Subscription Agency
Exploring and Distributing Knowledge Globally

Recent Advances in Big Data
Volume 1 Issue 1

Proposing New System for Securing Data Sharing in the Mobile Applications

Mohammad Hesamzadeh
Khayyam Institute of Higher Education, Mashhad, Iran
Corresponding Author
Email Id: hesamzadehm2020@gmail.com

ABSTRACT

A wide and increasing range of different technologies, devices, platforms, applications and services are being used every day by home users. In parallel, home users are also installing and using an enormous number of apps which collect and share a large amount of data. Users are also often unaware of what information apps collect about them, which is really valuable and sensitive for them. Therefore, users are becoming increasingly concerned about their personal information that is stored in these apps. While most mobile operating systems such as Android and iOS provide some privacy safeguards for users, it is unrealistic to manage and control a large volume of data. Therefore, this approach could impose an undue burden on the users. Accordingly, there is a need for an approach to predict many of a user's mobile app privacy preferences. A major contribution of this work is to assign users to the privacy profiles that most closely capture their privacy preferences. Applying privacy profiles as default settings for initial interfaces could significantly reduce the burden and frustration of the user. The result shows that possible to predict many of a user's mobile app privacy preferences by asking the user a small number of questions.

Keywords: *Data Privacy, Mobile applications, Security.*

INTRODUCTION

With the rapid growth of the devices, activities, services and information, the enormous amount of private and personal information that is stored has also increased. Therefore, users are becoming increasingly concerned about their personal information, how it is used, by whom and where it is stored [1]. For instance, a Consumer Report found that 92% of British and U.S.

Internet users are concerned about their privacy online [2]. When users became aware of online privacy issues, they were asked what made them most worried about their online privacy: 45% of British Internet users stated that it is personal information being shared between companies [3]. In addition, 89% of users avoided these companies because they believed that companies do not protect their privacy. It was also found that 76% of Internet users limited their online activity in the last 12 months due to these concerns [4]. This evidence indicates that users are sufficiently worried about their online privacy.

Due to the concerns of the users about privacy protection, most mobile operating systems such as Android and iOS provide some privacy safeguards for users [4]. Despite these provisions, there are several usability issues related to the functionality and interface. For instance, Kelley et al. found that users struggle to understand the permissions in Android due to the lack of usability [4]. Therefore, the Federal Trade Commission suggested privacy controls need more improvement to protect users' privacy [3].

A noticed focus has been given to the development of policies, procedures and tools that aid an end-user in managing and

understanding their privacy-related information. However, these approaches assume that users can correctly configure all resulting settings and they have uniform privacy requirements. In reality, users do have different privacy concerns and requirements as they have heterogeneous privacy attitudes and expectations [5]. For example, some users consider personal information such as age, address and gender in their profile on a social network being more sensitive than others [6]. Furthermore, in practice it is unrealistic to assume homogeneous privacy requirements across a whole population [7].

Accordingly, there is a need for an approach that considers individual requirements in a centralized and usable manner to meet users' needs. This paper shows that users are diverse and it is possible to divide users into small groups according to the users' privacy preferences. Applying privacy profiles as default settings for initial interfaces could significantly reduce the burden and frustration of the user. The result shows the optimal number of clusters based on the k-means method is four clusters.

This paper is organized into five sections. Section 2 presents an analysis of background literature. Section 3 shows how the data collected. Section 4 described different methods for determining the optimal number of clusters. The conclusions and future work are presented in Section 5.

BACKGROUND LITERATURE

This section provides an overview of current privacy solutions. In recent years, many studies and researches have been published in many areas in the privacy field such as mobile applications, web application, and social networks in order to protect users' privacy because privacy exists wherever personal information or sensitive information is disclosed. However, this section merely focuses on a mobile platform due to related to the proposed system.

Information Flow Analyzers

Many tools have been developed in recent years that aim to analyze and to detect personal information on mobile platforms. These tools analyze mobile apps regarding potential privacy breaches before they leave the system via untrusted apps. Some of the more prominent examples are, Taintdroid (Enck *et al*, 2014), AppIntent [11], Little BrothersWatching You [12], and PiOS [13].

