Temporal Privacy Scheme for End-to-End Location Privacy in Wireless Sensor Networks

Abdel-shakour Abuzneid, Tarek Sobh, and Miad Faezipour

Department of Computer Science and Engineering, University of Bridgeport
Bridgeport, Connecticut
Email: {abuzneid, sobh, mfaezipo}@bridgeport.edu

Abstract—Wireless sensor network (WSN) is built of hosts called sensors which can sense a phenomenon such as motion, temperature, and humidity. Sensors represent what they sense in data format. Providing an efficient end-to-end privacy solution would be a challenging task due to the open nature of the WSN. The key schemes needed for end-to-end location privacy are anonymity, observability, capture likelihood and safety period. On top of that, having temporal privacy is crucial to attain. We extend this work to provide a solution against global adversaries. We present a network model that is protected against passive/active and local/multi-local/global attacks. This work provides a solution for temporal privacy to attain end-to-end anonymity and location privacy.

Keywords—WSN; temporal privacy; traffic rate privacy; source location privacy; sink privacy

I. INTRODUCTION AND PROBLEM STATEMENT:

In this work, we shall focus on providing temporal privacy which is very curtail to providing source and sink location privacy. There are generally two ways to locate a sensor using passive attacks: *traffic analysis* [1] and *packet tracing* [2, 3]. The traffic analysis can determine location by analyzing the traffic. Packet tracing can be used to find the source location since adversaries may use radio-frequency localization techniques to perform a hob-by-hop trace. The adversary can move rapidly during packet trace. It could be used to trace mobile nodes due to its prompt response compared to traffic analysis [2, 4-6]. The rest of this paper is organized as follows: In section 2, we give some background and we explain our system, network, threat and traffic models. In section 3, we will explain suggested solution with a detailed analysis. In section 4, we will give some conclusion.

II. BACKGROUND AND SYSTEM, NETWORK, THREAT AND TRAFFIC MODELS

In this work, we assume bi-directional links where two nodes are considered neighbors if they can overhear one another [7].The WSN considers one sink which aggregates sensed data (stimuli) from all the sensor nodes. The sink works as an interface for WSN to the backbone network [8]. Data packets generated by SNs are ultimately sent uplink to the sink and never destined to another sensor. However, it would go through a multi-hop path. To enhance sink privacy, the sink acts like any other sensor in the network when it communicates with other sensors to make sure the sink is indistinguishable. All sensors are time synchronized [8]. The WSN will need a protocol for *topology discovery* that allows the sink to view the network global topology without revealing the location of the sink [7]. The adversary has very strong capabilities such as having sufficient energy supply, computation capabilities, and unlimited storage memory. An adversary could run both *passive* and *active* attacks. We adopts the *Kirchhoff's Principle*, where the adversary knows the system's protocols and behavior but not the *keys* and the *IDs*. The solution will be able to handle both passive and active attacks where we presume only few compromised nodes could coexist at one time due to the protection of intrusion detection system (*IDS*), able to detect compromised sensor nodes. We account for worst case global adversary, which can monitor the traffic of the entire network and can determine the node responsible for the initial transmission. We also assume that the adversary is capable of observing transmissions over spread periods of time. It is, however, not able to break the encryption algorithms or the hash functions. We presume *abundant* traffic model where sensors detect and transmit lots of packets.

WSN could suffer from time correlation attacks [1, 2, 9, 10] by observing the time between correlative packets sent and received in a certain neighborhood. Accordingly, hiding timing of the source is crucial for the anonymity and location privacy. Facilitating routing schemes to protect against time correlation is proven to be efficient to certain extent where local adversary usually has limited mobility and partial view of the traffic. However, routing based schemes do not work for global adversary where the traffic of the whole network can be easily observed with a full spatial coverage. Additionally, the adversaries could collude together to promptly detect the timing of the event [11, 12]. In summary, temporal privacy could be achieved

