# Sampling Based $\delta$-Approximate Data Aggregation in Sensor Equipped IoT Networks

Ji Li[1], Madhuri Siddula[1] Xiuzhen Cheng[2], Wei Cheng[3], Zhi Tian[4], and Yingshu Li[1]

[1] Department of Computer Science, Georgia State University, Atlanta, GA 30303, USA, Email: {jli30, msiddula1}@student.gsu.edu, yili@gsu.edu
[2] Department of Computer Science, The George Washington University, Washington DC, 20052, USA, Email: cheng@gwu.edu
[3] Department of Computer Science, Virginia Commonwealth University, Richmond, VA 23284, Email: wcheng3@vcu.edu
[4] Department of Electrical & Computer Engineering, George Mason University, Fairfax, VA 22030, Email: ztian1@gmu.edu

**Abstract.** Emerging needs in data sensing applications result in the usage of IoT networks. These networks are widely deployed and exploited for various efficient data transfer. Wireless sensors can be incorporated in IoT networks to reduce the deployment costs and maintenance costs. One of the critical problems in sensor equipped IoT devices is to design an energy efficient data aggregation method that processes the maximum value query and distinct set query. Therefore, in this paper, we propose two approximate algorithms to process the maximum queries and distinct-set queries in wireless sensor networks. These two algorithms are based on uniform sampling. Solid theoretical proofs are offered which can make sure the proposed algorithms can return correct query results with a given probability. Simulation results show that both $\delta$-approximate maximum value and $\delta$-approximate distinct set algorithms perform significantly better than a simple distributed algorithm in terms of energy consumption.

## 1 Introduction

With the ever-increasing population, problems of everyday sustainability have become onerous. According to the survey by United Nations, 54% of the world's current population lives in urban areas. It is also expected that the percentage increases to 66 by 2050. With this escalation in urban population, new challenges have emerged. These challenges include the constant power supply, public safety, disaster prediction, and traffic maintenance. Smart City (SC) has become inevitable to address these challenges. Many cities like New York, Detroit, Singapore, and London are working towards smart city development. These cities have adopted for various technologies like smart parking services, intelligent street light systems, sensors to redirect traffic, and water conservation. Although these applications are designed for urban living, they should also be incorporated in rural areas so that more resources could be preserved for future generations. SC networks should collect data from all over the city to provide better information. So, to collect such well spread data, they exploit various sensor equipped devices in the city to collect data and interpret information at the city level.

In today's world, all the devices from smart home devices to intelligent transport systems are well connected to the Internet. Such a network with well-connected devices is called Internet-Of-Things (IoT). As the network grows, our need for intelligent devices develop and so does the necessity to sense various activities for the convenient living of people in the cities. Some of the applications like transportation, healthcare, and seamless internet connection widely use IoT technologies. The main aim in the IoT is to reduce cost and provide faster access to the data [16] [21] [12]. But the primary challenge is that the deployment of IoT network is expensive as it requires a large number of sensing devices. Additionally, it is also important to collect data autonomously and provide intelligent methods that address the issues of dynamic traffic, accommodating new services, channel conditions, and ever-increasing user requirements.

Sensors are the building blocks for many IoT devices. Utilizing sensors as the communication media helps us resolve many of the problems discussed earlier. Previous researchers have studied the issues of routing, data management, link scheduling, coverage and topology control in networks that include sensors for communication [20], [3], [4], [7], [6], [5], [22], [23]. Although using sensors reduces the communication cost, it raises the issue of processing cost. Sensors collect data for a more extended period over vast networks. Therefore, we end up with massive data being processed at a single sensor node and thus increasing the processing cost. To address this issue, we need data aggregation at the sensor level. When data is aggregated, we only send the aggregated partial data into the network. This further raises the energy consumption issue as the aggregation costs much energy and the sensors are not equipped with a huge amount of power supply. According to [18], cost of transmitting one bit of data using wireless link is equivalent to the cost of executing 1000 instructions. So, reducing the data transmission is the major way to decrease the energy consumption. Hence, it is critical to design energy efficient data aggregation models for the sensor equipped IoT networks.

