Computer Communications 31 (2008) 4324–4332

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Group-based intrusion detection system in wireless sensor networks

Guorui Li [a,*], Jingsha He [b], Yingfang Fu [a]

[a] College of Computer Science and Technology, Beijing University of Technology, Beijing 100124, China
[b] School of Software Engineering, Beijing University of Technology, Beijing 100124, China

ARTICLE INFO

Article history:
Available online 2 July 2008

Keywords:
Wireless sensor networks
Security
Intrusion detection scheme
False alarm
Detection accuracy

ABSTRACT

Many mission critical wireless sensor networks require an efficient, lightweight and flexible intrusion detection algorithm to identify malicious attackers. In this paper, we propose a distributed group-based intrusion detection scheme that meets all the above requirements by partitioning the sensor networks into many groups in which the sensors in each group are physically close to each other and are equipped with the same sensing capability. Our intrusion detection algorithm takes simultaneously into consideration of multiple attributes of the sensor nodes to detect malicious attackers precisely. We show through experiments with real data that our algorithm can decrease the false alarm rate and increase the detection accuracy compared with existing intrusion detection schemes while lowering the computation and transmission power consumption.

1. Introduction

Wireless sensor networks have become one of the most interesting and promising research and development areas over the past few years. Such networks usually consist of hundreds or even thousands of small-sized, low power, inexpensive sensors to monitor some specific phenomenon cooperatively. The characteristics of sensor networks such as flexibility, self-organization, fault tolerance, high sensing fidelity, low cost and rapid deployment have created many new and exciting applications such as wildlife monitoring, disaster response, military surveillance, smart building and industrial quality control, to name a few [1].

In general, the sensors in a network are deployed in unattended environment or even hostile circumstance, and communicate with each other using wireless signals which can be eavesdropped very easily. The constrained capacity of wireless sensor nodes such as limitation in computation power, memory and battery lifetime further increases the insecurity of wireless sensor networks. Many different kinds of attacks against wireless sensor networks have been identified so far, e.g., bogus routing and sensed data attack, select forward attack, sink hole attack, worm hole attack, black hole attack and hello flood attack, etc. [2].

All the security solutions proposed so far can be classified into two main categories: prevention based techniques and detection based techniques. Prevention based techniques, such as encryption and authentication, are often regarded as the first line of defense against attacks. Detection based techniques are designed to identify and isolate attackers after prevention based techniques fail. Furthermore, there are two types of detection based techniques: signature based detection and anomaly based detection. Signature based detection techniques match the known attack profiles with suspicious behaviors whereas anomaly based detection techniques detect unusual deviations from pre-established normal profiles to identify the abnormal behaviors.

In this paper, we propose a new group-based intrusion detection scheme which is based on anomaly detection technique. In our scheme, we first partition the sensor nodes in a network into a number of groups using δ-grouping algorithm such that the nodes in a group are physically close to each other and their sensed data are dissimilar by at most δ. And then our intrusion detection algorithm is scheduled to run for each group. The monitor sensor node supervises multidimensional measurement promiscuously in each group in turn to average the power consumption among the group members. If an obvious deviation between monitored values is found using intrusion detection algorithm, an alarm will be issued and the corresponding outlier should be identified and segregated from the sensor network. Through experiments in which we use data released from the Intel Berkeley Research Lab, we show that our scheme can achieve a lower false alarm rate than that in the present intrusion detection schemes while consuming less power.

The two notable features of our scheme are: (1) the intrusion detection algorithm is executed in a number of groups in which the sensors in each group are physically close to each other and sense the similar observation. This feature makes it easier to detect outlier nodes and the intrusion detection results become more precise; and (2) we adopt the Mahalanobis distance measurement and the OGK estimators in the intrusion detection algorithm to take

* Corresponding author. Tel.: +86 10 67396607; fax: +86 10 67396061.
E-mail address: liguorui@emails.bjut.edu.cn (G. Li).

0140-3664/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.comcom.2008.06.020

into account the inter-attribute dependencies of multidimensional observed values and ensure a high breakdown point even with some missing data at a lower computational cost. The idea of partitioning the sensor network into many groups with the similar observation to detect the outlier nodes is first introduced in this paper and is shown to be effective in increasing the detection accuracy and decreasing the false alarm rate.

This paper is organized as follows. In the next section, we review some related work. In Section 3, we present our group-based intrusion detection scheme. In Section 4, we analyze our scheme and present some simulation results. Finally, we draw our conclusions in Section 5.

2. Related work

In this section, we review some related work in the security of wireless sensor networks in which we classify them into two categories: prevention based techniques and detection based techniques.

2.1. Prevention based techniques

Encryption and authentication are two primary techniques to secure wireless sensor networks against malicious access. The core ideas behind such techniques rely on key management. Eschenauer and Gligor proposed the basic probabilistic key predistribution scheme in which each sensor is assigned a random subset of keys from a key pool before the network is deployed so that any two sensor nodes will have a certain probability to share at least one key [3]. Chan et al. improved the above scheme and proposed the q-composite key predistribution scheme [4], which requires that two sensor nodes share at least q predistributed keys as the basis for the establishment of a pairwise key between the two nodes. Liu and Ning proposed a framework in which pairwise keys are predistributed by using bivariate polynomials [5]. They also proposed two efficient instantiations, i.e., a grid-based key predistribution scheme and a random subset assignment scheme, for the establishment of pairwise keys in a wireless sensor network. In addition, they proposed the closest pairwise key predistribution scheme and the closest polynomials predistribution scheme, which take advantage of sensor nodes' expected locations to predistribute appropriate keys to the sensors and thus can improve the performance of key establishment [6]. Li et al. proposed the hexagon-based key predistribution scheme that can improve the effectiveness of key management in sensor network by using the bivariate polynomial in a hexagonal coordinate system based on the deployment information about expected locations of the sensor nodes [7]. All the key management schemes mentioned above belong to the type of static key management schemes.

Another type of key management scheme is the dynamic key management scheme in which keys may be updated periodically or on demand as a response to node capture. By performing key update, the compromised nodes are segregated and the security of sensor networks can be enhanced. Moharrum and Eltoweissy compared dynamic key management with static key management and, as the result, proposed an EBS (exclusion basis system)-based dynamic key management scheme [8]. Eltoweissy et al. proposed a dynamic key management scheme called LOCK (LOcalized Combinatorial Keying) which is an EBS-based hierarchical key management scheme that can only be used in hierarchical wireless sensor networks [9]. Similar dynamic key management schemes also include SHELL [10] and identity-based symmetric keying [11], both relying on a central key server to perform rekeying. Li et al. proposed the group-based dynamic key management scheme that can solve the same problem without the need for such infrastructure as base station and cluster heads [12].

