ISSN (Online) : 2319 – 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization, Volume 3, Special Issue 1, February 2014

International Conference on Engineering Technology and Science-(ICETS'14)

On 10th & 11th February Organized by

Department of CIVIL, CSE, ECE, EEE, MECHNICAL Engg. and S&H of Muthayammal College of Engineering, Rasipuram, Tamilnadu, India

Secured Aggregation Based on Learning Procedure in Wireless Sensor Networks

K.Sudha[1], P. Divya Bala[2], D.Lavanya[3], S.Gajalakshmi[4]
Assistant Professor, CSE, Christ College of Engineering and Technology, Puducherry, India.[1]
Final year Student, CSE, Christ College of Engineering and Technology, Puducherry, India. [2, 3, 4]

Abstract—Wireless Sensor Networks (WSNs) is one of the important areas in research center, causing major impact on technology improvement. In a Wireless Sensor Network, attacks are usually based on the signature in a centralized approach which detects the anomalies. In this paper the existing Extended Kalman Filter (EKF) mechanism that can be used to find the malicious node which sends the false information with the constant threshold value. Each node in a network contains the normal value if any emergency event occurs then there will be a change in value of node. For this purpose system monitoring module (SMM) can be used, that will oversees the network behavior. In this paper, we proposed k-nearest neighbor (k-NN) algorithm for secure wireless network. This algorithm uses two methods such as distance and density in addition to this it also uses cluster summary aggregated from local and parent node to the base station which is used to identify the anomalies. We use centralized detection is useful for network anomalies, while network data can be piggybacked in packets, providing the centralized anomaly detector with a comprehensive view of network state.

Index Terms—Cumulative Summation (CUSUM), Extended Kalman Filter (EKF), Generalized Likelihood Ratio (GLR),k-Nearest Neighbor (k-NN), Wireless Sensor Networks(WSNs).

I.INTRODUCTION

A wireless sensor network (WSN) consists of number of sensor nodes in a network. These sensor node has the capacity of sensing, self-monitoring that communicate through wireless network and also the sensor nodes are autonomous and capable of monitoring the condition in atmosphere such as heat, temperature, motion, vibration, pressure at various locations. It is a challenging task to design system for anomaly detection. Traditional techniques used in wired network for intrusion detection is totally different for WSN because of limited resources in sensor node. Though many protocols have been proposed for the performance and aggregation of nodes only few are considered. To overcome and recovery drawbacks in the base paper, we proposed a new kNN-based algorithm for Anomaly Detection which is used to subdue the lazy-learning problem on hyper grid intuition. In this method, A scalable kNN-based Anomaly Detection scheme is proposed for supervising WSNs based on a hyper-grid, which can able to work online in a distributed manner.

In this paper, we proposed a novel scalable kNN-based Anomaly Detection scheme for supervising the sensor nodes based on a hyper-grid, which can work in a distributed manner on an online network. To correct the disadvantages in base paper, we put up a new kNN-based Anomaly Detection scheme based on hyper-grid insight for WSN applications to beaten the lazy-learning problem [5]. Anomaly detection secures WSNs from cyber attacks and unspecific faults. As a measureable and parameter-free unmonitored AD technique is k-nearest neighbor (kNN) algorithm has drawn a lot of interest for their usage in computer network and WSNs. Our proposed work is based on kNN algorithm.