There are two kinds of aggregation queries: maximum query and distinct set query. The maximum query is to calculate the maximum of all the readings in the sensory data while the distinct set query is to calculate the unique values in the sensory data. Both the queries are essential for a given network. For example, while monitoring pollution, maximum query results in the most polluted area along with its values. Similarly, distinct set query shows the pollution levels in all the regions. Hence, the energy efficient data aggregation model should accommodate both queries in its development.

In practice, exact query results are not always necessary while approximate query results may be acceptable for conservation [9]. Therefore, in this paper, we propose an algorithm to process $\delta$-approximate maximum queries and $\delta$-distinct-set queries in IoTs. This algorithm is based on uniform sampling. Proposed algorithm returns the exact query results with probability not less than 1-$\delta$ where the value of $\delta$ can be arbitrarily small.

The rest of the paper is organized as follows. Section 2 defines the problem. Section 3 provides the mathematical proof for the $\delta$-approximate aggregation algorithms. Section 4 explains the proposed $\delta$-approximate aggregation algorithms. Section 5 shows the simulation results and the related works is discussed in Section 6. Section 7 concludes the paper.

## 2 Problem Definition

Let us assume that we have a sensor equipped IoT network with $n$ sensor nodes and $s_{ti}$ is the sensory value of node $i$ at time $t$. $S_t = \{s_{t1}, s_{t2}, \ldots, s_{tn}\}$ is used to denote the set of all the sensory data in the network at time $t$. We use $Dis(S_t) = \{s^d_{t1}, s^d_{t2}, \ldots, s^d_{t|Dis(S_t)|}\}$ to denote the distinct set of $S_t$, which contains the distinct values in $S_t$. For example, if we have $S_t = \{s_{t1}, s_{t2}, s_{t3}, s_{t4}, s_{t5}\}$ and $s_{t1} = 1, s_{t2} = 1, s_{t3} = 2, s_{t4} = 3, s_{t5} = 3$; then the $Dis(S_t) = \{1, 2, 3\}$. In this paper, we assume that the data is distributed randomly in the network while the spatial and temporal correlation of the sensory data is ignored.

In this paper, we focus on two aggregation operations on $S_t$, which are *max* and *distinct set*. The definition of the maximum value and distinct-set are as follows:

1. The exact maximum value denoted by $Max(S_t)$ satisfies $Max(S_t) = \max\{s_{ti} \in S_t | 1 \leq i \leq n\}$.
2. The exact distinct-set of $S_t$ denoted by $Dis(S_t)$ satisfies that $\forall s \in S_t, \exists s^d \in Dis(S_t), s = s^d$ and $\forall s^d_x,\ s^d_y \in Dis(S_t), x \neq y \Rightarrow s^d_x \neq s^d_y$.

A naive method that solves the max and distinct set aggregation problems has three main steps.

1. Organize all the nodes in the network into an aggregation tree. The sink node broadcasts the aggregation operation in the network.
2. All the nodes in the network submit their sensory data to the sink node along the aggregation tree.
3. The intermediate nodes in the aggregation tree aggregate the partial results during the data transmission.

However, the above method will lead to an immense communication cost and computation cost for calculating exact aggregation result. Therefore, we propose a $\delta$-approximate result for the above two aggregation operations. Let $I_t$ and $\widehat{I_t}$ are the exact aggregation result and approximate aggregation result of $S_t$ at time $t$ respectively. The definition of the $\delta$-estimator is as follows.

**Definition 1 ($\delta$-estimator).** *For any $\delta$ $(0 \leq \delta \leq 1)$, $\widehat{I_t}$ is called the $\delta$-estimator of $I_t$ if* $\Pr(\widehat{I_t} \neq I_t) \leq \delta$, According to Definition 1, the problem of computing $\delta$-approximate maximum value and $\delta$-approximate distinct-set is defined as follows.

