University of Ibadan
Journal of Science and Logics in ICT Research

Data Over-Collection Reduction Model Using Security Level Analysis and Data Access Status Technique

[1]Akomolafe Patrick Oladeji and [2]Aasa Solomon Oluwadamilare
Department of Computer Science, University of Ibadan,
[1]akomspatrick@yahoo.com, [2]o.aasa@yahoo.com

Abstract
In smart Environments, different citizens tend to store data in electronic devices in order to make everything intelligent and easy. A smartphone for example is the most widely used electronic device and it is the hinge of all smart environs which is still not capable of handling users' sensitive data, and they (users) face the fear of data and privacy leakage due to data over-collection. We develop a data over-collection reduction model to reduce data over collection in smart phones while still within permission scope. We take emphasis to the current state of data over-collection and as well some of the most frequent data over-collected cases. We present a framework, which is an alternative approach to that of Li, *et. al.* [1] (mobile-cloud framework) to reduce data over-collection. Using our framework, the permission to data from different applications was limited to the data required by the application to perform its function. Also, the security risk posed by the application due to the probability of over collecting data reduced compared to that of Li, *et. al.* [1].

Keywords: Data over collection, Smart phones, Data privacy, Data security

1.0 INTRODUCTION

Many cities such as Barcelona, Amsterdam, Stockholm and Southampton amongst others around the world risk becoming barely livable within a few years as their infrastructures are stretched to their limits in terms of scalability, environment, and security while they adapt to support population growth (9.7 billion in 2050 according to the UN-Habitat United Nations [UN] program — http://unhabitat.org). Without innovative solutions economies will be under increased pressure; energy consumption will increase exponentially; the environment will be challenged; healthcare and education systems will demand new approaches; public safety will be further challenged; and the potential for future cyber-attacks against cities is high.

There are components that underpin most smart environ models: government, economy, mobility, environment, living, and people. All these components help a smart environ to achieve multiple benefits that include and not limited to: Energy Efficiency, Economic Development, Mobility Management, Information Management and Data Retrieval. Reflecting on Barcelona, March and Ribera-Fumaz [2] argue therefore that it is imperative to "put citizens back at the center of urban debate".

Various suggestions have been made and explored to integrate a wider group of citizens into smart environ design and policies, – for instance – through citizen participation [3], crowd sourcing [4] and citizen centered approaches [5]. Others have argued more generally for a stronger protection of the privacy of citizens living, working, shopping or travelling in a smart environ. Li, *et. al.* [1] identified over-collection of data as a severe security risk, especially when it comes to the sensitive data that people hold on their smart phones' Internet via everywhere Wi-Fi, take online courses, pay their bill online, sign a contract online, and receive medical treatment by tele-health. The smartphone not only stores users' data, but also generates data. These data may consist of users'

Akomolafe, P. O. and Aasa S. O. (2021). Data Over-Collection Reduction Model Using Security Level Analysis and Data Access Status Technique. *University of Ibadan Journal of Science and Logics in ICT Research (UIJSLICTR)*, Vol. 6 No. 2, pp. 99 – 109.

accounting numbers and passwords, emails and house addresses, photos, and other kinds of sensitive information. Applications within the smart phones communicate with their users as though they know what problems they would like to solve, what food they would like to eat and sometimes what games they would love to play. This makes the users go fond of the devices in a way that they give total control and access to data whenever the applications request for them.

Ideally, applications within smartphones should have access to data only required to perform their basic functions. However, that is not the case. Instead, they tend to over reach hereby having more access to what is required. This poses a security threat what has been classified as sensitive information to the data owner.

This study developed a data over-collection reduction model using a data class security risk level determination model in combination with class data usability status determination technique which is capable of eliminating the need for multiple permission and security risk computation. The objective here is to design an improved model that would reduce the consumption of computing resources during access control to requested data. Then also measure and evaluate the result of the research with comparable research work in the same problem domain

2.0 LITERATURE REVIEW

A number of researches have been carried out on data security and privacy issues posed to users. One in particular is that of Gaved, *et, al.* [6], on privacy and security in role of software. How smart environment have come up to counter act the ills of the industrial environment that generate a many bad side effects. Smart environment have been known to be environment friendly, better organized, have better mobility, and a more competent economy. To meet these necessary objectives, the role of software can make the smart environ more integrated and functional as a whole. But the use of smart software, as we call it, can pose problems pertaining to the security and integrity of the smart environ's data and privacy.

