Preserving Privacy in Cyber-physical-social Systems: An Anonymity and Access Control Approach

Tanusree Sharma
Informatics, University of Illinois at Urbana Champaign
Champaign, IL, USA
tsharma6@illinois.edu

John Christian Bambenek
Informatics, University of Illinois at Urbana Champaign
Champaign, IL, USA
bambenek@illinois.edu

Masooda Bashir
School of Information Sciences
University of Illinois at Urbana Champaign
Champaign, IL, USA
mnb@illinois.edu

ABSTRACT

With the significant development of mobile commerce, the integration of physical, social, and cyber worlds is increasingly common. The term Cyber Physical Social Systems is used to capture technology's human-centric role. With the revolutionization of CPSS, privacy protections become a major concern for both customers and enterprises. Although data generalization by obfuscation and anonymity can provide protection for an individual's privacy, over-generalization may lead to less-valuable data. In this paper, we contrive generalization boundary techniques (k-anonymity) to maximize data usability while minimizing disclosure with a privacy access control mechanism. This paper proposes a combination of purpose-based access control models with an anonymity technique in distributed computing environments for privacy preserving policies and mechanisms that demonstrate policy conflicting problems. This combined approach will provide protections for individual personal information and make data sharable to authorized party with proper purposes. Here, we have examined data with k-anonymity to create a specific level of obfuscation that maintains the usefulness of data and used a heuristic approach to a privacy access control framework in which the privacy requirement is to satisfy the k-anonymity. The extensive experiments on both real-world and synthetic data sets show that the proposed privacy aware access control model with k- anonymity is practical and effective. It will generate an anonymized data set in accordance with the privacy clearance of a certain request and allow users access at different privacy levels, fulfilling some set of obligations and addressing privacy and utility requirements, flexible access control, and improved data availability, while guaranteeing a certain level of privacy.

KEYWORDS

CPSS, Data privacy and security in CPSS, Access Control, Anonymity Model.

1 INTRODUCTION

With growing technological advances, Cyber Physical Social Systems (CPSS) have increasingly been used in automobile, chemical composition, robotics, and numerous other cloud-based and IoT applications. CPSS provide many features which enable us to leverage the potential of cloud scalability, context relevant experiences, network-based infrastructures, constructive documentation tools, and cross platforms, to name a few. The main advantage that CPSS offers is enabling human input in the loops. It generates a faster response with a shorter decision time because user- and customer-generated social data can be used as an unbiased sensor network for natural experimentation by extracting useful patterns and deploying intelligence to serve the entity to make predictions about future events and decision making [20]. CPSS can be utilized as a decision aiding framework and for designing architectural strategies for existing and new applications by tagging based systems or services where a human remains in the sensing loop or where social sensing data is a good option to train machine to make decisions with trained data set and fact-finding algorithms. However, these enabling technologies, which make the automatic design of CPSS feasible, also introduce multiple privacy and security challenges that need to be examined. One of the most important aspects that has not been researched well is how users' contributions to the system are protected from the privacy and security point of views. Due to the open network structure and service sharing scheme of the cloud, it imposes very challenging obstacles to security, as CPSS are relatively sophisticated systems, ranging from integratation of multiple devices to highly heterogeneous networks and the possible severity of the physical environment. Therefore, CPSS are more susceptible to targeted attacks since this system includes cyberspace, physical space and social space, where the malicious users can attack from multiple links and sources: for example, the location data that comes from GPS or the user's handheld device in social space or the user's authentication information in cyberspace. Malicious attackers may eavesdrop on sensitive information if there is lack of reasonable security and privacy mechanisms.

