Cluster Computing (2022) 25:3037–3049
https://doi.org/10.1007/s10586-021-03514-x

Achieving data privacy for decision support systems in times of massive data sharing

Rabeeha Fazal[1] · Munam Ali Shah[1] · Hasan Ali Khattak[2] · Hafiz Tayyab Rauf[3] · Fadi Al-Turjman[4]

Received: 1 May 2021 / Revised: 30 November 2021 / Accepted: 6 December 2021 / Published online: 10 January 2022

Abstract

The world is suffering from a new pandemic of Covid-19 that is affecting human lives. The collection of records for Covid-19 patients is necessary to tackle that situation. The decision support systems (DSS) are used to gather that records. The researchers access the patient's data through DSS and perform predictions on the severity and effect of the Covid-19 disease; in contrast, unauthorized users can also access the data for malicious purposes. For that reason, it is a challenging task to protect Covid-19 patient data. In this paper, we proposed a new technique for protecting Covid-19 patients' data. The proposed model consists of two folds. Firstly, Blowfish encryption uses to encrypt the identity attributes. Secondly, it uses Pseudonymization to mask identity and quasi-attributes, then all the data links with one another, such as the encrypted, masked, sensitive, and non-sensitive attributes. In this way, the data becomes more secure from unauthorized access.

Keywords Data privacy · Encryption · Blowfish · Data masking · Identity data · Sensitive data · Non-sensitive data

1 Introduction

As of July 2021, people all over the world are suffering from the novel Coronavirus called Covid-19. The outbreak of Covid-19 has introduced many difficulties. The researchers need Covid-19 records to help us tackle that situation, so the Decision Support System (DSS) is used to gather those records [1].

The researchers are trying to help us through collected data.[1] They access the patient's data through DSS and make predictions on the severity and effect of the Covid-19 disease. Here the main concern unauthorized access may become a privacy breach, so data privacy must be maintained.[2]

A medical organization cannot afford the data leakage of their Covid-19 patients.[3] Data breaches can affect patients' lives and the reputation of institutions.[4] In this pandemic, data privacy is necessary for patients of Covid-

✉ Hasan Ali Khattak
hasan.alikhattak@seecs.edu.pk

Rabeeha Fazal
rabeehafazal786@gmail.com

Munam Ali Shah
mshah@comsats.edu.pk

Hafiz Tayyab Rauf
h.rauf4@bradford.ac.uk

Fadi Al-Turjman
fadi.alturjman@neu.edu.tr

1 Department of Computer Science, COMSATS University Islamabad, Islamabad, Pakistan

2 School of Electrical Engineering and Computer Science (SEECS), National University of Sciences and Technology (NUST), H12, Islamabad, Pakistan

3 Department of Computer Science, Faculty of Engineering and Informatics, University of BRADFORD, Bradford, UK

4 Artificial Intelligence Department, Research Center for AI and IoT, Near East University, Nicosia, Mersin 10, Istanbul, Turkey

[1] https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html.

[2] https://www.ncbi.nlm.nih.gov/books/NBK9579/.

[3] https://www2.deloitte.com/be/en/pages/risk/articles/privacy-and-data-protection-in-the-age-of-covid-19.html.

[4] https://blog.netwrix.com/2019/11/05/data-privacy-trends-issues-and-concerns-for-2020/.

19, and for that reason, we proposed a privacy model. Our contribution is in two folds. First, we used the Blowfish algorithm for the encryption of data. It preserves data privacy and takes less time to execute. The Blowfish encryption is applied to identity data attributes such as name, phone number, and quasi attributes such as an address, gender, and age. The authorized user can access it using the key. The benefit is that the patient's data is safe in a repository securely. It can be accessed only by authorized users [2].

The problem arises here for research if medical organizations share actual data, e.g., only encrypted data, individual identity may reveal. For that reason, the second fold used the Pseudonymization masking technique. It published mask data that protect individual's privacy and remains used for research. Then masked data, non-sensitive and sensitive data, is associated with the help of a reference number to encrypted data. In that way, data may become secure and can be used for research or any medical purpose. If one publishes an individual's data without masking, it may chance a data breach. So, the aim is to protect that individual's privacy while using that data [3].

We analyzed these encryption algorithms: the Blowfish, AES (stands for advance encryption standards), DES (stands for data encryption standards), 3DES (stands for triple data encryption standards), IDEA (stands for the international data encryption algorithm), RSA (stands for Rivest, Shamir, Adleman), and RC6 (stands for Rivest cipher 6). Afterward, we proposed a hybrid model of the Blowfish and the Pseudonymization masking technique to protect data from malicious use. The Blowfish used the parameters that are execution time and best against known attacks. [4]

Our experimental evaluation through proof of concept implementation validates that Blowfish presents a suitable privacy-preserving mechanism for achieving data privacy especially considering healthcare data [5]. The Blowfish takes less time for execution among the algorithms as mentioned above. In the Pseudonymization masking technique, we masked the name, address, and gender, then associated sensitive attributes of covid+ information. After this masking, it is hard to reveal the patient's identity. At last, all encrypted, masked, sensitive, and non-sensitive attributes are associated with each other.

1.1 Related topic papers

The health insurance portability and accountability act (HIPPA) has been proposed by Leslie Lener et al. [6]. HIPPA rules and regulations provide guidelines for the data privacy of Covid-19 patients, but they have not presented any model to preserve data privacy. The guideline approach has been proposed by the Zwitter et al. [7]. That discusses how to deal with data to preserve privacy and provides guidelines but does not give an approach to handling such data. The regulation of the mobile positioning approach has been discussed by Iniobong Ekong et al. [8]. That limits unauthorized access and achieves data privacy. This approach only focuses on tracing Covid-19 users using the mobile tracing approach on the regulation provided by Nigeria's regulations. Functional encryption for securing data using the Spatio-temporal trajectory approach has been proposed by Wooil Kim et al. [9]. The Spatio-temporal approach tends to achieve security through contact tracing. However, their work is only focused on privacy for contact tracing that is not suitable when we require to ensure the privacy of Covid-19 patient's data in DSS.

1.2 Scope

This paper discusses the privacy model to protect the Covid-19 patient data from unauthorized access. If information leaks, it might cause data breaches. That's why it is vital to secure the patient's information. This proposed approach deals with privacy for Covid-19 Patients. We analyzed the existing solutions used for data protection in different research papers shown in Fig. 1. The other papers used HIPPA rules which only focus on providing guidelines, and some of them used contact tracing approaches to preserve data privacy. Our proposed hybrid model is used to achieve data privacy using a Blowfish encryption algorithm and data masking techniques and provides both privacy and efficacy for Covid-19 data uses in health organizations and research. The efficacy was measured in terms of minimum time utilization. So, by using this hybrid model, we achieved data privacy for Covid-19 patients.