Taintdroid was designed to detect sensitive data when it leaves the system via untrusted applications (Enck *et al*, 2014). It was designed based on a dynamic approach which is executed whilst a program is in operation. The system can track the flow of data through four levels: variable, method, message, and file. Although TaintDroid detects the sensitive data, the system assumes that users can correctly configure all the resulting settings. Therefore, this approach could impose an undue burden on the users. In addition, they do not examine the usability related to the interface displayed to users.

In comparison, Balebako *et al* [12] presented a solution that focuses on the user's awareness of privacy issues. The solution improves the user's understanding of potential privacy leakages.

It is built based upon the TaintDroid platform and helps users to know the frequency and destination of data being shared by an application. It also provides many user interfaces which can help to inform users about which privacy sensitive information leaves the phone. However, they do not provide users with control over their personal information to allow them to specify which type of information they prefer not to leave the mobile.

Another tool that aims to analyse programs for possible leaks of sensitive information from a mobile device to the third party is PiOS. It detected privacy leaks related to device ID, location and phone number. Moreover, PiOS considered the address book, browser history, and photos. PiOS uses static analysis to detect data flows. They have analysed more than 1,400 iPhone apps and they found that a majority of apps leak the device ID, which can provide detailed information about the habits of a user. However, PSiOS does not provide users with fine grain control over their personal information.

Finer Grain Privacy Controls

A number of research prototypes have also offered used fine grain controls in order to prevent potential privacy leakages. For example, AppFence [14], TISSA [15], AntMonitor [16] and ProtectMyPrivacy [4]. There are several techniques were used to enhance information flow control for mobile. AppFence uses replacing information approach in order to protect sensitive data (Hornyack et al, 2011). AppFence provides users with two privacy controls to protect sensitive resources: shadowing and blocking. Sometimes users do not want to provide application access to sensitive data. Therefore, AppFence sends shadow data instead of the actual data. For example, when an application requires access to user's contacts, AppFence may provide application shadow data that contains no contact entries, contains only those genuine entries not considered sensitive by the user, or that contains shadow entries that are entirely fictional. The second approach for protecting sensitive data is blocking sensitive data from being exfiltrated off the device. AppFence uses TaintDroid information flow tracking to track the sensitive data and prevent information from transmissions of these data out of the device. However, the system does not alert users about how applications use data and whether they will exhibit side effects if privacy controls are applied. In order to know whether side effects impact user-desired functionality, it needs to consult users each time. In this case, the system may place a high level of burden on users. The Taming Information Stealing Smartphone Applications (TISSA) provides users with fine-grained control over the disclosure of their personal information and consists of three main components (Zhou *et al*, 2011). TISSA was designed to protect four types of personal information: phone identity, location, contacts, and call log. The first one is the privacy setting content provider. It contains the current privacy settings for untrusted apps on the mobile device. It also provides users with an interface in order to query the current privacy settings for an untrusted app (*e.g.*, a location manager). In order to protect personal information, TISSA provides users with empty or bogus options for personal information that may be requested by the app. The second component is the privacy-setting manager. It allows users to manage or update the privacy settings for installed apps. The third component contains content providers or services to regulate the access for four types of personal information: phone identity, location, contacts, and call log. For example, when an app requires access to private data, the system will query the privacy settings, and response to the requests according to the current privacy settings for the app. However, it is difficult for an average user to determine which type of permission is high or low risk for the app because he does not know the reason about permission requirements for individual apps. Additionally, the system does not assist the user to make the right choice in order to reduce the burden on mobile users.