One important technique that is often used to protect private information (static or dynamic) in distributed systems is specifically tailored to support privacy policies. Securing private information cannot be easily achieved by traditional access management systems because traditional access management systems focus on which user is performing what action on which data object [18], and privacy policies are concerned with which data object is used for what purpose(s). When the users will be sharing or searching their location using apps like Foursquare and Swarm, which shopping mall/ hospitals they are visiting that might expose their data, including name, age, diseases, current location, and historical locations. If the privacy and data sharing policy is not defined clearly, including who will be using the data and for what purpose, then there will be complexities that might expose their data to unauthorized data collectors. Again, for hiding identifiable information, there are several anonymity and obfuscation techniques that have been developed by several researchers. However, anonymity is not enough to accomplish the purpose of CPSS. There is no doubt that

1st Workshop on Cyber-Physical Social Systems (CPSS2019), October 22, 2019, Bilbao, Spain.

users may be willing to participate in data aggregation but probably do not intend to have their private information leaked. For example, we can get our step numbers on WeChat everyday and share with our friends to establish a ranking list. However, the data collected by the cyber devices may contain personal information, which users may not want to be leaked. How to aggregate data with privacy preservation therefore has become a critical and challenging issue that hampers the advancement of data aggregation in CPSS [19]. Through our research, we have worked on how to preserve users' data privacy and security while maintaining the purpose of CPSS, which is data aggregation. Our combined approach proposes a comprehensive framework for purpose and data management where purposes are organized in a hierarchy. In our approach each data element is associated with a set of purposes, as opposed to the single security level in traditional secure applications. We have combined the anonymity model to hide the identification information from unauthorized third-party data collectors or other external users. In the following section, we will be presenting research work for preserving privacy in CPSS.

2 RELATED WORKS

Preserving privacy in the CPSS has been attracting attention from both academia and industries. Most of the prior research studies has focused on data privacy and has not considered the usability of CPSS.

Pitofsky [14] showed that 97 percent of web sites and distributed systems were collecting at least one type of identifying information such as name, home address, e-mail address, postal addressws of consumers, or current or historical locations. The fact that personal information is collected and can be used without any consent or awareness violates privacy for many people. Access control mechanisms for enforcing such policies have not been investigated [8]. Ni et al. [11] analyzed conditional privacy management with role-based access control, which supports expressive condition languages and flexible relations among permission assignments for complex privacy policies. It is important to note that simply removing identity information, like names or social-security numbers, from the released data may not be enough to anonymize the data. Many examples show that even when such information is removed from the released data, the remaining data, combined with other information sources, may still link the information to the individual [17]. Sweeney [15] proposed approaches based on the notion of k-anonymity as solutions to the problem. Another secure private information techniques such as density-based clustering algorithms happens in the context of data mining [10].

Trust-based security approaches are widely applied in CPSS. Privacy preservation in CPSSs has become increasingly important and thus attracts attention from both academic and industrial communities. This issue has drawn even more attention in the recent years due to pervasive embedded sensors in mobile devices. Privacy protections are also becoming a significant consideration for both customers and enterprises in today's corporate marketing strategies. This raises challenging questions and problems regarding the use and protection of private messages, especially for context-aware web services [4]. One principle of protecting private information is based on who is allowed to access private information and for what purpose [1]. For example, personal information provided by patients to hospitals may only be used for record keeping purposes not for advertising purposes. So, there must be a purpose for data collection and data access. The work in [7] proposes trust architecture for pervasive systems by extending SPKI and role-based access control. In particular, the framework is enabled based on a distributed model that employs various security agents to identify the authentication within their service domain. Additionally, ontology is utilized to specify the user's permission rule, and a delegation chain is used to deliver the access privilege between multiple users. In the work of [12], a cyber-physical-social security architecture is presented for future IoT, it is divided into three layers to protect the security of IoT, including information security, physical security and management security. The work in [3] presents a trust-based personalized privacy model for pervasive environments. It mainly includes a trust privacy manager and a user-centric privacy framework. A trust privacy manager is a user-centric abstraction in which the goal is to realize the balance between privacy protection, service usability and user manageability. Further, a user-centric privacy framework as a reference framework that not only offers privacy control but also gives the brokering context to interact with external parties. The component of a user-centric privacy framework is developed by a service- oriented architecture framework. To some extent, it is expected to enable the loose coupling of the holistic architecture and to achieve high flexibility for privacy management.