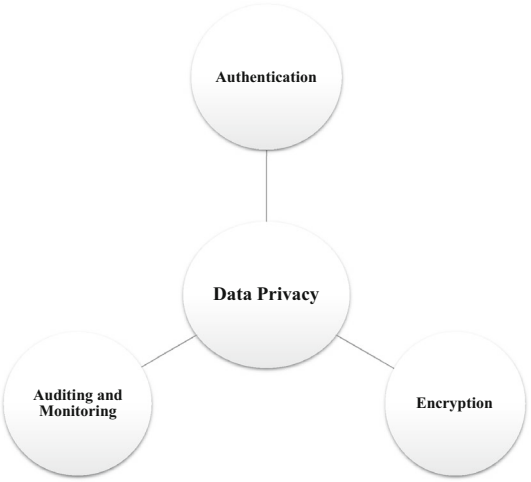


Fig. 1 Existing technologies used to achieve data privacy

2 Related work

In the medical setups, the DSS is used to collect and manage Covid-19 patient data. That data needs to be kept hidden so that individuals stay as protected a lot as could be expected. We performed an analysis of some technologies used to achieve privacy. Those are authentication, encryption, auditing, and monitoring. Here, we have presented them in a taxonomy diagram in Fig. 2. Below all those technologies are discussed to achieve data privacy [10].

2.1 Authentication

The oPass protocol proposed by Hung-Min Sun et al. that used for authentication [11]. Which controls stealing and reuse attacks of passwords for login so that data becomes safe. This model focuses on the one-time password for authentication. The model of Anak Agung Putri Ratna et al. measuring the time against the brute force attacks against the algorithms SHA-1 called secure hash algorithm, and MD5 called message digest5 [12]. It focuses only on measuring brute force against two algorithms. The OTP called as One-time-password used by the Rama et al. [13]. It's an authentication protocol, which prevents stealing the password and other attacks. But the danger is that the man in the middle attack can become a breach of privacy. The Secure-sockets-layer (SSL)/ Transport layer security(TLS) is used for further secure communication. The flaws and security issues about SSL and TLS are also considered by the Preeti Sirohi et al. [14]. There are also some open challenges like a logjam, Monstrosity, SSL, and stripping attack. The SSL/TLS is used for secure communication, but still, it has many vulnerabilities, i.e., dangers of attack.

The prevention from unauthorized access and proves the purpose of intent using factor authentication(FA) proposed by Atsushi Kogetsu et al. [15]. The comparison between factor authentication has been performed, but new technologies are ignored to prove which one is best in what type of authentication. The CAPTCHA stands for Completely Automated Public Turing test to tell Computers and Humans Apart OTP stands for the one-time password has been proposed by Thivanon Kansuwan et al. [16]. The email OTP is used for data protection instead of using the old procedures of login because if that email account is not logged in and the mobile phone is lost, in that situation, email OTP is secure than messages. Covid-19 Test Certification (CTC) ensured patients' privacy approach proposed by Untung Rahardja et al. [17]. The CTC is used for confirmation of test results in the distributed system. This method discussed data privacy while checking online test results after Covid-19 testing. The approach only deals with the privacy of test results which is not enough.

2.1.1 oPass

oPass named authentication protocol, that is used for authentication and attack protection. oPass protection is best against attacks like phishing, keylogger, password reuse, password guessing, and a MITM the man in the middle attack [11].

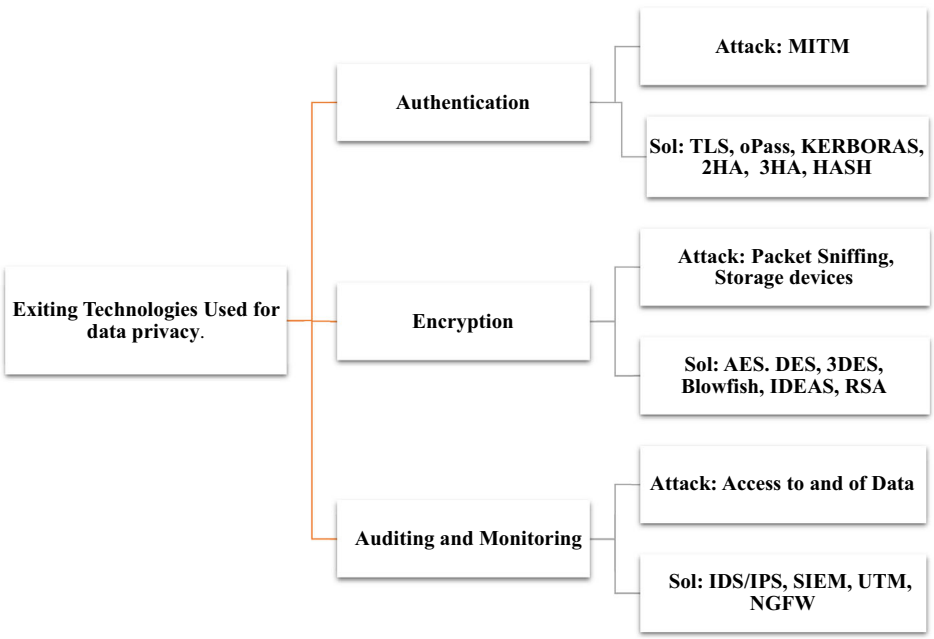


Fig. 2 Overview of existing technologies used for data privacy

2.1.2 Features of SHA-1 and MD5

The Secure hash algorithm 1 (SHA-1) and Message Digest 5 (MD5) is an authentication-based security system [12]. The SHA-1 is best against brute force attacks. The MD5 is best for memory usage and processing time.

2.1.3 OTP

The hash-based one-time password (HOTP), the time-based one-time password (TOTP), and challenge-response one-time password CROTP are used for the prevention to replay attacks. The HOTP prevents replay attacks but takes more processing time and a higher CPU overhead. The CROTP has used prevention from replay attacks, CPU overhead is high, and medium response time for the server. The TOTP also prevents replay attacks, takes less time to complete the process, and has low CPU overhead [13, 16].

2.1.4 TLS

The security in Google, Chrome, Firefox, and Safari support the TLS protocol for secure communication or safe searching [14, 18]. The key features required for TLS 1.3 version in any website; are Certificate, Key Exchange, Cipher Strengths, and Protocol Support[5] [19].

2.1.5 Factor authentication

Factor authentication is a method of security.[6] We analyzed different Factor authentication variants that are one-factor authentication (1FA), Two-factor authentication (2FA), and Three-factor authentication (3FA). That discussed their procedure, methods, and their best use. The 2FA has a predetermined code and security of more than 1FA and less than 3FA. The 1FA deals with what entity knows, the method used for, e.g., ID, password, and the security less than 2FA. The 2FA deals with what entity has, the method used codes or signed digital certificate or fingerprint, less security than 3FA. The 3FA deals with what entity remains, a method used for voice, hand, fingerprint, and retina scan, more secure than others [15].

2.2 Encryption

The cryptography algorithms; Data Encryption Standard DES, Triple Data Encryption Standard 3DES, and Advanced Encryption Standard AES discussed by Hamdan. O. Alanazi et al. [20]. These algorithms are compared for measuring effectiveness, adaptability, and security to secure data [21, 22]. It gives us brief information about the different algorithms, and the main focus is to find a secure algorithm. The comparison of encryption algorithms has been presented by Aggarwal [23]. The discussion about encryption is based on the best algorithm according to a situation to compare the effectiveness. Here the encryption algorithms are compared for two parameters. The encryption algorithms AES as advanced encryption standard, DES data encryption standard, T-DES triple-data encryption standard, and RSA Rivest, Shamir, Adleman, proposed by Pankaj Singh et al. [24]. These encryption algorithms were used for maintaining data privacy that gave a reasonable correlation between encryption and speed. However, these encryption algorithms are used to ensure privacy and speed, which impact the performance. The encryption algorithms AES, DES, RSA, RC6, 3DES, and Blowfish have been compared by Mohammed Nazeh Abdul Wahid et al. [25]. The algorithms are used to secure data from unauthorized access. This study focuses on the encryption algorithm where they can perform best.

