Computer Systems Science & Engineering
DOI: 10.32604/csse.2023.025796
Article

Tech Science Press

Evaluating Security of Big Data Through Fuzzy Based Decision-Making Technique

Fawaz Alassery[1], Ahmed Alzahrani[2], Asif Irshad Khan[2], Kanika Sharma[3], Masood Ahmad[4] and Raees Ahmad Khan[4,*]

[1]Department of Computer Science, College of Computers and Information Technology, Taif University, P. O. Box 11099, Taif 21944, Saudi Arabia
[2]Department of Computer Science, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
[3]Department of Computer Applications, Mangalmay Institute of Engineering & Technology, Greater Noida, 201310, Uttar Pradesh, India
[4]Department of Information Technology, Babasaheb Bhimrao Ambedkar University, Lucknow, 226025, Uttar Pradesh, India
*Corresponding Author: Raees Ahmad Khan. Email: khanraees@yahoo.com
Received: 05 December 2021; Accepted: 17 January 2022

Abstract: In recent years, it has been observed that the disclosure of information increases the risk of terrorism. Without restricting the accessibility of information, providing security is difficult. So, there is a demand for time to fill the gap between security and accessibility of information. In fact, security tools should be usable for improving the security as well as the accessibility of information. Though security and accessibility are not directly influenced, some of their factors are indirectly influenced by each other. Attributes play an important role in bridging the gap between security and accessibility. In this paper, we identify the key attributes of accessibility and security that impact directly and indirectly on each other, such as confidentiality, integrity, availability, and severity. The significance of every attribute on the basis of obtained weight is important for its effect on security during the big data security life cycle process. To calculate the proposed work, researchers utilised the Fuzzy Analytic Hierarchy Process (Fuzzy AHP). The findings show that the Fuzzy AHP is a very accurate mechanism for determining the best security solution in a real-time healthcare context. The study also looks at the rapidly evolving security technologies in healthcare that could help improve healthcare services and the future prospects in this area.

Keywords: Information security; big data; big data security life cycle; fuzzy AHP

1 Introduction

Nowadays, information and data are generated and handled at high velocity, creating huge amounts of data. This huge amount of data is known as "Big Data". Information and data are tools that users use to transfer information and data with them from the moment they start to live together. Information technology's type and nature have been transformed radically in the previous few years. This is the technical era, where information and data are created, managed, and processed at high velocity, creating a huge amount of data and information. This information is generated from various sources such as social

media, hospital databases, bank transactions, etc. Thus, it's totally up to the users to decide whether the information is worthless or valuable [1,2].

The most common definition of information security is that information cannot be disclosed to anyone. It is a combination of three main components, i.e., security, privacy, and accessibility [3,4]. Sharing information nowadays can pose jeopardy to any organisation and any user's privacy. Big data is a collection of a large amount of personal information that is easily accessible to be collected and analyzed. Organizations are rich sources of users' sensitive and personal information and epitomise supplementary opportunities for financially motivated cyber attackers [5,6]. Hacking and IT incidents, unauthorised access, theft and improper disposal were the main causes of data breaches.

India and other countries are facing cyber-related issues. The biggest data breaches occurred in India in 2021. Dominos India's 18 crore user data was compromised, Mobikwik's 10 crore, Facebook's 60 lakh, Air India's 45 lakh, Upstox's 25 lakh, and Bigbasket's 20 million customer data was exposed on the dark web [7]. From time to time, hackers appear in the cyber world and demand a ransom for data leaks. Percentage growth in 2020 data breaches compared to 2019 in different sectors is discussed and depicted in Fig. 1. Data breaches in the food and beverage industry increased by 1300% in 2020 compared to 2019 [8]. This was the highest growth compared to other industries in one year. Data security from hacking and other unauthorised access is more important.

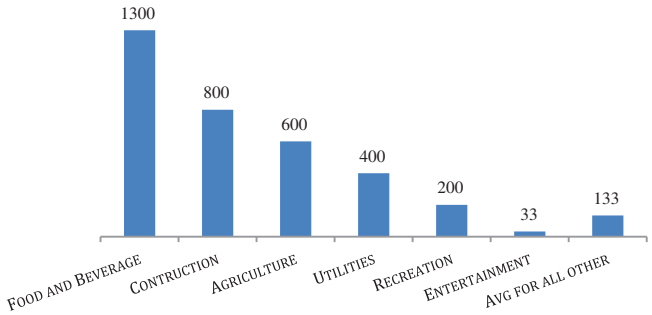

Figure 1: Data breaches growth in percentage across industries in 2020 compare to 2019

Data leaks enable data breaches, the cause of both events are intertwined hence the list below cloud pinpoint the sources of either data breaches or data leaks. When monitoring attack surfaces, information security programmes should choose a data leak detection perspective to increase the effectiveness of data breach prevention efforts. This will inevitably reveal and fix the security flaws that are at the root of both cyber threats. Managing vulnerabilities solely for the sake of preventing data breaches narrows the threat detection field, increasing the risk of crucial data leaks. In 2020, the following incidents will be some of the most common source of data leaks [9]. Misconfigured software; software settings that are incorrectly configured could disclose valuable client information. If the leaky software becomes widely used, millions of people could be vulnerable to cyberattacks. Social Engineering; Attackers seldom instigate data dumps, but when they do, it is usually as a results of social engineering tactics.