PrivacyGuard [6] and AntMonitor (Le *et al*, 2015) provide fine-grained privacy control and provide ground truth mapping of packets to applications. They used an approach which analyses actual network

traffic of Android using VPNService API to intercept traffic. This approach does not require root permissions and is portable to all devices with Android version 4.0 or later. The AntMonitor system consists of three components: an Android application, AnyClient, and two server applications, AntServer and LogServ. Whilst PrivacyGuard runs in its entirely on the local device. The purpose of the client-side analysis is to protect users in real time and provide fine-grained privacy control. However, LogServer works as the central repository to store and analyse all network traffic data and does not have to analyse a large amount of live traffic compared to AntServer. To evaluate AntMonitor system, they recruited student volunteers to use AntClient on their phones. The system collects the packets of the applications that the volunteers selected and stores them at LogServer in order to check whether any of the installed applications are sending the personal data out to the Internet. They found that 44% and 66% of the users have applications that leak their International Mobile Equipment Identity (IMEI) and Android Device ID respectively. However, both PrivacyGuard and AntMonitor assume that average users can correctly specify their personal information to allow the system to detect them when they leave their mobiles. In this case, these solutions do not help users to overcome the burden associated with managing such a large number of data. ProtectMyPrivacy (PMP) provides users with fine-grained privacy for each app in order to send the anonymized data instead of privacy sensitive information [4]. It detects privacy leaks on iOS Applications. The type of data that PMP protects is a unique device identifier, IMEI, Wi-Fi MAC address and Bluetooth MAC address. Another private data type that PMP protects is the user's address book. It includes names, addresses, phone numbers and emails because some apps upload this information to a server without user's permission. When the app wants to access to the private data, PMP allows the user to deny or allow the app to access private data in real time. Hence, PMP provides user two options to protect his address book: user can allow the app to access his address book or allow PMP to sends an alternative address book, filled with fictitious entries (names, emails and phone numbers). Additionally, they have developed a crowdsourcing system to help users to make informed decisions, which provides app specific privacy recommendations. However, the system just deals with mere access to private data but does not address privacy once the data leaves the app. Moreover, the system does not provide each user with personalized recommendations. Each user has its own privacy preferences. Therefore, it would be helpful to take account of user's profile when the system generates recommendations, in order to make a more personal recommendation.

PRIVACY PROFILES

A few studies have proposed modelling and predicting users' privacy preferences *(Frank et al,* 2012) [7]. Frank et al. cluster 188,389 Android apps and 27,029 Facebook apps to find patterns in permission requests [17]. They used a probabilistic method to extract permission request patterns from Android and Facebook apps. They identified over 30 common patterns of permission requests. However, they looked for permission request patterns in Android apps but they do not identify patterns in user privacy preferences. Liu *et al* analyzed the permissions users granted to mobile apps on Android and realized that the permission model is too complex and they can find patterns in permission requests [7]. They used machine learning clustering algorithms to split users into a small number of profiles based on their decisions to grant or deny apps access to different permissions. Their result showed that it is possible to significantly reduce user

burden while allowing users to better control their mobile app permissions. However, they do not elicit user's privacy preferences in a context where they are not just about the permissions requested by an app but also about the type of information, app categories, data location, time of access data, the entity access to data, data usage and the level of data.

DATA COLLECTION

Participants were recruited through different platforms such as Email, social network, and some communities' centre. Prior to displaying the survey questions, its aims and structure were briefed confirming that the respondents should be 18 years or older and they are free to withdraw up until the final submission of their responses. In total, 407 completed responses and the total responses are within the range of other surveys in the research domain and close to the expected and targeted figure. Demographic information was collected including questions related to gender, age, education, and the level of knowledge in order to analyze the data, though the age ratio or any other demographic composition of the participants were not specifically controlled. Among these participants, 30% of them were male; 70 of them were female. Regarding the age, almost half of the participants were between 25 and 34 which represent 47% of participants. The second large age group was between 35-44 which represent 35%. The vast majority of the populations of the participants were aged between 24-44 years old.

Predicting the User's Mobile App Privacy Preferences

This section examines how to predict many of a user's mobile app privacy preferences which could significantly reduce user burden with minimum questions. Accordingly, the machine learning technique was utilized to assign users to the privacy profiles that most closely capture their privacy preferences.