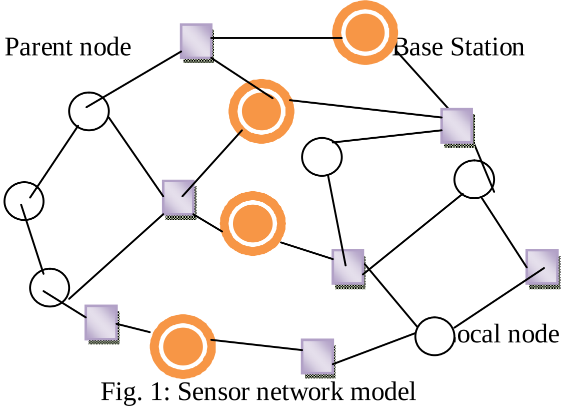

Fig. 1: Sensor network model

In this algorithm with the training data set for normal actions, the kNN based text characterization method can be easily get used for anomaly detection. For each new process, the audit data is scan for test and the sequence of system call is extracted. As previous case this new process is also converted to a vector by using the weighting method. This helps to know the correspondence between the each process and also between the new processes in the training data set is calculated for normal actions. We calculate the average correspondent value of the k nearest neighbors and set a threshold value. The new process is said to be normal when the average correspondence value is above the threshold. This scheme performs well automatically without any manual adjustment in argument. The new scheme characteristics are very low computation and communication raised, In This scheme the complexity is directly proportional to the sample logarithmical size. When complexity increases then the sample size also increases logarithmically, in place of quadratically. Complexity in computation is not high.

We have organized this paper in such a way that section II consists of existing system, section III consists of proposed scheme, section IV consists of implementation, section V consists of experiments and results, and finally section VI with the conclusion.

II.EXISTING SYSTEM

Extended Kalman Filter (EKF)

Depending upon the state-space technique used to represent a general problem of trying to evaluate a state of a dynamic system perturbed by Gaussian white noise, but measurements for linear functions will be corrupted by additive Gaussian white noise. EKF with linear estimation theory can be applied to many nonlinear applications by approximating effects of small perturbations linearly [11]. WSN application and utilizing time update and measurement update equations to recursively process data. In our technique, state denotes an actual value to be observed. In a given instant of time state can be specified by values with an attribute of interest. Such that actual temperature monitored by WSNs.

Cumulative Summation- Generalized Likelihood Ratio (CUMSUM-GLR)

EKF does not consider the attribute of attacks introduced at different times are not always independent. Therefore, an EKF based approach avoids the information provided by the entire sequence of measured values. We further illustrate an algorithm for joining both CUSUM and GLR, which uses the cumulative sum of the deviations between measured values and estimated values. Due to the usage of Extended Kalman Filter (EKF) based detection, this may provide false detecting false injected data because of the filtering nature. EKF based technique does not consider the fact that attacks introduced at different times are sometimes dependent. Therefore, an EKF based approach uses the complete sequence of calculated value that avoids the information. CUMSUM and GLR have high computational complexity [11].

III. PROPOSED SCHEME

In WSN environment there is a very large set of data and most of the elements are normal, some of intrusions are hidden in data set. One of major pros of unsupervised anomaly detection algorithms is for processing unlabeled data and detect intrusions that otherwise could not be detected. With reference to these type of algorithm can be a partially automate the manual monitoring of data in forensic analysis which helps on focusing the malicious element of the data. The k-Nearest neighbors helps us to define the point by determining the sum of the distances to the k- nearest neighbors in sparse region of the feature space. This volume is represented as the kNN score for a point. Intuitively, the points in tightly packed region will surely have n points near them and also the kNN score will be less. If the frequency of any given attack type in the data set is more the value of k the images of the attack elements are far from the images of the normal elements, then the kNN score is useful for detecting these attacks. Similarity between data points are measured which is used to detect anomalies and k-Nearest neighbor is also one such anomaly based method. With the assumption that anomalous nodes are mostly having a wide range of distance or vary from other points and this method also assumes that the normal nodes will have nearest neighbor. For measuring the similarity between the data points two methods are used which are as follows.

