Guarding Sensitive Sensor Data against Malicious Mobile Applications

Cynthia Claiborne
Cynthia.Claiborne@untsystem.edu

Ram Dantu
Ram.Dantu@unt.edu

Cathy Ncube
cymat1982@yahoo.com

University of North Texas
Denton, TX

Abstract—With increasing usage of sensor data for medical purposes, the ability to secure sensitive features in mobile sensor data from adversarial applications is a continuous challenge. This paper introduces a random anonymization algorithm, *SparCTym,* as a method for anonymizing sensitive features in walking accelerometer data while maintaining the utility of the data. *SparCTym* was implemented in the Android framework of a Nexsus S phone and tested with activity recognition applications**.**

Keywords

accelerometer, sensor data, Android, anonymized, sensitive, Sparse#, SparCTym, Nexsus S

This material is based upon work supported by the National Science Foundation under awards 1241768 and 1637291.

1. INTRODUCTION

Parallel with maintaining the utility of mobile sensor data, there is a growing concern and interest in the role that sensors play in the compromise of a mobile user's privacy. The potential risk of an adversarial application utilizing mobile sensor data to exploit a user's privacy is ever present and is a subject that has to be continuously analyzed and innovatively addressed.

While a user may be agreeable in allowing the use of his mobile sensor data for medical or research purposes, he does not want an adversarial application to capture sensitive information in his data.

Supervised training is a commonly used method for labeling data and building models to detect simple everyday activities [1][2][4][9]. It employs activity recognition algorithms and feature extraction to classify these activities. Over the years, classification techniques have proven to be very effective on mobile accelerometer sensor data. However, now, because of the accuracy with which activities are classified, the problem has arisen concerning the importance and the need to "unclassify" or hide features in a mobile user's accelerometer data that can be used to identify sensitive information about that mobile user.

Alongside the improvement in the ability to accurately classify a mobile user's activities, another privacy and security risk has arisen with Android's new 'Activity Recognition' permission. This permission is hidden under 'Other' permissions and does not require a mobile user's intervention [10].

The subject of [1] reveals how accelerometer sensor data is used to classify activities such as standing, walking, running, climbing up stairs, climbing down stairs, sit-ups and vacuuming. Other areas of research have taken this type of study further and show how, seemingly, innocent mobile accelerometer sensor data can be used to recognize a mobile user's daily activities such as walking, jogging, climbing stairs, sitting, and standing [3].

Using the accelerometer and gyroscope sensors, [9] focuses on recognizing the following activities; sitting, laying, standing, attaching to a table, walking, jogging,

This material is based upon work supported by the National Science Foundation under awards 1241768 and 1637291.

running, jumping, pushups, going down stairs, going up stairs and cycling. Person recognition is the study of [4] and is achieved by using supervised training techniques on a mobile user's accelerometer sensor data to determine specific activities.

Even though activity recognition techniques have enhanced and proven to be beneficial in areas such as health and medical research, these advancements, also, bring an increase in the risk of a mobile user sharing sensitive activity data with, potentially, adversarial applications. For example, a mobile user experiencing a limp, during the recovery from a recent knee surgery or another mobile user who has had a limp for years, both will not want to give an adversarial application their raw accelerometer data as it could reveal vulnerabilities in their walking data.

To address the problem of, potentially, sending sensitive features in activity data to calling applications, we propose a modification to the Android operating system with a feature anonymizing/utility preserving algorithm called *SparCTym.* The purpose of this algorithm is to randomly anonymize sensitive features in a mobile user's raw accelerometer walking data while maintaining its utility. Using the *SparCTym* algorithm, we aim to anonymize the following key attributes of accelerometer data: entropy, mean and correlation. There are some features, such as max and min that are expected to stay the same between the anonymized and original data, because of nature of the *SparCTym* algorithm.

