I. J. Computer Network and Information Security, 2012, 6, 10-18
Published Online June 2012 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijcnis.2012.06.02

Privacy Notions and Review of Privacy Preservation Mechanisms in Wireless Sensor Networks

Manjusha Pandey , Shekhar Verma
Indian Institute of Information Technology, Allahabad, India.
rs58@iiita.ac.in, sverma@iiita.ac.in

***Abstract*—** The current research work in wireless sensor networks has mostly focused on the resolving issues related to power consumption and computational resource constraints in the wireless sensor networks .To achieve the same various specific routing schemes ,MAC and cross layered protocols and techniques have been proposed and designed .But with recent advances the privacy issues related to the data collected and transmitted by the wireless sensor networks had taken the center stage .Privacy preservation in wireless sensor networks has become more challenging because of the wireless nature of communication in WSN as well as its self organizing architecture. The present paper provides a comparative review of various privacy preserving mechanisms proposed and implemented in wireless sensor networks with respect to the privacy notions of k-anonymity and L-diversity. Along with the discussion and analysis the present work is an effort for the pavement of a way towards the future research in the field of privacy preservation in WSN.

***Index Terms*—** wireless sensor networks, Privacy in WSN, K-anonymity, l-diversity

I. Introduction

With the recent year advancements in the rapid information exchange means and mechanisms, the security and privacy requisites have been added on with the ever improving technologies in the data and information storage, retrieval as well as exchange. New technologies both in hardware and software are radically advancing our freedoms, but they are also enabling unparalleled invasions of privacy. Privacy preservation in wireless sensor networks has been focused by the research community quiet a long tough the fool proof privacy preservation have yet not been achieved. The wireless sensor networks have inherent properties like resource limitation, erratic sized network exotic topology prior and post deployment and high risk of physical attacks due to unattended nature of the network, these constraints make WSN more vulnerable to privacy attacks.

Wireless sensor networks generally deal with the sensing and communication of very vital micro data that could be of great importance for security, research and various other purposes. Hence the privacy preservation of such significant data values is one of the primary concerns the needs to be addressed. Of the various privacy notions used for micro data privacy preservation in traditional networks some of the are equally viable for the wireless sensor networks. The rest of the paper is organized as first we introduce the need for privacy preservation in related fields to WSN then we introduce the privacy notions of k-anonymity and L-diversity for privacy preservation techniques in WSN. After that the experimental parameters and privacy preserving mechanisms considered for review are discussed respectively. Last but not the least a comprehensive evaluation and comparison of proposed and implemented privacy preservation mechanisms has been done, followed by selected references.

II. Need for privacy preservation in wireless sensor networks

The privacy preservation techniques of wired networks are not useful in sensor networks, firstly due to the fact that network being different the set of problems are different and secondly because many of the methods pose overheads which are too burdensome for the sensor networks. The shared wireless medium enhances the chances for the adversary to locate the origin of the radio transmission and thus facilitating the hop by hop trace back to the origin of the multi hop communication. Launch of physical attacks and node compromises by the adversary thus posing a menace to the whole wireless sensor networks is quiet evident due to the miniature size of the sensor nodes and very nature of the wireless communication environment. As we know a wireless sensor network is severely constrained by various resources as computation, storage, and wireless communication bandwidth and battery power. The adversary could monitor such activities of the sensor as the communication patterns to figure out the energy depletion or resource usage in order to spot the most vulnerable spots in the network and use them to attack the network as a whole. The mobile components of the network bring more challenges to the privacy preservation. Mobility makes communication more unstable and does not guarantee full coverage so that privacy is much easier to be threatened. A traffic pattern that is the distribution of traffic over the entire network is mainly determined by the topology of the network. As all the sensed data in the network has to be transmitted to the base station or sink there could be a fixed traffic pattern that adds to privacy threats in the network.

Though there is no well established notation for privacy preservation in wireless sensor networks some of the privacy preservation notations used in sensor networks in particular and adhoc networks in general may be dependent on the following privacy notations.

1) k –anonymity
2) ℓ-Diversity

III. Basic Notation

The basic notations considered for the present research work are Let N = { N_1, N_2. . . , N_m} be a table with attributes N1, . . . ,Nm. We assume that N to be subset of some larger population .In the considered superset each tuple represents an individual from the population. For example, if N is an example set of attributes for Battle field monitoring system having the node id of individual nodes as sensitive attribute for the present set of attributes. Types of activities in the battle field monitoring system being No Activity (NA) where no specific activity is sensed by the sensor nodes in the battle ground. Malicious Activity (MA) is the slightly different activity than the normal conditions. Alarming Activity (AA) is the condition when the network updates for some serious activity being updated by the networks nodes for quick action by the troupes. Let N denote the set of all attributes { N_1, N_2. . . , N_m} and t[Ni] denote the value of attribute Ni for tuple t . If C = {C1,C2, . . . ,Cp} ⊆N then we use the notation t[C] to denote the tuple (t[C1], . . . , t[Cp]).The tuple t[C] is the projection of t onto the attributes in C. In privacy-preserving data communication, there exist several important subsets of A. A *sensitive attribute* is an attribute whose value for any particular individual must be kept secret from people who have no direct access to the original data. Let S denote the set of all sensitive attributes. An example of a sensitive attribute is *Node Id* from Figure 1. The association between individual's sensor nodes should be kept secret; thus we should not disclose which particular node is the sender or receiver node, but it is permissible to disclose the information that there exist some sender and receiver node in the network. We assume that the base station knows which attributes are sensitive. All attributes that are not sensitive are called *non-sensitive* attributes. Let NS denote the set of non-sensitive attributes.

