

An Enhanced Source Anonymity Method Framework for Sensor Networks with Replica Detection Using Hypothesis Testing

^IK. Poongodi MCA., ^{II}S. Nivas MCA., M.Phil., Ph.D.,

^IResearch Scholar, Bharathiar University, Coimbatore, ^{II}Head of the Dept, CS,

^{I,II}Dept. of Computer Science, Maharaja Co-Education Arts and Science College,
Perundurai, Erode – 638052.

^IEmail id: poongodi.kandhasamy@gmail.com

^{II}Email id: nivasmaharaja@gmail.com

ABSTRACT: In certain applications, the locations of events reported by a sensor network need to remain anonymous. That is, unauthorized observers must be unable to detect the origin of such events by analyzing the network traffic. Known as the source anonymity problem, this problem has emerged as an important topic in the security of wireless sensor networks, with variety of techniques based on different adversarial assumptions being proposed. This thesis presents a new framework for modeling, analyzing, and evaluating anonymity in sensor networks. The novelty of the proposed framework is twofold: first, it introduces the notion of “interval in distinguish ability” and provides a quantitative measure to model anonymity in wireless sensor networks; second, it maps source anonymity to the statistical problem of binary hypothesis testing with nuisance parameters. The thesis shows how mapping source anonymity to binary hypothesis testing with nuisance parameters leads to converting the problem of exposing private source information into searching for an appropriate data transformation that removes or minimize the effect of the nuisance information. By doing so, it transforms the problem from analyzing real-valued sample points to binary codes, which opens the door for coding theory to be incorporated into the study of anonymous sensor networks. In addition, to mitigate the limitations of previous schemes, the thesis proposes a zone-based node compromise detection scheme in sensor networks. The main idea of the proposed scheme is to use sequential hypothesis testing to detect suspect regions in which compromised nodes are likely placed. A fast and effective mobile replica node detection scheme is proposed using the Sequential Probability Ratio Test.

Keywords: Source Anonymity, Evaluating Anonymity, Sensor Network, Hypothesis Testing, Replica, Traffic, Sequential Probability Radio Test.

1. INTRODUCTION

1.1 Sensor Networks

Sensor networks are deployed to sense, monitor, and report events of interest in a wide range of applications including, but are not limited to, military, health care, and animal tracking. In many applications, such monitoring networks consist of energy constrained nodes that are expected to operate over an extended period of time, making energy efficient monitoring an important feature for unattended networks. In such scenarios, nodes are designed to transmit information only when a relevant event is detected (i.e., event-triggered transmission).

Consequently, given the location of an event-triggered node, the location of a real event reported by the node can be approximated within the node’s sensing range. In the example depicted in Fig. 1.1, the locations of the combat vehicle at different time intervals can be revealed to an adversary observing nodes transmissions. There are three parameters that can be associated with an event detected and reported by a sensor node: the description of the event, the time of the event, and the location of the event.

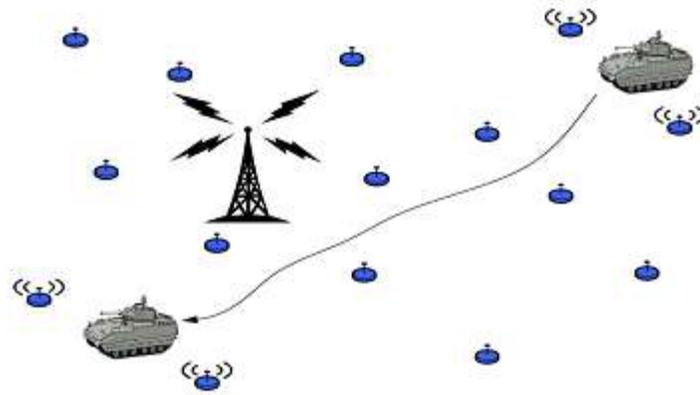


Figure 1.1 A sensor network deployed in a battlefield. Only nodes in close proximity to the combat vehicle are broadcasting information, while other nodes are in sleep mode.