Therefore, there's need to comprehend and assess the concerns for safety when developing smart software before they can be used in full scale. This paper concentrated on these issues. This paper talked on exchange of application data, problems related to hacking (with regards to collected data), effects of hacking, access to information in data centers and increase in personal data thefts.

Gaved, *et, al.* [6] research however, suggested a model for the devices poised to use these applications. This model can be used to increase the security of data collected by software and applications. Some of the solutions were authentication of data sets where the smart software fetches the datasets. But unless proper authentication procedure is set up for the software as well as the network to understand when the datasets are valid and trusted, it would be susceptible to disguised misinformation from as well as to the network from untrusted sources leading to a system failure for the device. Because the datasets sent is sensitive and involves the devices' functioning, unauthenticated datasets can breach overall functioning of the system and produce malfunctioning when the device is controlled by the information that is not from the network.

The second solution was Data scrutiny layer. Here, the devices in this model do not have express access to the server. The information sent by the devices to the network of the smart environ will first be sent to the data scrutiny layer. Now in this data scrutiny layer the information that is being sent from the devices are checked for consistencies and inconsistencies amongst others– the reason for which the data is to be sent to the network, if there is a reason, is it enough to provide access to data sending and also the amount of data to be sent (because the smart software will only sent a limited amount of data at a time), source of the data, previous problems that data from the same source have caused.

Li, *et. al.* [1] discussed how in a smart environ, all kinds of users' data are stored in different electronic devices in order to make everything intelligent. It mentioned how a smart phone was and is still the most widely used electronic device and as well is the pivot of all smart systems. His study brought about a mobile-cloud framework, which is an active approach to eradicate the data over-collection. Now by putting all these users' data into a cloud, the security of users' data can be greatly improved. Extensive experiments were

done and the experimental results demonstrated the effectiveness of their approach.

In this research we were made to understand that the sensitivity and security of data is the most difficult challenge to smart environ. Different users provide sensitive information using various medium. Now for a smart environ to be as effective as possible, it needs to be able to allow and accept data from these various kinds of systems.

Liesbet, *et. al.* [7] researched on how this ICT-driven society would challenge the privacy of users living within that geographic environment. A framework was constructed to theorize if and how smart environ technologies and metropolitan big data would produce privacy concerns among the people in these environment (such as residents, employees, guests, and otherwise). The framework is built on the basis of two recurring dimensions in research about people's concerns about privacy: The first one represents that people recognize particular data as more personal and sensitive than others, while the other dimension represents that people's privacy concerns differ according to the purpose for which data is collected, with the contrast between service and surveillance purposes most paramount.

3.0 METHODS/METHODOLOGY

The methodology comprises of three basic elements: application, roles and permissions where the application is the subject that needs to interact with a data object, roles defines the application's level of importance and permission states what can and cannot be viewed by a particular application. The first sub module (Data Security Level Determination) is responsible for the determination of the data security level an application is allowed to access and hence is the first module the application interacts with while the second sub module is responsible for the determination of the data class inside the data security level an application is allowed to access. This sub module assigns particular class groupings to the collection of data on the same security level. This is to ensure that an application requesting for access to a particular data to perform its primary function does not get access to all the data in the same security level but instead just the exact data it requires to perform its function

This solution is comprised of mainly two entities which interect which each other in a particular way to share information. This entities are:

1. The Smartphone, and
2. The Cloud.

Other sub components as depicted in Figure 1 are briefly discussed below:

1. **Smartphone:** this part of the model shows the different applications that will likely generate or use data belonging to an individual in the smart environ. It also holds the access control module which determines security risk level of data and the different status of the individual data in the security risk class.
2. **Access control module:** This module comprises of two other sub-modules which are: data security level determinator and the data status determinator. Each of this module plays an important part in the access control activity.
3. **Data security level determinator module:** This module assigns security risk level to data that is to be accessed by a smartphone in the smart environ.
4. **Data status determinator module:** This module handles the assignment of access status to each data in a particular data security risk class.
5. **App request router / Data router module:** This module is responsible for both the handling of request from a user app and the router of data from the storage service to the requesting app.
6. **Data security level and status list module:** This module handles the storage of the data security risk level data and the data access status for data of the same class and data of the same status in the same data security risk class.
7. **Encryption and Decryption module:** This module handles the encryption and decryption of data before entering the storage service and when leaving the storage service.
8. **Storage service module:** This module is responsible for holding data that is to be collected by a user app.