Generally the sensed data may have the three kinds of attributes namely: 1) attributes that clearly give the identity information of the individual nodes called as explicit identifiers, 2) attributes whose values when taken together may reveal the identity of the individual node,3) and attributes that are considered sensitive. The emphasis has to be on the preservation of accessibility, inference or leakage of sensitive information of the nodes in the network. The disclosure may be of two types: *identity discloser* or *attribute discloser*. Identity discloser occurs when the identification of individual node in the network is revealed. Attribute discloser occurs when the some existing of new information could be inferred in the network. Identity discloser generally led to attribute discloser. If the identity is disclosed for node in the network it becomes easier for the adversary to infer the sensitive attributes thus posing a great threat to the overall security and privacy of the network.

	Data Sensed	Parent Node ID	Node ID
1	NA(No Activity)	PN3	N11
2	AA	PN2	N26
3	MA(Malicious Activity)	PN1	N21
4	NA	PN3	N45
5	NA	PN3	N18
6	AA(Alarming Activity)	PN2	N51
7	MA	PN1	N28
8	AA	PN2	N68
9	MA	PN1	N32

Figure 1. Example Set of attributes for Battle Field Monitoring System.

The figure1 above presents an example set of attributes for Battle field monitoring system having the node id of individual nodes as sensitive attribute for the present set of attributes. Types of activities in the battle field monitoring system being No Activity (NA) where no specific activity is sensed by the sensor nodes in the battle ground. Malicious Activity (MA) is the slightly different activity than the normal conditions. Alarming Activity (AA) is the condition when the network updates for some serious activity being updated by the networks nodes for quick action by the troupes. We can now formally define the notion of a quasi-identifier.

Definition: (Quasi-identifier) A set of non-sensitive attributes {Q1, . . . ,Qw} of a table is called a *quasi-identifier* if these attributes can be linked with external data to uniquely identify at least one individual in the general population. For example a quasi-identifier is a primary key of any given table. We denote the set of all quasi-identifiers by QI .We can now formally define k-anonymity.

Definition: (k-Anonymity) A table T satisfies k-anonymity if for every tuple t ∈ N there exist k − 1 other tuples ti1 , ti2 , . . . , tik−1 ∈ N such that t[C] = ti1 [C] =ti2 [C] = • · ·= tik−1 [C] for all C ∈ QI.

Privacy preservation provided by k-anonymity is simple and could be understood without any efforts. The privacy preserving notion of k-anonymity presents the concept that if given set of attributes satisfy k-anonymity for some value k, then anyone who knows only the value of Quasi identifier for any individual node would not be able to identify the values corresponding to that individual node with a confidence greater than 1/ k. Thus k-anonymity preserves identity discloser but is not able to provide the preservation against attribute discloser. Hence two types of attacks have been identified a) Homogeneity attack and b) Background knowledge attack.

	Data Sensed	Parent Node ID	Node ID
1	N*	*3	N11
4	N*	*3	N45
5	N*	*3	N18
2	A*	*2	N26
6	A*	*2	N51
8	A*	*2	N68
3	M*	*1	N21
7	M*	*1	N28
9	M*	*1	N32

Figure 2. T*: 3- Anonymous version of example Set of attributes for Battle Field Monitoring System.

The Figure2 presents an anonymized version T* satisfying 3 -anonymity. The *Node Id attribute* is sensitive. Suppose the attacker knows that sensor node having node id 11 has PN3 as its parent node and also that node 45 has PN3 as its parent node the attacker may conclude that the Node 45 must belong to the first group of classification, this is called homogeneity attack .Furthermore if the attacker has the idea that the node having node id 45 has higher probability of sending NA signal. This background knowledge helps the attacker to know that the node has PN3 as parent node.

To effectively limit disclosure, we need to measure the disclosure risk of an anonymized Equivalence Class. Samarati and Sweeney introduced the concept of *k-anonymity* as a property in which each record is indistinguishable with at least *k-1* other records with respect to the quasi-identifier. In other words, *k*-anonymity requires that each equivalence class contains at least *k* records. While *k*-anonymity protects against identity disclosure, it is insufficient to prevent attribute disclosure.