2.2. Detection based techniques

Doumit et al. proposed a self-organized criticality and stochastic learning based intrusion detection scheme that takes advantage of self-organized criticality for a certain location based on an environment variable and uses a hidden Markov model to detect future anomalies [13]. Su et al. proposed eHIP – energy efficient hybrid intrusion prohibition system to improve the security of cluster-based sensor networks [14]. Such a system consists of authentication-based intrusion prevention (AIP) subsystem and collaboration-based intrusion detection (CID) subsystem to provide heterogeneous mechanisms that can meet the different demands of security levels in cluster-based wireless sensor networks to improve energy efficiency. Agah et al. proposed a non-cooperative game approach in which the key is to find the most vulnerable node in a sensor network and protect it [15]. Silva et al. defined multiple rules that can be used to determine if a failure has happened and to raise an intrusion alarm if the number of failures exceeds a predefined threshold [16]. A newly proposed scheme, called the insider attacker detection scheme, takes into consideration of multiple attributes simultaneously in node behavior evaluation without the requirement for prior knowledge about normal or malicious sensor activities [17]. It has high accuracy and low false alarm rate when some sensor nodes are misbehaving. We will compare our scheme with it in Section 4.

3. The group-based intrusion detection scheme in wireless sensor networks

The group-based intrusion detection scheme involves two phases: grouping the sensor networks and running the group-based intrusion detection algorithm in each group. We describe the assumptions in the proposed scheme first.

3.1. Assumptions

The group-based intrusion detection scheme can run in a flat homogeneous wireless sensor network. There is no need to include special nodes, such as cluster heads. We assume that there are no malicious nodes at the beginning of the sensor network deployment and all the sensors start the intrusion detection algorithm at the same time. Neither are there intense and unexpected varieties of sensed data at the grouping phase of our intrusion detection algorithm.

3.2. δ-Grouping algorithm

First, we partition the sensor nodes in the network into a set of groups with each group sensing a similar phenomenon. The sensors within the same group are physically close to each other and their sensed data are dissimilar by at most δ. This feature insures that the intrusion detection results are more precise than other schemes. According to [18], the δ-grouping problem is NP-complete and, therefore, it is impossible to derive an approximation to the optimal solution in polynomial time unless $P = NP$.

3.2.1. The original δ-grouping algorithm

The notations used in the original δ-grouping algorithm are described in Table 1.

The original δ-grouping algorithm in sensor node i works as follows:

Table 1
The notations used in the original δ-grouping algorithm

Notation	Meaning
grouped	A boolean variable (initially false) to indicate whether a sensor node has joined a group
r_i	The root of group to which node i belongs
p_i	Parent of node i in a group
f_i	Sensed value(s) at node i
f_{r_i}	Sensed value(s) at root r_i
$d(f_i, f_{r_j})$	Euclidean distance between f_i and f_{r_j}
$hops(i, r_j)$	Hops between node i and root r_j
h	Predefined maximum hops within a group
$T_{random}(i)$	Random time for node i to initialize a grouping message

Algorithm 1. Original δ-grouping algorithm

```
If (Not grouped) And (receive < "Grouping", r_j, f_{r_j} >) Then
    If (d(f_i, f_{r_j}) ⩽ δ/2) And (hops(i,r_j) ⩽ h/2) Then
        r_i:=r_j;
        grouped:=True;
        p_i:=j;
        f_{r_i} := f_{r_j};
        broadcast < "Grouping", r_i, f_{r_i} >;
    Endif
Else If (Not grouped) And (T_random(i) is up) Then
        r_i:=i;
        grouped:=True;
        p_i:=i;
        f_{r_i} := f_i;
        broadcast < "Grouping", r_i, f_{r_i} >;
    Endif
Endif
```

Each sensor i waits a random time $T_{random}(i)$ to initiate a grouping message. If it receives a grouping message from its neighbor within this time period and has not joined any group, it will calculate the Euclidean distance between its sensed data f_i and received sensed data f_{r_j} of group root r_j, as well the number of hops between them. If $d(f_i, f_{r_j}) \leqslant \delta/2$ and $hops(i, r_j) \leqslant h/2$, the Triangle Inequality ensures that the metric distance between any two sensor nodes in the group is at most δ and is h hops away from each other at the most. Therefore, the sensor nodes that are in the same group have the similar observation and are not far away from each other spatially. If this sensor has not joined any group after $T_{random}(i)$ time, it will form a new group with itself as the root and broadcast grouping message to its neighbors.

3.2.2. The refined δ-grouping algorithm

The original δ-grouping algorithm determines the root of the group randomly. It may be the node that holds the extreme value in the network. As a result, the number of sensor nodes that are selected to be included in this group is very small and, therefore, these nodes can no longer be included in any other appropriate groups anymore. This could lead to a situation in which a lot of small groups are formed, a situation that is not very suitable for running the group-based intrusion detection scheme and would consume more energy. So we improve the original δ-grouping algorithm by proposing a refined δ-grouping algorithm. The additional notations used in the refined grouping algorithm are described in Table 2.

Table 2
Additional notations used in the refined δ-grouping algorithm

Notation	Meaning
$f_{r_j\_\max}$	The current maximal sensed value(s) in the group
$f_{r_j\_\min}$	The current minimal sensed value(s) in the group
$f_{r_j\_avg}$	The current average sensed value(s) in the group
$f_{r_j\_count}$	The current count of sensors in the group

Algorithm 2. Refined δ-grouping algorithm

```
If (Notgrouped) And (receive < "Grouping",
   r_j, f_{r_j_max}, f_{r_j_min}, f_{r_j_avg}, f_{r_j_count} >)
Then
    f'_{r_j_avg} = (f_{r_j_avg} * f_{r_j_count} + f_i)/(f_{r_j_count} + 1)
    f'_{r_j_max} = max(f_{r_j_max}, f_i)
    f'_{r_j_min} = min(f_{r_j_min}, f_i)
    If d(f'_{r_j_max}, f'_{r_j_avg}) ⩽ δ/2) And (d(f'_{r_j_min}, f'_{r_j_avg}) ⩽ δ/2)
       And (hops(i,r_j) ⩽ h/2) Then
       r_i:=r_j;
       grouped:=True;
       p_i:=j;
       broadcast < "Grouping", r_i, f'_{r_j_max}, f'_{r_j_min}, f'_{r_j_avg}, r_{j_count} + 1 >;
    EndIf
Else If (Not grouped) And (T_random (i) is up) Then
        r_i:=i;
       grouped:=True;
       p_i:=i;
       f_{r_j_max} := f_{r_j_min} := f_{r_j_avg} := f_i;
       broadcast < "Grouping", r_i, r_j_max, r_j_min, r_{j_a}vg, 1 >;
    EndIf
EndIf
```

We show in the following that the refined δ-grouping algorithm presented above satisfies the condition required by any δ-grouping algorithm:

There is only one node in the group when the root of the group broadcasts a grouping message. So it is obvious that the metric distance between any two nodes is at most δ.

When a new node joins the group by following the refined δ-grouping algorithm, $d(f'_{r_j\_\max}, f'_{r_j\_avg}) \leqslant \delta/2$ and $d(f'_{r_j\_\min}, f'_{r_j\_avg}) \leqslant \delta/2$. So the metric distance between any two nodes i and j is

$$d(f_i, f_j) = |f_i - f_j| \leqslant |f_i - f'_{r_j\_avg}| + |f'_{r_j\_avg} - f_j|$$
$$\leqslant 2 \times \max(d(f'_{r_i\_\max}, f'_{r_j\_avg}), d(f'_{r_j\_\min}, f'_{r_j\_avg})) \leqslant \delta$$

Therefore, we can assure that the metric distance between any two sensor nodes in a group is at most δ and h hops far away at the most.