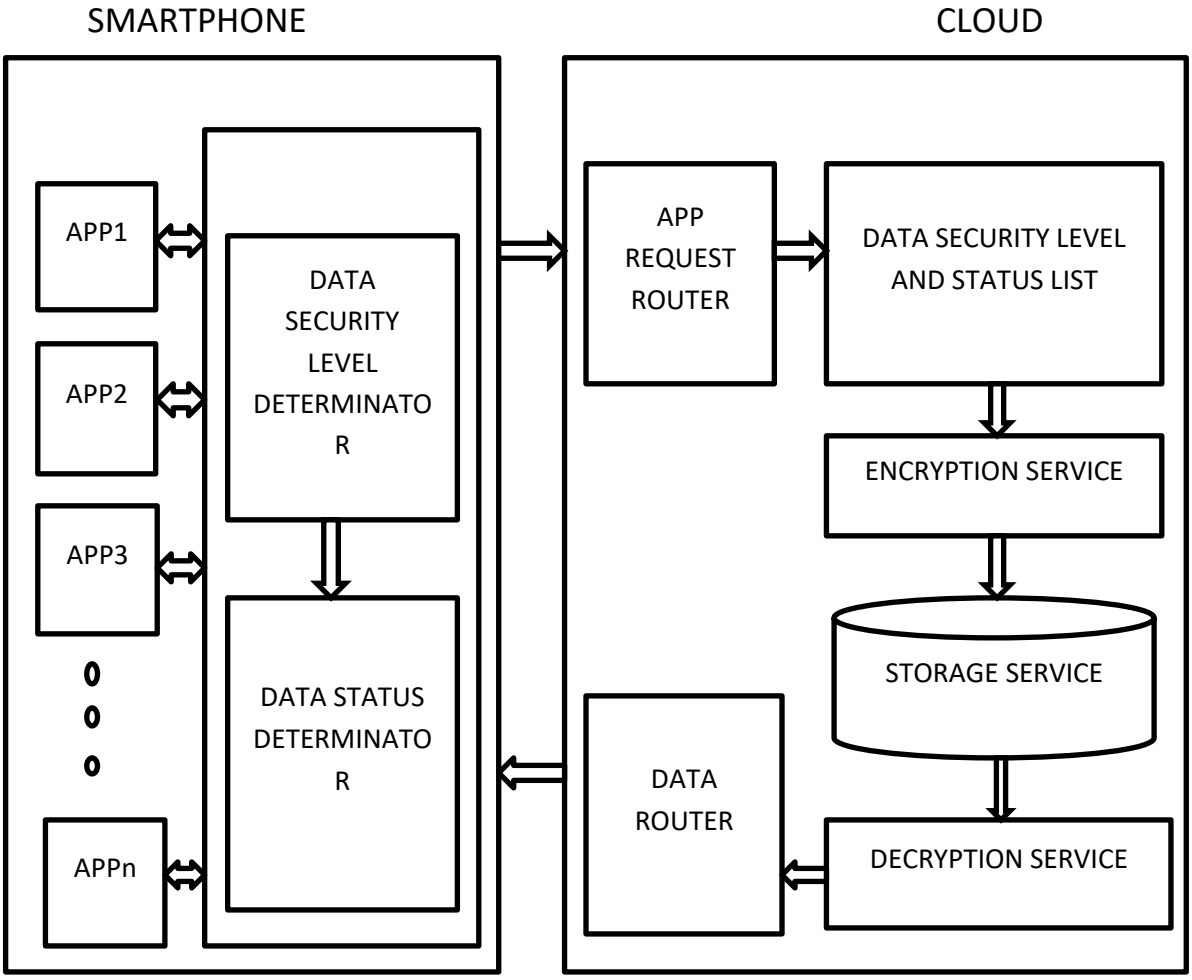

Figure 1: The service recommendation model

3.1 Risk Model for Data Over-Collection

Data over-collection is a kind of a potential risk, which is the direct result of security violation probability and the breakdown of security protocol. As a result, we model the security risk (SR) of an application *i* towards the data *d* as [1]:

$$SR_i^d = \frac{SL^d}{CL^d} \times Pro_i^d \quad (1)$$

Where SR_i^d means the security risk of the application *i* over-collects the data *d*. Meanwhile, SL^d is the security level of the data *d* and CL^d is the class of the data in the security level and Pro_i^d is the probability of the app *i* using the data *d* to do some security related damage to the owner of the data, which can be formulated as:

Note – N/M is amount of over-collected data

$$Pro_i^d = 1 - e^{-\gamma * N_i^d / M_i} \quad (2)$$

γ Is the security risk coefficient of the behavior of the application *i* over-collecting the data *d*, which can be adjusted by different applications and data, but fixed on a single scenario

Based on the equations above, we can use the amount of data over-collected (N/M) to formulate the security risk of app *i* towards data *d* as:

$$SR_i^d = \frac{SL^d}{CL^d} \times (1 - e^{-\gamma * N_i^d / M_i}) \quad (3)$$

We formulate the security risk of app *i* towards a user (U) as:

$$SR_i^U = \sum_{d=0}^{m} \frac{SL^d}{CL^d} * (1 - e^{-\gamma * N_i^d / M_i}) \quad (4)$$

Finally, the security risk of a smartphone belonging to a user U on which A amount of applications have been installed can be formulated as:

$$SR_i^U = \sum_{i=0}^{A} \quad \sum_{d=0}^{m} \frac{SL^d}{CL^d} * (1 - e^{-\gamma * N_i^d / M_i}) \text{ [1]}$$

After establishing necessary models for permission and security risks, the next step is to model the way applications will send data to the cloud for safe storage and also for allowing applications access such data.

3.2 CLOUD STORAGE FOR APPLICATION DATA

After establishing necessary models for permission and security risks, the next step is to model the way applications will send data to the cloud for safe storage and also for allowing applications access such data. The algorithm for modeling how data is sent to cloud is given in Figure 2.

3.3 THE USER DATA CLOUD

The user data cloud contains the stored data of the user using the cloud service. This part of the solution is comprised of five (5) main components, namely:

1. App Request Router,
2. Data Security Level and Status List,
3. Cryptography module,
4. Storage Service, and
5. Data Router.

Input: appID, userPassID, userData, DataClass.
1: Determine the request type made by appID,
return T;
2: **if** (T == hardware) **then**
3: Assign permission to the app using the appID;
4: **else**
5: Send request to Access Control Service including appID, userPassID,userData and DataClass;
Access Control Service judges whether this app has permission by accessing the app's request parameters and accessControlList,
return P;
7: **if** (P == true) **then**
8: Encryption Service encrypts data;
9: Store data into Cloud Storage with label appID,userPassID, userData and userDataClass;
10: **else**
11: return;
12: **end if**
13: **end if**

Figure 2: Algorithm for modeling data upload to the cloud [1]

Input: appID, userPassID, userDataClass preview information of requesting data PD. Output: concrete content of requesting data D

```
1: Send request to cloud request router including appID, userPassID, userDataClass and PD;
2: the request router interprets the request and sends the request to the
   security level and status list for confirmation, get the result P;
3:        if (P == true) then
4:       request status list sends request to cryptography service for encryption;
5:       cryptography service (after encryption) sends request with
         appID, userPassID, userDataClass and PD to Storage Service;
6:       Storage Service finds the encrypted data by userID, userDataClass and PD;
7:       Storage Service sends data with appID and userID to Decryption Service;
8:       Decryption Service checks the permission authorization again by matching
         appID with data and return P1;
9:       if (P1 == true) then
10:            Decryption Service decrypts data to data router D;
11:            return D;
12:      else
13:            return none;
14:     end if
15: else
16:     return
```

Figure 3: Algorithm for accessing user data in the cloud

3.4 TESTING THE ALGORITHM

The developed algorithm was implemented on an amazon cloud service server, namely an EC2 server. The processes was divided into two part,

- The client, which can be a device or application requesting for data in order to complete its task
- The server that handles requests and evaluates each request based on the algorithm.

3.5 THE CLIENT

The simulator is designed using the Oracle Java programming language and the Netbeans integrated development environment.

The client was designed to simulate a mobile device connection environment. The client tries to establish a connection between the server nodes and itself. This connection is the first part of the solution. The idea is to get the ID of the application requesting data. This ID is used in conjunction with the unique application ID to identify a requesting application that is requesting for data. The server runs a security parameter evaluation from a calculated list.

