Energy Efficient Privacy Preserved Data Gathering in Wireless Sensor Networks Having Multiple Sinks

Hayretdin Bahşi
Turkish National Research Institute of
Electronics and Cryptology
Email:bahsi@uekae.tubitak.gov.tr

Albert Levi
Faculty of Engineering and Natural Sciences
Sabancı University
E-mail: levi@sabanciuniv.edu

Abstract—Wireless sensor networks (WSNs) generally have a many-to-one structure so that event information flows from sensors to a unique sink. In recent WSN applications, many-to-many structures are evolved due to need for conveying collected event information to multiple sinks at the same time. This study proposes an anonymity method bases on k-anonymity for preventing record disclosure of collected event information in WSNs. Proposed method takes the anonymity requirements of multiple sinks into consideration by providing different levels of privacy for each destination sink. Attributes, which may identify of an event owner, are generalized or encrypted in order to meet the different anonymity requirements of sinks. Privacy guaranteed event information can be multicasted to all sinks instead of sending to each sink one by one. Since minimization of energy consumption is an important design criteria for WSNs, our method enables us to multicast the same event information to multiple sinks and reduce energy consumption.

I. Introduction

Recent technological advances lead to produce low cost wireless sensors for observing many physical phenomenons of world like temperature, humidity etc. As wireless sensor technology takes progress, missions of WSNs get complicated so that they are used in human, enemy, habitat, structure or traffic monitoring applications. With the advent of wireless body are networks, applications for health monitoring of patients outside the hospitals or home-caring of elderly people have designed and implemented widely.

Recent WSNs have began to collect much more information than simple WSNs observing temperature or humidity value of an environment. In an addition to spatio-temporal information of an event, especially in object monitoring applications, other attributes of monitored objects are gathered by sensors. For example, traffic monitoring applications collect velocity, direction and size information of a vehicle.

As the complexity of wireless sensor applications increase, structure of WSNs have evolved in order to meet the new application requirements. WSNs generally have many-to-one structure so that sensors collect event information from the area and send to a unique sink. Some recent sensor applications have begun to use many-to-many structure which actually means there exist multiple sinks in the deployed environment. WSN application may need to send the same event information to different sinks rather than a unique sink. For example, in a home-caring application for elderly people, information about the elderly person can be sent to a family member and a nurse at the same time.

As capability of WSNs are enhanced, privacy preserving is getting one of the major problems in these networks. Huge amount of information about an individual is collected and distributed. On the other side, individuals generally need to restrict the details of personal information. Therefore, countermeasures for privacy threats have to cover the both needs, enabling data collection and restricting the storage of some private parts. On the other side, in most of the WSNs, minimization of energy consumption is one of the primary criteria due to limited battery capacity or unavailability of battery replacements. All other security countermeasures as well as the privacy preserving solutions have to perform their works with minimum energy.

In this study, energy efficient privacy preserving method is proposed for WSNs having multiple sinks. Privacy requirement level of each sink is assumed to be different from each other which can be a realistic scenario in recent WSN applications. Our proposed method meets all the privacy requirements by consuming low amount of energy.

This paper is organized as follows: In Section II, motivation of the study and some background information are given. This section also includes the description of threat/network model and statement of our contributions. Section III discusses the details of proposed method. Section IV shows the experimental results of simulations. Literature review of the topic is presented in Section V. Section VI concludes the study.

II. Motivation and Background

Privacy is the ability of an individual or group to decide which information about themselves would seclude or which information would revealed to whom. Therefore, a privacy framework have to use methods which can be easily adapted to different requirements of applications and users.

Widely usage of WSNs in monitoring applications make people's life easy but they may cause violation of privacy. Untrusted parties can access to the collected information by eavesdropping, physical capturing of sensors or unauthenticated remote accessing to sensors or central databases [1]. Data encryption and authentication mechanisms can prevent these types of threats. Although applications of these mechanisms are not straightforward in WSNs, due to limitations like

physical capturing possibility of cryptographic credentials or limited battery usage, many studies proposed various methods in order to solve these security problems under the specified limitations [2], [3].

Restricting the details of information gathered by WSN is another effective mechanism for preserving of privacy. These restrictions may be required for prevention of privacy violations done by un-trusted parties since any privacy risk caused by a potential data loss incident can be reduced with the help of these restrictions. However, privacy requirements of individuals also force the designers of WSNs carry out some restrictions for data shared with trusted parties. These trusted parties are actually sinks in WSNs. There are mainly two reasons for these types of requirements: First one is trusted party can be captured physically or logically by un-trusted parties. Especially in applications, where attackers have high motivation or where major vulnerabilities exist including lack of physical security, data shared with trusted parties have to obey some privacy criteria. Enemy tracking, health monitoring, or wild-life monitoring applications are examples for these types of applications. The second reason is that individuals generally want to hide detail information of their personal lives directly from the trusted party. For example, in applications where health monitoring is done outside the hospital, individuals do not want their exact location and time information to sent hospital particularly in non-urgent times. However, they may want to weaken their privacy requirements in urgent times for getting appropriate medical helps. If there are many sinks in WSN, requirements of individuals may change for each sink as well. Revisiting health monitoring applications, information about individuals may be sent to the family members as well as to hospital. In these applications, individuals may willing to share more detail information with their family members but not detailed one with hospital.