Note that, in both grouping algorithms, we should choose the contention time $T_{random}(i)$ wisely to make sure that the sensor node waits enough time before starting a new group. The waiting time for sensor i, $T_{random}(i)$, can be computed by choosing a random number and multiplying it by the average message transmission time t between any two neighboring nodes. After $\max(T_{random}(i)) + ht/2$, all the sensor nodes in the entire network will complete the grouping algorithm.

3.3. Group-based intrusion detection algorithm

According to the δ-grouping algorithm, the resulting groups are comprised of sensors that are spatially close to each other and have similar observations. Therefore, if some sensors could find that there is noticeable difference between their sensed data and those of some other sensors in the same group through promiscuous monitoring, they could conclude that some group members have been compromised with high probability and should be segregated from the network to maintain the security of the entire sensor network.

3.3.1. Intrusion detection scheduling algorithm

We partition the δ-group into equal-sized sub-groups according to the node attributes such as sensor ID, sensor remaining power or the hops to the root of the δ-group. Assuming that there are N_i sensor nodes in the δ-group G_i and N_{S_i} sensor nodes in its sub-group, we partition G_i into $\lfloor N_i/N_{S_i} \rfloor$ sub-groups. Every sensor in the sub-group will monitor the network simultaneously using the intrusion detection monitoring algorithm described in Section 3.3.2. These sub-groups monitor the entire δ-group in turn to reduce the total power consumption resulting from monitoring to prolong the lifetime of the entire network. If one sensor found obvious deviation between its sensed data and its monitored data, it will alert other sensors in the same δ-group using the following message:

"Alert", Charged node, Monitor node, Timestamp.

When there are more than n_1 alert messages coming from the same monitor node or charging the same abnormal node in a sensor's alert table for a period of time, it will change to the promiscuous mode and only monitor the abnormal node or the monitor node by itself. If it found any abnormal behavior taking place in the charged node or the monitor node, it will record the observed alert in its own alert table and calculate the alert value: $n_1w_1 + n_2w_2$, where w_1 and w_2 are the weight value of the alert message coming from other sensor nodes and itself, respectively, and n_2 is the number of abnormal messages it has observed. When $n_1w_1 + n_2w_2$ is above a predefined threshold θ, it will confirm that the charged node is an abnormal node or the monitor node is a malicious charging node and will segregate it from the sensor network by removing it from the route table and/or removing the pairwise key between them.

Table 3
The collected information

Collected information	Detected attack behavior
Sensor sensed data	Fabricate information attack
Packet sending rate	Energy exhausting attack
Packet dropping rate	Select forward attack, black hole attack
Packet mismatch rate	Message alter attack
Packet receiving rate	Sink hole attack
Packet sending power	Hello attack, worm hole attack

3.3.2. Intrusion detection monitoring algorithm

The monitor sensor node can collect the following informations, which are shown in Table 3, to detect the malicious network attack behavior.

We assume that the multiple monitoring attributes of sensor x_i form a multiple dimension vector $f(x_i) = \{f_1(x_i), f_2(x_i), \ldots f_q(x_i)\}$ where q is the number of the monitoring attributes. And all the monitoring vectors of monitor sensors $x_1, \ldots, x_n$ form a matrix $F(x) = \{f(x_1), f(x_2), \ldots f(x_n)\}$, where n is the number of monitor sensor nodes in a sub-group.

We assume that all $f(x_i)(x_i \in \{x_1, \ldots, x_n\})$ form a sample of a multivariate normal distribution. And $f(x_i)$ is distributed as $N_q(\mu, \sum)$, following a multi-variate normal distribution with mean vector μ and variance–covariance matrix $\sum$. The Mahalanobis squared distance $(f(x_i) - \mu)^T \sum^{-1} (f(x_i) - \mu)$ is distributed as χ_q^2, where χ_q^2 is the chi-square distribution with q degree of freedom. Therefore, the probability that $f(x_i)$ satisfies $(f(x_i) - \mu)^T \sum^{-1} (f(x_i) - \mu) \succ \chi_q^2(\alpha)$ is α, where $\chi_q^2(\alpha)$ is the upper (100α)th percentile of a chi-square distribution with q degrees of freedom.

Let $\widehat{\mu}$ and $\widehat{\sum}$ be the estimates of μ and $\sum$, respectively. Then the probability that $f(x_i)$ satisfies $(f(x_i) - \widehat{\mu})^T \widehat{\sum}^{-1} (f(x_i) - \widehat{\mu}) \succ \chi_q^2(\alpha)$ is expected to be roughly α. Let $d(x_i) = ((f(x_i) - \widehat{\mu})^T \widehat{\sum}^{-1} (f(x_i) - \widehat{\mu}))^{1/2}$. Sensor x_i will be regarded as an outlier if $d(x_i)$ or $d^2(x_i)$ is unusually large. In our scheme, sensor x_i is regarded as an abnormal node if $d^2(x_i) \succ \chi_q^2(\alpha)$.

Rather than estimating μ and $\sum$ by the simple mean and the simple variance-covariance matrix:

$$\widehat{\mu} = \frac{1}{n} \sum_{i-1}^{n} f(x_i)$$

$$\widehat{\sum} = \frac{1}{n-1} \sum_{i=1}^{n} (f(x_i) - \widehat{\mu})(f(x_i) - \widehat{\mu})^T$$

in which the values from outlying sensors can easily distort the estimates of μ and $\sum$ and the detection via Mahalanobis distances may fail to identify true abnormal sensors, we adopt the OGK (Orthogonalized Gnanadesikan–Kettenring) estimators $\widehat{\mu}$ and $\sum$ [19].

The single-variate estimates $\widehat{\mu}$ and $\hat{\sigma}^2$ based on the single-variate sample set $Y = \{y_1, y_2, \ldots, y_n\}$ are:

$$\widehat{\mu} = \frac{\sum_{i=1}^{n} y_i W(v_i)}{\sum_{i=1}^{n} W(v_i)} \text{ for } v_i = \frac{y_i - \mu_0}{\sigma_0}$$

$$\widehat{\sigma}^2 = \frac{\sigma_0^2}{n} \sum_{i=1}^{n} \rho\left(\frac{y_i - \widehat{\mu}}{\sigma_0}\right)$$

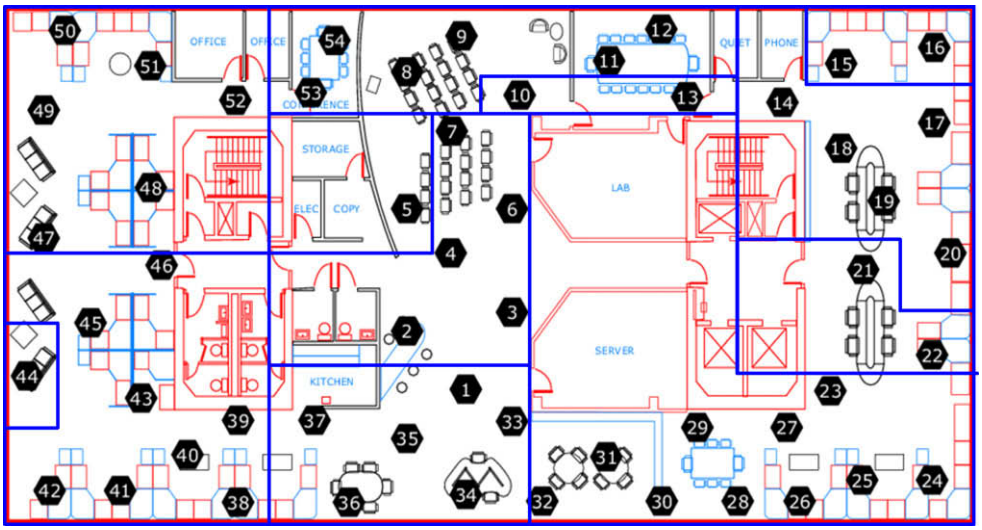

Fig. 1. An instance of the original δ-grouping algorithm.

