Privacy Preservation for Enterprises Data in Edge Devices

Aaloka Anant* and Ramjee Prasad

CTIF Global Capsule, Aarhus University, Herning, Denmark
E-mail: aaloka@anantprayas.org; ramjee@btech.au.dk
**Corresponding Author*

Received 06 July 2021; Accepted 07 November 2021;
Publication 11 February 2022

Abstract

Privacy becomes the most important topic as user's data gets more and more widely used and exchanged across internet. Edge devices are replacing traditional monitoring and maintenance strategy for daily used items in households as well as industrial establishments. The usage of technology is getting more and more pervasive. 6G further increases the importance of edge devices in a network as network speeds increase, making the edge device much more powerful element in the network. Edge devices would have massive store and exchange of personal data of the individual. Data privacy forms the primary requirement for accessing data of individuals. Paper presents a novel concept on combination of techniques including cryptography, randomization, pseudonymization and others to achieve anonymization. It investigates in detail how the privacy relevant data of individuals can be protected as well as made relevant for research. It arrives at an interesting and unique approach for privacy preservation on edge devices opening up new business opportunities and make the data subject in charge of their data.

Keywords: Privacy preservation, edge devices, anonymization, Data security, Enterprise data protection.

Journal of ICT Standardization, Vol. 10_1, 85–104.
doi: 10.13052/jicts2245-800X.1015

1 Introduction

Privacy is the freedom of an individual to let others know anything about themselves or not. Basic communication starts with self-introduction. In physical world, it's easy to forget an individual's identity. But in the digital world, once an identity is revealed, it may stay in digital format forever. The internet has enabled easy access to individuals across the globe. Individuals feel empowered to interact with people using internet and also transact to meet their basic needs using such channels of communication. The cost of such communication is possibly the lowest and its inviting even the poorest and secluded communities to interact with all and affirm their presence.

In modern day society with digital presence, privacy becomes more and more subject of discussion as its digital and easy to transmit, store, spread, analyse and in several ways to be misused or compromised. In his book in 1968, [1] Alan Westin, presented a detailed evaluation of the conflict between privacy and surveillance in modern society of those days. His observations indicated the shift in operating procedures of the governing body, policing structures, legal structures and highlighted that the privacy of individuals must be respected and supported with appropriate laws. Fifty plus years from the first major observation in this book, society is still looking for protection of privacy of the vulnerable individuals.

For example: An individual allowing the storage and consumption of cookies while browsing on internet. It can help in getting usable advertisements. Also, it can help in filling up forms, improving search results from search engines and so on. The data generated by the user, is sold by the search engines and the companies reading the cookies on browser to other companies. If the individual is asked to share his work on browser in public to anyone, they would never do so. But the same individual seldom gives a second thought to allow usage of his browsing history, which the company can process, derive information, and use this information to make money.

With new ways of communication, society is evolving into a never seen before operational structure, where individuals feel okay to give away their privacy to get some benefits and ease of doing things. The new methods are getting more and more pervasive and digging down to every aspect of life starting with smart watches recording heart beats to cars recording driving patterns and much more. With 6G enabling technologies like Telepresence, smart devices storing and exchanging biometric identity of individuals there would be massive data sets on edge devices. 6G would warrant not only security measures in data communication but privacy preservation as well while storing and exchanging personal data.

1.1 Privacy Preservation

Privacy preservation refers to storage, operations and transmission of data in such a way that privacy of the data subject is not exposed and not compromised. For example, data which is generated for a sales transaction, where an individual has purchased a few items in a store, contains the membership card and credit card information of the individual. In this case, providing access to this data in such a way that the individual, whose data is in discussion, cannot be identified by analysing this data, or cannot be traced back at the end of analysis of this data would mean that the privacy of the individual has been preserved in this operation.

When privacy of data is preserved, it becomes much more valuable as this data can be used to provide additional utility to the organization, which owns this data. Additional usage is possible, only ensuring the legal guidelines on data processing. With privacy preservation techniques the organization can fully ensure that there can be no misuse of this data to have any adverse effect on the individual.

Privacy preservation was often seen as a very technical topic with less impact on the business operations. With recent advancements in technology and Artificial Intelligence and Machine Learning coming into mainstream Enterprise applications, this topic has gained much attention.

