Available online at www.sciencedirect.com

SciVerse ScienceDirect

Physics Procedia 25 (2012) 2072 – 2080

Physics Procedia

2012 International Conference on Solid State Devices and Materials Science

Intrusion Detection for Wireless Sensor Network Based on Traffic Prediction Model

Han Zhijie, Wang Ruchuang

College of Computer
Nanjing University of Posts and Telecommunications
Nanjing, China

Abstract

In this paper, the authors first propose an efficient traffic prediction algorithm for sensor nodes which exploits the Markov model. Based on this algorithm, a distributed anomaly detection scheme, TPID(Traffic Prediction based Intrusion Detection), is designed to detect the attacks which make more influence on packet traffic, such as selective forwarding attacks, DOS attacks. In TPID, each node acts independently when predicting the traffic and detecting an anomaly. Neither special hardware nor nodes cooperation is needed. The scheme is evaluated and compared with other method in experiments. Results show that the proposed scheme obtain high detection ratio with less computation and communication cost.

Keywords: component; Wireless sensor networks; anomaly detection; Markov model; traffic prediction ; DOS attacks

1. Introduction

WSN, which is the abbreviation for wireless sensor network, can implement complex and large-scale environmental monitoring and tracking tasks in a wide range of application areas, so it has highly application[1,2]in the national defense and military, environmental monitoring, traffic management, disaster rescue and many other fields. Wireless transmission of WSN, characteristics which are no one care and other natural make it vulnerable to all kinds of attacks. One of the most damaging attacks is deceptive and denial of service (or DoS) attacks. Under the premise of the covering as much as possible, the former seeks to produce the monitoring results of a false so that the results can not be trusted to monitor. The latter seeks to damage local network or even the overall function of the network so that facilities are not available. So when WSN has been used in the scene with important mission, such as the residential

1875-3892 © 2012 Published by Elsevier B.V. Selection and/or peer-review under responsibility of Garry Lee
doi:10.1016/j.phpro.2012.03.352

wireless protection network of the commercial, battlefield surveillance of the military and so on, how to ensure that information can quickly and accurately transmit plays a key role in the success or failure of the whole mission.

A sensor node is a tiny and simple device with limited computational capability and broadcast power. Wireless sensor networks are generally provisioned to consist of a large number of inexpensive nodes reporting their data to a central, more capable sink node using multihop transmission. In general, it is assumed that sensors will be equipped with non-rechargeable batteries and will be left unattended after deployment. However, current and foreseeable future technology has put severe restraints on energy resources of sensor devices. Because long term operation of nodes with limited battery energy is the main design bottleneck of sensor networks, sensor network protocols have to be designed to operate with minimum resource utilization. Security solutions for sensor networks also have to be designed with the limited computational power, limited memory and limited battery life of sensor nodes in mind.

In general, network security solutions can be grouped into two main categories: prevention based techniques and detection based techniques. Intrusion prevention mechanisms, such as authentication, key management, security, routing protocols, and so on, can stop the deceptive attacks launched by external attackers, but it is difficult to confront the DOS attacks which have stronger concealment and destructive power and difficult to confront the deceptive attacks launched by captured node, these attacks must be found and handled through intrusion detection mechanism. The existing intrusion detection schemes of WSN [3-11] judge intrusion mostly through the analysis of data or flow characteristics, as well as collaboration between the nodes, putting forward the high requirements to the storage of the nodes and the computing power, additional communication costs will rapidly consume the limited energy of the nodes. Therefore, how to minimize the cost of resources as soon as possible under the premise of providing a higher detection rate is a key issue to be resolved.

The research content of this article is to design an intrusion detection scheme based on Traffic Prediction for the large-scale tiled WSN (Traffic Prediction Intrusion Detection Scheme, or TPIDS), detecting attacks which have considerable influence on the flow, such as the selective forwarding attacks and DOS attacks. As far as we know, this is first intrusion detection scheme for WSN at home and abroad using the traffic prediction model. The main contributions of the article are the following two points. First, according to the flow characteristics of WSN, it designs a WSN traffic prediction model using Markov which makes lower computational complexity and improves the forecast accuracy; Second, as the traffic prediction model used in WSN Intrusion Detection System, it designs a threshold value beyond judgment algorithm based on prediction to detect anomaly, followed by self-similar characteristics analysis mechanism to judge the reasons for the abnormal flow. If anomalies are caused by the attack, it will send the alarming message to the source node or sink node. There are two points that need to point out. Firstly, TPIDS which is based on anomaly detection technology can detect some unknown attacks, so it is more applicable to attack the WSN which has diversity of variant and unpredictability; secondly, TPIDS adopts a fully distributed scheme, so each node can independently complete traffic prediction, anomaly detection and intrusion judgment based on the analysis to characteristics of self-similar network traffic when collecting data at the same time. The scheme does not require additional hardware support and the cooperation between the nodes. Simulation results show that the improved Markov model has high prediction accuracy. TPIDS is a lightweight intrusion detection scheme, the calculation of each node with very low overhead, and compared with the intrusion detection systems based on cooperation , it can be faster to detect the behavior of the intrusions, while reducing the costs of the communications and energy.