Interpreting the Resulting Privacy Profiles

As mentioned in the previous section, this paper is an extension of a previous study, which divided users into different profiles by utilizing k-means method to determine the number of profiles. However, in order to determine the optimal number of clusters in the dataset three methods were used statistical testing methods, visual exploration, and precision of predicting users' preferences. The result shows that the optimal number of clusters is four clusters.

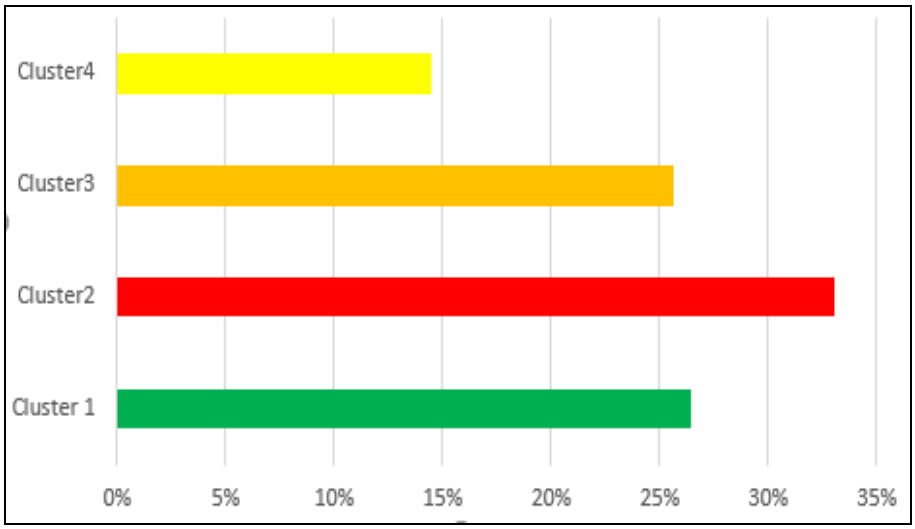

Fig. 1. The Percentage Of Each Cluster

The red chart in Figure 1 indicates a higher level of concern while the green chart indicates a lower level of concern. Cluster 2 represents the conservative group. However, it is clear from Figure 1 that cluster 2 is the largest cluster. Whilst the green chart indicates a lower level of concern who comfort to disclose their data (indicate of unconcerned) and represents 26% of users. The rest of the participants in the remaining clusters such as cluster 4 and 3 are covered in orange or yellow colures which indicates participants are moderately concerned or somewhat to share privacy-related information to the apps. When the centroid of each cluster was computed by averaging the feature vectors of instances within the cluster, cluster 4 and 3 indicate somewhat concerned (cluster 4: μ=2, cluster 3 μ= 3). Whilst the centroid of cluster 2 and 1 were μ=1 and μ=4 respectively. The characteristics of the clusters reveal a significant difference between each privacy profile, which means each cluster, has a unique profile.

Feature Selection

The machine learning approach is applied in order to assign users to these four privacy profiles. R program supports different machine algorithms such as Support Vector Machines (SVM), Random Forest (RF) and K-Nearest Neighbors (kNN). Before using classifier, it is important to determine which questions are the most important to ask users in order to minimize the questions. Therefore, the following techniques were used to minimize the 46 questions and help the models perform better and efficiently as shown in Figure 2.

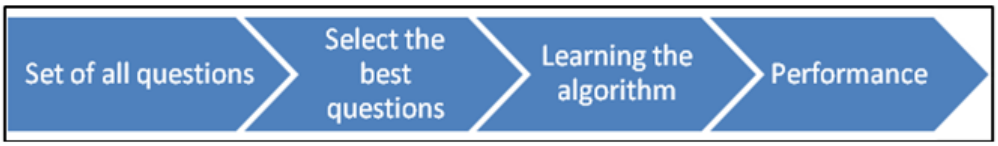

Fig. 2. Feature Selection methods

Machine learning assigns a weight to each question about privacy-related information. Hence, it could be easy to minimize the number of questions, which in turn could actually help reduce user burden. Table 1 shows the features ranked according to the ranking algorithm with their weight. 'Gini index' was performed to assign a score and rank the questions. These scores which are denoted as 'Mean Decrease Gini' by the importance measure indicate how much each question contributes to the homogeneity in the privacy-related- information. After ranking the questions, a top ten questions were selected in order to minimize the number of questions.

