An Architecture for Pull-based Public Health Interventions

Robert Steele
Medical University of South Carolina
steelerj@musc.edu

Andrew Clarke
The University of Sydney
acla6821@uni.sydney.edu.au

Abstract

Public health interventions consisting of information dissemination to affect behavior have long been a significant form of public health campaign. These interventions can be considered ‘push-based’ as they push information out to a population. New smartphone technologies for the first time provide a platform for a new type of informational public health intervention, which can be referred to as ‘pull-based’ interventions. In such interventions, it is the device via automatically collecting data relevant to the individual’s health that triggers the ‘request’ for and receipt of the informational public health intervention. This will enable far more targeted and personalized public health interventions than previously possible. Such techniques do however also pose privacy and security challenges. In this paper we introduce the architecture for pull-based public health interventions upon smartphone devices that also provides strong privacy support. We also evaluate its performance and scalability benefits relative to non-pull-based interventions.

1. Introduction

The use of informational public health interventions for behavioral adjustment has significant potential to evolve into a more targeted, measurable form of public health intervention through the use of new mobile computing platforms such as smartphones. Advantages include the collection of real time or near real time data on the behavior modification effectiveness of health interventions, long term measurement of the effectiveness of interventions and more precise targeting. By the term public health intervention in this paper, we refer to the dissemination of information to affect health behaviors, or dissemination of health promotion of health education information.

There are two key challenges to this type of public health intervention platform. Firstly, as the specific intervention is by necessity decided upon and applied at the local device level a large number of broader interventions need to be delivered to each device efficiently. As such systems enable more and more targeted and personalized public health interventions, the number of interventions being distributed will grow significantly. The pull-based intervention approach is a step towards improving efficiency of distribution so as to allow the accommodation of a greater level of targeted/personalized interventions. The greater personalization and targeting of interventions in fact blurs the line between public health intervention and individualized health intervention. Secondly, is the need to report with as much detail as possible, as to which intervention was performed and its effectiveness without breaching privacy, or as is a goal of our system to prevent inadvertently allowing individual re-identification at a later stage.

We propose as a solution to these problems which is an extension of our prior work in relation to health participatory sensing systems (HPSNs) [1-3] and query assurance [4]. These first three works have set out in detail the population health data collection capabilities of HPSNs, but not detailed the health intervention capabilities and in particular not introduced pull-based techniques. The pull-based intervention approach can improve upon intervention distribution efficiency. The query assurance architecture is adapted to reduce the quantity of health interventions that need to be delivered to participants and via the distribution network and hence reduce the resultant network load and computation load on the endpoint smartphone devices.

2. Related work

Current research into mobile device-based public health intervention architectures has either concentrated on advanced but non-health-specific participatory sensing technologies [5] or on older mobile technologies such as SMS/ MMS [6]. As such new capabilities such as pull-based public health
interventions or in particular the anonymity properties of such systems have not yet been addressed elsewhere.

The rich capabilities of participatory sensing have garnered interest in its usage for a range of quite disparate areas from air quality and pollution sensing [7], to urban area noise level data collection [8] and public health data collection [3] amongst many others. This has in turn spurred a number of different approaches to resolving or decreasing the implicit security and privacy concerns when involving individuals in sensing/data collection.

The more conventional approach would be to use a trusted server, then k-anonymity [9] or a variant, to anonymize the data before they are accessible for research/analysis. The main downside of this type of approach is the need for a fully trusted server, which creates a single point of failure in terms of privacy breaches. Alternatively, other approaches have improved upon this by removing some sensitive information before submission (removal of identifiers and communications anonymity) with a central point of trust [10] to provide an anonymous approach. While this is quite effective when the sensing is collecting data on something not specific to the individual, this alone is not well-suited to a model where information on the participant is a key submission component (such as in the case of collection of public health intervention data) as de-identification protection is still implemented at a central trusted point.