1.2 Enterprise Data

Data which is used to conduct business operations with an individual can be considered as the Enterprise Data. For the sake of defining boundaries for this paper, we refer to Enterprise data only and not in general any data like text in social media. Enterprise data would include the data, which contains information on the individual for example photos, address, credit card information, driving license, age, shopping history, GPS coordinates, membership information of professional bodies including shopping cards, other information stored by an enterprise/Company to conduct business operations with an individual.

1.3 Edge Devices

An edge device is any device which controls data flow at the boundary of a network. With IOT and Industry 4.0 based applications coming in use, the role of edge devices is becoming more and more crucial for privacy preservation topic. An edge device which is used to generate data at the edge of a network and provide this data to the other elements in the network in order to provide services to an individual has a direct relation to an individual. Even though the edge device may not contain any specific information about an individual like her name, age, driving license, its association to an individual can still reveal a unique individual, leaving the individual vulnerable.

A surprising example can be a microwave oven [2], which transmit information on when its run, if its operating well and at what settings. This information can be easily used to understand when the user of the microwave is in the house using the device. Also monitoring patterns in such data over a period of time, it can be easily used to identify if there are visitors in the house, when the house is vacant for a long time, what are food preferences of the family and visitors leading to information on ethnicity, religion and racial identity and many other sensitive information.

Edge devices can range from a router in a network, to a sensor in a house. To define the scope of this paper we focus on edge devices, where there is an individual in scope namely smart devices in a household like microwave oven, smart refrigerator, smart temperature/climate controller; smart speakers like Alexa, google assistant, Siri, Cortana and others; always connected smart sensors like health armbands, sensor watches like Apple watch, Pebble watch etc, smart shoes and other smart wearables on body of individuals; smart cars like Tesla, other versions of cars like BMW, Mercedes and others, which consistently communicate with a central server on location, speed, and other sensor in the engine of car for purpose of predictive maintenance and other reasons; smart surveillance devices like video doorbell, IP camera, motion sensors, smart smoke detectors, smart fire alarms and other devices for protection of individual establishments; last but not the least to complete this list, the smart phone. In the current paper, we generalize these devices and do not go into each specific device its operations and each device's privacy preservation.

Paper is organised into 7 sections. Introduction is followed by second section on specifics of Enterprise data, followed by third section on existing practices. Fourth section outlines the solution from research followed by the fifth section on new business opportunities, which can emerge with such a solution. Sixth section presents a critique of the presented approach in section four, followed by the last section with conclusions and future scope of work.

2 What is Different with Enterprise Data

Enterprise data is data required by organizations in general to work/interact with an individual. The organization stores the data in a form, where it can re-use the data. The data is stored in multiple tables with relationships. Data is used by different applications and stored by different applications with a different structure to the data. Sometimes in the same company data of a user is stored in multiple applications. Even though user provides data in one application company has to replicate this data into multiple applications to make use of this data for operational reasons.

One example can be data for purchasing of goods in an e-commerce platform. The company which manages the e-commerce platform, collects users data. This company maybe selling goods by one of the suppliers, to whom this user data would be provided. The data would be replicated from a web-application to another application, which may be completely different like an operational Enterprise Resource Planning (ERP) system. The supplier can ship the item purchased by the user via another courier company, where the users data including mobile phone number, address etc would be provided to the logistics company. The logistics company may be exchanging goods in transit with other logistics company for delivery to the customer. Hence the user's data provided is exchanged, stored, retained and further transferred to several systems and players without user's explicit knowledge. Data being mis-used/hacked/leaked by any of the players in this example can lead to loss of privacy for the individual. The mobile number and the courier details in this example, can be used to trick the user with a fraudulent text, asking him for extra payment in the name of customs or other reasons [3].

Enterprise data is more complex to anonymize as the applications which use this data have defined boundaries on how this data can be stored. Data is stored in multiple tables with relationships retained among the tables. In case of non-relational data storage methods, information is retained on how to derive the original information from the data sets and same data is stored in multiple places, making the challenge more difficult for data security, and also specifically for data privacy aspects. More details on this in Section 4, where the solution is presented.

To understand the challenge for Enterprise data, we must also consider the users who are affected with the mobile devices increasingly becoming the Edge devices [4]. For example, to use a smart phone, one has to configure the phone to an associated email, which was not the case till the mobile device was not an edge device to connect to internet but a normal phone to talk and message. The end user even being un-educated in several cases, easily falls prey to the terms and conditions and the default selections [4] while configuring the phone for first use, especially in less developed nations. Even terms and conditions are not available in a local language, which the end user can understand in several countries. The solutions to preserve privacy especially for Enterprise data, become more critical to be available on an edge device like mobile phone.

