ANNALS of Faculty Engineering Hunedoara
– International Journal of Engineering
Tome XIII [2015] – Fascicule 4 [November]
ISSN: 1584-2673 [CD-Rom; online]
a free-access multidisciplinary publication of the Faculty of Engineering Hunedoara

[1.] Nima SAADATMAND

A COOPERATIVE GAME THEORY APPROACH TO IDENTIFY EFFECTIVE FEATURES OF INTRUSION DETECTION IN WIRELESS SENSOR NETWORKS

[1.] Department of Computer Engineering, Islamic Azad University, Borujerd Branch, Borujerd, IRAN

Abstract: The continuous evolution of computer networks and mobile applications has drastically changed the nature of their security and privacy. This paper analyzes the problem of intrusion detection in a Gaussian-distributed wireless sensor network by characterizing the detection probability with respect to the application requirements and the network parameters including number of deployed sensors, sensing range, deployment deviation, maximal allowable intrusion distance, and intruder's starting distance. Effects of these parameters on the detection probability are examined in detail. In recent years, various information theoretic based measurements have been proposed to identify the importance of each feature from multi-dimensional data set. The aim of this paper is to introduce a cooperative game theory based framework to evaluate the power of each wireless sensor network feature. Results showed that among considered features, sensing range and deployment deviation had the most effect on the performance of intrusion detection in a wireless sensor network. This work can be used to guide the selection of an appropriate random sensor deployment strategy and help in the design of a wireless sensor network and determining critical parameters for intrusion detection.
Keywords: Intrusion detection, game theory, network deployment, sensing range, wireless sensor network

1. INTRODUCTION

The emergence of wireless sensor networks as one of the dominant technology trends in the coming decades has posed numerous unique challenges to researchers [1]. As networks play an increasingly important role in modern society, we witness the emergence of new types of security and privacy problems that involve direct participation of network agents. These agents are individuals, as well as devices or software, acting on their own behalf [2]. Consequently, there is a fundamental relationship between the decision making of agents and network security problems. Due to recent technological advances in wireless communication, manufacturing of small- and low-cost sensors has become economically feasible [3]. A large number of sensors can be deployed in an ad hoc fashion to form a wireless sensor network for many civil and military applications [4]. Intrusion detection has received a great deal of attention since it supports various applications such as environmental monitoring and military surveillance.

Recent studies on the intrusion detection problem fall into two major categories. First, it is considered as a system component for monitoring the security of a wireless sensor network and diagnosing compromised/vulnerable sensors to ensure the correct network behavior and avoid false alarm [5]. On the other hand, it is defined as monitoring or surveillance system for detecting a malicious intruder that invades the network domain [6]. This work focuses on the second category. In some references [2], examples are demonstrated in which a number of sensors are deployed in a circular area for protecting the central located target by sensing and detecting the presence of a moving intruder. Intrusion detection implies how effectively an intruder can be detected by the wireless sensor network. Obviously, sooner the intruder can be detected, better is the intrusion detection capability of the wireless sensor network [7].

Full sensing coverage means immediate intrusion detection. However, full sensing coverage demands for a large number of sensors and can be hardly feasible in an actual practice. Therefore, most intrusion detection applications do not have such a strict requirement of immediate detection [2]. Instead, a maximum allowable intrusion distance (D_m) is specified. Suppose the intruder moves a distance of D in the wireless sensor network before it is detected. If $D<D_m$, the wireless sensor network meets the performance requirements. Otherwise, the wireless sensor network needs to be reconfigured. Apparently, intrusion distance is a central issue in an intrusion detection application using a wireless sensor network [7].

Feature selection, also known as variable selection, is one of the fundamental problems in the fields of machine learning, pattern recognition and statistics [8]. For most feature selection algorithms based on information theory, feature (or subset) that has high relevance with the class and low redundancy among selected features will be selected in each iteration [9]. The major disadvantage

of these algorithms is that they disregard the dependencies between the candidate feature and unselected features. Consequently, interdependent features, weak as individuals but having strong discriminatory power as a group, are likely to be ignored. Several researchers also constructed examples to illuminate that some variables which are useless by themselves can be useful together. The main reason for this disadvantage is that Information-theoretic based measurements disregard the intrinsic structure among features [10].