The comparison of encryption algorithm has been presented by Patel et al. [26]. The discussion about encryption is based on execution time and memory usage to measure the performance. Here the encryption algorithms are compared for two parameters. These encryption algorithms (RSA, Blowfish, 3DES, and AES)compared for achieving data privacy, have been discussed by Daniel Commey et al. [27]. The main focus is upon choosing cipher, which provides more security. The encryption, auditing, and authentication solutions to achieve data privacy used SHA512, SHA256, and AES encryption techniques have been discussed by Arielle V. Luccal et al. [28]. The use of these algorithms provides security from unauthorized access. The Encryption applied to the prototype app was used to gather data. The prototype app was used to monitor a patient's condition and decide on discharge. This model focuses on condition-based monitoring securely. CryptoGA (GA) encryption is a Genetic cryptographic algorithm has been discussed by Muhammad Tahir et al. [29]. The GA compared with some algorithms, which denoted the effective transmission rate. It focuses upon providing security to be used regardless of location. K-Medoid and BLOWFISH encryption has been proposed by Dr. Sheena Hussaini et al. [30]. The K-Medoid algorithm is used for the reliability of clustering data, whereas Blowfish demonstrates more security for that data. However, it focuses on distance-based secure encryption.

This paper discusses the weakness of some encryption algorithms and proposes a new approach that ensures the reliability and security of data. Different algorithms (AES, RSA, MD5, DES, Blowfish, and SHA, regarding their execution time and memory) has been discussed by

[5] https://blog.cloudflare.com/rfc-8446-aka-tls-1-3/.

[6] https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/biometrics/case-study.

Ashwini P. Parka et al. [31]. A comparative analysis is performed in this paper for choosing the best algorithm regarding their performance.

2.2.1 Feature analysis

Encryption is to encrypt and hide data from unauthorized users. If any unauthorized attempt tries to access the data, it becomes unreadable. This section performed the analysis of encryption algorithms [20, 23–31].[7,8] In Table 1, there is an analysis of some encryption algorithms such as Blowfish, AES, DES, 3DES, Rivest cipher 6 RC6, RSA, IDEAS. The Table 1, shows different algorithms regarding the prevention from attacks and the effect of these algorithms.

Table 2 analyzes different algorithms regarding some parameters such as the block size (is denoted group of bits), key size (the length of the key), round (depends upon the key size), confidentiality, execution-time for encryption/decryption, power consumption, and memory usage. The confidentiality of Blowfish is more than the rest of all algorithms. The Blowfish encryption and decryption take less time, less power consumption, and low memory consumption. AES takes the shortest time for encryption and decryption, and it has more memory usage than Blowfish. DES takes the highest time for encryption and decryption, and it uses more memory than AES. The 3DES takes more execution time than DES for encryption and decryption. The IDEA, RC6, and RSA are high in all these mentioned parameters.

2.3 Auditing and monitoring

The unified threat management (UTM) has been discussed by Yin Chao et al. [32]. Although UTM features have been discussed here, the issue is that it's used for small security solutions. The Lidong Wang et al. paper discussed intrusion, typically done by individuals outside the association [33]. The IDS/IPS intrusion detection and intrusion prevention system are discussed to deal with intrusion detection; however, IDS/IPS are not useful in all situations. A detailed view of IDS/IPS technologies is given by Karim Abouelmehdi et al. [34]. IDS and IPS are used to observe, gather, and analyze the system to get the interruptions, keep up a log of each entrance, and adjust information. They are used for detection and prevention, but also they cannot handle any encrypted data which can be malicious [35]. The Next-generation firewall (NGFW) has been discussed by Kishan Neupane et al. [36]. It's analyzed against conventional firewalls, other security solutions and also discusses its objectives and danger. However, the configuration is not easy for NGFW. The security information and event management SIEM has been discussed by Sievierinov et al. [37]. It is used for monitoring, and it provides hardware, network, and application analysis. It is used as the generation of logs and reports. This security solution is used for mid-sized setups.

The overall analysis of the literature presents us with the conclusion that most of the capabilities in auditing technologies are present in NGFW as presented in Table 3. Table 3 is an analysis of all the technologies that are discussed in this section.

2.3.1 IDS/IPS

IDS passively monitors and detects intrusion activities in any system. IPS actively analyzes and prevents intrusion. The IDS variants, host-based intrusion detection system (HIDS), and the network-based intrusion detection system (NIDS) are used to detect anomalies on both host and networks. Moreover, an IPS has variants, included host-based intrusion prevention systems (HIPS) and network intrusion prevention systems (NIPS). The HIDS provides system-level protection, configuration changes, file changes, and registry changes. The NIDS provides features of network resources, network protection, denial of service attacks protection, and sniffs the network traffic continuously if irregularities found in traffic detect and generate an alarm. The applications of IPS and IDS variants are: HIDS have ISS, Symantec Enterasys, and HIPS have Cisco, McAfee Snort. The NIDS applications are ISS, Cisco, Enterasys Symantec, and McAfee Intrushied NetScreen TippingPoint are the application of NIPS. [34].

2.3.2 NGFW features

NGFW integrates all those deep inspections of the packet, IDS, and IPS, visibility of application regardless of protocol and ports, and access control policies. The NGFW is used for the following facilities because it provides high-intensity traffic environments, complex tasks, telecommunication, deep inspection of packets, cohesive architecture, and access control policies [36].

2.4 HIPPA

The HIPPA rules and CIA security trade have been discussed by Thapa et al. [38]. They focused on regulations because of ethical requirements; however, they only discussed rules and regulations. The guidelines for data protection of Covid-19 patient have been discussed by Bernier et al. [39]. They focus on reforms and guidelines for the data protection of Covid-19 patients. This paper discussed

[7] RC6 https://simple.wikipedia.org/wiki/RC6.

[8] IDEA https://en.wikipedia.org/wiki/International_Data_Encryption_Algorithm.

Table 1 An analysis of different encryption algorithms

Algorithm	Attack	Structure	Effect
IDEAS [23][a]	Narrow Bicliques	Feistel	Optional Algorithm in openpgp
Blowfish [25–27]	Dictionary	Feistel	Ideal for both domestic and exportable use
AES [20, 25–27]	Side Channel 1	Substitution, Permutation	Secured by TLS or SSH
3DES [20, 25, 27]	Brute force, Known Plaintext	Feistel	Insecure for modern application
DES [20, 25, 26]	Brute Force	Feistel	Insecure for modern application
RC6 [23][b]	Brute Force, Analytical	Substitution,Permutation	Essential for data dependent decisions
RSA [20, 25, 27]	Factoring the Public Key	Factorization	The exchange key establishes a secure connection.