However, while existing security and privacy approaches aim to address the security of embedded systems, Cyber Physical systems, and Cyber Social Systems, they are tricky to adapt to the multiple security requirements of CPSS. Currently there lacks a universal framework to integrate approaches for CPSS. Here, in our paper we are demonstrating with mathematical and logical expressions how to specify and enforce policies for authorizing purpose-based access management and combining anonymity techniques.

3 PROPOSED PRIVACY ARCHITECTURE

There are two parts of our security architecture to increase the level of user's privacy while maintaining the quality of data that can be shared with authorized data collector to make the CPSS a useful framework to deal with. Figure 1 shows the design architecture. Only obfuscation and anonymity can be reasons for hampering or disregarding the importance of data aggregation from CPSS that are useful in many ways. Hence, combining access control for authentication and verification to get information from users will be helpful for authorized third parties to gain their purpose and also restrict attackers.

For this use case, we will be considering external user attacks from three main adversaries who are more likely to want to access, or are prone to use, the online data or any relational query data. Addressing three kinds of adversaries at the same time would not be feasible from the perspective of privacy access control and k-anonymity mechanism. Yet while the motivation behind external users' attacks can be divergent, their approaches are often similar. They mostly seek to sow disruption misdirect by planting misleading data or taking down police and government systems. For now, using a single use case for this research will make the discussion more effective and we can say this integrated method can serve

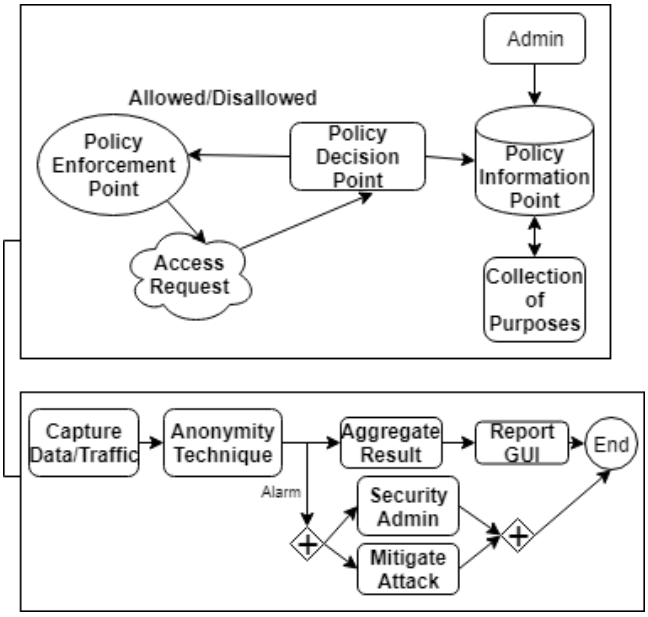

Figure 1: Combined Privacy preserving Model (Access Control Policy and Anonymity)

the purposes that we have mentioned above to fulfill mainly two main missions here which are data availability for authorized users and guarantee a certain amount of privacy access control of the information shared online or any relational databases by users.

3.1 Anonymity Technique

According to [13] [16] [2] [9], our privacy model could be consistent with the objective that of publishing truthful data against both reidentification and semantic attacks by satisfying criterions of $k-$ anonymity, $l-$ iversity and $t-$ closeness. First, $k-$ anonymity ensures the attacker cannot distinguish the victim from at least $k-1$ other individuals, which is used against reidentification attacks. We can also say that k-anonymity protects the privacy of individual persons by pooling their attributes into groups of at least k people, assuming the data set has N entries, and each entry includes attributes $X_i(i\epsilon[0,A])$ with information like age, gender, address, which are quasi identifiers. We are also assuming that the dataset only includes a single sensitive point of information like disease, income, or something what usually a person usually wants to protect. Our method will generalize the dataset with more than one sensitive data point, while there will be no indication of difference between quasi-identifier and sensitive information. Since we do not limit the attacker's knowledge about individual's trajectory, the victim's trajectory should be indistinguishable from at least $k-1$ other trajectories, which means these trajectories should be the same after generalization.