There has been some prior research to resolve the issue of requiring a fully trusted server, such as, decentralized participatory sensing networks [11] using user interaction/awareness as part of the approach or keeping the data managed by the participant [12], [13] and stringent user-definable access control mechanisms to manage sharing. The limitation of these approaches when considering HPSNs is that typically they have not incorporated support for public health interventions (or an equivalent), a capability that does not have a direct parallel in most participatory sensing systems and remains a key component of HPSNs.

3. Pull-based public health information systems

Smartphone-based public health interventions are distinct in their ability to reach individuals in a targeted and real-time manner. However, the greatest capabilities can be accessed when the smartphone as an intervention platform is coupled with the use of the smartphone as a data collection and analysis hub. This phenomenon of health data collection and analysis is becoming ubiquitous with many of the main smartphone manufacturers including health applications within their mobile operating systems such as Apple’s HealthKit (developer.apple.com/healthkit/) and ResearchKit (www.apple.com/researchkit/), Samsung’s s Health (shealth.samsung.com) and Google’s Fit (fit.google.com), and these capabilities can be further extended through the use of ‘wearables’ to provide additional data collection.

As such, this trend has created an incredibly rich and detailed data hub for individual’s health and wellness. It follows that the ability to utilize this data hub for public health interventions would be a significant opportunity. However, the nature of the type of data involved and the individual participant’s privacy and security concerns would make a typical centralized data server and direct communication approach problematic.

To realize the full capabilities of such public health interventions a pull-based public health information system is highly advantageous. Such a system would utilize health sensors and data processing at the mobile device level to instigate the pull-based intervention. We will subsequently use the term ‘mobile device’ to mean smartphone device or smart device in this paper. As such the overall platform would require an anonymous distribution layer to enable request and retrieval of public health interventions; an application layer to deploy the intervention; a reporting collection functionality to provide objective analysis and feedback on intervention suitability and effectiveness; an interventions module that infers context and analyses collected data to trigger a pull-based health intervention; and the sensor layer which provides context and detail to inform the reporting and intervention modules. These layers will be discussed in further detail in section 4.

These types of health intervention platforms can be utilized with the broader participatory sensing frameworks, which have experienced a growth in recent years that has been greatly increased through the high levels of smartphone adoption in many countries [14] and proliferation of commercial wearable devices and health sensors, leading to the pervasive availability of powerful sensing platforms that are highly human-centric, making them ideal as the center-points for health participatory sensing models.

In our previous work [1] a number of different classifications for participation in a HPSN are identified. The classification most relevant to public health interventions is 'active participatory sensing'.
Active participatory sensing differs from other types of participatory sensing by providing inputs to the individual to alter the actions they would have taken whilst participating in the HPSN. Active participatory sensing in the health context has a somewhat different goal to that of many other active participatory sensing contexts [15]. While an active participatory model for typical sensing might focus on affecting individuals to collect a more complete data set in terms of spatial/temporal range, health and epidemiological-related active participatory sensing would be more concerned with affecting a health-related action and hence have a component equating to a public health intervention. As such, the behavior change would be to firstly attempt to improve the sensing data captured in terms of risk and preventative factors. Additionally for public health goals, this allows for immediate and continuous feedback on the effectiveness of interventions on receiving groups. It is assumed that active participatory sensing would have similar levels of technical sensor capabilities to the other classifications [1], with the focus shifted to the potential two-way communication that can be built upon sensing data and an inherent feedback loop. This has the potential to be both a powerful data collection tool as well as a novel public health intervention platform. Its potential scope includes the ability to quantify precisely the effectiveness of public health interventions, in a timely and accurate manner.

4. Pull-based public health interventions: Architecture and process

The pull-based public health intervention platform will be incorporated into a larger health participatory sensing system which provides for anonymized public health data collection and also supports push-based public health interventions. This is because without such a larger capability the effectiveness of the utilized public health interventions could not be collected and analyzed in a timely manner. The pull-based public health intervention architecture is needed to accommodate a multiplication of the number of interventions that is implied by a system able to support more targeted and personalized interventions. Even without this larger system the intervention system can still provide a lesser but still significant improvement over traditional public health information/behavioral interventions. As such, we consider that public health interventions can be conducted as a component of a HPSN as described in section 3 and our previous work [1].