Methodologies:

This kNN algorithm uses major two methodologies which are briefly explained as follows. The first method is based on distance, such as Euclidean distance. The data points mostly have a wide range of distance or vary from other points are malicious. Thus the distance between the data points are always calculated from the beginning is supervised as results of variation malicious are detected. The second method is based on density which calculates the relative density of the neighbors of each data point to find the suspicious elements of the data. If a data point has low density region are declared as malicious node or else if the data point has a high density region then it is declared to be a normal or uninfected node .This method uses data points obtained from the sensor nodes has a region of varying densities and also vary with the distance least to worst performance worst. To compute the relative density or distance this method requires significant numbers of neighbors

In a vast dimensional data, It is difficult to compute the data sparse and similarity between the data point since the distance between different points may be similar. One of the best way to minimize the transmission overhead is by the data gathered from sensor node is clustered locally depending on the Euclidean distance between data vector and the centurion of the fixed width cluster. The computed distance is compared with the radius if the radius is greater than distance then the data vector and cluster are added if not a new cluster is created. Each and every local node will compute a summary and this summary is send to the parent node. The inter-cluster distance rule is used by the parent node to merge all the summaries received from the local nodes. The inter-cluster distance rule defines that the distance between threshold is greater than any two cluster if not as in previous case again a new cluster will be created. The cluster summary merged by the parent node will be send to the base station for detection

IV.IMPLEMENTATION

The computation involves the learning procedure and summation of data points at different position is calculated and this calculated value is spread to the parent node from child node. The major function of this child is to get the summary and sort them in position in diminish order. After this process over complete the calculated value is received by the parent node is detected online locally. In This algorithm given below o is set to value less than minimum which denotes the coefficient and the value of h is considered to be the value of hyper cube. The number of test data along with the hyper cube is much more than that of value of k then it is declared as normal node or if the value is less than k then it is suspected to be malicious. Suppose consider a critical situation in which the data is entirely new or fully changed in such situation the parent node sends a alert to the child in order to make the change or update to the data and also request the child node to start the learning procedure.

Table 1: Algorithm for k NN

```
Assumption: If k data exist in test data then it is normal, else it's abnormal or anomalous.
Input: parameter k and data value
Output: whether test data is normal or malicious
//Anomaly detection
get the data value in x
for w=1…m
 $n_w=x_w+o/h$,
 for w=1…m
if $n_w-|n_w|>0.5$ then s[w]=1 else s[w]=-1
Set  a=0
for $q_1=n_{1,}n_1+s[1]$…..$q_m-1=n_m-1,n_m-1+s[m-1]$
$q_m=n_{m,}$pos = 0
for w=1,…m
pos=$q_1$<<(m-w)/pos
if a<k then
 label as anomaly
else
label as normal
.return
```

V.EXPERIMENTS AND RESULTS

Our main goal is to attain reliable performance and also with accuracy detection must be achieved. If the standard deviation mean of a cluster is less than the average inter-cluster distance then it is identified as anomalous. With conclusion of result it is proved that compared to centralize clustering algorithm distributed clustering algorithm minimizes the communication overhead. Thus in this kNN algorithm the cluster size directly depend on the detection rate. It is also necessary to calculate the false positive rate as it is detected and recorded by the detection region. A detailed complexity analysis of the new scheme in terms of its requirements in computation, memory, and communications The proposed scheme is composed of three major procedures, including learning, detection, and update.

Thus the algorithm given above determines the value less than minimum which denotes the coefficient and the value of h is considered to be the value of hyper cube. The number of test data along with the hyper cube is much more than that of value of k then it is declared as normal node or if the value is less than k then it is suspected to be malicious. Suppose consider a critical situation in which the data is entirely new or fully changed in such situation the parent node sends a alert to the child in order to make the change or update to the data and also request the child node to start the learning procedure as explained in previous section.