1. RELATED WORK

The subject in [5] provides mobile sensor data that can be utilized by a requesting application, while protecting the privacy of users. It proposes the privacy framework, IPShield which accomplishes two major tasks: A mobile user defines a 'Blacklist' of inferences that should not to be shared with a calling application and a 'Whitelist' of inferences that can be shared with the application; a graphical model is created to reveal what an application already knows about a mobile user. This model is then used to determine what type of data will be sent to a calling application: suppressed, (no data are sent), perturbed, (noise is introduced into the data, prior to releasing it to an application or synthetic, (data unrelated to the sensor data)

This material is based upon work supported by the National Science Foundation under awards 1241768 and 1637291.

In another study, a modification to the Android framework called "Override" [6] is introduced to intercept raw sensor data, prior to it reaching a calling application and depending on rules set by the mobile user, either perturbs it or replaces it with synthetic data. Our study differs from the previous two, as there are no rules or anything to setup by the mobile user.

Activities in [7] were recognized using the K-nearest neighbor (K-NN) algorithm. A small database is created with the training data from activities such as walking, running, climbing up, etc. After training and classifying the data initially collected, K-NN is used to classify new records of each activity performed
for a specified time by comparing them with the already trained data to obtain the Euclidian distance between points. The new record is classified as its nearest neighbor. The *SparCTym* algorithm, also, creates a mini database. However, this database differs from that of [7], as it is created by logging, for a minute, each 'X' value with the first occurrence of a specified substring of 'X and its corresponding 'Y' and 'Z' values..

A new multi-objective loss function is used in [10] to train convolutional auto-encoders (CAEs) to provide a method for anonymizing accelerometer and gyroscope data from a mobile phone. In our study, an encoder is not used.

Prior work [8] explored randomly anonymizing mobile sensors' data on an Android mobile phone and testing the manipulated data on various Android sensor applications. This paper extends that study by exploiting the randomness of accelerometer data to obscure sensitive features in that data.

The remaining sections unveil the methodology and test results after implementing the *SparCTym* algorithm in the Android framework.

2. METHODOLOGY

The purpose of the *SparCTym* algorithm is to anonymize sensitive features in a mobile user's accelerometer walking data while maintaining the utility of the data. In this study, the features focused on for anonymization are entropy, correlation and mean. These elements were chosen because they are common features used in activity

This material is based upon work supported by the National Science Foundation under awards 1241768 and 1637291.

recognition algorithms.

***SparCTym* Algorithm:**

```
For time < = 1 minute
   For every unique substring(X)
      Insert X, Y, Z into ArrayB
      Count unique substring(X)
      Save 3 largest counts in max1, max2 and max3
      Anonymize X
      Output X,Y,Z
   End;
End;
For time > 1 minute
   For every new X value
      If substring(newX) in arrayB
           If X = max1, max2 or max3
                   If flag = 1
                     Randomly anonymize X
                   Else
                     Do not anonymize X.
                   End
           End;
      Else if substring(new X) not in array B and flag = 0
             Randomly anonymize X and use original Y and Z.
      End;
      Output X,Y,Z
   End;
End;
```

Figure 1

For time less than or equal to one minute, save each 'X' value with the first occurrence of a specified substring of 'X and its corresponding 'Y' and 'Z' values in an array (arrayB). Count the number of unique substrings of the "X" values that have occurred. These counts are known as *Sparse#*s, (Sparse numbers). Once the time is greater than one minute, select the three largest *Sparse#s* These values will be known as max1, max2 and max3.

For each new 'X' value, check if the substring of 'X' exists in arrayB. If the substring of the X value is in arrayB, select the 'X' value found and its corresponding 'Y' and 'Z' values in arrayB. If the substring of the X value is not in arrayB, then select the original "X', 'Y' and 'Z' values and randomly anonymize 'X'.

Note: If the *Sparse#* of an "X' value equals to max1, max2 or max3, then there is a greater probability that "X' will be randomly anonymized than

This material is based upon work supported by the National Science Foundation under awards 1241768 and 1637291.

if the *Sparse#* does not equal to max1, max2 or max3.