**Input:** (1) A sensor equipped IoT network with $n$ nodes; (2) The sensory data set $S_t$; (3) Aggregation operator $Agg \in \{Max,\ DistinctSet\}$ and $\delta$ $(0 \leq \delta \leq 1)$.

**Output:** $\delta$-approximate aggregation result of $Agg$.

## 3 Preliminaries

Let $u_1, u_2, ..., u_m$ denote $m$ simple random samplings with replacement from $S_t$, $U(m) = \{u_1, u_2, ..., u_m\}$ is used to denote a uniform sample of $S_t$ with sample size $m$, then we have the following conclusions.

1. $u_i$ and $u_j$ are independent of each other for all $1 \leq i \neq j \leq m$.
2. $\Pr(u_i = s_{tj}) = \frac{1}{n}$ for any $1 \leq i \leq m$, $1 \leq j \leq n$.

Based on the above conclusions, we have the following theorem.

**Lemma 1.** *For any given value $x \in Dis(S_t)$, we have*

$$\Pr(x \notin U(m)) = (1 - \frac{n_x}{n})^m$$

*where $n_x$ is the number of appearance of value $x$ in $S_t$.*

**Proof:** $\Pr(x \notin U(m)) = \Pr(u_1 \neq x \wedge u_2 \neq x \wedge \ldots u_m \neq x)$. Since all the samples $u_1, u_2, \ldots u_m$ are independent with each other, we have

$$\Pr(x \notin U(m)) = \prod_{i=1}^{m} \Pr(u_i \neq x) = (\Pr(u_1 \neq x))^m.$$

Moreover, we also have

$$\Pr(u_1 \neq x) = 1 - \Pr(u_1 = x) = 1 - \frac{n_x}{n}.$$

Then this lemma is proved. □

To obtain the $\delta$-approximate maximum value, we need the mathematical estimator. Let $\widehat{Max(S_t)}_u$ denote the uniform sampling based estimator of exact value $Max(S_t)$. Then $\widehat{Max(S_t)}_u$ is defined as

$$\widehat{Max(S_t)}_u = Max(U(m)) = \max\{u_i \in U(m) | 1 \leq i \leq m\}.$$

Based on Lemma 1, we have the following theorem.

**Theorem 1.** *$\widehat{Max(S_t)}_u$ is a $\delta$-estimator of $Max(S_t)$ if*

$$m \geq \frac{\ln \delta}{\ln(1 - \frac{n_{min}}{n})}$$

*where $n_{min}$ is the number of appearances for the least appearing data.*

**Proof:** Based on the condition, we have

$$m\ln(1-\frac{n_{min}}{n})\leq\ln\delta$$

$$(1-\frac{n_{min}}{n})^m\leq\delta.$$

According to Lemma 1, we have

$$\Pr(Max(S_t)\notin U(m))=(1-\frac{n_{Max(S_t)}}{n})^m$$

where $n_{Max(S_t)}$ is the number of times the maximum value appears in $S_t$. Since $n_{Max(S_t)}\geq n_{min}$, we have

$$\Pr(Max(S_t)\notin U(m))\leq(1-\frac{n_{min}}{n})^m\leq\delta.$$

Then this theorem is proved. □

Let $\widehat{Dis(S_t)}_u$ denote the uniform sampling based estimator of exact result $Dis(S_t)$. Then $\widehat{Dis(S_t)}_u$ is defined as

$$\widehat{Dis(S_t)}_u=Dis(U(m)).$$

Based on Lemma 1, we have the following theorem.

