Communications and Network, **2013, 5, 4-9**
doi:10.4236/cn.2013.53B1002 Published Online August 2013 (http://www.scirp.org/journal/cn)

The Internet of Things and Next-generation Public Health Information Systems

Robert Steele, Andrew Clarke
Discipline of Health Informatics, the University of Sydney, Sydney, Australia
Email: robert.steele@sydney.edu.au, andrew.clarke@sydney.edu.au

Received June, 2013

ABSTRACT

The Internet of things has particularly novel implications in the area of public health. This is due to (1) The rapid and widespread adoption of powerful contemporary Smartphone's; (2) The increasing availability and use of health and fitness sensors, wearable sensor patches, smart watches, wireless-enabled digital tattoos and ambient sensors; and (3) The nature of public health to implicitly involve connectivity with and the acquisition of data in relation to large numbers of individuals up to population scale. Of particular relevance in relation to the Internet of Things (IoT) and public health is the need for privacy and anonymity of users. It should be noted that IoT capabilities are not inconsistent with maintaining privacy, due to the focus of public health on aggregate data not individual data and broad public health interventions. In addition, public health information systems utilizing IoT capabilities can be constructed to specifically ensure privacy, security and anonymity, as has been developed and evaluated in this work. In this paper we describe the particular characteristics of the IoT that can play a role in enabling emerging public health capabilities; we describe a privacy-preserving IoT-based public health information system architecture; and provide a privacy evaluation.

Keywords: Internet of Things; Public Health; Population Health; Privacy; Anonymity

1. Introduction

The Internet of things can find particular applicability in the area of public health. This is because public health is a field where communication with large numbers of individuals is implicitly required, either for data capture or public health intervention. In addition, many of the data inputs required for public health information capture are increasingly available via the proliferation of consumer health and fitness sensors.

The recent rapid growth in both the capabilities and uptake of mobile devices or Smartphone's capable of acting as sensor platforms has the potential to enable a new generation of public health information systems. While increasingly, mobile devices and sensors are used as a tool for individual health data capture, tracking and feedback, the use of such technologies has not to-date substantially extended into use for public health purposes. In addition, the use of sensors for individual fitness and health tracking does not as critically require an IoT infrastructure as does public health, as individual fitness tracking does not necessarily require widespread interconnectivity between many sensors and processors – such individual fitness and health data only strictly needs to be made available to the individual user.

In this paper we describe how an IoT-based architecture can be utilized for population health data capture and public health intervention whilst still maintaining strong privacy and anonymity for all participating individuals. Prior work in relation to achieving privacy and security has relied on a trusted data collector or aggregation process, whereas our approach does not assume this. In addition, interestingly, the case for public health usage doesn't require the same level of precise data that would often be required in other IoT applications. For example the exact location and time of a measured sensor value is less important than the aggregate value over a period of time or the trend of change for a mass of people or community. Public health interventions [1], are a key component of future Health Participatory Sensing Networks (HPSNs) [2], and in our approach we describe capabilities whereby a targeted public health intervention can be distributed, performed and evaluated without the need for identifying details of an individual to ever leave their mobile device.

Also central to our approach is an anonymizng layer [2] within the IoT-based architecture, which utilizes either a MIX network [3] or Onion routing [4]. This anonymizing layer is one of the mechanisms to enforce privacy-preserving public health-related communications.

In Section 2 we discuss the relevant emerging capabilities of the IoT and their relevance and match to public

health goals. In Section 3, we describe the IoT-based public health information system architecture, including the resultant data capture and public health intervention capabilities, in Section 4 we describe privacy and anonymity and in Section 5 provide a privacy evaluation. This is followed by the Conclusion.

2. The Internet of Things and Public Health

In this section we identify novel IoT capabilities and overview their relationship to public health measures. Many public health measures can already be captured automatically via such IoT capabilities.

2.1. Internet of Things Health Sensor Capabilities

The proliferation of commercial fitness and health sensors provides new mechanisms for population health data capture. Emerging sensors also already able to capture many biomedical measures captured in public health data surveys. In addition, these have a number of characteristics quite distinct from traditional survey—based population health data capture approaches.