The use of psychological manipulation in order to get sensitive credentials from victims is known as social engineering [10,11]. Phishing is the most prevalent form of social engineering attack, which can be carried out verbally or online. Recycled passwords; Consumers tend to use the same password for all of their logins; a single leaked password often leads to the compromise of many digital solutions. Because stolen client data is commonly traded on dark web forums, this poor security approach results in a serious data loss. Physical theft; when company devices get into the wrong hands, sensitive information on them can be used to said security breaches or identity theft, resulting in data breaches. Software vulnerabilities;

Software flaws, such as zero-day exploits, provide easy access to sensitive information. This skips the first stage of cyberattack lifecycle, propelling attackers straight to the privilege escalation stage of the attack lifecycle, which is the only stage left before a data breach.

Use of default passwords; even attackers have access to many of the factory-standard login credentials that come with new gadgets. As a result, factory-standard credentials that haven't changed are considered data leaks [12,13]. Information security, privacy, and accessibility provide data security across the organisation and are crucial in the complete operation of an organization. Together, they involve the users and techniques needed to reduce unauthorised access [14,15]. Therefore, to accomplish this, various scientists are attempting to enhance the security by calculating it via various approaches. There is a lot of work available in this area, but in the literature, evaluating the attributes of security and accessibility characteristics by their applicability in actual situations didn't meet. Moreover, the user is the sole individual who is responsible for his or her information security and accessibility.

Hence, providing security with accessibility to information is a crucial task, but it's the demand for time in the big data environment that's the demand. These paper goals are to secure the information with accessibility through MCDM methods with the combination of fuzzy [1,6]. In general, the fuzzy approach enables users to analyse an IoT e-Health support system in a smart home and compute risk analysis to determine doubts in a real-world environment using soft computing. This study helps to develop and design a way that utilises fuzzy logic to analyse the hierarchy of security-accessibility attributes. For this evaluation, a hierarchy that describes the real attributes of security and accessibility was created.

Thus, in the next section, the authors addressed hierarchy attributes, and Fuzzy-AHP defined the demand for security in the big data world. The results obtained by this study may help users preserve their privacy during the big data security life cycle. The paper is organised as follows: The literature survey is part of the 2 section of this paper. In the next section, methodology will be discussed. Section 4 provides an overview of the obtained results and, finally, the conclusion in Section 5.

2 Literature Review

Almulihi et al., have discussed about big data breaches in healthcare perspective and evaluate the importance of various databases with the help of fuzzy logic [1]. Hence, the main aim of organisations should be to prevent unauthorised access and provide security with accessibility through proposed security-saving data mining. Data mining, security saving techniques, and cryptography approaches might be found in the blend study [16,17]. Multiple approaches, like randomization for grouping, utilising privacy-preservation techniques, life span of security have been studied [18,19]. Decision making-based methodologies offer greater security at the expense of the high computation and compatibility required in these cases. A detailed explanation of the relationship between various fields [20]. Further, this research work presented an extensive and similar analysis of mystery sharing mechanisms for Secure Multiparty Computation (SMC) based systems, as well as their effectiveness.

The authors proposed the preservation of privacy assessment in data mining [21,22]. The strategies of secure multiparty computation (SMC), homomorphic encryption (HE), and comparison have been useful. In the work of Sahu et al., current needs and importance of industry related to reliability have been discussed, which includes the various methods was recommended [23]. They included AHP in that research work with three parameters [24]. In a study, the approach was modified by Alzahrani [25]. In this method, the ANP approach was used to identify and prioritise the modes of failure and apply the fuzzy TOPSIS technique to get the risk factors' weight in the operations in a study.

Kumar et al., have underscored the details that in IS security, human factors' importance has been identified by both researchers and organisations, and they conducted a study survey on information security for user awareness [26,27]. However, prior research on user security behaviour has only focused on computer theft [28]. Kumar et al., have presented a model, which shows the judgments of respondents by defining the suitable functions of membership in a questionnaire with a 5-point Likert scale [29].

3 Proposed Methodology

3.1 Evaluation of Security Privacy Attributes

The authors simultaneously consider the risks and benefits of sharing personal information or reference information. Everything has two sides, so technology offers easiness via social media and mobile phones on one hand, but on the other hand, it has offered many issues related to privacy, security, and data availability. Between all the problems discussed, such as users' privacy and security, are of paramount importance, as users are disclosing their personal information on social media just for their own enjoyment [30]. Hence, when users expose, or we can say, provide the accessibility of their personal information on social media, then they simultaneously consider privacy risk and benefits. Security and accessibility attributes discussed here:

3.1.1 Security

Information security entails more than just protecting data from unauthorised access. The practise of preventing unauthorised access use, disclosure, disruption, alteration, inspection, recording, or destruction of information is known as information security. There are two types of information: physical and electronic. Information can be anything from your personal information to your social media profile, cell phone, biometric and so on. As a result, information security encompasses a wide range of academic topics, including cryptography, mobile computing, cyber forensics, and online social media, among others.

Confidentiality: The efforts of an organization to keep their data private or secret are referred to as confidentiality. In practise, it means restricting data access to prevent illegal disclosure. Typically, this entails ensuring that only authorised individuals have access to specified assets, and that unauthorised individuals are actively prevented from gaining access.