[a]IDEA

[b]RC6

Table 2 A features comparison between different encryption algorithms

Features	Blowfish [25–27]	AES [20, 25–27]	DES [20, 25, 26]	3DES [20, 25, 27]	IDEAS[23][a]	RC6 [23][b]	RSA [20, 25, 27]
Block size	64	128	64	64	64	128	1024-4096
Round	16	10,12,14	16	48	8.5	20	1
Confidentiality	Highest	High	Low	Low	Low	Low	High for key
Key-size (bits)	32-448	128,192,256	56	112,168	128	128-2040	1024-4096
Execution time	Shortest	Short	Highest	Highest	Highest	Highest	Highest
Power consumption	Less	Less	Highest	Highest	Highest	Highest	Highest
Memory usage	Low	>Blowfish	>AES	>DES	Highest	Highest	Highest

[a]IDEA

[b]RC6

Table 3 A comparison of auditing and monitoring technologies

Abilities	UTM [32]	SIEM [37]	IDS [33, 34]	IPS [33, 34]	NGFW [36]
Junk Mail Filtering	✓	✗	✗	✗	✓
Review	✓	✓	✓	✓	✓
Take Action	✓	✓	✗	✓	✓
Unresponsively Monitor	✗	✓	✓	✗	✓
Alert on Unusual Activity	✗	✗	✗	✓	✓
Integrate(firewall,IDS/IPS,AV)	✓	✓	✗	✗	✓
Firewall,AppControl,Web APPs	✓	✗	✗	✗	✓
Control filtering	✓	✗	✗	✗	✓

reforms to protect data privacy. The privacy for patients achieved via access control of this model has been discussed by Prince et al. [40]. This paper classified the system into three parts to determine confidential, medium, and low confidential data. That helps determine the protection of data. The data privacy guidelines for designing a system have been proposed by AlMarzooqi et al. It's specifically for Dubai [41]. This paper discussed guidelines to secure data. We analyzed all approaches discussed in various papers for data privacy. That helped us to choose the appropriate system for providing data privacy to Covid-19 patients.

3 Proposed model

The Covid-19 patient's data need protection from unauthorized access. The question arises here how to accomplish this. We proposed a hybrid approach for the privacy of Covid-19 patients. We analyzed some encryption algorithms to determine which technique is best for encrypting data for Covid-19 patients. We classified attributes of Covid-19 patients to identify what attributes need to be encrypted and masked. The classified attributes of Covid-19 patients are:

- Identity Attribute: Name, Phone number, and ID card.
- Quasi Attribute: Age, Gender, Address.
- Sensitive Attribute: Covid-19 test result positive.
- Non-sensitive Attribute: other than quasi attributes.

As shown in the taxonomy diagram Fig. 3, the Hybrid model divides into two parts: Blowfish and Pseudonymization. Blowfish encrypts the identity attributes, name, phone no, Id card, and quasi attributes, address. Pseudonymization masks identity data and quasi data such as random data, address with region and gender with the person and then associates this reference data with encrypted data of Blowfish for reference. In, that way researchers can use it for research purposes to overcome pandemic privacy breaches. The masked and encrypted attributes are:

- Identity Attribute = Blowfish Encryption + Associated with the reference number.
- Identity Attribute= Mask (Name, Phone Number) + Associated with the reference number.
- Quasi Attribute= Address masked with intervals, gender mask with the person.
- Sensitive Attribute = Covid-19 positive data remain the same.

The hybrid approach is used for data privacy as shown in Fig. 4. The Blowfish is used for encryption and decryption of identity data such as name and phone number. It is also used for quasi attributes, such as address. The parameter for this experiment is execution time and best against known attacks. The benefit of this approach is that the patient's data is saved in a repository securely. However, this data cannot be used for research purposes. If medical organizations share encrypted data, an attacker can make an attack. For that reason, the second fold used the pseudonymization masking technique. We masked identity attribute the name with random data, quasi attributes address, and age respectively with the region and intervals. After, masked, sensitive, non-sensitive, and encrypted data are associated with each other. In that way, researchers can use it for research to overcome the pandemic without privacy breaches. It may reduce the risk of privacy breaches because the patient's information is masked. Due to address changes with the region, the researcher can find spots in the particular regions without the privacy breaches. In this section, we discussed how we achieve data privacy for the Covid-19 patient. We described how things relate to each other.

4 Experiments and results

In this section, we performed experiments for the hybrid approach, and It gave us interesting results. The machine setup is in window 10 with 8 GB RAM. The data masking is done with a python script, and all encryption experiments were performed in the java language. The modified dataset used for this approach is the Adult dataset accessible a.[9] We added sensitive, i.e., covid-19, and gender columns in the adult dataset to measure our results. All the data is masked using Pseudonymization. We have masked data in a way it remains beneficial for medical and research. After this, we performed Blowfish encryption that took less time. Figure 5 shows the execution time for algorithms that are Blowfish, AES, DES,3DES, IDEAS, RC6, and RSA. The parameter for this experiment is execution time. Our hybrid approach showed efficient results.

The result is shown in Fig. 5. The Blowfish takes minimum time for encryption. The Covid-19 patient's data is confidential: name, address, phone, and sensitive attribute Covid-19 positive need to be encrypted for this, so we used Blowfish encryption, which took less time. The result shows it takes only 0.72 milliseconds during encryption and decryption of the data. We also have used online tools to run these algorithms for Blowfish. WE used blowfish.js encrypt/decrypt online for RSA and used RSA Encryption Decryption for IDEA, DES, 3DES, and RC6. We used 8gwifi Crypto Tool Playground for AES, encryption, and decryption Online.[10,11,12,13]

The eclipse and online tools are used for more results and measured execution time for encryption. Different cipher modes are used to measure execution time for the encryption algorithm. These cipher modes are electronic codebook (ECB), cipher block chaining (CBC), cipher feedback (CFB), output feedback (OFB), output feedback mode in-bits (NOFB), counter (CTR), and the propagating cipher block chaining (PCBC). The execution time is measured using those modes in a different algorithm. It gives us fantastic results.

[9] https://archive.ics.uci.edu/ml/datasets.
[10] http://sladex.org/blowfish.js/.
[11] https://8gwifi.org/CipherFunctions.jsp.
[12] https://8gwifi.org/rsafunctions.jsp.
[13] https://www.devglan.com/online-tools/aes-encryption-decryption.

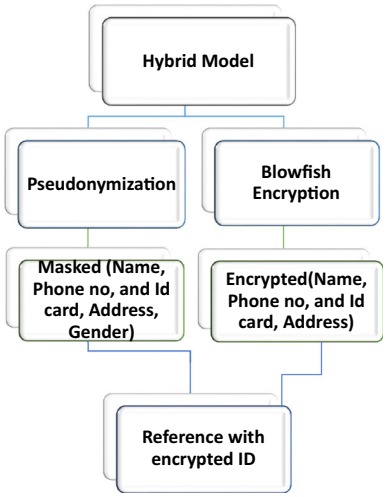


Fig. 3 The Hybrid model used to achieve data privacy

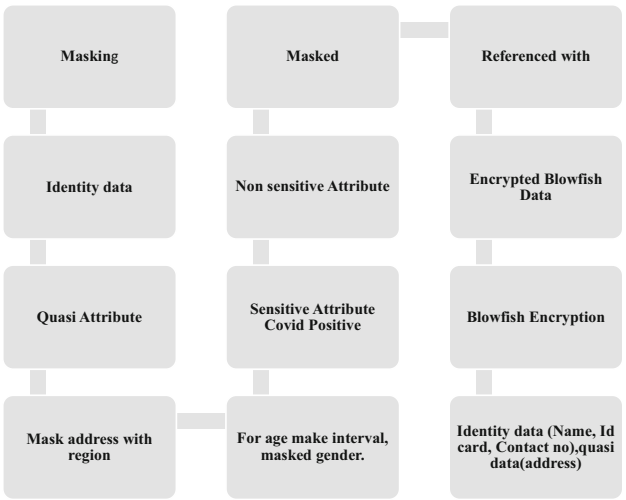


Fig. 4 Depicted how attributes are linked with each other to achieve data privacy

