A Provenance Framework for mHealth

Aarathi Prasad[1], Ronald Peterson[1], Jacob Sorber[2], and David Kotz[1]

[1]Department of Computer Science, Dartmouth College
[2]School of Computing, Clemson University

1 Provenance in mHealth

Consider Jane, who is using one or more mHealth devices, continuously or periodically collecting her health-related information into her mobile phone. The phone periodically uploads this information, along with other health-related information that Jane manually inputs to her phone, to her electronic health record (EHR). Jane can then share her health information with her health providers, family and friends, peers, employer, insurer, and researchers. But how can these data consumers know whether to trust the sensor-collected and human-entered data they receive? What confidence do they have that it is accurate and authentic?

Since personal and home-use devices are not maintained regularly, like hospital devices, Jane might not realize when her devices are malfunctioning or uncalibrated. If other people in her household use similar devices, they might accidentally confuse their own device for Jane's. mHealth devices, being mobile, can be stolen or misplaced – and then used by another person. Jane, or her caregivers, can deliberately fake or hide data collected using mHealth devices, particularly if there is an incentive to do so. In all these scenarios, the data collected and shared by the devices might be inaccurate, or about the wrong person. If so, its use could prove damaging or even fatal, especially if the data is used by health providers for diagnosis or treatment.

Data recipients might be able to verify the accuracy and authenticity of the data if they have information about its origin and about changes made to it, i.e., the *provenance* of the data. Previous work has looked into provenance in mobile healthcare. For example, Medially is a middleware for remote health monitoring, which supports collection of contextual provenance for reconstructing context to interpret the data streams [1].

Consider a specific scenario. Devi, a health worker, visits pregnant women in a village in India every week with a sensor kit that has mHealth devices like a blood-pressure cuff, heart-rate monitor, fetal monitor, spirometer, smoke sensor and weight scale. The pregnant women are enrolled as patients in the village health clinic and the data collected by Devi is uploaded to their patient records in the clinic's electronic health record system. Dr. Ravi is alerted by the electronic record system that Ritu, one of his patients, has had a gradual decrease in her lung capacity over a period of four weeks. He suspects that Devi's spirometer was not working correctly, but then he notices that the smoke measurements Devi collected at Ritu's house show a strong presence of nicotine at her home. The system also attests to the fact that all the devices Devi was using were working correctly and the spirometer readings were indeed from Ritu herself.

We define provenance in mHealth as contextual information that can attest to the authenticity and accuracy of the data and can help the recipient in interpreting the data. Provenance information could include who was using the sensor, what sensor was used, where the sensor was placed, when and where the data was collected, why it was collected, how it was collected and under what circumstances. In the above scenario, the spirometer measurement came with important contextual information (airborne nicotine, device calibration, and patient identity) for interpreting the lung-capacity data.

2 A Framework for Provenance

To realize this vision, we propose a provenance framework for mHealth. The primary function of the framework is to collect and share provenance metadata and help the data consumer verify whether certain provenance properties are satisfied by the data they receive. An mHealth *developer* uses the provenance framework in constructing an mHealth application. The developer works with a *domain expert*, i.e., someone who is familiar with the medical requirements of the application, to define the provenance properties desired by each class of data consumers (clinical staff may desire different properties than, say, family members), and to identify the provenance metadata required to verify these properties. Using tools in our framework, the developer and expert collaboratively prepare a *provenance model* for each class of data consumer in their application; other tools use this model to generate code for a provenance service to be installed on the patient's mobile phone. The model defines *rules* that determine how and when the data and metadata should be collected and when the patient or recipient should be alerted to

mHealthSys'12, November 2012, Toronto, ON, Canada.
Copyright © 2012 ACM 978-1-4503-1169-4 ...$10.00

possible problems with the data. Consider examples:

- A rule for collecting data: If the peak expiratory flow (PEF) occurs at a volume larger than 0.7 L, then collect another spirometer reading [2].
- A rule for collecting metadata: To interpret lung-capacity data, a doctor needs to know about the presence of smoke around the patient, and pollen count in the air.
- A rule for notifying recipient: If patient's lung capacity decreases over a period of four weeks, alert the doctor.

