The surge in battery consumption is one of the high risk factors to security, since a sudden drain of battery may occur due to illegal accessing of internal data or sending credentials to a remote node. The percentage of deviation from the expected value as given in Equation (10) will give a notion of threat to user.

$$V(n_i)\Big|_{BU}^{\alpha_i} = 1 - \frac{\left(\sum_{\alpha_i} BU(n_i) - E[BU(n_i)] \right)}{E[BU(n_i)]} \quad \text{where}, \alpha_i = \{AC, BG\} \subset \alpha \tag{10}$$

Finally, we have formulated a weighted linear equation of the three defined vulnerability factors, as demonstrated in Equation (11), where the weight factor is selected based on the sensitivity of the application.

$$V(n_i) = \eta_1 * RA(n_i) + \eta_2 * BU(n_i) + \eta_3 * MU(n_i) \text{ where, } \eta_1 + \eta_2 + \eta_3 = 1 \text{ and } \left.\begin{matrix}\eta 1 \\ \eta 2 \\ \eta 3\end{matrix}\right\} \geq 0 \tag{11}$$

4. PRINCIPAL COMPONENT ANALYSIS

The principal component analysis is a statistical technique, which is mainly applied to find patterns in a dataset, and to select the most significant feature producing high impact among the given features in the dataset. It provides mapping of a dataset with few variables into a new dataset with less uncorrelated variables, each of which is represented as a linear combination of the original variables, known as the principal components. The principal components are estimated from the Eigenvectors of the covariance matrix of the original variables and they are ordered by the amount of variation in the given data. The first principal component captures the largest variance and it is represented by drawing an axis in the direction of maximum variation of the given data. A second axis is added in the direction orthogonal to the first to display the next highest variation, which is the second principal component; then, the subsequent principal components capture the residual variations sequentially in a descending order.

4.1 Comparing Traffic Data Sets using PCA

Let X be a table of traffic datasets observed from a number of Apps, where each dataset has i observations, described by j variables (features) and each table is represented by a matrix $X_{i,j}$ having i rows and j columns. As a first step, we have standardized the given data matrix $X_{i,j}$ with zero mean and unit variance, since the variables in the dataset are measured on different scales and have different units. Then, a covariance matrix of the given data matrix $X_{i,j}$ is constructed with size ($j \times j$), where each row represents the dependence among the other selected variables. Here, for the given matrix of size ($i \times j$), having i observations and j variables, *where i=5, and j=3(a, b, and c),* its covariance matrix is obtained as a matrix of size (3×3) as shown in Equation (12) and it takes three values: positive, negative and zero. A positive covariance indicates that both the features are changing together, whereas a negative covariance indicates a converse relationship and a zero demonstrates no relation.

$$\begin{bmatrix} \text{cov}(a,a) & \text{cov}(a,b) & \text{cov}(a,c) \\ \text{cov}(b,a) & \text{cov}(b,b) & \text{cov}(b,c) \\ \text{cov}(c,a) & \text{cov}(c,b) & \text{cov}(c,c) \end{bmatrix} \tag{12}$$

Subsequently, we have estimated the Eigenvectors and Eigenvalues from the covariance matrix, where the Eigenvectors of the principal component represent the direction of new features space and the Eigenvalues represent the magnitudes. Then, a projection matrix is constructed with the selected r features from j features, such that $r < j$, where j is the original data matrix variables; the r features are selected after sorting the computed Eigen values in descending order. This facilitates the transformation of original data matrix through the projection matrix to obtain the new R-dimensional feature subspace.

5. PROPOSED METHODOLOGY TO DERIVE FRAMEWORK FOR APP LABEL

The proposed labeling may be embedded as a code construct within the App during App development to update the defined parameters periodically over a specified timespan *t*. The pseudocode construct is shown in Figure 2, and it is continuously recording the defined App parameters and exploring the data to update its popularity, energy consumption and memory consumption and maximum responsiveness to the users.