When sensor networks are deployed in untrustworthy environments, protecting the privacy of the three parameters that can be attributed to an event-triggered transmission becomes an important security feature in the design of wireless sensor networks. While transmitting the “description” of a sensed event in a private manner can be achieved via encryption primitives, hiding the timing and spatial information of reported events cannot be achieved via cryptographic means.

Encrypting a message before transmission, for instance, can hide the context of the message from unauthorized observers, but the mere existence of the ciphertext is indicative of information transmission. The source anonymity problem in wireless sensor networks is the problem of studying techniques that provide time and location privacy for events reported by sensor nodes.

1.2 Source Anonymity Problem

In the existing literature, the source anonymity problem has been addressed under two different types of adversaries, namely, local and global adversaries. A local adversary is defined to be an adversary having limited mobility and partial view of the network traffic. Routing-based techniques have been shown to be effective in hiding the locations of reported events against local adversaries.

A global adversary is defined to be an adversary with ability to monitor the traffic of the entire network (e.g., coordinating adversaries spatially distributed over the network). Against global adversaries, routing-based techniques are known to be ineffective in concealing location information in event-triggered transmission. This is due to the fact that, since a global adversary has full spatial view of the network, it can immediately detect the origin and time of the event-triggered transmission.

The first step toward achieving source anonymity for sensor networks in the presence of global adversaries is to refrain from event-triggered transmissions. To do that, nodes are required to transmit fake messages even if there is no detection of events of interest. When a real event occurs, its report can be embedded within the transmissions of fake messages. Thus, given an individual transmission, an observer cannot determine whether it is fake or real with a probability significantly higher than $1/2$, assuming messages are encrypted.