3 Existing Practices

Companies protect data collected from its customers, partners and vendors for access using data governance policies in the company. Based on the use case for data consumption, companies keep revising their agreement with the users. Different countries have different laws on what data can be consumed and how and what level of privacy must be provided to the data subject. Enterprise organizations roll out the devices and the applications consuming data from these devices, mostly in phased manner complying to the laws of given regions one by one.

Companies collecting data on edge devices like mobile phone or armbands etc. do have arrangements to ensure data security. Data collected on edge devices is synchronised to central servers in most cases using standard secure communication methods. Privacy of user is managed either by consent or by contract. Consent based management of privacy means the user has to consent to usage of their data to avail of a given service. The other way is via contract, which is primarily used, where the company collecting data from the user, gets a contract signed between the two parties for a continuing relationship. This contract has different options on how the data can be used by the company collecting the data. These contracts can be explicit or implicit, for example, while using a courier service. The user signs the slip paying money to the courier company and this has an implicit contract for the company to use, the data provided by the user. On the other hand, using a service like free photo storage provided by a company, the user must provide consent to terms and conditions, in order to continue usage. These terms and conditions for a consent-based approach can be amended and user can be notified, though contracts are valid for a limited period. In legal terms, both consent based agreement and contract can be binding on both the parties entering into the agreement.

One of the major examples to understand existing practices would be the usage of anonymization technique, Differential privacy, by Google to use the location data from its applications in order to maintain user's data privacy and data's accuracy for utility [5].

Several practices like [6] PEC – privacy aware edge computing, [7] Privacy-Preserving Asynchronous Federated Learning Mechanism for Edge

Network Computing (PAFLM) and many other techniques are being used to deal with privacy preservation of data in edge devices. [8] has more details on the existing practices for privacy preservation of Enterprise data in general.

4 Innovative Solution From Research

In order to preserve privacy of an enterprise relevant data, it is important to understand the nature of Enterprise data set in discussion. New data storage and retrieval methods using no-sql data bases and non-relational data modelling of information is increasing in use, though majority of Enterprise applications still rely on relational datasets. These datasets are in tabular form with relationship between different tables defined via a foreign key. Moreover, table also keeps information unique with a defined primary key to identify a unique record in the table. Details on relational data modelling are not presented in this paper though it is important to understand the basics as the challenge with Enterprise data sets becomes multi-fold due to these practices of relational data modelling.

4.1 Technical Challenge

State of the art methods for data anonymization like Differential privacy, k-anonymity, l-diversity can preserve privacy for standalone data sets in a table or view, though it cannot deal with multiple data sets. For example, if an enterprise application has n tables related via a data model design, which establishes relationship between these tables, Differential privacy method to anonymize data for privacy preservation cannot ensure that the data in the n tables after anonymization, can relate to each other with the foreign keys. Information derived in an application using these n tables and their defined relation may not ensure the benefits like retention of statistical distribution of the derived information. Hence methods like Differential Privacy [9] have limited usage for data anonymization for applications as they mostly deal with views on original data.

4.2 An Implementation

Presented in this section are results of an anonymization experiment performed on data from Kaggle [10]. These data sets have been anonymized to preserve privacy using a combination of steps as presented in the block diagram in Figure 1 on next page.

Table 1 Original data set sample for anonymization

Passenger	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
1	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.25	
2	1	Cumings, Mrs. John Bradley	female	38	1	0	PC 17599	71.2833	C85
3	3	Heikkien, Miss. Laina	female	26	0	0	STON/O2.	7.925	
4	1	Futrelle, Mrs. Jacques Heath	female	35	1	0	113803	53.1	C123
5	3	Allen, Mr. William Henry	male	35	0	0	373450	8.05	
6	3	Moran, Mr. James	male		0	0	330877	8.4583	
7	1	McCarthy, Mr. Timothy J	male	54	0	0	17463	51.8625	E46
8	3	Palsson, Master. Gosta Leon	male	2	3	1	349909	21.075	
9	3	Johnson, Mrs. Oscar W (Elisa	female	27	0	2	347742	11.1333	
10	2	Nasser, Mrs. Nicholas (Adele	female	14	1	0	237736	30.0708	
11	3	Sandstrom, Miss. Marguerite	female	4	1	1	PP 9549	16.7	G6
12	1	Bonnell, Miss. Elizabeth	female	58	0	0	113783	26.55	C103

Table 1 above presents the first 12 records out of thousand records from original data set. This data set has names and age, which are personal information. Additionally, ticket number can be used to identify the person, hence a personally identifiable field. Also, the fare in this table, is not consistent and can be used to identify a passenger with a certain accuracy, hence it can also be a potential Personally Identifiable Information (PII) in this case. Sex of the person if combined with other attributes can be used to reveal an individual as well. Cabin IDs are potential PIIs.

