Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Finding Abnormal Events in Home Sensor Network Environment Using Correlation Graph

Huey-Ming Lee
Department of Information Management
Chinese Culture University
Taipei, Taiwan
hmlee@faculty.pccu.edu.tw

Ching-Hao Mao
Department of Information Management
Chinese Culture University
Taipei,Taiwan
chmao2008@gmail.com

***Abstract*—Anomaly detection in sensor network seems a challenge when encountering the limitation of the energy requirement and dynamics environments. It is to rapidly analyze and identify the abnormal events among the extreme volume data. Using correlation graph representation to correlate the events generated by sensor networks is capable to find the intentional dependency behavior's insight for detecting home sensor network abnormal events. In this study, we proposed an anomaly detection mechanism based on correlation graphs of sensor networks for rapidly identifying abnormal home events. The proposed mechanism which makes the following contributions: (a) it is automatically identify the abnormal event under home sensor network environment (b) it eliminates irrelevant events for saving the computation power (c) it is easily to apply on different machine learning classifiers for enhancement. The evaluation from Intel Berkeley Research lab sensor network data set. The proposed mechanism performs well in sensor events elimination and abnormal event detection.**

***Index Terms*—Sensor network, correlation graph, anomaly detection, information applicances, artificial intelligent.**

I. Introduction

As the internet growing rapidly, we can use many devices to connect the internet for many kinds applications, such as: entertainment, E-commerce, healthy care, etc. Since information appliances (IAs) integrated with sensor network have become available to all in recent years, there are more and more varied IA products appeared. Having a complete and robust management mechanism is important to bring the functions into full play. In other words, sensor networks which always deploy an unattended area for remote monitoring and control of actuator in home network environments. Detecting in such enormous volume of recognized data and considering the performance of whole performance become important issues in both fields of sensor network and information appliances.

There are several studies to mention the information appliances controlling. Lee and Huang [11] proposed an IA controlling model (IACM) which can control IA devices through home management broker. Lee et al. [10] came up with the idea of IAs intelligent agent model (IAIA) that makes home environments more comfortable and convenient. Lee et al. [14] proposed a fuzzy neural network model of information appliances with the functions of self-learning and fuzzy inference; it enables IAIA to maximize efficiency of IAs in a more humane way. Lee et al. [15] proposed an intelligent control model of information appliances (ICMIA) which can not only collect the related users' information appliances preference messages automatically, but also generate the IA control rules by the fuzzy neural network learning. Lee et al. [13] proposed an emergency model of home network environment based on genetic algorithm. This model can not only adapt the home network environment by using genetic algorithm but also detect the emergency events automatically [12]. If there is a mechanism which can do the active response of emergency, then we can prevent the serious accident in home network environment.

Data analysis of sensor network seems a challenge work when encountering the limitation of the energy requirement and dynamics environments. In sensor networks, radio communication is the majority of consumed energy and is much more than the computation. In anomaly detection, reducing the data communication in sensor network is helpful to raise the performance efficiency in detection. Several works intend to do the anomaly detection in different applications, e.g. faults and malicious attacks [22], unusual changes in the monitored environment [23], elimination of erroneous measurement. In this study, we mainly focus on how to identify the abnormally home sensor network event, e.g., fire events, abnormal temperatures. Via reducing the amount of data which needs to be interchanges in the whole networks is as an important issue.