Integrity: Integrity refers to the property of something being full or complete in everyday usage. Integrity refers to ensuring that data has not been tampered with and so can be trusted. It is accurate, authentic and trustworthy.

Availability: If system programmes, and data are not accessible when authorised users need them, they are of limited use to a business and its customers. Availability indicates that data, networks, system, and application are operational it ensures that authorised users get timely and consistent access to resources when they required.

Severity: The severity of a flaw is defined as the degree to which it can affect the data. The severity of a defect is a metric that indicates how serious it is and how much it affects the data or software's functionality.

3.1.2 Accessibility

When we talk about data accessibility, we are talking about reducing restrictions to properly utilising the data in today's databases. Great software solutions that offer expanded data access can enable anyone in any function to use their data as a single source of truth to draw important insights and make educated decision in their work. Some of accessibility techniques discussed here:

Privacy Concern: Privacy of data, also known as information privacy, is a subset of data security that focuses on the proper handling of data, including consent, notification and regulatory requirements. Practical data privacy problems frequently focus around: whether or not data is shared with third parties and how data is shared with third parties.

Self-disclosure: Self-disclosure is a type of communication in which individual shares personal information with other. Thoughts, sentiments, aspirations, objectives, failures, achievements, fears and hopes, as well as one's likes, dislikes and favourites can all be included in the information.

Benefits: Accessibility benefits all members of society, including those with impairments, it is safe to state. Improving accessibility improves one's quality of life by allowing them to be more independent and socially integrated. It also leads to greater health and can save money in a variety of ways.

Privacy-Risk: The potential loss of control over personal information is referred to as a privacy risk. Any information about an character maintained by means of a corporation, such as (1) and information that may be utilized to differentiate or trace an identity, such as name, number, BOB, mother's name, biometric records and (2) every other information this is related to an person inclusive of clinical, educational, financial, and employment information, according to the privacy Act.

A collection of reference information can disclose the patterns of the user's actual life. The context information that has been disclosed on the internet can become a rich source for a hacker. That means users have been provided access to others [31,32]. Therefore, the researchers believe that this context information is a personal matter, and when disclosing it, users have to take privacy risks and benefits into consideration. There is a demand for time to evaluate security-accessibility for safeguarding the security of big data for ease of usage and satisfaction. The evaluation of the results of security accessibility attributes should be analysed in-depth and can be used to improve the utility of security services. Priority analysis is performed by the Multi-Criteria Decision Making (MCDM) method [33,34].

With expert input, this work donates security prioritisation and privacy attributes via Fuzzy AHP. It detects the ranks and weight of security-privacy attributes. Fig. 2 demonstrates security components, i.e., confidentiality, integrity, accessibility, and severity, and privacy attributes such as privacy concern, self-disclosure, benefits, and privacy risk, which are more important than affecting the user's security. Security and privacy can be improved by focusing on their attributes together.

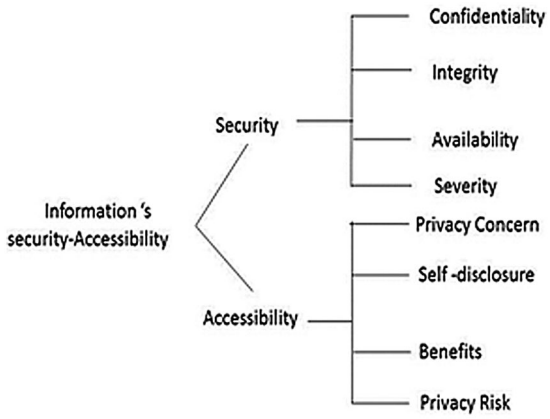

Figure 2: Hierarchy structure of attributes

3.2 Fuzzy AHP

A fuzzy set can be categorised via membership function. Fuzzy numbers can most commonly be used in two ways: the first is triangular numbers and the second is trapezoidal fuzzy numbers. The researchers considered fuzzy numbers to be triangular in this paper. Fig. 3 represents a triangular fuzzy number that is expressed as (f1, f2, f3) [35–37].

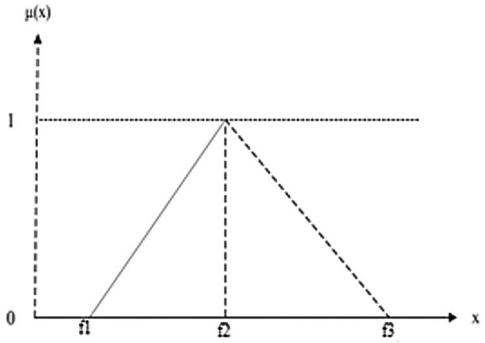

Figure 3: Triangular fuzzy number

Several academics have started a project that provides precise numbers. Fuzzy AHP is valuable because of its weights, whereas AHP is worthy of examining a decision in a group [34–39]. AHP is an essential apparatus that is often adopted by decision-makers. To handle the ambiguity and uncertainties of social decisions, the researchers came up with AHP's updated version, which is called Fuzzy AHP. The Fuzzy AHP, which includes a fuzzy theory with the AHP approach [6]. Hence, the priorities of security-accessibility attributes are essential to calculate the essential attributes of these eight factors. Also, the present of every attribute in security-accessibility is intended. The prioritisation of security-accessibility attributes to increase the security of information accessibility is discussed here. Researchers have proved that for a small-scale MCDM problem, AHP is the best arrangement approach [15–17].

