Privacy-preserved Mobile Sensing through Hybrid Cloud Trust Framework

Joy Ying Zhang†, Pang Wu†, Jiang Zhu†‡
Hao Hu‡, Flavio Bonomi‡
†Department of ECE, Carnegie Mellon University, Moffett Field, CA 94035
‡AA&R, Cisco Systems, San Jose, CA 95134
Email: {pang.wu, jiang.zhu, joy.zhang}@sv.cmu.edu; {hahu2, flavio}@cisco.com

***Abstract*—Mobile sensors embedded in smart phones and smart buildings enable mobile sensing and users' behavior modeling and thus open the doors for edge-cutting applications such as personalized intelligent computing, activity prediction, health/wellbeing monitoring and behavior intervention. One critical obstacle in mobile sensing and behavior modeling is privacy. Users do not always trust the public cloud to store and process their detailed personal data. In this paper, we propose a novel approach of using hybrid cloud to distribute the computing among mobile devices, personal cloud and public cloud. Raw sensor data is stored within the personal cloud where users have full, physical control. User can authorize analytic widgets (e.g., health monitor or marketing survey) to only collect user-approved data. We demonstrate that with this approach, users' privacy anxiety is significantly reduced and the acceptance rate of the mobile sensing technology increases from 23% to 60%.**

I. Introduction

With the advent of mobile sensing technology, collecting large scale of personal specific data is now becoming possible. Mobile sensing Apps such as the CMU MobiSens [2] and the MIT Funf[1] run as services in the background and can constantly collect sensor information from smart phones. These "sensors" can be either "hard" sensors (e.g., accelerometers) that are physical sensing devices or "soft" sensors that records information of a phone's running status (e.g., screen on/off). Typical sensor information include GPS, location, WLAN, accelerometers, gyroscope, magnetometer, bluetooth, cell tower ID, call log, SMS log, browsing history, microphone, camera, contacts, running apps, apps' network communication pattern, screen on/off state, battery status and so on. The Big Data generated by such detailed and 24/7 sensing can scale up to several Gigabytes per day for a user.

From this personal Big Data, it is easy to infer user's home address, office location, when and how the user usually go to work, who did the user talk to at a meeting, where did the user go for vacation and much detailed personal information. Through statistical modeling over the sensor time-series, we are able to infer behavior patterns of the user such as their outdoor [3] and indoor mobility pattern [4], mobile phone usage patterns [5] and even their stress level [6].

Revolutionary applications are made possible based on such personal, fine-grained, Big Data, for example, anomaly detection, passive authentication through behavior modeling, personalized intelligent computing, future activity prediction, trend forecasting, psychological status estimation, health behavior monitoring and intervention. However, such personal data, if not protected or used properly, has serious privacy consequences.

The consequences of potential privacy leaks hinder the wide adoption of mobile sensing technologies despite the benefit users would get. Preserving the privacy of the user while sensitive data is stored and processed in the public cloud is a non-trivial issue. According to [1], "privacy" is users' expectation of their private information on where the information resides, who has access to the information and how the information is used. Users become "anxious" of their private data when they are unsure who may have access to their personal data after being uploaded to the server and how it will be used or shared.

In this paper, we propose a novel sensing framework using hybrid cloud combined with a trusted third-party certification.

II. Hybrid Cloud Trust Framework

In this framework, raw sensor data is uploaded directly to users' own personal cloud and a user has physical and complete control of his/her own personal big data. Personal Big Data consumers such as health organizations and auto insurance companies can develop *data widgets*, applications to be run on users' Personal Cloud, to learn behavioral patterns (e.g., average daily commuting time) from the raw sensor data, which are then uploaded to the public cloud for different applications. A third-party certificate provider certifies such a widget and inform end users what statistics it is going to generate and then upload. In this framework, users own their data. No raw data is exposed to "data consumers". It is up to users' discretion to "sell" their statistics to "data consumers" by enabling the corresponding widgets. (Figure 1).

In this hybrid cloud approach, computation is distributed among mobile devices, personal cloud and public cloud. Raw sensor data is stored on user's personal cloud only where user has full physical control of his/her own data. We believe by doing this, users feel that the data is under control with full knowledge of where the data is stored and managed and will thus have less privacy-anxiety. We surveyed 55 users through online questionnaires[2]. 50 out of 55 users (91%) have smart phones and 48 of them (87%) are aware that smart phones "have sensors that can track your activities such as your location, your motion, your sound". Only 13 users (23.6%) are willing to share the raw sensor data to would-be personal information consumers such as car insurance companies if they are paid. However, with the proposed hybrid cloud solution, 33 users (60%) are comfortable using the mobile sensing software.

Information consumers such as health monitor organizations; car insurance companies and utility companies can develop application widgets to extract statistics from users' sensor data. A widget 1) declares what information is going to be extracted from users' data (e.g., "this widget will estimate your average commuting distance from home to work"); 2) is downloaded and

[1]http://funf.org/

[2]http://goo.gl/7TwC6

978-0-7695-5028-2/13 $26.00 © 2013 IEEE
DOI 10.1109/CLOUD.2013.108

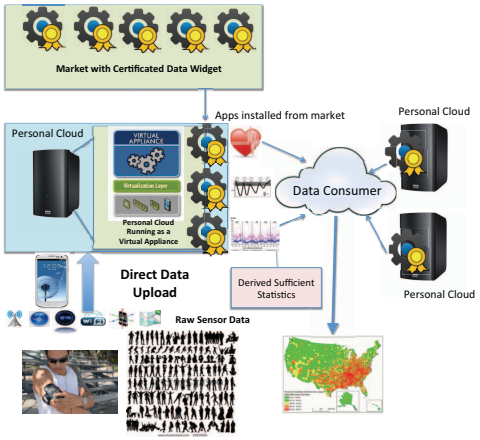

Fig. 1. Architecture of Privacy-preserved Mobile Sensing and Behavior Modeling framework using Hybrid Cloud Computing.

installed to user's personal cloud once a user grants information access to the information consumer; 3) runs on the personal cloud and process the raw sensor data using the data analytics API; and 4) uploads the extracted analytics to public cloud.