by delay which is only useful against local adversary. It will not suffice handle a global adversary. To handle global adversary, some schemes suggest issuing some dummy or fake messages while transmitting real messages to divert the adversary from detecting the source. This could work for local adversary but it will not work for global adversary. The cumbersome issue is the arrival of the real event-driven messages are not deterministic. The distribution of events could change which would be a reason for the adversary to detect the event and thereafter the source initiating the event. An adversary with reasonable statistical analysis engine can easily detect the norm of messages distribution [11]. This could be solved by delaying the transmission of the real message according to a mechanism which maintains fixed message distribution in the whole network. This approach, however, might not be suitable for time sensitive networks. Some literature clearly differentiates between two things: *event* and *interval* of the transmission. If every interval has one transmission, then event and interval are identical. However, this might not be the situation with multiple transmissions during one interval. So, the anonymity degree depends on the adversary's ability to distinguish between real and fake transmissions. This means, given multiple transmissions by a sensor, the adversary should be unable to distinguish, with significant confidence, between transmissions carry real data and transmissions carry fake data. Alomair et al. [11] suggested that indistinguishability is not enough. They claim that indistinguishability is achieved when adversary monitoring the network over multiple time intervals, in which some intervals contain real event transmissions and others do not, is unable to ascertain, with high confidence, which of the intervals contain the real data and vice versa. If intervals are indistinguishable, the individual transmissions within the interval should also be indistinguishable. We would need a mechanism to quantify anonymity. Let's presume ψ donates one adversary strategy for breaching the anonymity of a system among other strategies. Let's presume P_r is the probability that the adversary succeeds using the strategy (ψ). The anonymity A as defined in [11] with the existence of a strategy ψ, is presented in expression below:

$$A_\psi := 1 - P_r, \text{ where } \quad 0 \le P_r \le 1 \tag{1}$$

If Σ represents all possible strategies for the adversary to infract the anonymity of the WSN, then we can calculate the accumulated anonymity as:

$$A := \min(A_\psi), \text{ where } \psi \in \Sigma \tag{2}$$

It is very essential to increase anonymity for every sensor in the network especially with the presence of multi-local or global adversaries. Presence of colluding adversaries could cause the anonymity to drop exponentially [11]. For example, a moving Panda from point "a", to "b", to "c", and then "d" where each location has a sensor to report the Panda's location, if the anonymity of each sensor is $A=0.8$, then the anonymity at node "b" is $A=.8^2=0.64$ and at point "d" is $A=.8^4=0.41$.

III. Solution for Temporal Attacks Using Delay and Fake Messages

Some WSNs could have a single local adversary with a limited view of the traffic. Sensor node sends an event-driven message when the event is detected at the location of the sensor. The adversary can trace back the message to the source sensor or forward to the sink. Having few other transmissions in the network within the range of the adversary confuses it and prevents the adversary from having deterministic path to follow. In this work, we presume the lifetime of the network Ω is divided into a number of discrete intervals I and each interval time is ω_I, where:

$$\Omega = I.\omega_I \tag{3}$$

The sensors will send either a real or a fake message during one interval. The message is sent at the end of each interval or it is adjusted to be sent during the interval to create some variable delays throughout the routing of the message to the sink. This would obscure the adversary more and would prevent it from learning useful information about the network based on temporal correlation. A sensor which has sensed data or received real data from another sensor, will send the real message M_r through a hop-by-hop path to the sink, and some selected nodes will send fake messages M_f during the same time to confuse the adversary. M_r and M_f are sent at the end of the interval I. The withhold time (τ_w) expressed as in below:

$$\tau_w = \omega_i - t_a \text{ where: } t_0 \le t_a \le t_s \le \omega_I \tag{4}$$

Where: t_0 is the beginning of the interval I_i, t_a is the arrival time, $t_0 \le t_a$. Ideally, the message will be sent immediately after it is sensed or received which makes $\tau_w = 0$. Theoretically, τ_w could be a value: $0 \le \tau_w \le \omega_I$ as exhibited in **Figure 1**.

Sensors with no real messages need to send fake messages deterministically during the interval I_i, in the best case for anonymity and the worst case or energy conservation. Fake messages are usually sent probabilistically according to some protocol. Having deterministic or highly probabilistic fake message transmissions will reduce the lifetime of the network in favor of privacy. Doing the reverse will jeopardize the privacy of the sensor nodes. Having variable message withholding time τ_w is useful for privacy and for reducing the average network delay. The delivery time τ_d for a message presuming that the message is always sent at the end of the interval I_i is:

$$\tau_d = \tau_w + \tau_{trans} + \tau_{proc} \tag{5}$$

978-1-4799-7678-2/15/$31.00 ©2015 IEEE

Where: τ_d is delivery time, τ_{trans} is transmission time, τ_{proc} is processing time.

