A Novel Protection for Wireless Sensor Networks from Internal Attacks

Xu Huang, Muhammad R Ahmed, and Dharmendra Sharma

***Abstract*—Wireless sensor networks (WSNs) are becoming part of our daily life as they are widely used due to they are easy and rapid deployed, low cost, low power, self-organized, cooperatively collect the environmental information and realize the integration of the physical world and communication network. However, due to their open nature of the wireless medium an adversary can easily eavesdrop and replay or inject fabricated messages. Different cryptographic methods can be used to defend against some of such attacks. But for node compromised those methods can do little, which is another major problem of WSN security as it allows an adversary to enter inside the security perimeter of the network, which raised a serious challenge for WSNs. This paper is focusing on investigating internal attacks of wireless sensor networks with multi-hop and single sinker. Our novel algorithm, called 2 by 2 (2x2) method, enable to effectively protect WSN from internal attacks such as blackhole, Sybil attacks, node replication, etc.**

***Index Terms*—WSN security, insider attack, abnormal behavior, location identification, location detection, time difference of Arrival (TDoA)**

I. Introduction

WIRELESS sensor network (WSN) consists of spatially distributed autonomous sensors and provide a theoretical basis for many different applications range military implementation in the battlefield, environmental monitoring, health sector as well as emergency response of surveillance. It is an application dependent technology which can be changed with additional sensor nodes deployed based on the necessity. A typical WSN is composed of a large number of sensor nodes responsible for sensing data and a sink node responsible for collecting and processing data as shown in Figure 1. The sensor nodes consists a transceiver unit (combination of transmitter and receiver), a restricted memory processing unit, a sensing unit as well as a battery with limited power. Thus, for any application overhead of computation and communication is low.

In order to ensure the efficient functionality of a WSN, security mechanism is essential, especially in the field of emergency response or battlefield implementations. Indeed, security in the wireless sensor network is challenging and important task because of the construction of the node. Many algorithms have developed in order to secure WSNs.

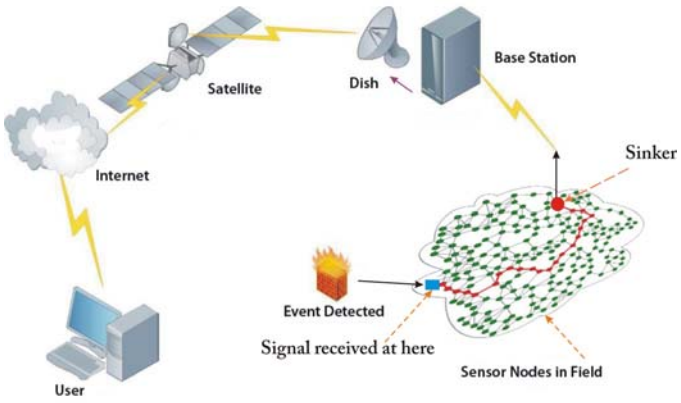

Figure 1: A Typical WSN architecture.

Most of the works have focused on the pair wise key establishment, authentication access control and defense against attacks. These works mainly focused on the traditional cryptographic information, data authentication in order to build the relationship between the sensors the opened unreliable communication channels through wireless connecting make the techniques vulnerable by allowing the sensor nodes to compromise and release the security in formation to the adversary [1]. Through this type of access, adversaries can easily attack the network internally with data alteration, message negligence, selective forwarding as well as by jamming the network.

Theoretically adversaries can be determined through the abnormal behavior of the sensor. Unfortunately, the internal attack (the sensor behaves abnormally) remains unsolved through the conventional way of WSNs security which implements the encrypting method or authentication. Thus, it is important to detect comprised nodes and their location information to provide the security to WSNs [2]. In this research work we proposed a two-step with two-level method, called 2 by 2 (2x2) method, to overcome the security issue of internal attacks.