Many studies about data privacy has been done in the database field under the name of "privacy preserving data publishing" [4]. The main aim is to provide privacy of data tables which are exchanged by other parties. A generic example is the application where hospitals share medical records with medical research institutions. At first glance, it may be assumed that privacy problem can be easily solved by stripping off the attributes which identifies individuals like name, social security number etc. However, it is shown that it is possible to identify the owner of a record by using the residual data and other public information sources. This attack is called "re-identification attack" [5]. Re-identification attack bases on the assertion that some attributes, called quasi-identifiers, can easily help to identify the individuals although they do not uniquely identify them. Anonymity, which is defined as being not identifiable of an individual within set of individuals [6], is used as a privacy criterion in order to make data resistant to "re-identification attacks". "k-anonymity" brings a specific restriction to anonymity so that an individual is hided among at least k other individuals [5]. Quasi-identifier attributes are generalized or suppressed in order to meet the requirements of anonymity. In a generalization operation, attribute is re-
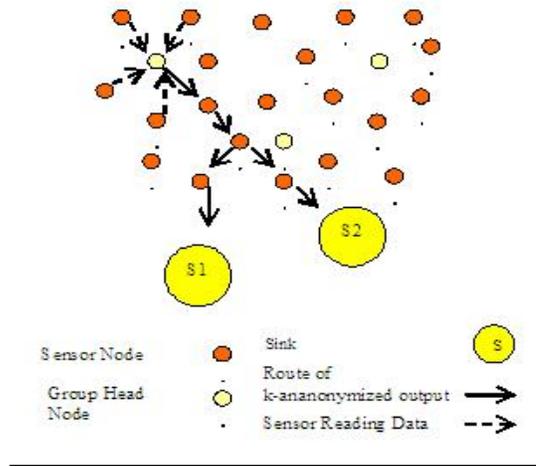

Fig. 1.   Network Model

placed by a more general one like replacement of birth data "04.05.1977" by 1977. One attribute or all attributes of a record are deleted in a suppression operation. These operations cause information loss so anonymity solutions intrinsically try to solve a trade-off between information loss and privacy. They try to cause minimum information loss while achieving the required level of privacy.

A. Threat and Network Model

Our threat model bases on the requirement that the individuals do not want sinks to identify their records among other records of k individuals within a specified time-frame through the quasi-identifier fields of records. For simplicity, it is assumed that one event record is generated for each individual within that time-frame. The required privacy levels of each sink differs so that suppose that there are n sinks, each ith sink has a privacy level p_i where each level requires to share k_i-anonymous data with i^{th} sink and inequality of $k_1 < k_2 ... < k_n$ is valid.

Sensors are clustered in separate sensor groups according to sensor localizations where each group has a group head sensor. In our method, each sensor conveys its readings to group heads, they k-anonymizes data and multi-cast the anonymized output to all sinks. Network model is shown in Figure1.

B. Our Contribution

In this paper, k-anonymity notion is adapted as a privacy framework for WSN applications having multiple sinks. Collected event information is iteratively k-anonymized for all sinks each having different privacy levels. Encryption operation with appropriate key management schema is used in addition to generalization in order to meet the different requirements in one k-anonymized output. Achieving all privacy requirements in one output considerably decreases the energy consumption so that this output can be multi-casted to multiple sinks instead of sending different outputs for each sink. It is shown that proposed method can make WSN save energy up to

48% while preserving the required privacy levels. Bottom-up clustering idea is used during k-anonymization process.

III. PROPOSED ANONYMIZATION METHOD

Proposed k-anonymization method, iterative k-Anonymization (IKA), basically produces a common k-anonymous data that meets each requirements of sinks by the help of encryption operation in addition to generalization operation. The main aim is to meet the privacy requirements with the minimum information loss. IKA bases on hierarchical bottom-up clustering idea. Quasi-identifier attributes of records are extracted from event data and they are feed as input vectors to iterative hierarchical clustering process. Basic idea is partitioning the input vectors into clusters where each cluster has at least k vectors. During the clustering, each cluster generates a representative vector which contains common generalized values or encrypted versions of attributes of all vectors belonging to that cluster. Vectors of clusters are all replaced with this representative vector in the anonymous output. Since an appropriate distance function is used during clustering and appropriate generalizations, this replacement is expected to create minimum information loss. k-anonymization work takes place in group head sensors.

Subsection III-A explains how collected information is represented in our proposed method. In Subsection III-B, distance metric which is used in the clustering process is described. Subsection III-C briefs how a common output is formed for meeting the needs of each sink. Subsection III-D gives the details of bottom-up clustering process which is the core of the proposed method.