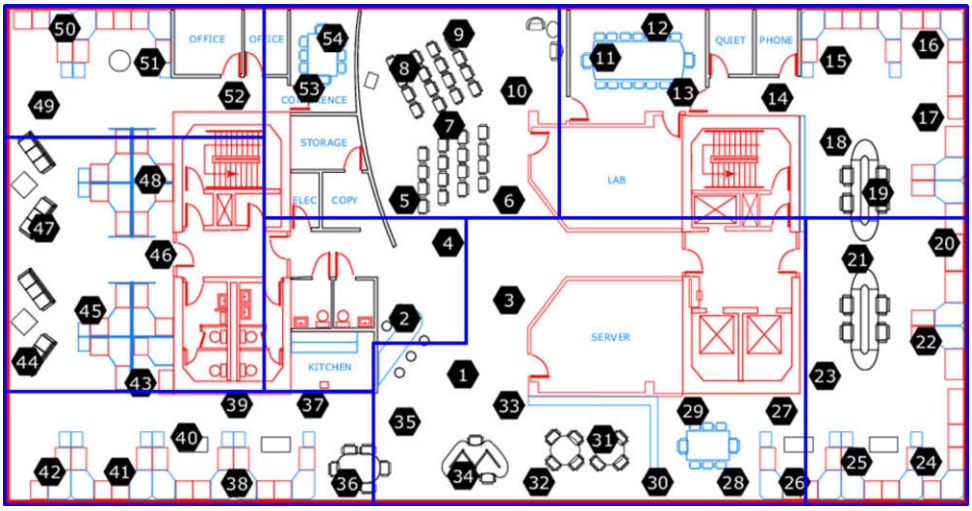

Fig. 2. An instance of the refined δ-grouping algorithm.

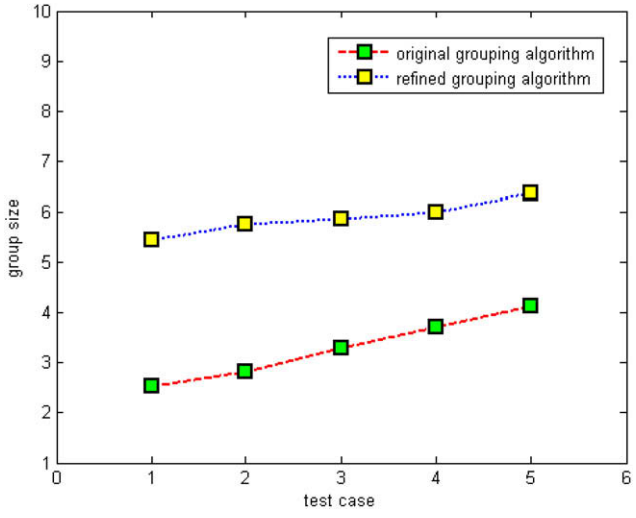

Fig. 3. The average group size of the two grouping algorithms.

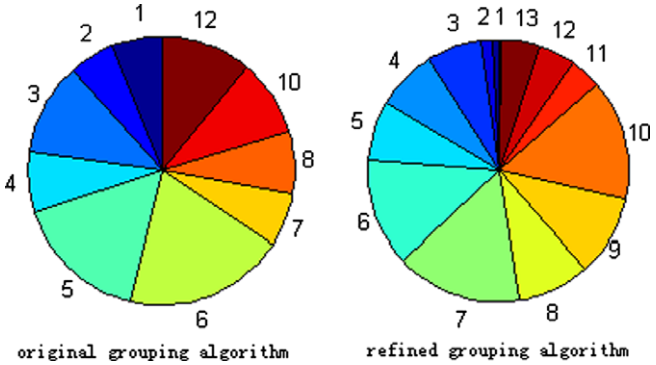

Fig. 4. The percentage of the sensor nodes in each group partitioned by the two grouping algorithms.

where μ_0 and σ_0 are the median and $MAD = median(|Y - median(Y)|)$ of Y, weight function $W(x) = (1 - (x/c_1)^2)^2 I(|x| \leqslant c_1)$ and ρ function $\rho(x) = \min(x^2, c_2^2)$ where $c_1 = 4.5$ and $c_2 = 3$.

The multivariate estimates $\widehat{\mu}$ and $\widehat{\sum}$ based on the multivariate sample set $F(x) = \{f(x_1), f(x_2), \ldots f(x_n)\}$ where $f(x_i) = (f_1(x_i), f_2(x_i), \ldots, f_q(x_i))^{\mathrm{T}}$ are computed as follows:

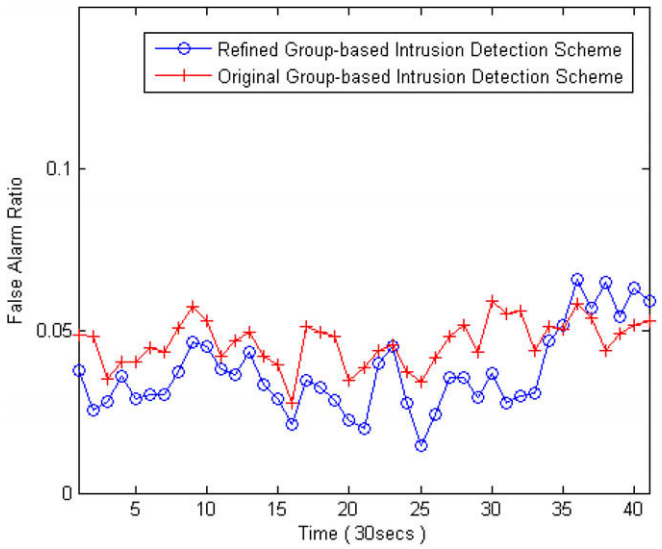

Fig. 5. False alarm ratio in the refined group-based intrusion detection scheme vs. that in the original group-based intrusion detection scheme.

(1) Compute $G(x) = \{g(x_1), g(x_2), \ldots, g(x_n)\}$ where $g(x_i) = P^{-1} f(x_i)$, $P = diag(\widehat{\sigma}(\widetilde{F}_1(x)), \widehat{\sigma}(\widetilde{F}_2(x)), \ldots, \widehat{\sigma}(\widetilde{F}_q(x)))$ and $\widetilde{F}_j(x)$ is the j row of $F(x)$.