2. Related work

On at proposed a distributed cooperation anomaly detection scheme, which has assumed the attack packets is remarkably different from the normal message in the energy and speed. Each node inserts a detection engine and the engine statistics for two eigenvalues, energy and packet arrival rate, of each node on the neighbors. When abnormal is discovered, the engine radios alarm information. If alarming message ,which is collected contrary to node B by node A, arrived the predestinate threshold, node A was sure that node B was the intrude node. The limitations of the scheme are the following two points. First, detection rate is closely related to the energy and speed of the message sent by attackers, when the attacker uses energy and speed which is close to normal to send the message to avoid detection, detection rate will be greatly reduced. Second, nodes cooperate to judge the abnormal introducing a large number of communication overhead, so it will be rapidly depleting the limited energy of nodes and limited network bandwidth, thus shortening the life of the network. Loo suffered an anomaly detection scheme to detect routing attack. Each node configurates an IDS agent, and at first the agent establishes normal characteristics space of the traffic in the training process. Then in the test stage, if the characteristic information disilled from the packet flowing the node appears in the sparse region of the feature space. In this scheme, nodes detect anomalies independently, thus reducing the cost of communication; however, in the training process, it needs to collect twelve different network traffic characteristics, and computes the mean and variance of each characteristic, so the computational complexity of the scheme is higher. Rajasegarar proposed a hierarchical anomaly detection scheme, against the same hierarchical structure WSN, to reduce communication costs and the calculation overhead of the sink node. Specifically, a structure system of nodes, sink node acts as root and the father node is elected regularly. Each leaf node collects cyclical flow of data acquisition properties, when a time window closing, the nodes according to data collected by similar degree of division, and Packet Data sent to the parent node, node father of the merger of all sub-node packet transfer to the next level, this iterative process repeated until the sink node, node sink summary of all groups and to detect abnormal. With the deployment of a small number in the network with more resources and energy in the high-end nodes, Doumit proposed based on hidden Markov models WSN anomaly detection scheme, low-end nodes will collect the data sent to the high - node, node by the high-end data analysis, modeling and abnormal judgment. Ren based on consideration RTS / CTS MAC layer protocol mechanism of the security issues, presents a detection and defense scheme of DOS attack data link layer.

Yu bo proposed, which is based on the detection point Multi-hop recognition scheme to detect attacks caused by the choice of transmitting the abnormal packet loss, in the scheme, a part of the transmission path nodes will be randomly selected for testing, Detection point will be generated for each incident packet a confirmation packet, and packet confirmed to the upstream transmission, transmission paths, any node in the middle, if not adequately recognized package, will generate warning information of abnormal packet loss, and to submit a multi-hop to the source node. Were randomly selected for testing at the way the enemy can not predict the next point of your selection to avoid a node to become part of the goal of the enemy's capture of high robustness, but the article only consider selecting transmission attacks, while transmission of recognizing packets and messages decrypted on the increase, the introduction of a larger communications, and computing costs. Zhou da quan and Agah will introduce the game theory to Clustering WSN Intrusion Detection, through only two participants in a non-zero; non-cooperative game model describes the attacker and WSN between offensive and defensive issues to prove that the model can be Nash equilibrium reached. Zeng Peng proposed based on the principle of WSN biological immune Intrusion Detection System, but the article is only a conceptual framework structure, the lack of physical implementation and application.

Similar to the literature [3-8], TPIDS also judged weather network was intruded by the attacker or not by detecting abnormal. However, the scheme adopt a different approach, which is based on the ARMA traffic anomaly detection model, each node in the transmission of the data acquisition, at the same time, independent of the completion of flow projections, analysis and anomaly detection. Simulation analysis showed that TPIDS is a lightweight anomaly detection scheme, whether calculation overhead of node or network communication costs are minimal, suitable for large-scale tiled WSN, and it has high detection rate to all attacks which can lead to traffic abnormal changes.