**Theorem 2.** *$\widehat{Dis(S_t)}_u$ is a $\delta$-estimator of $Dis(S_t)$ if*

$$m\geq\frac{\ln(1-(1-\delta)^{n_{min}/n})}{\ln(1-\frac{n_{min}}{n})}$$

*where $n_{min}$ is the number of appearances for the least appearing data.*

**Proof:** Based on the condition, we have

$$(1-\frac{n_{min}}{n})^m\leq 1-(1-\delta)^{n_{min}/n}$$

$$(1-(1-\frac{n_{min}}{n})^m)^{n/n_{min}}\geq 1-\delta$$

$$1-\prod_{i=1}^{|Dis(S_t)|}(1-(1-\frac{n_{min}}{n})^m)\leq\delta$$

Let $n_{s^d_{ti}}$ denote the number of appearances for $s^d_{t_i}$, then we have

$$1-\prod_{i=1}^{|Dis(S_t)|}(1-(1-\frac{n_{s^d_{ti}}}{n})^m)\leq\delta$$

since $n_{min} \leq n_{s^d_{ti}}$. Moreover, according to Lemma 1, we have

$$1 - \prod_{i=1}^{|Dis(S_t)|} (1 - \Pr(s^d_{ti} \notin U(m))) \leq \delta$$

$$1 - \prod_{i=1}^{|Dis(S_t)|} \Pr(s^d_{ti} \in U(m)) \leq \delta$$

$$1 - \Pr(\widehat{Dis(S_t)}_u = Dis(S_t)) \leq \delta$$

$$\Pr(\widehat{Dis(S_t)}_u \neq Dis(S_t)) \leq \delta$$

Then this theorem is proved. □

## 4 δ-Approximate Aggregation Algorithm

The theorems in Section 3 show how to calculate the required sampling size and sampling probability according to a given $\delta$. However, we still have the following problems to be solved.

1. How does the sink node broadcast the sampling information in the whole network.
2. How to sample the sensory data from the entire network.
3. How to transmit and aggregate the partial aggregation results.

When the sample size $m$ is calculated using the theorems in Section 3, there is a simple method to sample the sensory data.

1. The sink generates $m$ random numbers from the set $\{1, 2, 3, \ldots, n\}$ and broadcasts them in the whole network.
2. The sensor node whose id is one of the $m$ randomly selected ids sends its sensory data to the sink node.

However, the above algorithm has a huge energy cost during the first step since a significant amount of sampling information needs to be transmitted. To further reduce the energy cost, we divide the whole network into $k$ disjoint clusters $C_1$, $C_2$, ... , $C_k$. Each cluster randomly selects one of its node as the cluster head. By using the method, proposed in [15], all the cluster heads in the network are organized as a minimum hop-count spanning tree rooted at the sink node. We then adopt the uniform sampling algorithm proposed by [8], described as follows.

1. The sink generates a series of random numbers $Y_i$ with the probability $\Pr(Y_i = l) = \frac{|C_l|}{n}(1 \leq i \leq m)$,
2. Let $m_l$ be the sample size of $C_l$. Then $m_l$ is calculated by $m_l = |\{Y_i | Y_i = l\}|$.
3. The sink node sends the sample size $\{m_l \mid 1 \leq l \leq k\}$ to each cluster head. Each cluster head samples the sensory data in the cluster using the above naive sampling algorithm.

When the cluster head of the $l$-th cluster receives all the sampled sensory data, $U(ml)$, it calculates the partial aggregation result $R(U(m_l))$ according to aggregation operation $Agg$ by using the following method.

$$R(U(m_l)) = \begin{cases} Max(U(m_l)) & \text{if } Agg = Max \\ Dis(U(m_l)) & \text{elsewhere} \end{cases}$$

The above process is explained in Algorithm 1. Then the partial aggregation result $R(U(m_l))$ is transmitted along the spanning tree to the sink node. To further reduce the transmission cost, the intermediate nodes in the spanning tree aggregate the received partial result while transmitting the data. The above process is explained in Algorithm 2.