(2) Calculate $q \times q$ matrix R where

$$R_{jk} = \begin{cases} \frac{1}{4}[\widehat{\sigma}^2(G_j + G_k) - \widehat{\sigma}^2(G_j - G_k)] & j \neq k \\ 1 & j = k \end{cases}.$$

(3) Apply the spectral decomposition to obtain $R = Q \Lambda Q^{\mathrm{T}}$ where Q is R's eigen-matrix and Λ is the diagonal matrix composed of R's eigen-values.

(4) Compute $H = \{h(x_i) | h(x_i) = Q^T g(x_i)\}$. Then calculate $\Lambda = (\widehat{\mu}(H_1(x)), \widehat{\mu}(H_2(x)), \ldots, \widehat{\mu}(H_q(x)))^{\mathrm{T}}$ and $\Gamma = diag(\hat{\sigma}^2(H_1(x)), \hat{\sigma}^2(H_2(x)), \ldots, \hat{\sigma}^2(H_q(x)))$.

(5) Let $V = PQ$. Then the robust multivariate estimates are $\widehat{\mu} = V\Lambda$ and $\widehat{\Sigma} = V\Gamma V^{\mathrm{T}}$.

We choose the Mahalanobis distance measurement because it includes the inter-attribute dependencies so we can compare the attribute combinations and get more precise results [20]. The reason why we decide to choose OGK is because it ensures a high

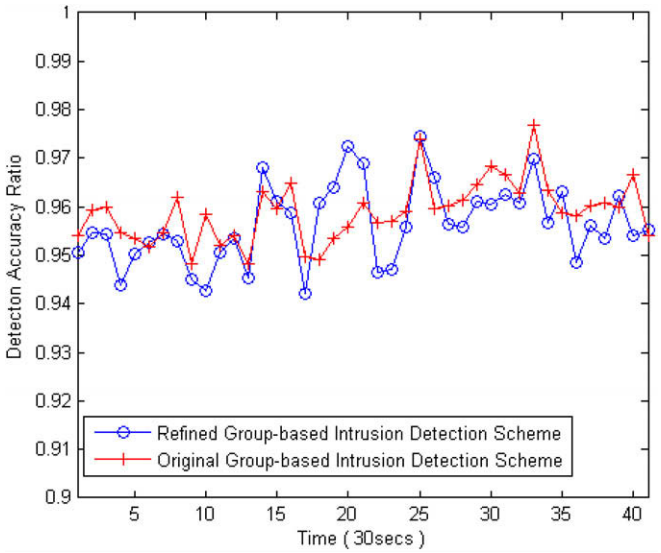

Fig. 6. The detection accuracy ratio between refined group-based intrusion detection scheme and original group-based intrusion detection scheme.

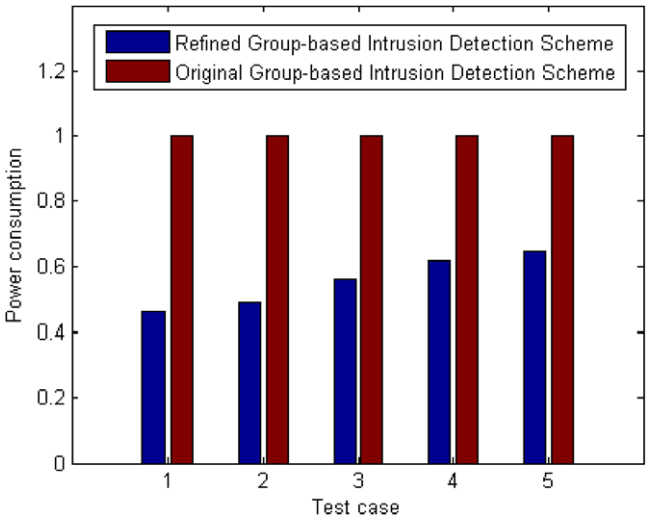

Fig. 7. Comparison of power consumption between the refined and the original group-based intrusion detection schemes.

breakdown point with some missing data and can compute quickly with a lower computational cost [17].

4. Analysis

We use the real sensed data collected from 54 Mica2Dot sensors deployed in the Intel Berkeley Research Lab between 28 February and 5 April 2004 to demonstrate the performance of our group-based intrusion detection scheme. The collected data include humidity, temperature, light and voltage values along with timestamp information collected once every 31 s. We randomly add some noise following the normal distribution to the tested data in order to simulate the abnormal sensors.

We compare our group-based intrusion detection scheme with insider attacker detection scheme proposed in [17] using R 2.6.0 statistical computing software with robustbase and rrcov packages.

4.1. Comparison between the refined and the original group-based intrusion detection schemes

Figs. 1 and 2 show two instances of the original δ-grouping algorithm and the refined δ-grouping algorithm in the Intel Berkeley Research Lab, respectively. We can see that the refined δ-grouping algorithm can partition the sensor networks more neatly, uniformly and larger than what the original δ-grouping algorithm could do. We now conduct a more thorough analysis in this section.

Fig. 3 shows the average group size of the two δ-grouping algorithms in different judgment condition sets of temperature, humidity and light. We can see that the average group size of the refined grouping algorithm is much larger than that of the original grouping algorithm. And the average group size of both algorithms is growing when the condition sets of grouping get looser. We can conclude that the refined grouping algorithm can partition the sensor networks into larger groups than the original grouping algorithm.

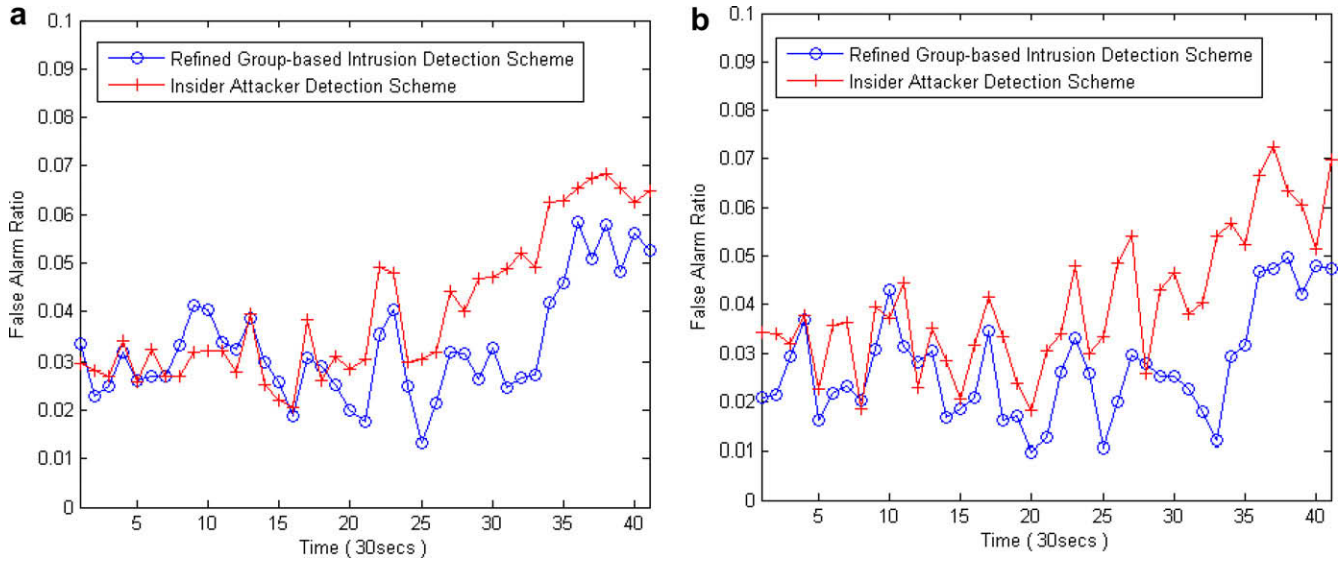

Fig. 8. Comparison of false alarm ratio between the refined group-based intrusion detection scheme and the insider attacker detection scheme with different back retrieve numbers. (a) Back retrieve number = 3 and (b) back retrieve number = 4.

