The Protection of Data Privacy of College Students

Zhiheng Tao

107 S Indiana Ave, Bloomington, IN 47405, US

Abstract. Privacy data belongs to big data, and it is personal information from our lives. The common problem is privacy leaking and cyber attacks. The current solution uses a complex algorithm to encrypt data, but it is expensive and inefficient. In the paper, we combine students' information at university to make four layers based on a reliable framework named FMM. The proposed framework FMM, according to different privacy levels, will choose other encryption methods to protect these data and increase efficiency to keep trade-offs.

Keywords: **Privacy, Big data analytics, Security, Encryption, Protection.**

1. Introduction

With the development of the Internet, data privacy protection is taken more seriously for people. Many people always store their personal information on the Internet, and it is now prime and common data globally. Privacy protection has many layers to filter people's information and distribute it in the correct layers for encryption. Through internet development, many places support privacy protection methods and combine them with their Internet of Things (loT) infrastructure, such as medical facilities, stores, transit, etc. In the loT scenario, users' information will be transmitted through the loT infrastructure to the database. These processes will have many dangerous factors to influence (cyberattacks). Since this personal information will become plaintext, and there are many encryption methods (symmetric, asymmetric, and hashing) to encrypt, it leads to different encryption methods with different speeds and security strengths.

Currently, people use three standard techniques to protect their private data: Anonymous [5], Encryption [5], and Access Control [5]. Indeed, each technique has its advantages and methods, such as anonymity, which is suitable for hiding users' names, encryption can use a different type of encryption to encrypt, and Access Control can check each access user's condition and identity at the first time. Even successful privacy protection methods still need to be built to meet users' requirements: applicable requirements to them. However, all the conditions must contain confidentiality, integrity, and availability. Data also needs to make more highly efficient encryption, to match what type of data is more suitable and which type of encryption to encrypt.

In contrast to other privacy protection models, our FMM advantages are flexible and innovative; FMM can, according to different layer features, use different encryption to encrypt each layer. To improve efficiency, we screen suitable encryption methods for each layer. In addition, our FMM can change students' requirements, and if students only want to filter their records information, they can only use our Record Layers from our FMM to filter their required information. In the paper, we will focus more on the privacy model and propose a new four-layer framework-the Face Mask Model (FMM)-at universities. To analyze how to flexibly and deeply protect students' privacy data at universities and solve how to store data from universities more securely. It is essential that the FMM reduces the risk of data privacy leaks and increases the encryption speed and flexibility to satisfy different types of students. There are three contributions to the paper.

1. We propose a new privacy framework (FMM) to help students protect their university information and provide different layers using different encryption algorithms to increase effectiveness and safety.

2. The privacy framework (FMM) contains four interaction layers: Sensitive, Bias, Record, and Base. To give us a big data scenario, we will fulfill the need to use different encryption methods to achieve high reliability and efficiency trade-offs according to that security level.

3. Flexible to match the student's requirements. This (FMM)model can satisfy different students' requirements by providing other protection solutions. FMM can further protect our data security level, thus reducing the risk of private data leaks.

2. RELATE WORK

2.1 Data Analytics

Nowadays, Zainuddin [2] proposes different potential problems regarding privacy in the Internet of Things

environment. It analyzes many intelligent environments, like homes, meters, medical centers, cities, and clouds. These innovative areas [2] work remotely. It will have potential problems like people can steal information on the transmission, collect other users' data to predict their private information, directly attack physical devices: make them invalid, and also use tools to collect unencrypted user information from intelligent homes devices. These attacks are hazardous to the Internet of Things; they can use these methods to collect users' big data quickly. Also, Alshboul's model method [7] combines big data security lifecycle and security attacks, which improves data protection; it also prevents data mining influence because it is a widespread attack method and causes significant damage to the database. Then, Mehmood analyzes many new model ideas [6] that can explore different areas or types of data at the same time. People don't have to use multiple models to protect or analyze this data; it still has some areas that need strengthening in different periods, like decentralized storage and data anonymization.