To untie this knot, a practicable method is needed to retain the useful intrinsic structures for feature selection. The solution of cooperative games represents the contribution of each feature as a player to the game by constructing a value function, which assigns a value to each player. Banzhaf power index was proposed by Banzhaf, which yields a unique outcome in coalitional games, to measure the contribution of players in the game. It is based on counting, for each player, the number of coalitions to which the player is crucial to winning [11].

The main contributions of this work include develop an analytical model for intrusion detection in a Gaussian-distributed wireless sensor network, and investigate the interplays between the network parameters and the intrusion detection capability of the network, and validate theoretical derivations and results by probabilistic simulations. The role of each network parameter in intrusion detection is investigated.

The rest of this paper is organized as follows. Section 2 introduces the system model and definitions. Section 3 presents the proposed algorithm to identify effective network parameter in intrusion detection using cooperative game theory. Section 4 illustrates and explains the results of proposed method in intrusion detection. Finally, the paper is concluded in Section 5.

2. SYSTEM MODEL AND DEFINITIONS

We consider a wireless sensor network with randomly deployed N sensors around a target point (i.e., the central red star) following a 2D Gaussian distribution. The region of interest A is assumed to be a square area with side length L. Without loss of generality, we assume the coordinate of the target point as $G=(0,0)$ and the same standard deviation (i.e. $\sigma_x = \sigma_y = \sigma$) along X and Y dimensions in the deployment field [12]. It is possible to imagine that different deviations lead to different sensor distribution [2]. Furthermore, the closer the location is to the center, the higher is the probability of deploying sensors there. When the standard deviation is increased to some extent, some sensors may be deployed outside the region of interest A. If all sensors ought to be deployed inside A, a Gaussian distribution can be used. When σ increases toward infinity, the truncated Gaussian distribution tends toward a uniform distribution.

All sensors are assumed to be equipped with the same sensing range r_s, and their sensing coverage is assumed to be circular and symmetrical following a Boolean sensing model [13]. In a wireless sensor network, there are two ways to detect an intruder: single-sensing detection and multiple-sensing detection [7]. In single-sensing detection, the intruder can be successfully detected by a single sensor when entering its sensing range. On the other hand, in the multiple-sensing detection model, an intruder has to be sensed by at least m sensors and m depends on a specific application [14]. These sensors need not sense the intruder simultaneously in the considered model.

The intruder is assumed to be aware of its target (i.e., the hot spot), and follows the shortest intrusion path D toward the target. The straight-line intrusion path model was adopted in [15]. It is due to the fact that abstractions and assumptions are inevitable to conduct theoretical analysis [16] and make influencing factors tractable. Further, we assume that the intruder can enter the wireless sensor network from an arbitrary point with distance R to the target. The corresponding intrusion detection region S_D is indirectly determined by the sensor's sensing range r_s and intrusion distance D.

It is important to observe that in a single-sensing detection at least one sensor should be located in the region S_D for detecting the intruder. Similarly, in multiple-sensing detection at least m sensors should reside in the region S_D for recognizing the intruder.

In order to evaluate the performance of intrusion detection in a Gaussian-distributed wireless sensor network, we use Intrusion distance and detection probability metrics [17]. Intrusion distance D is the distance that the intruder travels before it is detected by a wireless sensor network for the first time. Specifically, it is the distance between the point where the intruder enters the wireless sensor network and the point where the intruder gets detected by any sensor(s). Detection probability $P[D \leq D_m]$ is defined as the probability that an intruder is detected within the maximal allowable intrusion distance D_m, specified by a wireless sensor network application. To be specific, if the intruder moves less than or equal to D_m, i.e., $D \leq D_m$ before it is detected, the intrusion detection of the wireless sensor network is regarded as a successful case. Otherwise, the intrusion detection is considered as a failed one when $D > D_m$. The detection probability for single-sensing and multiple-sensing detection scenarios are derived in [2].