Although many anomaly detections have already proposed in sensor network fields [24-25], however, it is not straightforward to apply into sensor network environment due to the constraint of computation resources. Most of the algorithms which have been proposed are assuming that data is stored centrally. Especially in large sensor network, even in home network environment, it is difficult and time consuming to detect these abnormal events. In many sensor nodes environment, correlating these nodes is quite important for keeping the relations and understanding the context of events. In order to find long term dependency in whole sensor data, we use n-gram models to profile the relations of sensor nodes and keep the distribution of subsequence patterns. According to the subsequence pattern finding by n-gram model, we can construct the correlation graphs to keep the relations among the alerts. Even correlation graphs can keep the relations among the alert at a given time interval, we need capture non-proximity, nevertheless, coming

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

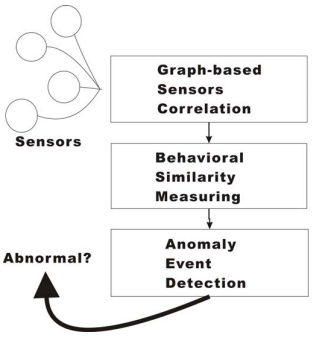


Fig. 1. The architecture of proposed method.

with related temporal and statistical patterns. We attempt to extract the features to construct the long-term dependency relationship among correlation graphs for identifying abnormal events in home network. It is a general technique that can be applied to any application of home network management in which abnormal behavior comes from diverse of sources.

In this study, we analyze sensor network data for a set of related sensor data rather than for individual one. When the source actor is an abnormal event with high temperature, the scenario would be abnormal behaviors; otherwise, it is a normal event. In order to profile behavior path among different correlated graphs, dimension reduction method is used to find the meaningful low-dimensional structures hidden in their high-dimensional observations. We hope to find out the potential behaviors of attacks by extracting their sequential behavior for identifying abnormal behaviors. The remainder of this paper is organized as follows. Section 2 contains the proposed system architecture and algorithms. Also, we introduce the construction of correlation graph, and use manifold learning method to profile the event paths for anomaly detection. In Section 3 contains an evaluation of our approach's performance. Then, in Section 4, we summarize our conclusions.

II. Correlation Graphs and Anomaly Detection

In order to identify the sequential malicious behaviors, we proposed Trajectory Finding mechanism based on Correlation Graph which captures the alert temporal relations. In this Section, we illustrate our proposed mechanism in fully details. The proposed mechanism consists of three modules: Correlated Graph Constructer, Malicious Trajectory Profiling and Malicious Behavior Detector. The correlated graph is used to profile the potential abnormal behaviors based on sequential alerts events given by a specified time interval. Using constructed correlated graph, we develop a graph behavioral similarity method and use dimension reduction technique to extract the potential behavior trajectories. Finally, we apply state-of-the-art machine learning classification algorithm to identify malicious behavior in network environment.

A. Sensors graphs correlation

The goal of correlation graph which uses tree structure to profile sequences of sensor network events. Since diverse network behavior often occurs simultaneously, constructing the correlation graphs should tolerance the noise and redundant data. Under these correlation graphs, capturing the co-occurred relations of proximity sensor events is helpful for finding the temporal relations among whole sensor network environment.

For constructing of correlation graph, three main processes are needed. Initially, it is given the sensor data SA, the time frame length l and window size w. First, we extract the proximity relations in the sequence of sensor data in the specified time interval. We use sliding window approach, that is, given the window size w and behavior thread separated by former step, by capturing the bigram relations between sensor data to capture the alert relations keep in adjacency matrices W. Finally, in order to finding the representative structures among alerts bigram relations, we try to find out the maximum spanning tree given the original bigram frequency graph based on mutual information. Let X and Y be two random variables represent as two alert type in this study, with joint distribution p(x, y) and marginal distribution p(x) and p(y) respectively. The Mutual Information is defined as Equation 1:

$$I(X;Y) = \sum_{x}\sum_{y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}, \tag{1}$$

Mutual Information is attractive because it is not only easy to compute, but also takes into consideration of context statistics. The Mutual Information between two terms can be calculated using preview's alert proximity adjacency because it just compute pair-wise alert distribution.