3. Sensor node data ARMA traffic Predection model

Different from the traditional network, WSN is data-centric and the data flow shows that the apparently uneven character; at the same time, WSN network is application-related, mainly on the query of event-driven cyclical data, so the data flow represents the randomness and periodicity correspondingly. Accurate traffic prediction model can capture the statistical characteristics of the actual network. The existing traffic prediction algorithms of WSN mainly include Poisson [12], the Markov process (Markov) [13], the auto-regressive model (AR) [14] and the autoregressive moving average model (ARMA) [15]. Poisson has been proven unsuitable for the flow characteristics of WSN. The paper adopts a Markov model to predict WSN network traffic, and the specific prediction model is shown in the following：

In each node, a random variable sequence X_0, X_1, X_2… is used to denote the state of the node during this period of time. Since each node has a random variable sequence, then at the same time, different nodes can be in different modes. In the possible set of the operational modes, if $X_n = i$, the sensor nodes are in the operational mode i when it is in the time domain n, a time domain is a short time. Assuming that all the states transition take place at the beginning of the any time domain, each node has some fixed probability in the state i . If the next state is j , it used to be denoted as P_{ij} . This probability can be denoted using the following formula:

$$P_{ij} = P\{X_{m+1} = j \mid X_m = i\} \tag{1}$$

P_{ij} Is denoting the probability of entering the state j when a node in the operational state i . The migration probability of Second-order is defined as the $P_{ij}^{(2)}$, expressing that a node in the current state of i , then entering the state j after experiencing two state transition. Namely,

$$P_{ij}^{(2)} = P\{X_{m+2} = j \mid X_m = i\} \tag{2}$$

$P_{ij}^{(2)}$ Can be calculated from P_{ij} by the following formula.

$$P_{ij}^{(2)} = \sum_{k=1}^{M} P_{ik} P_{kj} \tag{3}$$

The migration probability of n order is denoted as $P_{ij}^{(n)}$. Chapman-Kolmogorov equation is defined as follows,

$$P_{ij}^{(n)} = \sum_{k=1}^{M} P_{ik}^{(r)} P_{kj}^{(n-r)} \tag{4}$$

For arbitrary values of 0<r<n, another notation of Markov chain probability is to use an M*M matrix of P which is called migration probability matrix. In this matrix, the element P_{ij} which in the i Th row and the j Th column denote the probability.

$$\begin{bmatrix} P_{11} & P_{12} & . & . & P_{1M} \\ P_{21} & P_{22} & . & . & P_{2M} \\ . & . & . & . & . \\ . & . & . & . & . \\ P_{M1} & P_{M2} & . & . & P_{MM} \end{bmatrix} \tag{5}$$

$P_{ij}^{(2)}$ Is the elements of Matrix P and its own product line matrix section i in Section j.Said with P^2 $P*P$ means that $P*P$ is the matrix P^2 i j Bank in the elements. Similarly, the $P_{ij}^{(n)}$ is the element of matrix P the n-th power of i to j .If $P^{m+n} = P^m * P^n$, then this means that

$$P_{ij}^{(m+n)} = \sum_{k=1}^{M} P_{ik}^{(m)} P_{kj}^{(n)} \tag{6}$$

In the model, migration probability matrix acts on behalf of sensor nodes. Under migration probability matrix and the initial state X_0 of each node, we can build the energy consumption sequence of the entire wireless sensor networks.

How many time-domains does a node experience that arriving at the state S after the number of T time-domains. Now assume that nodes in the state i , namely $X_0 = i$. Because $P_{is}^{(t)}$ represents that a node is in the current state of i , and reaches the state s after experiencing the number of t state transitions, then for the arbitrary state of s , the number of the time-domains that a node remain in the state s can be calculated by the formula as follows:

$$\sum_{t=1}^{T} P_{is}^{(t)} \tag{7}$$

Assuming that B_s is the transmissible data quantity of a note stays in the state s for a time domain, and the T -node domain arrived at the state of the expectations of s domain can be calculated by the number, that is, if the current state of a node is i, then after the number of T time-domains, the total transmissible data quantity, namely B^T is:

$$B^T = \sum_{s=1}^{M}\left(\sum_{t=1}^{T} P_{is}^{(t)}\right) * B_S \tag{8}$$

The data of each node in the time T can be calculated from the formula of $\sum_{t=1}^{T} P_{is}^{(t)}$, with the total number of nodes by the following formula to calculate:

$$B_{total} = \sum_{C_{k-1}}^{C_{k-n}} \sum_{s=1}^{M}\left(\sum_{t=1}^{T} P_{is}^{(t)}\right) * B_s \tag{9}$$