The main aim of the paper is to find out the priority of security-accessibility factors. A questionnaire has been planned for this purpose. Thus, to answer the questionnaire, it is necessary to have a number of users who consider the information accessibility and its security. To evaluate the significance of security-accessibility factors, Fuzzy AHP is used because it can control the fuzzy decision-related inputs given by participants [6]. For the better valuation of security accessibility in the form of rankings and weights, then convert qualitative results into quantitative results [15,16]. Additionally, the pair-wise comparison matrix is arranged with the support of a questionnaire for the Fuzzy AHP approach. For calculating the weight of security-accessibility attributes, the applicant's inputs are converted into numeric values. The numeric values have been converted into Triangular Fuzzy Numbers (TFN) with the help of Eqs. (1) and (2) [17,18] and indicated as (Lo, Mo, Up), where Lo is the lowest possible, Mo is the most probable, and Up is the highest possible action.

With the help of Eqs. (1)–(4) transfiguring the linguistic terms into Triangular Fuzzy Number (TFNs).

$$\eta_{ij} = (\text{Lo},\ \text{Mo},\ \text{Up}) \tag{1}$$

where $\text{Lo} \leq \text{Mo} \leq \text{Up}$

$$\text{Lo} = (\text{J}_{\text{ijd}}) \tag{2}$$

$$\text{mi}_{\text{ij}} = (\text{J}_{\text{ij1}},\ \text{J}_{\text{ij2}},\ \text{J}_{\text{ij3}})^{\frac{1}{x}} \tag{3}$$

$$\text{and}\ \ \text{Up} = \ \ (\text{J}_{\text{ijd}}) \tag{4}$$

By the above situation, assumed two TFNs T1 and T2, T1 = (Lo_1, Mo_1, Up_1) and T2 = (Lo_2, Mo_2, Up_2). Further, Eqs. (5)–(7) help to aggregate TFN values.

$$T1 + T2 = (Lo_1,\ Mo_1,\ Up_1) + (Lo_2,\ Mo_2,\ Up_2) = (Lo_1 + Lo_2,\ Mo_1 + Mo_2,\ Up_2 + Up_2 \tag{5}$$

$$T1{*}T2 = (Lo_1,\ Mo_1,\ Up_1) \times (Lo_2,\ Mo_2,\ Up_2) = (Lo_1 \times Lo_2,\ Mo_1 \times Mo_2,\ Up_1 \times Up_2) \tag{6}$$

$$(T)^{-1} = (Lo_1,\ Mo_1, Up_1)^{-1} = \left(\frac{1}{Up_1}, \frac{1}{Mo_1}, \frac{1}{Lo_1}\right) \tag{7}$$

Step 2: Pair-wise decision matrix designed after the receiving the expert's response. The matrix is calculated by Eq. (8).

$$\widetilde{A^d} = [\tilde{k}^d_{11}\tilde{k}^d_{12}\ldots\tilde{k}^d_{1n}\tilde{k}^d_{21}\tilde{k}^d_{22}\ldots\tilde{k}^d_{2n}\cdots\tilde{k}^d_{n1}\tilde{k}^d_{n2}\ldots\tilde{k}^d_{nn}] \tag{8}$$

where, $\widetilde{k^k_{ij}}$ represents the dth decision makers' if more than one expert's are present, ith criteria preference over jth criteria.

The average of the preferences of each expert is assessment by the Eq. (9).

$$\tilde{k}_{ij} = \sum_{d=1}^{d} \tilde{k}^d_{ij} \tag{9}$$

Stage 3: In this stage authors update the pair-wise comparison matrixes for all factors in Fig. 3 on the behalf of results obtained from Eq. (10) with the help of Eq. (11).

$$\tilde{A} = \lfloor\widetilde{k_{11}}\ldots\widetilde{k_{1n}}\cdots\ddots\cdots\widetilde{k_{n1}}\cdots\tilde{k}_{nn}\rfloor \tag{10}$$

We obtained the geometrical mean and fuzzy weights of all factors by giving below Eq. (11).

$$\tilde{p}_i = \left(\prod_{j=1}^{n}\tilde{k}_{ij}\right)^{\frac{1}{n}},\quad i = 1,\ 2,\ 3\ldots n \tag{11}$$

Stage 4: We assessment the fuzzy weight of all factors by the Eq. (12).

$$\tilde{w}_i = \tilde{p}_i \otimes (\tilde{p}_1 \oplus \tilde{p}_2 \oplus \tilde{p}_3 \ldots \oplus \tilde{p}_n)^{-1} \tag{12}$$

After that, average weights are calculated and normalized weight of criteria by the Eqs. (13), (14).

$$M_i = \frac{\tilde{w}_1 \oplus \tilde{w}_2 \ldots \oplus \tilde{w}_n}{n} \tag{13}$$

$$Nr_i = \frac{M_i}{M_1 \oplus M_2 \oplus \ldots \oplus M_n} \tag{14}$$

Stage 5: In this stage, we have done the de-fuzzification of the values to obtain again the crisp values. The BNP (Best Non-fuzzy Performance) values of the fuzzy weights of every measurement by the Eq. (15).

$$BNPwD1 = \frac{[(uw1 - lw1) + (miw1 - lw1)]}{3} + lw1 \tag{15}$$

The values between two parameters have comparative significance from step 1 to step 6, where i and j represent the pair of norms determined by participants. For specific comparison, the value of aij is evaluated based on the geometric mean. The geometric mean can correctly collect and represent the values [6] and signify the lower and upper scores correspondingly for the comparative position among the two criteria. A pair-wise comparison matrix of fuzzy is evaluated as an n1 n2 matrix, after calculating the values of

TFN for every pair-wise evaluation. The size of the comparison matrix is 4 4, with a size of group limit to attain an adequate reliability level.