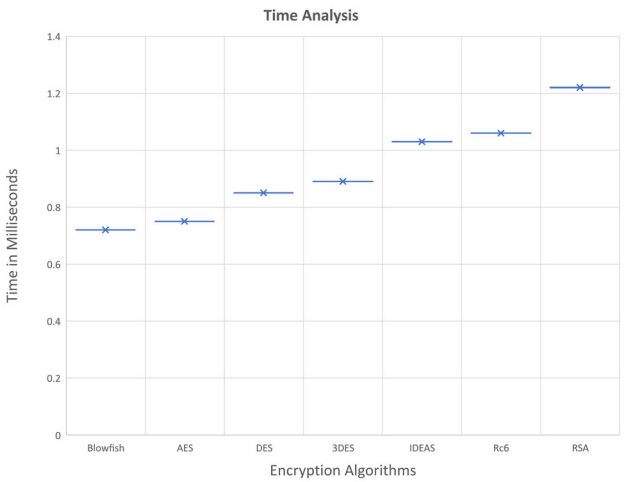


Fig. 5 The execution time of all encryption algorithm

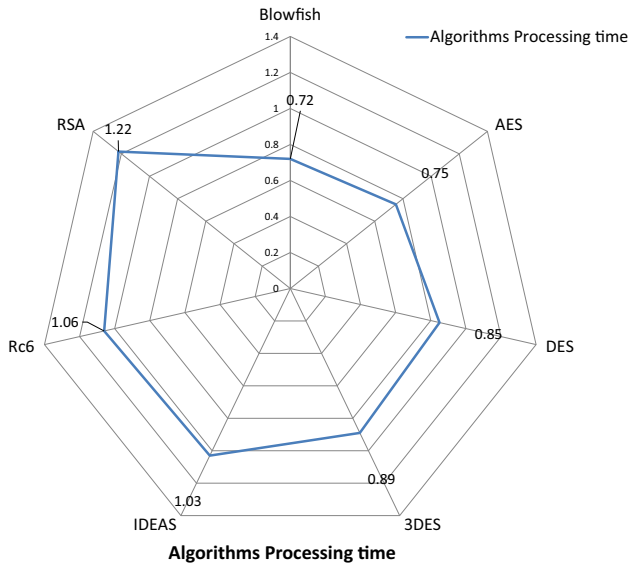


Fig. 6 The execution time for AES encryption and decryption

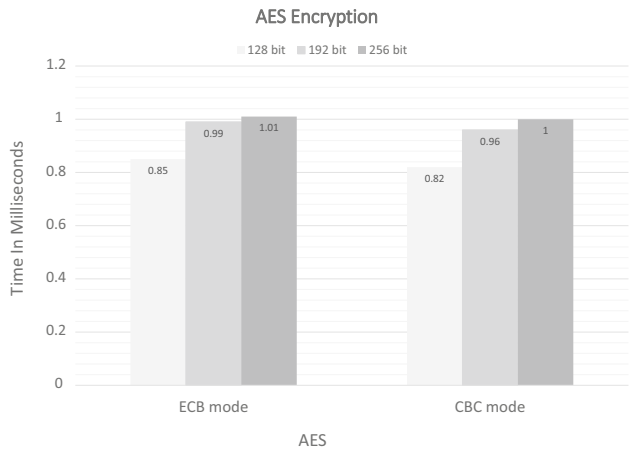


Fig. 7 The DES and 3DES execution time with different modes

We took the result as an average time execution of all these Encryption algorithms. The result shows execution times for different modes. Figure 6 shows AES encryption execution time with the mode in milliseconds such as ECB and CBC, and block sizes of 128,192, and 256. Figure 7 shows the encryption execution time for DES and 3DES regarding different modes and where results depended upon encryption modes; ECB, CBC, CFB, OFB, and NOFB.

Figure 8 shows encryption and decryption execution time for RC6 and IDEA. It shows for both algorithms, time taken for encryption and decryption.

Figure 9 shows Blowfish encryption execution time with modes such as ECB, CBC, PCBC, CFB, OFB, and CTR modes. As results show, it takes less time.

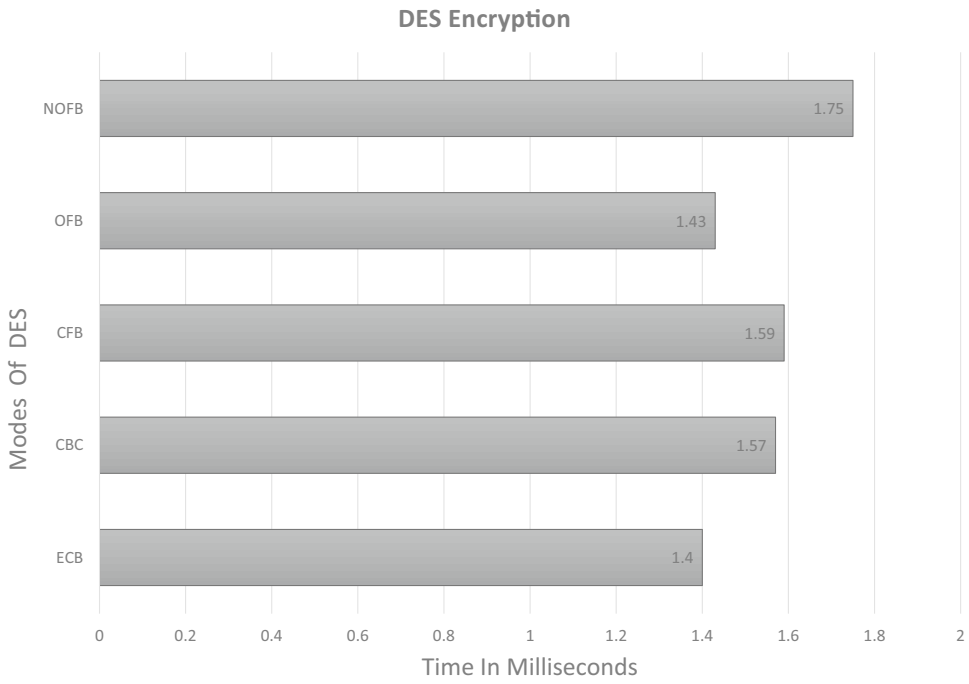


Fig. 8 The RC6 and IDEA execution time of encryption and decryption

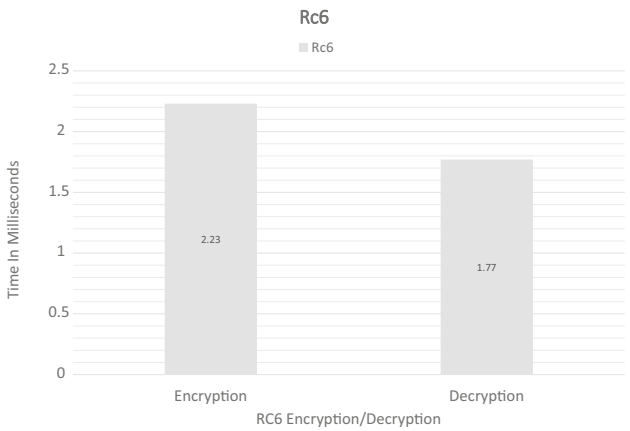


Fig. 9 The Blowfish execution time for encryption and decryption with different modes

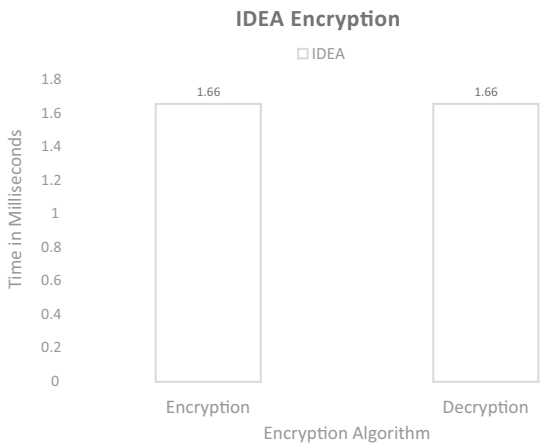


Fig. 10 RSA encryption and decryption execution time

RSA execution time shows in Fig. 10. It shows a different variation of execution for RSA: RSA, RSA ECB, RSA with padding, RSA with SHA padding and RSA ECB SHA padding, and RSA ECB padding with 256 and the block sizes of 512, 1048,2048, and 4096.