Table 2 below presents the first 12 record from original data set in anonymized form. Names have been replaced with pseudo names. Sex has been shuffled. Age has been added with noise using differential privacy concepts. Ticket Id has been replaced with pseudonymized IDs using cryptographic methods. Fare has been added with noise using differential privacy concepts. Cabin IDs has been shuffled in data randomly.

Table 2 Anonymized data set

PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
914	3	Person_1	female	33	0	0	ADCF1235	7.925	B36
915	3	Person_2	male	45	1	0	ADCF1236	7.225	
916	2	Person_3	female	58	0	0	ADCF1237	59.4	A21
917	3	Person_4	male	25	0	0	ADCF1238	3.1708	
918	3	Person_5	female	26	1	1	ADCF1239	31.6833	
919	3	Person_6	male	13	0	0	ADCF1240	61.3792	
920	3	Person_7	male	28	0	0	ADCF1241	262.375	
921	2	Person_8	male	24	1	1	ADCF1242	14.5	
922	3	Person_9	male	14	0	0	ADCF1243	61.9792	C78
923	3	Person_10	male	22	2	0	ADCF1244	7.225	
921	3	Person_11	female	1	0	0	ADCF1245	30.5	
925	1	Person_12	female	45	0	0	ADCF1246	21.6792	

Using different techniques to each field of data make the implementation of this approach more complex at the same time provides capability to keep data relevance while providing privacy preservation. Any one method used alone would not be sufficient to retain data utility as well as provide anonymization.

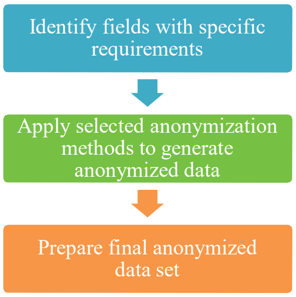
Identify fields with specific requirements
Apply selected anonymization methods to generate anonymized data
Prepare final anonymized data set

Figure 1 Block diagram for the steps performed in proof-of-concept.

4.3 Algorithmic Approach in the Solution

Figure 1 above has the block diagram, which represents the steps performed to achieve the above results. These result look very simple, though they present a completely different approach to solve the problem of privacy preservation with a combination of industry renowned techniques.

This approach has several advantages over other established approaches and most importantly this approach enables any application using one or multiple tables, to work and analyse anonymized data. Some of the major disadvantage is the inability to identify outliers.

4.4 Improving the Approach

Outliers in a dataset expose the data for privacy as sparse attributes combined with intuitive information can lead to revealing of personal information. For example Date of birth combined with post-codes [11]. To correct that problem other step was added as below in Figure 2 to use another method for data anonymization l-diversity on top of the other methods used in previous step.

This then becomes a generic approach to deal with further such use cases in the block diagram which is not specific to the presented proof-of-concept but in general can be used to apply this logic for privacy preservation for any dataset in any device. It would need to be adjusted with appropriate methods of anonymization to suit the data set and the given use case to retain data utility for the needs of the application.

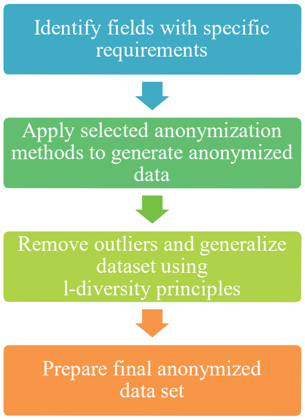
Identify fields with specific requirements
Apply selected anonymization methods to generate anonymized data
Remove outliers and generalize dataset using l-diversity principles
Prepare final anonymized data set

Figure 2 Block diagram for the added steps performed in proof-of-concept.