B. Behavior similarity measuring

Measuring the distance between activity graphs is quite crucial for applying machine learning approach. Here we proposed a behavior distance measuring mechanism for comparing arbitrary two activity graphs. The behavioral distance means measuring the deviation of temporal network behaviors in different time periods. The goal of measuring behavioral distance is to provide detecting the coming up abnormal behaviors by computing the diversity with normal behaviors. The behavior distance measuring uses two kinds of measures, one is structure difference measure, the other is distribution difference measures. Structure difference measure is defined as the sum of missing links (a link in graph G_A but not in graph G_B) and extra links (a link in graph G_B but not in graph G_A) shown as Equation 2 and Equation 3. Using structure difference can find the degree of both graphs casual interactions.

$$\delta = \sum_{i=1}^{n} \delta_i \tag{2}$$

$$\delta_i = (\pi_i(A) \cup \pi(B) \setminus \pi(A) \bigcap \pi(B)) \tag{3}$$

Distribution difference measures adopt Kullback-Leibler divergence (KLD) or relative entropy to measure the similiarity from the graphs distribution. The KLD arises in many contexts as an appropriate measurement of the distance between two models. In the relational activity graph, we would like to

measure the distance between two relational activity graphs distribution. Each node can be regarded as a random variable, whole relational activity graphs can be regarded as a dependency tree come with an statistical model. Given two probability distribution P of G_A and Q of G_B, we can calculate the KLD using Equation 4.

$$D_{KL}(P\|Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)} \tag{4}$$

Finally, we combine these two measures as Behavior Distance(BD) Measuring according to the proportion of matchted edges between two graphs, shown as Equation 5.

$$\begin{aligned} BD &= \alpha(\delta) + (1 - alpha)(D_{KL}(P\|Q)) \\ where \quad \alpha &= \frac{\#\, matched\, edges}{\#\, total\, edges} \end{aligned} \tag{5}$$

Finally we proposed a new behavior distance measuring mechanism shown as Equation5 in Isomap for extracting the meaningful features and finding the behavior trajectory at low dimension space.

C. Anomaly detection in correlation graphs

In this study, we use machine learning supervised learning classification algorithm, Support Vector Machines (SVMs), for detection of malicious behaviors. SVMs are well suited for solving binary classification problems like intrusion detection. Also, we compare the result with other machine learning classification methods, for examples, Naive Bayes (NB) and k-Nearest-Neighbor (kNN) algorithms.

III. Experimental Result

This section presents the experimental results based on a case of abnormal indoor temperature management, and discusses the self-adaptive evolutionary negative selection model for home abnormal events detection. In sub-section A, we describe the used dataset from Intel Berkeley Research lab [7]. Furthermore, the experimental results will be presented in sub-section B.

A. Dataset Description

To analyze the availability of proposed model, we employ the room controlled data generate from Intel Berkeley Research February 28th and April 5th, 2004. This dataset is also a benchmark for the studies of sensor network. It uses Mica2Dot [16] sensors with weather boards collected time stamped topology information, along with humidity, temperature, light and voltage values once every 31 seconds. This dataset includes a log of about 2.3 million readings collected from these sensors. The great quantity of this dataset is suitable for verifying the function of adapting and stabilizing our proposed methods.

B. Experimental Results

In this subsection, we show the experimental results to evaluate our proposed method both in redundant sensor data reduction and abnormal behavior detection. To demonstrate the effectiveness of our proposed approach in meaningfully finding the behavior path, we illustrate the example from data set first. Then, the effectiveness of behavior thread separation mechanism is demonstrated. Finally, we compared the error rate, false positive rate, false negative rate given different time frame length and analyzed the characteristics of proposed methods.

Figure 2 is the 2-Dimension plots for this experiment using Intel Lab data set from Feb. 28th to Mar. 5th behavior path finding respectively. After using trajectory finding mechanism, we can recognize normal and abnormal behavior even in the 2-Dimension spaces obviously.