IC is the index of coincidence and technique of cryptoanalysis. In this paper, we measured IC for all algorithms for secure cipher. The Eq. 1 has been used to measure IC for all encryption algorithms.

In Eq. 1 the C is an index of coincidence, i_m denoted as repetition of a letter, the T denoted as the total number.

$$C = \sum_{m=a}^{m=z} \frac{i_m(i_m - 1)}{T(T-1)} \tag{1}$$

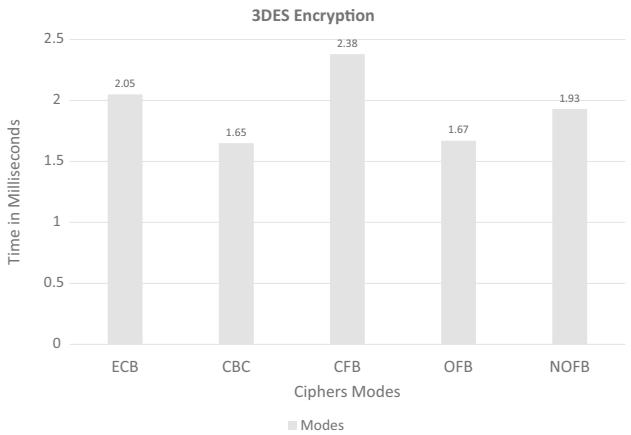


Fig. 11 The IC of the IDEA encryption algorithms

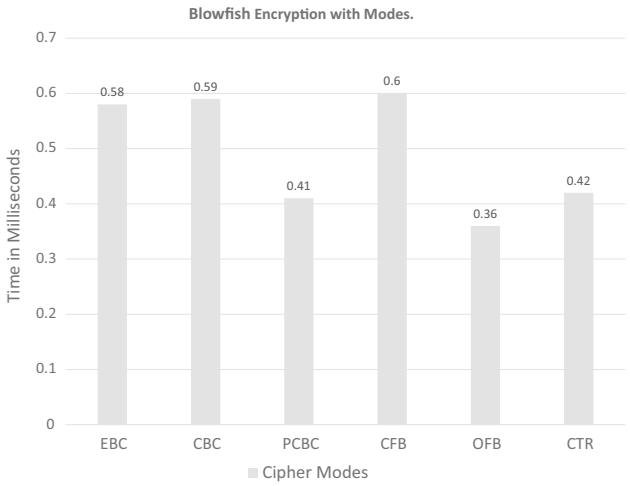


Fig. 12 IC of the RSA encryption algorithms

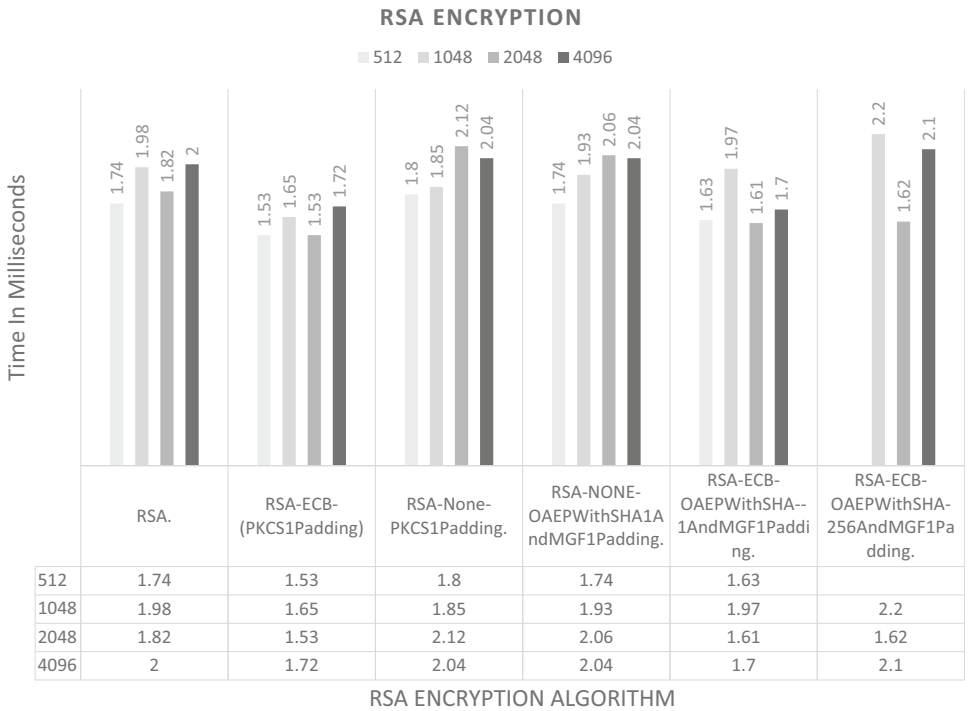
IC is measured using Eq. 1 for all algorithms discussed in this paper.

If the value is closer to 0.7, its means it nears plain text, and if it's near to 0.3850 value, its means it nears encrypted text, and if it's near to 0.3850 value, it is more secure. We checked it for all those algorithms simulated above, and the result is shown in a graph. The Figs. 11, 12, and 13 shows the results of measured IC's for different algorithms such as AES, DES, 3DES, Blowfish, IDEAS, RC6, and RSA. The result showed we achieved data privacy for Covid-19 patients using our proposed hybrid model Blowfish encryption and Pseudonymization.

5 Discussion

In this paper, we analyzed some research papers related to these technologies, authentication, encryption, auditing, and monitoring, to explore the weaknesses and strengths. Our proposed hybrid model of Blowfish and data masking is used to achieve data privacy for identity, quasi attributes, and sensitive attributes. It associated the encrypted data, masked and sensitive data, to achieve data privacy. We have performed some experiments for this hybrid model using python script and Java language. Also, all Encryption algorithms were run online and on eclipse. The result showed that Blowfish is an efficient algorithm for achieving data privacy of Covid-19 patients. Our proposed model used Blowfish encryption because it is best against known attacks. It takes less time and uses minimum memory consumption compared to other algorithms. We used the

Fig. 13 IC of different encryption algorithms



	RSA.	RSA-ECB-(PKCS1Padding)	RSA-None-PKCS1Padding.	RSA-NONE-OAEPWithSHA1AndMGF1Padding.	RSA-ECB-OAEPWithSHA--1AndMGF1Padding.	RSA-ECB-OAEPWithSHA-256AndMGF1Padding.
512	1.74	1.53	1.8	1.74	1.63	
1048	1.98	1.65	1.85	1.93	1.97	2.2
2048	1.82	1.53	2.12	2.06	1.61	1.62
4096	2	1.72	2.04	2.04	1.7	2.1

cryptoanalysis technique to measure the IC value and found the attack surface for all algorithms. The repeated alphabets show a low IC value, non-repeated alphabets show a high IC value. The result and experiment section show that's Blowfish IC value is high. The Blowfish is the best algorithm for encryption. This paper achieved data privacy for the Covid-19 patients using a Hybrid model of Blowfish Encryption and data masking technique.