In this evaluation, participants include online users such as academicians and students with both security and accessibility experience. These candidates were selected to confirm the reliability of the AHP test. After the qualitative calculation, the pair-wise membership of TFN and the decision matrix have been created after the function is performed. The matrix was arranged by the authors after evaluating the responses [15]. In the third step, the TFN function and pair-wise comparison matrix are evaluated to generate a fuzzy decision matrix which has been established. Additionally, after creating the comparison matrix, to generate measurable values that depend on the TFN calculated values, de-fuzzification is performed. In this work, the de-fuzzification approach has been adopted from [15–19] as formulated in Eq. (15).

4 Result Evaluation

The researchers classify the security attributes (confidentiality, integrity, availability, and severity) as well as accessibility (privacy concerns, self-disclosure benefits, and privacy risk) attributes. Accessibility is a vital attribute for providing security at the time of disclosure of information. As a result, people and originators must understand the impact of both attributes in order to obtain secure information. So the researchers are using the Fuzzy Analytic Hierarchy Process (Fuzzy-AHP) method for the assessment in this paper. For the evaluation, a hierarchy needs to be established that describes the affected attributes. With the help of the hierarchy and Fuzzy AHP method, the security-accessibility of information has been calculated [6]. Outcomes obtained from the evaluation can support the security of information at the time of the data generation/creation phase in the big data life cycle [15].

Tab. 1 represents the security pair-wise comparison matrix of Fuzzy. The values of CR (Consistency Ratio) are less than 0.1. Tab. 2 shows the local weight of the de-fuzzified level 1 and pair-wise comparison matrix characteristics. Through the results, it is clear security is more important than accessibility to balance the accessibility of information. Tab. 3 represents the level 2 attributes for security, independent weight, and the de-fuzzified pair-wise comparison matrix. Further, CR values and the security-accessibility attributes' independent weight are represented in Tabs. 4–6, respectively. Fig. 4 represents the graphical representation of local weights at level 1. Fig. 5 also represents the weight levels 2 and 3. Tab. 7 represents the independent weights and attributes of level 2's de-fuzzified pair-wise comparison matrix for accessibility. Tab. 8 represents the overall ranking and dependent weights of the hierarchy.

Table 1: TFNs vale with their Linguistic terms

Saaty scale	TFNs value	
1	Equally	(1, 1, 1)
3	Poor	(2, 3, 4)
5	Good	(4, 5, 6)
7	Better	(6, 7, 8)
9	Best	(9, 9, 9)
2	Intermittent values	(1, 2, 3)
4		(3, 4, 5)
6		(5, 6, 7)
8		(7, 8, 9)

Table 2: For securities level 2 pair-wise comparison matrix of fuzzy AHP

	Confidentiality	Integrity	Accessibility	Severity
Confidentiality	1, 1, 1	0.433, 0.614, 0.866	0.117, 0.197, 0.309	1.41, 0.107, 0.79
Integrity		1, 1, 1	0.080, 0.119, 0.184	0.32, 0.67, 0.84
Accessibility			1, 1, 1	0.047, 0.064, 0.118
Severity				1, 1, 1

Table 3: Level 1's security and accessibility pair-wise comparison matrix

	Security	Accessibility	Weights
Security	1	0.877	0.725
Accessibility	0.122	1	0.275

Table 4: Level 2's security pair wise comparison matrix of fuzzy AHP

	Privacy concern	Self-disclosure	Benefits	Privacy risk	Weights
Privacy concern	1	0.750	0.574	0.432	0.459
Self-disclosure	0.285	1	0.236	0.195	0.187
Benefits	0.132	0.126	1	0.112	0.097
Privacy risk	0.063	0.179	0.741	1	0.257

Table 5: Results evaluation via fuzzy AHP

Attributes of 1st level	Level 1's local weights	Attributes of 2nd level	Level 2's local weights	Total weights
S1	0.725	S11	0.528	0.382
		S12	0.107	0.077
		S13	0.171	0.123
		S14	0.194	0.140
S2	0.275	S21	0.459	0.126
		S22	0.187	0.014
		S23	0.097	0.026
		S24	0.257	0.066