Adversary’s Background Knowledge. One of the major limitations of K-anonymity is the background knowledge attack, that is due to the adversary’s additional knowledge about the table .Some of the type of background knowledge an adversary may have are: First, the adversary has access to the published table T ⋆ and she knows that T ⋆ is a generalization of some base table T . The adversary also knows the domain of each attribute of T. Second, the adversary may know that some individuals are in the table. This knowledge is often easy to acquire. In addition, the adversary could have knowledge about the sensitive attributes of specific individuals in the population and/or the table. Such knowledge is called “instance-level background knowledge,” since it is associated with specific instances in the table. Third, the adversary could have partial knowledge about the distribution of sensitive and non-sensitive attributes in the population called as “demographic background knowledge”. Now armed with the right notation, let us start looking into principles and definitions of privacy that leak little information.

Bayes-Optimal Privacy

The ideal notion of privacy is called *Bayes-Optimal Privacy*. Bayes-Optimal Privacy involves modeling background knowledge as a probability distribution over the attributes and uses Bayesian inference techniques to reason about privacy. We first introduce the tools for reasoning about privacy ,after that we use them to discuss theoretical principles of privacy, which helps us to point out the limitations needed to be overcome to arrive at a practical definition of privacy.

Changes in Belief

In order to simplify the discussion, we combine all the non-sensitive attributes into a single, multi-dimensional quasi-identifier attribute Q whose values are generalized to create the anonymized table T ⋆ from the base table N.

Following two simplifying assumptions have been made

First: N is a simple random sample from some larger population (a sample of size n drawn without replacement is called a *simple random sample* if every sample of size n is equally likely).

Second: There is a single sensitive attribute.

We would like to emphasize that the above two assumptions will be dropped in the practical definition of privacy. We consider that in our attack model, the adversary has partial knowledge of the distribution of the sensitive and non-sensitive attributes. Considering a worst case scenario where the adversary knows the complete joint distribution f of Q and S (i.e. she knows their frequency in the population). Then the adversary knows that a Nodeid corresponds to a record t ∈ T that has been generalized to a record t∗ in T ⋆, and she also knows the value non-sensitive attributes (i.e., she knows that t[Q] = q). The adversary’s goal is to use her background knowledge to discover the nodes’s sensitive information— the value of t[S]. We gauge the adversary’s success using two quantities: Adversary’s *prior belief*, and her *posterior belief*. Adversary’s *prior belief*, P(q,s), that node’s sensitive attribute is s given that its non-sensitive attribute is q, is just her background knowledge:

$$P_{(q,s)} = P_f\ (t[S] = s \mid t[Q] = q) \qquad (1)$$

As the adversary observes the table T ⋆ her belief about node’s sensitive attribute changes. This new belief,

$P'_{(q\ ,s,T*)}$, is her *posterior belief* :

$$P'_{(q\ ,s,T*)} = P_f\ (t[S] = s \mid t[Q] = q \wedge \exists t^* \in T\ ^*,\ t^* \longrightarrow t^*) \qquad (2)$$

Given f and T *, we can derive a formula for $P'_{(q\ ,s,T*)}$ which will help us formulate our new privacy definition in . The main idea behind the derivation is to find a set of equally likely disjoint random worlds such that the conditional probability P(A|B) is the number of worlds satisfying the condition A ∧ B divided by the number of worlds satisfying the condition B. We avoid double counting because the random worlds are disjoint. In our case, a random world is any permutation of a simple random sample of size n that is drawn from the population and which is *compatible* with the anonymized table T*.

Theorem 1 Let q be a value of the non-sensitive attribute Q in the base table N ; let q* be the generalized value of q in the anonymized table T *; let s be a possible value of the sensitive attribute; let n(q⋆,s′) be the number of tuples t* ∈ T* where t*[Q] = q* and t*[S] = s′; and let f(s′ | q*) be the conditional probability of the sensitive attribute conditioned on the fact that the non-sensitive attribute Q can be generalized to q*. Then the following relationship holds:

$$P'(_{(q\ \ ,s,T*)} = \frac{n(q*s)\frac{f\ (s \mid q)}{f(s \mid q*)}}{\sum s'\in Sn(q*,s')\frac{f(s' \mid q)}{f(s' \mid q*)}} \qquad (3)$$

After calculating adversary's belief about node's private data after she has seen T *, let us now examine some principles for building definitions of privacy.

IV. Privacy Principles

Based Given the adversary's background knowledge, an anonymized table T* discloses the information in two important ways : *positive disclosure* and *negative disclosure*.

Definition (Positive disclosure) The table T ⋆ that was derived from N results in a positive disclosure if the adversary can correctly identify the value of a sensitive attribute with high probability; i.e., given a δ > 0, there is a positive disclosure if $P'(_{(q\ \ ,s,T*)}> 1 - P'$ and there exists t ∈ N such that t[Q] = q and t[S] = s.