6 Conclusion

The data privacy of Covid-19 patients has been proposed by this paper using a hybrid model. The researchers are using Covid-19 patient data, along with that the adversary can also access that data for malicious purposes, which may cause a privacy breach. This paper used a hybrid encryption and data masking approach to secure the COVID-19 patients' data. We proposed a hybrid algorithm approach of Blowfish and AES for future work to give a more secure framework for achieving data privacy. They can also achieve privacy for the 1:M data set of Covid-19 Patients. The topic of data privacy of Covid-19 patients is innovative for exploration for the researchers.

Author contributions All authors have contributed equally.

Data availibility Data Used for this work has been mentioned where needed in the paper.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

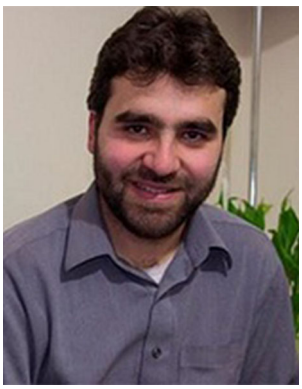
Ethical approval Ethical statement has been followed and fully endorsed as attached

References

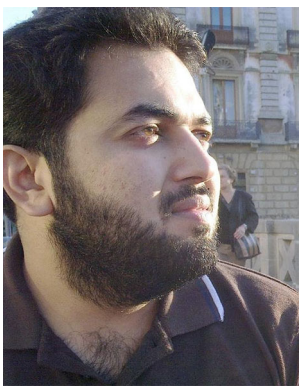
1. Govindan, K., Mina, H., Alavi, B.: A decision support system for demand management in healthcare supply chains considering the epidemic outbreaks: a case study of coronavirus disease 2019 (COVID-19). Transp. Res. E **138**, 101967 (2020)
2. Afzal, M., Hussain, M., Lee, S., Khattak, H.A.: Redesign of clinical decision systems to support precision medicine. In: TENCON 2018–2018 IEEE region 10 conference, pp. 2259–2263. IEEE (2018)
3. Khattak, H.A., Imran, M., Abbas, A., Khan, S.U.: Maintaining fog trust through continuous assessment, in world congress on services, pp. 129–137. Springer, Cham (2019)
4. Ahmad, I., Shah, M.A., Khattak, H.A., Ameer, Z., Khan, M., Han, K.: FIViz: forensics investigation through visualization for malware in internet of things. Sustainability **12**(18), 7262 (2020)
5. Gheisari, M., Najafabadi, H.E., Alzubi, J.A., Gao, J., Wang, G., Abbasi, A.A., Castiglione, A.: OBPP: an ontology-based framework for privacy-preserving in IoT-based smart city. Future Gener. Comput. Syst. **123**, 1 (2021)
6. Lenert, L., McSwain, B.Y.: Balancing health privacy, health information exchange, and research in the context of the COVID-19 pandemic. J. Am. Med. Inform. Assoc. **27**(6), 963 (2020)
7. Zwitter, A., Gstrein, O.J.: Big data, Big data, privacy and COVID-19–learning from humanitarian expertise in data protection. J. Int. Hum. Action **5**(1), 4 (2020). https://doi.org/10.1186/s41018-020-00072-6
8. Ekong, I., Chukwu, E., Chukwu, M.: COVID-19 mobile positioning data contact tracing and patient privacy regulations: exploratory search of global response strategies and the use of digital tools in Nigeria. JMIR mHealth uHealth **8**(4), e19139 (2020)
9. Kim, W., Lee, H., Chung, Y.D.: Safe contact tracing for COVID-19: a method without privacy breach using functional encryption techniques based-on spatio-temporal trajectory data. PLoS ONE **15**(12), e0242758 (2020)
10. Asghar, A., Abbas, A., Khattak, H.A., Khan, S.U.: Fog based architecture and load balancing methodology for health monitoring systems. IEEE Access **9**, 96189 (2021)
11. Sun, H.M., Chen, Y.H., Lin, Y.H.: oPass: a user authentication protocol resistant to password stealing and password reuse attacks. IEEE Trans. Inform. Forens. Secur. **7**(2), 651 (2011)
12. Ratna, A.A.P., Purnamasari, P.D., Shaugi, A., Salman, M.: Analysis and comparison of MD5 and SHA-1 algorithm implementation in Simple-O authentication based security system. In: 2013 international conference on QiR, pp. 99–104, IEEE, (2013)
13. Rama, M., Raja, S.S.: Web based security analysis of OPASS authentication schemes using mobile application. In: 2013 international conference on emerging trends in VLSI, embedded system, nano electronics and telecommunication system (ICEVENT), pp. 1–3, IEEE, (2013)
14. Sirohi, P., Agarwal, A., Tyagi, S.: A comprehensive study on security attacks on SSL/TLS protocol. In: 2016 2nd international conference on next generation computing technologies (NGCT), pp. 893–898, IEEE, (2016)
15. Kogetsu, A., Ogishima, S., Kato, K.: Authentication of patients and participants in health information exchange and consent for medical research: a key step for privacy protection, respect for autonomy, and trustworthiness. Front. Genet. **9**, 167 (2018)
16. Kansuwan, T., Chomsiri, T.: Authentication model using the bundled CAPTCHA OTP instead of traditional password. In: 2019 joint international conference on digital arts, media and technology with ECTI northern section conference on electrical, electronics, computer and telecommunications engineering (ECTI DAMT-NCON), pp. 5–8, IEEE, (2019)
17. Rahardja, U., Bist, A.S., Hardini, M., Aini, Q., Harahap, E.P.: Authentication of Covid-19 patient certification with blockchain protocol. Int. J. Adv. Sci Technol. **29**(8s), 4015 (2020)
18. Rauf, H.T., Malik, S., Shoaib, U., Irfan, M.N., Lali, M.I.: Adaptive inertia weight bat algorithm with Sugeno-function fuzzy search. Appl. Soft Comput. **90**, 106159 (2020)
19. Gao, J., Wang, H., Shen, H.: Smartly handling renewable energy instability in supporting a cloud datacenter. In: 2020 IEEE international parallel and distributed processing symposium (IPDPS), pp. 769–778, IEEE, (2020)
20. Alanazi, H., Zaidan, B.B., Zaidan, A.A., Jalab, H.A., Shabbir, M., Al-Nabhani, Y., et al.: New comparative study between DES, 3DES and AES within nine factors, arXiv preprint arXiv:1003.4085 (2010)