We have proposed a structure of App label frame as illustrated in Figure 3, where we have defined the labels, such as features, popularity, energy consumption, and security. The features detail the functions of the App, and it should be given by the developer to know how they match with users' requirements. The popularity is classified into three, based on the range of connected domains: i) local, if the users are from the local hub (domain), ii) extended, if the users are connected widely and iii) global, if connectivity is across the borders and the range is high. Then, the second label, energy consumption, details the average energy consumed over the stipulated timespan during the execution and the third label, security, is defined in a scale from zero to one. The App is unsecure, when the value is close to zero and its moderate if it is around the midpoint and highly secure when it is close to one.

Dynamic Code Construct for App_Label ( )

{

Generate maximum connectivity pattern for the App( );

Observe and log parameters (node_deg, traffic, domain, battery_use, and memory_ use);

Set thresholds of battery_use and memory_use parameters from_observations ( );

Estimate_popularity over defined timespan t ( ):

Estimate_security of App ( );

Update_observation_period ( ) (t = t+ Δt);

Update_App_labels ( );

}

Figure 2. Pseudocode construct for App labels generation.

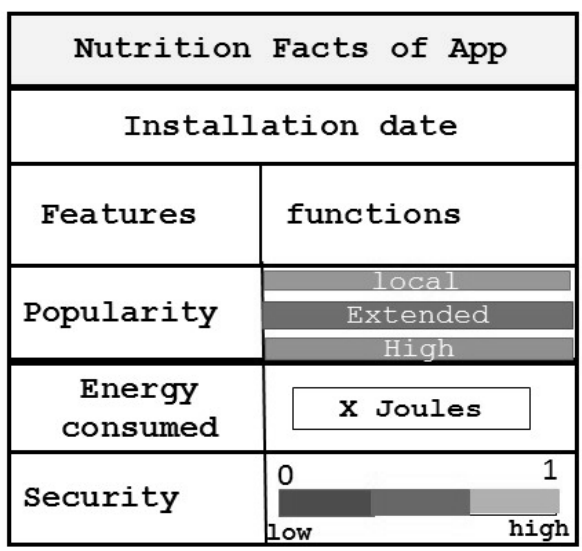

Figure 3. App label frame.

6. RESULTS AND DISCUSSION

Our experiment test scenario is comprised of sample data comprising traffic pattern recorded for 30 days period (Caida, 2019), which is added with features, such as the node degree, connectivity, domain information to reflect an App behavior. We have carried out a set of experiments to estimate the parameters pertaining to initial phase of labeling the App.

6.1 Popularity

The node degree of connected users in the selected range, as explained in section 5, the frequency of connectivity, and the duration of connectivity *(t)* are the three parameters utilized to conclude the popularity of a given App, and the popularity of an App is defined as in Equation (13). The popularity of App in various domains is demonstrated in Figure 4, where each section represents range of connectivity and the number of connections. Here, there are more numbers of connectivity observed in higher connectivity range (70%), thereby indicating high popularity.

$$P = \sum(\deg(App) * t * domain) \tag{13}$$

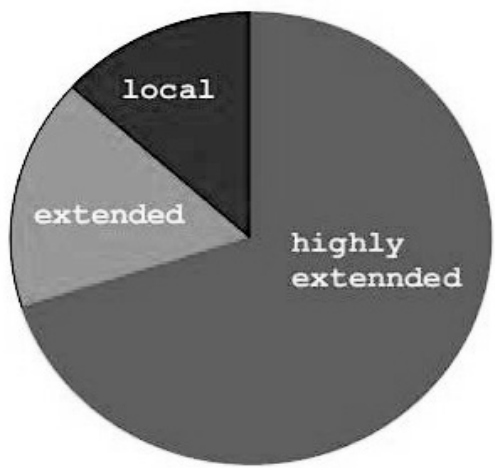

Figure 4. Popularity of the selected App.

