# The Effective Encrypt and Decrypt Methodology for Secure Data Aggregation in Mobile Sensing

[1]Prathipati Prasanna Kumar, [2]G. Appa Rao
[1,2]Dept. of CST, Gitam University, Rushikonda, Visakhapatnam, India

**Abstract**
Throughout the years ability of the portable detecting gadgets like advanced mobile phones are expanded as far as catching and sharing the data. This data can be helpful if examined as an accumulated esteem or values. Great illustrations resemble movement pattern specifically territory, therapeutic data specifically region and so on. At the point when such data is shared to an aggregator, client's personality ought to be traded off and just data should be shared. To accomplish this present client's personality Li, Cao and Porta has recommended effective convention to get total quality utilizing Homomorphic encryption and novel key administration. Sadly, the utilization of cryptography in the anticipated convention plan presented the win big or bust decoding model. In this manner, the aggregator adapts nothing if a solitary client comes up short. In a portion of the basic applications, information must be gathered from all clients to get right results. Any spillage in information will prompt off base hostility and sub-sequent figuring's. Adaptation to internal failure must be taken care of to get right results and D. Tune [5] have proposed a strategy to handle adaptation to internal failure. In proposed strategy, a twofold interim tree over n clients, and permit the aggregator to gauge the entirety of coterminous interims of clients as spoke to by hubs in the interim tree. The parallel tree method permits taking care of client disappointments joins and leaves, with a little logarithmic (or polylog) cost as far as correspondence and estimation mistake. In the event that clients have no impetus, or feel that their protection may be jeopardized, it is likely that they won't partake. In this article, we concentrate on security insurance in participatory detecting and present a reasonable protection improved foundation. To start with, we give an arrangement of meanings of protection prerequisites for both information makers (i.e., clients giving detected data) and buyers (i.e., applications getting to the information). At that point we propose an effective arrangement intended for cell telephone clients, which acquires low overhead. At last, we examine various open issues and conceivable examination bearings.

**Keywords**
Encryption, Multi-hop network, Mobile Sensing, Data Aggregator, Privacy

## I. Introduction

Participatory detecting is novel worldview to gather information from cell phones and other cell phones conveyed by an intensely expanding number of individuals. In view of this worldview (otherwise called astute, individuals driven, or urban detecting), an extensive variety of utilizations have been recommended that gather and process data on, for instance, ecological conditions like activity, urban air and commotion contamination, free stopping openings, or earth shudders, on business sector perspectives like fuel costs, or concerning individual wellbeing like weight control plans. Every one of these applications influence the high and expanding conveyance and accessibility of cellular telephones, whose number of memberships surpassed 5 billion with a surpassing offer of cell phones with adequate calculation power for—at any rate little—detecting assignments. Rather than remote sensor systems, where sensors are claimed, sent, and kept up by a solitary association, singular clients go about as the proprietor of sensors in the setting of participatory detecting who benefit their cellular telephones all alone and add to a typical pool of information, as a rule put away by a focal administration supplier. In any case, the utilization of individuals' cellular telephones as sensors likewise presents new security and protection viewpoints that must be dealt with. Most noticeably, sensors in a participatory detecting situation are no more stationary gadgets, yet are rather conveyed by their proprietors constantly, subsequently uncovering touchy information about their area or even picture or sound catches if utilized for agreeing errands. Moreover, the detected information is not from the earlier openly reachable and may hence be security touchy and require suitable insurance when distributed or answered to a focal information pool, though information gathered in remote sensor systems as a rule is really acquirable by the separate association. Participatory detecting subsequently acquaints the testing undertaking with handle the acquired information in a safe and protection saving way while accomplishing the best conceivable advantage from the detected information. In the most recent years, numerous methodologies have been made to accomplish protection in participatory sensor systems (cf. our treatment of related work in Chapter 2 for a broad diagram). In spite of this impressive corpus of work just a solitary late work by De Cristofaro and Soriente [6] drew closer a formally exact meaning of security and protection in participatory detecting for their plan called PEPSI, however prohibited essential angles as, e.g., assaults by different conspiring parties. This work henceforth presents the primary complete and cryptographically exact meaning of protection safeguarding participatory detecting. To this degree, we take up the design model of De Cristofaro and Soriente in light of the perception, that basic frameworks for participatory detecting include the accompanying negligible arrangement of gatherings:

### A. Sensing Devices

Devices (e.g., cell phones) conveyed by individuals, vehicles, or different elements that sense and report information (e.g., temperature, clamor level, and so forth.) utilizing proper sensors, shaping the premise for participatory detecting. We in this manner allude to those detecting gadgets as versatile hubs.

### B. Queriers

Individuals, establishments, or different substances keen on detected information (e.g., "clamor level on Time Square, New York") that subscribe for such data and get relating sensor reports.

### C. Network Operators

Entities that give the correspondence foundation of the participatory detecting application. Also, most participatory detecting foundations incorporate a middle person administration supplier, putting away information reports got by portable hubs

and handling the information for or transferring it to intrigued queriers. The administration supplier is when all is said in done a crucial gathering in a participatory sensor systems, as portable hubs are asset obliged gadgets being not for all time associated with the system and in this way unequipped for giving every single intrigued querier their information reports independent from anyone else, particularly not in a period deferred way. Be that as it may, a middle person administration supplier presents yet promote security challenges, as it gets all information reports as well as takes in the data hobbies of all queriers in a participatory detecting application. We along these lines present a model of a protection safeguarding participatory detecting foundation (PPP|■) in light of the portrayed engineering and refine the security prerequisites proposed by De Cristofaro and Soriente in the new model so as to give three principle security targets: hub security, question security, and report unlinkability. Hub security goes for the insurance of both the substance and reason for an information report issued by a versatile hub against unapproved queriers, the administration supplier, and other portable hubs, regardless of the possibility that every one of them intrigue. Inquiry protection formalizes the inverse security prerequisite that neither the administration supplier, nor different queriers or portable hubs should have the capacity to decide to what detecting data a querier subscribes. Report unlinkability at long last guarantees the in noticeability of versatile hubs by requiring, that information reports can't be followed back to the issuing portable hub by any gathering. We consider our protection safeguarding participatory detecting framework PPP|■ as a free building square, abstracting from the hidden

## II. Related Work

Personalization has been effectively actualized in an assortments of zones including Web pursuit and promoting. Late work has raised security worries about personalization taking into account private data in light of the fact that a tick on an advertisement/ Web page can release some private data about the client. In this manner, protection saving personalization has as of late gotten a considerable measure of consideration in the scholastic group as well as in the media. (1) Targeted Advertising. Before we talk about past work on protection mindful focused on promoting, let us quickly audit how existing web search tools customizes advertisements to be shown nearby with Web list items: The publicists offer cash on watchwords. For an approaching question, the subset of advertisements offering on watchwords in the inquiry is resolved. From this subset the promotions with the most elevated offers duplicated with their quality score are picked. The most essential element of the quality score is the active clicking factor. The setting considered envelops the precise question furthermore geographic data accessible from the client presenting the inquiry. Different elements of the quality score are the nature of the point of arrival and its page stacking time. (2) Our work develops on this methodology by joining private information from sensors on cellular telephones into the connection and adding security sureties to the general plan. A few late works have tended to different parts of protection safeguarding focused on promoting. Adnostic considers doing personalization at the customer and proposes a security mindful bookkeeping apparatus to accurately charge the promoters without spilling which client tapped on what advertisements. RePriv [9] bolsters checked mineworker through a program module which permits a client to control the amount of private data abandons her program and to which site. Privad proposes camouflaging client's personality and muddling/ conglomerating private data before discharging it. All these works are orthogonal to our work in that they don't determine how the personalization is done and some of them consider getting along the personalization in the customer as it were. Our personalization calculation can without much of a stretch be fused into these current frameworks. Customized Search. Here a client's advantage profile is set up in light of her skimming history and indexed lists are being re-positioned in view of how well the substance of the page coordinates her advantage. The re-positioning should either be possible by the client [10] or the internet searcher [3-4]. On the off chance that the re-positioning is finished by the web crawler then clients.

