J Shanghai Univ (Engl Ed) (2008) 12(1): 66–70
Digital Object Identifier(DOI): 10.1007/s 11741-008-0113-2

TDOA-based Sybil attack detection scheme for wireless sensor networks

WEN Mi (温　蜜), LI Hui (李　辉), ZHENG Yan-fei (郑燕飞), CHEN Ke-fei (陈克非)
Lab of Cryptography and Information Security, Shanghai Jiaotong University, Shanghai 200240, P. R. China

Abstract As wireless sensor networks (WSN) are deployed in fire monitoring, object tracking applications, security emerges as a central requirement. A case that Sybil node illegitimately reports messages to the master node with multiple non-existent identities (ID) will cause harmful effects on decision-making or resource allocation in these applications. In this paper, we present an efficient and lightweight solution for Sybil attack detection based on the time difference of arrival (TDOA) between the source node and beacon nodes. This solution can detect the existence of Sybil attacks, and locate the Sybil nodes. We demonstrate efficiency of the solution through experiments. The experiments show that this solution can detect all Sybil attack cases without missing.

Keywords attack detection, Sybil attack, time difference of arrival (TDOA), wireless sensor networks (WSN).

1 Introduction

Wireless sensor networks (WSN) have recently emerged as an important application resulting from the fusion of wireless communications and embedded computing technologies. It has been wildly applied in many fields including monitoring, location, tracking and foresting. However, the nature of wireless sensor network makes them vulnerable to security attacks. Especially, without a trusted centralized authority, the Sybil attack is always possible. The Sybil attack introduced in [1] denotes an attack that the Sybil node tries to forge multiple identifications to broadcast messages in a certain region. Broadcasting messages with multiple identifications can be extremely harmful to many important functions of the sensor network such as voting, fair resource allocation, group based decisions, routing, data aggregation, and misbehavior detection.

A number of protocols for Sybil attack prevention have been proposed in recent years. But most of them are too costly for the resource-poor sensors. Douceur[1] proposes a resource testing method. It assumes that each physical entity is limited in some resource. The verifier tests whether identities correspond to different physical entities by verifying that each identity has as much of the tested resource as a physical device. It is unsuitable for wireless sensor networks because the attacker may use a physical device with several orders of magnitude more resources than a resource-starving sensor node. Karlof, *et al.*[2] used a Needham-Schroeder like protocol to verify each other's identity and establish a shared key. Consequently, it can limit the number of neighbors a node allowed to have and send an error message when a node exceeds it. But this method just limits the capability of the Sybil attack and cannot locate the Sybil node and remove it. Newsome, *et al.*[3] adopts key validation for random key pre-distribution and registration. However, they consume precious memory space as every node is required to store pair-wise keys with neighbors. Bazzi[4] prevents Sybil attacks *via* geometric distinctness certification, which tests that amongst a group of identities a large enough subset resides on a set of distinct entities. It is too complex and energy consumptive. Demirbas, *et al.*[5] presents a scheme based on the received signal strength indicator (RSSI) readings of messages to detect the Sybil attack. This is the one most close to ours. Zhang, *et al.*[6] proposes a suite of location-based compromise-tolerant security mechanisms based on a new cryptographic concept called pairing. To our knowledge, pairing is energy-consuming and it is not suitable for the sensor networks.

The major contribution of this paper is that, it proposes a time difference of arrival (TDOA) based solution to Sybil attack detection and demonstrates its efficiency by experiments. This solution can not only detect the existence of Sybil attacks but also locate the Sybil nodes. It requires minimal storage and communication overhead for sensors, as they are listened by three beacon

Received Jul.17, 2006; Revised Oct.16, 2006
Project supported by the Specialized Research Foundation for the Doctoral Program of Higher Education (Grant No.20050248043)
Corresponding author WEN Mi, PhD Candidate, E-mail:superwm@sjtu.edu.cn

nodes in each cluster, which are assumed to know their own locations (*e.g.*, through GPS receivers or manual configuration). It also does not burden the WSN with shared keys or piggy backing of keys to messages. The essential point of the TDOA-based solution is to associate the TDOA ratio with the sender's identity (ID). Once the same TDOA ratio with different ID is received, the receiver knows there is a Sybil attack. To use TDOA ratio instead of TDOA to associate the ID is to avoid the sensors at the circle centered at one of the beacon nodes being misdiagnosed.