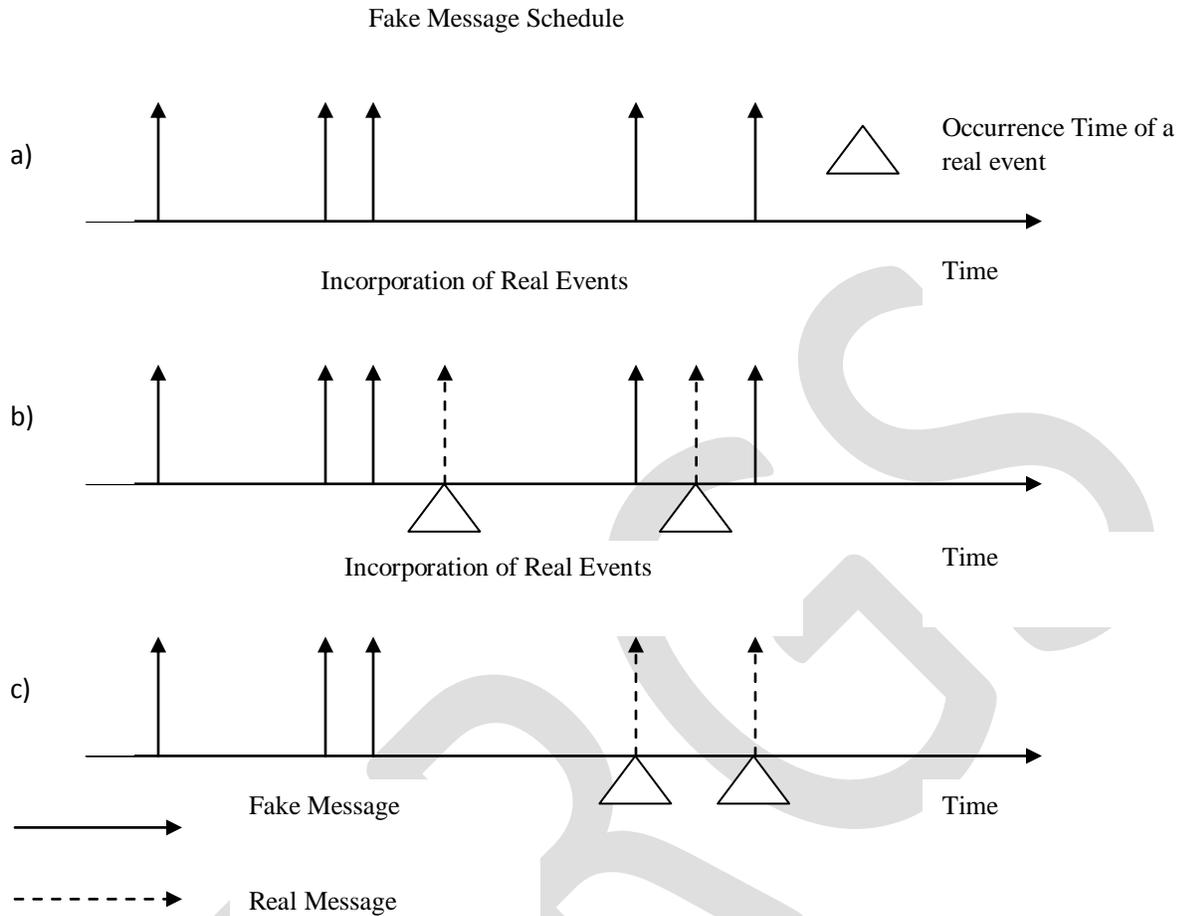


Figure 1.2. Different approaches for embedding the report of real events within a series of fake transmissions; (a) shows the prespecified distribution of fake transmissions, (b) illustrates how real events are transmitted as soon as they are detected, (c) illustrates how nodes report real events instead of the next scheduled fake message.

1.3 Probabilistic Distribution

In the above approach, there is an implicit assumption of the use of a probabilistic distribution to schedule the transmission of fake messages. However, the arrival distribution of real events is, in general, time-variant and unknown a priori. If nodes report real events as soon as they are detected (independently of the distribution of fake transmissions), given the knowledge of the fake transmission distribution, statistical analysis can be used to identify outliers (real transmissions) with a probability higher than 1/2, as illustrated in Fig. 2b. In other words, transmitting real events as soon as they are detected does not provide source anonymity against statistical adversaries analyzing a series of fake and real transmissions.

One way to mitigate the above statistical analysis is illustrated in Fig. 2c. As opposed to transmitting real events as they occur, they can be transmitted instead of the next scheduled fake one. For example, consider programming sensor nodes to deterministically transmit a fake message every minute. If a real event occurs within a minute from the last transmission, its report must be delayed until exactly 1 minute has elapsed.

This approach, however, introduces additional delay before a real event is reported (in the above example, the average delay of transmitting real events is half a minute). When real events have time-sensitive information, such delays might be unacceptable. Reducing the delay of transmitting real events by adopting a more frequent scheduling algorithm is impractical for most sensor network applications since sensor nodes are battery powered and, in many applications, unchargeable. Therefore, a frequent transmission scheduling will drastically reduce the desired lifetime of the sensor network.

1.4 Main Contributions

The main contributions of this thesis are.

- The notion of “interval in distinguish ability” is introduced and illustrated how the problem of statistical source anonymity can be mapped to the problem of interval in distinguish ability.
- A quantitative measure is proposed to evaluate statistical source anonymity in sensor networks.
- The problem of breaching source anonymity is mapped to the statistical problem of binary hypothesis testing with nuisance parameters.
- The significance of mapping the problem is demonstrated in hand to a well-studied problem in uncovering hidden vulnerabilities. In particular, realizing that the SSA problem can be mapped to the hypothesis testing with nuisance parameters implies that breaching source anonymity can be converted to finding an appropriate data transformation that removes the nuisance information.
- Existing solutions under the proposed model is analysed. By finding a transformation of observed data, the problem is converted from analyzing real-valued samples to binary codes and a possible anonymity breach is identified in the current solutions for the SSA problem.

2. PROBLEM FORMULATION

2.1 Main Objectives

- To detect same identity based multi adversaries
- To Localization the hacker node
- Cluster based victim node detection in the network
- Detect the presence of spoofing attacks
- Determine the number of attackers
- Localize multiple adversaries and eliminate them.
- To create Mobile Node Network.
- To make Mobile movement (Random walk) within given speed.
- To update location information to its neighbors.
- To update location information of all nodes to Base Station.
- To make replicate node attack.
- To make Base station identifies the mobile replication attack.