We presume τ_{proc} is much smaller than τ_{trans}. If the message needs to go through U hops, and if we assume that the transmission only happens at the end of the Interval I_i, the total delivery time *(τd-$_{total}$)* can be calculated according to the expression below:

$$\tau_{d-total} = \sum_{u=1}^{U} \tau_{w_u} + \tau_{trans_u} \quad (6)$$

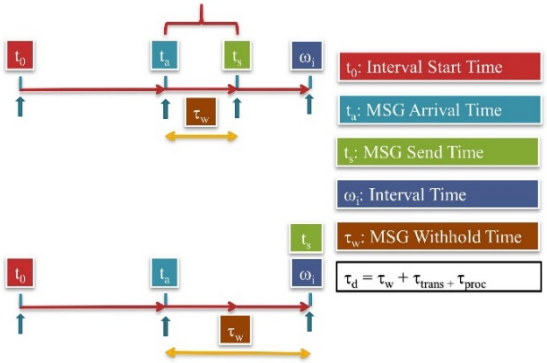

Figure 1: Total delay for a transmitted message.

Transmitting messages at the end of the interval, will increase the delay of the delivery presuming that every τ_{trans} is equal. Thus, optimizing $\tau_{d\text{-}total}$ is a function of τ_w according to the expression below:

$$\tau_{d-total} = f(\tau_w) = \sum_{u=1}^{U} \tau_{w_u} \quad (7)$$

A. Security analysis:

For the best case for anonymity, the adversary sees every sensor node sending a message at a fixed data rate at any one time. It also cannot distinguish any message from the rest of the messages in the network due to the implementation of ID anonymity solution. If we have N nodes in the WSN, the probability that one adversary can locate the sending node or the BS equals to:

$$P_i = \frac{1}{N} \quad (8)$$

One of the methods to quantify for the degree of anonymity is calculating the rational entropy as expressed below:

$$EN(X) = -\sum_{i=1}^{N} [P_i \times \log_2 P_i] \quad (9)$$

The maximum entropy is when the adversary believes every sensor node has the same probability to be the transmitter (uniform distribution):

$$EN_{max} = -\sum_{i=1}^{N} [\frac{1}{N} \times \log_2 \frac{1}{N}] = \log_2 N \textbf{(Error! Bookmark not defined.)}$$

Where $\sum_{i=1}^{N} P_i = 1$, P_i is the probability that the sensor node is the transmitting node. To calculate the degree of anonymity (A):

$$A = 1 - \frac{EN_{max} - EN(X)}{EN_{max}} = \frac{EN(X)}{EN_{max}} \textbf{(Error! Bookmark not defined.)}$$

If all the nodes would send messages at every interval then the anonymity:

$$A = \frac{EN(X)}{EN_{max}} = \frac{\log_2 N}{\log_2 N} = 1 \quad \textbf{(Error! Bookmark not defined.)}$$

However, it is not realistic to have transmissions by all N in one time. If we presume the minimum number of transmissions per one interval (N_{min}) and the average transmissions per one interval (N_r), then anonymity ranges:

$$A = \frac{\log_2 N_{min}}{\log_2 N} < \frac{\log_2 N_r}{\log_2 N} < 1 \quad \textbf{(Error! Bookmark not defined.)}$$

With the presence of global adversary the anonymity will range $0 > A \geq 1$ depending on the number of node transmitting (N_t) at one interval time as exhibited in Figure 2. However, local adversary can detect transmission within its range, let's say 50 nodes density. If the transmissions within the adversary range is more than 50 transmissions, then it would not contribute to the anonymity against that particular adversary. The fake messages transmission could be reduced to have all the transmissions (real and fake) equals to 50.

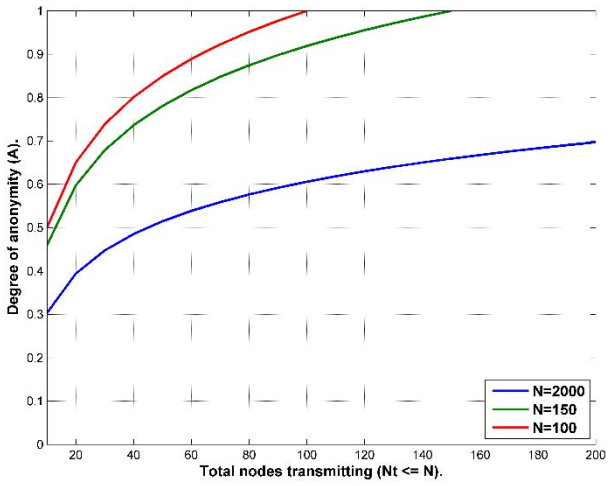

Figure 2: Degree of anonymity (A).