This paper is organized as follows. In Section 2 we discuss the network model and methodology. In Section 3 we present our Sybil attack detection schemes. In Section 4 we discuss the experiments of our solution. In Section 5 we analyze the performance of our solution. Finally, in Section 6 we give our conclusion and propose the future work.

2 Network model and methodology

2.1 Network model

We assume a static network, where all nodes are deployed randomly over a 2-dimensional monitored area (it can be easily expand to 3 dimensions). If the nodes are deployed too dense, on one hand, the position of more than one node may be located at the same place. This location error may influence the detection of the Sybil attack. On the other hand, in a large scale network deployed outdoor, it is expensive to deploy nodes too densely. So, we assume that the density of the sensor network is beyond 10 m (10 m is the position accuracy of MTS420C, which is the MICA2 GPS sensor board of crossbow). We assume there is time synchronization between the source nodes and the beacon nodes. Three beacon nodes S_1, S_2, S_3 with known coordinates (X_1, Y_1), (X_2, Y_2), and (X_3, Y_3), respectively, are placed at the boundary of the monitored area (usually a cluster), as shown in Fig.1. Let (x, y) be the Sybil node's location, which will be determined by time-based positioning schemes[7,8]. Each node can reach all beacon nodes in the cluster. Note that Sybil node can forge non-existent multiple identities.

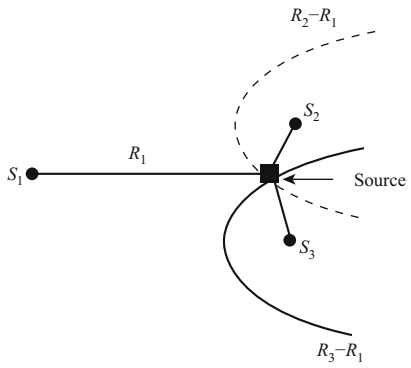

Fig.1 Hyperbolic position location (redrawn from [8])

2.2 Time difference of arrival principle

The TDOA of a message can be estimated by the hyperbolic position location solution (HPL)[9]. Assume that S_1 is the master beacon node. The distance between the source and the ith beacon node is

$$R_i = \sqrt{(X_i - x)^2 + (Y_i - y)^2}. \tag{1}$$

Now, the distance difference between beacon nodes with respect to S_1 is given as

$$R_{i,1} = cd_{i,1} = R_i - R_1, \tag{2}$$

where c is the signal propagation speed, $R_{i,1}$ is the range distance difference between S_1 and $S_{i(i>1)}$, and $d_{i,1}$ is the estimated TDOA between S_1 and $S_{i(i>1)}$. This defines the set of nonlinear hyperbolic equations whose solution gives the 2-D coordinates of the source. From (2) we know that

$$R_i = R_{i,1} + R_1. \tag{3}$$

Subtracting (1) at $i = 1$ from (3) results in

$$R_{i,1}^2 + 2R_iR_1 = X_i^2 + Y_i^2 - 2X_{i,1}x - 2Y_{i,1}y - X_1^2 - Y_1^2, \tag{4}$$

where $X_{i,1}$ and $Y_{i,1}$ are equal to $X_i - X_1$ and $Y_i - Y_1$ respectively. Without loss of generality, we assume that the beacon node S_1 is located at (0, 0). From (2) we obtain

$$R_1^2 = x^2 + y^2. \tag{5}$$

For a three base station system, Chan's method[9] producing two TDOA to render solution for x and y in terms of R_1 is in the following form:

$$\begin{bmatrix} x \\ y \end{bmatrix} = -\begin{bmatrix} X_{2,1} & Y_{1,2} \\ X_{3,1} & Y_{1,3} \end{bmatrix}^{-1} \left(\begin{bmatrix} R_{2,1} \\ R_{3,1} \end{bmatrix} R_1 + \frac{1}{2} \begin{bmatrix} R_{2,1}^2 - K_2 + K_1 \\ R_{3,1}^2 - K_3 + K_1 \end{bmatrix} \right), \tag{6}$$

where

$$K_1 = X_1^2 + Y_1^2, \quad K_2 = X_2^2 + Y_2^2, \quad K_3 = X_3^2 + Y_3^2,$$
$$R_{2,1} = cd_{2,1}, \quad R_{3,1} = cd_{3,1}.$$

On the right side of the above equation, all quantities are known except R_1. Therefore solution of x and y will be determined by R_1. When these values of x and y are substituted into (5), a quadratic equation in terms of R_1 is produced. Once the roots of R_1 are known, values of x and y can be determined.

3 TDOA-based Sybil node detection

Here we first present a basic TDOA-based Sybil attack detection protocol in Section 3.1, and in Section 3.2, we propose an advanced one by considering the environment errors.

3.1 Basic algorithm

We are going to use the localization algorithm in Section 2.2 to detect the Sybil attack as follows. Once hearing a message $m_i = \{\text{data}, D_x\}$ from source S, the three beacon nodes record its arriving time respectively, for example, t_1, t_2, t_3 at S_1, S_2, S_3. The master beacon node S_1 can compute the time difference of arrival when receive t_2 and t_3 from S_2 and S_3 derive the location of the source using (2) and (6). Then S_1 associates this location with the source-ID included in the message. Later, when another message with a different source-ID is received and the location of the source is computed to be the same as the previous one, the beacon nodes detect a Sybil attack.

But, it is costly and very inconvenient to calculate the location of every node at every communication session using (6). In fact, we do not need this computation for Sybil node detection because it is possible to detect Sybil attack by just recording and comparing the ratio of TDOA for the received messages. Only after the Sybil attack is found can we use (6) to locate the Sybil one.

Suppose that a Sybil node forge its ID as $D_1, D_2, \cdots, D_x$ and so on. Considering at session 1, a Sybil node broadcasts message $m_1 = \{\text{data}, D_1\}$ with D_1. When beacon nodes hear the message from source node, they transmit their own ID and the arriving time of message m_1 as $\text{report}_1 = \{S_{i(i>1)}, D_1, t_1\}$ to S_1. S_1 will use

$$d_{i,1}^{D_1} = t_i - t_1 = (R_i - R_1)/c \tag{7}$$

to denote the TDOA value between $S_{i(i>1)}$ and itself. Then, S_1 computes the ratio

$$\text{tr}_1 = d_{2,1}^{D_1}/d_{3,1}^{D_1}, \tag{8}$$

and stores it locally.

Similarly, at session 2, the source node broadcasts another message $m_2 = \{\text{data}, D_2\}$ with D_2. When beacon nodes hear the message, they also transmit their own ID and the arriving time of message m_2 as $\text{report}_2 = \{S_{i(i>1)}, D_2, t_i\}$ to S_1. S_1 computes the ratio

$$\text{tr}_2 = d_{2,1}^{D_2}/d_{3,1}^{D_2}. \tag{9}$$

Now, S_1 can test by comparing the ratio at session 1 and session 2.

Definition 1 (Sybilly) A Sybilly is defined as a phenomenon in which the TDOA ratios in two session i and j are same, *i.e.*, $|\text{tr}_i - \text{tr}_j| = 0$.

If difference between two ratios is very close to zero, the beacon node S_1 concludes that the Sybil attacking is happening in the cluster and the Sybil one is at the location of (x, y), which can be calculated using the method of Section 2.2. The received TDOA ratio is the same, meaning that the sources' locations are the same. In other words, the same source node broadcasts messages with multiple IDs. That is a typical behavior of the Sybil attack. On the contrary, S_1 can tell there is no Sybil node. That means, here if

$$|\text{tr}_1 - \text{tr}_2| = 0 \tag{10}$$

is true, we can detect a Sybil attack.