Table 6: Level 1's security and accessibility pair-wise comparison matrix

	Security	Privacy	Weights
Security	1	0.352	0.529
Accessibility	0.106	1	0.471

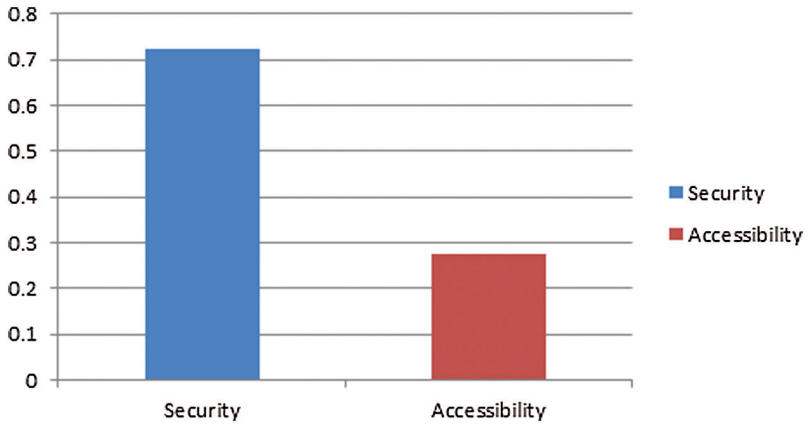

Figure 4: Local's weight of level 1

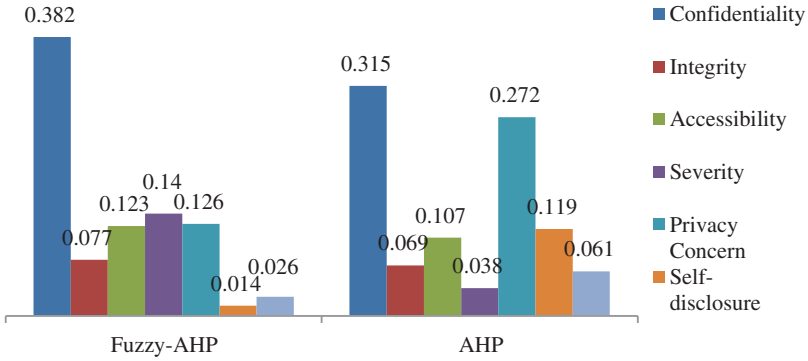

Figure 5: The comparison of fuzzy-AHP and AHP results

Table 7: Aggregated pair-wise comparison matrix for accessibility at level 2

	Privacy concern	Self-disclosure	Benefits	Privacy risk	Weights
Privacy concern	1	1.258	0.3876	0.4368	0.5576
Self-disclosure	0.1115	1	0.3876	0.3876	0.2516
Benefits	0.1840	0.0830	1	0.1292	0.1292
Privacy risk	0.0780	0.0503	0.0503	1	0.0624

Insecurity accessibility, security has 0.725 and accessibility has 0.275, which means security is more essential than accessibility and needs stability between accessibility and security. After implementing the Fuzzy AHP approach, another method has been used, known as the AHP method, to prove the complete accuracy of the result and estimation. To prove the results' accuracy, the security and accessibility effects of various types were evaluated by AHP. Data collection is a process in classical AHP, but there is only one difference in classical AHP is that no de-fuzzification is needed. Therefore, for classical AHP, the data comes in the form of crisp [16,17]. Fig. 5 depicts the graphical representation of the results.

A decision hierarchy is developed in the traditional AHP. Next, a pairwise matrix of participants' opinions has been established, but in this technique, numeric values are used directly rather than TFN values. Another phase is to combine the comparison matrixes pairwise of participants' opinions while

checking CR. Tab. 6 shows the level 1's independent weights and the combined comparison matrixes pairwise As a result, it is clear that security is more important than accessibility in improving overall security.

Further, Tab. 7 depicts the combined comparison matrix of pairwise and independent weights of the security attributes of level 2. Tab. 8 shows the level 2's local weights and the combined comparison matrix pair-wise of security factors. The dependent weights and the overall ranking of the hierarchy are shown in Tab. 9. The proposed algorithm is also compared with the existing approach based on security and accessibility attributes such as confidentiality, integrity, availability, severity, privacy concerns, self-disclosure benefits, and privacy risk (Tab. 10). In Tab. 11, the comparison results of security and accessibility evaluation via both the fuzzy AHP and traditional methods of AHP have been represented. It can be seen that while using the Fuzzy-AHP, results are better and more efficient in comparison to the AHP. The difference between the results from both the methods, Fuzzy AHP and AHP, is shown in Fig. 5 and Tab. 11.

Table 8: Level 2's security pair-wise comparison matrix

	Confidentiality	Integrity	Availability	Severity	Weights
Confidentiality	1	0.650	0.8040	0.3595	0.5973
Integrity	0.1194	1	0.0663	0.2157	0.1300
Availability	0.1493	0.3900	1	0.1438	0.2010
Severity	0.0830	0.0429	0.1005	1	0.0719

Table 9: Results evaluation via AHP

Level 1 attributes	Level 1 local weights	Level 2 attributes	Level 2 local weights	Total weights
S1	0.529	S11	0.5973	0.315
		S12	0.1300	0.069
		S13	0.2010	0.107
		S14	0.0719	0.038
S2	0.471	S21	0.5760	0.272
		S22	0.2516	0.119
		S23	0.1292	0.061
		S24	0.0624	0.030

Table 10: Comparison between the attributes

Security and accessibility attributes	Our approach	[1]	[6]
Confidentiality	High	High	Medium
Integrity	High	High	Medium
Accessibility	High	High	Medium
Severity	High	Less	Less
Privacy concern	High	Less	Less
Self-disclosure	High	Less	High
Benefits	High	Less	Less
Privacy risk	High	Less	Less

Table 11: Comparison between the fuzzy-AHP and AHP results

S. no	Security and accessibility attributes	Fuzzy-AHP	AHP
1.	Confidentiality	0.382	0.315
2	Integrity	0.077	0.069
3.	Accessibility	0.123	0.107
4.	Severity	0.140	0.038
5.	Privacy concern	0.126	0.272
6.	Self-disclosure	0.014	0.119
7.	Benefits	0.026	0.061
8.	Privacy risk	0.066	0.030