Definition (Negative disclosure) The table T* that was derived from N results in a negative disclosure if the adversary can correctly eliminate some possible values of the sensitive attribute (with high probability); i.e., given an Q > 0, there is a negative disclosure if $P'(_{(q\ \ ,s,T*)} <$ Q and there exists a t ∈ N such that t[Q] = q but t[S] = s.

Note that not all positive disclosures are disastrous. If the prior belief was that P(q,s) > 1−P, the adversary would not have learned anything new. Similarly, negative disclosures are not always bad. Thus, the ideal definition of privacy can be based on the following principle:

Principle1 (Uninformative Principle) The anonymized table must provide the adversary with little additional information beyond the background knowledge. In other words, there should not be a large difference between the prior and posterior beliefs.

- To instantiated uninformative principle we may consider for example if the (P, P_b) are privacy breach definition [14]. Then under this definition, privacy is breached either when $P_{(q,s)} < P \wedge P'_{(q\ \ ,s,T*)} > P_b$ or when $P_{(q,s)} > 1 - P \wedge P'_{(q\ \ ,s,T*)} < 1 - P_b$. The alternative privacy definition on the basis of uninformative principle would bound the maximum difference between $P_{(q,s)}$ and $P'_{(q\ \ ,s,T*)}$ using any of the functions commonly used to measure the difference between probability distributions.
- Any privacy definition based on the uninformative principle, and instantiated either by a (P, P_b) -privacy breach definition or by bounding the difference between $P_{(q,s)}$ and $P'_{(q\ \ ,s,T*)}$ is a Bayes-optimal privacy definition.
- The specific choice of definition depends on the application. Note that any Bayes-optimal privacy definition captures diversity as well as background knowledge.

Limitations of the Bayes Optimal Privacy: In particular, Bayes-optimal privacy has several drawbacks that make it hard to use in practice.

Insufficient Knowledge**.** The data collector sink node is unlikely to know the full distribution f of sensitive and non-sensitive attributes over the general population from which N is a sample.

The Adversary's Knowledge is Unknown. It is also unlikely that the adversary has knowledge of the complete joint distribution between the non-sensitive and sensitive attributes. However, the data collector sink node does not know how much the adversary knows.

Instance-Level Knowledge. The theoretical definition does not protect against knowledge that cannot be modeled probabilistically.

Multiple Adversaries. There will likely be multiple adversaries with different levels of knowledge, each of which is consistent with the full joint distribution. Thus, although additional knowledge can yield better inferences on average, there are specific instances where it does not. Thus the data collector sink node must take into account all possible levels of background knowledge. To address this limitation of *k*-anonymity, Machanavajjhala et al. recently introduced a new notion of privacy, called l-diversity, which requires that the distribution of a sensitive attribute in each equivalence class has at least l "well represented" values.

V. ℓ-Diversity: Evolution to Practical Privacy

This section discusses how to overcome the drawbacks and limitations outlined above for privacy preservation in wireless sensor networks .We discuss the ℓ-diversity principle first then show how to instantiate it with specific definitions of privacy, outline how to handle multiple sensitive attributes and then discuss how ℓ-diversity addresses the issues raised related to privacy preservation in wireless sensor networks.

The ℓ-Diversity Principle

The Theorem 1 allows us to calculate the observed belief of the adversary. Let us define a q*-block to be the set of tuples in T* in which the non-sensitive attribute values generalize to q*. If we consider the case of positive disclosures with very high probability. As per Theorem 1, this can happen only when:

$$\exists s\ ,\forall\ s\ ',\ \ n(q*s)\frac{f\ (s \mid q)}{f(s \mid q*)} \ll\ Sn(q*,s')\frac{f(s' \mid q)}{f(s' \mid q*)} \qquad (4)$$

The Equation (2) may hold true only due to a combination of two factors discusses as follows : (i) a lack of diversity in the sensitive attributes in the q*-block, and/or (ii) strong background knowledge. .

Lack of Diversity. Lack of diversity in the sensitive attribute manifests itself in the condition if

$$\forall s' \neq s \quad n(q*, s') \ll \quad n(q*s) \tag{5}$$

In such a case, almost all tuples have the same value s for the sensitive attribute S, and thus $P'_{(q\ ,s,T*)} \approx 1$. To be noted is that this condition can be easily checked since it only involves counting the values of S in the anonymized table T*.We can ensure diversity by requiring that all the possible values s′ ∊ domain (s) occur in the q*-block with roughly equal proportions. This, however, is likely to cause significant loss of information: if domain(s) is large then the q*-blocks will necessarily be large and so the data will be partitioned into a small number of q*-blocks. Another way to ensure diversity and to guard against Equation 3 is to require that a q*-block has at least $\ell \geqslant 2$ different sensitive values such that the ℓ most frequent values (in the q*-block) have roughly the same frequency. We say that such a q*- block is well-represented by ℓ sensitive values.