B. Delivery time:

Message follows hop-by-hop path until it gets to the BS as exhibited in Figure 3. In this scheme, the message waits until the end of the interval and before it's sent out. The delay will be calculated according to expression below:

978-1-4799-7678-2/15/$31.00 ©2015 IEEE

$$\tau_{d-total} = \tau_{w_0} + (\mathrm{HC} - 1) * \omega_i + \tau_{trans_u} \quad (10)$$

It axiomatic that most delay accumulates from holding the message until the end of the interval periods.

C. Energy cost:

In our work, we will assume a simple energy dissipation model. The radio dissipates □ nJ/bit for both transmission and reception by the sensors circuitry. And it consumes □ nJ/bit/m^2 for the transmitter amplifier to achieve an acceptable signal to noise ratio. So, to transmit *k* bits for *r* distance, the total transmission energy dissipation will be:

$\mathrm{E_{trans}} = \mathrm{k} *\ + \mathrm{k} *\ \mathrm{r}^2 * \varepsilon$ **(Error! Bookmark not defined.)**

And the receiver would consume for reception of k-bit message:

$\mathrm{E_{receiv}} = \mathrm{k} *$ **(Error! Bookmark not defined.)**

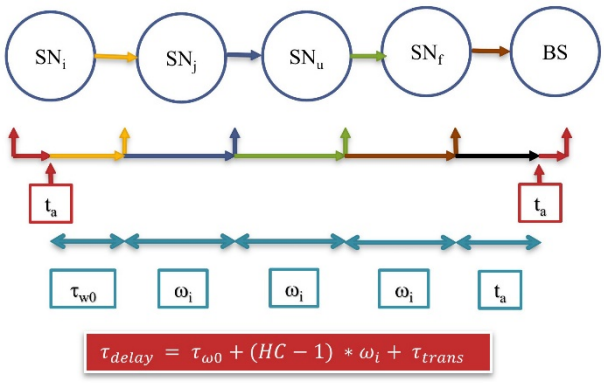

Figure 3: Total delay required to send a message from source to the BS through U hopes.

Presuming that every node would send one message at the end of each interval, the message could be either real or fake. If we have *N* nodes in the WSN, then we expect *N* messages during each interval I_i. The energy spent for transmission or reception is almost constant per a message since we have fixed size messages to prevent size correlation attacks. If we have *p* percent of the nodes issue or forward real data at each interval, then *1-p* percent of the energy and the bandwidth is wasted. We can adjust the amount of energy consumed by increasing the interval time period ω_i. However, increasing ω_i, would increase the delay.

A SN could generate new message and forward received messages during one interval time. If it receives multiple messages, then it will *queue* the messages for transmission. Because the SN needs to wait till the end of the interval I_i, it could arrange the messages in a queue and send them randomly at the end of interval. This approach is energy-expensive due to sending fake/real messages by every node per each interval of time. The consumption of transmitting fake messages is a double fold since the transmitter will consume $\mathbf{E_{trans}}$ for every message and all the neighbors n will consume $(n * E_{trans})$. When the transmission range increases, n increases. The total energy consumed in the network to send real messages in one interval [13]:

$\mathrm{E_R} = (\mathrm{p} * \mathrm{N})(\mathrm{k} *\ + \mathrm{k} *\ \mathrm{r}^2 * \varepsilon) + (\mathrm{p} * \mathrm{N} * \mathrm{n})(\mathrm{k} *\ )$ **(Error! Bookma**

The total energy consumed in the network to send fake messages in one interval:

$\mathrm{E_F} = (1 - \mathrm{p}) * \mathrm{N} * (\mathrm{k} *\ + \mathrm{k} *\ \mathrm{r}^2 * \varepsilon) + (1 - \mathrm{p}) * \mathrm{N} * \mathrm{n} * (\mathrm{k} *\ )$ **(Error! Bookmark not defined.)**

Energy transmission efficiency *ETE* can be calculated as:

$\mathrm{ETE} = \frac{\mathrm{E_R}}{\mathrm{E_R} + \mathrm{E_F}}$ **(Error! Bookmark not defined.)**

To optimize for energy consumption, we need to reduce n and increase p. The number of neighbors usually is a function of the transmission rage where increasing the transmission range would make the message reaches out to more neighboring sensors. The sensor needs to read the message anyways to be able to distinguish between real and fake messages. On the other hand, p is not predictable in most of the applications such as monitoring and tracking since it is event-driven. Unless we have a known distribution for the events, in advance, we cannot optimize for it. Figure 4 exhibits three networks where the minimum number of nodes required to transmit is 30%, 40% and 60% consecutively. If the minimum required transmissions achieved by the real messages then there is no need to send fake messages. Figure exhibits a network consists of 200 sensor nodes, the average number of neighbors is 10 sensors, and the range of transmission is 3m, the size of the message is 1000 byes, the sensor consumes 50 nJ/bit for both reception and transmission and 100 pJ/bit/m2. The network is required to have 60% of nodes send messages at each interval where the messages are a combination of fake and real messages.