However, if all the persons in the group of data have the same sensitive attributes, the adversaries will still be able to learn about the sensitive attributes. In order to fix this problem, privacy criterions of $l-$ diversity and $t-$ closeness should be met, which requires that the sensitive attribute of a $k-$ anonymous set contains at least l well-represented values for the sensitive attribute. With probabilistic reason, adversaries can still access about a person's information where t-closeness is significant. It demands that the statistical distribution of the sensitive attribute values in each k-anonymous group is "close" to the overall distribution of that attribute in the entire dataset. Closeness can be measured using e.g. the Kullback-Leibler (KL) divergence.

Third, in order to maintain the truthfulness of the dataset, we only use spatiotemporal generalization and suppression to process the trajectory data. Spatial generalization means merging nearby base stations, and temporal generalization means increasing temporal granularity to combine different trajectories into one. When merging some spatiotemporal points causes a huge loss of spatiotemporal granularity, we just delete them, which is called suppression.

Turning a dataset into a anonymous dataset is a difficult problem and even finding the optimal partition into k-anonymity is NP-Hard. We have used greedy search technique "Mondrian" to partition the original data.

Quasi Identifier: pieces of information that are not of themselves unique identifiers, but are sufficiently well correlated with an entity

Sensitive Attribute: Information related to a specific individual that can cause a privacy breach.

Algorithm 1: Partitioning data to k-anonymous group

```
Result: Complete Set of Partitions
initialize complete set to empty set, P_com = 0
Initialize working set of partition to set containing a
  partition with entire dataset
P_working = (1, 2, 3, 4, ...n)
while Partition in working set do
    pop;
    Calculate span(columns in partition);
    Sort Resulting columns;
    Split with median;
    if partition with anonymity then
        add new partition;
    else
        add original partition to complete partitions;
    end
end
```

3.2 Purpose-Based Access Control

This paper bridges the gap between private information protecting mechanisms and access control models. We propose a purpose-based access control framework with an anonymity technique. This section develops the purpose-based access control framework, which includes extended access control models and supports purpose hierarchy by introducing the intended and access purposes and purpose associated data models.

The purpose explains the reason(s) for collecting data and accessing it [5]. If there is a set of purposes P that is organized in a tree structure, then each node represents a purpose in P and each

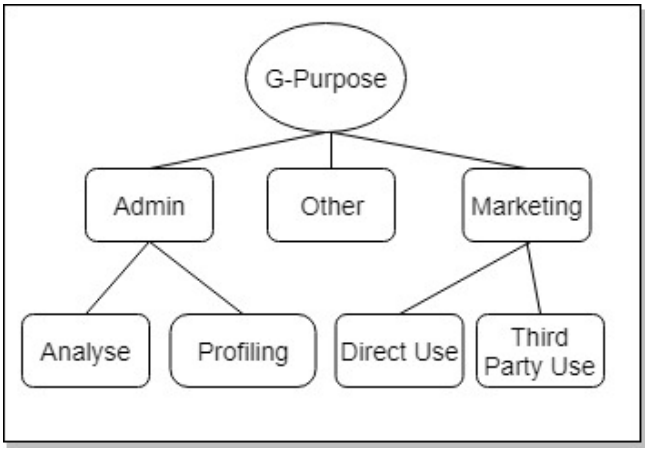

Figure 2: Example of purpose structure (inspired by [5])

edge represents relations between two purposes. Figure 2 shows the purpose structure tree.

Assuming P_i and P_j be two purposes in the hierarchical purpose tree where P_i is the predecessor of P_j. There remains some relationship among the purposes based on the tree structure of purposes. Suppose in the purpose tree, while P is a set of purposes, $P_k \epsilon P$ is a purpose, the predecessor purposes of P_k which is the set of all nodes that are senior to P_u. On the above tree structure, Predecessor (Direct Use) = Marketing, G-Purpose in figure 2. The junior purposes of P_k, is the set of all nodes that are junior to P_k. For instance, Successor (Admin) = Analyze, Profiling. We have followed the research work of [5] to design an access control model with a stated privacy policy by adding purposes for data objects to be confirmed if a particular data element is allowed to be shared. Access purpose authorizations are granted to users based on the access purpose on the data, obligations and conditions. In order to perform privacy-preserving access control, a system needs to enforce privacy requirements stated by data owners.