For the first step, the first level is the identification of insider attacker based on the judgement of abnormal behaviour of the node with various parameters of the defined performances in an interesting network. For the first step, the 2nd level is further confirming the abnormal behaviours by the collected parameters the by means of Dempster-Shafer Theory. For the 2nd step, the first level is dividing the network area into reasonable size that the distributed beacons can fairly carry out the task that is identification of the location of the nodes with the abnormal behaviours. For the 2nd level of the 2nd step is making the

Xu Huang is with Faculty of Information Science and Engineering, University of Canberra, Australia, (e-mail: xu.huang@canberra.edu.au)

Muhammad R Ahmed is with the Faculty of Information Science and Engineering, University of Canberra, Australia, (e-mail: muhammad.ahmed@canberra.edu.au)

Dharmendra Sharma is with Faculty of Information Science and Engineering, University of Canberra, Australia, (e-mail: dharmendra.sharma@canberra.edu.au)

ISBN: 978-988-19251-1-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

detection to the insider attacker with designed performance. However, sometimes a further action will be taken to make the network secure by reprogramming the node or obsolete the node form the network.

The paper is organized as follows: section 2 is comprised of the overview of the related work followed by a description of the proposed framework in section 3. This section covers the details of insider attacker identification process and location detection. The results will be presented in section 4. The efficiency of the framework is presented in Result section followed by conclusion section 5.

II. RELATED WORKS

Numerous ways and solution have been proposed to secure WSNs. However, it is noted the fact that WSNs have the open nature of the wireless medium an adversary can easily eavesdrop, which is called a "passive attacker", and replay or inject fabricated messages, so-called an "active attacker." It is well known that for the protection from the WSNs attacks there are various cryptographic methods can be used and sometimes are very efficient and effective [8-10]. Moreover, because of WSN deployments in open and possibly hostile environments, attackers can easily lunch "denial-of-service" (DoS) attacks, cause physical damage to sensors or capture them to extract sensitive data such as identities, encryption keys, address, other privacy data, etc. However, internal attacks attracted great attentions to the people who have been working in the fields as it allows an adversary to enter inside the security perimeter of the network to attack the network. For example, a node is so-called compromised, the attack can produce internal attack such as Sybil attacks, node replication or black-grey-worm-sink holes. As mentioned above that cryptography to secure routing functionalities is relevant against the aforementioned internal attacks, as it can introduce false topological, neighborhood, control, and routing information. It may just simply drop message as a black hole. So far, there is little research literature in investing and analyzing against those type attacks. In our current paper, we are focusing on investigating internal attacks of wireless sensor networks.

WSNs use multi-hop communication to increase network capacity, in multi-hop routing, messages may traverse many hops before reaching their destinations. However, simple sensor nodes are usually not well physically protected because they are cheap and can be deployed in open or hostile environments where they can be easily captured and compromised, which is known as the fact that an adversary can extract sensitive information. When a node is compromised, an adversary gains access to the network and can produce malicious activities. The attacks are involved in corrupting network data or even disconnecting major part of the network.

Karlof and Wagner have discussed attacks at the network layer in and mentioned altered or replayed routing information and selective forwarding, node replication, Sybil attacks or black-grey-sink holes, and HELLO flooding. Some papers discussed various attacks in term of network's resiliency, such as, discussed how to keep WSN routing protocols as stateless as possible to avoid the proliferation of specific attacks and provide for a degree of random behavior to prevent the adversary from determining which the best nodes to be compromised are. They defined three items, namely (a) average delivery ratio, (b) average degree of nodes, and (c) average path length to describe the networks resiliency.

Unlike traditional routing, where intermediate nodes just forward input packets, in network coding intermediate nodes actively mix or code input packets and forward the resulting coded packets. The very nature of packet mixing also subjects network coding systems to a severe security threat, knows as a pollution attack, where attackers inject corrupted packets into the network. Since intermediate nodes forward packets coded from their received packets, as long as least one of the input packets is corrupted, all output packets forwarded by the node will be corrupted. This will further affect other nodes and result in the epidemic propagation of the attack in the network.