2.2 Specific Objectives

- A fast and effective mobile replica node detection scheme is proposed using the Sequential Probability Ratio Test.
- To tackle the problem of spoofing attacks in mobile sensor networks.
- The scheme detects mobile replicas in an efficient and robust manner at the cost of reasonable overheads.
- The number of attackers when multiple adversaries masquerading as the same node identity

2.3. System Methodology

2.3.1. Proposed Framework for SSA

In this section, we introduce our source anonymity model for wireless sensor networks. Intuitively, anonymity should be measured by the amount of information about the occurrence time and location of reported events an adversary can extract by monitoring the sensor network. The challenge, however, is to come up with an appropriate model that captures all possible sources of information leakage and a proper way of quantifying anonymity in different systems.

2.3.2. Statistical Goodness of Fit Tests and the SSA Problem

2.3.2.1. SSA Solutions Based on Statistical Goodness of Fit Tests

The statistical goodness of fit of an observed data describes how well the data fits a given statistical model. Measures of goodness of fit typically summarize the discrepancy between observed values and the values expected under the statistical model in question. Such measures can be used, for example, to test for normality of residuals, to test whether two samples are drawn from identical distributions, or to test whether outcome frequencies follow a specified distribution.

2.3.2.2. Statistical Goodness of Fit under Interval In distinguish ability

In this section, they are analyzing for statistical goodness of fit-based solutions under the proposed model of interval in distinguish ability. As before, let X_i be the random variable representing the time between the i^{th} and the $(i + 1)^{\text{st}}$ transmissions and let the desired mean of these random variables be μ ; i.e., $IE [X_i] = \mu$, for all i (since the X_i 's are iid). We now examine two intervals, a fake interval and a real one.

2.3.3 Sequential Probability Ratio Test

The enhanced Sequential Probability Ratio Test (SPRT) which is a statistical hypothesis testing. SPRT has been proven to be the best mechanism in terms of the average number of observations that are required to reach a decision among all sequential and non-sequential test processes. SPRT can be thought of as one dimensional random walk with lower and upper limits. Before the random walk starts, null and alternate hypotheses are defined in such a way that the null one is associated with the lower limit and the alternate one is associated with the upper limit. A random walk starts from a point between two limits and moves toward the lower or upper limit in accordance with each observation.

Algorithm process for enhanced SPRT:

DECLARATION: $n=0, w_n=0$

INPUT: location information L and time information T

OUTPUT: accept the hypothesis H_0 or H_1

curr_loc= L

curr_time= T

if $n > 0$ **then**

 compute $T_0(n)$ and $T_1(n)$

 compute speed 0 from curr_loc and prev_loc, curr_time and prev_time

if $0 > V_{max}$ **then**

$w_n = w_n + 1$

end if

if $w_n \geq T_1(n)$ **then**

 Accepts the hypothesis h_1 and terminate the test

end if

if $w_n \leq T_0(n)$ **then**

 initialize n and w_n to 0 and accepts the hypothesis H_0

 return;

end if

end if

$n = n + 1$

prev_loc=curr_loc

prev_time=curr_time

Algorithm Steps

- 1) Create a network of 'n' nodes and save the information in the database table.
- 2) Draw the network with the available node information.
- 3) Random walk procedure is worked out so that the nodes' mobility is carried out by just moving its location with 'n' pixels below (the given speed) in both x and y direction. For example, if the speed is given as 10 units, then a random value below 10 is chosen, and the node is moved in x or y direction. This is carried out for all nodes. For simulation, the timer is set to 5 seconds. So once each 5 seconds, all the nodes are moved within the given speed horizontally or vertically.
- 4) The nodes are sending their location to their neighbor nodes. The node is treated as neighbor to one, if it is within the given pixel units. For example, the unit is given as 50, then a node with left position in the space with 150 x value and another node with 180 x value is treated as neighbor nodes. This is applicable to y axis also. So in the rectangular area of 50 units (side), when the two nodes fall inside, then they are treated as neighbor nodes.
- 5) The nodes are updating their location information once in 10 seconds. The arrow lines are drawn during the animation such that from all nodes, the line is drawn to the base station. The area located at left bottom corner of the drawing space in the form.
- 6) Replica Attack: When a button is clicked, a node is chosen randomly which behaves as attacker node; a node is chosen randomly which behaves as affected node. The attacker node through sends the current location information, it sends its

id as the affected node. So the base station receives updates with two ids at single update. Now, the base station needs to identify which node is correct and which is attacker.