3.2.1 Definition: According to the basic privacy access control model, there are a few components in it. Mainly there are three entities that are used in a basic access control system: subjects, objects, and operations. Based on the access control model, purposes, role, policy would be added. A set S of Subjects, a set D of Data, a set P of purposes, a set A of actions, a set of O for obligations and a set of C for conditions.

- `Set of data access right`: (d, a)|a ϵ A, d ϵ D
- `Private data access right`: (da,a,p,c,o)|daϵ DA,pϵ P,cϵ C,oϵ O,aϵ A
- `Assignment of private data subject`: access of private information.
- `Purpose`: The reason for access.

For example:

- `Subjects`:*Amazon*, *eBay*, *Fedex*, *Customer − service*
- `Data`:*InfoOrder*, *ContactInfo*, *MailingAdd*, *EmailAdd*
- `Action`:*check*, *Update*, *Delete*
- `Purpose`:*Order*, *Complaint*, *Billing*, *Shipping*, *ProblemSolving*

The following privacy policies:
1. "Amazon can check customers' MailingAdd for shipping purpose".
2. "eBay can only check customers' EmailAdd for sending further alert if they allow to do so".
3. "Fedex may check customers' InfoOrder for Billing purpose and customers will be informed by Email".
4. "Customer-service can check customers' ContactInfo for Problem solving if it is approved by Amazon".

Hence, these policies are expressed as follows in a privacy access control model: P1: (Amazon, (MailingAdd, check), Shipping, N/A, ϕ); $P2$: (eBay, (EmailAdd, check), Purchase, OwnerConsent = 'Yes', ϕ); $P3$: (Fedex, (InfoOrder, Check), Billing, N/A, Notify(ByEmail)); $P4$: (customer-service, (ContactInfo, check), Problemsolving, 'Approved by Amazon', N/A)

3.2.2 Policy Operation: With the change of technological and regulatory affairs, new policies need to be added. This section analyzes the impact of generating new policies to add to an existing privacy access control model. Sometimes, a new policy for privacy protection is raised, but it might not be addressed. For example, when eBay moves to the complaint section, a new policy need to be addressed.

5. "eBay can only check Email address of customers, for complaint purpose if they are allowed by customers" The corresponding expression will be reflected in the model:
P5:(eBay,(EmailAdd, check), Complaint, OwnerConsent = 'Yes', ϕ).

Algorithm 2: Component Checking for access by [6]

Comp-Check1(ap,AIP, PIP)
1. if $ap \epsilon PIP then$
2. return False;
3. else if $ap \epsilon AIP \downarrow then$
4. return True;
5. end if
Comp-Check2(ap, CIP, PIP)
1. if $ap \epsilon PIP then$
2. return False;
3. else if $ap \epsilon CIP \downarrow then$
4. return True;
5. end if
Where, AIP: Allowed Intended Purpose; PIP: Prohibited Intended Purpose; CIP: Conditional Intended Purpose

Now compared with previous purpose $P2$, these are two policies for eBay to access email addresses. There are two different purposes: one for purchase and one for compliant. Now how would the system verify Complaint to access email addresses within consent conditions? To make it simpler and for polices to be updated, we can use a conjunction here for two different purposes. That is, if a user wants to access right a_r on data d for purpose Pu, all access polices related to $((d, a_r), P)$ need to be checked. So, in the example above, eBay can check the email address if there exists at least one policy (purchase or compliant) that will satisfy all policies. If a new access policy is related to the same user, same data, same right and same conditions of some existed private policies, it is not used to relax the access situations but to make the access stricter. If privacy

admin wants to ease/modify the access environments, they can do so by revising the existed access policies instead of creating a new one. For Policy checking, here we utilized the algorithms by [6]. Finally, the access decision is constructed based on the Comp-Check and intended purposes of a specific attribute.