## III. Problem Statement

There is a need to collect data with various advanced sensors some of them are fitted into latest smart phones. Data collected by individual sensor won't be always helpful. Data has to be collected and aggregated to get right results. To perform this task, basic assumption is "Trusted Aggregator" which is not true. There is a need of protocol to provide data aggregation without compromising privacy with "Untrusted aggregator" and with an enhancement to handle fault tolerance.

## IV. The PDAAS Protocol

In this section, we present the PDAAS protocol, which can protect the privacy of a node against any other sensor node, the aggregator and the sink, if the aggregator and the sink don't collude. The basic idea is that, each sensor node owns two keys, one shared with the sink and the other shared with the aggregator (cell header). When queried to submit a data, each sensor node computes two keyed values from the two keys, adds the keyed values to the raw data to perturb it, and submits the perturbed data.

### A. Key Distribution

In PDAAS, each sensor node perturbs its raw data using some keyed values before submitting the data, which necessitates a key distribution process. This key distribution process includes two steps: key pre-distribution and keyed value establishment. Note that, the key distribution process must take the sensor mobility into consideration.

Before deploying the WSN, there is a key pre-distribution phase, in which a trusted key server distributes some necessary key information. This key pre-distribution is executed in an offline manner. Specifically, the key server needs to do the following:

1. For the sink, the key server generates a master key, MSK, and loads a pseudo-random function (PRF), f.
2. For each cell header $CH_j$, the key server also generates a master key, $MCK_j$, which has an ID same as the cell header's ID, and load the same PRF f. As sensor nodes are mobile, each cell header has no idea which sensor node will moves into its cell. In addition, the number of sensor nodes is large. Thus, letting each cell header shares a key with each sensor node is not efficient, as it will consume a lot of storage. In our solution, each cell header is loaded with a master key and a PRF, by which it can generate the shared key with each sensor node
3. For each sensor $s_i$, the key server first generates a long-lived key that $s_i$ shares with the sink, $lsk_i$ □ $f_{MSK}(s_i)$. As sensor nodes are mobile and can move between different cells, there is no binding between senor nodes and cell headers. Thus, the key server generates another m keys, one for each cell header, $lek_i^j = f_{MCKj}(s_i)$. The key server loads a secure hash function, Hash,

to the sink, the cell headers, and all the sensor nodes.
Note that, the sink and the cell headers need not to store the keys shared with sensor nodes, as they can reconstruct them when required. The keyed value establishment is executed at each sensor node after the WSN is deployed in the interested area and begin to operate, to derive two short-lived keyed values, one shared with the current cell header and the other shared with the sink. These two keyed values will be used in data aggregation.
The sensor nodes collect data from the environment, with some kind of mobility. In order for a sensor node to know the cell it currently stays, each cell header $CH_j$ will periodically broadcast a hello message, which includes its ID, to all the sensor nodes currently stayed in the cell. This way, each sensor that just

When queried by the user, the sink will broadcast a "data collection" request message to all the cell headers. This message includes a seed $S_{current}$ (e.g., a timestamp), which serves as an identification of the current aggregation process. Each cell header forwards this message, plus a nonce $n^{current}_j$, to all the sensor nodes currently stayed in the cell it manages. After receiving the request, each sensor node $s_i$ will derive two short-lived keyed values, one shared with the sink, $sk_i$ □ Hash($lsk_i$ || Scurrent ) , and the other shared with the cell header $CH_j$, $ek_i^j$ □ Hash($lek_i^j$ || $n_{curret}^j$). Note that, || denotes the concatenating operation.

## B. Data Aggregations

The basic idea of this protocol is that each node uses its two short-lived keyed values to compute a perturbed value, and the cell header can remove one to get the intermediate result, while the sink can remove the other to get the final aggregation result, both without accessing the raw sensor data.