6.2 Energy Consumption

We carried out an energy calculation for the selected App traffic, where we made few assumptions. In the experimental setup, we have considered a fixed data rate of 1500 kbps and the average power consumed during the selected data rate as 300 mill watt, according to (Caida, 2019). We have considered a fixed packet size of 1472 bytes, according to (Fall & Stevens, 2011) and we have estimated the duration of each packet and the total number of packets generated for a specific traffic. The battery capacity is taken as 1369 mAh, rated at 5V and the corresponding energy is 24642 J. We estimated the energy consumption against the number of packets and its ratio against the given maximum energy limit during the App execution. The energy consumption specific to day traffic is found to be varied between 14 J and 39 J, which exceeds the upper limit of 35 J in 10 percent (3 days /30 days) of the total observations.

6.3 Vulnerability Index

We have utilized three parameters, that is, closeness centrality, battery consumption and memory consumption to estimate the vulnerability index of a data transfer. Equation (11) in Section 3.4 is utilized to compute the vulnerability index, where the parameters η_1, η_2, and η_3 are selected as 0.4, 0.3 and 0.3 respectively to give importance to the domain of connectivity. The maximum node degree for each data transaction is restricted to 50 and a distance matrix detailing the node distances corresponding to all connectivity is included. In this experiment scenario, the closeness

centrality is given more weightage, as the sparsely distributed nodes may be prone to personal attack easily. In each category, the outcomes of the experiment are assigned with lower and upper threshold values and three ranges are derived as in Equation (14) based on the selected threshold values. Figures 5 and 6 demonstrate the observed data and its vulnerability index. It is observed from Figure 6 that 6% of the observed data falls near the upper limit of the 0 to 1 scale of vulnerability index and 23% data fall near the lower limit and the rest 71% is in the normal range, thus showing the risk index is low and security is high.

$$\begin{aligned} range\_1 &= 0 \prec range1 < lower \\ range\_2 &= lower \le range2 < upper \\ range\_3 &= \ge upper \end{aligned} \tag{14}$$

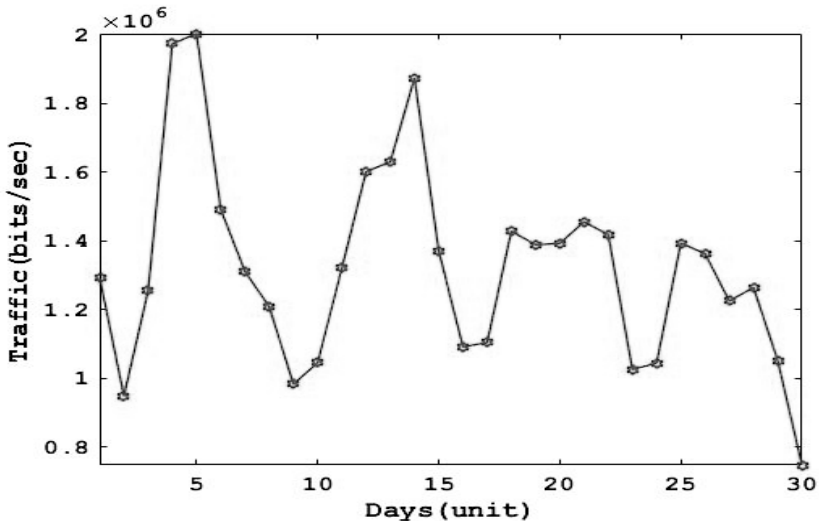

Figure 5. App traffic pattern.