3.2 Advanced one for error tolerance

Theoretically, TDOA should stay the same if the locations of the two transceivers are fixed, but in practice, there are three major sources of errors[7] for our time based location detection scheme: the receiver system delay, the wireless multipath fading channel, and the nonline-of-sight transmission. These factors make the measurement of $R_{2,1}$ and $R_{3,1}$ fluctuate a lot and correspondingly they will cause fluctuation of the TDOA ratio. So we should consider the variance of the TDOA ratio and quantify the range of it.

For simplicity and error tolerance, we let the threshold used to detect a Sybil node be α, and apply the algorithm described in Section 3.1. We can detect Sybil attacks robustly using the following formula.

Definition 2 (α-Sybilly) A α-Sybilly is defined as a phenomenon in which the TDOA ratios in two sessions i and j are only different within the probability of α, *i.e.*, $|\text{tr}_i - \text{tr}_j| < \alpha$.

If the source nodes' identities D_1 and D_2 are different but their location is adjacent, we can infer the Sybil attack by noticing that the difference of TDOA ratio for both cases is within the threshold α. That means, if

$$|\text{tr}_i - \text{tr}_j| < \alpha \tag{11}$$

is true, we can detect Sybil attack.

In the next section we will quantify the range of the TDOA ratio variance and evaluate the effect of different factors on α by experiments.

4 Experiments

Since the basic scheme only has theoretic value, in this section, we just base our experiments on an advanced scheme. From (2), (6) and (7), we can see that to determine the TDOA from source S, $R_{2,1}$ and $R_{3,1}$ are the main factors. The inaccuracies of $R_{2,1}$ and $R_{3,1}$ will cause TDOA errors. We assume that calculation of $R_{2,1}$ and $R_{3,1}$ will be repeated for enough times to average their errors. Also certain measures will be taken

to reduce the effect of the three major sources of errors. For example, as [7] mentioned, beacon nodes can be placed well above the surrounding objects to avoid the multi-path fading and nonline-of-sight transmission, and the system delay can be predetermined to calibrate the time measurements. Therefore, the errors of $R_{2,1}$ and $R_{3,1}$ are approximately normally distributed. So the measuring errors of TDOA are approximately normally distributed. Without loss of generality we assume that errors of $R_{2,1}$ and $R_{3,1}$ are distributed according to $N(0, \sigma_1^2)$ and $N(0, \sigma_2^2)$ respectively.

The beacon node we used in our simulations is MTS420CA, which is the MICA2 GPS sensor board of crossbow. The source node is MDA300-CA, which is the MICA2 data acquisition board of crossbow. In our simulation, we let the three beacon nodes locate at $S_1(0, 0)$, $S_2(0, 2000)$ and $S_3(2000, 0)$. We write the detection program according to the proposed method stated in Sections 2.2 and 3.2. The TDOA ratio was computed from (7). For each trial, we try our method 10000 times.

4.1 Dynamic source

In this scenario, we study distribution of TDOA ratio errors over a 2D planar monitored area. We let the source node move according to the diagonal of the rectangle, determined by the coordinates of the three beacon nodes and the point (2000, 2000). We compare the difference of the TDOA ratio. Now the source is placed at the point $(100i, 100i)$, where $i = 1, 2, 3, \cdots, 20$. We let the measuring errors distribute according to $N(0, \sigma^2)$. For easy comparison, we set $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Here we let $\sigma^2 = 0.05$ corresponding to the TDOA deviation of ± 0.22 unit[7]. In the simulation, we observe that with the distance from the source to three beacon nodes augmenting, the variance of TDOA ratio becomes larger and larger. Also when the source is close to any of the three beacon nodes, the error becomes larger. Especially, when the source arrives at (1000, 1000), the intersection of the perpendicular bisector of isosceles right-angled triangle $\Delta S_1 S_2 S_3$ variance of TDOA ratio becomes the smallest. This is because the distance from the source to the three beacon nodes is the same at this point. The measuring error influence is almost the same. This result is similar to that of [7]. Fig.2 shows the results.