Among the formula, C_{k-i} denotes that it is the i Th sensor nodes of cluster of C_k , and B_S is the transmissible data quantity of a node stays in the state s for a time domain. P_{is} Is the probability from the state i migrating to the state s

4. Based on the traffic prediction model intrusion detection scheme

4.1 Assumption

This paper presents that the test scheme has an important mission for WSN applications Scene. We assume that during the deployment phase, WSN is completely safe, the existing security agreement WSN [8] hold the same assumptions, and that during the actual operation if we adopt the appropriate measures of protection and the deployment plan, it is easy to achieve. At the same time, we assume that the sensor nodes on the link layer has been realized, such as the routing layer protocol (such as WMACA), we can put the security agreement on the operation of these agreements, but it does not depend on specific agreements.

WSN Intrusion Detection System typically includes two stages, detection and response , the former transmission of the data acquisition completed on the network anomaly detection, alarm information will be sent to the source node or nodes sink ;the latter refers to the source node or nodes sink collected adequate warning information, can run more complex decision-making algorithms to IDS and counterattack, for example, can adjusted notice routing protocol routing, the nodes can be reduced through traffic, and even manual inspections and remove malicious nodes. This study focused on considering how to detect anomalies caused by the invasion, does not consider decision-making and counter-measures.

TPIDS is based on the flow prediction and analysis, a major target for a greater flow of attacks, such as radio frequency interference, select E-attacks, hello flooding attacks, sinkhole attacks, and black hole attacks. For other types of attacks, such as spoofed and altered packets, Sybil attacks, the wormhole attack and so on. This article does not consider corresponding preventive measures; interested readers can References [19].

4.2 Analysis on detection rate and false alarm rate

In this section, we TPIDS different schemes on the threshold of detection accuracy of the theoretical analysis, detection accuracy by Detection Rate and False Positive Rate two indicators to assess. (They are detected in the number of malicious packet loss and the number of all the malicious packet loss ratio, as well as non-detected and malicious packet loss to all of the reported text of the malicious packet loss ratio)

Figure 1 shows the detection rate and false alarm rate of the different threshold TPIDS. Further show that the optimal threshold in the range of accuracy of the advantages

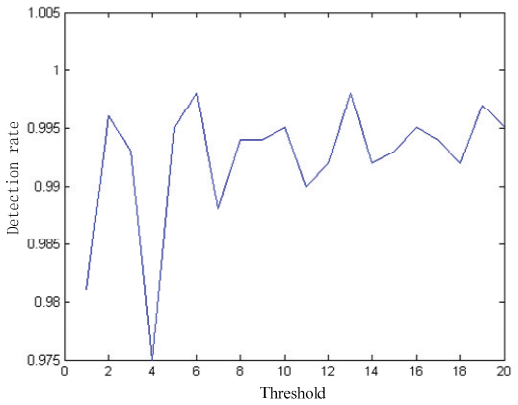

Fig.1 Detection rate

5. Simulation Experiment

In this section, TPIDS by simulating experiments carried out in-depth assessment. NS2 as a simulation tool used in the NS2 achieve detection mechanisms, and in the Paper [9], the simulation include reported malicious attacks, for example, in testing conditions under different packet loss rate, TPIDS detection scheme accuracy and communication overhead. Experimental parameters as shown in table 1

Table 1 Table simulation parameters

Nodes	400
Area	2000m×2000m
Node coordinates of sink	（0，0）
Channel bandwidth	19.2Kbps
Power initial nodes	10J
Periodic Sampling	10s

Perception of each node sampling of the data generated for 4096 byte.

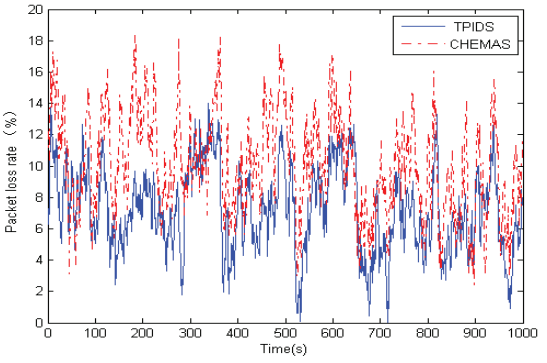

Fig.2 Packet loss rate comparation of TPIDS and CHEMAS

As Figure 2 was shown that compared with CHEMAS， packet loss rates of TPIDS relatively was low. as the rate time decreases, performance of the detection algorithm gets better. However, this also increases the false alarm rate. Again, the selection threshold of in actual sensor networks is a design decision that relies on traffic and network properties.

