International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 / https://doi.org/10.14445/22315381/IJETT-V70I6P213

Volume 70 Issue 6, 106-113, June 2022

Original Article

Merkle Tree-based Access Structure for Sensitive Attributes in Patient-Centric Data

B. Ravinder Reddy[1], T. Adilakshmi[2]

[1]*Department of CSE, UCE (A), OU.*
[1]*Department of CSE, Anurag University, Hyderabad-500088.*
[2]*Department of CSE, Vasavi College of Engineering (A), Hyderabad-500089.*

ravinderreddycse@cvsr.ac.in

Received: 01 April 2022 Revised: 24 May 2022 Accepted: 03 June 2022 Published: 27 June 2022

Abstract *- Wearable health care [1] is a prominent technology that allows people to get medical services. Medical data obtained by wearables necessitates immediate and accurate information sharing from any location for improved healthcare decisions. However, during sharing, the security and confidentiality of such patient information will become a key problem. The security of patients' data is critical in medical settings when unwanted access to information is unavoidable. To close the security gap, this study proposes the Merkle Tree-based Access Structure Construction for the CP-ABE model* ***[2]****, a unique distributed authority and policy-hiding capability approach that provides secure access to patient-sensitive data characteristics. We improve the CPABE access structure by building two access trees, one for sensitive patient information and the other for non-sensitive qualities, to provide fine-grained access control, data privacy, and integrity [1]. We compared the proposed framework to other CP-ABE-based systems that were already in use. It demonstrates that the suggested paradigm provides improved security for clinical services.*

Keywords *- cpabe, Hash, Merkle tree, pcd, Privacy.*

1. Introduction

In general, anybody may access data and services through a communication network, and once the data is uploaded to a server, the owners lose control. This data may contain sensitive information, so it's critical to specify [3] a specific access control strategy for such data in a distributed system. Health wearable gadgets, for example, have become key components of remote patient monitoring systems in recent years. These devices may automatically gather and process personal health data elements such as a user's identity, location, PID, symptoms, diagnosis, procedures, and results in real-time. Health data is classified as "special and sensitive category data" under the GDPR 2016/679 [4]. Because of this, wearable health care technologies can be dangerous for people's privacy and access control.

1.1. Sensitive Patient Data

It may be combined with information about the patient, such as their name, email address, phone number, and location, to ensure that it reaches the correct account. Patient vital information can be captured while being transmitted from local storage to a system application using a smartphone in the health care business. Patient data may be split into two groups in this situation: sensitive and non-sensitive. Sensitive data is information that might, directly or indirectly, reveal patients' identities. Phone books, the Internet, and corporate directories are good places to look for non-sensitive information.

We propose a novel technique for hiding patient-sensitive information features in this study, which involves hashing the attributes at each level to create a Merkle tree. The destination can confirm a Merkle tree's root hash value without revealing the tree's nodes or attributes. When used with the CPABE public access tree, this novel solution can increase access control and system efficiency.

2. Related Work

Sahai et al. [5] proposed ABE, which allows for fine-grained ciphertext access control. ABE [6] defines CP-ABE as a public-key cryptography system for enhanced access control while sharing sensitive information in one-to-many interactions. In an open setting, it can provide data security and privacy. The policy is contained in the ciphertext [7,8]. Each user's private key is a collection of characteristics that can only be decrypted successfully if the attributes or user's key satisfy the policy key. One key flaw in this technique is that the policy is transmitted with a ciphertext for the decryptor to discover which characteristics comply, exposing the policy's privacy. As a result, it is unsuitable for applications involving sensitive data, such as those in the medical, industrial, and financial areas. On the other hand,

the decryption process becomes impossible if the policy characteristics are unknown.

Nishide et al. [9] developed a partially concealed access strategy in CP-ABE to avoid the mentioned issue. The attribute is specified in two parts in this scheme: the attribute name and the attribute values. This system preserves the policy's privacy to some level but also has downsides. In some circumstances, the attribute name, for example, also contains sensitive information. Because the user must guess which attribute value is contained in the ciphertext if the decryptor has numerous values for each attribute, the decryption time may be super-polynomial [10]. As a result, attribute names must likewise be kept hidden. Another issue with present ABE [11] systems is that decryption computation costs a lot of money. A few cryptographic processes with a significant computing burden might be outsourced to third-party services to decrease the decryptor's computational cost.