- Real-time
- Larger participant numbers
- More detailed data
- Captured electronically
- Direct measurement, not human response
- Anonymized

The area of IoT personal health sensor and software development [5] is one of the most active areas of the IoT ecosystem. This is possibly due to the relevance of these individual sensors to both consumer-centric phone technologies and the increasing interest to leverage such technologies for improved personal wellness, health and healthcare [6,7].

Fitness and Activity Sensors

Commercial implementations such as Nike Fuel and Jawbone Up [8] demonstrate the achievability and potential for continuous physical activity sensing. Jawbone Up extends beyond physical activity monitoring to include sleep pattern and quality, and a nutritional diary. Other well-known examples of such sensors include FitBit, RunKeeper, myFitnessPal, Pebble Watch, the Basis Watch and Google Glass. Such fitness and health sensors are the most contemporarily available component of the IoT that can be utilized for public health as such sensors are already achieving widespread interest and adoption.

Also of significant relevance is Google Now's, Activity Summary [9] which provides a monthly estimate of how far an individual has walked and cycled, and comes as part of Google Android – hence is already extremely widely deployed.

Vital Signs Sensors

Smartwaches such as the Mio Active are able to capture heart rate. The Amiigo wristband captures blood oxygen levels, So maxis provides ECG and EMG sensors and the mc10 stretchable electronic tattoo can transmit heart rate and brain activity [5]. The capturing of vital signs is often more beneficial for individual health care, but it also adds new capabilities to public health information systems. As another example, the Sense A/S monitoring patch is able to measure blood pressure [5].

Blood Constituent Sensors

Increasingly there are wireless-enabled patch technologies emerging that may be able to capture the levels of some blood constituents. Examples include the Sano Intelligence [10] wearable patch which is touted to allow the capture of blood glucose and potassium levels, with further blood constituent capture planned to be forthcoming. Numerous continuous blood glucose monitoring systems are also currently available.

Such sensor capabilities in a cheap and accurate form have the potential to revolutionize individual health care, early detection and preventative health; and also public health.

It should be noted that such capabilities may be so beneficial in terms of individual health monitoring and health maintenance that they could achieve wide adoption. If so, their possible role in public health data capture will also be proportionally significant.

Ambient sensors

Other initiatives such as Riderlog [11] and the Copenhagen Wheel [12] are moving towards capturing physical activity levels and at the same time, additional contextual data. The Copenhagen wheel goes beyond physical activity sensing, to urban environment monitoring with air quality and noise sensors included in the implementation to provide additional data beyond just the activity of the individual.

2.2. Public Health Measures

The various types of data that can be collected via the above-mentioned IoT sensors, already relate to a majority of public health measures:

• Physical Activity Levels – This is one of the most important lifestyle factors for chronic health conditions and other health risks [13]. This can now be quite accurately captured with already available sensors and even in-built Smartphone capabilities [8].

• Caloric Burn and Caloric Intake – Caloric burn information can be captured by a range of activity sensors as described, and caloric intake can also be increasingly automatically captured [14].

• Nutritional Data – As mentioned wearable patches have the ability to measure potassium levels, one of the markers of nutrition status [15].

• Blood Pressure – Blood pressure is a public health marker of cardiovascular disease [15]. As described, blood pressure can be captured via a wearable patch such as the Sense A/S.

• Blood Glucose – a marker of diabetes [15] can be captured by wearable patches and other continuous glucose monitoring (CGM) devices.

• Body Mass Index (BMI) – Height is roughly invariant for adults and Bluetooth-enabled scales are increasingly available.

• Sleep Pattern and Regularity – Sleep patterns are both an indicator and a preventative/risk factor for a number of conditions. Sleep quality can be captured by currently available wristband sensors.

3. A Proposed Internet of Things-based Public Health Information System Architecture

The preceding section indicates that even current IoT capabilities have a significant match to many public health measures of interest. We now describe a privacy-preserving public health information system architecture.

3.1. Architecture

The overall system architecture (**Figure 1**) involves one or many central Servers that communicate with mobile devices through a MIX network or Onion routing network to provide communications anonymity, and mobile devices that incorporate local processing and privacy thresholds to maintain data anonymity/privacy/de-identification.