The platform components and their inter-relationships are illustrated in Fig. 1, and serve to support the capabilities of anonymous distribution, local application of public health interventions, data collection for reporting, and analysis of results. These are described in further detail in the following subsections. The overall architecture (Fig 1) is that of a HPSN [1-3], but here the architectural figure is modified to show the two modes of interventions, both push-based and the introduced pull-based intervention. Figure 2 shows the internal workings at both the Health Participatory Sensing Server and the mobile devices specific to the pull-based public health interventions.

4.1. Distribution

The distribution of public health interventions in the HPSN is comprised of two main components, the distribution network and the distribution approach.

Fig. 1. Pull-based public health intervention architecture

The distribution network consists of a mix network [16] or onion network [17], which provides for anonymity of the submitter as well as secure communication. Such approaches utilize a chain of proxy servers between the participant and HPSN, which can provide anonymity for both parties, though
in this case it is only required for the mobile device user. Though this creates additional implementation complexity the potential benefit to real privacy is significant, with the only remaining significant privacy threats being: insecure storage of data on the local device which we consider outside the HPSN network; and re-identification via the content of the data submitted and pull-based interventions requested as discussed below.

Fig. 2. Pull-based public health intervention process

The distribution approach utilizes two distinct methodologies for receiving public health interventions. The more common public health interventions can be fairly efficiently packaged and distributed and sorted groupings remove the risk of re-identification through knowledge of the distribution of specific types of interventions to small subsets of the participant community. This approach was demonstrated in our previous work [3] which utilizes a query assurance approach [4] to provide granular completeness, correctness and freshness assurance of the public health interventions that are distributed to the HPSN clients. This approach uses an implementation of one or many sorted and digitally signed merkle hash tree/s utilizing expiring timestamps, retrieved alongside the requested data to verify the content of the retrieved data. This allows for a hash of each possible granule of retrieved data to be efficiently distributed with a single digital signature and expiring time stamp for the overall request, reducing verification overhead of both computation time and data. This is effective even where only subsets of the overall data are retrieved through a third party or untrusted distribution network. This allows for high levels of certainty of the validity of data, while allowing for flexibility in request size even though the data is distributed through untrusted nodes, while keeping verification data overhead size and processing time to acceptable levels.

However, though this approach is effective for the common types of interventions distributed to the participant community as a whole, the development of advanced sensors that provide greater detail and significantly more complete and complex digital collections of individual details, means that the range of public health interventions possible is likely to grow beyond what is feasible to distribute in a common batch approach. By common batch we refer to the data messages sent out through the distribution network that can be activated at the various smartphone devices for which the locally held information matches the conditions for the intervention to be applicable to that individual. As such, we propose that an additional pull-based approach for highly targeted interventions is required that can fulfill this need without breaching individual privacy. Improving distribution efficiency so as to enable utilization in a broader range of network environments is key to utilizing ubiquitous computing in healthcare [18].

This pull-based approach, detailed in Figure 2 allows the individual’s data as determined locally on their device, including dynamic health sensor data [3, 19], to trigger the request of an intervention to be sent to their device. This represents a highly personalized and dynamic level of public health intervention. For example the confluence of a particular individual’s demographic data, with weight information and blood sensor measures could indicate this individual to be at risk of heart disease. Their device would then trigger or ‘pull’ an informational health intervention for those at elevated risk of heart disease.

This request would be sent via the mix network to the central server. Once a sufficient number of such requests have been received, the intervention will be distributed back through the mix network.