A. Data Representation

Suppose input data is a table T having m attributes, r records. T_{ij}, represents the j'th attribute of the i'th record where, $\{i : 1 \leq i \leq r\}$ and $\{j : 1 \leq j \leq m\}$. Table T is represented by a set of bit strings B, where B_{ij} is bit string representation of j'th attribute of i'th record. k'th bit of B_{ij} is shown as $B_{ij}(k)$. Suppose that j'th attribute of table is categorical and there are d_j distinct values. These values are indexed by k and shown as $V_j(k)$ where $\{k : 1 \leq k \leq d_j\}$. Bit string of this categorical attribute has a size of d_j and formed as follows:

$If \ T_{ij} = V_j(k) \ then \ B_{ij}(k) = 1 \ else \ B_{ij}(k) = 0 \ as \ \forall k : 0 \leq k \leq d_j$,

If attribute is numerical, the range of attribute is divided into equal-sized intervals. Assume that j'th attribute is numeric and attribute range is divided into d_j number of intervals. Each interval is indexed by k. Bit string representation of this numeric attribute has a size of d_j and formed as follows:

$If \ T_{ij} \ intersects \ with \ k'th \ interval, \ then \ B_{ij}(k) = 1 \ else \ B_{ij}(k) = 0 \ as \ \forall k : \ 0 \leq k \leq d_j$

B. Information Loss Metric

Calculating the data loss of k-anonymous data is needed to predict the performance of our proposed method under different k-anonymity parameters. In our study, we use the entropy

TABLE I
A SAMPLE BIT STRING REPRESENTATION SET

Records	B_{i1}	B_{i2}	B_{i3}
T_1	00010	01000	10000
T_2	01100	11100	01111

TABLE II
A SAMPLE NORMALIZED VERSION OF BIT STRING REPRESENTATION SET

Records	$\overline{B_{i1}}$	$\overline{B_{i2}}$	$\overline{B_{i3}}$
T_1	00010	01000	10000
T_2	$0\frac{1}{2}\frac{1}{2}00$	$\frac{1}{3}\frac{1}{3}\frac{1}{3}00$	$0\frac{1}{4}\frac{1}{4}\frac{1}{4}\frac{1}{4}$

concept of information theory to measure the information loss [26]. The difference of entropies between the k-anonymous data and the original data constitutes the information loss. Suppose that T is the input data set having r records and m attributes, B is the bit string representation of this data set and C is the random variable that gets the probability value of an attribute value in a k-anonymous data entry being the actual attribute value in the original data. Assume that all the entries of B is normalized according to the number of bits having value "1" in that entry (from now on we refer "true bit" to a bit having value "1") and normalized version forms data set $\overline{B}$. A sample data set is shown in Table I. Here, there are two records; each record has three attributes; each attribute is categorical and each has five distinct attribute values. Table II shows the normalized version of data. During normalization, each entry is divided by the number of true bits in the corresponding bit string entry.

Information loss of a data table T, $IL(T)$, is equal to the conditional entropy, $H(C \mid B)$. Here, conditional entropy gives the uncertainty about the prediction of the original attribute values of a record when we have the knowledge of corresponding k-anonymous bit strings of that record. Original data has only one true bit in each bit string because each original data entry corresponds to one attribute value. However, in k-anonymous data, each entry may have more than one attribute value and each attribute value is represented by an additional bit. Therefore, if an entry has only one true bit, that entry does not have information loss. In this situation, we have no doubt that this true bit is the true bit that comes from the original data. As the number of true bits increases, disorder of the data increases because it is harder to predict which one of them is the original true bit. Prediction gets harder because information is lost due to the increase in the number of true bits. Conditional entropy, which is used in order to calculate the disorder of the data, is a well measurement tool for the information loss. Conditional entropy $H(C \mid B)$, which is equal to information loss of table T, $IL(T)$, can be found as follows:

$$IL(T) = H(C \mid B) = \sum_{B_{ij} \in B} p(B_{ij}) H(C \mid B = B_{ij})$$

$$IL(T) = - \sum_{B_{ij} \in B} p(B_{ij}) \sum_{k \in \{1..z\}} p(C = k \mid B_{ij}) \log p(C = k \mid B_{ij})$$

(1)

In Equation 1, it is assumed that each attribute is con-

verted to bit strings having size z. This means all categorical attributes have z distinct attribute values and all numerical attributes have z number of interval ranges. Also, it is assumed that all k's, where the equalities of $p(C = k \mid B_{ij}) = 0$ are true, are excluded from the summation. C random variable can take values from the set $\{1..z\}$. Actually, $\overline{B}$ is calculated for finding the value of this random variable.

$$p(C = k \mid B = B_{ij}) = \overline{B}_{ij}(k) \; for \; each \; k : \; 1 \leq k \leq z \quad (2)$$

In Equation 1, it is assumed that each record has equal probability to be chosen and each attribute of record has the same probability, therefore probability mass function of j'th attribute of i'th record, $p(B_{ij})$, is calculated as $p(B_{ij}) = \frac{1}{m.r}$. Equation 1 can be rewritten as follows:

$$IL(T) = - \sum_{B_{ij} \in B} \frac{1}{m.r} \sum_{k \in 1..z} \overline{B}_{ij}(k) . \log \overline{B}_{ij}(k) \quad (3)$$