4.5 Advantages of the Proposed Approach

Anonymized data prepared using this approach has several advantages. The advantages of this approach distinguish the proposed method from other standalone methods of anonymization like Differential Privacy, Pseudonymization, k-anonymity, homomorphic encryption and others. Below bulleted points highlight the importance of such an approach for anonymization of data on edge devices

- Anonymized data retains utility for application to keep running as with original data sets.
- Multiple tables in the dataset can be related using primary keys and foreign keys
- Mobile device or the edge device, where the data resides, may retain copy of the original data with high security controls and no external access, while the anonymized data is used for external communication and provided to other applications which seek data from the edge device.
- Mobile/edge device may get rid of the original dataset altogether, where not necessary to retain only anonymized data for pending operations on the device.
- Anonymization can be automated as a process step while storing the data in the edge device or while sending data for communication in the network.

4.6 Mobile Phone as the Edge Device

In this section we present a mobile phone as the edge device with ample computing power. A smart phone, with data management and several other features built into the operating system like android or iOS. Figure 3 below presents a visual demonstration of how this approach can be used for privacy preservation of data in a mobile device. Four icons represent 4 applications: a privacy app, a fitness app, a calendar app and a social media app.

The privacy app can be the one, which anonymizes the data in any other application on the mobile phone. The social media app with a smiley and the calendar app with a scheduler image are only for display purpose in Figure 3. The flowchart in Figure 3 has steps for privacy preservation of data in the fitness app. As per the flowchart, personal information in the fitness app is first identified. This marks the tables and fields, which need to be anonymized. Then in the next step, the storage locations, where such data is stored is accessed by the privacy app.(all approved applications in the official app store must comply to MSTG guidelines for storing sensitive data [12]). The privacy app creates anonymized data using the algorithm as presented in Figure 2. In the next step, original data may be replaced by anonymized data, with the option in the privacy app, to revert this replacement, if user desires so.

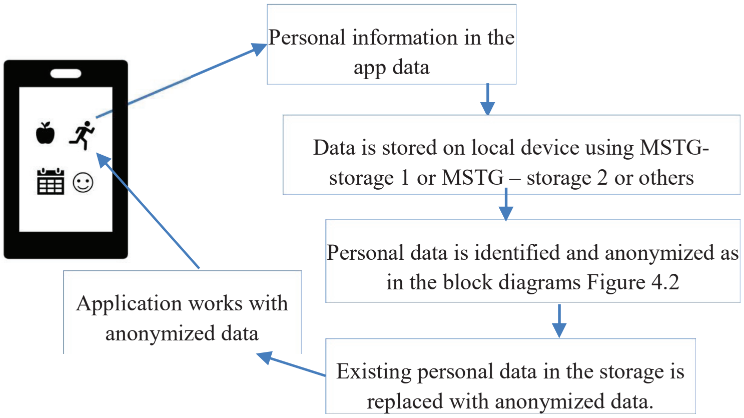
Personal information in the app data
Data is stored on local device using MSTG-storage 1 or MSTG – storage 2 or others
Personal data is identified and anonymized as in the block diagrams Figure 4.2
Existing personal data in the storage is replaced with anonymized data.
Application works with anonymized data

Figure 3 Process flow diagram for privacy preservation of data on a mobile phone.

After this step is performed by the user using the privacy app, the fitness app works with anonymized data. The fitness app stores the anonymized data of the user on cloud storage and hence, the original data of the user, never leaves the device. This protects the privacy of the user and prevents any misuse of such information by any other party including the fitness app company. This also prevents and avoids any risk of the privacy app being hacked for data or any such accidents of data being stolen from the fitness app.

The above example can be extended to any edge devices including IOT devices which work with android or iOS operating system. Such an arrangement can enable communication of data to any outside storage, with anonymized data for user. While the local device can store original data providing all the services to the user and at the same time keeping her privacy fully intact.

4.7 Evaluating Different Approaches for Enterprise Data Anonymization

Existing approaches and the presented approach above are being compared in this section on different attributes which are relevant to privacy preservation of data on edge devices. Important to note is also the fact that currently data security doesn't not cover privacy preservation of data in general. Hence the approached listed below are also a conceptual comparison, if implemented on mobile device. These approaches are used by different companies for selected data only and not for all personal data.

Table 3 below presents a comparison of different approaches in the first column of the table against three major challenges, discussed in this paper. Privacy preservation of data on the edge device, can be ascertained by getting a favourable rating for any approach against these three measures. Individual methods evaluated in the table below are not explained in detail in this paper. More on these methods can be found in the paper reference [8].