2.2 Data Privacy

Schaub [8] proposes using many different policies and rules, such as transparency, security, organizational safeguards, etc., to protect students' privacy. It also protects students' confidentiality and prevents unfair analysis of students. Reidenberg's new policies suggest [8] that people care more about students' privacy and data. They notice students' privacy rights. Educational data mining and learning analytics [8] will positively support the study process and negatively hurt users' privacy. Also, El Ouazzani can block people who want to attack or leak, even if they successfully attack this data, but they still don't know which data matches which owner: which means nothing. However, the weakness of this method is that hackers can still use keywords to match this privacy. Additionally, the innovation of this method is it can block people who want to attack or leak data. Even if they successfully attack this data, but they still don't know which data matches which owner: which means nothing. Also, the weakness of this method is that hackers can still use keywords to match this privacy. And people can use this method to attack back, like the utilization of t-closeness [9] against some similar attack styles in treating categorical sensitive attributes.

2.3 Internet of Things

Internet of Things (loT) is when the user uses physical equipment connected to the Internet and has services from there, it often happens that people change the equipment without the users' authorization: get data from that method and steal data in the data transmission part. Recently, Kabalci [3] proposed how the energy Internet connects with the Internet of Things and how they combine. It [3] can provide two advantages in the energy Internet area, the first advantage is it can provide specific communication and complex network structures for different communication scenarios, and the other is that it is used to reduce power and cost to make the devices more efficient. So we can see it is an excellent idea to combine them and reduce the Internet's original weakness, making it more flexible. In early research, Medaglia [4] proposed how current technological and technical trends influence our security and privacy and analyzed the future impact. It [4] explores the IoT infrastructure between users' connections and remotely managed relationships will increase security and privacy problems because it needs data transmission: between infrastructure and users. Whereas IoT is not very safe in their infrastructure because people overlook that and invest time to upgrade, loT security design [4] tries to open pervasive and interoperable infrastructure to follow this trend.

2.4 Application Scenarios

Ziegeldorf [1] proposes a basic new model and law. It is based on the International Telecommunications Union and the IoT European Research Council to make users quickly and safely connect any devices and services on the Internet and needs a complete privacy law to support it. Primarily, they still think [1] that profiling is very dangerous to their model because people can, though, collect data to increase their attack privacy possibility. Sun [5] proposed that medical loT needs to strengthen security and privacy because medical loT will remotely watch and collect patient medical data: heart rate, etc. Medical loT [5] usually has three-layer: perception, network, and application; each layer will store important privacy data in there. It needs to promise its integrity, validity, and authenticity because this private data is only accessible to people who have authorization from the owner. The problem for the future in medical loT [5] is an insecure network, lightweight protocols for devices, and data sharing; these areas are challenging to manage and have low-security protection methods, and they need more focus on their devices and network security techniques on this area to prevent these factors from influencing it.

3. Methodology

Based on different encrypting features, different data styles need suitable encryption methods to protect them. There are four other encryption methods below.

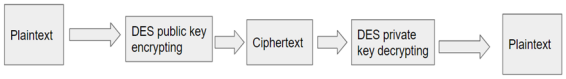

Figure 1. DES algorithm process

According to FIGURE 1 shows, the DES encryption method is a symmetrical encryption, it uses the same key to encrypt and decrypt. It has a simple encryption algorithm: short key, and manages well to extensive data. Contrast the AES method, the AES has overprotection to these types of data.

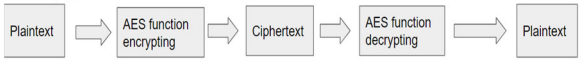

Figure 2. AES algorithm process

We describe the AES details in FIGURE 2, and the AES encryption is a symmetrical-based method. It has a

high-speed and a new algorithm for encryption. Contrast the DSA method, and the DSA is not a very high-speed encryption method because DSA has high-level verification by users.

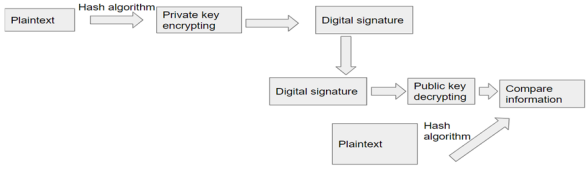

Figure 3. DSA algorithm process

The FIGURE 3 illustrates the DSA encryption method, and it is asymmetric, this method uses the signature to verify the key. It has high-level security protection, and needs a digital signature for decryption and encryption, to improve the data transmission process security. Contrast the RSA method, the RSA has low-speed decryption.

Figure 4. RSA algorithm process

As process content in the FIGURE 4, the RSA encryption method is asymmetric. The public key is that everyone can view it and the sender to encrypt, but the private key is only for the receiver to decrypt. Then, it will form different encrypt and decrypt keys to exchange. It has the highest security level protection, this method needs a public and private key exchange because the public and private decryption keys are different, it makes each key individual. Besides, the RSA key is longer than other encryption methods, so it will be difficult to break.