Strong Background Knowledge. One more factor that could lead to a positive disclosure is strong background knowledge. Even though a q*-block may have ℓ "well-represented" sensitive values, the adversary may still be able to use her background knowledge to eliminate sensitive values when the following holds true:

$$\exists s \ , \qquad \frac{f(s' \mid q)}{f(s' \mid q*)} \approx 0 \tag{6}$$

This equation states that a node with quasi-identifier t [Q] = q is much less likely to have sensitive value s′ than any other individual node in the q*-block. For example, Adversary may know that node having *NODEID*: 1 lies in the center of the battle field, may never be near the sink node as all the sink nodes are situated at the corners of the battle field. It is not possible for the data collector to guard against attacks employing arbitrary amounts of background knowledge. However, the data collector can still guard against many attacks even without having access to adversary's background knowledge. In the present model, adversary might know the distribution f (q, s) over the sensitive and non-sensitive attributes, in addition to the conditional distribution f(s|q). The most damaging type of such information has the form f(s|q) ≈ 0, e.g., "sink nodes never die", or the form of e.g., "sensor nodes may never behave as data collectors" Note that *a priori* information of the form f(s|q) = 1 is not as harmful since this positive disclosure is independent of the table T*. Adversary can also eliminate sensitive values with instance-level knowledge such as "some node is at the center of the battle field". In spite of such background knowledge, if there are ℓ "well represented" sensitive values in a q*-block, the adversary needs $\ell - 1$ damaging pieces of background knowledge of Non-Sensitive sensitive attributes. to eliminate 1 possible sensitive values and infer a positive disclosure! Thus, by setting the parameter ℓ 1, the data collector can determine how much protection is provided against background knowledge even if this background knowledge is unknown to the publisher. Putting these two arguments together, we arrive at the following principle.

	Data Sensed	Parent Node ID	Node ID	Route Followed
1	NA	PN3	N11	R14
2	AA	PN2	N26	R11
3	MA	PN1	N21	R33
4	NA	PN3	N45	R28
5	NA	PN3	N18	R36
6	AA	PN2	N51	R05
7	MA	PN1	N28	R04
8	AA	PN2	N68	R16
9	MA	PN1	N32	R13

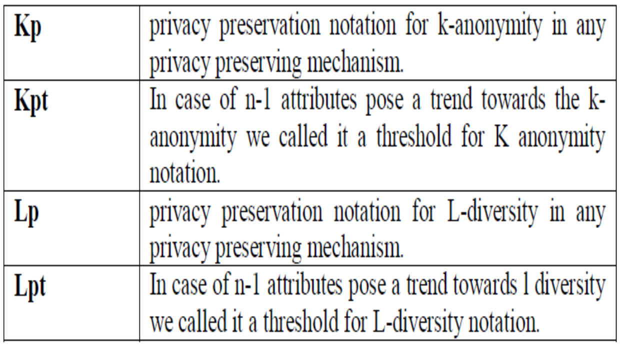
Figure 3. Example table of Route Followed / Node ID set of attributes for battle field monitoring System.

	Data Sensed	Parent Node ID	Node ID	Route Followed
1	N*	*3	N11	R14
4	N*	*3	N45	R28
5	N*	*3	N18	R36
2	A*	*2	N26	R11
6	A*	*2	N51	R05
8	A*	*2	N68	R16
3	M*	*1	N21	R33
7	M*	*1	N28	R04
9	M*	*1	N32	R13

Figure 4. A 3-diverse version of Table 3.

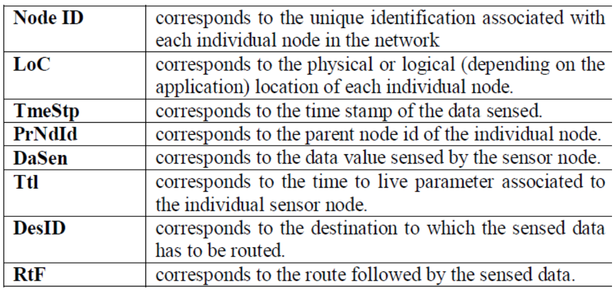
VI. Discussion and Analysis

The simulations For the review of various privacy preservation mechanisms proposed and implemented in wireless sensor network on the basis of the above described notation has been done. Forthe parameterized analysis of various mechanisms we have defined the privacy preservation notation parameters on the basis of which the mechanisms have been evaluated both quantitatively as well as qualitatively these parameters include:



Kp	privacy preservation notation for k-anonymity in any privacy preserving mechanism.
Kpt	In case of n-1 attributes pose a trend towards the k-anonymity we called it a threshold for K anonymity notation.
Lp	privacy preservation notation for L-diversity in any privacy preserving mechanism.
Lpt	In case of n-1 attributes pose a trend towards l diversity we called it a threshold for L-diversity notation.