It has addressed pollution attacks against network coding systems in wireless mesh networks. They proposed a lightweight scheme, DART, which uses time-based authentication in combination with random liner transformations to defend against pollution attacks. So far, security using the information from attackers that defined as abnormal behavior of the sensor to make the identification of location discovery of the comprised nodes did not given significant attention. Even though number of localization process has been proposed in different research but main focus was given on preventing and securing routing from attacks.

As the study was done by [3], however, most of the scheme proposed are needed to have special device, such as SeRLoc [4] the improved version of SeRLoc is HiRLoc [5] requires directional antenna, SPINE requires nano second timing scale. Attack resilient location estimation method [6] proposed by Lui fails if the attacker is compromised. ROPE is combination of SeRLoc and SPINE [7], it require extra hardware and pair wise key with every locator. Recently, a few papers published regarding the protections WSNs from internal attacks [19-21], but they are not fully showing the picture of the 2 stages with 2-level method. This paper is based on our previous research results to extend them to the 2x2 method.

These developments somehow solve the mathematical problems with certain constrain but does not take the insider attacker identification and location detection in consideration. In our paper we have come up with the approach to identify and detect the internal attacker.

III. NETWORK MODEL OR FRAMEWORK DESCRIPTION

A. Conditions and Assumptions

In our experimental works we use the following parameters: a network with N uniformly distributed sensor node over the area of 500m * 500m squared field in a 2D scenario. Sensors and channels are stationary after deployment of the network with transmission radius of 200m. Sensing nodes are responsible to collect and forward

ISBN: 978-988-19251-1-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

the monitored data around them. The collected data is then sent to the sinker through channel. In order to detect the abnormal behavior of the sensor node we use the false massage detection. We will consider the system is synchronized.

B. Definitions of Abnormal Behaviour/Attacker Identification

The abnormal behavior or exponential message detection process in a channel is detected with one stationary sinker in this paper. It is well known that insider attacker or abnormal behavior is not possible to detect only based on the cryptographic based technique, as the unreliable opened wireless channel makes it very easy to be compromised, the sensors will break the trust relationship established, so the security foundation become insufficient [2]. The false message detection mechanism focuses on the contingency of the message which defines as the exponential message attack. WSN is densely deployed and continuously observe the phenomenon, this characteristics drive the senor nodes network normally encounter the spatial-temporal correlation. In our research we considered the message generated from the nodes is similar for a defined period. In normal massage delivery of the nodes the probability of different message are negligible or rare. If D is the length of the message with restricted memory, M_i is the message and F_i is considered as the frequency of the message we can write the equation as below

$$D = \{(M_i, F_i) \mid (M_1, F_1); (M_2, F_2); ..., (M_n, F_n)\} \quad (1)$$

It is a set that will store the latest massage that is sent to the network recently. If a new message sent to the network than that is M_{new} arrives at the channel than that is authenticated using the false message detection process [8].

Hence, it can be expressed as the equation shown below:

$$mach\,(M_{new}, M_i) = \frac{(V(M_{new}) \times V(M_i))}{(|V(M_{new})| \times |V(M_i)|)} \quad (2)$$

Based on the k-nearest neighbour algorithm we can find the normal message from the equation (2), this is the simple algorithm that classifies the data based on neighbour training example [9].

Here M_i will be equivalent to D, and if the result of the equation (2) matches with the designed threshold then it will be considered as normal massage. The value of the M_{new} will be compared with the whole set of Ds to determine whether does it match with M_i or not. If it matches it will increase the frequency F_i, or else it will be considered as false message or abnormal behaviour and will be hold in to the buffer until it is authenticated. If it does not match than it will be considered as fake massage and it is the candidate of an attack and the related node is most likely to be comprised. If the authentication process is not passed it will be considered as a fake message and will be identified as the attacked or abnormal sensor.

In this method the calculation is simpler, the latency is smaller as well as less parameter is considered which is supported by the limited memory sensor nodes [8].