**Algorithm 1** Uniform Sampling Based Aggregation Algorithm

**Input:** $\delta$, aggregation operator $Agg \in \{Max, DistinctSet\}$
**Output:** $\delta$-approximate aggregation results

**if** $Agg = Max$ **then**
  $m = \lceil \frac{\ln \delta}{\ln(1 - \frac{n_{min}}{n})} \rceil$
**else**
  $m = \lceil \frac{\ln(1-(1-\delta)^{n_{min}/n})}{\ln(1 - \frac{n_{min}}{n})} \rceil$
**end if**
generate $Y_i$ following $\Pr(Y_i = l) = \frac{|C_l|}{n}$,
$m_l = |\{Y_i \mid Y_i = l\}|$ $(1 \le i \le m, 1 \le l \le k)$, the sink sends $m_l$ to each cluster head by multi-hop communication
**for** each cluster head of the clusters $C_l$ $(1 \le l \le k)$ **do**
  generates random numbers $k_1$, $k_2$, $\ldots$, $k_{m_l}$ then broadcast inside the cluster
**end for**
**for** each cluster member of $C_l$ $(1 \le l \le k)$ **do**
  send sensory value to cluster head if its $id \in \{k_1, k_2, \ldots, k_{ml}\}$
**end for**
**for** each cluster head of the clusters $C_l$ $(1 \le l \le k)$ **do**
  receive sample data $U(m_l)$ and calculate partial result $R(U(m_l))$
**end for**

According to the analysis in Section 3, for the sample size $m$, we have

$$m = \begin{cases} \lceil \frac{\ln \delta}{\ln(1 - \frac{n_{min}}{n})} \rceil & \text{if } Agg = Max \\ \lceil \frac{\ln(1-(1-\delta)^{n_{min}/n})}{\ln(1 - \frac{n_{min}}{n})} \rceil & \text{if } Agg = Dis \end{cases}$$

Therefore, we have

$$m = \begin{cases} O(\ln \frac{1}{\delta}) & \text{if } Agg = Max \\ O(\ln(\frac{1}{1-(1-\delta)^{n_{min}/n}})) & \text{if } Agg = Dis \end{cases}$$

**Algorithm 2** Partial Data Aggregation Algorithm

**for** each node $j$ in the spanning tree **do**
  **if** $j$ is the leaf node **then**
    Send $R_j$ to its parent node
  **else**
    Receive partial results $R_{j1}, R_{j2}, \ldots, R_{jc}$ from its children
    **if** $Agg = Max$ **then**
      $R_j = \max(R_{j1}, R_{j2}, \ldots, R_{jc})$
    **else**
      $R_j = \bigcup_{i=1}^{c} R_{ji}$
    **end if**
    **if** $j$ is the sink node **then**
      **return** $R_j$
    **else**
      Send $R_j$ to its parent node
    **end if**
  **end if**
**end for**

In practice, $|R_j|$ can be regarded as a constant. According to [8], the communication cost and the energy cost of the uniform sampling based $\delta$-approximate aggregation algorithm is $O(\ln \frac{1}{\delta})$ if $Agg = Max$, while the cost is $O(\ln(\frac{1}{1-(1-\delta)^{n_{min}/n}}))$ if $Agg = Dis$.

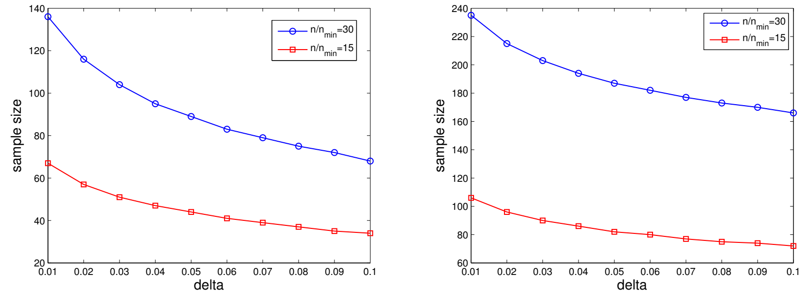
## 5 Simulation Results

To evaluate the proposed algorithms, we have simulated a network with 1000 nodes. All the nodes are randomly distributed in a rectangular region of size $300m \times 300m$, and the sink is in the center of the region. The following strategy is used to define the clusters.

