Random Anonymization of Mobile Sensor Data

Modified Android Framework

Cynthia L. Claiborne
University of North Texas
Denton, USA

Ram Dantu
University of North Texas
Denton, USA

Cathy Ncube
Pensacola, USA

***Abstract*— With the increasing ability to accurately classify activities of mobile users from what was once viewed as innocuous mobile sensor data, the risk of users compromising their privacy has risen exponentially. Currently, mobile owners cannot control how various applications handle the privacy of their sensor data, or even determine if a service provider is adversarial or trustworthy. To address these privacy concerns, third party applications have been designed to allow mobile users to have control over the data that is sent to service providers. However, these applications require users to set flags and parameters that place restrictions on the anonymized or real sensor data that is sent to the requestor. Therefore, in this paper, we introduce a new framework, RANDSOM, that moves the decision-making from the application level to the operating system level.**

***Keywords*—privacy; provider; pervasive; Android; anonymized; RANDSOM; smart phone**

I. Introduction

In today's society, data privacy is an evolving research discussion and effort. Even more, privacy concerns are greater than ever because of the ease of global accessibility to electronic devices which contain a variety of common sensors.

As noted in [3] and [4], seemingly innocuous sensor data can be used to accurately identify the activities of cell users. Activity identification may be acceptable by mobile users in cases of research, for medical purposes or even to get a lower insurance rate. However, a mobile user may, unknowingly, compromise privacy by communicating with adversarial applications.

One way in which raw sensor data is used to help determine a person's activities is by labeling the data through a method called supervised training. Supervised training uses labeled data to build models that can recognize simple activities, such as sitting, walking, standing and jogging, of individuals through activity recognition algorithms.

Once activity recognition models are built, more sensor data is collected while individuals are doing the same activities. The additional data are used to test the accuracy of the activity recognition models.

Following are some common activity recognition algorithms and methods: [5] Hidden Markov Model, Naïve Bayes classifiers, Conditional Random Field, Skip Chain, Conditional Random Field, Emerging Pattern and Decision Trees.

As outlined above, activity recognition is a common method used for determining simple activities of an individual. If desired, this process can be applied toward monitoring a mobile user's movements throughout the day. However, activity recognition also increases the privacy risks of unveiling activities that mobile users do not want to reveal. Thus, we propose a modification to the Android operating system framework, RANDSOM, which randomly anonymizes raw sensor data based on permissions granted to an application. RANDSOM serves three purposes.

A. *Pervasively anonymize data with no user interface.*
B. *Pervasively anonymize data for applications that are considered to be in a state of "high-security-risk".An application is in a state of "high-security-risk' if permission from at least 3 of the following categories is requested.*
   - WIFI connectivity - (Normally checked before making a data transfer.)
   - Internet – (Needed for network connectivity.)
   - Network state – (Check network availability.
   - Write external storage – (Large files are usually written to storage before a data transfer.)
C. *Maintain data utility for "non-high-security-risk" applications.*

II. Related Work

The study in [9] provides mobile sensor data that can be utilized by requesting applications, while protecting the privacy of users. It proposes the privacy framework, IPShield, which accomplishes two major tasks:

- *Define two lists of possible inferences which are known as the ''Blacklist", inferences that the user would like to prevent, and the "Whitelist", inferences that the application is requesting.*
- *Create a graphical adversarial model to unveil how much an application initially knows about a user and as future data is collected, reveal any change in an application's knowledge of the data provider. This graphical model helps determine the level of distortion performed on data selected for transmission to applications.*

This research is supported by the National Science Foundation under grants 0751205 and 1229700.

978-1-4799-9889-0/15/$31.00 ©2015 IEEE

According to an application’s knowledge of a user, there are 3 types of data distortion that can be performed: *Suppression*- data is not revealed; *Perturbation*- noise is introduced to the data before trafficking it to the requestor; *Synthesis*- Supplying the requestor with data not related to the actual sensor data. This data is also called “Synthetic” data.

Another study [2] is centered on modifying the Android framework as a way to help balance data utility and user privacy with the current, rapid, growth of context-aware applications. Context-aware applications have the ability to not only provide useful services for mobile users, but they also, via sensor data, are capable of making unwanted medical, personal, social and physiological inferences about the state or activities of mobile customers.

A modification to the Android framework called “Override” is presented as a solution to this issue. “Override” securely intercepts raw sensor data requested by applications and either perturbs it, according to rules setup by the user, or replaces it with synthetic data before releasing it to the requestor.

Each of the preceding techniques requires user interaction. Our approach is different in that it is pervasive. The dependency on the interaction of a mobile user with an anonymization application is eliminated by bringing the decision to anonymize sensor data to the Android framework.