6. Conclusion

In this paper, we have introduced a novel anomaly detection based security scheme for large scale sensor networks based on Markov model. If each node can build a simple Markov model of traffic predict, these statistics can later be used to detect changes in them. We have shown that, by looking at a relatively small number of received packet features, a node can effectively identify an intruder impersonating a legitimate neighbor.

In our implementation, we considered the anomaly detection algorithm executed at each node separately. Low-complexity cooperative algorithms may improve the detection and containment process. Different routing, medium-access and distributed control algorithms will introduce different features. More research is needed to determine better node features addressing specific vulnerabilities and to develop improved detection algorithms with sensor node capabilities in mind.

References

[1] Ren Feng Yuan,Huang Hai Ning,Lin Chuang.Wireless sensor networks.Journal of Software, 2003.

[2] Sun LM, Li JZ, Chen Y, Zhu HS. Wireless Sensor Networks. Beijing: Tsinghua University Press, 2005 (in Chinese).

[3] Onat I , Miri A. An int rusion detection system for wirelesssensor networks/ / Proceedings of t he IEEE International Conference on Wireless and Mobile Computing , Networking and Communications (WiMOB′ 05) . Mont real , Canada , 2005 : 2532259

[4] Doumit, S. and Agrawal, D.P. "Self-organized criticality & stochastic learning based intrusion detection system for wireless sensor networks", MILCOM 2003 - IEEE Military Communications Conference, vol. 22, no. 1, pp. 609-614, 2003

[5] Rajasegarar S , Leckie C , Palaniswami M , Bezdek J C. Dist ributed anomaly detection in wireless sensor networks/ / Proceedings of t he 10t h IEEE Singapore International Conference on Communication System (ICCS′ 06) . Singapore , 2006 : 125

[6] Agah A , Das S K, Basu K, Asadi M. Int rusion detection in sensor networks : A non cooperative game approach/ /

Proceedings of t he 3rd IEEE International Symposium on Network Computing and Application (NCA′ 04) . Cambridge ,MA , 2004 : 3432346

[7] 42-04-Secure media access control (MAC) in wsn-- intrusion detections and countermeasures

[8] Yu Bo , Yang Min , Wang Zhi , Gao Chuan Shan. Identify abnormal packet loss in selective forwarding attacks. ChineseJournal of Computers , 2006 , 29 (9) : 154221552 (in Chinese)

[9] Agah A , Das S K, Basu K, Asadi M. Int rusion detection in sensor networks : A non cooperative game approach/ / Proceedings of t he 3rd IEEE International Symposium on Network Computing and Application (NCA′ 04) . Cambridge ,MA , 2004 : 3432346

[10] Zeng Peng , Liang Wei , Wang J un , Yu Hai Bin. Research on security system of wireless sensor network based on biological immunity principle. Mini2micro System , 2005 , 26 (11) :190721910 (in Chinese)

[11] Y. Ma and J. H. Aylor, System Lifetime Optimization for Heterogeneous Sensor Networks with a Hub-Spoke Topology, IEEE Transactions on Mobile Computing, Vol. 3, No. 3, July-September 2004:286-294

[12] Y Ma,J H Ayor, System lifetime optimization for heterogeneous sensor network with a hub-spoke topology[J],IEEE Trans on Mobile Computing,3(3):286-294

[13] Lisa A, Shay, The wireless network environment sensor: a technology independent sensor of faults in mobile wireless network links [D]. Rensselaer Polytechnic Institute Troy, New York ,USA, 2002

[14] Jing Deng, Richard Han and Shivakant Mishra. Defending against path-based DoS attacks in wireless sensor networks [C],ACM press ,New York, NY, USA:89-96

[15] Ilker Demirkol,Fatih Alag¨oz,Hakan Delic,Cem Ersoy. Wireless Sensor Networks for Intrusion Detection: Packet Traffic Modeling[EB/OL]. www.cmpe.boun.edu.tr/~ilker/IlkerDEMIRKOL_COMML_ext_abstract. pdf

[16] Zhang ShuJing , Qi Li2Xin. Time Series Analysis Concise Guide. Beijing : Tsinghua University Press , 2003 (in Chinese)

[17] Zou Bai2Xian , Liu Qiang. ARMA based t raffic prediction and overload detection of network. Journal of Computer Research and Development , 2002 , 39 (12) : 164521652 (in Chinese)

[18] Karlof C., Wagner D.. Secure Routing in Sensor Networks: Attacks and Countermeasures. In: Proceedings of the 1st IEEE International Workshop on Sensor Network Protocols and Applications, Anchorage, 2003,113~12