C. Dempster-Shafer Theory and Judgment of Abnormal Behavior

The Theory of Evidence is a branch of mathematics that is concerned with combining evidence to calculate the probability of an event. The Dempster-Shafer theory (D-S theory) is a theory of evidence used to combine separate pieces of evidence to calculate the probability of an event. The Dempster-Shafer theory was introduced in the 1960's by Arthur Dempster [1968] and developed in the 1970's by Glenn Shafer [1976]. According to Glen Shafer the D-S theory is a generalization of the Bayesian theory of subjective probability [22].

The Frame of Discernment (Θ):
A complete (exhaustive) set is describing all of the sets in the hypothesis space. Generally, the frame is denoted as Θ. The elements in the frame must be mutually exclusive. If the number of the elements in the set is n, then the power set (set of all subsets of (Θ) will have 2^n elements.

The theory of evidence assigns a belief mass to each subset of the power set. It is a positive number between 0 and 1. It exists in the form of a probability value.

If Θ is the frame of discernment, then a function
m: $2\Theta \rightarrow [0, 1]$ is called an *BPA*, whenever
$m(\emptyset) = 0$ and
$\Sigma\, m(A) = 1$ and
$A \subseteq \Theta$
Here, *BPA* is Basic Probability Assignment.

Given a frame of discernment Θ and a body of empirical evidence $\{m(B_1), m(B_2), m(B_3)....\}$, the belief committed to A and ε is

There is also a definition of *Belief (Bel)*, which is defined as below:

$$Bel\,(A) = \Sigma\, m(B_i)$$
$$B \subseteq A$$

Also, we have $Bel\,(\Theta) = 1$

The plausibility (*Pl*) is the sum of all the masses of the sets B that intersect the set of interest A:

$$Pl\,(A) = \Sigma\, m\,(B_i),\ B \mid B \cap A \neq \emptyset$$

Here, the interval [*Bel* (*A*), *Pl*(*A*)] is called the belief range. It is noted that Plausibility (*Pl*) and Belief (*Bel*) are related as follows:

$$Pl\,(A) = 1 - Bel\,(\bar{A}) \quad (3)$$

Dempster 's Combination Rule:

The combination called the joint mass (m_{12}) is calculated from the two sets of masses m_1 and m_2.

$$m_{12}(A) = \frac{B \cap C = A, \sum m_1(B) m_2(C)}{1 - [B \cap C = 0, \sum m_1(B) m_2(C)]} \quad (4)$$

where, $m_1(B)$ and $m_2(C)$ are evidence supporting hypothesis B and C respectively as observed by m_1 and m_2

Dempster-Shafer theory gives a rule of combining sensor S_i's observation m_i and sensor S_j's observation m_j:

ISBN: 978-988-19251-1-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$$(m_i \oplus m_j)(A) = \frac{\sum_{E_k \cap E_{k'} = A} m_i(E_k)\, m_j(E_{k'})}{1 - \sum_{E_k \cap E_{k'} = \phi} m_i(E_k)\, m_j(E_{k'})} \quad (5)$$

This combining rule can be generalized by iteration: if we treat m_j not as sensor S_j's observation, but rather as the already combined (using Dempster-Shafer combining rule) observation of sensor S_k and sensor S_l [23].

D. Identifying Attack / Abnormal Sensor Location

Location estimation is a complex process that involves multifaceted numerical operations. Unfortunately there is no simple process exists for the efficient computation of a location estimation of wireless sensor nodes. It is uncertain that a more complex mathematical computation will increase the accuracy of the estimation, conversely if a reduction of the complexity that would compromise with the efficiency of location estimation.

In this paper we used the Time Difference of Arrival (TDoA) signal rather than absolute time of Time of Arrival (ToA). As mentioned above that first level for the 2nd step, we need to make sure the designed beacons can cover the area that we are interested to obtain the locations we are going to find out. A signal is sent to the node by at least three antennas at an unknown and different time. The most common trilateration method is used in order to get the sensor node location [11]. For each TDoA measurement, the transmitter must lie on a hyperboloid with a constant range difference between the two measuring units. If we consider B_1 is the Master Beacon node. The distance between the source and *i-th* Beacon node is

$$R_i = \sqrt{(X_i - x)^2 - (Y_i - y)^2} \quad \textbf{(6)}$$

In the 2-D scenario the target location can be estimated from the intersection of two TDoA measurements. Beacon nodes (B_1, B_2 and B_3) are considered as a measuring unit from which intersection point is determined and that locates the target point A, as shown in Figure 2 below:

In Figure 2, Three sensors are defined as B_i with the location as (x_i, y_i), where i = 1, 2 or 3. For any point A = (x, y) in the plane [11].