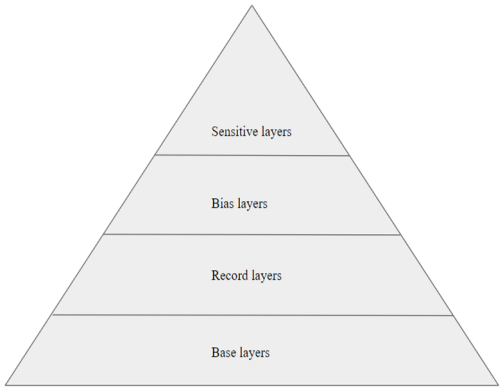

Figure 5. FMM Model Filter

Based on these four encryption algorithms, we propose a new model in FIGURE 5, to better protect and flexibly fill our private data. It has a sensitive layer, bias layer, record layer, and base layer. The base layer uses the DES method because this layer has large data and these data are not important, we only need standard protection. Second, the record layer uses the AES method because this layer also has large data, but these records data are important to users, we need safe and fast speed to protect them. Third, the bias layer uses the DSA encryption because this data will influence the user's living quality. We need signatures to verify by users. Fourth, the sensitive layers use the RSA method because this layer of data is very important to users' social credits, and some data is from the government, we need to use high-level encryption to protect them.

Furthermore, TABLE 1 illustrates each layer's details. For instance, the Base layers include full name and phone number information, the full name is Jack Duke Li and his phone number is 812-453-2433, these informations are searchable by the internet, we only need standard encryption to encrypt. The Record layers have an expense record and bus record. The expense record information is the student spends 10 dollars in the dining hall and he takes the route 9 bus back to home because he swiped his student card before entering the bus, most records information will be from that reader maschine. Then, the Bias layers include transcript and sex, if the student has 2.4 GPA and he is a male, these inforamtion will influence by jobs market: need agree by user permission. Last, the Sensitive layers cover SSN and IP address, the student's SSN is 154-75-3541 and his IP address is 11.18.238.174, these information are very important to this student's living, so we need the highest security level to protect them.

Table 1. Details of Each Layers

Layers	Function	Definition
	Full name	A person's legal name
	Phone number	Student's phone number and emergency contact
	Major	Major and minor information
	Enrollment date	Enrollment date information for the academic team
	School name	Currently university name
Base Layers	Student ID number	Student ID number infromation
	School email	Student's school emaill information
	Tuition	Student yearly tuition
	Class	Taken and in-progress classes
	Degree time	How long have you been stidying for your degree at univerisity
	Transfer	School transfer history information
	Spending record	Student spending on campus, like dining, vending machines, stores, etc.
Record Layers	Campus bus record	Campus bus record information
	Attendance record	Attendance record information from your classes
	Dorm record	Dorm card reader information
	Transcript	Transcript information
	Addresses	Local address and Home address
Bias Layers	Nationality	Nationality information
	Race	Race information
	Sex	Sex information
	High school privacy	High school privacy information
	Bank account	Bank account information
	IP address	Student's devices IP address information
Sensitive Layers	Medical history	Medical history information
	Criminal history	Criminal history information
	SSN	SSN information
	Gender	Gender information

3.1 DES Definition(DES Symmetric-key algorithm)

This layer is the basic information layer, there is no need for complex encrypting, but it needs to be moderate speed and efficient, and DES meets the moderate speed and large areas of data encryption requirements.

3.2 DES Example

Jake is a university student, and his legal name is Jake Duke Li, his phone number is 812-453-2433 and his email is jakeli@iu.edu, he is taking computer science, a major, at Indiana University Bloomington. His student ID is 20003598041 and his enrollment date was August 20th in 2018, he paid 56000 dollars for his second-year tuition. And he is taking 15 credits, in-person classes, for this semester and he already finished 30 credits last year. Also, he has no transfer history from another school, and he has

been studying for his degree for one year at his university: currently.

3.3 AES Definition (AES Symmetric-key algorithm)

The information at this layer basically comes from records, it needs some security protection. AES has certain security, high speed, and high efficiency.

3.4 AES Example

Jake is taking route 9 of the campus bus every day morning for his first class, and his instructor will record his attendance every time. After his classes are down, he will go to school dining for his lunch: spend 10 dollars on fast food, because he needs to go back to his class at 1 p.m. to prepare for his afternoon classes. When his 5 p.m. class is down, he will take the route 9 bus again to go back to his dorm, and swipe his student ID to enter his dorm: back to his room.