Table 1. The 10 Most Important Questions

N	Category	Data	weight
1	Shopping	Identity	12.41
2	Navigation	Approximate location	12.33
3	Shopping	Contacts	12.31
4	Navigation	Contacts	11.57
5	Navigation	Exact location	11.1
6	Productivity	Calendar	11.00
7	Navigation	Phone	10.76
8	Productivity	Identity	10.35
9	Lifestyle	Phone	9.70
10	Shopping	Approximate location	9.14

Accuracy of Predictions

This section aims to create some models of the data and estimate their accuracy. However, there are different algorithms therefore; the following steps were performing in order to determine the best algorithm:

Set-up the test harness to use 10-fold cross-validation.

Build five different algorithms to predict the user's mobile app privacy preferences Select the best algorithm. Ten folds cross-validations were performed to estimate accuracy. This divides the data into 10 groups, train in 9 and test on 1 and release for all combinations of train-test splits. The process was repeated three times for each algorithm with different splits of the data into 10 groups; in an effort to get a more accurate estimate to predict the user's mobile app privacy preferences. Five different machine-learning algorithms were evaluated in order to determine the best algorithm would be good on this problem. The following list shows the five algorithms:

- Linear Discriminate Analysis (LDA)
- Classification and Regression Trees (CART).
- K-Nearest Neighbors (kNN).
- Support Vector Machines (SVM) with a linear kernel.
- Random Forest (RF)

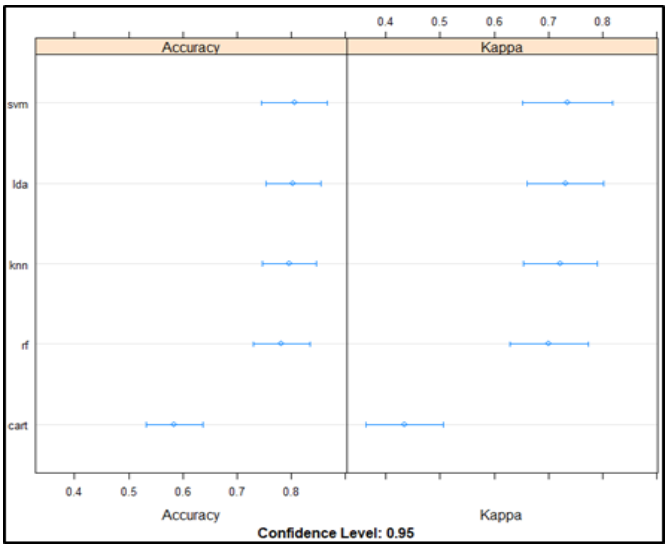

Fig.3. Feature Selection Methods

Figure 3 shows the evaluation results of each algorithm and compare the spread and the mean accuracy of each model. It is clear from Figure 3 that the most accurate model, in this case, was SVM. SVM classifiers achieve the highest accuracy 86% with 10 questions.

Overall accuracy for 46 questions and 4 clusters is 99%. When asking users to answer ten of questions related to privacy decisions, the accuracy decreases to 86%. This result reflects the exploration of tradeoffs between accuracy and the number of questions – in other words, tradeoffs between accuracy and user burden. The result also decreases by 80 % (46 to 10 questions) of the user's effort.

CONCLUSION

The outcomes of this research show that it could significantly reduce the burden and frustration of the user while allowing users to better control information. In particular, the classifiers could be built to assign users to the privacy profiles that most closely capture their privacy preferences. In order to reduce user burden in terms of

decisions, machine learning assigns a weight to each question about privacy-related information. Hence, it could be easy to minimize the number of questions, which in turn could actually help reduce user burden. Five different machine-learning algorithms were evaluated in order to determine the best algorithm. SVM algorithm achieved the highest accuracy, which is 86% for just 10 questions. The research as particularly encouraging as they offer the prospect of remarkably alleviating user burden.