1. Divide the whole region into $10 \times 10$ grids.
2. Group the nodes in the same gird into the same cluster.
3. Randomly chose the cluster head among the nodes of the same grid.

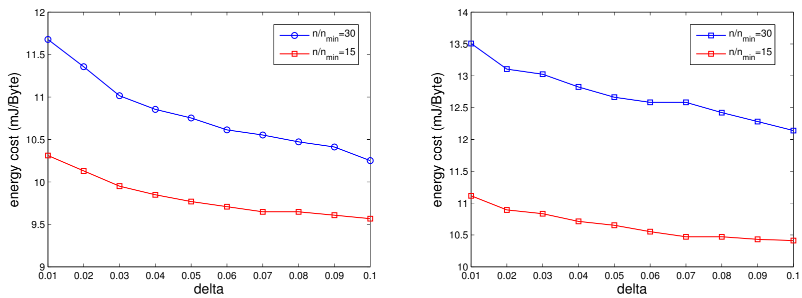
For each node, the energy cost to send and receive one byte is set as 0.0144mJ and 0.0057mJ [10]. The communication range of each sensor node is set to be $30\sqrt{2}$m [1]. This kind of simulation setting can make every sensor node communicate with its cluster head by a one-hop message.

The first group of simulations is about the relationship between $\delta$ and the sample size. The results are presented in Fig.1. These results show that the sample size increases with the decline of $\delta$. Moreover, the sample sizes are much smaller than the size of the network. For example, when $\delta = 0.01$, the sample size is about 67 for deriving $\delta$-approximate maximum value. If $\frac{n}{n_{min}} = 15$, which indicates that we just need to sample 6.7% sensory data from the network to guarantee that the estimated maximum value being equal to the actual maximum




(a) Maximum Value (b) Distinct-Set

**Fig. 1.** The relationship between $\delta$ and the sample size.



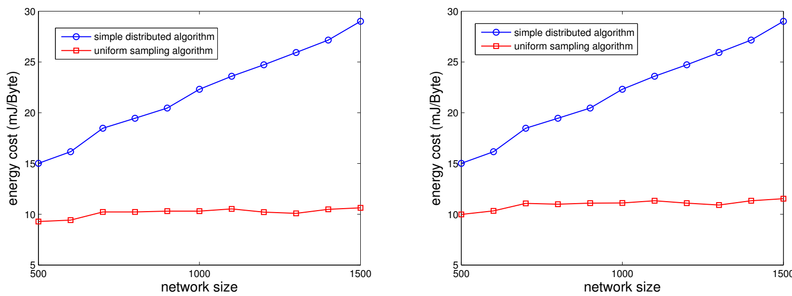

(a) Maximum Value (b) Distinct-Set

**Fig. 2.** The relationship between $\delta$ and the energy cost for the uniform sampling based aggregation algorithm.

value with the probability greater than 99%. Therefore, our uniform sampling based algorithm saves a tremendous amount of energy as it only needs a little amount of sensory data to be sampled and transmitted in the network. Moreover, we can see that in the same condition, the required sample size for the distinct-set aggregation is greater than that of the maximum value aggregation since the distinct-set aggregation needs to make sure all the distinct values are being sampled.

The second group of simulations is about the relationship between $\delta$ and the energy cost. The results are shown in Fig.2. These results indicate that the energy cost increases with the decline of $\delta$. We can also see that for the same condition, the energy cost for the distinct-set aggregation is higher than that of the maximum value aggregation, as the distinct-set aggregation has a greater sample size.