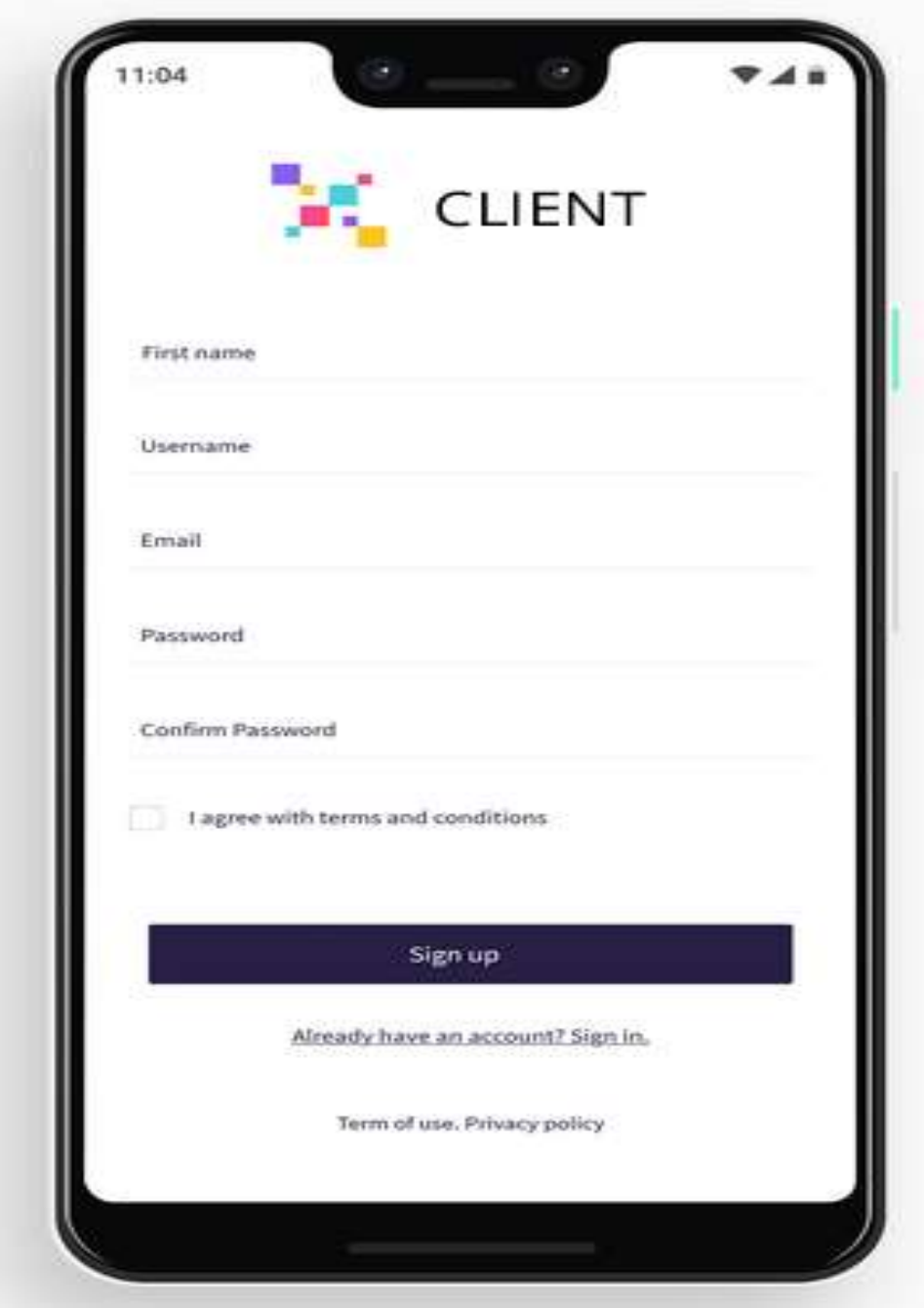

Figure 4: Client View

3.6 THE EXPERIMENT AND THE RESULTS

To simulate the data over-collection behaviors in smart environ environment, we use four real smartphones and one simulative cloud to build a simple mobile-cloud environment. Then we evaluate the feasibility and the performance of our approach through extensive experiments.