Table 3 Comparison of different approaches and the presented appraoch

	Privacy Preservation	Data Utility	Enterprise Data on Edge Device
Pseudonymization	Data subject may be re-identified	Flexibility to retain utility of data	Suitable for enterprise data. Easy to implement
Differential privacy	Most adopted method for anonymization	Utility of data is highly retained	Not suitable for (relational) Enterprise data
Homomorphic encryption	Highly reputed privacy preservation technique	Only with very high computing power	Limited usage due to high requirement of computing power
K-Anonymity/ L-diversity	Reputed technique for privacy preservation	Loss of data utility	Not sufficient for Enterprise data. Cannot be applied incrementally
Presented approach in this paper	Good on privacy preservation	Retains data utility	Easy to implement for Edge devices. Suitable for Enterprise data

5 New Business Opportunities

In the current situation, data anonymized using the most advanced methods is only available for access to companies who generate and collect this data from the data subjects. As such small companies doing research work do not have ‘the data’, which is the most needed element for smart Enterprise applications, including capabilities for Machine Learning and Artificial Intelligence. As the approach proposed in this paper, changes the privacy relevance of data in the edge device itself, it opens up several new business opportunities with the data generated in the edge device.

Data available in public domain for research is the data published by government and other public bodies. Such data is neither granular nor has any dynamic mechanisms to accommodate a request from a researcher. It is plain vanilla available or not available and researcher has to suit his work, based on available data. In many cases researcher has to collect their own data by designing surveys for the end users – their new data subject. Depending upon the reach of the researcher, which is again depending upon association of the researcher, results in data sets, which in many cases do not reflect the actual situation. Large datasets which are possible to have for collaboration among researchers are limited in availability and need separate contracts between researchers and hence cause delays and slow down the speed of having quality research output.

With anonymization of data available in the edge devices itself, the user (data subject) who generates the data has actual freedom to protect his privacy. Rather than selecting terms of engagement and receiving or not receiving a service from a company by signing "Terms and Conditions" for use of a device, the user would be in-charge of the data generated on her edge device. The user can choose levels of privacy preservation and still provide data from an edge device to outside network players. The process of collecting data and distributing data for research would be simpler.

The user would not be forced to part away with the data with a contract by the device/service provider and leave it up to the service provider or the device provider on how they protect user's privacy. Users can update their privacy settings for anonymization of all data right at the source, the edge device, where this data is being generated.

There can be new service providers, which provide anonymization service for data in the edge devices. Having independent service providers for data anonymization, rather than the forced use of data by the service/device provider, can open up more flexibility in the "Data economy". Personal Data Economy (PDE) and Pay-for-Privacy (PFP) models [13] can be substantiated with such an approach.

Instead of the limited and large players, which currently are into the business of trading data, every individual, who is generating data, can be in a position to have a choice. These companies which currently trade data are increasingly under scrutiny. Laws in Switzerland [14], Europe (GDPR), USA (CCPA) and other countries are getting more and more strict in how the data stored by company can be used.

With the proposed solution in this paper, the user sees a direct benefit in sharing her information. The user can choose to share their privacy protected data with one or more party/researcher/analyst/company. Such an approach can make the society [15] reap the benefits of the new innovations in Artificial

Intelligence and other advanced technology areas rather than making the society a surveillance state, where a handful of players rule the data and use or misuse the same.

6 Critique of the Presented Approach

One-size fits all is not true for the presented approach in this paper. There are shortcomings in the presented approach and limitations on what use cases can be addressed. Data in images and video or in general un-structured data is majority space consuming data on a mobile device for example. Photographs, videos and text in social media applications on the mobile device cannot be covered by the implemented proof-of-concept. For example blurring the image of a person, blurring the car number plate and other personally identifiable information in a surveillance camera recording is one of the ways to protect the privacy of the individual, whose activity is being recorded and stored though not needed to achieve the goal of surveillance. Different levels of access to a video footage taken from a surveillance camera and transmission or copy of data only after privacy preservation, can be some of the steps, which can protect privacy and avoid any misuse of the same.

Apart from the suitability of the proposed approach with images and media data, some of the limitations of such an approach are in dealing with automation of such processes. Implementation of such an approach would need extensive work on the deeper layer of technology development for example operating system. Not only on the technology side, there would be required changes on the legal side as well, for all the user end applications which collect personal information to open up interfaces, so that anonymized data can replace the original data captured by those applications.