2.1. Wearable Technology and Healthcare

Many benefits may be gained by using wearable health care devices for remote patient health monitoring, including considerable cost savings per patient, reduced staff workload, improved diagnosis, disease prevention, faster clinical decision making, and an overall improvement in patient care. Patients, physicians, hospitals, nurses, manufacturers, security researchers, regulatory agencies, and insurance companies are among the stakeholders [12] who play a key role in enhancing access control and safeguarding the privacy of such sensitive health care information.

2.2. Securing Patient-Centric Data

Many wearable devices retain data in local storage without encryption when privacy and security [13] are addressed in wearable technology. As a result, there's a good chance that secret and sensitive information concerning patient health data may be lost. On the other hand, a comprehensive solution to the majority of the issues remains elusive. This study seeks to produce actual patient-centric data by addressing the primary issues about wearable devices' generated health care data access and privacy. CPABE has substantially influenced PCD's ability to provide comprehensive access control.

3. Motivation

Patients might encrypt data in various methods to improve access control in health care [14], such that anybody who obtains the encrypted text can only comprehend the publicly visible qualities of the patient while the sensitive attributes stay hidden. The solution(s) is to encrypt the complete access policy with attributes using CP-ABE or only the policy section that needs to be concealed using CP-ABE. However [9], because approved end-users can only acquire the ciphertext, the system cannot determine if the end-user has adequate authorization to access it. The second drawback of this strategy is that decryption overhead cannot be outsourced. Finally, CPABE encryption alone will not safeguard the system if the policies are partially or completely hidden. It is critical to make data patient-centric in a healthcare situation since it may be shared across various domains. It may be accomplished by masking the attribute of access policy by hashing at multiple layers in CPABE, which can provide fine-grained access control.

3.1. Contributions

As a result of the preceding finding, this study significantly contributes to developing a Merkle Tree-based access policy for sensitive characteristics in patient-centric data. The health attributes associated with the data owner (patient) are detected and categorized in our system depending on kind and sensitivity. The non-sensitive, public features have a direct role in policy development. The root hash of sensitive characteristics, on the other hand, is computed and added to the public access policy. Only permitted users can access the ciphertext, and the sensitive characteristics' privacy is protected.

4. Preliminaries

This section discusses the symbols used in Appendix A for the proposed model: access structure, access tree, hash function, sensitive attributes, Merkle tree T_S, and Merkle Proof

4.1. Access Tree and Access structure

Definition 1: Access structure: An access structure [10,15] is a monotone collection. Assume that {a1,a2,…,an} are a set of attributes..

a1, a2,..., and For a given collection [16] A ⊆ 2 {a1,a2,…,an} is monotone if ∀ B, C : if B ∈ A and B ⊆ C then C ∈ A .

A of non-empty subsets of {a1,a2,…,an}, [10,12] that is A ⊆ 2 {a1,a2,…,an} \{∅}.

The sets in A and not in A are authorized and unauthorized, respectively.

Definition 2: Access tree: T_S represents a tree in which every non-leaf node indicates a threshold gate, defining its children and a threshold value.

In T_S, every non-leaf node is a threshold gate, which displays its children and the threshold value.

Table 1. Access Tree Algorithm

Step 1:	If (num_s = number_of_children(s))
Step 2:	ks =t, where t is the threshold value
Step 3:	then $0 \leq ks \leq num_s$.
Step 4:	while(t == 1)
Step 5:	"the threshold gate is an OR gate" [16]and
Step 6:	while(t == num_s),
Step 7:	The threshold gate is an AND gate.
Step 8:	Each leaf node [8] T_S is described by an attribute a_s && t == 1.