First, we set two scenarios: in original environment and in mobile-cloud framework environment. Then we choose some typical apps from each of the nodes or devices under consideration to score their security threat level by location, photos, contacts, username and password in the two scenarios. Finally, to simulate the prototype of Mobile-Cloud framework, we use one computer as the cloud and four smartphones as the experimental objects

API SERVER FILE STRUCTURE

The folders & files structures were used to write the API SERVER.

API:

├——— config/

├————— database.php – file used for connecting to the database.

├——— objects/

├————— user.php – contains properties and methods for “user” database queries.

├——— User/

├————— signup.php – file that will accept user data to be saved to the DB.

├————— login.php – file that will accept username & password and validate

Database & Users Table:

The SQL codes used to create the User table.

```
1. CREATE TABLE `users` (
2. `id` int(11) NOT NULL PRIMARY KEY AUTO_INCREMENT,
3. `username` varchar(255) NOT NULL,
4. `password` varchar(255) NOT NULL,
5. `created` timestamp NOT NULL DEFAULT CURRENT_TIMESTAMP ON UPDATE
   CURRENT_TIMESTAMP
6. )
```

Database Connectivity:

The php class that handle database connection.

```
1. <?php
2. class Database{
3.
4. // specify your own database credentials
5. private $host = "localhost";
6. private $db_name = "PHPLearning";
7. private $username = "root";
8. private $password = "";
9. public $conn;
10. ……./continue yourself/
```

A back-up function in each smartphone was used to extract and transmit photos, contacts, calendar, notes, mail, and other data into the simulated cloud; meanwhile, we delete all these data in smartphones. Due to lack of standard and universal benchmark about application over-collecting data in mobile, we formulate a scoring system to assess security risks of apps based on the model we developed in chapter three of this work. Aiming at presenting various data over-collection behavior, we list ID, location, photo, contacts, username & password into our evaluation mechanism and set various security levels and status for them as listed in Table 1.

The level of security is determined by the sensitivity of the data to be accessed. These are divided into Security Level 1, 2 and 3.

They are also divided into Class Level 1, 2 and 3 (Class Level means the class of the data within that security level)

For instance: (3, (1, 2, 3)) means Security Level 3 Class Level 1, Class Level 2, Class Level 3

The solution was evaluated using the same amount of devices and the same amount of mobile apps. The mobile apps where evaluated based on their security risks with their model and their security risks with our model. The evaluation result using four mobile devices is shown in the Table 2.

We take the average of the result of model to arrive at a figure for the model.

(22.35 + 20.23 + 26.87 + 21.56) / 4 = 22.75

where 22.75 is the percentage of data that was over collected from each device using the model.

Figure 5 shows a graphical representation of security risk for data over collected without the model for each device compared to an average of data over collected while using the model

Table 1: SECURITY EVALUATION

SECURITY	ID	LOCATION	PHOTO	CONTACT	U & P
LEVEL & STATUS	(3,(1,2,3))	(3,(1,2,3))	(2,(1,2,3))	(1,(1,2,3))	(2,(1,2,3))

Table 2: DEVICE EVALUATION

MOBILE DEVICE	WITHOUT MODEL	WITH MODEL
DEVICE A	41.50	< 22.35
DEVICE B	38.05	< 20.23
DEVICE C	38.91	< 26.87
DEVICE D	41.51	< 21.56