5 Conclusions

Nowadays, security is one of the essential quality attributes for all users. The purpose of this paper is to evaluate the security accessibility of information at the data creation phase of the big data security life cycle. The model proposed in this paper will help to calculate the security of information. In this contribution, the author has examined eight security and accessibility factors throughout the big data life cycle. In this research, the attributes of security and accessibility are recognized. Users want to access information as well as the security of their personal data. But providing access to security is a tedious task. In this paper, the calculation of security accessibility is a multi-purpose measure of decision difficulty. Fuzzy AHP has been used to calculate the security accessibility. Most essential characteristics have also been calculated. In the big data environment, if we want to provide a secure environment to any user, then we have to restrict access to information. For the assurance of security accessibility, at the data creation phase, developers need to restrict access to information and focus on the information's security.

Acknowledgement: We deeply acknowledge Taif University for supporting this study through Taif University Researchers Supporting Project Number (TURSP-2020/150), Taif University, Taif, Saudi Arabia.

Funding Statement: Funding for this study was received from the Taif University, Taif, Saudi Arabia under the Grant No. TURSP-2020/150.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] A. H. Almulihi, F. Alassery, A. I. Khan, S. Shukla, B. K. Gupta *et al.,* "Analyzing the implications of healthcare data breaches through computational technique," *Intelligent Automation & Soft Computing*, vol. 32, no. 3, pp. 1763–1779, 2022.

[2] F. M. Behlen and S. B. Johnson, "Multicenter patient records research: Security policies and tools," *Journal of the American Medical Informatics Association*, vol. 6, no. 6, pp. 435–443, 2018.

[3] K. Sharma, A. Agrawal, D. Pandey and R. A. Khan, "RSA based encryption approach for preserving confidentiality of big data," *Journal of King Saud University-Computer and Information Sciences*, article in press, vol. 8, pp. 1–18, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1319157819305592.

[4] M. Mohammadian and D. Hatzinakos, "A hierarchical fuzzy logic systems frame work for data security," *International Journal of Information Technology*, vol. 14, no. 5, pp. 147–157, 2017.

[5] R. Agrawal and R. Srikant, "Privacy preserving data mining," *Journal of Data Management*, vol. 17, no. 6, pp. 439–450, 2000.

[6] A. Attaallah, H. Alsuhabi, S. Shukla, R. Kumar, B. K. Gupta *et al.,* "Analyzing the big data security through a unified decision-making approach," *Intelligent Automation & Soft Computing*, vol. 32, no. 2, pp. 1071–1088, 2022.

[7] A. I. Shaik, 5 Major Data Breaches in India, 2021. [Online]. Available: https://www.91mobiles.com/hub/5-major-data-breaches-india-2021/.

[8] E. Kost, 6 Most Common Causes of Data Leaks in 2021. [Online]. 2021. Available: https://www.upguard.com/blog/common-data-leak-causes.

[9] G. Kapil, A. Agrawal, A. Attaallah, A. Algarni, R. Kumar *et al.,* "Attribute based honey encryption algorithm for securing big data: Hadoop distributed file system perspective," *PeerJ Computer Science*, vol. 6, no. 6, pp. 65–74, 2020.

[10] S. T. Duygu, T. Ramazan and S. Seref, "A survey on security and privacy issues in big data," *Internet Technology and Secured Transactions*, vol. 11, no. 16, pp. 1439–1450, 2015.

[11] K. Sharma, A. Agrawal and R. A. Khan, "An improved security threat model for big data life cycle," *Asian Journal of Computer Science and Technology*, vol. 2, no. 8, pp. 33–39, 2018.

[12] K. Sahu, F. A. Alzahrani, R. K. Srivastava and R. Kumar, "Hesitant fuzzy sets based symmetrical framework of decision-making for estimating the durability of web application," *Symmetry*, vol. 12, no. 6, pp. 1770–1792, 2020.

[13] J. Brickell and V. Shmatikov, "Privacy-preserving classifier learning," in *Proc. of 13th Int. Conf. on Financial Cryptography and Data Security*, Accra Beach, Barbados, pp. 1–6, 2009. [Online]. Available: https://dblp.org/db/conf/fc/index.html.

[14] K. Sahu, F. A. Alzahrani, R. K. Srivastava and R. Kumar, "Evaluating the impact of prediction techniques: Software reliability perspective," *Computers, Materials & Continua*, vol. 67, no. 2, pp. 1471–1488, 2021.

[15] R. Kumar, A. I. Khan, Y. B. Abushark, M. M. Alam, A. Agrawal *et al.,* "An integrated approach of fuzzy logic, AHP and TOPSIS for estimating usable-security of web applications," *IEEE Access*, vol. 8, no. 8, pp. 50944–50957, 2020.

[16] S. A. Khan, M. Alenezi, A. Agrawal, R. Kumar and R. A. Khan, "Evaluating performance of software durability through an integrated fuzzy-based symmetrical method of ANP and TOPSIS," *Symmetry*, vol. 12, no. 4, pp. 1–15, 2020.

[17] M. T. J. Ansari, A. Baz, H. Alhakami, W. Alhakami, R. Kumar *et al.,* "P-STORE: Extension of store methodology to elicit privacy requirements," *Arabian Journal for Science and Engineering*, vol. 64, no. 3, pp. 1–25, 2021.