The following table defines four functions:

Table 2. Functions defined in Sensitive attribute access tree T_S

S.No.	Function (s)	Description
1	*parent(s)*	Indicates [17] the parent of the node s in the sensitive attribute access tree T_S
2	T_S	Defines [18,19] an ordering between the child of every node, that is, numbered between 1 to num
3	*index(s)*	Returns a number associated with the node s, where the index values are arbitrarily unique to nodes for a given key.
4	*leaf_node (s)*	Corresponds to an attribute[10] a_s

4.2. Merkle Tree and Blockchain

The Merkle Tree sensitive attribute, T_S, was created to preserve the access structure's privacy. A Merkle Tree serves as the foundation for the creation of a blockchain.

Definition 3 (Hash Function): It's a deterministic function that can take any vast quantity of data and turn it into a fixed-size number. To do so, the hash function must adhere to several rules.

a. A change in the input must imply a full change in the output.
b. The output range must be consistent.
c. The same output is produced through a regulated and limited amount of collisions among the inputs.
d. In reality, the non-invertible feature means it is impossible to derive the hash output from the input.

SHA256 is a well-known hash function that turns a message (string) into a 256-bit integer.

Table 3. Illustration of Hash values for multiple input messages

S.No.	Input Message (String)	SHA256 Hash Value
1	DIAGNOSIS	784a50b9a834760ce88a45cbeb07ae00cd4be06fed7d0da256e610e688b09120
2	Diagnosis	d7c674a1348768db91ec90a10d88dae9ff9e93e2debacf06a9b34ca6cd1101c3
3	Diagnosis	d44eb2bfbbd35a7aacf19f98db4454e809e934de14d26026d695235a4abd81a6

Definition 4 (Sensitive attributes Merkle tree T_S): It's a method of storing enormous volumes of data while allowing the user to verify that the data hasn't been changed easily. The T_S has the same structure as conventional access trees, except that it does not expose [10] any information about the relationship between sensitive leaf nodes and attributes. While sensitive attributes are still hashed, it is relatively simple to evaluate whether a combination of leaf nodes and hash root value fits the access structure. Consider the hash function H and attributes set to build the tree [11]:

A = {a1, a2, . . . , an}

The hash values of elements in set A are the leaves of the Merkle tree.

Where A: H(a1), H(a2),…, H(an), And the tree is built recursively.

Definition 5 (Merkle root): RH(A) is the root for A, and it equals the root hash of the associated Merkle tree.

Definition 6 (Merkle Proof):

Consider the access tree [10] T_S with root r.

Denote by T_{SS} the subtree of T_S rooted at the node ss.

Hence, $T_S == T_r$.

If $A(\gamma) \in T_{SS}$,

Denote [17] it as $T_{SS}(\gamma) = 1$.

We compute $T_{SS}(\gamma)$ recursively as follows.

if(ss ≠leaf node)

Evaluate $T_{SS}'(\gamma) = 1$ ∀ children ss′ of node ss.

$T_{SS}(\gamma) \leftarrow 1$

iff ≥ ks children ← 1.

if(ss == leaf_node),

Then $T_{SS}(\gamma) \leftarrow 1$

iff $A(ss) \in \gamma$.

5. Construction of the proposed Merkle Tree-based Access Structure in CPABE

The main goal of our approach is to provide a Merkle Tree-based access structure for defining the access control policy for Patient-Centric data by distinguishing sensitive information qualities over the general access policy. The access tree lists all stakeholders or qualities that have permission to access the PCD.

5.1. System Model

Our proposed Merkle Tree-based access structure for sensitive access policy Tree T_S are shown in Figure 1, which includes sensitive attributes as leaf nodes with public attributes in Tree T_{NS}. Before being transmitted to the decryptor, a hash root $H(T_S)$ for the tree is discovered and appended to the public non-sensitive tree T_{NS}.

Our proposed scheme includes five algorithms:

5.1.1. Setup(γ,S) → (PK,MK)

The KGC creates PK and MK, respectively, given by γ and S.

5.1.2. KeyGen(PK, MK, S) → SK

For M. PK, MK and S, the KGC generates SK.