As described in more detail in Section 5, fundamentally the improved efficiency and scalability of the pull-based approach arises as not all the interventions will need to be distributed to all of the nodes in the mix distribution network as is the case in the push-based approach.
These batches will also differ from the generic intervention intervention in that they won’t include any additional data collection rules – rather just the pull-based health interventions and the required query assurance verification information. The interventions are batched so as to allow the retrieval of the intervention by individuals without any risk of matching a particular individual to a specific intervention. An additional safeguard would be the implementation of a protocol on the client devices whereby the client device in a randomized manner retrieve a pull-based intervention distribution that was not requested by their device and discard. This essentially acts as an ‘obfuscation request’, by making the number of retrievals and retrieval times of real retrievals less certain. This additional protocol would further obfuscate the exact individuals that are actually receiving the sensitive (potentially more identifying) pull-based interventions, so that even if the security of the mix network is compromised there is an underlying level of uncertainty that could provide an additional layer of anonymity protection.

4.2. Application

The public health interventions are performed on the local device. The decision as to the intervention to request also must be made locally as more specific information about the individual is not transmitted to the HPSN server. As such, the specific intervention is chosen locally to most closely match the individual’s demographic and health profile details, even if those details cannot be fully disclosed to the server.

4.3. Reporting collection

To provide an anonymous public health intervention system that also collects outcomes and the effectiveness of those executed interventions, a level of data collection is a necessity. However, if the necessary limitations on data collection are not considered, this could result, even in cases where de-identification of data is performed locally, in unwanted re-identification of data at a later stage using data external to the HPSN [2]. This potential scenario is a significant concern of HPSNs and by extension public health interventions systems on such networks. We consider that the most effective way to mitigate this risk that doesn’t require a trusted server or aggregation in some form is the use of local processing of data reporting at a suitably conservative privacy setting to minimize risks.

As such our system, by applying granular and modular restrictions upon data reporting [2], reduces real privacy risks through a threshold approach to privacy and submissions. The local processing approach considers the potential for re-identification before submission and reduces or modifies the number or detail of the demographics submitted. Additionally, the use of a local processing approach to data submission and health interventions policies allows the on-device adaptation to achieve a data submission which matches the reporting request as closely as possible without breaching variable user defined privacy conditions [2]. It should be noted that the fully detailed data can still be utilized locally on the device for individual health care stored via portable personal health record [20].

For public health interventions this is resolved by submitting aggregate data that is not time or location sensitive, with restrictions on the specificity of the intervention reported to be performed. That is, for example if an intervention was targeted at the entire population a certain level of demographic detail could be returned as well as the intervention type and the effectiveness of the intervention as a measure, such as any measurable change in behavior or health indicators. Alternatively, if the intervention was tightly focused on a small subset of the community, the specific intervention type may need to be reported as a broader type that is inclusive of the specific type and limited additional demographic details as prioritized by the intervention request.

In the specific case of pull-based public health interventions, due to their much more specific contents, the level of data collection will have to be more stringent. However, the nature of the pull-based health interventions are health status rather than demographic in nature, so it would still be possible to in most cases report on the broad demographics and the success of the intervention.

4.4. Interventions

The analysis of public health participatory sensing data relies on collection of sufficient data for public health uses [2], which differs from what would be required in most other participatory sensing systems. Generally aggregate nonspecific demographic level data is needed as well as the measured values and the types of interventions performed.

5. Implementation

Our implementation provides an approach that addresses the key challenges of efficient and privacy preserving public health intervention distribution and the reporting of the application and effectiveness of
public health interventions. This implementation utilizes a combination of pull-based public health interventions to provide condition-based more individual-specific health interventions and push-based public health interventions. In our evaluation we consider the combination of these different types of interventions at specific relative levels e.g. 10%, 20%, 50% and 70% pull-based and the effect that will have on overall data and verification overhead.

Additionally, we consider the scalability of an intervention approach from 10,000 public health intervention definitions to 10x, 100x and 1000x greater levels.