Another important aspect in preserving privacy is the loss of utility of data. The data elements which retain uniqueness of a record for example a customer number, if removed from a dataset, may render data useless for certain purposes, where a company wants to analyse specific information about that customer. But despite this information is removed, the company would be able to analyse data for other use cases, like doing a churn analysis, where the user behaviour can be simulated using an anonymized customer number. This would enable the company to identify potential factors related to a customer's behaviour. The company may not be able to trace back the customer, whose data is in question, but can get a good understanding of consumer behaviour and the benefits of such analysis can be passed on as services to all the effected consumers.

All the use cases, possible with the original data, may not be possible with anonymized data. In the experiment presented in this paper with the Kaggle dataset in Section 4, the loss of utility of data for the Kaggle Challenge was minimal. Though the loss of utility of data can be high for use cases, where relational data models are built on top of anonymized data. As discussed in Section 4, such loss of utility would be far less compared to other approaches like Differential privacy, where anonymization of relational data-sets is even impossible in most cases.

A very positive aspect of the current approach is to bring all the applications on an edge device at par with the latest requirements on privacy protection of individuals by different regulatory bodies across the globe. Current approach of applications like Covid tracker, which do not store person names and other personal information of the user outside of the mobile phone, presents a promising picture of the possibilities for privacy preservation. At the same time, it also reflects the awareness among technology developers on the importance of privacy and opening up of technology stack in a way to preserve privacy of the data subject [16].

Though all the apps across the globe do not meet the same standards on privacy, increased awareness indicates a good trend. The proposed approach in this paper can further support the need of development of privacy preserving technology, not only in contact tracing applications but in all applications, which capture data for end users. It would give a feasible option for Enterprise applications to provide this option to the user of the edge device.

7 Conclusions and Future Scope

Data including Enterprise sensitive information, which can be used to conduct business with the individual, are pretty much lying around with companies, which run for profits. Without safeguards built into the edge device itself on detecting, protecting, and preserving personal information, a complete solution to protecting privacy of individuals is impossible.

Paper presents a unique approach for privacy preservation for the purpose of user satisfaction and control on her own data. It highlights the weakness in the existing approaches, which are mainly used by companies for making profits and suggests an approach to empower data subjects. The user would be aware that her data can generate profits for other companies and other companies may have to pay the user a part of the profits they make with the users' data. Privacy preservation at the edge device can open several new business models.

Role of regulatory body in defining standards for privacy preservation is duly noted in this paper and also emphasised with the adoption of privacy preservation by contact tracing apps, which are supported by various governments across the globe.

Further work would be needed to improve on anonymization techniques. Presented algorithm can be enhanced based on requirements of application and the storage guidelines for data on a device. Also, special methods and algorithms need to be developed based on proposed approach to deal with images and video data. How anonymized data can be used in the network and other applications may impose more requirements for this approach to deliver good quality of anonymization and adoption in commercial applications.

References

[1] A. Westin, "Privacy And Freedom," *Wash. Lee Law Rev.*, vol. 25, no. 1, p. 166, Mar. 1968.

[2] T. Takenaka, Y. Yamamoto, K. Fukuda, A. Kimura, and K. Ueda, "Enhancing products and services using smart appliance networks," *CIRP Ann.*, vol. 65, no. 1, pp. 397–400, Jan. 2016, doi: 10.1016/j.cirp.2016.04.062.

[3] A. K. Jain and B. B. Gupta, "A survey of phishing attack techniques, defence mechanisms and open research challenges," *Enterp. Inf. Syst.*, vol. 0, no. 0, pp. 1–39, Mar. 2021, doi: 10.1080/17517575.2021.1896786.

[4] "Automatically Granted Permissions in Android apps |Proceedings of the 17th International Conference on Mining Software Repositories." https://dl.acm.org/doi/abs/10.1145/3379597.3387469 (accessed Nov. 29, 2020).

[5] "Tackling Urban Mobility with Technology," *Google Europe Blog*. https://europe.googleblog.com/2015/11/tackling-urban-mobility-with-technology.html (accessed Dec. 18, 2019).

[6] J. Zhao, R. Mortier, J. Crowcroft, and L. Wang, "Privacy-Preserving Machine Learning Based Data Analytics on Edge Devices," in *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, New York, NY, USA, Dec. 2018, pp. 341–346. doi: 10.1145/3278721.3278778.