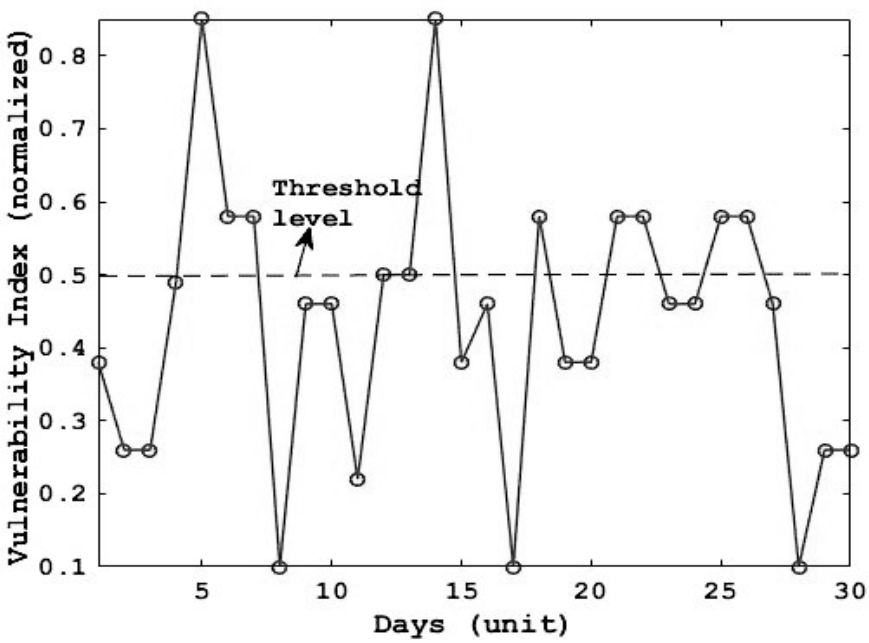

Figure 6. Vulnerability Index of traffic pattern in Figure 5.

Finally, we draw the App label frame for the above selected application, as in Figure 7, which summarizes the statistical constituents of the App from user point of view.

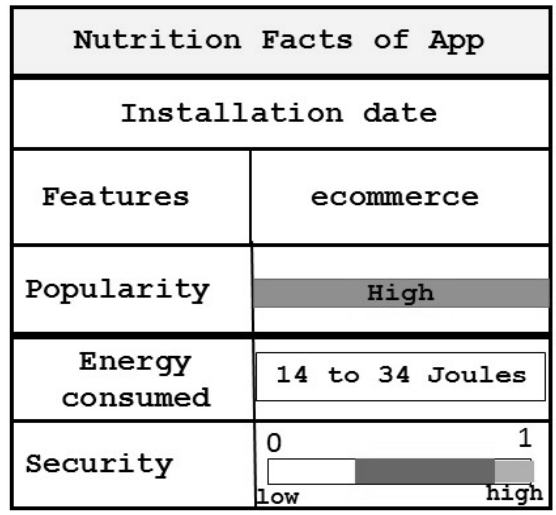

Figure 7. App Label Frame.

6.4 Validation of Proposed Framework

We carried out a second set of experiments on a real-time traffic dataset (kaggle, 2017) to validate the proposed framework, where the traffic dataset is comprised of frequency of transactions, amount of data transfer, duration of connection, number of forwarded packets, number of backward packets, length of packets, and energy usage for various Apps. We extracted the data of three Apps for 100 observations. We applied PCA to select the prominent features out of seven given parameters in the dataset. The prominent features, showing high variances are: traffic flow and connection frequency. We compared the selected Apps with respect to the traffic flow and the resultant covariance and scores of PCA analysis are presented as a three dimensional graph using biplot, as shown in Figure 8. The graph is used to distinguish the prominent features in a particular run. The parameters component 1, component 2 and component 3 represent the first three principal components of the dataset; where the lines labelled by App-1 to App-3 represent the correlations of Apps with each other. App-1 and App-2 are closely related, as those Apps showed positive correlation and an angle deviation less than 90°. App-3 showed a large angle deviation, thus demonstrating negative correlation with App-1 and App-2. The scores (dots in the graph) on each row demonstrate the positive or negative impact of selected features on the given App.