Public health interventions are likely to include a combination of text, images, video and audio components. Additionally, even when considering only targeting broad demographics this can result in potentially tens of thousands of different combinations for targeted interventions, when extending this to specific data about the individual stored at the mobile device level and multiple public health groups/organizations involved in the system.

When the consideration of more specific health related interventions are taken into account, the potential overall data overhead, would likely be detrimental to the effectiveness of the HPSN.

While much of this overhead could be reduced through conditional approaches which make a single intervention relevant to multiple targets the core problem remains. That is, conditional approaches will also lead to significantly larger sized intervention data messages needing to be distributed, to allow for the required multi-target personalization.

To address this scalability problem we utilize a combination of pull-based interventions for the most specific types of public health interventions and a query assurance approach to allow for modular retrieval of the common health intervention set.

As such, we created an example data set that includes different data types and compares the data overhead of retrieving the entire data set, to retrieving a subset and a verification tree for data quality assurance and the previous approach that utilizes a more efficient verification tree [4].

We then evaluate this data set against a spectrum of possible configurations, whereby we investigate the interaction of the utilization of pull-based interventions vs push-based for relative proportions of the overall data set.

The data setup involved 2000 components typical of an audio/visual intervention size and 10000 components of a text and intervention details size. These components were verified by a single verification tree [4] (see Section 4.1). Even with this limited size dataset it is apparent that it wouldn’t be feasible to distribute the entirety of the interventions to any particular user, as this would represent hundreds of megabytes. Our proposed approach instead involves a user requesting a subset (approximately 8-10 megabytes). The request is broad enough that it does not expose any personally identifiable details, then uses the verification tree to authenticate the subset. This removes the need for direct communication with the source, or for the source to hash and digitally sign every possible requested combination.

6. Evaluation

We assess the pull-based component through the retrieval of an additional batched pull-based intervention sets for a proportion of the requests.

With the pull-based batch distribution it is assumed that the pull-based recipients will retrieve the entire batch and then verify, rather than retrieving a subset. A batch of interventions will be the collection of a group of interventions that have been requested for pull-based distribution, with a minimum and maximum batch size. These batches will be distributed once and have a set expiry, after which they are discarded from the distribution node, that is, the pull-based interventions are only temporarily held within the distribution network as compared to the push-based distributions which are held by the distribution nodes and continually refreshed to keep the interventions up to date. As such configuration of the size of the batch and the number of pull-based interventions on average that will be delivery to an individual in a particular batch is key to the data distribution efficiency.

![Fig. 3. Public health intervention distribution data volume](image-url)
Fundamentally, this isn’t too dissimilar to the push-based configuration, whereby, clients retrieve more public health interventions than are actually performed, with the algorithmic decision to apply a particular public health intervention on the mobile device, made on the device based on the more detailed data than could be communicated through the HPSN.

In Figure 3, we illustrate the trade-off of the data retrieved by a number of clients through a combination of pull and push based distribution. We assumed a batch size of approximately 10 MB, and an average number of interventions for each individual per batch as 7. We then plot the data retrieval total in bytes, while increasing the pull-based intervention component from 0% to 30% and keeping the number of targeted public health interventions at a constant 10000 per implementation benchmark. As can be seen in Figure 3, these levels of pull-based intervention incorporated at the batch level previously discussed results in data retrieval levels that are consistent to push-based approaches, with potential for slight improvements as in the 30% configuration.

Additionally, the switch from push-based to pull-based decreases the overall volume of interventions that need to be held at distribution nodes. Essentially, if 30% of distinct public health interventions are no longer distributed with the main push distribution there is an effective decrease in the amount of data distributed to the distribution nodes, which in a scaled out system could number in the hundreds. Instead pull-based batches that are only distributed when needed and can be discarded by the distribution nodes after a short timeframe are utilized. Additionally, push-based interventions need to be kept fresh, with distributed updates periodically, while the same update process is not required for pull-based interventions as they are only retrieved when needed.

The combination of the above three efficiencies allows for pull-based distribution to be utilized and the higher capabilities that are entailed without a negative efficiency consequence.