- 7) If two nodes send same id, then the base station, collects the previous location information of the same id. Any one of the entry will have wrong previous location. At the same time, the neighbor nodes location data is also used such that, the affected nodes neighbors update correct location of suspected id whether the attacker nodes neighbor nodes update wrong location and the attacker node will be identified.
- 8) Then the node is revoked from the network.

Techniques to Detect Compromised Nodes In Zones

Reputation-based trust management schemes do not stop compromised nodes doing malicious activities in the network. Also, the existing schemes based on software attestation require each sensor to be periodically attested because it cannot be predicted when attacker compromises sensors. The periodic attestation of individual nodes will incur large overhead in terms computation and communication overhead.

To mitigate the limitations of both approaches, a zone-based node compromise detection scheme is proposed which facilitates node compromise detection and revocation by leveraging zone trust information. Specifically, the **network is divided into a set of zones, establish trust per zone, and detect untrustworthy zones in accordance with zone trust values.**

Once a zone is determined to be untrustworthy, the network operator attests the software modules of all sensors in the untrustworthy zone, and detects and revokes compromised nodes in that zone.

A straightforward approach for untrustworthy zone detection is to decide a zone as untrustworthy by observing a single evidence that its trust value is less than a pre defined threshold. However, this approach does not consider the zone trust measurement error. Due to the error occurrence in the zone trust measurement, trustworthy (resp. untrustworthy) zone could be detected as untrustworthy (resp. trustworthy).

To minimize these false positive and negatives, it needs to make a decision with multiple pieces of evidence rather than single evidence. To meet this need, the **Sequential Probability Ratio Test (SPRT)** is used, which is a **statistical decision process** that makes a decision with multiple pieces of evidence. The SPRT benefits in the sense that the **SPRT reaches a decision with a small number of evidences** while achieving the low false positive and negative rates. The SPRT can be thought of as one-dimensional random walk with lower and upper limits.

It is believed that SPRT is well-suited for tackling the **compromised node detection problem in the sense that a random walk with two limits can be constructed in such a way that each walk is determined by the trust value of a zone**; the lower and upper limits are properly configured to be associated with the excess and shortfall of a predefined trust threshold, respectively.

Protocol Operation

The proposed protocol to find the compromised zones proceeds in three phases:

1) Phase I:

Zone Discovery and Trust Aggregator Selection: After deployment, every sensor node u finds out its location and determines the zone to which it belongs. This zone is called the home zone. From u 's point of view, other zones are called as the foreign zones. Node u discovers every other node residing in the same zone. After the zone discovery process, the Trust Aggregator (TA) is selected in a round robin manner. Specifically, the time domain of a zone is partitioned into time slots. An initial duty time slot is assigned to each node u in the zone according to the ascending order of the nodes' IDs. Each node u then acts as trust aggregator every S time slots starting from its initial duty time slot, where S is the number of nodes residing in the zone.

2) Phase II:

Trust Formation and Forwarding: For each time slot T_i , each node u in zone Z computes neighborhood-trust that is defined in accordance with the difference between the probability distributions of the information generated by u and the information sent to u by u 's neighboring nodes in zone Z .

3) Phase III:

Detection and Revocation: Upon receiving a zone-trust report from a TA in zone Z , the base station verifies the authenticity of TA's report with the secret shared key between TA and itself and discards the report if it is not authentic. The base station also maintains a record per TA associating each TA's ID with its home zone. This prevents compromised TAs from claiming multiple home zones.