Figure 5. Privacy Preservation Notions Parameters



Node ID	corresponds to the unique identification associated with each individual node in the network
LoC	corresponds to the physical or logical (depending on the application) location of each individual node.
TmeStp	corresponds to the time stamp of the data sensed.
PrNdId	corresponds to the parent node id of the individual node.
DaSen	corresponds to the data value sensed by the sensor node.
Ttl	corresponds to the time to live parameter associated to the individual sensor node.
DesID	corresponds to the destination to which the sensed data has to be routed.
RtF	corresponds to the route followed by the sensed data.

Figure 6.The attribute set related to the individual nodes of the sensor network.

The privacy preservation mechanisms under consideration were compared and analyzed for the above eight network packet parameters on the basis of k-anonymity and L-diversity notions of privacy.

The first set of proposed privacy preservation mechanism included mechanisms based on Data Aggregation based schemes as proposed by Claude Castelluccia [12] that uses additively homomorphism stream cipher which allows efficient aggregation of encrypted data. New cipher that uses modular additions (with very small moduli) is well suited for CPU constrained devices like sensor nodes. The present scheme has as advantage of efficiently computed statistical values, and the disadvantage is that the scheme is slightly less bandwidth efficient than the generally preferred hop-by-hop aggregation scheme. Aldar C-F. Chan et al [13] discussed Privacy of Concealed Data Aggregation. Standard security notions for public key encryption schemes, including semantic security and indistinguishability against chosen ciphertext attacks, have been refined to cover the multi-sender nature and aggregation functionality of CDA in the security model. A generic CDA construction based on public key homomorphic encryption has been proposed, along with a proof of its security in the proposed model. The security of two existing schemes has also been analyzed in the proposed model. Gelareh Taban et. Al [14] aimed for integrity-assured data aggregation with efficiency and privacy as a joint objective. The mechanism uses Integrity Verification Aggregation Functions homomorphism and Message authentication codes (MAC) construct authenticated encryption scheme. This provides a Limited cost of aggregation functions privacy-preserving computations included. But the disadvantage is that it is a complex algorithm and hence energy consuming for implementation. Thus making it less suitable of energy constrained networks like sensor networks.

Second set of privacy preserving mechanisms being considered are the encryption based mechanisms. These set of mechanisms concentrate on the content privacy through encryption and decryption techniques. The privacy preservation is achieved through encryption and decryption of data based set of protocols. Much work has been going on in the field of security for WSNs. Cryptographic techniques such as Skipjack, RC5, and Elliptic Curve Cryptography (ECC) and Identity Based Encryption (IBE) are found to be very promising for WSNs. Steffen Peter et.[15] Al proposed end-to-end encryption solutions for converge cast traffic hop-by-hop based encryption approaches. Here aggregator nodes can perform in-network processing on encrypted data. A privacy homomorphism (PH) is an encryption concealed data aggregation is implemented in the scheme. KealanMcCusker et. Al [16] proposed Identity Based Encryption (IBE) implemented by the usage of Tate pairing, in 90nm CMOS and obtained area. Hardware implementation of IBE would meet the strict energy constraint of a wireless sensor network node. But the Tate pairing is the most computationally expensive process in IBE.Roberto Di Pietro[17] implemented Energy efficient node-to-node authentication and communication Confidentiality having a smart attacker model novel pseudo-random key pre-deployment strategy ESP. The scheme provided energy efficient key discovery requiring no communications highly resistant to the smart attacker as well as node to node authentication. Limitation to the scheme being Message encryption might reduce the effectiveness of in network processing if the keying mechanism is not carefully devised.

Next set of privacy preservation protocols was those based on key management in the sensor networks. Niu Dou et. Al [18] proposed the key management based privacy preservation in wireless sensor networks. Each node update the original key at given intervals, and discards the original key and cluster head nodes used out of ordinary nodes to regenerate the keys. Even if the enemy captures the sensor nodes, they could only get some old keys, so it can’t pose a threat to the network. But the energy consumption increases because the ordinary node needs to send a confirmation to the cluster head node.Yong Ho Kim et. Al[19] talked about a secure and efficient key management based mechanism having a key distribution scheme as an advancement over pair-wise key establishment in sensor networks. The scheme improves the resilience against node capture and reduces communication cost supports efficient node addition, with the limitation of security and efficiency is trade-off. Sajid Hussain[20] proposed a key distribution scheme based on random key pre-distribution for heterogeneous sensor network (HSN).With the central concept of instead of generating a large pool of random keys, a key pool is represented by a small number of generation keys, one-way hash function generates a key chain that collectively make a key pool, Thus reducing the storage requirements while maintaining the security strength. With the disadvantage of being limited to heterogeneous sensor networks based applications only.