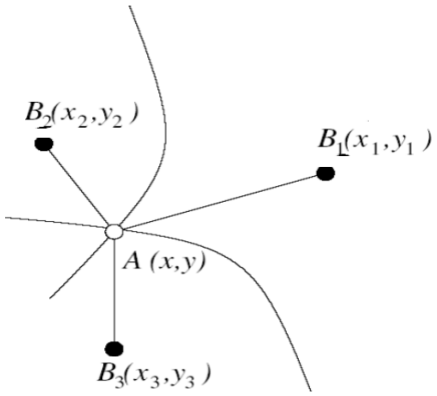

Figure 2: Attacker Location Detection by beacon nodes

The range difference between beacons with respect to the beacon B_1 where the signal arrives first, is

$$R_{i,1} = cd_{i,1} = R_i - R_1 \quad (7)$$

where c is the signal propagation speed, $R_{i;1}$ is the range difference distance between the first beacon B_1 and the *i-th* beacon ($B_{1(i>1)}$), R_1 is the distance between the first beacon and the source, and $d_{i;1}$ is the estimated TDOA between the first beacon B_1 and the *i-th* beacon ($B_{1(i>1)}$) . This defines the set of nonlinear hyperbolic equations whose solution gives the 2-D coordinates of the source.

Solving the nonlinear equations of (7) is dificult. Consequently, linearizing this set of equations is commonly performed. One way of linearizing these equations is through the use of Taylor-series expansion and retaining the first two terms [12,13]. A commonly used alternative method to the Taylor-series expansion method, presented in [14, 15, 16, 17], is to first transform the set of nonlinear equations in (7) into another set of equations. Rearranging the form of (7) into

$$R_{i,1}^2 = (R_{i,1} + R_1)^2 \quad (8)$$

Subtracting (6) at i = 1 from (8) results in

$$R_{i,1}^2 + 2R_{i,1}R_1 = X_i^2 + Y_i^2 - 2X_{i,1}x - 2Y_{i,1}y + x^2 + y^2 \quad (9)$$

where $X_{i;1}$ and $Y_{i;1}$ are equal to X_i - X_1 and Y_i and Y_1 respectively. The set of equations in (9) are now linear with the source location A(x; y) and the range of the first receiver to the source R_1 as the unknowns, and are more easily handled.

Inorder to solve the R_1 we use Chan's method, in this method A non-iterative solution to the hyperbolic position estimation problem which is capable of achieving optimum performance for arbitrarily placed sensors was proposed by Chan [18]. The solution is in closed-form and valid for both distant and close sources. When TDOA estimation errors are small, this method is an approximation to the maximum likelihood (ML) estimator.

Following Chan's method [18], for a three beacon node system (B = 3), producing two TDOA's, x and y can be solved in terms of R_1 from (9). The solution is in the form of

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} X_{2,1} & Y_{2,1} \\ X_{3,1} & Y_{3,1} \end{bmatrix}^2 \times \left\{ \begin{bmatrix} R_{2,1} \\ R_{3,1} \end{bmatrix} R_1 + \frac{1}{2} \begin{bmatrix} R_{2,1}^2 - K_2 + K_1 \\ R_{3,1}^2 - K_3 + K_1 \end{bmatrix} \right\} \quad (10)$$

where, we have the following notations:

$$K_1 = X_1^2 + Y_1^2;\ K_2 = X_2^2 + Y_2^2;\ K_3 = X_3^2 + Y_3^2$$

ISBN: 978-988-19251-1-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

hen (10) is substituted into (6), with $i = 1$, a quadratic equation in terms of R_1 is produced. Substituting the positive root back into (10) results in the final solution. Therefore, we can find the location of the abnormal node which is $A(x; y)$

IV. RESULT

In the experiment we have considered the temperature measurement field; the sensors are randomly deployed in the field and assumed that in the field the temperature would be 8 to 14. At the beginning the test was done with 10 sensors where node 4 has different data and secondly with 20 sensors in which node 13 has different data to detect the false message in the MATLAB environment. The result shows in Figure 4 and Figure 5.