Patients, or in our scenario, health workers like Devi, install the mHealth app on their phone. During installation, the provenance service seeks to pair with specific types of devices that collect the metadata specified by the provenance model. There is a link between the app and the service, such that the service knows what measurement is being taken by the app. When the mHealth app collects the patient's health information, the provenance service collects metadata from the paired devices. The data collected by the mHealth app and the metadata collected by the provenance service are cryptographically bound and sent to the patient's electronic health record (EHR).

When Devi collects Ritu's lung capacity reading, the smoke sensor measures airborne nicotine and the phone connects to a weather website to retrieve the pollen count in the neighborhood. All these readings are cryptographically bound by the provenance service along with information about the mHealth devices (and weather forecast website), Devi and Ritu's identities, and time and location of Devi's mobile phone. This package is sent to Ritu's electronic record in the hospital's EHR database.

To make best use of provenance information, the clinic's EHR system should be aware of the provenance information and be able to display it to users like Dr. Ravi. We envision presenting him with a *confidence metric* that represents how well the data satisfies the provenance properties desired by his class of recipients (clinical staff). The provenance service computes confidence using the metadata and rules provided by the domain expert during the modeling process.

In another use, an auditor can later verify the authenticity of a patient's records. The auditor can leverage the provenance information, and confidence metrics, to mark certain records as suspect or invalid. The audit may examine a history of records from a given patient, sensor device, or caregiver; if there is a pattern of concern (e.g., a particular device has become uncalibrated, or a caregiver is suspected of poor application of the sensors) this information can be used to mark all relevant records suspect.

Data and metadata can be voluminous. In low-resource environments, bandwidth may be insufficient to upload all of it to the EHR. Data and metadata will be saved on the phone, for days or weeks, until there is sufficient connectivity available to transfer the full information to the patient's records. Meanwhile, the phone uploads summary information to the record, for immediate use. Later audits can compare the summary and the raw data at the server side to check whether the condensed version of the data is an accurate and authentic summary of the detailed version.

3 Challenges

As we pursue this vision we recognize many challenges.

A provenance framework needs interoperability between mHealth devices and EHR systems. We will explore ways to leverage existing standards for mHealth and EHR systems, but new standards may be needed.

mHealth devices will not be able to collect all types of health information, especially psychological information, automatically. Such information will have to be obtained as manual input, but verifying self-reported data is hard. Trend analysis using statistical techniques might be one approach to finding problems in manual input.

We need to develop a general approach to computing a confidence metric from data and metadata.

The framework collects provenance metadata to attest to the accuracy of the health data. But how can we determine whether the provenance metadata is itself accurate?

How do we reliably determine the identity of the patient, or person being sensed? Identity could be verified using basic biometric techniques like face and speech recognition.

How can the domain expert decide how much metadata is sufficient for the recipient to verify their desired provenance properties? One alternative is to collect information from all available metadata sources – a big challenge in environments with resource-constrained mobile devices. Another approach could be to provide a feedback loop from the data consumer to alter the rules of data collection.

How should the data, metadata, and confidence values be presented, clearly, to various recipients?

We are exploring different ways to address the above challenges and we expect our framework to provide extensive techniques to help data consumers verify whether desired provenance properties hold for the data they receive.

Acknowledgments

We wish to thank Ashutosh Sabharwal and the Dartmouth TISH group for their continued feedback.

This research results from a research program at the Institute for Security, Technology, and Society at Dartmouth College, supported by NSF under award number 0910842, award number CNS-1143548, and by the Department of HHS (SHARP program) under award number 90TR0003-01. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the sponsors.

4 References

[1] A. R. Chowdhury, B. Falchuk, and A. Misra. Medially: A provenance-aware remote health monitoring middleware. In *IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 125–134. IEEE, Mar. 2010. DOI 10.1109/PERCOM.2010.5466985.

[2] S. Gupta, P. Chang, N. Anyigbo, and A. Sabharwal. mobileSpiro: accurate mobile spirometry for self-management of asthma. In *Proceedings of the First ACM Workshop on Mobile Systems, Applications, and Services for Healthcare (mHealthSys)*. ACM, Nov. 2012. DOI 10.1145/2064942.2064944.