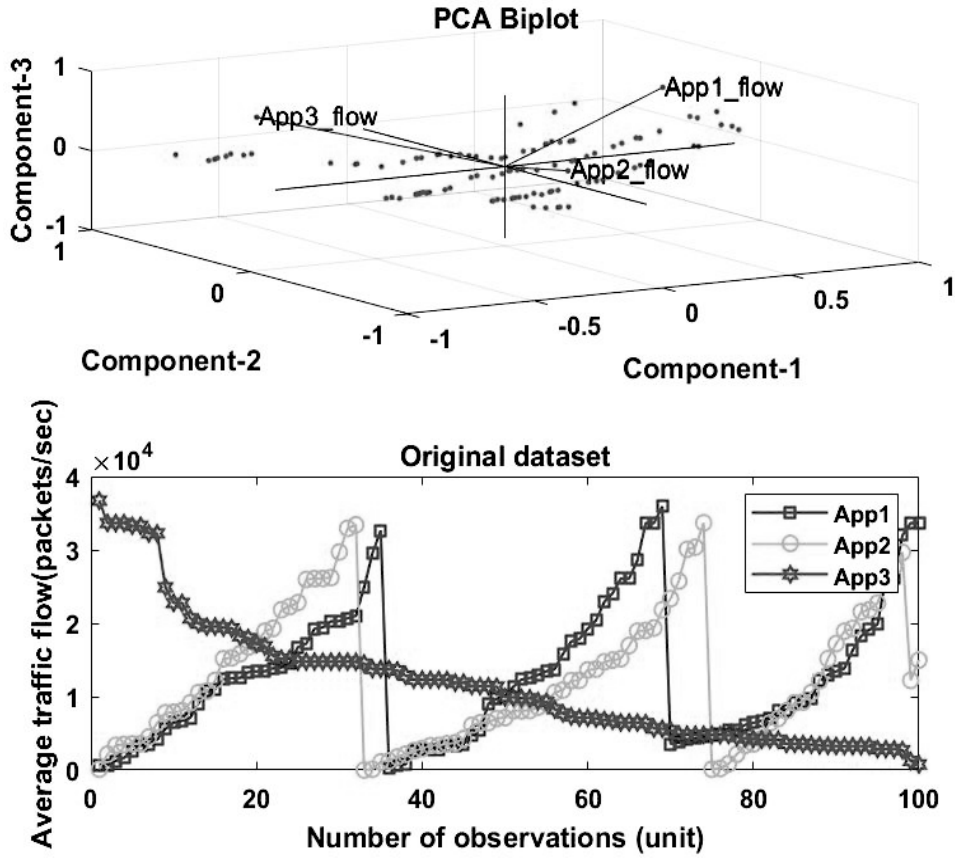

Figure 8. Principle component analysis.

The second plot in Figure 8, demonstrated the behavior of original dataset. On comparing the biplot with the original dataset exhibiting the Apps behavior, it is observed that App-3 showed a negative trend which is very well matched from the biplot conclusions.

Further, we computed the variability index of the Apps, as presented in Table 1. The variability index showed that the App, which is showing highest variation is contributing more in the assessment of the selected parameters, in this case, the traffic flow. This might be due to increase in the popularity and more usage of the App.

Table 1. Variability Index of the Apps with respect to traffic flow.

App	Variations
App-1	59.67
App-2	22.95
App-3	17.37

The connection frequency is presented in Figure 9. It is observed that App1 is highly connected and it is popular among the three. It also showed higher variations during data transfer, as tabulated in Table 1.