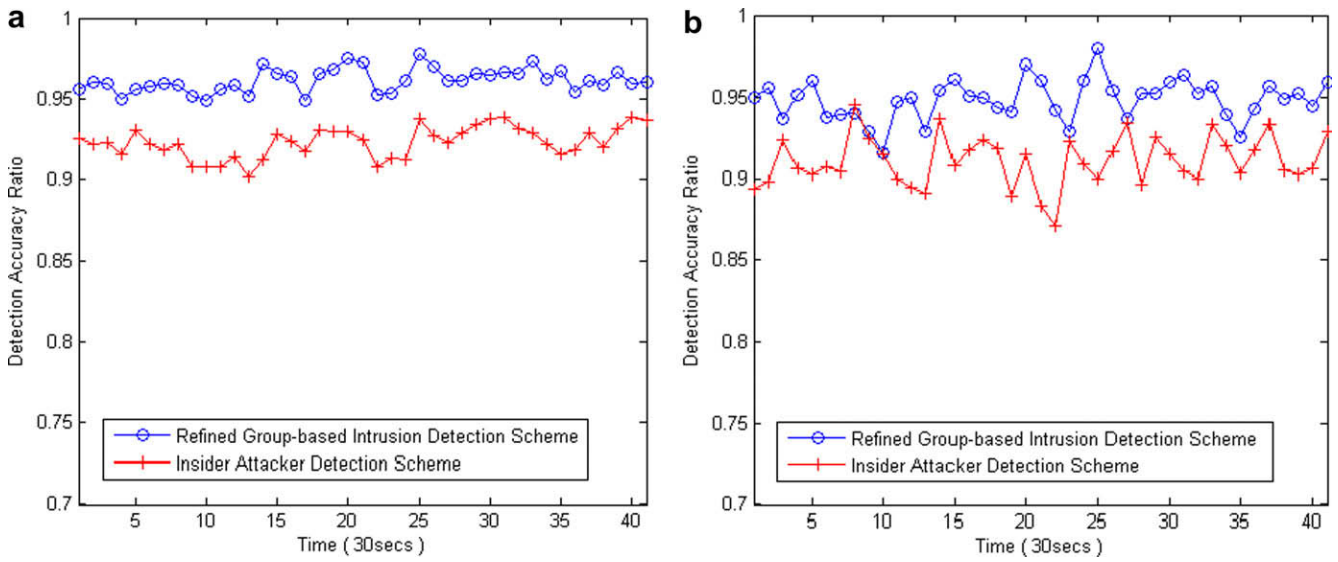

Fig. 9. Comparison of detection accuracy ratio between the refined group-based intrusion detection scheme and the insider attacker detection scheme with different back retrieve numbers. (a) Back retrieve number = 3 and (b) back retrieve number = 4.

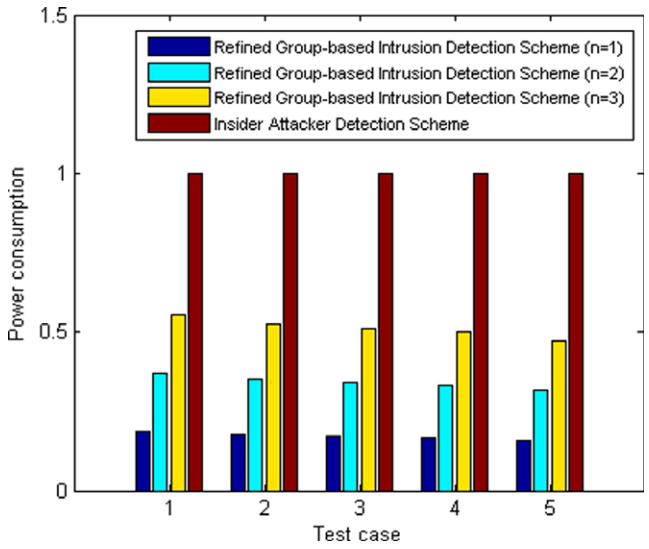

Fig. 10. Comparison of power consumption between the refined group-based intrusion detection scheme and the insider attacker detection scheme.

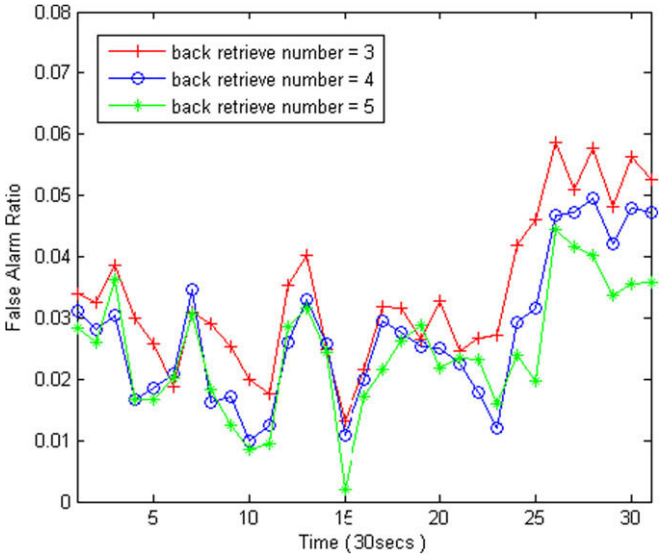

Fig. 11. False alarm ratios of the refined group-based intrusion detection scheme with different back retrieve numbers.

Fig. 4 shows the percentage of the sensor nodes in each group partitioned by the two grouping algorithms. The numbers around the pie shows the size of each group. We can see that the percentage of the small groups in the original δ-grouping algorithm is much larger than that in the refined δ-grouping algorithm. These groups do not fit in well with the group-based intrusion detection algorithm and should be as small as possible. The refined δ-grouping algorithm reduces the small groups by modifying the root of the each group in the grouping stage and can thus achieve more efficient performance than the original δ-grouping algorithm.

False alarm ratio is defined as the percentage of the normal sensors that are claimed as abnormal sensors. And detection accuracy ratio is defined as the percentage of abnormal sensors that can be successfully detected. These are the two key performance indicators for any intrusion detection scheme. We can see from Fig. 5 that the false alarm ratio in the refined group-based intrusion detection scheme is lower than that in the original group-based intrusion detection scheme and from Fig. 6 that the detection accuracy ratio of these two schemes is similar.

