7959, 2015.

[49] D. Golmohammadi and M. Mellat-Parast, "Developing a grey-based decision-making model for supplier selection," *International Journal of Production Economics*, vol. 137, no. 2, pp. 191–200, 2012.

[50] Z. Li, G. Wen, and N. Xie, "An approach to fuzzy soft sets in decision making based on grey relational analysis and Dempster-Shafer theory of evidence: an application in medical diagnosis," *Artificial Intelligence in Medicine*, vol. 64, no. 3, pp. 161–171, 2015.

[51] R. Bhattacharyya, "A grey theory based multiple attribute approach for R&D project portfolio selection," *Fuzzy Information and Engineering*, vol. 7, no. 2, pp. 211–225, 2015.

[52] G. Kou, Y. Lu, Y. Peng, and Y. Shi, "Evaluation of classification algorithms using MCDM and rank correlation," *International Journal of Information Technology and Decision Making*, vol. 11, no. 1, pp. 197–225, 2012.

[53] G.-D. Li, D. Yamaguchi, and M. Nagai, "A grey-based decision-making approach to the supplier selection problem," *Mathematical and Computer Modelling*, vol. 46, no. 3-4, pp. 573–581, 2007.

[54] H.-H. Wu, "A comparative study of using grey relational analysis in multiple attribute decision making problems," *Quality Engineering*, vol. 15, no. 2, pp. 209–217, 2002.

[55] Y. Kuo, T. Yang, and G.-W. Huang, "The use of grey relational analysis in solving multiple attribute decision-making problems," *Computers and Industrial Engineering*, vol. 55, no. 1, pp. 80–93, 2008.

[56] W.-S. Lee and Y.-C. Lin, "Evaluating and ranking energy performance of office buildings using Grey relational analysis," *Energy*, vol. 36, no. 5, pp. 2551–2556, 2011.

[57] C.-L. Chang, C.-C. Wei, and Y.-H. Lee, "Failure mode and effects analysis using fuzzy method and grey theory," *Kybernetes*, vol. 28, no. 8-9, pp. 1072–1080, 1999.

[58] G. Wei, J. Shao, Y. Xiang, P. Zhu, and R. Lu, "Obtain confidentiality or/and authenticity in big data by ID-based generalized signcryption," *Information Sciences*, vol. 318, pp. 111–122, 2015.

[59] B. Glavic, "Big data provenance: challenges and implications for benchmarking," in *Specifying Big Data Benchmarks*, pp. 72–80, 2014.

[60] J. Park, D. Nguyen, and R. Sandhu, "A provenance-based access control model," in *Proceedings of the 10th Annual International Conference on Privacy, Security and Trust (PST '12)*, pp. 137–144, Paris, France, July 2012.

[61] H.-C. Chen, I. You, C.-E. Weng, C.-H. Cheng, and Y.-F. Huang, "A security gateway application for End-to-End M2M communications," *Computer Standards and Interfaces*, vol. 44, pp. 85–93, 2016.

[62] R. A. Oliveira, N. Laranjeiro, and M. Vieira, "Assessing the security of web service frameworks against Denial of Service attacks," *Journal of Systems and Software*, vol. 109, pp. 18–31, 2015.

[63] K. Kambatla, G. Kollias, V. Kumar, and A. Grama, "Trends in big data analytics," *Journal of Parallel and Distributed Computing*, vol. 74, no. 7, pp. 2561–2573, 2014.

[64] G. Lafuente, "The big data security challenge," *Network Security*, vol. 2015, no. 1, pp. 12–14, 2015.

[65] National Institute of Standards and Technology—NIST, *Big Data Interoperability Framework: Security and Privacy*, vol. 4, NIST, Gaithersburg, Md, USA, 2015.

[66] R. Iqbal, F. Doctor, B. More, S. Mahmud, and U. Yousuf, "Big data analytics: computational intelligence techniques and application areas," *International Journal of Information Management*, 2016.

[67] J. Chen, Y. Tao, H. Wang, and T. Chen, "Big data based fraud risk management at Alibaba," *The Journal of Finance and Data Science*, vol. 1, no. 1, pp. 1–10, 2015.

[68] J. H. Purba, "A fuzzy-based reliability approach to evaluate basic events of fault tree analysis for nuclear power plant probabilistic safety assessment," *Annals of Nuclear Energy*, vol. 70, pp. 21–29, 2014.

[69] R. Ferdous, F. Khan, R. Sadiq, P. Amyotte, and B. Veitch, "Handling data uncertainties in event tree analysis," *Process Safety and Environmental Protection*, vol. 87, no. 5, pp. 283–292, 2009.

[70] T. V. Garcez and A. T. De Almeida, "Multidimensional risk assessment of manhole events as a decision tool for ranking the vaults of an underground electricity distribution system," *IEEE Transactions on Power Delivery*, vol. 29, no. 2, pp. 624–632, 2014.

[71] T. V. Garcez and A. T. De Almeida, "A risk measurement tool for an underground electricity distribution system considering the consequences and uncertainties of manhole events," *Reliability Engineering and System Safety*, vol. 124, pp. 68–80, 2014.

[72] E.-S. Hong, I.-M. Lee, H.-S. Shin, S.-W. Nam, and J.-S. Kong, "Quantitative risk evaluation based on event tree analysis technique: application to the design of shield TBM," *Tunnelling and Underground Space Technology*, vol. 24, no. 3, pp. 269–277, 2009.

[73] Y. Peng, G. Kou, Y. Shi, and Z. Chen, "A descriptive framework for the field of data mining and knowledge discovery," *International Journal of Information Technology and Decision Making*, vol. 7, no. 4, pp. 639–682, 2008.

[74] D. Feledi and S. Fenz, "Challenges of web-based information security knowledge sharing," in *Proceedings of the 7th International Conference on Availability, Reliability and Security (ARES '12)*, pp. 514–521, Prague, Czech Republic, August 2012.