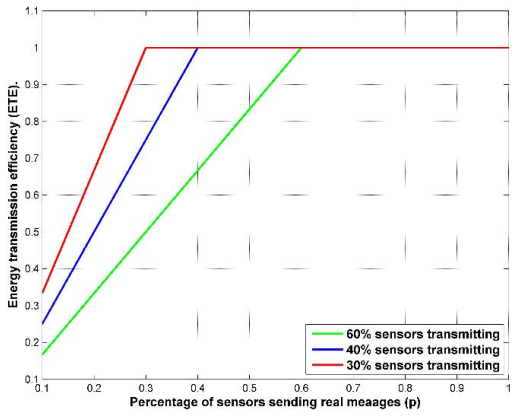

Figure 4: Energy transmission efficiency.

D. Handling rate attack:

One issue that WSN with one sink could endure is having

higher transmission rate nearby the sink where messages ultimately need to reach out to sink as terminal destination. **Figure 6** exhibits the issue. This would endanger the location privacy of the sink. The network is required to maintain similar average rate among all the sensors. This could be accomplished by increasing the volume of fake messages transmitted by lesser busy nodes which means increasing the bandwidth usage and the power consumption. In contrast, we need also to reduce the volume of fake messages sent by busy nodes. The latter is achieved automatically since the sensors don't send fake messages when they have real messages. However, this could be better tuned for average busy nodes as well. Having balanced rate in the WSN could help to maintain balanced lifetime for the nodes in the network. If all the sensor nodes initially are heterogeneous in terms of energy, this would mean that busy nodes would deplete sooner. This scenario could create an empty coverage area or a buffer zone between the sink and the peripheral sensors. This makes it a double fold problem. The first approach is to select a suitable location for the sink in the network map.

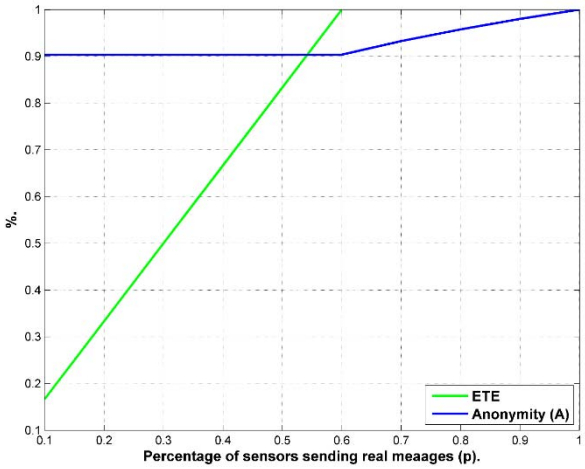

Figure 5: The relationship between entropy and ETE.

Most of the literature shows a side location for the sink. It is maybe due to the fact that it is more suitable for the applications in hand where it is connected to the backbone network in an approachable area and sensors are unattended at out of reach areas. The density of nodes closer to the sink should be higher. We could have multiple density areas around the sink where the density is reduced as it gets distant from the sink. If the storage of the sensor is not big enough, which is not expected with increasing storage technology in the sensors, the sensor does not need to include all the neighbors in the tables. The WSN will be divided into two areas, near (A_{near}) and far (A_{far}). The WSN will set transmission average rate (ATR) thresholds, D_{max} and D_{min}. Sensors in A_{near} will be loaded with D_{max} where the sensors need to queue messages to maintain the threshold. On the other hand, sensors in A_{far} will be loaded with D_{min} to maintain the lower threshold by sending more fake messages as needed.

IV. Conclusions and Future Work

We have provided a solution for temporal attacks which is very important for location privacy. Most of the previous work assumed local adversary view and passive attack model. This work addressed local and global adversary network view. To provide temporal privacy, the global adversary needs to see a maze of transmissions happening all over the network. Fake messages were introduced. However, using fake messages needs to be adjusted to manage the energy consumption.