3. SYSTEM DESIGN

3.1. Module Description

The following modules are present in the thesis

- **Real interval identification using interval in distinguish ability**
- **Fake interval**
- **Mobile node network creation.**
- **Mobile movement (random walk) within given speed.**
- **Update location information to its neighbors.**
- **Base station updates location information of all nodes.**
- **Replicate node.**
- **Base station identifies the mobile replication attack.**

1. Real Interval Identification Using Interval In Distinguish Ability

In this module, sender node C chooses two intervals I_R and I_F , in which I_R is a real interval and I_F is a fake one. C draws a bit $b \in \{0, 1\}$ uniformly at random and sets $I_R = I_b$ and $I_F = I_{b'}$, where b' denotes the binary complement of b . C gives I_b and $I_{b'}$ to receiver A . A makes any statistical test of her choice on I_b and $I_{b'}$ and outputs a bit b' . If $b' = b$, A wins the transmission.

2. Fake Interval

In the absence of real events, nodes are programmed to transmit fake messages according to a pre-specified probability distribution. Nodes transmit fake messages according to a pre specified probabilistic distribution and maintain a sliding window of inter transmission times. When a real event occurs, it is transmitted as soon as possible under the condition that the samples in the sliding window maintain the designed distribution.

3. Mobile Node Network Creation

In this module, a form is generated which contains a text box to get node id and the id is saved in to 'Nodes' table. During network creation, the nodes with id will be displayed in random X and Y position. The base station node is need not be displayed as it is programmatically listens and updates the location information of all the nodes when they are in movement.

4. Mobile Movement (Random Walk) Within Given Speed

In this module, all the nodes are roaming in any directions (their walk is updated by incrementing x-axis or y-axis or both at a movement with any number of pixels within the specified maximum limit. In practical situation, the nodes can move with their physical capabilities. For sake of convenience, if the nodes reach the picture box limit, then they move in opposite direction so that they roam in the rectangular boundary of the picture box control.

5. Update Location Information to Its Neighbors

In this module, all the nodes are calculating the neighbor nodes with their transmission range (specified in 'n' units common for all nodes. It means than all the sensor nodes are having homogeneous transmission ranges). Then it gives the location information i.e., its position to all of its neighbors. It occurs for all the nodes at regular intervals. The timer control is provided and the time is considered in global aspect. All the nodes are having unique time values.

6. Base Station Updates Location Information Of All Nodes

In this module, the base station is collecting the location information from all nodes. It occurs for all the nodes at regular intervals. It is assumed that no two nodes are in same location since the nodes purpose is to serve individually a specific area.

7. Replicate Node

In this module, the node is updating its location information to base station with one of the remaining nodes. It means that it is replicating some other node. This results in, at a given time, both the nodes are sending same location information to the base station of which one is true and other is false.

8. Base Station Identifies the Mobile Replication Attack

This module presents the details of the technique to detect replica node attacks in mobile sensor networks. In static sensor networks, a sensor node is regarded as being replicated if it is placed in more than one location. If nodes are moving around in network, however, this technique does not work, because a benign mobile node would be treated as a replica due to its continuous change in location.

The base station computes the speed from every two consecutive claims of a mobile node and performs the SPRT by considering speed as an observed sample. Each time the mobile node's speed exceeds (respectively, remains below) V_{max} , it will

expedite the random walk to hit or cross the upper (respectively, lower) limit and thus lead to the base station accepting the alternate (respectively, null) hypothesis that the mobile node has been (respectively, not been) replicated. Once the base station decides that a mobile node has been replicated, it revokes the replica nodes from the network.