Algorithm 1. PDAAS_Sensor_Aggregate.

**Input:**

Raw data item $d_i$; keyed values $ek_i^j$ and $sk_i$

**Output:**
A perturbed data item di with sensor ID Method:
1. $d_i = d_i$;
2. $d_i = d_i + sk_i$;
3. $d_i = d_i + ek_j$ i;
Sends <$s_i$, $d_i$> to the cell header.

A sensor node $s_i$ follows steps in Algorithm:
1. It simply adds the two keyed values, $ek_i^j$ and $sk_i$, to the raw data and gets the perturbed data. Then the sensor sends the perturbed data di and its ID to the cell header. Note that, "+" means modular addition as in [13]. That is, the perturbed value is computed as follows:

*Perturbed_value = sensed_value + keyed_values* mod M

We select a sufficiently large value, i.e., $M > n*D_{max}$, for M, where $D_{max}$ is the upper bound for the sensed raw data. This is necessary to remove all keyed values from the aggregation result to obtain the precise sum of the sensed data. For simplicity, we will use "+" and "−" as modular addition and subtraction in this paper.

Each cell header, $CH_j$, acts as an aggregator of the cell, and follows Algorithm

2. It first initializes the intermediate aggregation result $D_j$ to be 0, and the set $S_j$ of sensor IDs that contribute to the intermediate aggregation result to be empty. Suppose in current aggregation process, $CH_j$ receives $N_j$ perturbed data items. For each perturbed data item di received from sensor node $s_i$, $CH_j$ first computes the short-lived keyed value shared with si, using its master key, sensor ID, the hash function Hash and the PRF f. Then $CH_j$ subtracts this value from the perturbed data item, and adds the result to $D_j$.
In addition, $CH_j$ puts the IDs of the sensors that contribute into the set $S_j$. Finally, $CH_j$ sends <$S_j$, $D_j$> to the sink.

## C. Analysis of PDAAS

Correctness: A data item is perturbed by two keyed values, one shared with the cell header and the other with the sink. At the cell header, it uses the ID of the sensor that collects the data item to derive one of the keyed values, and remove it. At the sink, it derives the other keyed value and removes it. Thus, the sink can obtain an accurate sum of all the data items [7].

Privacy: First, before sending the data item, each sensor node perturbs it using two keyed values. The two keyed values are computed locally and not communicated with any other. Thus, each sensor has a distinct pair of keyed values, which can't be eavesdropped. As a result, PDAAS can protect the data privacy of each sensor node against any other node and an external passive eavesdropper. Second, PDAAS can protect the data privacy of each sensor node against the cell header or the sink, because the data item is perturbed with two keyed values, one shared with the cell header and the other with the sink.

If the cell header and the sink don't collude, neither can recover the raw data correctly. Third, PDAAS can protect the data privacy of each sensor node against a powerful external eavesdropper that can compromise a portion of the network (but not the sink and cell headers simultaneously). This is because, without compromising the sink and the cell header at the same time, the adversary can't get both the two keyed values used in perturbing data item. As a result, the adversary can't recover the raw data item. However, if the adversary compromises the sink and a cell header at the same time, or if the sink and a cell header collude, PDAAS fails to protect privacy of any sensor node using the cell header as the aggregator. Let $P_{ch}$ and $P_{sink}$ denote the probability that a cell header or the sink is compromised (or intends to collude), and let $P_{disclosure}$ denote the probability that private data of a sensor node is disclosed:

$$P_{disclosure} = P_{ch} * P_{sink} \quad (1)$$

### 1. Efficiency

In PDAAS, each sensor node keeps a constant (m + 1) number of long-lived keys, one shared with the sink, and others shared with cell headers. Thus, the storage overhead per sensor node is bounded by the constant m + 1. In each aggregation, each sensor node needs to compute two short-lived keyed values using a hash function, and adds the two keyed values to the raw data item. This will incur a constant computation overhead bounded by 2. As each sensor node sends its ID and a perturbed data item to the cell header, the communication overhead is also constant, which is $\log_2 n$ □ $\log_2 M$ bits. For the cell header $CH_j$, it collects $N_j$ perturbed data items, each with an ID, computes $N_j$ keys, and conducts Nj subtractions. Then the cell header sends an intermediate aggregation result and $N_j$ IDs to the sink.