3. RESULTS

X-axis	*Entropy-O*	*Entropy-A*	*Corr(x,x')*	*Max-O*	*Max-A*	*Min-O*	*Min-A*	*Mean-O*	*Mean-A*	*Approx Time(min)*
1	9.413628	10.29177	0.833334	19.34515	19.05785	-5.47793	-5.42047	5.306627	4.456562	3
2	9.357947	10.39957	0.816822	19.44092	19.44092	-4.86502	-4.86502	5.67368	4.76427	4
3	9.163893	10.34524	0.613192	14.5759	14.5759	-18.6173	-18.6173	-3.43012	-1.74553	2.4
4	9.195563	10.21246	0.905513	17.27656	17.27656	-7.53484	-7.53484	2.701332	2.494276	3
5	9.525144	9.851024	0.823844	17.27656	17.27656	-7.53484	7.53484	7.500744	5.953381	3.5
6	8.588724	10.09314	0.625372	17.27656	17.27656	-7.53484	-7.53484	-1.66484	-0.81871	4.5
7	10.54151	9.705886	0.786994	17.27656	17.27656	-7.53484	-7.53484	7.427767	5.847743	4.2
8	9.717024	10.41499	0.792897	17.27656	17.27656	-7.53484	-7.53484	7.850074	6.087719	3
9	9.634911	10.20446	0.821203	17.27656	17.27656	-7.53484	-7.53484	6.725572	5.320417	2
10	9.856614	10.6956	0.76593	17.27656	17.27656	-7.53484	-7.53484	8.657008	6.65247	6

Table 1

Y-axis	*Entropy-O*	*Entropy-A*	*Corr(x,x')*	*Max-O*	*Max-A*	*Min-O*	*Min-A*	*Mean-O*	*Mean-A*	*Approx Time(min)*
1	8.55523	8.568587	0.238515	19.6133	19.6133	-5.76524	-5.76524	7.73322	7.073151	3
2	9.140095	8.44794	0.200646	12.718	19.6133	-2.18351	-2.18351	7.596848	6.993391	4
3	9.277036	8.447172	0.376482	-18.6173	19.6133	-2.48997	-2.48997	-9.02831	8.479211	2.4
4	9.387105	8.499581	0.452979	19.6133	19.6133	-1.89621	-1.89621	8.917103	8.773177	3
5	9.254084	8.413321	0.395204	19.6133	19.6133	-1.89621	-1.89621	4.74731	-1.89621	3,5
6	8.333823	7.742663	0.191221	19.6133	19.6133	-1.89621	-1.89621	9.075225	6.587124	4.5
7	9.476378	8.559606	0.192524	19.6133	19.6133	-1.89621	-1.89621	6.181217	6.376716	4.2
8	9.627792	8.471075	0.213658	19.6133	19.6133	-1.89621	-1.89621	5.157546	5.804249	3
9	9.553783	8.610077	0.056359	19.6133	19.6133	-1.89621	-1.89621	6.617608	5.746302	2
10	9.500363	8.507697	0.044827	19.6133	19.6133	-1.89621	-1.89621	4.534509	6.14974	6

Table 2

This material is based upon work supported by the National Science Foundation under awards 1241768 and 1637291.

Z-axis	*Entropy-O*	*Entropy-A*	*Corr(z,z')*	*Max-O*	*Max-A*	*Min-O*	*Min-A*	*Mean-O*	*Mean-A*	*Approx Time(min)*
1	9.161894	8.562146	0.304118	15.32289	15.32289	-14.6142	-11.9135	-0.5093	0.487325	3
2	9.012137	8.469517	0.230011	12.718	12.718	-13.6374	-13.6374	-0.65951	0.315711	4
3	9.01959	8.368485	0.473301	13.48414	13.48414	-18.6748	-16.453	-0.42243	0.275771	2.4
4	9.036708	8.271748	0.396542	11.6454	11.6454	-11.3581	-11.3581	0.2228	1.500955	3
5	9.388446	8.495653	0.559651	11.6454	11.6454	-11.3581	-11.3581	0.943359	-11.3581	3.5
6	8.420772	7.736031	0.271345	11.6454	11.6454	-11.3581	-11.3581	2.488453	3.631723	4.5
7	9.455842	8.516423	0.161946	11.6454	11.6454	-11.3581	-11.3581	-0.10294	0.646943	4.2
8	9.522114	8.351078	0.140925	11.6454	11.6454	-11.3581	-11.3581	5.157546	0.971274	3
9	9.222059	8.450561	0.011765	11.6454	11.6454	-11.3581	-11.3581	-0.05821	1.134066	2
10	9.498556	8.381469	0.056228	11.6454	11.6454	-11.3581	-11.3581	-0.04764	0.643115	6