Anti-traffic analysis based privacy preservation mechanisms proposed and implemented in wireless sensor networks. In order to make it harder for an attacker, Deng et al. proposed a set of advanced techniques to counter the traffic analysis attacks [21, 22]. The rate monitoring attack can be partially prevented by the multiple parents routing scheme since traffic spreads along multiple paths. In this scheme, each node has multiple parent nodes, which route messages to the base station. In order to forward a message, a node randomly selects one of its parent nodes. This scheme can be extended by the controlled random walk. A node forwards a message to one of its parent nodes with probability p. With probability 1 - p the node forwards the message randomly to one of its neighbors including the parent nodes. This technique introduces delivery time delays, which are proportional to extra hops taken by the messages. This technique is still vulnerable to the time correlation attack. Therefore, the authors propose a new technique called the multi-parent routing scheme with fractal propagation. When a node hears that a neighbor forwards a message to the base station, the node generates a fake message with probability pf and forwards it to one of its neighbors. The main problem with this technique is that it generates a large amount of traffic near the base station, because nodes near the base station usually forward more messages. This can be solved by the Differential Fractal Propagation technique (DFP). When a node forwards messages more frequently, it sets a lower probability for creating new fake messages. In order to make the traffic analysis more difficult, the authors propose to generate also artificial areas (called hot-spots) of high a communication activity. We have encountered a problem with the DFP [23]. Deng et al. consider an internal adversary in their work; however the probability of creating fake messages that can be discovered by the internal adversary, and leaks significant information on the distance from the base station. By capturing few nodes, the adversary can easily estimate the location of the base station. In order to defend against the time correlation attack, Hong et al. [11] propose to add random delays to message retransmission at each forwarding node. Their approach does not introduce any dummy traffic; nonetheless it is not suitable for the networks with minimal network traffic.

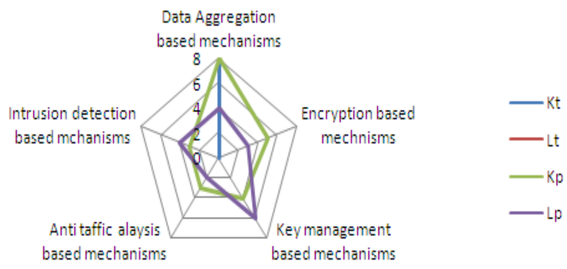


Figure 7. Comparative efficiency of various privacypreserving mechanisms with fixed number of network parameters

VII. Conclusion

Efficiency We compare the efficiency and data quality of four privacy measures through the privacy notions of : (1) k-anonymity denoted by Kp; (2) k-anonymity threshold denoted by Kt; (3)L-diversity denoted by Lp;(4) L-diversity threshold denoted by Lt. Results of efficiency experiments are shown in Figure 1.Again we use the Node Id attribute as the sensitive attribute. Figure 7 shows the comparative efficiency of various privacy preserving mechanisms with fixed number of network parameters i.e. 8 and varied quasi-identifier size s, where 2 ≤ s ≤ 7. A quasi-identifier of size s consists of the first s attributes listed. The results presented a k-anonymity threshold for the data aggregation techniques under consideration while the L-diversity threshold is not obtained by any of the privacy preservation mechanisms under consideration. Data aggregation techniques present L-diversity privacy notion by 4 attributes. Though the encryption based mechanisms pose L-diversity notion with 3 of the considered attributes and k-anonymity notion by 5 attributes. Key management based privacy preservation mechanisms pose L-diversity notion by 6 parameters and k-anonymity notion by 4 parameters. Anti traffic analysis attack based mechanisms pose L-diversity notion with only 2 parameters lowest of all the notions and k-anonymity with 3 parameters. The Intrusion detection based privacy and security measures provide L-diversity notion with 4 parameters and k-anonymity notion with 3 parameters.

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci.A survey on Sensor networks. *IEEE Communications Magazine*, 40(8):102–114, August 2002.

[2] Na Li, Nan Zhang, Sajal K. Das, Bhavani Thuraisingham , “Privacy preservation in wireless sensor networks: A state-of-the-art survey” in Ad Hoc Networks 7 (2009), p. 1501–1514, 2009.

[3] J. Deng, R. Han, and S. Mishra. Security, privacy, and fault tolerance in wireless sensor networks. Artech House, August 2005.

[4] C. Karlof and D. Wagner. Secure routing in wireless sensor networks: Attacks and countermeasures. Elsevier's AdHoc Networks Journal, Special Issue on Sensor Network Applications and Protocols, 1(2–3):293–315, September 2003.

[5] W. Xu, W. Trappe, Y. Zhang, and T. Wood. The feasibility of launching and detecting jamming attacks in wireless networks. In ACM MobiHoc, 2005.

[6] W. Xu, T. Wood, W. Trappe, and Y. Zhang. Channel Surfing and Spatial Retreats: Defenses against Wireless Denial of Service. In ACM WiSe, pages 80–89, 2004.

[7] G. Gaubatz, J.P. Kaps, and B. Sunar. Public key cryptography in sensor networks - revisited. In 1st European Workshop on Security in Ad-Hoc and Sensor Networks *(ESAS 2004)*, 2004.