The third group of simulations is to compare the energy cost between the uniform sampling based aggregation algorithm and the simple distributed algorithm. The simple distributed algorithm is to collect all the raw sensory data and




(a) Maximum Value　　(b) Distinct-Set

**Fig. 3.** Energy cost comparison between the uniform sampling based aggregation algorithm and the simple distributed algorithm.

to aggregate the partial results during the transmission, which can always return accurate aggregation results. For the uniform sampling based aggregation algorithm, we set $\delta = 0.1$, $\frac{n}{n_{min}} = 15$, and the network size varies from 500 to 1500. The results are listed in Fig.3. We can see that for all the proposed algorithms, the energy cost increases with the increase of the network size. Moreover, for the same network size, the energy cost of the uniform sampling based aggregation algorithm is much lower than that of the naive distributed algorithm. These results indicate that the uniform sampling based aggregation algorithm performs better in terms of energy consumption. It is also to be observed that with an increase in the network size, the energy cost of the naive distributed algorithm proliferates, while the energy cost of the uniform sampling based aggregation algorithm almost remains the same. The above phenomenon indicates that the uniform sampling based aggregation algorithm has even better performance when the network size is large.

## 6 Related Works

The sampling technique has been widely used in many fields, such as quantile calculation, data collection and top-k query. For example, [14] and [13] introduce approximate algorithms to calculate the quantiles in wireless sensor networks. This algorithm reduces energy cost by using the sampling technique. By using the sampling technique, [11] develops ASAP, an adaptive sampling approach to energy-efficient periodic data collection in sensor networks, whose basic idea is to use a dynamically changing node set as samplers. [19] uses samples of past sensory data to formulate the problem of optimizing approximate top-k queries under an energy constraint. However, all the above techniques cannot be used in our problem directly since the above operations differ a lot with the maximum query and distinct-set query.

The distinct-count query in wireless sensor networks has been widely studied in many existing works, such as [17] and [2]. [17] proposes an algorithm to calculate approximate distinct-count based on approximate frequency query results. [2] also proposes an algorithm to compute the approximate distinct-count. However, this algorithm is centralized and not appropriate for large scale wireless sensors. Moreover, all the above works are for the distinct-count query, which can only reflect the size of the distinct set instead of all the content of the distinct set. Therefore, the above works still cannot be used in our problem directly.

## 7 Conclusions

In this paper, the $\delta$-approximate algorithms for the maximum value and distinct-set aggregation operations in sensor equipped IoT networks are proposed. These algorithms are based on the uniform sampling. Mathematical proofs have been made for better understanding of these algorithms. Additionally, we have also proposed mathematical estimators for the two algorithms. Moreover, we have derived the values for the sample size and the sample probability which satisfies the specified failure probability requirements of the final result. Finally, a uniform sampling based algorithm is provided.

Experiments are conducted for various delta values and the network sizes. The results are then compared between the naive method and the proposed algorithms. The simulation results indicate that the proposed algorithms have high performance with respect to the energy cost.

## Acknowledgment

This work is partly supported by the NSF under grant No.1741277, No.1741279, No.1741287, No.1741338 and the National Natural Science Foundation of China under Grant NO. 61632010, 61502116, U1509216, 61370217.

## References

1. Anastasi, G., Falchi, A., Passarella, A., Conti, M., Gregori, E.: Performance measurements of motes sensor networks. In: Proceedings of the 7th ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems. pp. 174–181. ACM (2004)
2. Beyer, K., Haas, P.J., Reinwald, B., Sismanis, Y., Gemulla, R.: On synopses for distinct-value estimation under multiset operations. In: Proceedings of the 2007 ACM SIGMOD international conference on Management of data. pp. 199–210. ACM (2007)
3. Cai, Z., Chen, Z.Z., Lin, G.: A 3.4713-approximation algorithm for the capacitated multicast tree routing problem. Theoretical Computer Science 410(52), 5415–5424 (2009)
4. Cai, Z., Lin, G., Xue, G.: Improved approximation algorithms for the capacitated multicast routing problem. In: In Proceedings of the 11th International Computing and Combinatorics Conference (COCOON 2005). vol. 8881, pp. 136–145. Springer (2005)