There are two primary data transmissions from and to the Server respectively: 1) data requests and public health interventions are distributed from the server; and 2) anonymized data collection submissions are sent to the server. The core functionality components of the public health system's Server are Data Aggregation, Analysis and Intervention/Data Requests. The Server interfaces with Public Health Groups, which could include state or federal health departments, public health research institutions or other public health organizations.

The fundamental architecture can support different levels of data collection and/or potentially public health intervention, depending largely on the capabilities of the end user's mobile devices and preferences of the individual users of these devices. As described in the previous section these IoT capabilities range from those built into Smartphone to wristband sensors, smartwatches, wearable patches and tattoos and other sensors.

3.2. Anonymous Public Health Data Capture

We have developed our approach such that it does not require a fully trusted server-based approach that would likely prove impractical on population-scale applications. Instead it utilizes an architecture incorporating an anonymous communications network (MIX network or Onion routing) in combination with de-identification of data submitted, to provide anonymous submission/inter- action. However, this alone would still allow the risk of re-identification based on quasi-identifiers, in the form of information known about individuals outside the system that could be used to match with and re-identify the submitted data. The most common approach to address this type of risk is to use a trusted server or aggregation point to combine and obfuscate data to the point where k-anonymity [16] is provided, such that any individual is indiscernible from k other records based on quasi-identifiers.

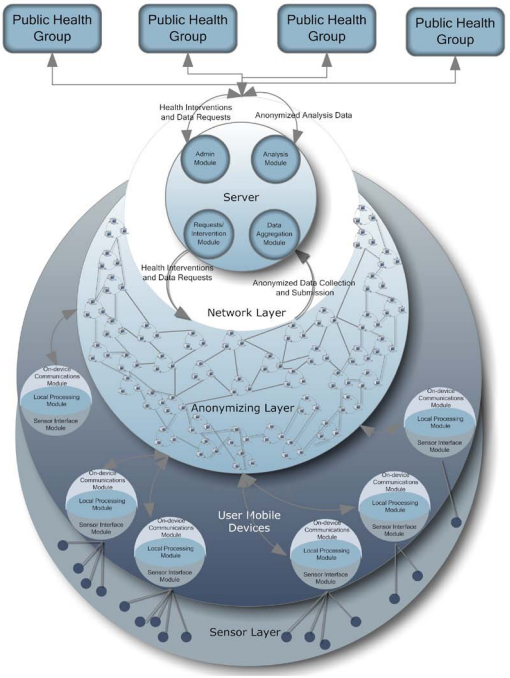

Figure 1. Internet of Things-based Public Health Information System Architecture.

To provide an approach that doesn't require a trusted server component we propose that a suitable level of anonymity can be provided by locally processing on the user's mobile device collected data into an aggregated generalized form that can still meet the purposes of public health data collection, as described in our previous work [17]. By utilizing quasi-identifier scores (QIS) and a threshold approach to privacy limits, the level of privacy disclosure an individual agrees to can be easily managed without requiring a case-by-case approval. Additionally, our approach involves the specification of and weighting of the data to be submitted to allow the local device to alter the resolution and breadth of data submitted to preserve privacy and anonymity, while still submitting the data needed for public health data purposes.

3.3. Anonymous Public Health Intervention Capabilities

A major area of potential usefulness of such an IoT-based public health information system is the ability to distribute targeted or personalized public health interventions to individuals.

Additionally, it seems likely that there will be a number of public health groups (Figure 1) that would be interested in participating in these types of networks and with individuals able to subscribe or opt-in to partake in passive or active participation with each such group.

We propose a novel approach in relation to public health intervention. In line with the local aggregation and processing approach to preserve privacy when submitting sensing data, it appears appropriate to use a similar approach for communication from the public health body to the individual. This novel approach broadcasts larger generalized public health intervention packages from the Server to the entirety or subsets of the participants and then based on local processing, the correct information is displayed or auctioned on individual devices.

This would allow for communication with individuals that could be meaningful and personalized without risk of re-identification of the individual. This approach could also be used for the dissemination of micro-surveys to individuals for additional human-entered data collection.

However, this increases the overhead of data distributed to individuals since in all cases the data required for local processing would need to be received rather than specific targeted data for each individual.