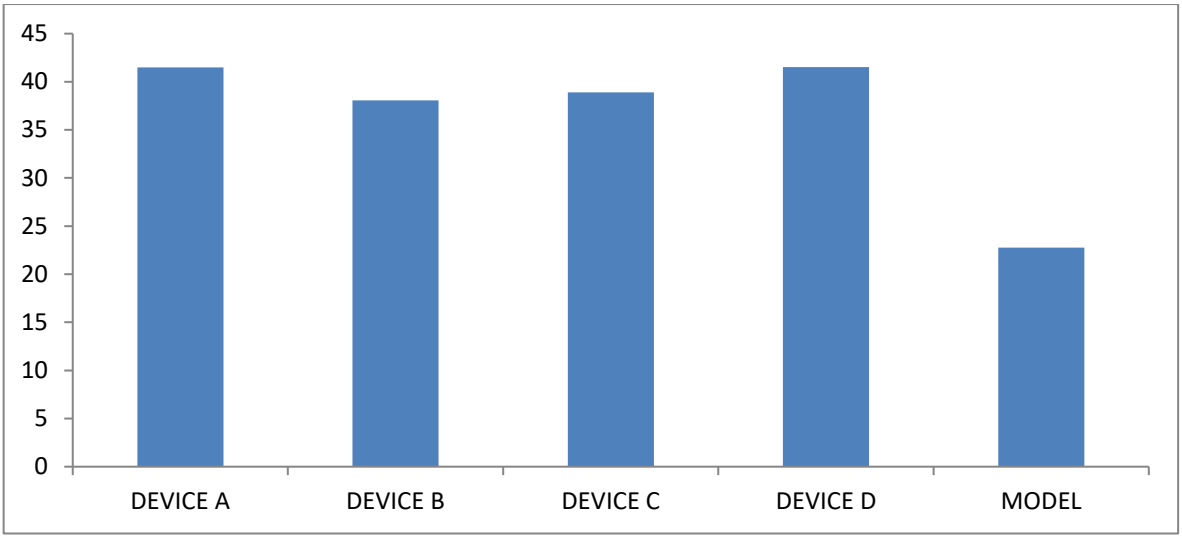

Figure 5: Graphical representation of security risk for data over collected without the model for each device compared to an average of data over collected while using the model

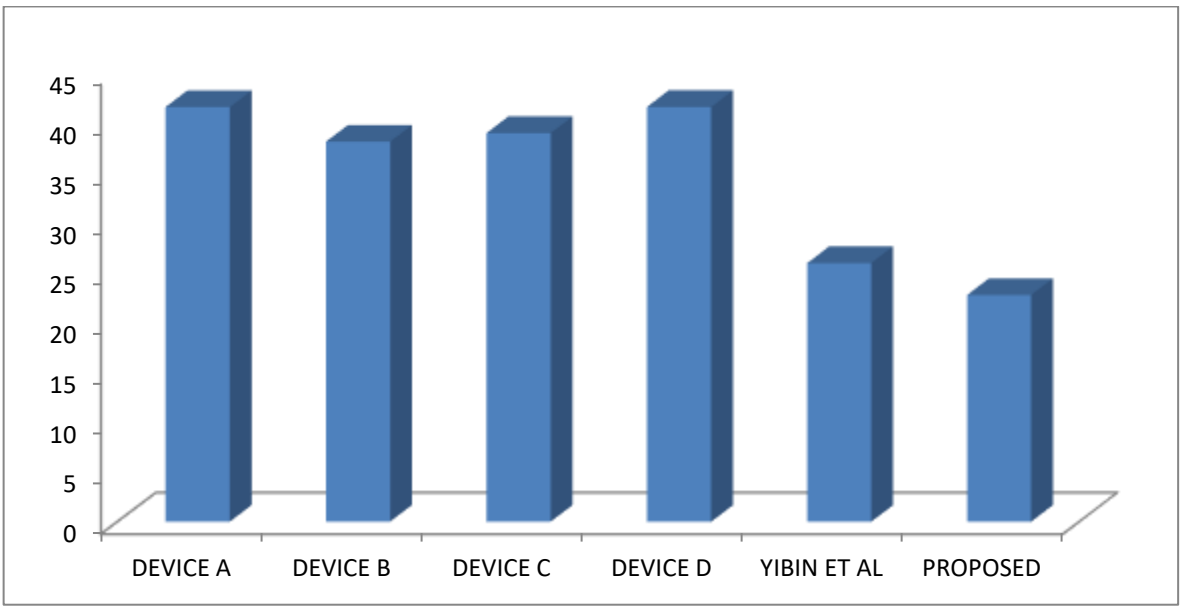

Figure 6: A graphical representation in comparison of security risk for each device, Yibin's result and proposed model: (in percentage %)

4.0 CONCLUSION

Data privacy and what to protect from unwanted access is a global issue. Liesbet, *et. al.* [7] researched on how our ICT-driven society (smart phones inclusive) would challenge the privacy of users living within the geographic environment where they are mostly used. Individuals now use DIY (Do it yourself) to set up their environments to match or mimic smart environments which has in turn increased the amount of security risks posed by over collection of data. They [7] created a framework to theorize if and how the use of technologies in smart devices would produce privacy concerns among the people living in these cities. Models have been targeted to increase security protecting the mobile devices and applications within the mobile devices from external threats, however, studies have shown that in recent times internal threats have become high risk as well.

These have driven the research towards reducing the security threats posed by applications within mobile devices and the want to over collect data from storage areas where they have been provided access.