4. RESULT AND DISCUSSION

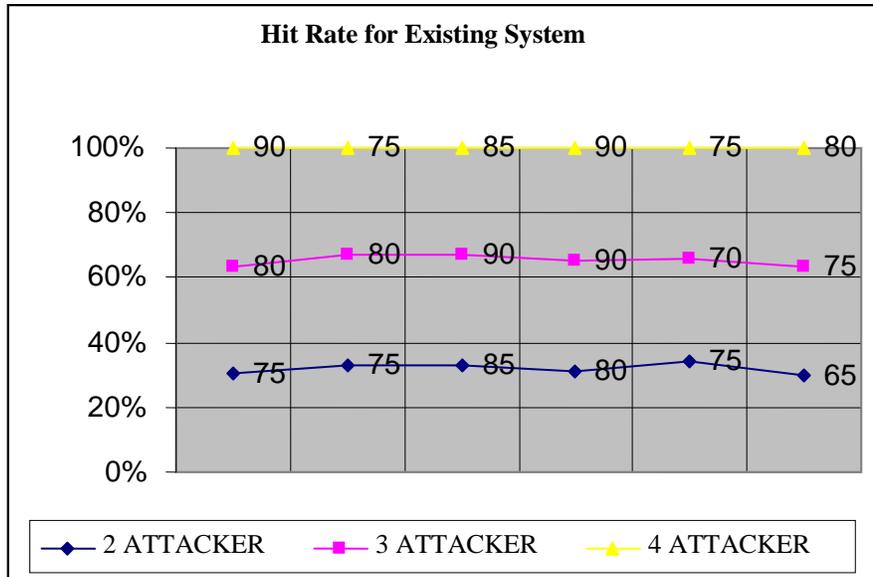
Provided the training data collected during the offline training phase, we can further improve the performance of determining the number of spoofing attackers. In addition, given several statistic methods available to detect the number of attackers, such as System Evolution and combine the characteristics of these methods to achieve a higher detection rate.

In this section, we explore using hypothesis testing to classify the number of the spoofing attackers. The advantage of using hypothesis is that it can combine the intermediate results (i.e., features) from different statistic methods to build a model based on training data to accurately predict the number of attackers.

Hit Rate %	Hit Rate Existing System	Hit Rate Proposed System
75	80	90
75	80	75
85	90	85
80	90	90
75	70	75
65	75	80

The training data set can be obtained through regular network monitoring activities. Given a training set of instance-label pairs and the label, the support vector machines require the solution of the following optimization problem:

Error Rate for Existing System



$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^l \xi_i$$

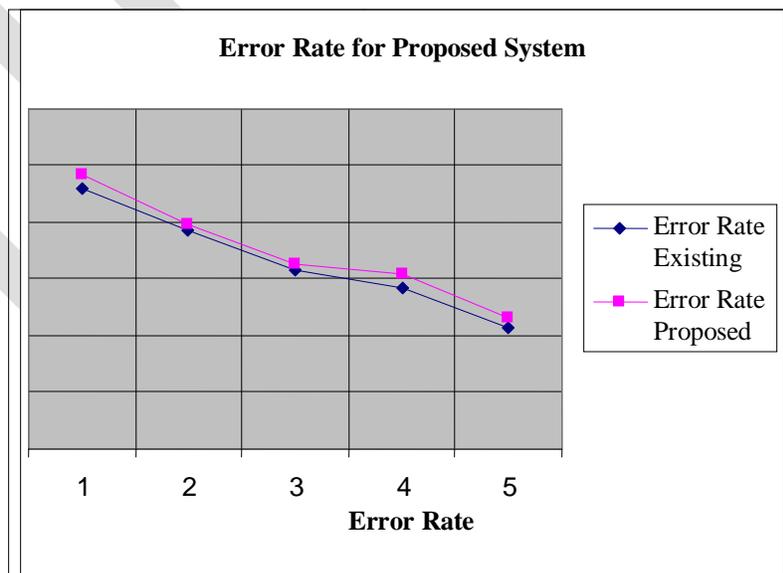
Subject to $y_i(\mathbf{w}^T \phi(x_i) + b) \geq 1 - \xi_i,$
 $\xi_i \geq 0.$

Its dual is

$$\min_{\alpha} \frac{1}{2} \alpha^T Q \alpha - \mathbf{e}^T \alpha$$

Subject to $\mathbf{y}^T \alpha = 0,$
 $0 \leq \alpha_i \leq C, \quad i = 1, \dots, l,$

Error Rate for Proposed System



5. CONCLUSION AND FUTURE ENHANCEMENTS

This thesis proposed to use received signal strength-based spatial correlation, a physical property associated with each wireless device that is hard to falsify and not reliant on cryptography as the basis for detecting spoofing attacks in wireless networks. It provided theoretical analysis of using the spatial correlation of RSS inherited from wireless nodes for attack detection. It derived the test statistic based on the cluster analysis of RSS readings. The approach can both detect the presence of attacks as well as determine the number of adversaries, spoofing the same node identity, so that we can localize any number of attackers and eliminate them.