The remaining sections unveil the methodology and test results after implementing the RANDSOM framework.

III. METHODOLOGY

Currently, accelerometer and GPS data are the most commonly used sensor data in activity recognition. However, the RANDSOM framework offers more privacy to mobile phone users, during activity recognition, by randomly anonymizing data from four sensors: accelerometer, gyroscope, orientation and geomagnetic field (magnetometer). The data from all four sensors are anonymized because some mobile sensors or interdependent. For instance, if the orientation sensor data are not anonymized, then they might reveal the true values of the accelerometer and the magnetometer sensors [8]. The orientation sensor derives its data from these two sensors.

The RANDSOM framework consists of three major functions:

1) Determine if at least 3 of the following permissions are requested by an application: WIFI connectivity, INTERNET permission, NETWORK state and WRITE external storage. This is done by setting a unique app level system variable for each permission requested.. [1]

2) Once 3 of the permissions in “A” have been requested, set the “anony_time” flag to ‘Y’.

3) Once, the “anony_time” flag is set to “Y”, start sending anonymized data to an application.

- Calculations involved in RANDSOM’s anonymization algorithm are shown in Fig. 1.

A. Obfuscate obvious influences of gravity and the geomagnetic field.

The force of gravity will show on the x, y and z axes when a mobile phone is at rest [8]. This fact can assist an adversarial entity who may desire to use gravitational information to compromise sensor data. Therefore, in order to secure data, 9.80665 m/s^{2}, the Earth’s gravity, is subtracted from the axis of the accelerometer sensor data that has the greatest magnitude. Doing so will cause the values on the x, y and z axis to be more uniform.

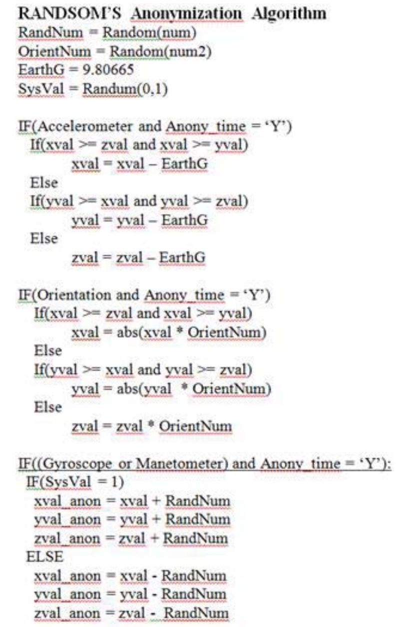
Relative to the magnetic field, the x, y and z values for orientation can vary, significantly, within the range of -180 to 270. Thus, based on this knowledge, the absolute values of anonymized orientation data are taken to add even more anonymity to the orientation data.

B. Randomly select a value by which to anonymize data.

In the RANDSOM framework, there is no dependency on the mobile user to choose an anonymization factor to alter sensor data. Instead, RANDSOM randomly generates these values which are called random factors. Moreover, the application of randomly produced anonymization factors to the mobile sensor data, decreases the probability that the data would be trained for illicit/undesired activity recognition

C. Randomly generate the application of the random factor on the raw sensor data.

To add more unpredictability to deriving the activities of mobile users from raw sensor data, the application of the random factor on raw sensor data is also randomized. This is done by setting a randomly generated system variable. [1] If the system variable is ‘1’, the random factor is added to the raw sensor data. If the system variable is ‘0’, the random factor is subtracted from the raw sensor data.



```
RANDSOM'S Anonymization Algorithm
RandNum = Random(num)
OrientNum = Random(num2)
EarthG = 9.80665
SysVal = Randum(0,1)

IF(Accelerometer and Anony_time = ‘Y’)
  If(xval >= zval and xval >= yval)
        xval = xval – EarthG
  Else
  If(yval >= xval and yval >= zval)
        yval = yval – EarthG
  Else
        zval = zval – EarthG

IF(Orientation and Anony_time = ‘Y’)
  If(xval >= zval and xval >= yval)
        xval = abs(xval * OrientNum)
  Else
  If(yval >= xval and yval >= zval)
        yval = abs(yval  * OrientNum)
  Else
        zval = zval * OrientNum

IF((Gyroscope or Manetometer) and Anony_time = ‘Y’):
 IF(SysVal = 1)
  xval_anon = xval + RandNum
  yval_anon = yval + RandNum
  zval_anon = zval + RandNum
 ELSE
  xval_anon = xval - RandNum
  yval_anon = yval - RandNum
  zval_anon = zval -  RandNum
```

Fig. 1 RANDSOM Framework’s Anonymization Algorithm

IV. RESULTS

Using the new RANDSOM framework, sensor data for applications that reached a "high-security-risk" state were effectively anonymized. The following graphs are a sampling of accelerometer data from two applications executed on the RANDSOM framework.