5.1.3. Construction of Merkle Tree (T_S) for sensitive attributes

The Merkle tree for sensitive characteristics is developed to experiment by considering the IBM Clinical Hub [20], which provides clinical data pieces that, when joined with patient identity information, enable the storage, production, and access of a longitudinal patient record. The IBM Clinical Hub is a more comprehensive workbench model that contains patient demographics, clinical data aspects, and access to consuming systems and processes. The longitudinal patient record is built from a stream of HL7 events and messages relating to patient visits, lab orders, results, admittance, discharge, and transfer events across numerous acute and ambulatory environments. Medication, allergies, test findings, difficulties, procedures, and family history are just a few items captured in the longitudinal patient record. The classification of sensitive and non-sensitive attributes is provided in Tables 5 and 6.

Table 4. IBM Clinical Hub Patient Health Record for public or Non-Sensitive attributes

S.No.	Non-Sensitive Attributes	Transaction IDs	S.No.	Non-Sensitive Attributes	Transaction IDs
1.	Patient Name	P_N	10.	Next of Kin Home Phone	K_P
2.	Home Address	H_A	11.	Next of Kin Relationship	K_R
3.	Billing Address	B_A	12.	Gender	G
4.	Work/Other Address	W_A	13.	Marital Status	M_S
5.	Email	E_m	14.	Patient Maiden Name	M_N
6.	Home Phone	H_P	15.	Patient Mother's Maiden	M_{MN}
7.	Mobile Phone	M_P	16.	Previous / Alias Name	A_N
8.	Work/Other Phone	W_P	17.	Language	L
9.	Next of Kin Name	K_N			

Table 5. IBM Clinical Hub Patient Health Record private or sensitive attributes

S.No.	Sensitive Attributes	Transaction IDs	S.No.	Sensitive Attributes	Transaction IDs
1.	Primary Physician	PP	16.	Vitals	V
2.	Date of Birth	DOB	17.	Family History	FH
3.	Driver's License Number	DLN	18.	Visit info	VI
4.	E-Patient ID	PID	19.	Diagnosis	DI
5.	Medical Record Number	MRN	20.	Exam Date	ED
6.	Social Security Number	SSN	21.	Patient Drugs	PD
7.	Patient Blood Group	PBG	22.	Patient Measurements	PM
8.	Insurance	IN	23.	Patient Risk Factors	PRF
9.	Guarantor Name	GN	24.	Patient Study	PS
10.	Guarantor Phone	GP	25.	Patient Symptoms	PSY
11.	Guarantor Relationship	GR	26.	Patient Treatment	PT
12.	Last Activity/Service Date	LA	27.	Patient Outcomes	PO
13.	Allergies	A	28.	Organizational Info	OI
14.	Immunizations	IM	29.	Physician Info	PHI
15.	Habits	HA	30.	Staff Info	SI

Fig. 1 Our Proposed Merkle tree Construction for Sensitive Attributes of Patient Health Data

5.1.4. Encrypt (PK, T_{NS}, T_S, M) → CT

The sender outputs the CT with input PK, T_{NS}, T_S, and M, where T_{NS} is a non-sensitive and public tree, and T_S is a sensitive and hash tree.

5.1.5. Decrypt (CT, SK) → M

When given the inputs CT and SK, the receiver decrypts and outputs a message, CT and M, respectively. Merkle proofs can show that a certain attribute is present by climbing the hash tree T_S from a transaction to the root. The decryptor checks the public tree T_{NS} and finds the Merkle Proof for T_S to check the message's integrity through hidden attributes.