REFERENCES

1) M. Abedi; "involving Kalman filter technique for increasing the reliability and efficiency of cloud computing", International WORLD COMPETITION 2012; Los Vegas, USA.
2) P. Shahbazi; "New Novel idea for Cloud Computing: How can we use Kalman filter in security of Cloud Computing", International IEEE Conf. AICT., Oct. 2012, Georgia, Tbilisi. DOI: 10.1109/ICAICT.2012.6398466.
3) F. Kashefi "Perusal about influences of Cloud Computing on the processes of these days and presenting new ideas about its security", International IEEE Conf. AICT., Dec. 2011, Baku, Azerbaijan. DOI: 10.1109/ICAICT.2011.6111007.
4) Alshehri, A. and Alotaibi, F. (2019) 'Profiling Mobile Users Privacy Preferences', International Journal of Digital Society (IJDS), 10(1), pp. 1436–1441. Available at: https://infonomics-society.org/wp-content/uploads/Profiling-Mobile-Users-Privacy-Preferences.pdf (Accessed: 4 September 2019).
5) Bashirov, A. E., & Norozpour, S. (2018). On an alternative view to complex calculus. Mathematical Methods in the Applied Sciences, 41(17), 7313-7324.
6) BASHIROV, A. E., & Norozpour, S. (2017). On complex multiplicative integration. TWMS Journal of Applied and Engineering Mathematics, 7(1), 82-93.
7) Anton, A. I., Earp, J. B. and Young, J. D. (2010) 'How Internet Users ' Privacy Concerns Have Evolved', IEEE Privacy & Security, 1936(February), pp. 21–27. doi: 10.1109/MSP.2010.38.
8) Balebako, R. et al. (2013) '"Little Brothers Watching You": Raising Awareness of Data Leaks on Smartphones', SOUPS '13: Proceedings of the Ninth Symposium on Usable Privacy and Security, pp. 12:1--12:11. doi: 10.1145/2501604.2501616.
9) Hornyack, P. et al. (2011) 'These Aren't the Droids You're Looking for: Retrofitting Android to Protect Data from Imperious Applications', In Proceedings of the 18th ACM conference on Computer and communications security, pp. 639–652. doi: 10.1145/2046707.2046780.
10) Enck, W. et al. (2014) 'TaintDroid: an information-flow tracking system for realtime privacy monitoring on smartphones', ACM Transactions on Computer Systems (TOCS), 32(2), p. 5. doi: 10.1145/2494522.
11) Norozpour, S. (2018, September). Existence and uniqueness results for Multiplicative Fractional differential equation with three point integral boundary value problem. In THE ABSTRACT BOOK (p. 28).
12) ERGÜN, C., & Norozpour, S. (2019). Farsi document image recognition system using word layout signature. Turkish Journal of Electrical Engineering & Computer Sciences, 27(2), 1477-1488.
13) Ramtin, O. Sharafi, "Tasks Mapping in the Network on a Chip Using an Improved Optimization Algorithm", Published by International Journal of Pervasive Computing and Communications, Vol. 16, Issue 2, PP. 165-182, 2020.

https://doi.org/10.1108/IJPCC-07-2019-0053.

14) S. Norozpour, "Proposing New Method for Clustering and Optimizing Energy Consumption in WSN"; Published by International Journal of Talent Development & Excellence (ISSN: 1869-0459), Vol. 12, No. 3S, PP. 2631-2643, 2020.

15) G. Prakash, N. Gafar, N. H. Jabarullah, M. Jalali, "A New Design of 2-bit Universal Shift Register Using Rotated Majority Gate Based on Quantum-dot Cellular Automata Technology"; Published by International Journal of Theoretical Physics, (ISSN: 0020-7748), PP. 1-19, June 2019. *DOI:10.1007/s10773-019-04181-w.*

16) Frank, M. et al. (2012) 'Mining permission request patterns from Android and Facebook applications', Proceedings - IEEE International Conference on Data Mining, ICDM, pp. 870–875. *doi: 10.1109/ICDM.2012.86.*

17) Egele, M. *et al.* (2011) 'PiOS Detecting privacy leaks in iOS applications', Proceedings of the 18th Annual Network & Distributed System Security Symposium, NDSS 2011, p. 11.