5. Cheng, S., Cai, Z., Li, J.: Curve query processing in wireless sensor networks. IEEE Transactions on Vehicular Technology 64(11), 5198–5209 (2015)
6. Cheng, S., Cai, Z., Li, J., Fang, X.: Drawing dominant dataset from big sensory data in wireless sensor networks. In: The 34th Annual IEEE International Conference on Computer Communications (INFOCOM 2015). pp. 531–539 (April 2015)
7. Cheng, S., Cai, Z., Li, J., Gao, H.: Extracting kernel dataset from big sensory data in wireless sensor networks. IEEE Transactions on Knowledge and Data Engineering 29(4), 813–827 (April 2017)
8. Cheng, S., Li, J.: Sampling based (epsilon, delta)-approximate aggregation algorithm in sensor networks. In: Distributed Computing Systems, 2009. ICDCS'09. 29th IEEE International Conference on. pp. 273–280. IEEE (2009)
9. Considine, J., Li, F., Kollios, G., Byers, J.: Approximate aggregation techniques for sensor databases pp. 449–460 (2004)
10. Crossbrow Inc: MPR-Mote Processor Radio Board User's Manual
11. Gedik, B., Liu, L., Philip, S.Y.: Asap: An adaptive sampling approach to data collection in sensor networks. IEEE Transactions on Parallel and distributed systems 18(12), 1766–1783 (2007)
12. Han, M., Li, L., Xie, Y., Wang, J., Duan, Z., Li, J., Yan, M.: Cognitive approach for location privacy protection. IEEE Access (2018)
13. He, Z., Cai, Z., Cheng, S., Wang, X.: Approximate aggregation for tracking quantiles and range countings in wireless sensor networks. Theoretical Computer Science 607, 381–390 (2015)
14. Huang, Z., Wang, L., Yi, K., Liu, Y.: Sampling based algorithms for quantile computation in sensor networks. In: Proceedings of the 2011 ACM SIGMOD International Conference on Management of data. pp. 745–756. ACM (2011)
15. Lachowski, R., Pellenz, M.E., Penna, M.C., Jamhour, E., Souza, R.D.: An efficient distributed algorithm for constructing spanning trees in wireless sensor networks. Sensors 15(1), 1518–1536 (2015)
16. Li, J., Cai, Z., Wang, J., Han, M., Li, Y.: Truthful incentive mechanisms for geographical position conflicting mobile crowdsensing systems. IEEE Transactions on Computational Social Systems (2018)
17. Li, J., Cheng, S., Cai, Z., Yu, J., Wang, C., Li, Y.: Approximate holistic aggregation in wireless sensor networks. ACM Transactions on Sensor Networks (TOSN) 13(2), 11 (2017)
18. Li, J., Li, J.: Data sampling control, compression and query in sensor networks. International Journal of Sensor Networks 2(1-2), 53–61 (2007)
19. Silberstein, A.S., Braynard, R., Ellis, C., Munagala, K., Yang, J.: A sampling-based approach to optimizing top-k queries in sensor networks. In: Data Engineering, 2006. ICDE'06. Proceedings of the 22nd International Conference on. pp. 68–68. IEEE (2006)
20. Wang, Y.: Topology control for wireless sensor networks. In: Wireless sensor networks and applications, pp. 113–147. Springer (2008)
21. Yan, M., Han, M., Ai, C., Cai, Z., Li, Y.: Data aggregation scheduling in probabilistic wireless networks with cognitive radio capability. In: Global Communications Conference (GLOBECOM), 2016 IEEE. pp. 1–6. IEEE (2016)
22. Yu, J., Huang, B., Cheng, X., Atiquzzaman, M.: Shortest link scheduling algorithms in wireless networks under the sinr model. IEEE Transactions on Vehicular Technology 66(3), 2643–2657 (2017)
23. Yu, J., Wan, S., Cheng, X., Yu, D.: Coverage contribution area based $k$-coverage for wireless sensor networks. IEEE Transactions on Vehicular Technology 66(9), 8510–8523 (2017)