Table 3

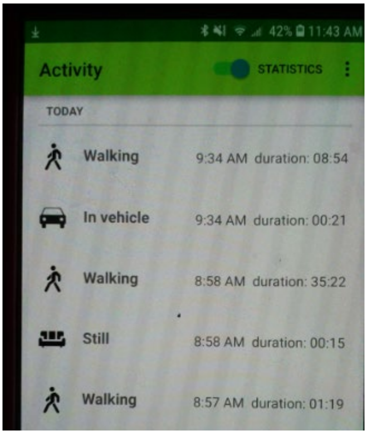
Over a period of three days and for different lengths of time, a Nexus S phone user captured recordings as he walked. Tables 1, 2 and 3 hold the results of calculations performed on the data.

From these tables, it can be seen that the entropy of the anonymized accelerometer walking data (*Entropy-A*) is different from that of the entropy of the original data (*Entropy-O*). Also, the anonymized accelerometer mean (*Mean-A)* is notably different than that of the original mean (*Mean-O).*

The results, also, show that the Max (*Max-A*) and min (*Min-A*) anonymized X, Y and Z values were sometimes different from those of the original Max (*Max-O*) and Min (*Min-O*) original values. This would be expected, because if the substring(X) is found in array(B), there is a good chance that the original Max and Min values will be reused in the anonymized data.

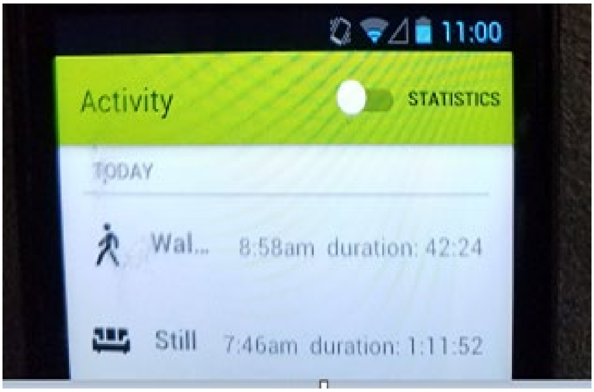
Correlation of the anonymized X, Y and Z values with the original X, Y and Z values portrays how significantly different the anonymized data are different from the original data.

This material is based upon work supported by the National Science Foundation under awards 1241768 and 1637291.



(Galaxy S7)

Picture 1



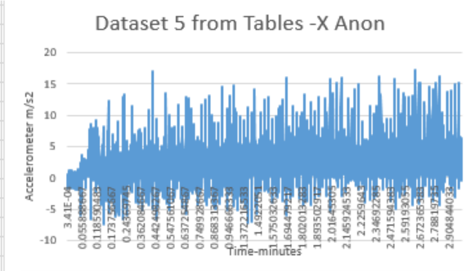
(Nexus S)

Picture 2

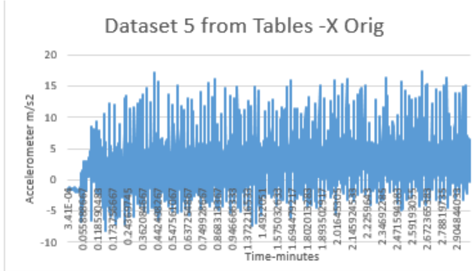
Two activity recognition applications were installed on the Nexus S phone to capture the mobile user's activity when walking [11][12]. One application was installed on a Galaxy S7 phone to capture the user's walking data at the same time that the Nexus S phone was recording data. Pictures 1 and 2 are sample screen

This material is based upon work supported by the National Science Foundation under awards 1241768 and 1637291.