4 IMPLEMENTATION

We implement a simple algorithm multi-dimensional k-anonymity to produce a k-anonymous dataset. k-anonymity protects the privacy of individual persons by pooling their attributes into groups of at least k people. We explore the "Mondrian" algorithm which uses a greedy search algorithm to partition the original data into smaller and smaller groups. (If we plot the resulting partition boundaries in $2D$ they resemble the pictures by Piet Mondrian, hence the name.) The algorithm assumes that we have converted all attributes into numerical or categorical values and that we're able to measure the "span" of a given data attribute X_i.

The algorithm proceeds then to partition the data into k-anonymous groups. After obtaining the partitions, we still need to aggregate the values of the quasi identifiers and the sensitive attributes in each k-anonymous group. For this, we can e.g. replace numerical attributes with their range (e.g. "age: 24-28") and categorical attributes with their union (e.g. "employment-group: [self-employed, employee, worker]"), though other aggregations are possible. Methods like [5] even preserve the micro-data in each group, which increases re-identification-risk.

We are using text data of Adult with different quasi identifiers like age, work class, education, marital status, occupation, race, and age, and also containing sensitive attribute income. For our implementation purpose, first we have considered two columns from the dataset to apply partition to speed up the execution. With that execution, 500 partitions have been created. The results after creating partitioning functions to divide datasets are below for better visualization. After partitioning and sorting the resulting data frame using features columns and sensitive attributes, we have a k-anonymous dataset with age, count, education and income.

For generating k-anonymous data that contains one row for each partition and value of sensitive data, we aggregate the columns of each partition.

We implement l-diversity in order to protect the privacy of the persons in the dataset even better. The image below cam make it more understandable.

For t closeness: As we can see, for regions where the value diversity is low, our l-diverse method produces partitions that contain a very large number of entries for one value of the sensitive attribute and only one entry for the other value. This is not ideal because while there is "plausible deniability" for a person in the dataset (after all the person could be the one "outlier"),an adversary can still be very certain about the person's attribute value in that case. t-closeness solves this problem by making sure that the distribution of sensitive attribute values in a given partition is similar to the distribution of the values in the overall dataset. We generate the global frequencies for the sensitive columns.

In our model, customers are given three more possible options for using their data. These make them comfortable to release their data fully or conditionally, knowing the private information will be