## V. Proposed System

We propose a new protocol for mobile sensing to obtain the sum aggregate of time-series data in the presence of an untreated aggregator with fault tolerance. We propose an efficient protocol to obtain the Sum aggregate, which employs an additive homomorphism encryption and a novel key management technique to support large plaintext space and fault tolerance by creating binary tree. We propose a scheme that utilizes the redundancy in security to reduce the communication cost for each join and leave. We also propose a scheme that employs the redundancy in security to reduce the communication cost of dealing with dynamic joins and leaves. One building block of our solution is the additive homomorphism encryption scheme proposed by Castelluccia [8]. Advantages:

1. It reduces the Communication cost of dealing with dynamic joins and leaves
2. Users may frequently join and leave in mobile sensing
3. In each time period, a mobile user sends her encrypted data to the aggregator via Wi-fi, 3G or other available access networks
4. No peer-to-peer communication is required among mobile users, since such communication is nontrivial in mobile sensing scenarios due to the high mobility of users and users may not be aware of each other for privacy reasons

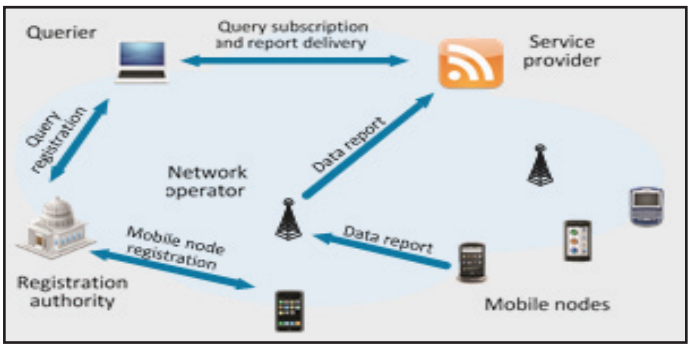


Fig. 1: Proposed Architecture Diagram

### A. Encryption & Decryption Methodology

One building block of our solution is the additive homomorphic encryption scheme proposed by Castelluccia [8]. This scheme works as follows:

**Encryption:**

1. Represent message m as an integer within range [0, M-1], where M is a large integer.
2. Let k be a randomly generated key, k $\epsilon$ {0, 1} $\lambda$, where $\lambda$ is a security parameter.
3. Output cipher c = (m + h(fk(r))) mod M, where fk is a pseudorandom function (PRF) that uses k as a parameter, h is a length-matching hash function and r is a nonce for this message.

**Decryption:**

Output plaintext m = (c - h($f_k$(r)))mod M

Protocol Implementation

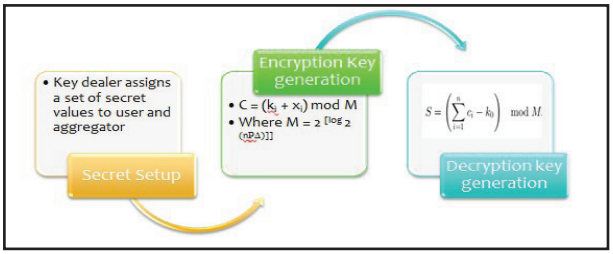


Fig. 2: Protocol Implementation

Above diagram depicts the protocol implementation for projected solution which consists of three main steps –

Secret setup -- The key dealer assigns a set of secret values (secrets for short) to each user and the aggregator. Any standard key distribution algorithm can be used for this. A unique key will be shared to all users which will be used to encrypt the date. This encrypted data will be shared to aggregator.