3. COOPERATIVE GAME THEORY APPROACH

The existence of intrinsic correlative structures among variables results in different importance of every individual. Our contributions focus on evaluating the importance (or power) of each feature using the Banzhaf power index. The original definition of Banzhaf power index is described as follows [11]: A winning coalition is one for which $v(S)=1$ and a losing coalition is one for which $v(S)=0$.

Each coalition $SU\{i\}$ that wins when S loses is called a swing for player i, because the membership of player i in the coalition is crucial to the coalition winning. Let $\sigma_i(N,v)$ be the number of swings for i, and let $\sigma_o(N,v)$ be the total number of swings of all players in the game [18]. Then, the normalized Banzhaf index is $b_i(N,v)= \sigma_i(N,v)/ \sigma_o(N,v)$ [11]. Calculation of Banzhaf power index is presented in [8] in detail. The idea is motivated by the observation that every subset of features can be regarded as a candidate subset for the final selected optimal subset, thus, the power of each feature can be measured by averaging the contributions that it makes to each of the subset which it belongs to. Details of the feature evaluation framework based on cooperative game theory are presented in [8]. The output of this evaluation framework is a vector P_v of which each element $P_v(i)$ represents the normalized Banzhaf power index of feature f_i. It is noticed that Banzhaf power index is only a metric estimating the importance of every feature based on the intrinsic correlative structures among features. Thus, to select features based on our evaluation framework, a metric reflecting the feature's relevance to target class and a heuristic search strategy are also needed. By calculating normalized Banzhaf power index for number of deployed sensors, sensing range, deployment deviation, maximal allowable intrusion distance, and intruder's starting distance, the role of each parameter can be obtained. To define target classes for investigation of the performance of proposed method, 10 classes are considered for intrusion detection probability index including 1 to 10 for probability values from 0 to 1 with 0.1 steps.

Table 1. Levels of network input parameters in creating the dataset

Parameter	Levels
Number of deployed sensors (10 levels)	50, 100, 150, 200, 250, 300, 350, 400, 450, 500
Sensing range (15 levels)	2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30
Deployment deviation (10 levels)	10, 20, 30, 40, 50, 60, 70, 80, 90, 100
Maximal allowable intrusion distance (10 levels)	5, 10, 15, 20, 25, 30, 35, 40, 45, 50
Intruder's starting distance (10 levels)	20, 40, 60, 80, 100, 120, 140, 160, 180, 200

Table 2. Levels of network output parameter in creating the dataset

Parameter	Levels
Intrusion detection probability index (10 levels)	1, 2, 3, 4, 5, 6, 7, 8, 9, 10

Tables 1 and 2 show the levels of network parameter in creating the dataset. Based on the derivations in [2], we theoretically examine the effect of network parameters on the intrusion detection probability under both single-sensing detection and multiple-sensing detection cases in a Gaussian-distributed wireless sensor network using MATLAB. Therefore, 10×15×10×10×10 runs are performed for generating the simulation results and obtained intrusion detection probability index from each run classified in 10 classes between 1 and 10.

4. RESULTS AND SIMULATION VERIFICATION

We employed four representative classifiers, i.e., Naïve Bayes, SVM, 1-Nearest Neighbor and C4.5, which are the most influential algorithms that have been widely used in the data mining community. The experimental workbench is Weka (Waikato environment for knowledge analysis), which is a collection of machine learning algorithms for data mining tasks. The parameters of classifiers for each experiment are set to default values of Weka. For estimating the performance of classification algorithms, 10-fold cross-validation is used. In 10-fold cross-validation the data is first partitioned into10 nearly equally sized folds. Subsequently, ten iterations of training and validation are performed such that within each iteration, a different fold of the data is held-out for validation while the remaining nine folds are used for learning.