The rate of detection accuracy curve with the values of detection accuracy and false positive rate is given as shown in fig. 2

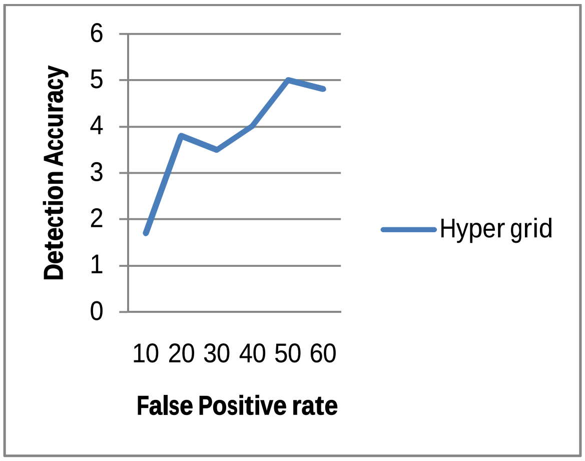

Fig. 2: Rate of Detection accuracy curve

VI. CONCLUSION

In This paper we propose a kNN algorithm used to determine the sum of the distance between the data points. This was proposed to pros of simplicity and scalability. Similarity between data points are measured which is used to detect anomalies; this method is also based on density which calculates density of the neighbors of each data point to find the suspicious elements of the data. If a data point has low density region are declared as malicious node. A cluster is identified as anomalous if the standard deviation is less than the average inter-cluster distance.

REFERENCES

[1] R. Pon, M. Batalin, M. Rahimi, Y. Yu, D. Estrin, G. J. Pottie, M.Srivastava, G. Sukhatme, and W. J. Kaiser,2004"Self-aware distributed embedded systems," in Proc. IEEE FTDCS, May 2004,

[2] D. Wagner, "Resilient aggregation in sensor networks," in Proc. ACM SASN, 2004, pp. 78–87.

[3] H. Cam, S. Ozdemir, P. Nair, and D. Muthuavinashiappan, 2003,"Espda: Energy efficient and secure pattern-based data aggregation for wireless sensor networks," in Proc. IEEE Sensors, pp. 732–736.

[4] C. Castelluccia, E. Mykletun, and G. Tsudik,2005, "Efficient aggregation of encrypted data in wireless sensor networks," in Proc. MOBIQUITOUS, pp. 109–117.

[5]http://seit.unsw.adfa.edu.au/staff/sites/hu/Sample_Publication/Tpds_scalable.pdf

[6] S. Zhu, S. Setia, S. Jajodia, and P. Ning,2004, "An interleaved hop-by-hop authentication scheme for filtering false data injection in sensor network"s, pp. 260–272.

[7] B. Przydatek, D. Song, and A. Perrig, 2003,"SIA: Secure information aggregation in sensor networks," in Proc. ACM Sensys, pp. 255–265.

[8] L. Hu and D. Evans, 2003,"Secure aggregation for wireless networks," in Proc. Workshop Security Assurance Ad Hoc Network.

[9] Y. Yang, X. Wang, S. Zhu, and G. Cao,2006, "SDAP: A secure hop-by-hop data aggregation protocol for sensor networks," in Proc. ACM MOBIHOC, pp. 356–367.

[10] K. Wu, D. Dreef, B. Sun, and Y. Xiao,2007, "Secure data aggregation without persistent cryptographic operations in wireless sensor networks," Elsevier Ad Hoc Networks, vol. 15, no. 1, pp. 100–111.

[11] Bo Sun , Xuemei Shan, Kui Wu, and Yang Xiao, 2013, "Anomaly Detection Based Secure In-Network Aggregation for Wireless Sensor Networks", vol. 7.

[12] B. Krishnamachari and S. Iyengar, 2004,"Distributed Bayesian algorithms for fault-tolerant event region detection in wireless sensor networks," IEEE Trans. Comput., vol. 53, no. 3, pp. 241–250.

[13]T. Clouqueur and K. Saluja,2004, "Fault tolerance in collaborative sensor networks for target detection," IEEE Trans. Comput., vol. 53, no. 3, pp.

[14] D. Dong, Y. Liu, and X. Liao ,2008 , "Self-monitoring for sensor networks," in Proc. ACM MobiHoc.

[15] K. Premkumar , A. Kumar , and J. Kuri, 2009, "Distributed detection and localization of events in large ad hoc wireless sensor networks," in Proc. 47th Annu. Allerton Conf. Communications Control Computer, pp. 178–185.