Moreover, the original group-based intrusion detection scheme consumes more energy than that of the refined group-based intrusion detection scheme. This is because only the nodes that belong to the monitoring sub-group will examine the entire group while other sub-group members remain in the sleep state until the schedule cycle concludes. Each sub-group takes charge of the monitoring periodically. Thus, the total monitoring power consumption of the group-based intrusion scheme is $\frac{N \cdot N_s}{N_{avg}} \Delta \cdot t$, where N, N_s and N_{avg} represent the total number of the sensor nodes, the number of monitoring nodes in each sub-group and the average group size, respectively, Δ is the monitoring power consumption of each sensor node in unit time and t is the monitoring time. Therefore, the larger the average group size, the less the total monitoring power consumption. A comparison of the total monitoring power consumption between these two schemes corresponding to Fig. 3 is shown in Fig. 7, which testifies our conclusion.

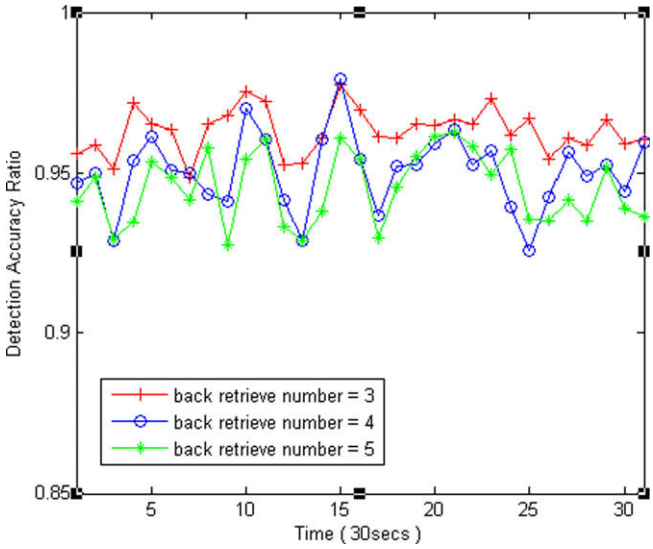

Fig. 12. Detection accuracy ratios of the refined group-based intrusion detection scheme with different back retrieve numbers.

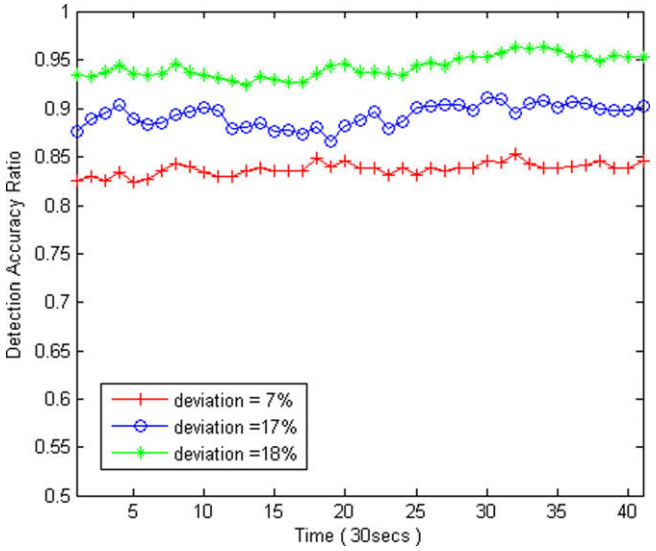

Fig. 14. Detection accuracy ratios of the refined group-based intrusion detection scheme with different outlying deviations.

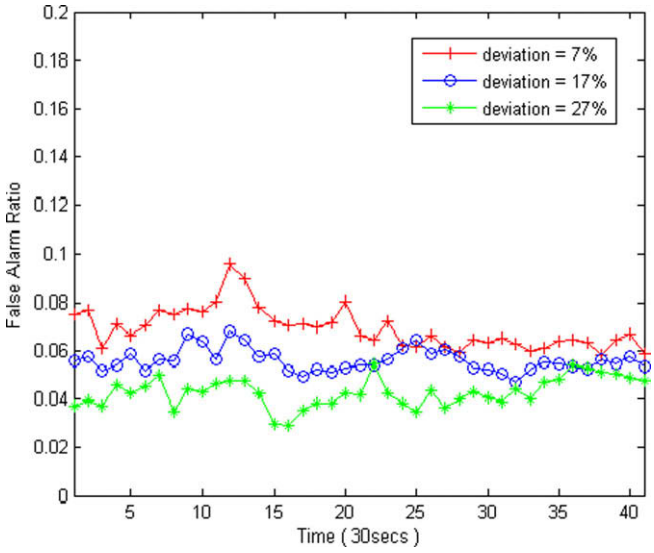

Fig. 13. False alarm ratio of the refined group-based intrusion detection scheme with different outlying deviation.

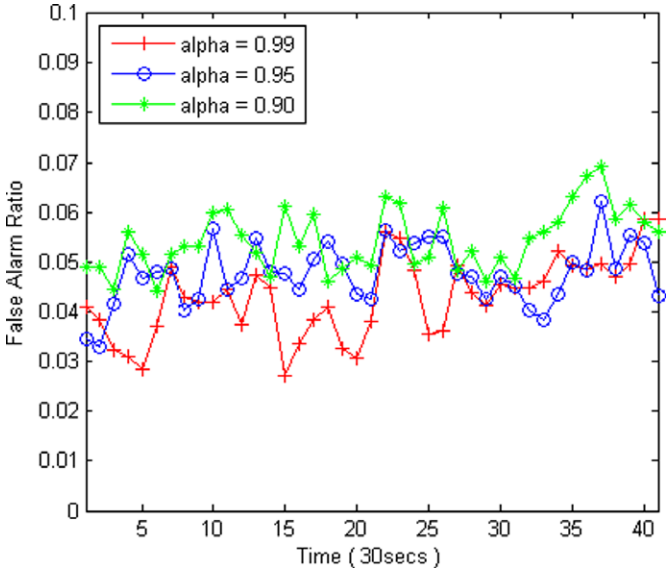

Fig. 15. False alarm ratios of the refined group-based intrusion detection scheme with different alpha-quantile of the chi-squared distribution.

4.2. Comparison between the refined group-based intrusion detection scheme and the insider attacker detection scheme

Fig. 8 shows the false alarm ratio between the refined group-based intrusion detection scheme and the insider attacker detection scheme with back retrieve number 3 and 4, respectively, where the back retrieve number is the number of history messages reserved in the monitoring nodes to facilitate the detection process. We can see that the refined group-based intrusion detection scheme produces less number of false alarm alerts than that of the insider attacker detection scheme.

Fig. 9 shows the detection accuracy ratio between the refined group-based intrusion detection scheme and the insider attacker detection scheme with back retrieve number 3 and 4, respectively. We can see that the detection accuracy of the refined group-based intrusion detection scheme is much higher than that of the insider attacker detection scheme. The decrease in false alarm ratio and increase in detection accuracy lie in the fact that the grouping algorithm eliminates the difference between sensed data within the monitoring sensor's neighborhood so that the detection results are less influenced by the variety of the observed values.