6. Results and Discussion

Table 6: Results

S.No.	*Number of Leaf Nodes*	*Number of iterations*	*Leaves size in bytes*	*Internal nodes' size (bytes)*
1	24	47	1344	1656
2	90	47	5040	6408
3	165	47	9240	11808

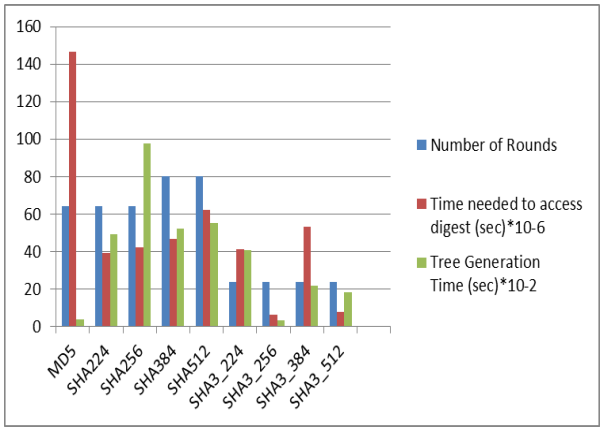

Fig. 2

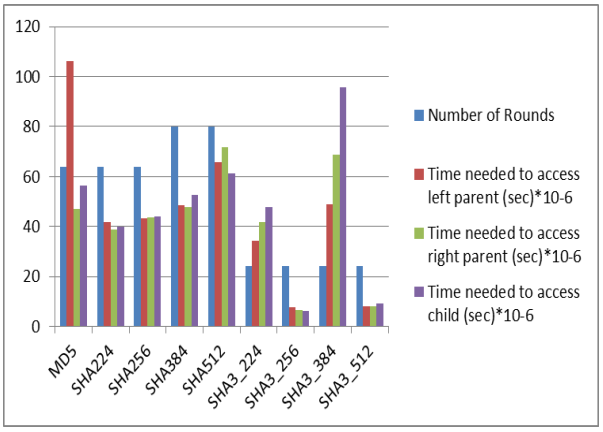

Fig. 3

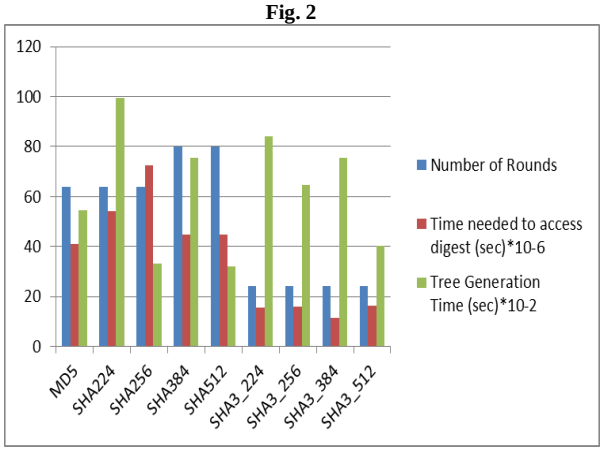

Fig. 4

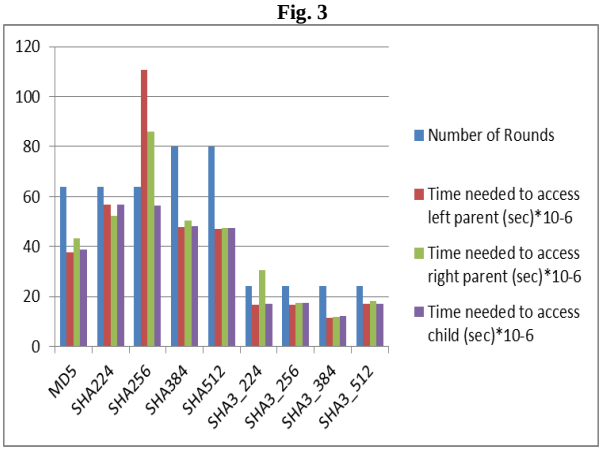

Fig. 5

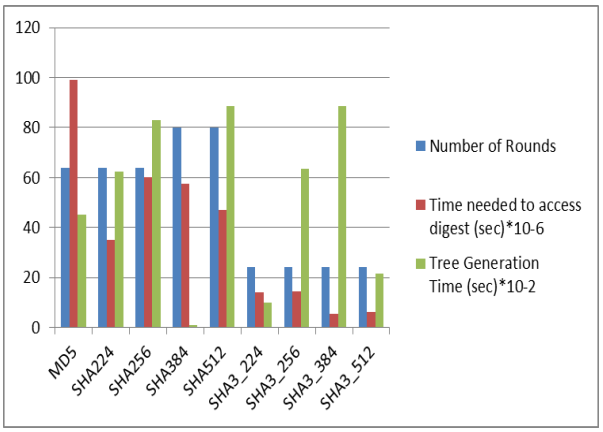

Fig. 6

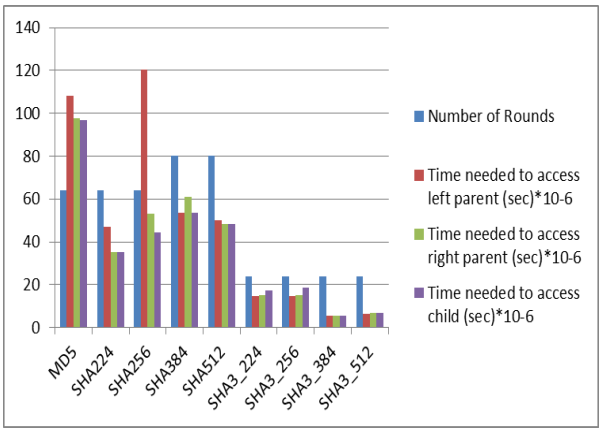

Fig. 7

Figures 2nd, 4th, and 6th show the hash algorithms vs. the time it takes to retrieve the digest and the time it takes to generate the tree. The 3rd, 5th, and 7th figures are time spent contacting the left parent, the right parent, and the child

6.1. Comparison with other frameworks

We provide a full comparison based on the findings in the above results, covering policy type, tree creation time, privacy-preserving, data integrity, fine-grained access control [5], and time to access nodes. The proposed Merkle Tree-based paradigm is compared to several different techniques. By creating a Merkle tree, guaranteeing integrity via a hashing method, and offering fine-grained access control, our proposed CP-ABE scheme is capable of implementing concealed sensitive attribute access policies.

Table 7. Comparison of the proposed model with other frameworks

Ref. No.	Policy Type	Efficiency	Privacy-Preserving	Tree Generation time	Data Integrity	Time to Access Nodes	Fine-Grained Access Control
[1]	Hides public and sensitive access	Low	Y	Moderate	Y	Moderate	**Y**
[17]	No Hidden policy	Low	Y	Moderate	N	Moderate	**Y**
The Proposed Model	Hides only Sensitive attributes	Moderate	Y	Less	Y	Less	**Y**

7. Conclusion