To improve the flexibility of this approach, we propose a technique using verification objects that incorporates a granular approach to hashing and digital signing of distributed content, by including timestamps and expiry rates to ensure the quality of the distributed data without direct communication to the associated public health groups. Our previous work found through implementation that user CPU and data overheads for this type of implementation can be quite minimal [18], without significant additional overheads for the data owners/distributer.

This approach would additionally allow for dissemination and retrieval of data through the anonymous communication network with users retrieving policy updates and interventions relevant to them without breaching their anonymity.

4. Privacy and Security

Our system, by applying granular restrictions on data collection controlled by the user, allows perceived and real privacy concerns to be alleviated.

The core concept of local processing (on the user mobile device) of health data for anonymized submission requires that individual components of a data submission have an associated quasi-identifier score (QIS). To avoid the QIS exceeding a privacy threshold, data components can be modified to be more generalized (see Section 5) such as for example a submission including the county of submission rather than postcode, and the QIS would reflect the increased generality. The approach also takes into account the case where multiple quasi-identifiers are submitted together as such a group of quasi-identifiers would have a combined QIS value that is assessed against privacy thresholds. The four core data components in determining the combined QIS are (i) Measures, (ii) Location, (iii) Temporal; and (iv) Demographic and are described below

Measures are aggregate or calculated values that refer to a specific value to be collected. Examples are listed in Section 2.2. Location refers to the specific location a measure occurred, Temporal refers to the period of time in which a measure occurred and Demographic refers to the other characteristics of an individual.

Assuming that the data submitted is aggregated across a relatively long temporal period, and not submitted with exact physical location data, the only likely source of re-identification might be the individual's particular demographic data.

The types of demographic data needed for the public health data capture system, such as age or age range, gender, major ethnicities, city or zip/postcode are also generally non-identifying based on population distributions. The population demographics of regions and countries are already collected and known in many cases such as where national census data is collected, and in some cases are known for specific activities that may be used in measures, such as cycling-based activity [19]. As such, the probability of a combination of demographics can be calculated and compared against a privacy threshold setting. Such a formula for the QIS, D_{QIS}, is below where the λs are the individual demographic details.

$$D_{QIS} = 1 - Pr(\lambda_1, \lambda_2, \quad \lambda_3, \ldots, \lambda_n)$$

We consider how non-identifying such demographics might be in the following evaluative section.

5. Privacy Analysis

To demonstrate the operation of this approach we evaluate an example data submission for the New South Wales geographical area based on real distributions data. This area has a population of 6,917,658 as of last census. Using the Australian Bureau of Statistics census population statistics [20] we generated a random data set based on the relative size of the demographics, specifically looking at gender, age bracket and local government area. Additionally, to create plausible activity measures we then generated activity averages and cycling participation

based on previous research [19].

Assessing our local processing approach we generated the data set out to a specific number of participant submission numbers: 10000, 50000, 100000, 200000 and 400000. We then tallied the number of individuals that had a *k* value under the threshold of 20, 10 and 5. Having a small *k* value for a specific demographic category is undesirable, as it can allow for potential re-identification or inference-based attacks to be used against the data set.

As can be seen in **Figure 2**, at 10,000 submissions, there were high numbers of individual submissions that had low *k* values with 4782 submissions having a *k* value lower than 20 and 929 having a *k* value lower than 5. In practice this would be extremely problematic in ensuring anonymity and privacy of data submissions. As for example, if additional knowledge that an individual participates in the population data submission is available, it may be enough to perform re-identification of some individuals. As the data submissions are increased to 400,000 these risks diminish but there is still a reasonable chance of re-identification even at significant data collection levels of 400,000.

To improve this result we implemented our demographic formula and set a reasonably conservative threshold value for D_{QIS}. As local government area was the optional value in this submission that was adjusted, rather than just withholding the submission. If a D_{QIS} value for an individual was over the threshold based on known population demographics, locale government area details were excluded from the submission.

As demonstrated in **Figure 3** this resulted in a decrease in the number of unique submissions that had low *k* values. This differentiation increased as the number of data submissions increased with the adjusted submission approach reaching a safe level much sooner at ~200,000 and comprising as low as 1.6% of the submissions below the *k* threshold at the 100,000 submission level compared to 4.1% in the unadjusted data set.