Information consumers aggregate statistics collected and derive behavior models for their own applications.

III. System Design

A. Private Cloud

On of the contribution of the proposed framework is decentralized data aggregation. Unlike most the conventional sensing framework which has a centralized server for user data aggregation, each user uploads direct raw data to his/her private cloud. Without the user's permission, raw data will stay on the private cloud without any leakage. In our approach, private clouds are virtual machines deployed on users' own physical machines. Each virtual machine has it's built-in sensor data management system. It provides a web UI for users to manage their raw data; create/delete user specific sensing profile and authorize feature extraction app developed by the third parties through an authorization procedure(III-C).

Data Chunk: The raw data are split into data chunks before being uploaded to the private cloud. Each chunk contains a certain amount of sensing data with corresponding timestamp when collected. In our implementation, we use 1 hour as the chunk split period. Related information, i.e., start/end timestamp, upload time, location and the owner device of the data chunk file, is stored in a MySQL database. The private data cloud thus can be viewed as an indexed file storage system.

B. Data Consumer and Data Widget App

Data Consumers are third parties (trusted or untrusted) who want to leverage the statistics from user's sensor data. They cannot access users' raw data without users' permission in our proposed framework. They can develop feature extraction application, called *Data Widget*, which can run locally in the private cloud to extract the desired statistics and transmit the results to the cloud, only after authorized by the user. The benefit of running data widget in the private cloud is two-fold: 1) *Scalability* the computation is distributed to private clouds. 2) *Privacy* the users have full control in what information to provide.

When publishing a data widget, the data consumer has to explicitly inform users the following *Sensitivity Level Description (SLD)* : 1) which part of the raw data is used for feature extraction. 2) What kind of extracted feature will be further transmitted to the data consumer. SLD is certified by a trusted third party to guarantee it aligns with the widget's actual behavior.

C. Trusted Third-party Data Widget Certification and App Market

Trusted third-party data widget certification is key to the framework. We have two observations: 1) Users usually do not have the knowledge to identify the features that could be reverse engineered to reveal the personal identity. 2) Data consumers release widgets that behave differently from the SLD (a *Data Malware*). But, the third party should share no interest with all data consumers (e.g., Apple and all iOS app developers). It defines different sensitivity levels and categorizes data widgets into different levels. This trusted third-party also needs to make sure the behavior of a data widget agrees with the widget's SLD.

Upon verification, the widget will be certificated by the third-party and published in the *Widget Market*. The Widget Market is a virtual place that allows the user to view and install data widgets. The trusted third-party operates the market and maintains its trustworthy. In this design, users trust the *behavior* described by the SLD through an independent trusted party.

IV. Conclusion

In this paper, we propose a hybrid cloud trust framework to address the privacy concern of mobile sensing systems. By introducing a trusted third-party to certificate the behavior of applications developed by data consumers, the framework can increase the difficulty of publishing data malware. We believe this trust paradigm is suitable for addressing the privacy concerns in behavior modeling mobile sensing systems.

Acknowledgment

The project is supported in part by DARPA/NAVY DCAP project and by Northrop Grumman Cyber Security Consortium and Cisco Systems Inc.

References

[1] D. Norman, *The Design of Everyday Things*. Basic Books, 2002. [Online]. Available: http://books.google.com/books?id=w8pM72p_dpoC

[2] P. Wu, J. Zhu, and J. Zhang, "Mobisens: A versatile mobile sensing platform for real-world applications," *Mobile Networks and Applications*, vol. 18, pp. 60–80, 2 2013. [Online]. Available: http://dx.doi.org/10.1007/s11036-012-0422-y

[3] S. Buthpitiya, Y. Zhang, A. Dey, and M. Griss, "n-gram geo-trace modeling," in *Proceedings of Ninth International Conference on Pervasive Computing*, San Francisco, CA, June 12-15 2011.

[4] J. Zhu and Y. Zhang, "Towards accountable mobility model: A language approach on user behavior modeling in office wifi networks," in *Proceedings of The IEEE International Conference on Computer Communications and Networks (ICCCN 2011)*, Maui, Hawaii, July 31 - August 4 2011.

[5] J. Zhu, P. Wu, X. Wang, A. Perrig, J. Hong, and J. Y. Zhang, "Sensec: Mobile application security through passive sensing," in *Proceedings of International Conference on Computing, Networking and Communications. (ICNC 2013)*, San Diego, CA, USA, January 28-31 2013.

[6] J. Zhu, Y. Tian, S. Hu, X. Hu, E. LaRue, and Y. Zhang, "Human stress evaluation through passive sensing and behavioral modeling," in *Proceedings of 2013 AMIA Summit on Clinical Research Informatics (CRI)*, San Francisco, CA, 2013.