In addition, a zone-based node compromise detection scheme is proposed using the Chronological Likelihood Fraction Test (CLFT). Furthermore, several possible attacks are described against the proposed scheme and proposed counter-measures against these attacks. The scheme is evaluated in simulation under various scenarios. The experimental results show that the scheme quickly detects untrustworthy zones with a small number of zone-trust reports.

6. REFERENCES

- [1] B. Alomair, A. Clark, J. Cuellar, and R. Poovendran, "On Source Anonymity in Wireless Sensor Networks," Proc. IEEE/IFIP 40th Int'l Conf. Dependable Systems and Networks (DSN '10), 2010.
- [2] K. Mehta, D. Liu, and M. Wright, "Location Privacy in Sensor Networks Against a Global Eavesdropper," in ICNP 2007. IEEE International Conference on Network Protocols. , 2007.
- [3] B. Alomair, A. Clark, J. Cuellar, and R. Poovendran, "Statistical Framework for Source Anonymity in Sensor Networks," Proc. IEEE GlobeCom, 2010.
- [4] P. Kamat, Y. Zhang, W. Trappe, and C. Ozturk, "Enhancing Source-Location Privacy in Sensor Network Routing," ICDCS 2005. The 25th IEEE International Conference on Distributed Computing Systems.
- [5] C. Ozturk, Y. Zhang, and W. Trappe, "Source-location privacy in energy-constrained sensor network routing," in Proceedings of the 2nd ACM workshop on Security of ad hoc and sensor networks, 2004.
- [6] Y. Xi, L. Schwiebert, and W. Shi, "Preserving source location privacy in monitoring-based wireless sensor networks," in IPDPS 2006. The 20th International Parallel and Distributed Processing Symposium, 2006.
- [7] B. Hoh and M. Gruteser, "Protecting Location Privacy Through Path Confusion," in SecureComm 2005. First International Conference on Security and Privacy for Emerging Areas in Communications Networks., 2005.

- [8] Y. Ouyang, Z. Le, G. Chen, J. Ford, F. Makedon, and U. Lowell, "Entrapping Adversaries for Source Protection in Sensor Networks," in Proceedings of the 2006 IEEE International Symposium on World of Wireless, Mobile and Multimedia Networks, 2006.
- [9] K. Mehta, D. Liu, and M. Wright, "Location Privacy in Sensor Networks Against a Global Eavesdropper," in ICNP 2007. IEEE International Conference on Network Protocols., 2007.
- [10] M. Shao, Y. Yang, S. Zhu, and G. Cao, "Towards Statistically Strong Source Anonymity for Sensor Networks," INFOCOM 2008. The 27th IEEE Conference on Computer Communications., 2008.
- [11] Y. Yang, M. Shao, S. Zhu, B. Urgaonkar, and G. Cao, "Towards event source unobservability with minimum network traffic in sensor networks," in Proceedings of the first ACM conference on Wireless network security, 2008.
- [12] N. Li, N. Zhang, S. Das, and B. Thuraisingham, "Privacy preservation in wireless sensor networks: A state-of-the-art survey," Ad Hoc Networks, 2009.
- [13] A. Perrig, R. Szewczyk, J. Tygar, V. Wen, and D. Culler, "SPINS: Security Protocols for Sensor Networks," Wireless Networks, vol. 8, no. 5, pp. 521-534, 2002.
- [14] D. Hong, J. Sung, S. Hong, J. Lim, S. Lee, B. Koo, C. Lee, D. Chang, J. Lee, K. Jeong, H. Kim, J. Kim, and S. Chee, "HIGHT: A New Block Cipher Suitable for Low-Resource Device," Proc. Eighth Int'l Workshop Cryptographic Hardware and Embedded Systems (CHES '06), pp. 46-59, 2006.
- [15] National Institute of Standards and Technology (NIST), FIPS-197: Advanced Encryption Standard, November 2001.
<http://www.itl.nist.gov/fipspubs/>