Encryption key generation -- In each time period, user i (i $\epsilon$ [1, n]) generates encryption key ki using the secrets that it is assigned. It encrypts its data xi by computing ci = ($k_i$ + $x_i$) mod M where M = 2[$\log_2$ (nΔ)]. Then, it sends the ciphertext $c_i$ to the aggregator. [9]

Decryption key generation -- In each time period, the aggregator generates decryption key k0 using the secrets that it is assigned, and decrypts the sum aggregate S = $\Sigma x_i$ by computing

$$S = (\Sigma c_i - k_0)$$

The keys are generated using a PRF family and a length matching hash function.

## VI. Conclusion

To facilitate the collection of useful aggregate statistics in mobile sensing without leaking mobile users' privacy, we proposed a new privacy-preserving protocol to obtain the Sum aggregate of time-series data. The protocol utilizes additive homomorphic encryption and a novel, HMAC based key management technique to perform extremely efficient aggregation. Implementation-based measurements show that operations at user and aggregator in our protocol are orders of magnitude faster than existing work. Thus, this protocol can be applied to a wide range of mobile sensing systems with various scales, plaintext spaces, aggregation loads, and resource constraints. Based on the Sum aggregation protocol, we also proposed two schemes to derive the Min aggregate of time-series data. One scheme can obtain the accurate Min, while the other one can obtain an approximate Min with provable error guarantee at much lower cost. To deal with dynamic joins and leaves, we proposed a scheme that utilizes the redundancy in security to reduce the communication cost for each join and leave. Using binary tree various blocks are formed. One idea is to form user groups, and run the Block Aggregation Scheme for each block. The aggregator is then able to estimate the sum for each block. If a subset of the users fails, we must be able to find a set of disjoint blocks to cover the functioning users. In this way, the aggregator can estimate the sum of the functioning users.

## References

[1] E. S. Cochran et al.,"The QuakeCatcher Network: Citizen Science Expanding Seismic Horizons," Seismological Research Letters, Vol. 80, 2009, pp. 26–30.

[2] C. Cornelius et al.,"AnonySense: Privacy-Aware People-Centric Sensing," 6th Int'l. Conf. Mobile Systems, Applications, and Services, 2008, pp. 211–24.

[3] D Cuff, M. H. Hansen, J. Kang,"Urban Sensing: Out of the Woods," Commun. ACM, Vol. 51, No. 3, 2008, pp. 24–33.

[4] E. De Cristofaro, C. Soriente,"Privacy-Preserving Participatory Sensing Infrastructure," [Online] Available: http://sprout.ics.uci.edu/PEPSI/index.php?page=projects.php.

[5] E. De Cristofaro, C. Soriente,"Privacy-Enhanced Participatory Sensing Infrastructure," [Online] Available: http:// sprout.ics.uci.edu/PEPSI/TR-2011-01.pdf.

[6] P.T. Eugster et al.,"The Many Faces of Publish/Subscribe," ACM Computing Surveys, Vol. 35, No. 2, 2003, pp. 114–31.
[7] R. K. Ganti et al.,"PoolView: Stream Privacy for Grassroots Participatory Sensing," 6th Int'l. Conf. Embedded Networked Sensor Systems, 2008, pp. 281–94.
[8] P. Gilbert et al.,"Toward Trustworthy Mobile Sensing," 11 Wksp. Mobile Computing Systems and Applications, 2010, pp. 31–36.
[9] M. Ion, G. Russello, B. Crispo,"Supporting Publication and Subscription Confidentiality in Pub/Sub Networks," 6th Int'l. ICST Conf. Security and Privacy in Communication Networks, 2010, pp. 272–89.
[10] D. H. Kim et al.,"Discovering Semantically Meaningful Places from Pervasive RF-Beacons," 11th Int'l. Conf. Ubiquitous Computing, 2009, pp. 21–30.

Prathipati Prasanna Kumar Pursuing M.Tech (CST) From Gitam University, Gitam University, Rushikonda, Visakhapatnam, India.

Sri G. Appa Rao M.Tech.,(Ph.D.) is working as Asst. Professor in GITAM University at Visakhapatnam. Areas of research interest includes Cryptography & network security, TC, DAA Bioinformatics.