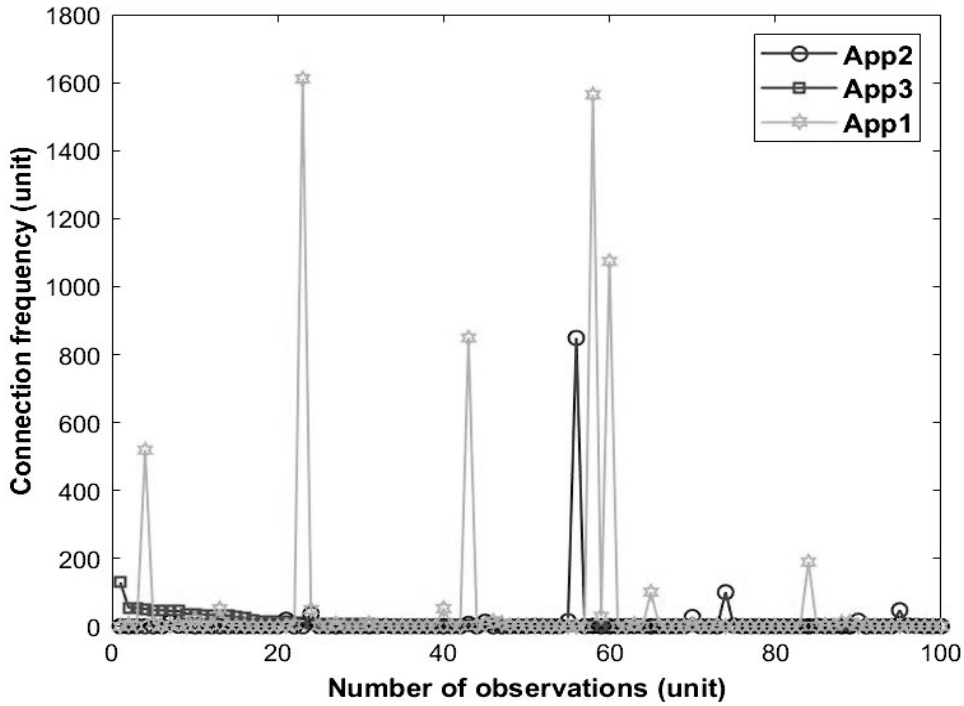

Figure 9. Connection frequency variations observed among the Apps.

We estimated the energy consumed by the three Apps, where we utilized the parameters, such as, frequency of connection, duration of each connections, and the number of packets transferred to estimate the energy. We assumed k units of energy consumed by each packet transfer, and the average energy transfer for 100 observations is shown in Figure 10. App1 consumed 12% more than App2 and 35% more than App3. This demonstrated that App3 is not popular and with lower transactions.

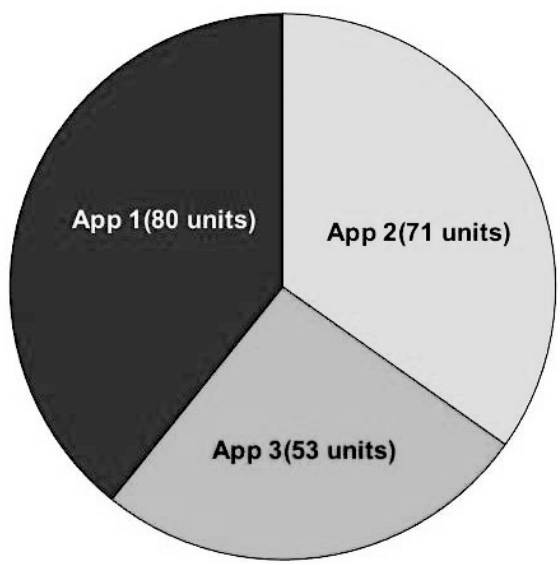

Figure 10. Energy consumption observed among the Apps.

The security level is computed by analyzing the variance in the connectivity, which is listed in Table 2. More variations (90%) are observed in App1 compared to others, while App2 and App3 contributed only 10% variations. On consolidating the connection frequency and energy parameters, App1 tops the security. The summary of App labels for the selected comparison study is presented in Figure 11, which guides the users in choosing right App according to their requirements.

Table 2. Variability Index of the Apps with respect to connection frequency.

App	Variations
App-1	90.06
App-2	9.49
App-3	0.44

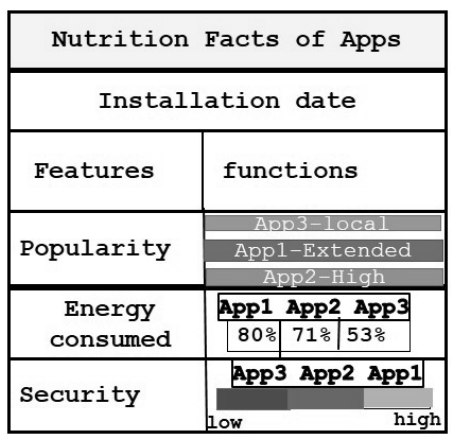

Figure 11. Comparison of App labels.