3.5 DSA Definition (DSA asymmetric-key algorithm)

This layer of information is actually searchable, such as transcript and address, and information leakage will cause psychological burden on users. Also, DSA can avoid these problems because each decryption needs user approval. It lets users know how secure their information is. The following is the flow chart of encryption and decryption.

3.6 DSA Example

Jake is a male international student, his nationality is Spain and he is an Asian. He currently: local address, live in 300 Pelo Varde Dr, Bloomington, IN 47401 and his home address is B. Diego de Velázquez, 1, 35449 Pozuelo de Alarcón, Madrid, Spain. He goes to the office to see his transcript: overall GPA and each semester credits condition, etc. Also, he is checking his high school privacy, like behaviors record and GPA, etc.

3.7 RSA Definition (RSA asymmetric-key algorithm)

This layer contains sensitive information, such as bank accounts and SSN. Information leakage can lead to users losing money and credit because others use their accounts and SSN for illegal transactions. However, RSA can avoid this situation because RSA has a high security level, and only needs the consent of both parties to decrypt and encrypt. Therefore, information security has effective protection.

3.8 RSA Example

Jason's gender is male and he has a heart attack history in his medical history, he needs to make an appointment to the hospital to check every year. Before going to the hospital, he wants to see how much money he has right now, so he goes to log in to his bank account: 5247-5874-3574-7164. But, his computer warned him, he is using VPN to hide his IP address: 11.18.238.174. He has to immediately close his VPN because he has a criminal history: he used VPN to hide his location and DDOS attacked other people's computers. After this process, he fills the medical form and gives the hospital his SSN number: 154-75-3541.

4. Conclusion

FMM resolves students' privacy security potential problems from their school. Students spend lots of time on campus to stay, and the school grasps students' mass personal information. A potential problem is a data leak. Our innovation in our proposal utilizes different features of encryption methods to combine a model, and this model will, according to the encryption data type, automatically adjust the security level. Additionally, this model has four independent layers to filter the different types of data because students leave many different types of data records at school. These four independent layers can work together or independently to satisfy user preferences. Compared to other models, our model has DES, AES, DSA, and RSA advantages to protect our different types of data: flexibility. This model also has high efficiency in dealing with these data and cooperates with students' data features. In the future, FMM will develop in the school area to deeply improve the security speed and accuracy. It also could merge in the education area or sociality because these areas' data is diverse. Significantly, loT will apply in our society in the future, and it will collect multiple data to protect from the database. This model could add an artificial intelligence element to more automatically give to different types of data or area solutions.

References

1. Ziegeldorf, J. H., Morchon, O. G., & Wehrle, K. (2014). Privacy in the Internet of Things: threats and challenges. Security and Communication Networks, 7(12), 2728-2742.
2. Zainuddin, N., Daud, M., Ahmad, S., Maslizan, M., & Abdullah, S. A. L. (2021, January). A study on privacy issues in internet of things (IoT). In 2021 IEEE 5th International Conference on Cryptography, Security and Privacy (CSP) (pp. 96-100). IEEE.
3. Kabalci, Y., Kabalci, E., Padmanaban, S., Holm-Nielsen, J. B., & Blaabjerg, F. (2019). Internet of things applications as energy internet in smart grids and smart environments. Electronics, 8(9), 972.
4. Medaglia, C. M., & Serbanati, A. (2010). An overview of privacy and security issues in the internet of things. The internet of things, 389-395.
5. Sun, W., Cai, Z., Li, Y., Liu, F., Fang, S., & Wang, G. (2018). Security and privacy in the medical internet of things: a review. Security and Communication Networks, 2018.
6. Mehmood, A., Natgunanathan, I., Xiang, Y., Hua, G., & Guo, S. (2016). Protection of big data privacy. IEEE access, 4, 1821-1834.

7. Alshboul, Y., Nepali, R. K., & Wang, Y. (2015, August). Big Data LifeCycle: Threats and Security Model. In AMCIS.
8. Reidenberg, J. R., & Schaub, F. (2018). Achieving big data privacy in education. Theory and Research in Education, 16(3), 263-279.
9. El Ouazzani, Z., & El Bakkali, H. (2020). A classification of non-cryptographic anonymization techniques ensuring privacy in big data. International Journal of Communication Networks and Information Security, 12(1), 142-152.