18) Lin, J. *et al.* (2014) 'Modeling Users' Mobile App Privacy Preferences: Restoring Usability in a Sea of Permission Settings', 12th USENIX security symposium., pp. 199–212.

19) S. Seyedi, N. J. Navimipour; "Designing an efficient fault tolerance D-latch based on quantum-dot cellular automata nanotechnology"; Published by Optik Journal, (ISSN: 0030-4026), Vol. 185, PP. 827-837, May 2019. *DOI:10.1016/j.ijleo.2019.03.029*

20) M. Darbandi; "Proposing New Intelligent System for Suggesting Better Service Providers in Cloud Computing based on Kalman Filtering"; Published by HCTL International Journal of Technology Innovations and Research, (ISSN: 2321-1814), Vol. 24, Issue 1, PP. 1-9, Mar. 2017, *DOI: 10.5281/Zenodo.1034475.*

21) M. Darbandi; "Proposing New Intelligence Algorithm for Suggesting Better Services to Cloud Users based on Kalman Filtering"; Published by Journal of Computer Sciences and Applications (ISSN: 2328-7268), Vol. 5, Issue 1, 2017; PP. 11-16; DOI: 10.12691/JCSA-5-1-2; USA.

22) Le, A. et al. (2015) 'AntMonitor: A System for Monitoring from Mobile Devices', (1), pp. 15–20.

23) Liu, B., Lin, J. and Sadeh, N. (2013) 'Reconciling Mobile App Privacy and Usability on Smartphones: Could User Privacy Profiles Help?' Available at: *http://reports-archive.adm.cs.cmu.edu/anon/anon/home/ftp/usr0/ftp/2013/CMU-CS-13-128.pdf (Accessed: 25 December 2017).*

24) Pew Research Center (no date) App Permissions: An Analysis of Android Phones | Pew Research Center. Available at: *https://www.pewinternet.org/2015/11/10/an-analysis-of-android-app-permissions/ (Accessed: 4 September 2019).*

25) M. Darbandi; "Kalman Filtering for Estimation and Prediction Servers with Lower Traffic Loads for Transferring High-Level Processes in Cloud Computing"; Published by HCTL International Journal of Technology Innovations and Research, (ISSN: 2321-1814), Vol. 23, Issue 1, pp. 10-20, Feb. 2017, DOI: 10.5281/Zenodo.345288.

26) S. Haghgoo, M. Hajiali, A. Khabir, "Prediction and Estimation of Next Demands of Cloud Users based on their Comments in CRM and Previous usages", International IEEE Conference on Communication, Computing & Internet of Things; Feb. 2018, Chennai. DOI: 10.1109/IC3IoT.2018.8668119.

27) Song, Y. (2015) 'PrivacyGuard: A VPN-Based Approach to Detect Privacy Leakages on Android Devices', pp. 15–26. doi: 10.1145/2808117.2808120.
28) Song, Y. and Hengartner, U. (2015) 'PrivacyGuard: A VPN-based Platform to Detect Information Leakage on Android Devices', Proceedings of the 5th Annual ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices - SPSM '15, pp. 15–26. doi: 10.1145/2808117.2808120.
29) TRUSTe (2016) 2016 TRUSTe/NCSA Consumer Privacy Infographic - US Edition | TRUSTe. Available at: *https://www.truste.com/resources/privacy-research/ncsa-consumer-privacy-index-us/ (Accessed: 11 March 2017).*
30) Yang, Z. et al. (2013) 'AppIntent: analyzing sensitive data transmission in android for privacy leakage detection', Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security - CCS '13, pp. 1043–1054. *doi: 10.1145/2508859.2516676.*