From this analysis, we can conclude that variance of TDOA ratio relates to the position of the beacon nodes and the source sensor. This result can help us do better deployment of beacon nodes for better performance in other studies.

4.2 Static source

Next we consider the scenario when sensors are randomly deployed in a square region with lower-left corner (0, 0) and upper-right corner (2000, 2000). In this case, the error model is a normal distribution $N(0, \sigma^2)$. We consider the influence of σ^2 on TDOA ratio variance. The source is randomly placed in the cluster. Once it is placed, its location is fixed. We let σ^2 equal to 0.10 and 0.05 respectively.

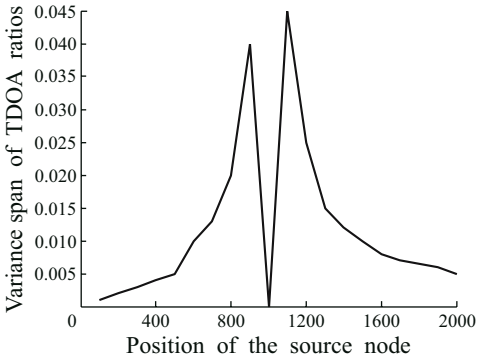

Fig.2 Variance span of TDOA ratio with the changing of the source location

From the simulation, we can make the following observations. First, Figs.3, 4 show that the TDOA ratio variance increases with variance σ^2. This is rational as σ^2 corresponds to the error of $R_{2,1}$ and $R_{3,1}$. In fact, if σ^2 increases, σ_1^2 and σ_2^2 will increase. From (7) and (8), TDOA variance will increase and TDOA ratio difference will increase correspondingly.

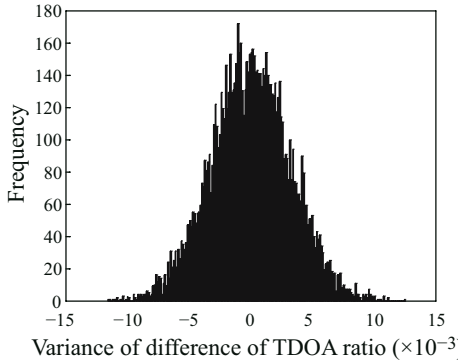

Fig.3 Measurement errors are normally distributed ($\sigma_1^2 = \sigma_2^2 = 0.1$).

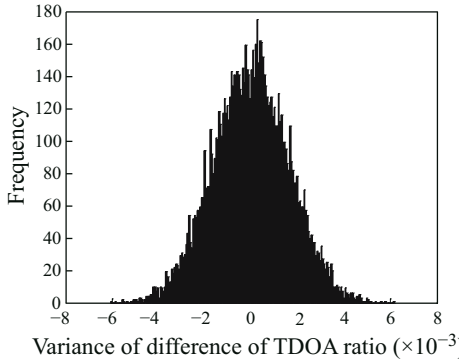

Fig.4 The measurement errors are normally distributed ($\sigma_1^2 = \sigma_2^2 = 0.05$).

In Fig.3, the values -0.010 and 0.010 occurred only at most 5 times out of 10000 times (0.05 %). It is similar to those of -0.006 and 0.006 in Fig.4. Therefore, if we let the threshold used to detect a Sybil node be $\alpha = k\sigma^2, (k \geqslant 1)$, and apply (11) to do the detection, we will succeed. The reason is that the standard deviation covers almost 100 % of the values in Figs.3, 4, that is, if we set $\alpha = 0.1$, it can detect the Sybil node with probability of success approaching 100 %. This result is also shown in Fig.2. This indicates that the solution is efficient.