Suppose that F is the array that contains the number of true bits of the bit string array B. Total number of true bits in B_{ij} is F_{ij}. Total number of elements in $\overline{B}_{ij}(k)$ that has the value of $\frac{1}{F_{ij}}$ is equal to F_{ij}, and the rest is zero. Therefore, the second sum operation of Equation 3 yields the value, $\log \frac{1}{F_{ij}}$. The simplest equation for the information loss of data table T, $IL(T)$, can be calculated as follows:

$$IL(T) = - \sum_{F_{ij} \in F} \frac{1}{m.r} \log \frac{1}{F_{ij}} = \frac{1}{m.r} \sum_{F_{ij} \in F} \log F_{ij} \quad (4)$$

C. Iterative Anonymization Model

In the WSN, there are n sinks and $n-1$ symmetric encryption keys which are labelled as $e_1, e_2..e_{n-1}$. $i'th$ sink contains list of the keys as $e_i, e_i..e_{n-1}$. Each group head sensors store all the $n-1$ keys. In IKA, anonymization is completed in n iterative steps as shown Figure 2. In the first step, by using only generalization operation, input data is $k1$-anonymized. In the second step, $k1-$anonymous data is $k2$-anonymized by encrypting the chosen data parts by e_1. For each ith step to n^{th} step, anonymization is done by encryption using key, e_{i-1}. The output after n^{th} iteration is multi-casted to all sinks.

After the arrival of anonymous data, each sink decrypts the data with their keys. The result data after decryption actually has the level of privacy required for that sink. i^{th} sink can only recover the encrypted operations done after the ith iterations because it has the corresponding keys. Data parts encrypted by the keys, $e_1..e_{i-2}$, are not decrypted therefore they can be considered as suppression operations for that sink. 1st sink, which has to get data with lowest privacy criteria, can decrypt all the encrypted parts and the result data is actually k_1-anonymized. In the other side, n^{th} sink has no any key and gathers data as k_n-anonymized.
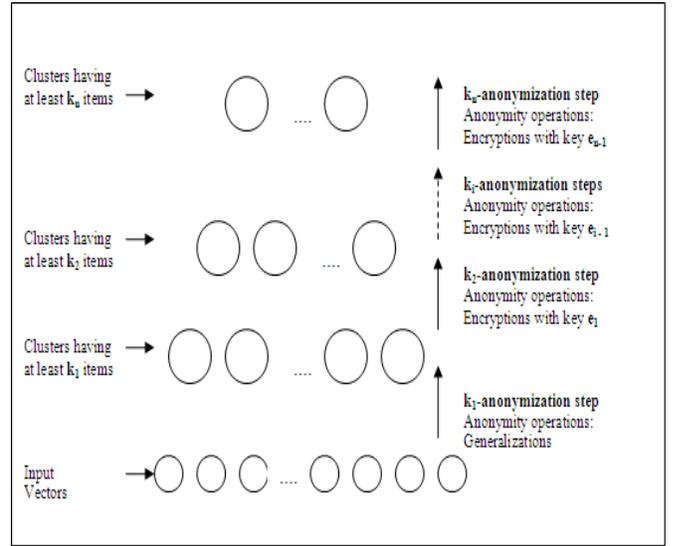

Fig. 2. Steps of Iterative Anonymization

D. Bottom-Up Hierarchical Clustering Process

Method bases on forming clusters of input vectors iteratively. Each cluster numerated as C_j^l in each epoch, l, contains a number of input vectors, N_j^l, and a representative vector, R_j^l where j is index number of cluster. Suppose that k^{th} data item of representative vector is denoted as $R_j^l[k]$. Representative vector is actually anonymized output of input vectors belonging to that cluster which is formed by generalization and encryption operations of some data parts of vectors.

Hierarchical clustering process starts with the assumption that each input vector constitutes separate cluster and that vector is also representative vector of the cluster. In each epoch, by using the information loss metric described in Section III-B, distances between each cluster are calculated. Distance between any two clusters is actually equal to the information loss that may occur if both clusters are merged. Two clusters having smallest distance, assume that clusters, C_s^l and C_t^l, are chosen for merging. New bigger cluster, C_u^{l+1} which contains the vector items of both clusters is formed and old two clusters are deleted. N_u^{l+1} is equal to the sum of N_s^l and N_t^l. If the anonymity operation is generalization, $R_u^{l+1}[k]$, is equal to the XOR of $R_s^l[k]$ and $R_t^l[k]$. If operation is encryption, representative vector, $R_u^{l+1}[k]$, is calculated as follows (Encryption function, E, input to function, x, encrypted output, E(x), concatenation operation, $\parallel$):

If $R_s^l[k] \neq R_t^l[k]$ then $R_u^{l+1}[k] = E(R_s^l[k] \parallel R_t^l[k])$ otherwise $R_u^{l+1}[k] = R_s^l[k]$