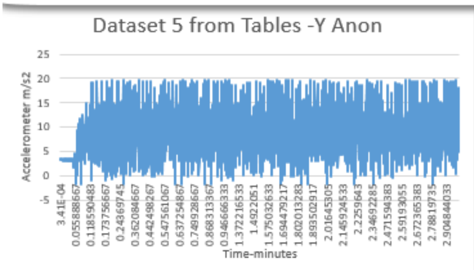
shots from the Galaxy S7 and Nexus S phones, respectively, when capturing data for data set 7 in Tables 1, 2 and 3. Referencing these pictures, it can be seen that both phones started recording the mobile user's walking data at the same time for, approximately, the same amount of time.



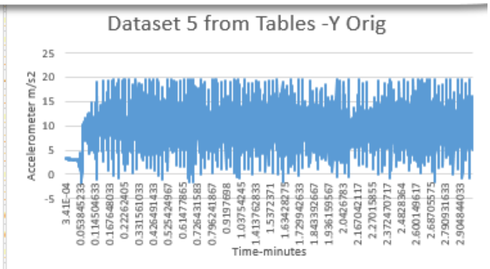
Graph 1



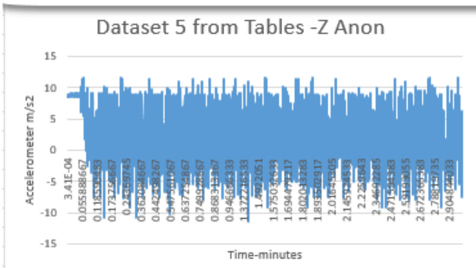
Graph 2



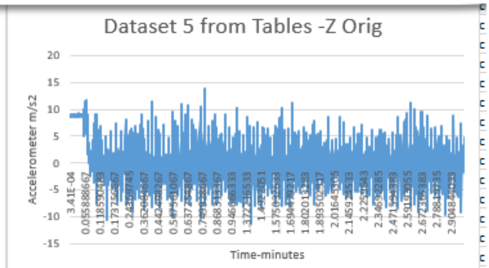
Graph 3



Graph 4

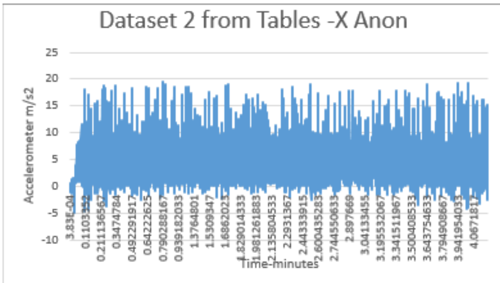


Graph 5

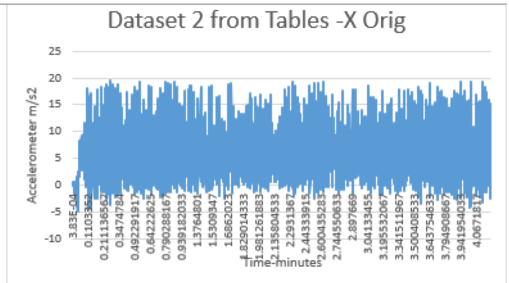


Graph 6

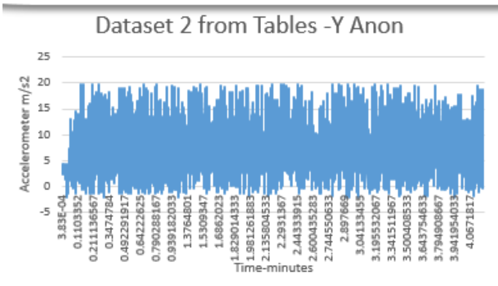
This material is based upon work supported by the National Science Foundation under awards 1241768 and 1637291.