21. Gao, J., Wang, H., Shen, H.: Task failure prediction in cloud data centers using deep learning. IEEE Trans. Serv. Comput. (2020). https://doi.org/10.1109/TSC.2020.2993728
22. Gao, J., Wang, H., Shen, H.: Machine learning based workload prediction in cloud computing. In: 2020 29th international conference on computer communications and networks (ICCCN), pp. 1–9, IEEE, (2020)
23. Aggarwal, K., Saini, J.K., Verma, H.K.: Performance evaluation of RC6, blowfish, DES, IDEA, CAST-128 block ciphers. Int. J. Comput. Appl. **68**(25), 10–16 (2013)
24. Singh, P., Kumar, S.: Study and analysis of cryptography algorithms: RSA, AES, DES, T-DES, blowfish. Int. J. Eng. Technol. **7**(1.5), 221 (2017)
25. Wahid, M.N.A., Ali, A., Esparham, B., Marwan, M.: A comparison of cryptographic algorithms: DES, 3DES, AES, RSA and blowfish for guessing attacks prevention. J. Comput. Sci. Appl. Inform. Technol. **3**(2), 1 (2018)
26. Patel, K.: Performance analysis of AES, DES and blowfish cryptographic algorithms on small and large data files. Int. J. Inform. Technol. **11**(4), 813 (2019)
27. Commey, D., Griffith, S., Dzisi, J.: Performance comparison of 3DES AES, blowfish and RSA for dataset classification and encryption in cloud data storage. Int. J. Comput. Appl. **177**(40), 17–22 (2020). https://doi.org/10.5120/ijca2020919897
28. Lucca, A.V., Luchtenberg, R., de Paula Conceicao, L.G., Silva, L.A., Ovejero, R.G., Navarro-Cáceres, M., Leithardt, V.R.Q.: System for control and management of data privacy of patients with COVID-19. Europe PMC **20**(1), 1–19 (2020)
29. Tahir, M., Sardaraz, M., Mehmood, Z., Muhammad, S.: CryptoGA: a cryptosystem based on genetic algorithm for cloud data security. Clust. Comput. **24**, 739–752 (2020)
30. Hussaini, S.: Cyber security in cloud using blowfish encryption. Int. J. Inform. Technol. (IJIT) **6**(5), 13–19 (2020)
31. Parkar, A.P., Gedam, M.N., Ansari, N., Therese, S.: Performance level evaluation of cryptographic algorithms. In: Ann, R. (ed.) Intelligent computing and networking, pp. 157–167. Springer, Singapore (2021)
32. Chao, Y., Bingyao, C., Jiaying, D., Wei, G.: The research and implementation of UTM. In: IET international communication conference on wireless mobile and computing (CCWMC 2009), pp. 389–392, IET, (2009)
33. Wang, L.: Big data in intrusion detection systems and intrusion prevention systems. J. Comput. Netw. **4**(1), 48 (2017)
34. Abouelmehdi, K., Beni-Hessane, A., Khaloufi, H.: Big healthcare data: preserving security and privacy. J. Big Data **5**(1), 1 (2018)
35. Awan, K.A., Din, I.U., Almogren, A., Khattak, H.A., Rodrigues, J.J.: EdgeTrust-a lightweight data-centric trust management approach for green internet of edge things. Wirel. Pers. Commun. (2021). https://doi.org/10.21203/rs.3.rs-453986/v1
36. Neupane, K., Haddad, R., Chen, L.: Next generation firewall for network security: a survey. In: SoutheastCon 2018, pp. 1–6, IEEE, (2018)
37. Liang, D.: Security information and event management, security information and event management. US Patent 10,616,258. (2020)
38. Thapa, C., Camtepe, S.: Precision health data: requirements, challenges and existing techniques for data security and privacy. Comput. Biol. Med. **129**, 104130 (2020)
39. Bernier, A., Knoppers, B.M.: Pandemics, privacy, and public health research. Can. J. Public Health **111**(4), 454 (2020)
40. Prince, P.B., Lovesum, S.J.: Privacy enforced access control model for secured data handling in cloud-based pervasive health care system. SN Comput. Sci. **1**(5), 1 (2020)
41. AlMarzooqi, F.M., Moonesar, I.A., AlQutob, R.: Healthcare professional and user perceptions of eHealth data and record privacy in Dubai. Information **11**(9), 415 (2020)

Rabeeha Fazal is a Masters of Computer Science Student at COMSATS University Islamabad, Islamabad Campus. Her research interests are Data Privacy in Modern Applications.



Munam Ali Shah received B.Sc and M.Sc degrees, both in Computer Science from University of Peshawar, Pakistan, in 2001 and 2003 respectively. He completed his MS degree in Security Technologies and Applications from University of Surrey, UK, in 2010, and has passed his Ph.D. from University of Bedfordshire, UK in 2013. Since July 2004, he has been an Assistant Professor, Department of Computer Science, COMSATS University Islamabad, Pakistan. His research interests include internet of things (IoT), information security in modern networks and energy harvesting. Dr. Shah received the Best Paper Award of the International Conference on Automation and Computing in 2012. Dr. Shah is leading the research group Performance Evaluation and Enhancements of Computing Systems (PEECS) within his department. Currently, Dr. Shah is supervising 3 Ph.D. students and 9 masters students. 31 masters students already finished their dissertation under his supervision. Dr. Shah is the author of more than 200 research articles published in national and international conferences and journals. Dr. Munam Ali Shah is a HEC approved supervisor.



Hasan Ali Khattak currently working as Associate Professor at National University of Sciences and Technology (NUST), Islamambad graduated as full time Ph.D. student from Polytechnic University of Bari, Italy where I was pursuing Ph.D. in Electrical and Computer Engineering from 2012 - 2015. He, recently, has started working on Autonomous Vehicles and Cyber Physical Systems for enabling the envisioned Smart Cities through future internet architectures through mainly application of data sciences and artificial intelligence in fog computing and vehicular networks. Dr. Hasan can be contacted at hasan.alikhattak@seecs.edu.pk

Hafiz Tayyab Rauf received the bachelors and masters degrees in computer science from the University of Gujrat, Gujrat, Pakistan. He is currently working as a Research Assistant independently with several research institutes. He is the author of more than 20 research articles published in reputed journals. His research interests include evolutionary computing, swarm intelligence, neural networks, image processing, computer vision, and machine learning. He is a Reviewer of various high impact factor journals.

Fadi Al-Turjman received his Ph.D. in computer science from Queens University, Canada, in 2011. He is a full professor and a research center director at Near East University, Nicosia, Cyprus. Prof. Al-Turjman is a leading authority in the areas of smart/intelligent IoT systems, wireless, and mobile networks architectures, protocols, deployments, and performance evaluation in Artificial Intelligence of Things (AIoT). His publication history spans over 350 SCI/E publications, in addition to numerous keynotes and plenary talks at flagship venues. He has authored and edited more than 40 books about cognition, security, and wireless sensor networks deployments in smart IoT environments, which have been published by well-reputed publishers such as Taylor and Francis, Elsevier, IET, and Springer. He has received several recognitions and best papers awards at top international conferences. He also received the prestigious Best Research Paper Award from Elsevier Computer Communications Journal for the period 2015-2018, in addition to the Top Researcher Award for 2018 at Antalya Bilim University, Turkey. Prof. Al-Turjman has led a number of international symposia and workshops in flagship communication society conferences. Currently, he serves as book series editor and the lead guest/associate editor for several top tier journals, including the IEEE Communications Surveys and Tutorials (IF 23.9) and the Elsevier Sustainable Cities and Society (IF 5.7), in addition to organizing international conferences and symposiums on the most up to date research topics in AI and IoT.