In Figure 3, a sample merging operation is shown. Two clusters having representative vectors, $(0011, 1010, 1000)$ and $(0011, 1100, 1000)$ are merged. If anonymization operation is chosen as generalization, by using of XOR operation, representative vector of new cluster is computed as $(0011, 1110, 1000)$. In the case of encryption operation, first and third items remains as the same value in the new representative vector because they are identical in both clusters.
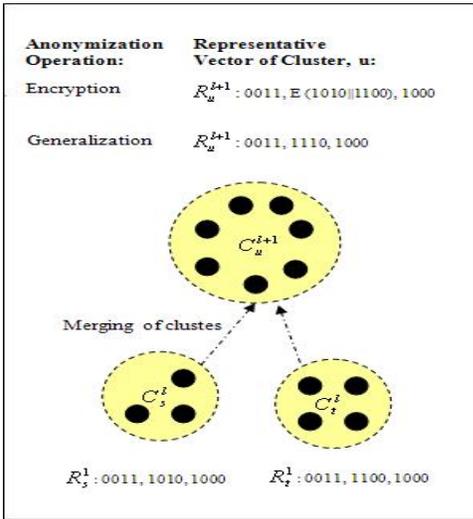
Fig. 3. Merge Operation of Clusters

Second item is encryption of $1010||1100$.

Clustering process occurs in each iteration of anonymization model described in Section III-C. In that model, each iteration takes k_i-anonymized output, clustering operations are completed until data is k_{i+1}-anonymized. In the first iteration, , where raw data is k_1-anonymized by generalization operations. In the second one and the rest of all iterations data is anonymized to a higher level by encryption operations where different key is used in each iteration.

IV. PERFORMANCE EVALUATION

Main aim of k-Anonymity solutions is providing the required privacy level with minimum information loss. However another factor, minimization of energy consumption, is an important criteria in WSNs. A sensor node consumes energy for different processes like event sensing, CPU processing, or transmitting/receiving data packets. Among these processes, transmission/reception operations consumes much of the energy so that studies [27] show that energy consumption rates for transmission/reception is over three orders of magnitude greater than the energy consumption rates for encryption. Since each sensor node acts as a router for the messages of other nodes and one message goes over many hops in the network, energy saving for transmission/reception operations becomes a crucial design criterion. Shortening the length of messages and decreasing the number of travelled hops would help to reduce energy enormously.

In a WSN topology where there are multiple sinks and each sink has different privacy criteria, the basic solution of anonymization is that group head sensor anonymizes the data, produces different outputs for each sink and sends each output to related sink in different paths as shown in Figure 4 (In this figure, there are two sinks in WSN). However, IKA produces unique output which is ready for multicasting. One anonymized output is sent to a multicast point. After reaching to multicast point, one copy of data is sent to sink1

and the other copy is sent to sink2 as presented in Figure 5. Multicasting schema decreases the number of travelled hops. Assume that the number of hops in the shortest route from group head sensor, G, to Sink1 and Sink2 is represented as $h_{G,Sink1}$, $h_{G,Sink2}$ respectively. Also assume that the hop distance between G and multicast point, M, is $h_{G,M}$, distances from M to Sink1 and Sink2 are $h_{M,Sink1}$, $h_{M,Sink2}$. Our method finds the appropriate node for M so that $h_{G,M} + h_{M,Sink1} + h_{M,Sink2}$ is minimum and the following inequality holds:

$$h_{G,Sink1} + h_{G,Sink2} > h_{G,M} + h_{M,Sink1} + h_{M,Sink2}$$

In order to prove the decrease of number of hops that messages travel, expected number of message relaying is calculated as below. Suppose the WSN field has the size of $X_{region}.Y_{region}$ and WSN has two sinks having different privacy criteria. Group head nodes are uniformly deployed in this area. Sink1 is located at $X_{sink1}, 0$ and sink2 is located at $(X_{sink2}, 0)$. The group head nodes are uniformly distributed. Expected distance value of a group head node to sink1, d_{sink1}, is calculated as follows:

$$d_{sink1} = \int_{x=0}^{X_{region}} \int_{y=0}^{Y_{region}} \sqrt{(x - X_{sink1})^2 + (y-0)^2} f(x)f(y)dxdy \tag{5}$$

Here, $f(x)$ and $f(y)$ are the probability distribution functions of group head node coordinates. Since they are uniformly distributed, they are chosen as $f(x) = 1/X_{region}$ and $f(y) = 1/Y_{region}$. The expected number of hops an event message travels from a group head node to sink1 is:

$$h_{sink1} = d_{sink1}/R \tag{6}$$

where R is the distance of each hop. From group head node to sink1, an event message travels h_{sink1} hops which is calculated as follows:

$$h_{sink1} = \int_{x=0}^{X_{region}} \int_{y=0}^{Y_{region}} \frac{\sqrt{(x - X_{sink1})^2 + (y-0)^2}}{X_{region}Y_{region}R} dxdy \tag{7}$$

h_{sink2}, the number of hops for reaching sink2 can be calculated with nearly the same formula with the exception that sink2 is located at the coordinate value, $(X_{sink2}, 0)$.