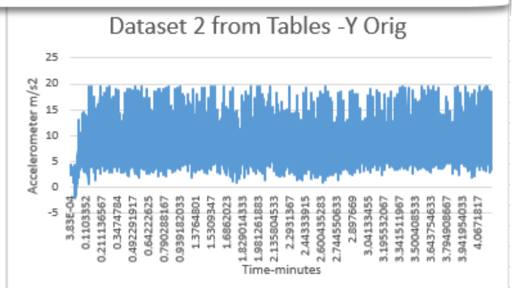
Graph 7



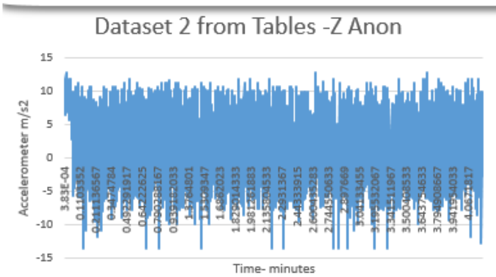
Graph 8



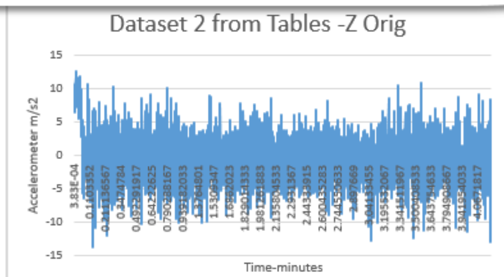
Graph 9



Graph 10

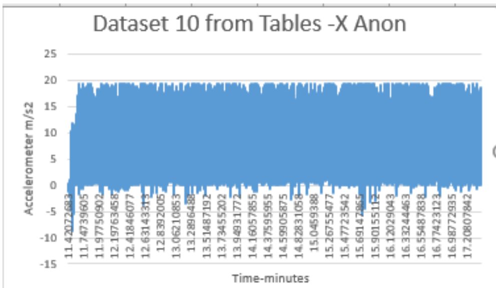


Graph 11

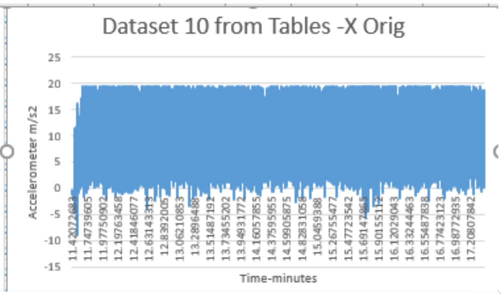


Graph 12

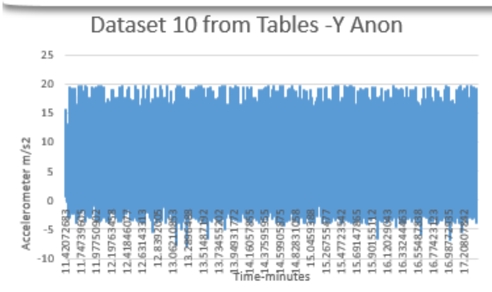
This material is based upon work supported by the National Science Foundation under awards 1241768 and 1637291.



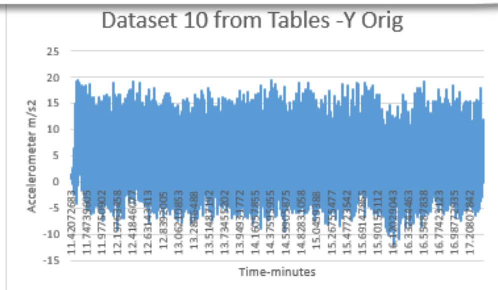
Graph 13



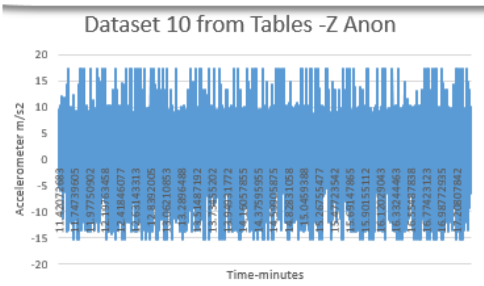
Graph 14



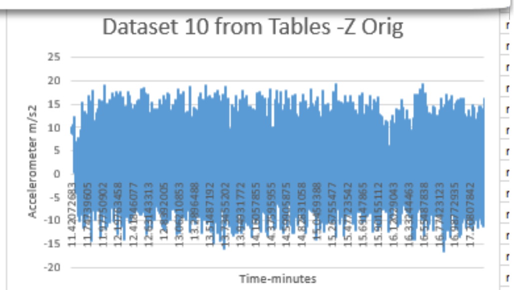
Graph 15



Graph 16



Graph 18



Graph 18

Graphs 1-18 show how the anonymized data for data sets 2, 5 and 10 look when compared to the data of original counterparts. The graphs show that, even though the Max and Min values in Tables 1 – 3 are sometimes equal to each, how

This material is based upon work supported by the National Science Foundation under awards 1241768 and 1637291.

significantly the anonymized data for the Y and Z axes differ from that of the original Y and Z axes data.