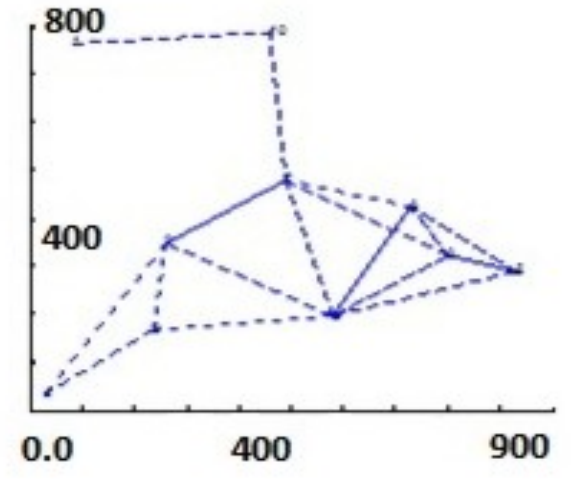

Figure 4: 10 nodes deployed in the sensor field

The Figures are clearly showing that it can detect the false massage is detected efficiently with the output massage neighbor node number to determine the neighbor Euclidian distance used.

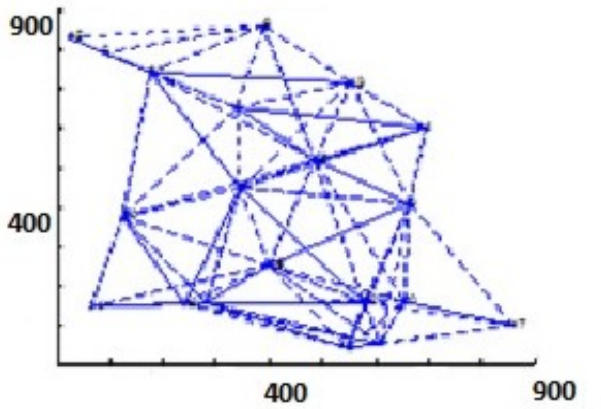

Figure 5: 20 node deployed in the sensor field

When we determined with the false massage we can use the TDOA process to get the location by using equation (10), which is discussed in section 3.

V. CONCLUSION

In this paper we have presented a novel framework to identify and locate the insider attacker that behave abnormally in the network in the wireless sensor network by using false massage detection and Time Difference of Arrival (TDoA) method, the most common practice in wireless communication to detect the location. Due to the simplicity of the process this method may be useful for the small scale deployment.

This process in for the stationary sensors in future we will implement the process for mobile scenario.

REFERENCES

[1] Y. Zhou, Y. Fang, and Y. Zhang, "Securing wireless sensor networks: a survey," *IEEE Communications Surveys & Tutorials*, 3rd Quarter 2008.

[2] W. T. Zhu, Y. Xiang, J. Zhou, R. H. Deng, and F. Bao, "Secure localization with attack detection in wireless sensor networks," *International Journal of Information Security*, vol. 10, no. 3, pp. 155-171, 2011.

[3] A. Srinivasan, and J. Wu, "A Survey on Secure Localization in Wireless Sensor Networks," *Encyclopedia of wireless and mobile communications*, 2008.

[4] L. Lazos, and R. Poovendran, "SeRLoc: Secure range independent localization for wireless sensor networks," in *ACM workshop on Wireless security (ACM WiSe '04)*, Philadelphia, 2004.

[5] L. Lazos and R. Poovendran, "HiRLoc: High-Resolution Robust Localization for Wireless Sensor Networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 2, February 2006.