Let's assume that the size of WSN field is 500m x 500m, distance of each hop, R, is $10m$, location of sink1 is at $(100, 0)$ and location of sink2 is at $(400, 0)$. Expected number of hops for reaching to each sink from group head nodes is computed as 32.86 by using Equation 7. Since different anonymized output is sent to each sink in the first option, total number of hops is 65.72. Minimization of length of multicast route can be achieved when multicast point has an expected distance of 28.2 hops from group head node. Total length amount of multicast route is 30.07. This is considerably lower than the previous alternative. We assume that each group head node covers a region having 50m x 50m and each sensor is uniformly distributed through the region. The expected number
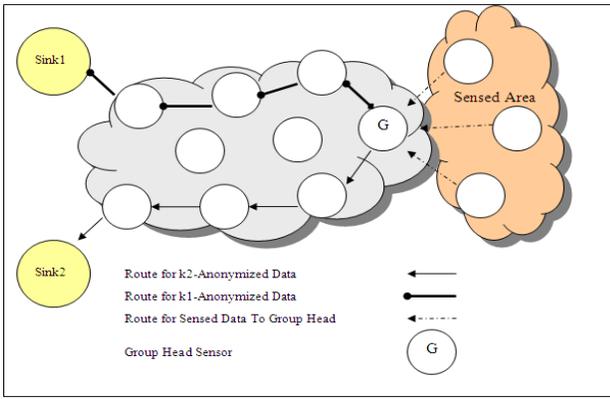
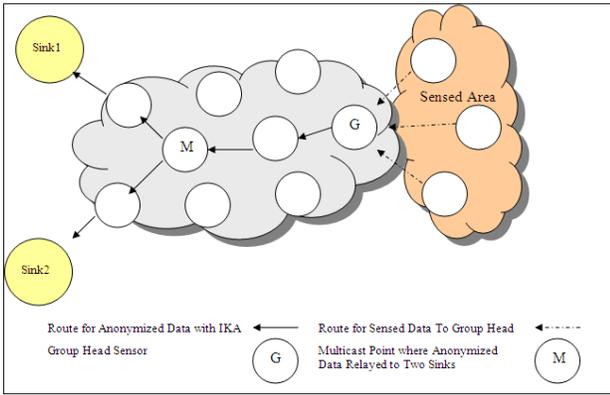Fig. 4. Routes when multiple k-anonymized outputs are generated

Fig. 5. Routes when IKA anonymized output is multicasted to sinks

of hops from sensor node to group head node is calculated and they are taken into consideration in calculation of energy consumption. Energy consumption is not only depend on the number of hops, length of the messages are also important for the final results. Message lengths are taken into consideration during energy calculations.

Energy consumption parameters are determined according to the experimental results presented in [27]. We assume that the data is processed in Sensoria's WINS NG RF subsystem with MIPS R400 processor where encryption algorithm is AES. The transmission/reception, transmission/encryption and encryption/decryption energy consumption ratios for the same length of data are shown in Table 10. The transmission and reception rate is taken as 10 Kbps and power is 10mW. In all energy calculations, only event data processes are taken into consideration. We assume that transmission energy of each byte T_T is 1.5 units (the actual unit is not so important since we eventually calculate energy saving as a ratio), reception energy T_R is 1 unit, encryption and decryption energy, T_E and T_D, are 4.29e-4 units.

Assume that energy consumption of method named as "multipath method" where each anonymized output is sent to each sink separately is denoted as $E_{multipath}$ and energy consumption of multicast method is represented as $E_{multicast}$.

Energy Consumption Ratios	Ratio Value
Transmission/Reception	1.5
Transmission/Encryption	2333.34
Encryption/Decryption	1

TABLE IV
RESULTS OF MULTIPATH METHOD WITH DATA SETS HAVING VARIOUS
RECORD NUMBERS

Number of Records	Info. Loss For Sink1	Info. Loss For Sink2
100	0.59	1.08
300	0.46	1.05
500	0.53	1.04
1000	0.44	0.88

Energy gain ratio of IKA, EG_{IKA}, is computed as follows:

$$EGA_{IKA} = 1 - \frac{E_{multicast}}{E_{multipath}} \qquad (8)$$

Table IV and Table V give the results of anonymization process according to multipath and multicast methods respectively. Information loss results for sink1 are the same in both methods. Multipath directly sends k_1-anonymized output to sink1. On the other side, IKA generates an output which is generalized to reach anonymity level $k1$ and $k1-$anonymized output is converted to k_2-anonymous data by encryption operations. Sink1 decrypts the encrypted parts and gets the k_1-anonymized data which is exactly the same data received by sink1 in multipath method. However, information loss of multipath method for sink2 is greater than loss of multicast method in each experiment. Encrypted parts show suppression behavior for sink2 due to lack of decryption key in sink2. Suppression causes more information loss than generalization operation so that IKA suppresses all the columns of vectors belonging to one cluster and multipath method uses only generalization operation for sink2. However, encryption enables us to multicast the data which results with high amount of energy savings as shown in Figure 6. Energy gain increases up to 48% when the record number is 1000. Energy gain increases as the number of records gets bigger so that this result shows the effectiveness of IKA in data sets having high number of records.