An additional component of our approach is optimization of the verification tree based on historic usage [4], this is only possible to achieve over the push-based interventions, the pull-based even though they are batched due to their transient nature and content will not be possible to optimize based on historic usage. As such we perform our implementation pre and post optimization incorporating 10000 requests for each. The results of the verification overhead are displayed in Figure 4.

As can be seen in Figure 4, the incorporation of pull-based distribution has a net positive effect on the verification data overheads – with pre-maintenance efficiency being substantially improved by the inclusion. Post-maintenance received more mixed results, however, this is compared to the best case scenario for the maintenance optimization approach.

Further discussion of the other impacts of introducing pull-based public health interventions on the existing push-based system are covered in the following scalability section.

### 6.1. Scalability

The size of the returned verification object that is packaged alongside the public health interventions will have the largest impact on the scalability of the pull-based public health intervention system. The other components of the verification module being: 1. Hashing the retrieved data O(1), 2. Verifying the digital signature and timestamp O(1), 3. Searching the balanced verification tree and retrieving results – similar complexity to the size of the verification object but less computationally expensive. Specifically, the depth of the verification tree is decisive in the number of hashing and comparison operations that need to occur. As such the hash tree size can be evaluated as a min/max value (1) based on the whether the leaves of the tree are full (f entries) or in the worst case all half full (f/2 entries).
Leafmin = \( [n/f] \), Leafmax = \( [2*n/f] \)                         (1)

Based on these calculations we can evaluate the height of the tree (2) and (3).

\[
\text{Heightmin} = \log f (\text{Leafmin}) + 1 \quad \text{(2)}
\]

\[
\text{Heightmax} = \log f/2 (\text{Leafmax}) + 1 \quad \text{(3)}
\]

So for a given \( n \) number of elements in the verification tree there is an associated height based on the number of leaves per branch \( f \).

This using the max height from (3) gives a \( O(\log f/2(2*n/f)) \) value for verification, where that number of hashes will have to be calculated and compared in addition to \( O(1) \) digital signatures operations. Additionally, the efficiency improvements aim to decrease the height of the tree that needs to be retrieved and evaluated, however as this is a dynamic process (and heavily reliant on usage patterns) it is only evaluated in the implementation.

Additionally, for the implementation of our combined pull-based and push-based approach, there is the additional consideration of the batch pull-based public health interventions, that are also packaged with a fixed size verification object due to the nature of the batch approach. As such the scalability of the overall public health information system becomes a combination of these factors. Since the batches will have an approximate size BatchLeaf the following formula can be used to calculate the height of the verification tree.

\[
\text{BatchHeightmin} = \log f (\text{BatchLeaf}) + 1 \quad \text{(4)}
\]

\[
\text{BatchHeightmax} = \log f/2 (\text{BatchLeaf}) + 1 \quad \text{(5)}
\]

As such both the verification computational time and the verification data overhead for the pull-based batches will be static based on the configuration of the batch size and the proportion of obfuscation requests made by clients to aid privacy preservation.

As such for an example situation where 10000 public health interventions need to be distributed by a combination of push and pull based mechanism we can see the following interaction in Figure 5 and Figure 7; the increase in nodes returned as the number of public health interventions available/stored in the HPSN grows. In both Figure 5 and Figure 7, 10000 public health interventions are being distributed. The x-axis in Figure 5, refers to the size of the HPSN by number of public health interventions available/stored. This differs from the number of interventions being distributed to the mobile devices as in a push-based approach the HPSN needs to distribute out its entire health intervention collection and the associated verification tree to distribution nodes, from which the 10000 health interventions are delivered as subsets to each individual device. While the pull-based approach in Figure 7 only needs to distribute out the 10000 requested interventions in small batches with smaller verification trees. Additionally, this is considered where there is a node per branch value of 5, 10 and 20 defined in the underlying verification tree, which utilizes a B-tree type structure. B-trees are a form of sorted, balanced tree commonly used in indexes in databases and file systems. A key characteristic of B-trees is that branch nodes are always kept at least half full, where full is indicated by the node per branch value. Where a branch would exceed the node per branch value after an insertion a split and rebalance is required. Similarly when a deletion would reduce a branch to having child nodes less than half of the node per branch value a merge and rebalance is initiated.