[7] X. Lu, Y. Liao, P. Lio, and P. Hui, "Privacy-Preserving Asynchronous Federated Learning Mechanism for Edge Network Computing," *IEEE*

Access, vol. 8, pp. 48970–48981, 2020, doi: 10.1109/ACCESS.2020.2978082.

[8] A. Anant and R. Prasad, "State-of-the-art in Privacy Preservation for Enterprise Data," in *2020 23rd International Symposium on Wireless Personal Multimedia Communications (WPMC)*, Oct. 2020, pp. 1–6. doi: 10.1109/WPMC50192.2020.9309459.

[9] S. L. Garfinkel, J. M. Abowd, and S. Powazek, "Issues Encountered Deploying Differential Privacy," in *Proceedings of the 2018 Workshop on Privacy in the Electronic Society*, New York, NY, USA, Jan. 2018, pp. 133–137. doi: 10.1145/3267323.3268949.

[10] "Titanic - Machine Learning from Disaster." https://kaggle.com/c/titanic (accessed Dec. 30, 2020).

[11] L. Sweeney, A. Abu, and J. Winn, "Identifying Participants in the Personal Genome Project by Name (A Re-identification Experiment)," *ArXiv13047605 Cs*, Apr. 2013, Accessed: Apr. 07, 2021. [Online]. Available: http://arxiv.org/abs/1304.7605

[12] "Introduction to the Mobile Security Testing Guide." https://mobile-security.gitbook.io/mobile-security-testing-guide/overview/0x03-overview (accessed Apr. 08, 2021).

[13] S.-A. Elvy, "Paying for Privacy and the Personal Data Economy," *Columbia Law Rev.*, vol. 117, p. 1369, 2017.

[14] "CC 235.1 Federal Act of 19 June 1992 on Data Protection (FADP)." https://www.admin.ch/opc/en/classified-compilation/19920153/index.html (accessed Dec. 30, 2020).

[15] A. Anant and R. Prasad, "Data Privacy Technology for Society," *River Publ.*, vol. Series in Information Science and Technology, p. 16, doi: 10.13052/rp-9788770222174.

[16] H. Wen, Q. Zhao, Z. Lin, D. Xuan, and N. Shroff, "A Study of the Privacy of COVID-19 Contact Tracing Apps," in *Security and Privacy in Communication Networks*, Cham, 2020, pp. 297–317. doi: 10.1007/978-3-030-63086-7_17.

Biographies

Aaloka Anant is a researcher at CTIF Global Capsule, Aarhus University, Denmark since October 2019. He attained his Post Graduate degree in Enterprise Management from Indian Institute of Management Bangalore and B.Sc in Electronics and Communication Engineering from BIT Sindri in India. Aaloka has held leadership and senior positions in SAP and Honeywell since 2004 and also worked with start-ups like Idea Device Technologies, MovidDLX, NGeneR and co-founded a non-profit organization Anant Prayas.

He teaches students for Masters program in Data Science subjects for over 3 years as guest Lecturer in National College of Ireland, Dublin. He also has teaching experience at Furtwangen University, Germany. He is actively pursuing research on the topic of Privacy Preservation. His work focusses on new approaches for privacy preservation of Enterprise data and missing technology and structural framework for achieving end-to-end data privacy.

Ramjee Prasad, Life Fellow IEEE, Fellow IET, IETE, and WWRF, is a Professor of Future Technologies for Business Ecosystem Innovation (FT4BI) in the Department of Business Development and Technology, Aarhus University, Herning, Denmark. He is the Founder President of the CTIF Global Capsule (CGC) and Founder Chairman of the Global ICT Standardisation Forum for India.

He has been honoured by the University of Rome “Tor Vergata”, Italy as a Distinguished Professor of the Department of Clinical Sciences and Translational Medicine. He is Honorary Professor of University of Cape Town, South Africa, and University of KwaZulu-Natal, South Africa. He has received Ridderkorset af Dannebrogordenen (Knight of the Dannebrog).

He has received several international awards such as: IEEE Communications Society Wireless Communications Technical Committee Recognition Award.

He has published more than 50 books, 1000 plus journal and conference publications, more than 15 patents, over 145 PhD Graduates. Several of his students are today worldwide telecommunication leaders themselves.