5 Performance analysis

In this section, we show that the given solution is also a lightweight one in comparison to the random key predistribution in [3]. For simplicity, we let the length of the messages and the keys be L bytes and each transmission cost E energy unit. K is the number of the nodes in the network. In Section 3, we have shown that, except for the normal message transmission in each session, the solution only needs two more transmissions for beacon nodes S_2 and S_3 to report the message's source-ID and arriving time to S_1. Only S_1 needs to record the TDOA ratio (also let TDOA ratio be L bytes long) in each session. So the total overhead to detect a Sybil one is $K(2E + L)$.

By comparison, in [3], in order to detect the Sybil attack it needs a key pool in the network, with the size of m keys. Each node needs to store n keys, m and n being related to the size of the network (in [3], $m = 20000$ and $n = 200$). The node identity is associated with the keys assigned to the node. The network needs to verify part or all of the keys that an identity claims to have. So, during each verification section, the node must claim what keys it has, this claim may be $knE\,(0 < k \leqslant 1)$. The total overhead of [3] to detect if a node is the Sybil one is $KknE + (m + n)L$.

From the above discussion, we can see that the solution given in this paper requires less storage and lower energy consumption. Therefore it is a lightweight solution.

6 Conclusions and future work

In this paper, we have proposed an efficient and lightweight TDOA based Sybil attack detection scheme. It includes a basic scheme and an advanced scheme to tolerate the measuring errors. We demonstrate efficiency of the solution by experiments. In the future work we will study another factor that can affect the effectiveness of the advanced scheme, namely accuracy of the TDOA computation model. If this model cannot accurately model the actual deployment, there will be extra errors (both on false positive and detection rate) in the Sybilly detection. We will extend the scheme to tolerate existing nodes in the network. Also we will try the other location algorithms such as AOA, *etc.*

References

[1] Douceur J R. The Sybil attack [C]//*Proceedings of the First International Workshop on Peer-to-Peer Systems* (*IPTPS'* 02), Cambridge, MA, USA. 2002, **2429**: 251–260.

[2] Karlof C, Wagner D. Secure routing in wireless sensor networks: attacks and countermeasures [C]//*Proceedings of the* 1*st IEEE International Workshop on Sensor Network Protocols and Applications*, Anchorage, AK, USA. 2003.

[3] Newsome J, Shi E, Song D, Perring A. The Sybil attack in sensor networks: analysis & defenses [C]//*Proceedings of the Third International Symposium on Information Processing in Sensor Networks* (*IPSN'*04), Berkeley, CA, USA. 2004: 259–268.

[4] Bazzi R A, Goran K. On the establishment of distinct identities in overlay networks [C]//*ACM Symposium on Principles of Distributed Computing*, Las Vegas, NV, USA. 2005: 312–320.

[5] Demirbas M, Song Y W. An RSSI-based scheme for sybil attack detection in wireless sensor networks [C]//*International Workshop on Wireless Mobile Multimedia* (*WOWMOM'*06), New York, USA. 2006: 564–570.

[6] Zhang Y C, Liu W, Lou W J, Fang Y G. Location-based compromise-tolerant security mechanisms for wireless sensor networks [J]. *IEEE Journal on Selected Areas in Communications*, 2006, **24**(2): 247–260.

[7] Cheng X, Thaeler A, Xue G, Chen D. TPS: a time-based positioning scheme for outdoor wireless sensor networks [C]//23*rd Annual Joint Conference of the IEEE Computer and Communications Societies*, (*INFOCOM'04*), San Francisco, CA, USA. 2004, **4**: 2685–2696.

[8] Reza R I. Data Fusion for Improved TOA/TDOA Position Determination in Wireless Systems [D]. Virginia Polytechnic Institute and State University, July 2000.

[9] Chan Y T, Ho K C. A simple and efficient estimator for hyperbolic location [J]. *IEEE Transactions on Signal Processing*, 1994, **42**(8): 1905–1915.

(Editor HONG Ou)