It is expected to offer a security response for secure sensitive data and attributes in modern patient health care. The Merkle tree-based CP-ABE system is the model proposed in this paper. It can also improve user and provider security and integrity in medical care and clinical associations.

Our framework is robust to attacks like collusion if in case any attacker obtains multiple private keys and permits policies to be transmitted as a monotonic access structure. Finally, we demonstrate how to use our system with various hash types. One constraint of our architecture is that it must be proven secure using a set of static inputs. Work has started on encrypting huge amounts of data and making a Merkle proof that can be done quickly.

Appendix A: Symbols Table

Table 8. Symbols Table

S.No.	Symbol	Description
1	a1,a2,a3…an, a_s	Attributes
2	A,B,C,S	Attribute Set
3	T_S	Sensitive attribute Access Tree
4	T_{NS}	Non-Sensitive attribute Access Tree
5	T_{SS}	Sub Tree
6	T	Threshold value

7	ks	Child node
8	num_s	Number of children of s
9	H	Hash function
10	RH	Root Hash
11	γ	Security Variable
12	PK	Public Key
13	MK	Master Key
14	SK	Secret Key
15	KGC	Key Generation Centre
16	PCD	Patient-Centric Data
17	M	Input Message
18	HL7	Health Level 7
19	GDPR	General Data Protection Regulation
20	ABE	Attribute Based Encryption
21	CPABE	Cipher-Text Policy ABE
22	CT	Cipher Text

References

[1] Siti Dhalila Mohd Satar, Mohamad Afendee Mohamed, Masnida Hussin, Zurina Mohd Hanapi and Siti Dhalila Mohd Satar, Cloud-based Secure Healthcare Framework by using Enhanced Ciphertext Policy Attribute-Based Encryption Scheme International Journal of Advanced Computer Science and Applications (IJACSA), (2021) 12(6).