[8] J. Hwang and Y. Kim. Revisiting random key pre-distribution schemes for wireless sensor networks. In Proceedings of the 2nd ACM workshop on Security of Ad hoc and Sensor Networks (SASN '04), pages 43–52, New York, NY, USA, 2004. ACM Press.

[9] R. Zhang, Y Zhang, and K. Ren,"DP 000b2;ac: Distributed privacy preserving access control in sensor networks", in INFOCOM 2009, IEEE, pp. 1251-1259, April 2009

[10] A. Perrig et. al. SPINS: Security Protocols for Sensor Networks. Wireless Networks, 8(5):521–534, 2002.

[11] Aysal, T.C.; Barner, K.E.; Sensor Data Cryptography in Wireless Sensor Networks. In IEEE Transactions on Information Forensics and Security, Volume: 3 Issue:2 On page(s): 273 – 289,2008

[12] Claude Castelluccia Einar Mykletun, Gene Tsudik .Efficient Aggregation of encrypted data inWireless Sensor Networks. Proceedings of the Second Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services (MobiQuitous'05), 2005.

[13] Kashif Kifayat ,Madjid Merabti, Qi Shi, David Llewellyn-Jones. Applying Secure Data Aggregation techniques for a Structure and Density Independent Group Based Key Management Protocol. Proceeding of IAS '07 Proceedings of the Third International Symposium on Information Assurance and Security *(IAS 2007)*, Manchester, UK, 29-31 August 2007.

[14] G. Taban and V.D. Gligor.Privacy-Preserving Integrity-Assured Data Aggregation in Sensor Networks. In Proc. CSE (3), 2009, pp.168-175.

[15] Steffen Steffen Peter, Krzysztof Piotrowski, Peter Langendörfer.In-network-aggregation as case study for a support tool reducing the complexity of designing secure wireless sensor networks. In proceedings of 33rd IEEE Conference on Local Computer Networks, 2008. LCN 2008. Pp 778 - 785.

[16] McCusker, K. O'Connor, N.E. CLARITY: Centre for Sensor Web Technol., Dublin.Low-Energy Symmetric Key Distribution in Wireless Sensor Networks.IEEE Transactions on Dependable and Secure Computing, Volume: 8 Issue: 3 ,pp. 363 - 376.

[17] Roberto Di Pietro, Luigi V. Mancini, Alessandro Mei.Energy efficient node-to-node authentication and communication confidentiality in wireless sensor networks. Wireless Networks 12 Volume: 12, Issue: 6 ,pp. 709-721 (2006)

[18] Niu Dou; Yao Yan-yan; Wang Ting-ting.Improved group-key-management scheme for WSN.Computing, Communication, Control, and Management, 2009. CCCM 2009. ISECS International Colloquium on , 8-9 Aug. 2009 ,pp 330 – 333.

[19] Yong Ho Kim Hwaseong Lee Dong Hoon Lee .A Key Distribution Scheme for Wireless Sensor Networks. Pervasive Computing and Communications, 2008. PerCom 2008. Sixth Annual IEEE International Conference on 17-21 March 2008 pp. 572 – 577.

[20] Sajid Hussain , Firdous Kausar , Ashraf Masood. An Efficient Key Distribution Scheme for Heterogeneous Sensor Networks. In Proceeding of IWCMC '07 Proceedings of the 2007 international conference on Wireless communications and mobile computing, pp. 388 – 392.

[21] Jing Deng, Richard Han, and Shivakant Mishra. Countermeasures against traffic analysis attacks in wireless sensor networks. In SECURECOMM '05: Proceedings of the First International Conference on Security and Privacy for Emerging Areas in Communications Networks, pages 113-126, Washington, DC, USA, 2005. IEEE Computer Society.

[22] Jing Deng, Richard Han, and Shivakant Mishra. Decorrelating wireless sensor network traffic to inhibit traffic analysis attacks. Pervasive and Mobile Computing, 2(2):159-186, April 2006.

[23] Jir Kur and Andryi Stetsko. Location privacy issues in wireless sensor networks. In The Future of Identity in the Information Society, pages 160-169. Springer Boston, 2009.

Authors

Ms. Manjusha Pandey She is pursuing Ph.D. from Indian Institute of Information Technology, Allahabad, India in Information and Technology, has done her M. Tech in Computer Science. Her research interest areas include Wireless Sensor Networks, Privacy in Wireless Communication, Privacy and security in Digital & Mobile Communication, Signal Processing and Vehicular Technology.

Dr. Shekhar Verma He received his Ph.D. degree from IT, Banaras Hindu University, Varanasi, India in Computer Science and Engg. He is Associate Professor in Information Technology at Indian Institute of Information Technology, Allahabad, India. His research interest areas are Computer Networks, Wireless Sensor Networks, Vehicular Technology, Cryptography, Information and Network Security.