7. CONCLUSION

We have proposed a quantification labeling to assess mobile/web App's quality to facilitate the users to know the features, responsiveness, security, and energy consumption before selecting an App from a group of similar Apps in the market. Our labeling accounted the series of states the App undergoes during its execution by quantifying the changes in the mobile network components. The degree and domain of connectivity, energy consumption and vulnerability were the set of parameters selected to quantify the state changes, and subsequently, utilized to form an App label frame, which is comprised of four labels: features, popularity, energy consumption, and security. Further, the labeling applied principal component analysis, a statistical technique, to compare the traffic pattern of two Apps running simultaneously to categorize the prominent one impacting the network parameters. The experimental results in quantifying a real time traffic data validated the proposed framework by deriving information necessary to form the labels. The principal component analysis carried out on two sets of traffic data enables the selection of the prominent App features, which revealed the possibility of embedding of the framework within an App for dynamic monitoring. We validated the proposed framework using a real dataset, and the estimated quantification parameters revealed the suitability of the proposed framework for framing App labels.

ACKNOWLEDGMENT

This work was supported by Kuwait University under a research grant no. QE01/17

REFERENCES

Akhawe, D., and Finifter, M. 2012. Product labels for mobile application markets. In the Proceedings of Mobile Security Technologies Workshop, 24 May, San Francisco, CA, USA.

Alderson, D., Li L., Willinger W., and Doyle J. 2005. Understanding internet topology: principles, models, and validation. IEEE/ACM Transactions on Networking. vol. 13, no. 6, pp. 1205-1218.

Alotaibi, A., Clause, J., and Halfond, W.G.J. 2020. Mobile App Energy Consumption: A Study of Known Energy Issues in Mobile Applications and their Classification Schemes – Summary Plan, In the Proceedings of IEEE International Conference on Software Maintenance and Evolution, Adelaide, Australia, pp. 854-854.

Chevalier, J.A, and Mayzlin, D. 2006. The effect of word of mouth on sales: online book reviews, J. Marketing Research, vol. 43, no. 3, pp. 345-354.

Chun, B.G., and Maniatis, P. 2009. Augmented smart phone applications through clone cloud execution, In the Proceedings of the 12th Workshop on Hot Topics in Operating Systems, USENIX Association, May 18-20, Monte Verita, Switzerland, pp. 1–8.

Cho, T., Kim, J-H., Cho, H-J., Seo, S-H., and Kim, S. 2013. Vulnerabilities of android data sharing and malicious application to leaking private information, In the Proceedings of Fifth International Conference on Ubiquitous and Future Networks, 2-5 July, Da Nang, Vietnam, pp. 37-42.

Chen N., Hoi S.C.H., Li S., and Xiao X. 2016. Mobile app tagging. In the Proceedings of 9th ACM international conference on web search and data mining, 22 – 25 February, San Francisco, California, USA, pp. 63-72.

Chen, S, Fan, L., Meng, G., Su, T, Xue, M., Xue, Y., Liu, Y., and Xu, L. 2020. An Empirical Assessment of Security Risks of Global Android Banking Apps, In the Proceedings of IEEE/ACM 42nd International Conference on Software Engineering (ICSE), Seoul, pp. 1310-1322.

Dean, G. 2013. Understand the states and transitions of an iOS App. Downloaded from https://www.techrepublic.com on 12-September 2019.

Fall K.R and Stevens W.R. 2011. TCP/IP Illustrated-volume I. Addison Wesley, Michigan, USA.

Harman, M., Jia, Y., and Zhang, Y. 2012. App store mining and analysis: MSR for App stores, In the Proceedings of 9th IEEE Working Conf. Mining Software Repositories, June 2-3, Zurich, Switzerland.

Hayes, T. 2016. Mobile Apps for 21st Century Skills: A quantitative analysis of educational mobile apps on graphite.org. In Proceedings of 2016 World Conference on Educational Media and Technology. 28-30 June, Vancouver, BC, Canada, pp. 1630-1637.

Habib S.J., Marimuthu P.N., and Rajasundari, T. 2018. Quantification of new web applications within enterprise networks. In the Proceedings of International Conference on Information Integration and Web-based Applications and Services, 19-21 November, Yogyakarta, Indonesia.

Habib S.J., Marimuthu P.N. 2019. App nutrition label. In the Proceedings of World Conference on Information Systems and Technologies, 16-19 April, Galicia, Spain.