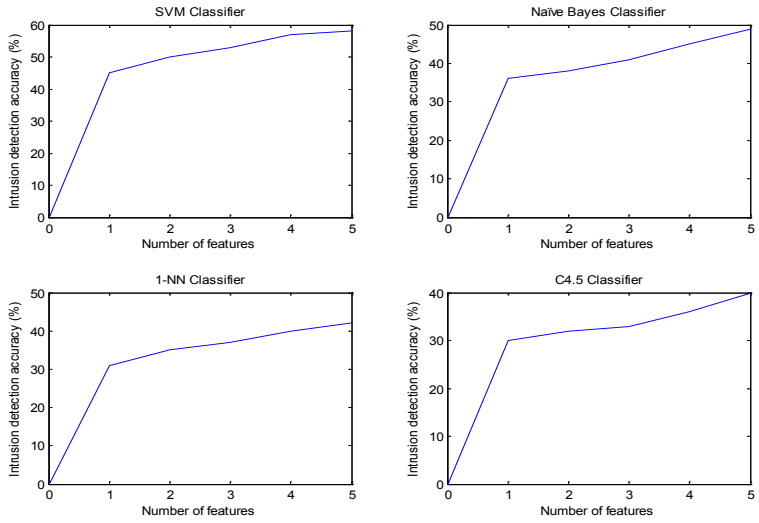

Figure 1. Intrusion detection accuracies vs. different numbers of selected features using four different classifiers for single-sensing detection

According to the obtained results, the effective network parameters in order to having the biggest Banzhaf index to the smallest are: deployment deviation σ, sensing range r_s, number of deployed sensors N, maximal allowable intrusion distance D_m, and intruder's starting distance R. Figures 1 and 2 show the classification accuracies on the obtained datasets for four classifiers using proposed algorithm for single-sensing and multiple-sensing detections, respectively. In these figures, the priority in number of features belongs to parameters with higher Banzhaf indexes. According to these results, the intrusion detection accuracy was obtained more than 85% by having the values of deployment deviation σ and sensing range r_s parameters in multiple-sensing detection. Results showed that SVM classifier had more appropriate performance in intrusion detection in comparison with other methods.

Advantages of filter selectors are that they are fast and easy to interpret. There are also some disadvantages of using filters, such as (i) redundant features may be included and (ii) some features which as a group have strong discriminatory power but are weak as individual features will be ignored. To cope with these problems, a new method for feature evaluation and selection has been proposed in this paper. As is known to all, it is difficult to discover the association relationship among features exactly. We cannot guarantee that our method retains all useful interdependent groups or the whole interdependent group; however, the method suggested an effective way to retain useful interdependent features and groups as many as possible.

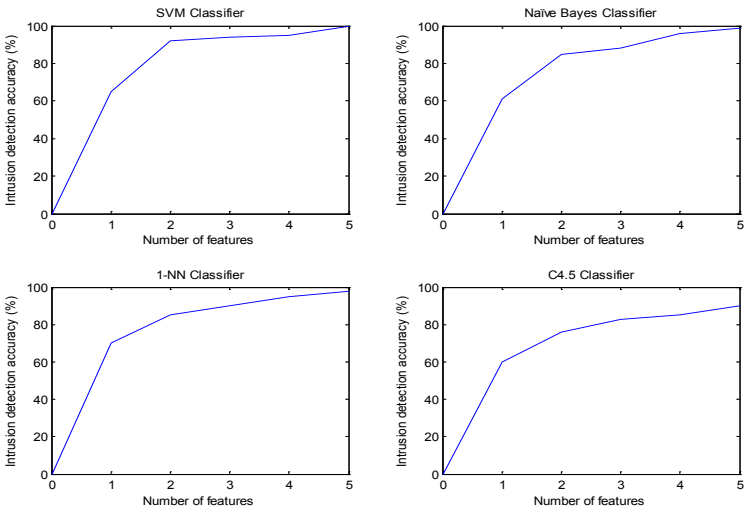

Figure 2. Intrusion detection accuracies vs. different numbers of selected features using four different classifiers for multiple-sensing detection

5. CONCLUSION

In this study, we have presented an overview of security and privacy problems that are addressed and analyzed within a game-theoretic framework. The aim of this paper was to introduce a cooperative game theory approach to evaluate the power of each wireless sensor network feature. Results showed that among considered features, sensing range and deployment deviation had the most effect on the performance of intrusion detection in a wireless sensor network. By having the values of these two parameters in multiple-sensing detection, the intrusion detection accuracy was obtained more than 85%. This work can be used to guide the selection of an appropriate random sensor deployment strategy and help in the design of a WSN and determining critical parameters for intrusion detection.