![Fig. 5. Push-based public health intervention verification scalability](image)

The result of this underlying structure means that this configuration value has an impact of the depth and breadth of the verification tree, with a higher node value resulting in a shallower tree but greater node retrieval in a merkle hash tree approach as all the nodes of each branch up the direct tree hierarchy to the root needs to be retrieved for verification.

As above the growth rate is a log function and could be further improved through the maintenance function provided in the implementation. In figure 7 we consider the scenario where a proportion of those 10000 interventions are instead distributed by a pull-based batch approach. As the size of the batch is linear we plot the number of nodes retrieved for verification for \( x\% \) of the overall intervention distribution. The number of nodes required to be retrieved for verification is the nodes directly related to queried content, sibling nodes and the parental hierarchy to the root/signed node. These nodes need to be retrieved to hash the intervening levels of the
tree and finally verify the digital signature; this process is illustrated in Figure 6. This tree description and figure and further detail can be found in our previous work on query assurance [4].

Fig. 6. Verification tree retrieved node structure

While if compared directly it is obviously less efficient for small scale public intervention platforms to utilize an entirely pull-based distribution approach as compared to a push-based approach. The greatest advantage is achievable through the combined approach of pull and push based interventions which can by removing rarer public health interventions from the general distribution improve the overall efficiency. Indeed, for more large scale public health intervention platforms that for example have greater than 500000 public health interventions available/stored as displayed in Figure 5 it can be seen from the comparison to Figure 7’s 100% pull-based intervention evaluation that a pull-based batch approach can be more efficient even as a standalone component. That is, the pull-based approach varies from 343000 to 390000 nodes retrieved for a 100% pull-based approach, based on the number of nodes per branch. While a push-based approach at 500000 public health interventions available/stored will have a verification node retrieval size of 333029 to 469897 nodes varied by the nodes per branch, with the overall public health intervention distribution number remaining the same i.e. 10000 public health interventions.

The overall efficiency can be improved at a much earlier point through the incorporation of pull-based public health interventions. This is possible through as previously described providing less common interventions only through the pull-based approach and reducing the overall size of the available/stored interventions on the distribution nodes and by extension the height and breadth of the verification tree.

The final consideration for scalability of pull and push hybrid public health intervention systems, is the actual volume of public health intervention data that is distributed to mobile devices and through the overall network in general. As previously discussed the comparable advantage of either approach beyond the capability enhancement garnered through a pull-based interaction, is determined by the configuration of the pull-based intervention batch size, and the average amount of interventions per individual distributed in a single batch. This is contrasted to the push-based approach where the configuration is focused on deciding the level of retrieval from the HPSN to retrieve all the relevant public health interventions for an individual that may be needed on the device, without missing key data. The advantage of pull-based components is that public health interventions and the associated overheads are only incurred related to a direct request for an intervention as opposed to the overheads incurred through retrieving and keeping a large subset of the public health interventions locally and up to date.

7. Conclusion

This paper describes an approach for pull-based public health interventions within a smartphone-based public health information system. In particular, we describe the new and powerful capability that public health interventions can be distributed, performed and evaluated without the need for identifying details of an individual participant to ever leave their mobile device. Additionally we have considered and evaluated the efficiency, privacy and scalability of the pull-based public health intervention capabilities. The smartphone based public health information systems include an approach based on local processing to aggregate data for public health use that utilizes privacy thresholds and an adaptable approach to public health interventions and reporting. To this end we have provided a detailed evaluation of the overheads, efficiency and scalability of the public health intervention distribution model.
8. References