Huckvale, K., Prieto, J.T., Tilney, M., Benghozi, P.J., and Car, J. 2015. Unaddressed privacy risks in accredited health and wellness Apps: A cross-sectional systematic assessment. BMC medicine, vol. 13, no. 214, pp.1-13.

I. de la Torre-Díez, B.O. Trinchet, J.J.P.C. Rodrigues and M. López-Coronado. 2017. Security analysis of a mHealth app in Android: Problems and solutions, In the Proceedings of IEEE 19th International Conference on e-Health Networking, Applications and Services, Dalian, China, pp. 1-6.

Internet traffic traces, https://www. Caida.org/data accessed on September 2019.

Jebari C., and Wani, M.A. 2012. A multi-label and adaptive genre classification of web pages. In the Proceedings of 11th International Conference on Machine Learning and Applications, December 12-15, Boca Raton, FL, USA, pp. 578-581.

Jiang Z., Kuang R., Gong J., Yin, H., Lyu, Y. and Zhang, X. 2018. What makes a great mobile app? A quantitative study using a new mobile crawler. In the Proceedings of IEEE Symposium on Service-Oriented System Engineering, 26-29 March, Bamberg, Germany, pp. 222-227.

Lewis, T.L., and Wyatt, J.C. 2014. MHealth and mobile medical apps: A framework to assess risk and promote safer use, Journal of Medical Internet Research, vol. 16, no. 9. pp. 1-10.

Mojica Ruiz, I.J, Nagappan, M., Adams, B., Berger, T., Dienst, S., and Hassan, A.E. 2016. Examining the rating system used in mobile-App stores, IEEE Software, vol. 33, no. 6, pp. 86-92.

Nayga R.M, Lipinski D, and Savur N. 1998. Consumers' use of nutritional labels while food shopping and at home. Journal of Consumer Affairs, vol. 32, no. 1, pp. 106-120.

Pei, Y. 2015. Linear principal component discriminant analysis. In the Proceedings of IEEE International Conference on Systems, Man, and Cybernetics, October 9-12, Hong Kong, pp. 2108-2113.

Silitonga A.S., Atabani A.E., and Mahlia, T.M.I. 2012. Review on fuel economy standard and label for vehicle in selected ASEAN countries, Renewable and Sustainable Energy Reviews, vol. 16, no. 3, pp. 1683-1695.

Sun Z., Ji Z., Zhang P., Chen C., Qian X., Du X., and Wan Q. 2017. Automatic labeling of mobile apps by the type of psychological needs they satisfy. Telematics and Informatics, vol. 34, no. 5, pp. 767-778.

Traffic dataset: https://www.kaggle.com/jsrojas/ip-network-traffic-flows-labeled-with-87-apps, accessed on 12 December 2020.

Van den Wijngaart A.W. 2002. Nutrition labelling: purpose, scientific issues and challenges, Asia Pacific Channel of Clinical Nutrition, vol. 11, no. 2, pp. 68 -71.

Venkataraman, H., and Muntean G-M. 2012. Green mobile devices and networks energy optimization and scavenging techniques. CRC Press. NY, USA

Xiao, Y., Cui, Y., Savolainen, P., Siekkinen, M., Wang, A., Yang, L., Yla-Jaaski, A., and Tarkoma, S. 2014. Modeling energy consumption of data transmission over Wi-Fi, IEEE Transactions on Mobile Computing, vol. 13, no. 8, pp. 1760-1773.

Zakeri, V., Tavakolian, K., Arzanpour, S., Zanetti, J.M., Dumont, G.A., and Akhbardeh, A. 2014. Preliminary results on quantification of seismo-cardiogram morphological changes, using principal component analysis. In the Proceedings of 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, August 26-30, Chicago, IL, USA, pp. 6092-6095.

16 Metrics to Ensure Mobile App Success. Downloaded on 10-September-2019 https://www.appdynamics.com/media/uploaded-files/1432066155/white-paper-16-metrics-every-mobile-team-should-monitor.pdf