In addition, the mobile user recorded running for approximately 2 minutes and driving a vehicle for 3 minutes. These activities were recorded by the activity applications installed on the Nexus S phone and anonymized by the *SparCTym* algorithm.

4. CONCLUSION

From the results, we have shown with the implementation of the *SparCTym* algorithm in the Android operating system, that we were able to anonymize accelerometer walking data while maintaining the utility of the data. The entropy and mean of the original data significantly changed when the data were anonymized. Also, correlation between the original and anonymized Y and Z data was very small. Low correlation is an indication that, potentially, sensitive data has been obscured.

The initial detection of walking activity by the applications on the Nexus S was slower than that of the Galaxy S7. However, once detected, subsequent activities registered in a reasonable amount of time.

The CPU (1 GHz single-core ARM Cortex-A8) on the Nexus S could be a factor as to why its initial recognition of an activity was slower than that of the Galaxy S7 (Smapdragon 820/Exynos 8890 with 4 GB of RAM).

REFERENCES

[1] N. Ravi, N. Dandekar, P. Mysore, and M. Littman. 2005. "Activity recognition from accelerometer data. In Proceedings of the 17th conference on Innovative applications of artificial intelligence" - Volume 3 (IAAI'05), Bruce Porter (Ed.), Vol. 3. AAAI Press 1541-154s

[2] J.Kwapisz, G. Weiss and S. Moore, "Activity recognition using cell phone accelerometers," in SensorKDD '10, July 25, 2010.

[3] C. D´ıaz, and S. Seys, and J. Claessens, and B. Preneel, Towards measuring anonymity, Dingledine and Syverson (Eds.), Designing Privacy Enhancing Technologies, LNCS 2482, pp. 54-68, 2002

[4] T.B. Singha, R.K. Nath, & A.V. Narsimhadhan. "Person Recognition using

This material is based upon work supported by the National Science Foundation under awards 1241768 and 1637291.

Smartphones' Accelerometer Data." *CoRR* abs/1711.04689 (2017).

[5] A. Raij, A. Ghosh, S. Kumar and M. Srivastava, "Privacy Risks Emerging from the adoption of innocuous wearable sensors in the mobile environment," CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

[6] S. Chakraborty, K. Raghavan, M. Johnson and M. Srivastava, "A framework for context-aware privacy of sensor data on mobile systems," ACM HotMobile'13, February 26–27, 2013, Jekyll Island, Georgia, USA.

[7] K. Raghavan, S Chakraborty, M. Srivastava and H. Teague, "Override: A mobile privacy framework for context-driven perturbation and synthesis of sensor data streams," PhoneSense'12, November 6, 2012, Toronto, ON, Canada.

[8] T. Brezmes, J. Gorricho, J. Cotrina, A. Raij, M. al'Absi, S. Shah, S. Mitra, T. Kwon and J. Jeong. - "Activity Recognition from Accelerometer Data on a Mobile Phone." S. Omatu et al. (Eds.): IWANN 2009, Part II, LNCS 5518, pp. 796–799, 2009.

[9] C. Claiborne, R. Dantu and C. Ncube. "Random Anonymization of Mobile Sensor Data" Intelligence and Security Informatics (ISI), 2015 IEEE International Conference. 27-29 May 2015.

[10] Shdefat AY, Halimeh AA, Kim HC (2018) "Human Activities RecognitionVia Smartphones Using Supervised Machine Learning Classifiers." Prim HealthCare 8: 289. doi: 10.4172/2167-1079.1000289.

[11] "Activity" simpleprojects.org

[12] "Physical Activity Recognition" pop.studio.hk@gmail.com

This material is based upon work supported by the National Science Foundation under awards 1241768 and 1637291.