References

[1.] Manshaei, M. H.; Zhu, Q.; Alpcan, T.; Bacşar, T. and Hubaux, J. P.: Game theory meets network security and privacy, ACM Computing Surveys (CSUR), 45(3), 25, 2013.
[2.] Wang, Y.; Fu, W. and Agrawal, D. P.: Gaussian versus uniform distribution for intrusion detection in wireless sensor networks. Parallel and Distributed Systems, IEEE Transactions on, 24(2), 342-355, 2013.
[3.] Sohraby, K.; Minoli, D. and Znati, T.: Wireless Sensor Networks: Technology, Protocols, and Applications. John Wiley and Sons, Inc., 2007.
[4.] Tilak, S.; Abu-Ghazaleh, N. B. and Heinzelman, W.: A taxonomy of wireless micro-sensor network models. ACM SIGMOBILE Mobile Computing and Communications Review, 6(2), 28-36, 2002.
[5.] Giruka, V. C.; Singhal, M.; Royalty, J. and Varanasi, S.: Security in wireless sensor networks. Wireless communications and mobile computing, 8(1), 1, 2008.
[6.] Arora, A.; Dutta, P.; Bapat, S.; Kulathumani, V.; Zhang, H.; Naik, V.; ... and Miyashita, M.: A line in the sand: a wireless sensor network for target detection, classification, and tracking. Computer Networks, 46(5), 605-634, 2004.
[7.] Wang, Y.; Wang, X.; Xie, B.; Wang, D. and Agrawal, D. P.: Intrusion detection in homogeneous and heterogeneous wireless sensor networks.Mobile Computing, IEEE Transactions on, 7(6), 698-711, 2008.
[8.] Sun, X.; Liu, Y.; Li, J.; Zhu, J.; Chen, H. and Liu, X.: Feature evaluation and selection with cooperative game theory. *Pattern recognition*, *45*(8), 2992-3002, 2012.
[9.] Guyon, I. and Elisseeff, A.: An introduction to variable and feature selection. Journal of Machine Learning Research, 3, 1157-1182, 2003.
[10.] Liu, H. and Motoda, H. (Eds.).: Computational methods of feature selection, CRC Press, 2007.
[11.] Banzhaf III, J. F.: Weighted voting doesn't work: A mathematical analysis. Rutgers Letters Review, 19, 317, 1964.
[12.] Wang, D.; Xie, B. and Agrawal, D. P.: Coverage and lifetime optimization of wireless sensor networks with gaussian distribution. Mobile Computing, IEEE Transactions on, 7(12), 1444-1458, 2008.
[13.] Hsin, C. F. and Liu, M.: Network coverage using low duty-cycled sensors: random & coordinated sleep algorithms. In Proceedings of the 3rd international symposium on Information processing in sensor networks (pp. 433-442). ACM, 2004.
[14.] Banerjee, S.; Grosan, C.; Abraham, A. and Mahanti, P. K.: Intrusion detection on sensor networks using emotional ants. International Journal of Applied Science and Computations, 12(3), 152-173, 2005.
[15.] Lazos, L.; Poovendran, R. and Ritcey, J. A.: Probabilistic detection of mobile targets in heterogeneous sensor networks. In Proceedings of the 6th international conference on Information processing in sensor networks (pp. 519-528). ACM, 2007.
[16.] Bai, X.; Yun, Z.; Xuan, D.; Jia, W. and Zhao, W.: Pattern Mutation in Wireless Sensor Deployment, Proc. IEEE INFOCOM, 1-9, 2010.
[17.] Zhang, Y.; Meratnia, N. and Havinga, P.: Outlier detection techniques for wireless sensor networks: A survey. Communications Surveys & Tutorials, IEEE, 12(2), 159-170, 2010.
[18.] González-Díaz, J.; García-Jurado, I. and Fiestras-Janeiro, M. G.: An introductory course on mathematical game theory (Vol. 115). Providence: American Mathematical Society, 2010.