Fig. 2 and Fig. 3 show real and anonymized accelerometer data, respectively, for a "high-security-risk" application, App1. This application checks for WIFI connectivity and requests INTERNET permissions, the state of the NETWORK and WRITE permissions to storage. Thus, the accelerometer data are anonymized for App1.

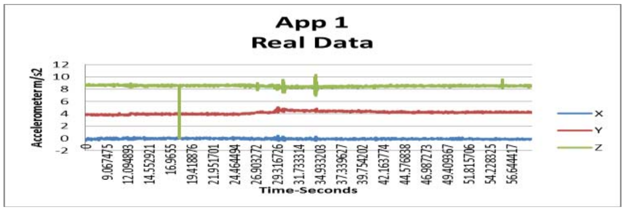


Fig. 2

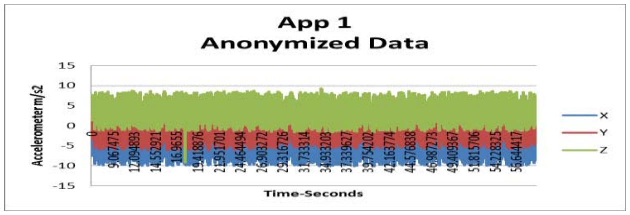


Fig. 3

The accelerometer data for another application, App2, are depicted in Fig. 4 and Fig. 5. Fig. 4 displays the real data for this application. Fig. 5 is the non-anonymized data, after executing App2 on the RANDSOM framework. The application only requests WRITE permissions to storage. Thus, its sensor data is not anonymized.

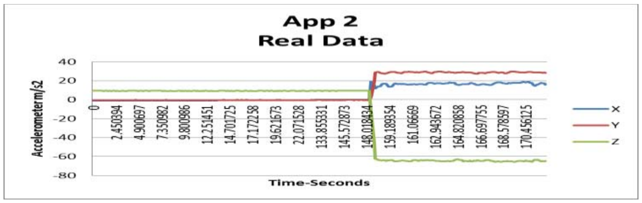


Fig. 4

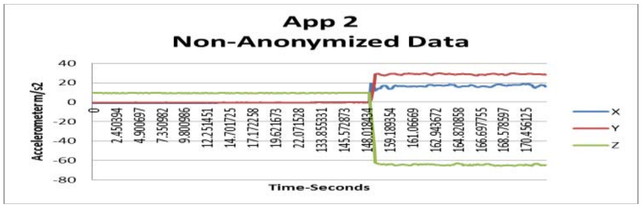


Fig. 5

The RANDSOM framework was implemented in Android version 4.1.2 and tested with "sensor" applications that read and display raw sensor data for one or more sensors.

Data anonymization did not affect the functionality of an application. This includes applications that simultaneously display sensor data for more than one sensor.

Because they never reached a "high-security-risk" state, 84% of the applications tested on the RANDSOM framework, had their data utility preserved. This was accomplished without the necessity of user intervention or involvement.

V. FUTURE WORK

Future work will, possibly, focus on expanding this preliminary study to include analyzing the effect of other permissions on sensor data anonymization and utility. Also, the identification of "high-security-risk" applications that process streaming data will be explored.

VI. CONCLUSION

Our results show that the RANDSOM framework pervasively produced anonymized sensor data for applications in a state of "high-security-risk". Thereby, preserving data utility for applications in a "non-high-security-risk" state.

One of the cons of not including user intervention in determining data anonymization is that the mobile user is limited in deciding which applications will display real or anonymized data.

REFERENCES

[1] C. Claiborne, M. Fazeen, and R. Dantu, "Android sensor data anonymization," In S.J. Stolfo, A. Stavrou, and C.V. Wright (Eds.): RAID 2013, LNCS 8145, pp. 469–471, 2013.

http://www.springer.com/computer/security+and+cryptology/book/978-3-642-41283-7

[2] 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.

[3] 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.

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

[5] E. Kim, S. Helal, and D.Cook, "Human activity recognition and pattern discovery IEEE Pervasive Computing. 2010;9(1):48–53.

[6] L. Sankar, S. Rajagopalan and H. Poor, "A theory of utility and privacy of data sources," Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on 13-18 June 2010, Austin, TX.

[7] E. Ertin, N. Stohs, S. Kumar, A. Raij, M. al'Absi, S. Shah, S. Mitra, T. Kwon and J. Jeong. "AutoSense: unobstusively wearable sensor suite for inferring the onset, causality and consequences of stress in the field", SenSys'11, November 1–4, 2011, Seattle, WA, USA.

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

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