Fig. 10 is the comparison of power consumption between the refined group-based intrusion detection scheme and the insider attacker detection scheme, where n is the sub-group size. We can see that the refined group-based intrusion detection scheme consumes less power than that of the insider attacker detection scheme. The reason behind this improvement lies in the fact that the intrusion detection scheduling mechanism distributes the monitoring power consumption to each node equally and the sensor nodes within each group collaborate with each other.

4.3. Characteristics of the refined group-based intrusion detection scheme

Figs. 11 and 12 compare the false alarm ratios and the detection accuracy ratios of the refined group-based intrusion detection scheme with different back retrieve number 3, 4 and 5, respectively. We can see that both the false alarm ratio and the detection

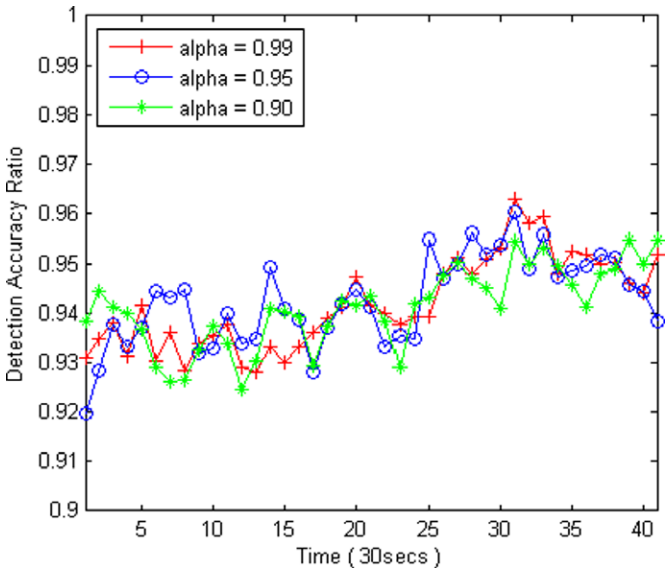

Fig. 16. Detection accuracy ratios of the refined group-based intrusion detection scheme with different alpha-quantile of the chi-squared distribution.

accuracy ratio of the scheme decrease slightly with the increase of the back retrieve number.

Figs. 13 and 14 compare the false alarm ratio and the detection accuracy ratios of the refined group-based intrusion detection scheme with outlying deviation 7%, 17% and 27%, respectively. We can see that the false alarm ratio decreases with the increase of the outlying deviation and the detection accuracy ratio increases with the increase of the outlying deviation. In another word, the larger the outlying deviation is, the better the detection performance that the scheme can achieve.

Figs. 15 and 16 compare the false alarm ratios and the detection accuracy ratios of the refined group-based intrusion detection scheme with alpha-quantile of the chi-squared distribution 0.99, 0.95 and 0.90, respectively. We can see that the false alarm ratio increases slightly with the decrease of the alpha-quantile. However, the change of the detection accuracy ratio is not obvious.

5. Conclusion

In this paper, we proposed the group-based intrusion detection scheme for wireless sensor networks with the goal of detecting malicious attackers. Our experiment results show that our scheme can achieve a lower false alarm rate and a higher detection accuracy rate than the existing detection schemes. At the same time, it can also reduce the monitoring power consumption with the requirement of grouping the sensor nodes in the network only once.

References

[1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, Wireless sensor networks: a survey, Computer Networks 38 (4) (2002) 393–422.
[2] C. Karlof, D. Wagner, Secure routing in wireless sensor networks: attacks and countermeasures, Ad Hoc Networks 1 (2) (2003) 293–315.
[3] L. Eschenauer, V.D. Gligor, A key-management scheme for distributed sensor networks, in: Proceedings of the Ninth ACM Conference on Computer and Communications Security, November 2002, pp. 41–47.
[4] H. Chan, A. Perring, D. Song, Random key predistribution schemes for sensor networks, in: Proceedings of the IEEE Symposium on Research in Security and Privacy, 2003, pp. 197–213.
[5] D. Liu, P. Ning, Establishing pairwise keys in distributed sensor networks, in: Proceedings of the 10th ACM Conference on Computer and Communications Security, October 2003, pp. 52–61.
[6] D. Liu, P. Ning, Location-based pairwise key establishments for static sensor networks, in: Proceedings of the 2003 ACM Workshop on Security in Ad Hoc and Sensor Networks, 2003, pp. 72–82.
[7] G. Li, J. He, Y. Fu, Key management in sensor networks, in: Proceedings of the International Conference on Wireless Algorithms, Systems and Applications 2006, August 2006, pp. 457–466.
[8] M. Moharrum, M. Eltoweissy, A study of static versus dynamic keying schemes in sensor networks, in: Proceedings of the Second ACM international workshop on performance evaluation of wireless Ad Hoc, sensor, and ubiquitous networks, 2005, pp. 122–129.
[9] M. Moharrum, M. Eltoweissy, R. Mukkamala, Dynamic key management in sensor networks, IEEE Communications 44 (4) (2006) 122–130.
[10] M. Younis, K. Ghumman, M. Eltoweissy, Location-aware combinatorial key management scheme for clustered sensor networks, IEEE Transaction on Parallel and Distributed System 17 (8) (2006) 865–882.
[11] G. Jolly, M.C. Kuscu, P. Kokate, M. Younis, A low-energy key management protocol for wireless sensor networks, in: Proc. of Eighth IEEE international Symposium on Computers and Communications, June, 2003, pp. 335–340.
[12] G. Li, J. He, Y. Fu, A group-based dynamic key management scheme in wireless sensor networks, in: Proceedings of the 21st International Conference on Advanced Information Networking and Applications, 2007, pp. 127–132.
[13] S. Doumit, D.P. Agrawal, Self-organized criticality & stochastic learning based intrusion detection system for wireless sensor network, in: Proceedings of the 2003 IEEE Military Communications Conference, vol. 22, no. 1, 2003, pp. 609–614.
[14] W. Su, K. Chang, Y. Kuo, eHIP: an energy-efficient hybrid intrusion prohibition system for cluster-based wireless sensor networks, Computer Networks 51 (4) (2007) 1151–1168.
[15] A. Agah, S. Das, K. Basu, M. Asadi, Intrusion detection in sensor networks: a non-cooperative game approach, in: Proceedings of the Third IEEE International Symposium on Network Computing and Applications, August 2004, pp. 343–346.
[16] A. da Silva, M. Martins, B. Rocha, A. Loureiro, L. Ruiz, H. Wong, Decentralized intrusion detection in wireless sensor networks, in: Proceedings of the First ACM international workshop on Quality of service & security in wireless and mobile networks, 2005, pp. 16–23.
[17] F. Liu, X. Cheng, D. Chen, Insider attacker detection in wireless sensor networks, in: Proceedings of the 26th IEEE International Conference on Computer Communications, 2007, pp. 1937–1945.
[18] A. Meka, A.K. Singh, Distributed Spatial Clustering in Sensor Networks, Lecture Notes in Computer Science 3896, Springer Verlag, Heidelberg, Germany, 2006, pp. 980–1000.
[19] R.A. Maronna, R.D. Martin, V.J. Yohai, Robust Statistics: Theory and Methods, Wiley Publisher, 2006.
[20] V.J. Hodge, J. Austin, A survey of outlier detection methodologies, Artificial Intelligence Review 22 (2) (2004) 85–126.