The threshold at the local device level could of course be adjusted either higher or lower based on the expected submission numbers. However, it performed quite respectably at the initial level with a significantly lower level of risk at the 10,000 submission level and close to no statistical risk at the 400,000 level which represents 5.8% of the area total population.

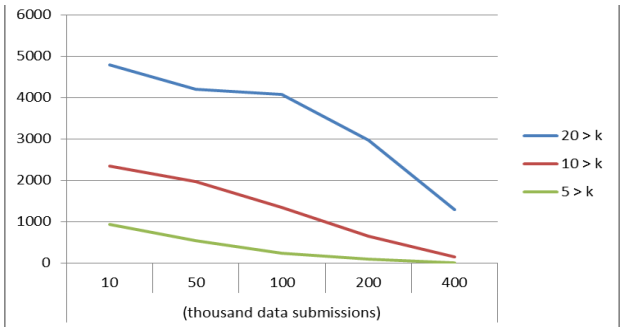

Figure 2. Demographic *k* value without local processing.

The limitation of this local processing approach as compared to a trusted server approach that performs *k*-anonymity is that the number of other submissions cannot be known with certainty by the local device. As such, the privacy threshold is set at a conservative value to preserve privacy. However this means that when there are high levels of submissions more records are adjusted than was required. This relationship is displayed in **Figure 4** where for 10,000 records the percentage of records adjusted was less than the low *k* value percentage and the miss rating was extremely low. This diverges as the number of data submissions increases, since the adjustment level remains fairly constant at around 39.5% of data submissions for the example data set. In this case due to the high number of local government areas (153) with a significant proportion having extremely low populations, the adjustment rate was quite high.

Overall, this wouldn't pose a serious problem, as the priority of the demographic detail is controlled by the data requestor and trade-offs are to be expected for increased detail in other sections.

In summary, for the example data set the local processing aggregate data approach performed favourably for the defined public health and privacy requirements.

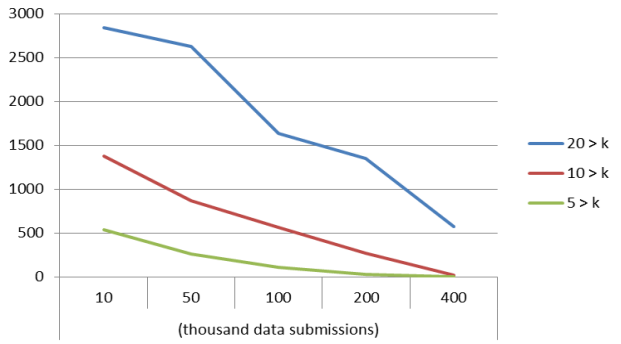

Figure. 3. Demographic *k* value with local processing.

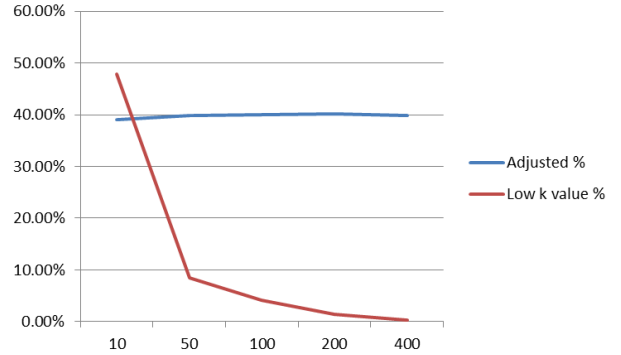

Figure 4. Adjusted submission compared to low *k* value submissions.

6. Conclusions

We have described that the current state of IoT in relation to commercial health and fitness sensors is well matched to capturing data relevant to numerous public health measures. We have described a novel IoT-based public health information system architecture that allows the completely new capabilities of both anonymous public health data collection and anonymous public health intervention. We have described its privacy and anonymity mechanisms and provided a privacy evaluation.

REFERENCES

[1] P. Klasnja and W. Pratt, "Methodological Review: Healthcare in the Pocket: Mapping the Space of Mobile-Phone Health Interventions," *Journal of Biomedical Informatics*, Vol. 45, No. 1, 2012, pp. 184-198. doi:10.1016/j.jbi.2011.08.017

[2] A. Clarke and R. Steele, "Health Participatory Sensing Networks," Mobile Information Systems, 2013 (in press).