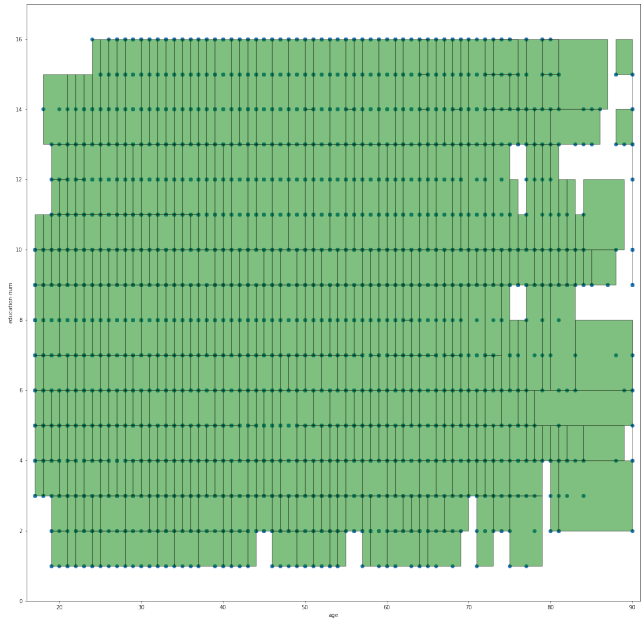

Figure 3: anonymous data visualization after partitioning

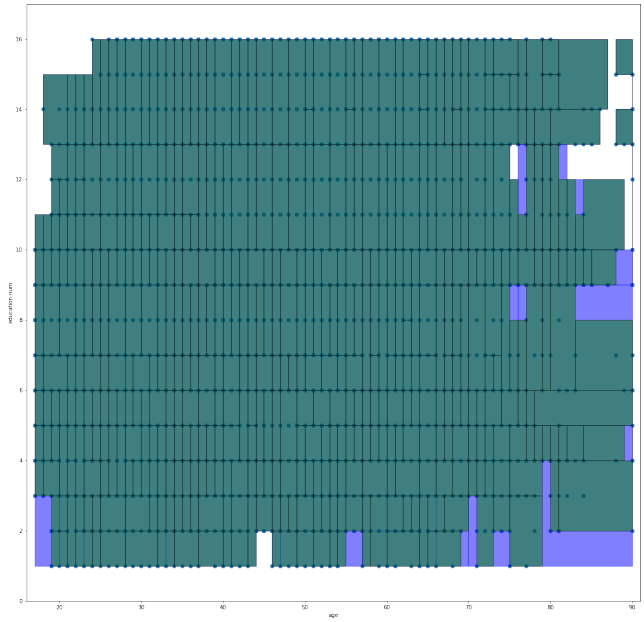

Figure 4: After Applying I-diversity

protected. After data are collected, intended purposes with three different levels will be associated with data. As the intended purpose is assigned to every data element, an intended-purposes table (IPT) is formed. Data providers (customers) are able to control the release of their data by adding privacy levels to the IPT which will not affect data in the database. After authorizing an access purpose, users get access to purpose permissions from the access control

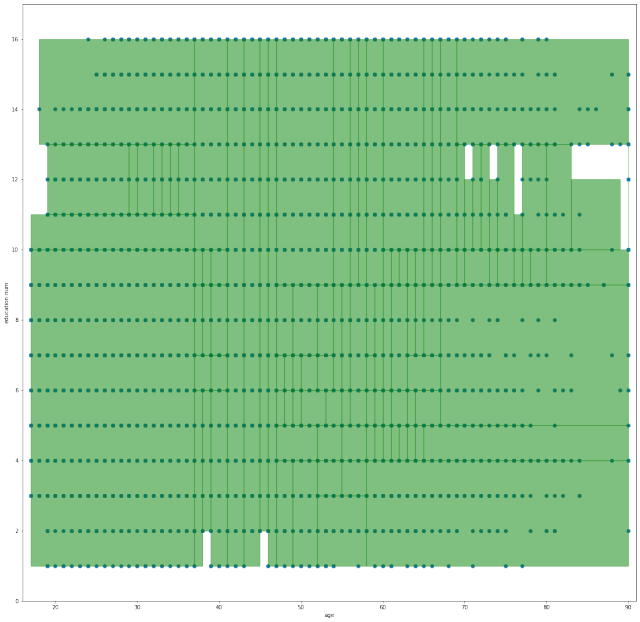

Figure 5: After applying t-closeness

engine. The access control engine needs a match process to finish the compliance computation fully or conditionally in accordance with access purposes and intended purposes. If the requester's access purpose is fully compliant with the intended purposes of requested data, the engine will release full data to the requester. On the other hand, if the access purpose is conditionally compliant, the engine will release conditional data to the requester; otherwise returned data will be null. Thus, in this model the search engine needs to evaluate two compliance checks, the first one is for full compliance and the second one is for conditional compliance.

5 CONCLUSION

From our research, we can conclude that a combined approach of anonymity and a purpose-based access control policy foster a privacy preserving environment for personal information. Formulating the interaction between these two mechanisms make the cyber physical social system more usable and at the same time preserve a certain level privacy. We have also analyzed the impact of adding new policies and the conflicts that can result. Algorithms have been developed to help a system detect and solve these problems. Furthermore, the experimental results demonstrate the practicality of the algorithms. The evaluation of the dataset validates the effectiveness of the algorithm, and the component check for purpose-based privacy paves the way to a direct, proper policy for access control. For our future work, we will evaluate more datasets with this method and extend this model to incremental data.

REFERENCES

[1] Rakesh Agrawal, Jerry Kiernan, Ramakrishnan Srikant, and Yirong Xu. 2002. Hippocratic databases. In *VLDB'02: Proceedings of the 28th International Conference on Very Large Databases*. Elsevier, 143–154.