[2] Xu, R. and Lang, B., A CP-ABE Scheme with Hidden Policy and its Application in Cloud Computing. International Journal of Cloud Computing, 4(4) (2015) 279.

[3] Aldeen, Y. and Salleh, M., Techniques for Privacy-Preserving Data Publication in the Cloud for Smart City Applications. Smart Cities Cybersecurity and Privacy, (2019) 129-145.

[4] (2022) Irwin, L., GDPR | Personal Data vs Sensitive Data: What's the Difference? IT Governance UK Blog.

[5] Meng, F., et.al. Ciphertext-Policy Attribute-Based Encryption with Hidden Sensitive Policy from Keyword Search Techniques in Smart City. EURASIP Journal on Wireless Communications and Networking, 1 (2021).

[6] Meng, F., Cheng, L. and Wang, M. ABDKS: Attribute-Based Encryption with Dynamic Keyword Search in Fog Computing. Frontiers Of Computer Science, 15(5) (2021).

[7] Goyal, V., Pandey, O., Sahai, A. and Waters, B. Attribute-based encryption for fine-grained access control of encrypted data. Proceedings of the 13th ACM conference on Computer and communications security - CCS '06, (2006).

[8] Greene, E., Proctor, P. and Kotz, D. Secure sharing of mHealth data streams through cryptographically-enforced access control. Smart Health, 12 (2019) 49-65.

[9] Nishide, T., Yoneyama, K. and Ohta, K. Attribute-based encryption with partially hidden encryptor-specified access structures | Proceedings of the 6th international conference on Applied cryptography and network security, (2022).

[10] Doshi, N. and Jinwala, D. Constant Ciphertext Length in Multi-Authority Ciphertext Policy Attribute Based Encryption. 2011 2nd International Conference on Computer and Communication Technology (ICCCT-2011),

[11] Helil, N. and Rahman, K. CP-ABE Access Control Scheme for Sensitive Data Set Constraint with Hidden Access Policy and Constraint Policy. Security and Communication Networks, (2017) 1-13.

[12] Hur, J., Park, C. and Hwang, S. Fine-Grained User Access Control in Ciphertext-Policy Attribute-Based Encryption. Security and Communication Networks, 5(33) (2011) 253-261.

[13] Lewko, A. and Waters, B. Decentralizing Attribute-Based Encryption. Advances in Cryptology – EUROCRYPT, (2011) 568-588.

[14] Vijayan, V., Connolly, J., Condell, J., McKelvey, N. and Gardiner, P. Review of Wearable Devices and Data Collection Considerations for Connected Health. Sensors, 21(16) (2021) 5589.

[15] Doshi, N. and Jinwala, D. Fully secure ciphertext policy attribute-based encryption with constant length ciphertext and faster decryption. Security and Communication Networks, 7(11) (2013) 1988-2002.

[16] V. Echeverría, L. M. Liebrock and D. Shin, Permission Management System: Permission as a Service in Cloud Computing, IEEE 34th Annual Computer Software and Applications Conference Workshops, (2010) 371-375.

[17] J. Bethencourt, A. Sahai and B. Waters, Ciphertext-Policy Attribute-Based Encryption, 2007 IEEE Symposium on Security and Privacy (SP '07), (2007) 321-334.

[18] D. Meng, E. Luo and G. Wang, A Privacy-Preserving Multi-Authority Attribute-Based Encryption Approach for Mobile Healthcare, IEEE 13th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), (2016) 299-306.

[19] Liu, X., Ma, J., Xiong, J., Li, Q. and Ma, J. Ciphertext-Policy Weighted Attribute Based Encryption for Fine-Grained Access Control. 5th International Conference on Intelligent Networking and Collaborative Systems, (2013).

[20] Longitudinal Patient Records Artifacts. Longitudinal Patient Records Artifacts, www.ibm.com, 12 Apr. (2021).