This security parameter is defined using the permission model defined in chapter three which was subsequently implemented. This work does not in any way disregard previous research work in data reduction with the use of applications, however, majority of the work in this area was focused on use of encryption of the data (still in access control). In this case we took into consideration the individual classification of the individual data. This provided us the opportunity to leverage on the previous model to reclassify the data that is being requested.

The permission model can be implemented on other types of data different from what was used for this work. The data router ensure that the information is sent to the right data, however, the data router can be provided with an encryption model whereby, when the data is being sent out to the requesting application, it does not get tampered with before it gets to its destination address. This would help improve the integrity of the data at destination.

REFERENCES

[1] Li, D., Cao, J., & Yao, Y. (2015). Big data in smart cities. Science China Information Sciences, 58(10), 1–12. https://doi.org/10.1007/s11432-015-5396-5

[2] March, H., & Ribera-Fumaz, R. (2016). Smart contradictions: The politics of making Barcelona a Self-sufficient environ. European Urban and Regional Studies. https://doi.org/10.1177/0969776414554488

[3] Johannessen, M. R., & Berntzen, L. (2016). Smart Cities Through Implicit Participation: Using Gamification to Generate Citizen Input for Public Transport Planning. Innovation and the Public Sector. https://doi.org/10.3233/978-1-61499-670-5-23

[4] Schuurman, D., Baccarne, B., De Marez, L., & Mechant, P. (2012). Smart ideas for smart cities: Investigating crowdsourcing for generating and selecting ideas for ICT innovation in a environ context. Journal of Theoretical and Applied Electronic Commerce Research. https://doi.org/10.4067/S0718-18762012000300006

[5] Gaved, M., Jones, A., Kukulska-Hulme, A., & Scanlon, E. (2012). A Citizen-Centred Approach to Education in the Smart Environ: Incidental Language Learning for Supporting the Inclusion of Recent Migrants. International Journal of Digital Literacy and Digital Competence. https://doi.org/http://dx.doi.org/10.4018/jdldc.2012100104

[6] Sen et al (2013) Ethics of Public Use of AI and Big Data: The Case of Amsterdam's Crowdedness Project (Big Data and artificial intelligence) https://doi.org/10.29297/orbit.v2i1.101

[7] Liesbet van Zoonen, Mehdi Sookhak; Helen Tang; F. Richard Yu(2016). Privacy concerns in smart cities. https://doi.org/10.1016/j.giq.2016.06.004

[8] Khan, Z., Anjum, A., & Kiani, S. L. (2013). Cloud based big data analytics for smart future cities. In Proceedings - 2013 IEEE/ACM 6th International Conference on Utility and Cloud Computing, UCC 2013 (pp. 381–386). https://doi.org/10.1109/UCC.2013.77

[9] Khatoun, R., & Zeadally, S. (2017). Cybersecurity and privacy solutions in smart cities. IEEE Communications Magazine, 55(3), 51–59. https://doi.org/10.1109/MCOM.2017.1600297CM

[10] Kitchin, R., Lauriault, T. P., & McArdle, G. (2017). Data and the environ. In Data and the Environ. https://doi.org/10.4324/9781315407388

[11] March, H., & Ribera-Fumaz, R. (2016). Smart contradictions: The politics of making Barcelona a Self-sufficient environ. European Urban and Regional Studies. https://doi.org/10.1177/0969776414554488

[12] Martinez-Balleste, A., Perez-Martinez, P., & Solanas, A. (2013). The pursuit of citizens' privacy: A privacy-aware smart environ is possible. IEEE Communications Magazine. https://doi.org/10.1109/MCOM.2013.6525606

[13] O., S., T., P., & F., K. (2014). Smart cities as corporate storytelling. Environ. https://doi.org/10.1080/13604813.2014.906716

[14] Rebollo-Monedero, D., Bartoli, A., Hernández-Serrano, J., Forné, J., & Soriano, M. (2014). Reconciling privacy and efficient utility management in smart cities. Transactions on Emerging Telecommunications Technologies. https://doi.org/10.1002/ett.2708

[15] Schuurman, D., Baccarne, B., De Marez, L., & Mechant, P. (2012). Smart ideas for smart cities: Investigating crowdsourcing for generating and selecting ideas for ICT innovation in a environ context. Journal of Theoretical and Applied Electronic Commerce Research. https://doi.org/10.4067/S0718-18762012000300006