[2] Machanavajjhala Ashwin, Kifer Daniel, Gehrke Johannes, and Venkitasubramaniam Muthuramakrishnan. 2007. l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data* 1, 1 (2007), 1–52.

[3] Susana Alcalde Bagüés, Andreas Zeidler, Ignacio R Matias, Cornel Klein, and Carlos Fernández-Valdivielso. 2010. Enabling Personal Privacy for Pervasive Computing Environments. *J. UCS* 16, 3 (2010), 341–371.

[4] Elisa Bertino, Pierangela Samarati, and Sushil Jajodia. 1997. An extended authorization model for relational databases. *IEEE Transactions on Knowledge and Data Engineering* 9, 1 (1997), 85–101.

[5] Gabriel Ghinita, Panagiotis Karras, Panos Kalnis, and Nikos Mamoulis. 2009. A framework for efficient data anonymization under privacy and accuracy constraints. *ACM Transactions on Database Systems (TODS)* 34, 2 (2009), 9.

[6] Md Enamul Kabir, Hua Wang, and Elisa Bertino. 2010. A role-involved conditional purpose-based access control model. In *E-Government, E-Services and Global Processes*. Springer, 167–180.

[7] Lalana Kagal, Tim Finin, and Anupam Joshi. 2001. Trust-based security in pervasive computing environments. *Computer* 34, 12 (2001), 154–157.

[8] Min Li, Xiaoxun Sun, Hua Wang, Yanchun Zhang, and Ji Zhang. 2011. Privacy-aware access control with trust management in web service. *World Wide Web* 14, 4 (2011), 407–430.

[9] Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian. 2007. t-closeness: Privacy beyond k-anonymity and l-diversity. In *2007 IEEE 23rd International Conference on Data Engineering*. IEEE, 106–115.

[10] Jinfei Liu, Joshua Zhexue Huang, Jun Luo, and Li Xiong. 2012. Privacy preserving distributed DBSCAN clustering. In *Proceedings of the 2012 Joint EDBT/ICDT Workshops*. ACM, 177–185.

[11] Qun Ni, Elisa Bertino, Jorge Lobo, Carolyn Brodie, Clare-Marie Karat, John Karat, and Alberto Trombeta. 2010. Privacy-aware role-based access control. *ACM Transactions on Information and System Security (TISSEC)* 13, 3 (2010), 24.

[12] Huansheng Ning and Hong Liu. 2012. Cyber-physical-social based security architecture for future internet of things. *Advances in Internet of Things* 2, 01 (2012), 1.

[13] Zahid Pervaiz, Walid G Aref, Arif Ghafoor, and Nagabhushana Prabhu. 2013. Accuracy-constrained privacy-preserving access control mechanism for relational data. *IEEE Transactions on Knowledge and Data Engineering* 26, 4 (2013), 795–807.

[14] Matthias Schunter and C Powers. 2003. The enterprise privacy authorization language (epal 1.1).

[15] Latanya Sweeney. 2002. Achieving k-anonymity privacy protection using generalization and suppression. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10, 05 (2002), 571–588.

[16] Latanya Sweeney. 2002. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10, 05 (2002), 557–570.

[17] Vicenç Torra. 2010. Towards knowledge intensive data privacy. In *Data Privacy Management and Autonomous Spontaneous Security*. Springer, 1–7.

[18] Hua Wang, Yanchun Zhang, and Jinli Cao. 2008. Effective collaboration with information sharing in virtual universities. *IEEE Transactions on Knowledge and Data Engineering* 21, 6 (2008), 840–853.

[19] Jiahui Yu, Kun Wang, Deze Zeng, Chunsheng Zhu, and Song Guo. 2019. Privacy-preserving data aggregation computing in cyber-physical social systems. *ACM Transactions on Cyber-Physical Systems* 3, 1 (2019), 8.

[20] Daniel Zeng, Hsinchun Chen, Robert Lusch, and Shu-Hsing Li. 2010. Social media analytics and intelligence. *IEEE Intelligent Systems* 25, 6 (2010), 13–16.