V. RELATED WORK

Studies on privacy problem mostly concentrated on achieving sharing of databases under the required privacy constraints in order to make efficient knowledge-based decisions. Generic name, "Privacy Preserving Data Publishing", is given to these

TABLE V
RESULTS OF MULTICAST METHOD WITH DATA SETS HAVING VARIOUS
RECORD NUMBERS

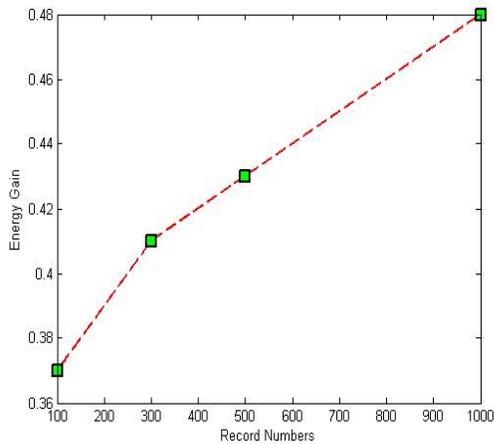| Number of Records	Info. Loss For Sink1	Info. Loss For Sink2
100	0.59	1.67
300	0.46	1.59
500	0.53	1.55
1000	0.44	1.40

Fig. 6. Energy Gain vs Record Numbers by Multicast Method

efforts [4]. k-anonymity notion is introduced by Samarati and Sweeney in [5]. It is shown that k-anonymization with minimum number of suppression is NP-hard [11]. Some optimal k-anonymization algorithms have been presented which may be feasible for small sized data sets [12], [13]. Greedy heuristics algorithms are proposed to find approximate solutions for large data sets [14], [15].

All these k-anonymity solutions solve the prevention of "record linkage attack" which is actually finding the owner of a record through quasi-identifier attributes. However, it is shown that without finding the exact owner of a record, if sensitive attribute exists in a record, it may be possible to identify sensitive attribute of an individual in some circumstances by an attack called "attribute linkage attack" [4]. This problem is also named as "attribute disclosure" [16]. In order to prevent attribute linkage and record linkage together, k-anonymity notion extended in some studies. Machanavajjhala et. al. extended k-anonymity with a l-diversity notion that also prevents the identity disclosure when attackers have background knowledge [16]. Notion of p-sensitive is introduced so that p of k-anonymized records having identical quasi-identifier attribute values have to have distinct sensitive attribute values [17]. Generalization hierarchies are constructed for sensitive attributes and extended version of p-sensitive notion is adapted in [18]. An additional requirement, t-closeness, for l-diversity is defined in [19]. In this study, distribution of sensitive attributes in a record set having identical quasi-identifier attribute values are adjusted so that it is close to the distribution of that attribute in overall data set.

Anonymity is considered as hiding the identities of sender or receiver of a communication in data and communication networks for many years. DC-Net and mix-net solutions are proposed for achieving sender or receiver anonymity [8], [7]. Especially, mix-net idea have been used in many practical Internet applications like web and e-mail [9], [10]. In ad-hoc networks, routing protocols for anonymous transmission of the data packets are designed.

Studies about the anonymity problem in WSNs basically try to hide location or time information of the events. Gruteser et al. [20], [21] proposed anonymity solutions for providing high degree of privacy in a sensor network that gives location-based services. Ozturk et al. [22] proposed phantom routing method for hiding location information of originator sensor node in a sensor network. Threat model is based on an existence of only one movable adversary node in the environment. Location privacy protection of receiver in a WSN is provided by a routing protocol in [23]. Proposed routing protocol prevents the eavesdropper to identify the receiver by tracing the wireless packets. It randomizes the routing paths and injects fake packets in order to mislead eavesdroppers. Wadaa et al. [24]studied on providing anonymity of coordinate system, cluster and routing structures during the network setup of a WSN. Protection of location privacy is guaranteed by k-anonymity in location based services those are given on mobile networks [25]. None of these studies do not propose solution for anonymity problem in WSNs having multiple sinks.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, a k-anonymization model for WSNs having multiple sinks are proposed. Study bases on a realistic threat model which states that each sink has different level of privacy requirements. Proposed method reduces energy consumption while fulfilling the required different privacy levels. Method uses encryption operations with generalization operations in order to have one common anonymized output. Multicasting of this output enables WSN to reduce a great amount of energy so that in some experiments energy reduction can increase to 48%. Multicasting method can degrade the data quality of some sinks. Owner of WSNs has to decide about the trade-off between energy saving and information loss. The intelligence of choosing data parts for encryption can be enhanced for decreasing the information loss caused by the proposed method as a future work.