[6] D. Liu, P. Ning, and W. Du, "Attack-Resistant Location Estimation in Sensor Networks," in *Proc. of The Fourth International Conference on Information Processing in Sensor Networks (IPSN '05*), 2005, pp. 99-106.

[7] S. Capkun and J.-P. Hubaux, "Secure positioning of wireless devices with application to sensor networks," in *Proc. of IEEE INFOCOM '05*, 2005.

[8] Y. Zhang, W. Yang, K. Kim, and M. Park, "Inside attacker detection in Hierarchical Wireless Sensor Networks," in *Proc. of the 3rd International conference on innovative computing information and control (ICICIC)*, 2008.

[9] C. Haiguang, C, XinHua, and N. Junyu, "Implicit Security Authentication Scheme in Wireless Sensor Networks," in *Proc. of 2010 International Conference on Multimedia Information Networking and Security*, 2010.

[10] Y. Chraibi, "Localization in wireless sensor networks," Masters' degree project submitted to KTH signal and sensor systems, Stockholm, Sweden (2005)

[11] X. Xiaochun, R. Nageswara, and S. Sartaj, "A computational geometry method for DTOA triangulation," in *Proc, of 10th International Conference on Information Fusion*, 2007, pp. 1–7.

[12] W. H. Foy, "Position-Location Solutions by Taylor-Series Estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-12, pp. 187-194, March 1976.

[13] D. J. Torrieri, "Statistical Theory of Passive Location Systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-20, no. 2, pp. 183-198, March 1984.

[14] B. Friedlander, "A Passive Localization Algorithm and Its Accuracy Analysis," *IEEE Journal of Oceanic Engineering*, vol. OE-12, no. 1, pp. 234-244, January 1987.

[15] H. C. Schau, and A. Z. Robinson, "Passive Source Localization Employing Intersecting Spherical Surfaces from Time-of-Arrival Differences," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-35, no. 8, pp. 1223-1225, August 1987.

[16] J. O. Smith, and J. S. Abel, "The Spherical Interpolation Method for Source Localization," *IEEE Journal of Oceanic Engineering*, vol. OE-12, no. 1, pp. 246-252, January 1987.

[17] J. S. Abel and J. O. Smith, "The Spherical Interpolation Method for Closed-Form Passive Localization Using Range Difference Measurements," in *Proc. ICASSP-87*, Dallas, TX, 1987, pp. 471-474.

[18] Y. T. Chan and K. C. Ho, \A Simple and Efficient Estimator for Hyperbolic Location," *IEEE Transactions on Signal Processing*, vol. 42, no. 8, pp. 1905-1915, August 1994.

[19] X, Huang, M, Ahmed, D, Sharma, "The node became compromised when an attacker gain control of the node that acts as a legitimate

ISBN: 978-988-19251-1-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

node, after network deployment", 2011 Ninth IEEE/IFIP International Conference on Embedded and Ubiquitous Computing. Oct 2011 Melbourne, Australia.

[20] X, Huang, M, Ahmed, D, Sharma, "Timing Control for Protecting from Internal Attacks in Wireless Sensor Networks", Accepted for ICOIN 2012 conference to be held in Bali Indonesia Feb (1-3), 2012.

[21] M, R, Ahmed, X, Huang, D, Sharma, "A Novel Framework for Insider Attacker Identification and Detection for Wireless Sensor Networks", Accepted for ICIS 2012 conference to be held in Kuala Lumpur, Malaysia Feb (19-21), 2012

[22] K. Sentz, "Combination of Evidence in Dempester-Shafer Theory", System Science and Engineering Department, Binghamton University, SAND 2002-0835, April 2002.

[23] H. Wu, M. Siegel, R. Stiefelhagen, J. Yang, "Sensor Fusion Using Dempster-Shafer Theory", IEEE Instrumentation and Measurement Technology Conference Anchorage, AK, USA, 21-23 May 2002

ISBN: 978-988-19251-1-4
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)