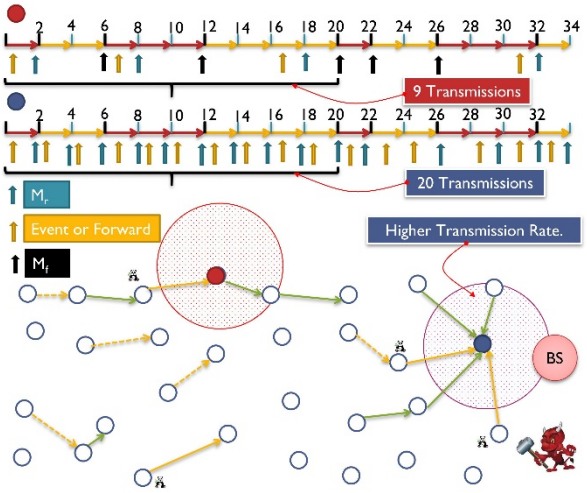

Figure 6: Higher transmission rate next to the sink. The figure exhibits about 20 transmissions nearby the sink and only 9 transmissions at a middle distant sensor.

References

[1] D. Jing, R. Han, and S. Mishra, "Countermeasures Against Traffic Analysis Attacks in Wireless Sensor Networks," in *Security and Privacy for Emerging Areas in Communications Networks, 2005. SecureComm 2005. First International Conference on*, 2005, pp. 113-126.

[2] J. Ying, C. Shigang, Z. Zhan, and Z. Liang, "A novel scheme for protecting receiver's location privacy in wireless sensor networks," *Wireless Communications, IEEE Transactions on,* vol. 7, pp. 3769-3779, 2008.

[3] L. Xinfeng, W. Xiaoyuan, Z. Nan, W. Zhiguo, and G. Ming, "Enhanced Location Privacy Protection of Base Station in Wireless Sensor Networks," in *Mobile Ad-hoc and Sensor Networks, 2009. MSN '09. 5th International Conference on*, 2009, pp. 457-464.

[4] L. Yao, L. Kang, P. Shang, and G. Wu, "Protecting the sink location privacy in wireless sensor networks," *Personal and Ubiquitous Computing,* vol. 17, pp. 883-893, 2013/06/01 2013.

[5] Q. Jing, A. Vasilakos, J. Wan, J. Lu, and D. Qiu, "Security of the Internet of Things: perspectives and challenges," *Wireless Networks,* vol. 20, pp. 2481-2501, 2014/11/01 2014.

[6] Z. Yan, P. Zhang, and A. V. Vasilakos, "A survey on trust management for Internet of Things," *Journal of Network and Computer Applications,* vol. 42, pp. 120-134, 6// 2014.

[7] A. A. Nezhad, A. Miri, and D. Makrakis, "Location privacy and anonymity preserving routing for wireless sensor networks," *Computer Networks,* vol. 52, pp. 3433-3452, 2008.

[8] S. M. a. G. Xue, "Efficient anonymity schemes for clustered wireless sensor networks," *Int. J. Sensor Networks,* vol. 1, 2006.

[9] P. Kamat, Y. Zhang, W. Trappe, and C. Ozturk, "Enhancing Source-Location Privacy in Sensor Network Routing," in *Distributed Computing Systems, 2005. ICDCS 2005. Proceedings. 25th IEEE International Conference on*, 2005, pp. 599-608.

978-1-4799-7678-2/15/$31.00 ©2015 IEEE

[10] L. Xi, J. Xu, and P. Myong-Soon, "Location Privacy against Traffic Analysis Attacks in Wireless Sensor Networks," in *Information Science and Applications (ICISA), 2010 International Conference on*, 2010, pp. 1-6.
[11] B. Alomair, A. Clark, J. Cuellar, and R. Poovendran, "Toward a Statistical Framework for Source Anonymity in Sensor Networks," *Mobile Computing, IEEE Transactions on,* vol. 12, pp. 248-260, 2013.
[12] H. Xiaoyan, W. Pu, K. Jiejun, Z. Qunwei, and L. jun, "Effective probabilistic approach protecting sensor traffic," in *Military Communications Conference, 2005. MILCOM 2005. IEEE*, 2005, pp. 169-175 Vol. 1.
[13] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on*, 2000, p. 10 pp. vol.2.

978-1-4799-7678-2/15/$31.00 ©2015 IEEE