REFERENCES

[1] H. Chan, A. Perrig, Security and Privacy in Sensor Networks, Computer, vol. 36, no. 10 pp.103-105, 2003
[2] D. Boyle, T. Newe, Security Protocols for Use With Wireless Sensor Networks: A Survey of Security Architectures, In Wireless and Mobile Communications, 2007
[3] Y. Xiao, V. K. Rayi, B. Sun, X. Du, F. Hu, M. Galloway, A Survey of Key Management Schemes in Wireless Sensor Networks, In Computer Communications, Volume 30, Issue 11-12, pages: 2314-2341, 2007
[4] B. C. M. Fung, K. Wang, R. Chen, P. S. Yu, Privacy-Preserving Data Publishing: A Survey on Recent Developments, In ACM Computing Surveys, 2009
[5] L.Sweeney. k-anonymity: A model for protecting privacy. *Int'l Journal on Uncertainty, Fuziness, and Knowledge-based Systems,* 10(5):557-570, 2002.
[6] A.Pfitzmann, M. Khntopp. Anonymity, Unobservability, and Pseudonymity- A Proposal for Terminology. *In H. Federrath, editor, DIAU'00, Lecture Notes in Computer Science* 2009/2001: 1-9, 2000.
[7] D. Chaum. The Dining Cryptographers Problem: Unconditional Sender and Receipent Untraceability. *In Journal of Cryptology*, 1(1):65-75, 1988.
[8] D. Chaum. Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms. *In Communications of the Associations for Computing Machinery*,24 (2):84-88, 1981.

[9] C. Gulcu, G. Tsudik. Mixing E-mail with BABEL. *In Symposium on Network and Distributed Systems Security (NDDS '96)*, San Diego,California, 1996.

[10] M. K. Reiter, A.D. Rubin Anonymous Web Transactions with Crowds. *In Communications of the ACM*, 42(2):32-48, Feb 1999.

[11] A. Meyerson, R. Williams. On the complexity of optimal k-anonymity.*In Proc. of the 23rd ACM SIGMOD-SIGACT-SIGART Symposium on the Principles of Database Systems,* June 2004.

[12] K. Lefevre, D. J. Dewitt, R. Ramakrishnan, Incognito: Efficient full-domain k-anonymity, In Proc. of ACM SIGMOD, Baltimore, 49-60, 2005

[13] P. Samarati, Protecting respondents' identities in microdata release, In IEEE Transactions on Knowledge and Data Engineering 13, 6, 1010,1027, 2005

[14] G. Aggarwal, T. Feder, K. Kenthapadi, R. Motwani, R. Panigraphy, D. Thomas, A. Zhu. Anonymizing tables.*In Proc. of the 10th Int'l Conference on Database Theory,* January 2005.

[15] Datafly: A System for providing anonymity in medical data. In Proc. of the IFIP TC11 WG11.3 11th International COnference on Database Security 11: Status and Prospects. 356-381, 1998

[16] A. Machanavajjhala, J. Gehrke, D. Kifer, M. Venkitasubramaniam, l-diversity: Privacy beyond k-anonymity. In Proc. 22nd Intnl. Conf. Data Engg. (ICDE), page:24, 2006

[17] T.M. Truta, V. Bindu, Privacy Protection: p-Sensitive k-Anonymity Property. In Proceedings of the Workshop on Privacy Data Management, In Conjunction with 22th IEEE International Conference of Data Engineering (ICDE), Atlanta, Georgia, 2006

[18] A. Campan, T.M. Truta, Extended p-Sensitive k-Anonymity, Studia Univ. BABE-BOLYAI, INFORMATICA, Volume LI, Number 2, 2006

[19] N. Li, T. Li, S. Venkatasubramanian, t-Closeness: Privacy Beyond k-Anonymity and l-Diversity, CERIAS Tech. Report 2007-78, Purdue University

[20] M. Gruteser, D. Grunwald. Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking, In First International Conference On Mobile Systems, Applications, Services (MobiSYS), USENIX, 2003

[21] M. Gruteser, G. Schelle, A. Jain, R. Han, D. Grundwald. Privacy-Aware Location Sensor Networks, In Proceedings 9th USENIX Workshop on Hot Topics in Operating Systems (HotOS), 2003.

[22] C. Ozturk, Y. Zhang, W. Trappe. Source-Location Privacy in Energy-Constrained Sensor Network Routing, In Proceedings of the 2004 ACM Workshop on Security of Ad Hoc and Sensor Networks, pp.88-93, 2004

[23] Y. Jian, S. Chen, Z. Zhang, L. Zhang, Protecting Receiver-Location Privacy in Wireless Sensor Networks, In Proceedings of IEEE INFOCOM 2007

[24] A. Wadaa, S. Olariu, L. Wilson, M. Eltoweissy, K. Jones. On Providing Anonymity in Wireless Sensor Networks , Proceedings of the Tenth International Conference on Parallel and Distributed Systems (ICPADS'04) 1521, 2004

[25] B. Gedik, L. Liu, Protecting Location Privacy with Personalized k-Anonymity: Architecture and Algorithms, In IEEE Transactions on Mobile Computing, Volume. 7, No. 1, January 2008

[26] P. Andritsos, V. Tzerpos, Software clustering based on information loss minimization. *In Proceedings of 10th Working Conference on Reverse Engineering(WCRE'03),* page:334, 2003.

[27] D. W. Carman, P.S. Kruus, B. J. Matt. Constraints and approaches for distributed sensor network security. NAI Laboratories, Tech. Rep. 00-010, 2000.