[3] K. Sampigethaya and R. Poovendran., "A Survey on Mix Networks and Their Secure Applications," *Proceedings of the IEEE*, Vol. 94, No. 12, 2006, pp. 2142-2181. doi:10.1109/JPROC.2006.889687

[4] S. Mauw, J. H. S. Verschuren and E. P. Vink, "A Formalization of Anonymity and Onion Routing," in (Samarati, P., Ryan, P., Gollmann, D., and Molva, R., 'eds.'): Computer Security – Esorics 2004, Springer Berlin Heidelberg, 2004, pp. 109-124.

[5] M. Swan, "Sensor Mania! The Internet of Things, Wearable Computing, Objective Metrics, and the Quantified Self 2.0," *Journal of Sensor and Actuator Networks*, Vol. 1, No. 3, 2012, pp. 217-253. doi:10.3390/jsan1030217

[6] R. Steele, "Social Media, Mobile Devices and Sensors: Categorizing New Techniques for Health Communication." Proceedings of the Fifth International Conference on Sensing Technology (ICST), 2011, pp. 187-192.

[7] R. Steele, A. Lo, C. Secombe and Y. K. Wong, "Elderly Persons' Perception and Acceptance of using Wireless Sensor Networks to Assist Healthcare," *International Journal of Medical Informatics*, Vol. 78, No. 12, 2009, pp. 788-801. doi:10.1016/j.ijmedinf.2009.08.001

[8] Jawbone, "UP™ by Jawbone® with MotionX® Technology Empowers You to Live a Healthier Life," 2011 [accessed 2013 April 20]; Available from: http://content.jawbone.com/static/www/pdf/press-releases/up-press-release-110311.pdf.

[9] Mobi Health News, "Google Adds Activity Tracking to Android App," 2012, [accessed 18th Jun 2013]; Online: http://mobihealthnews.com/19551/google-adds-activity-tracking-to-android-app/

[10] Sano Intelligence [Online], Accessed 10th June 2013, http://rockhealth.com/accelerator/portfolio-companies/sano-intelligence/

[11] B. Network, "Riderlog," 2011; Available from: http://www.bv.com.au/general/ride-to-work/91481/.

[12] C. Outram, C. Ratti and A. Biderman, "The Copenhagen Wheel: An Innovative Electric Bicycle System that Harnesses the Power of Real-time Information and Crowd Sourcing," *in EVER Monaco International Exhibition & Conference*, 2010, pp. 8.

[13] D. E. Warburton, W. C. W. Nicol and S. S. Bredin. "Health Benefits of Physical Activity: The Evidence," *Canadian Medical Association Journal*, Vol. 174, No. 6, 2006, pp. 801-809. doi:10.1503/cmaj.051351

[14] R. Steele, "An Overview of the State of the Art of Automated Capture of Dietary Intake Information," *Critical Reviews in Food Science and Nutrition*, 2013.

[15] AIHW, "Biomedical Component of the Australian Health Survey: Public Health Objectives," 2011, [Online]. Available:http://www.health.gov.au/internet/main/publishing.nsf/Content/health-pubhlth-strateg-food-monitoring.htm/$File/Biomedcal%20component%20AHS-public%20health%20objectives.pdf

[16] P. Kalnis and G. Ghinita, "Spatial K-Anonymity", in (Liu, L., and Özsu, M.T., 'eds.'): Encyclopedia of Database Systems, Springer US, 2009, pp. 2714-2714.

[17] A. Clarke and R. Steele, "Summarized Data to Achieve Population-Wide Anonymized Wellness Measures," *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, 2012, pp. 2158-2161.

[18] A. Clarke and R. Steele, "Secure and Reliable Distributed Health Records: Achieving Query Assurance across Repositories of Encrypted Health Data," *in System Science (HICSS), 2012 45th Hawaii International Conference on*, 2012, pp. 3021-3029.

[19] Austroads, "Australian Cycling Participation "2011, http://www.austroads.com.au/abc/images/pdf/AP-C91-11.pdf.

[20] Australian Bureau of Statistics, "Census Community Profiles Greater Sydney", 2011, http://www.censusdata.abs.gov.au/census_services/getproduct/census/2011/communityprofile/1GSYD,accessed 28/03/2013