[18] R. Kumar, M. Zarour, M. Alenezi, A. Agrawal and R. A. Khan, "Measuring security durability of software through fuzzy-based decision-making process," *International Journal of Computational Intelligence Systems*, vol. 12, no. 2, pp. 627–642, 2019.

[19] M. Duckham and L. Kulik, "A formal model of obfuscation and negotiation for location privacy," *Pervasive Computing*, vol. 8, no. 8, pp. 243–251, 2005.

[20] K. Sahu and R. K. Srivastava, "Soft computing approach for prediction of software reliability," *ICIC Express Letters*, vol. 12, no. 12, pp. 1213–1222, 2018.

[21] R. Kumar, A. I. Khan, Y. B. Abushark, M. M. Alam, A. Agrawal *et al.,* "A knowledge-based integrated system of hesitant fuzzy set, AHP and TOPSIS for evaluating security-durability of web applications," *IEEE Access*, vol. 8, no. 8, pp. 48870–48885, 2020.

[22] H. Deng, "Multicriteria analysis with fuzzy pairwise comparison," *International Journal of Approximate Reasoning*, vol. 5, no. 11, pp. 215–231, 2019.

[23] K. Sahu and R. K. Srivastava, "Needs and importance of reliability prediction: An industrial perspective," *Information Sciences Letters*, vol. 9, no. 1, pp. 33–37, 2020.

[24] K. Sahu and R. K. Srivastava, "Predicting software bugs of newly and large datasets through a unified neuro-fuzzy approach: Reliability perspective," *Advances in Mathematics: Scientific Journal*, vol. 10, no. 1, pp. 543–555, 2021.

[25] F. A. Alzahrani, "Evaluating the usable-security of healthcare software through unified technique of fuzzy logic, ANP and TOPSIS," *IEEE Access*, vol. 8, no. 6, pp. 109905–109916, 2020.

[26] R. Kumar, M. T. J. Ansari, A. Baz, H. Alhakami, A. Agrawal *et al.,* "A Multi-perspective benchmarking framework for estimating usable-security of hospital management system software based on fuzzy logic, ANP and TOPSIS methods," *KSII Transactions on Internet and Information Systems*, vol. 15, no. 1, pp. 240–263, 2021.

[27] R. Kumar, S. A. Khan and R. A. Khan, "Durability challenges in software engineering," *CrossTalk*, vol. 42, no. 4, pp. 29–31, 2016.

[28] R. Kumar, S. A. Khan and R. A. Khan, "Revisiting software security: Durability perspective," *International Journal of Hybrid Information Technology*, vol. 8, no. 2, pp. 311–322, 2015.

[29] R. Kumar, A. Baz, H. Alhakami, W. Alhakami, A. Agrawal *et al.,* "A hybrid fuzzy rule-based multi-criteria framework for sustainable-security assessment of web application, *Ain Shams Engineering Journal*, vol. 12, no. 2, pp. 2227–2240, 2021.

[30] P. Bunn and R. Ostrovsky, "Secure two-party k-means clustering," *Computer and Communications Security*, vol. 42, no. 4, pp. 486–497, 2007.

[31] K. Sahu and R. K. Srivastava, "Revisiting software reliability," *Advances in Intelligent Systems and Computing*, vol. 802, pp. 221–235, 2019. https://doi.org/10.1007/978-981-13-1402-5_17.

[32] O. Mohsen and N. Fereshteh, "An extended VIKOR method based on entropy measure for the failure modes risk assessment–A case study of the geothermal power plant (GPP)," *Safety Science*, vol. 12, no. 14, pp. 160–172, 2017.

[33] Y. Ozdemir, Y. M. Gul and E. Celik, "Assessment of occupational hazards and associated risks in fuzzy environment: A case study of a university chemical laboratory," *Human and Ecological Risk Assessment: An International Journal*, vol. 6, no. 10, pp. 895–924, 2017.

[34] R. Kumar, S. A. Khan, A. Agrawal and R. A. Khan, "Measuring the security attributes through fuzzy analytic hierarchy process: Durability perspective," *ICIC Express Letters*, vol. 12, no. 6, pp. 615–620, 2018.

[35] R. Kumar, S. A. Khan and R. A. Khan, "Analytical network process for software security: A design perspective," *CSI Transactions on ICT*, vol. 4, no. 2, pp. 255–258, 2016.

[36] J. Zhang, B. Reithel and H. Li, "Impact of perceived technical protection on security behaviors," *Information Management &Computer Security*, vol. 5, no. 6, pp. 330–340, 2009.

[37] A. Azar and Z. A. Darvishi, "Development and validation of a measure of justice perception in the frame of fairness theory-fuzzy approach," *Expert System Application*, vol. 8, no. 2, pp. 7364–7372, 2011.

[38] Y. Chan, M. Woon and A. Kankanhalli, "Perceptions of information security at the workplace: Linking information security climate to compliant behavior," *Journal of Information Privacy and Security*, vol. 11, no. 12, pp. 18–41, 2005.

[39] R. Fattahi and M. Khalilzadeh, "Risk evaluation using a novel hybrid method based on FMEA, extended multimoora, and AHP methods under fuzzy environment," *Safety Science*, vol. 5, no. 7, pp. 290–300, 2018.