Furthermore, we analyze the abnormal event detection performance of proposed method. Figure 2 shows the precision rate and recall rate for each result respectively. After finding the behavior path based on correlated, the error rate is under the 10

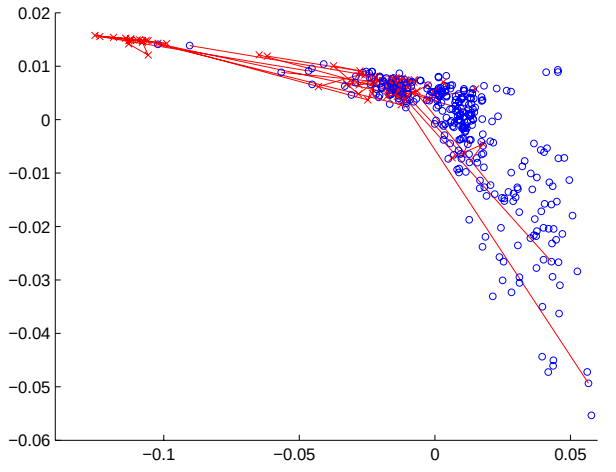
The longer length of time window, the much more relations can be observed by the correlation graphs. In other words, the abnormal behaviors come with relationships is capable to observed when the length of time frame could cover the abnormal behavior. In our experiments, Intel data exists in several relational abnormal behavior come with the relations among different steps. In Figure 3, the recall and precision rates in Intel data set decrease with the enlargement of the time frames. That is because increasing the length of time frame just correlated more un-related events which can be regarded as noise data. Consequently, the proposed mechanism is much more suitable to find the abnormal behaviors which come with relationships.

IV. Conclusion

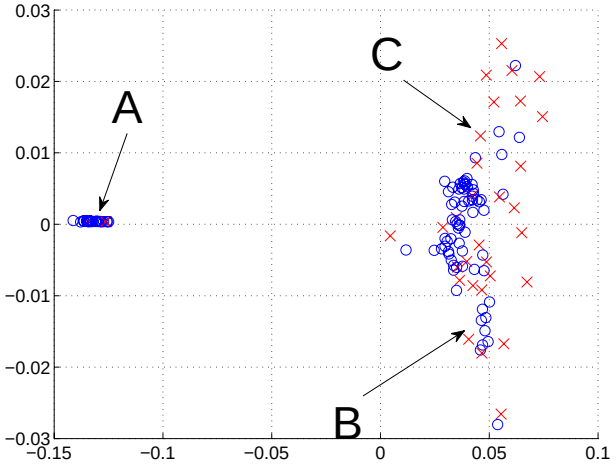
In this study, we proposed an anomaly detection mechanism based on correlation graphs of sensor networks for rapidly identifying abnormal home events. The proposed mechanism which makes the following contributions: (a) it is automatically identify the abnormal event under home sensor network environment (b) it eliminates irrelevant events for saving the computation power (c) it is easily to apply different machine learning classifiers for enhancement. The evaluation from Intel Berkeley Research lab sensor network data set is performed positive results. The proposed mechanism performs well in sensor events elimination and abnormal event detection.

Acknowledgment

This work was partially supported by the National Science Council, Taiwan, under Grant NSC-97-2221-E-034-016- . The authors gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.



(a)



(b)

Fig. 2. (a) According to time series, the abnormal nodes are connected for analysis the anomaly trends,(b) The area (A) denotes the temporal changes states, (B) and (C) are combines several different sensor behaviors.

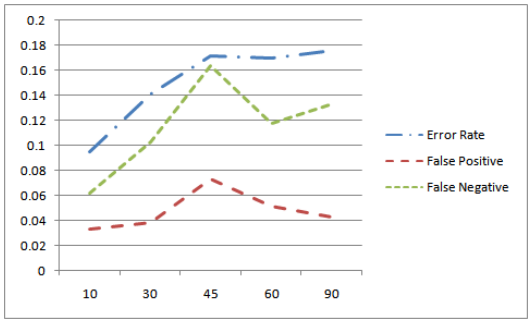


Fig. 3. The comparison result given different time frames in Intel data set.

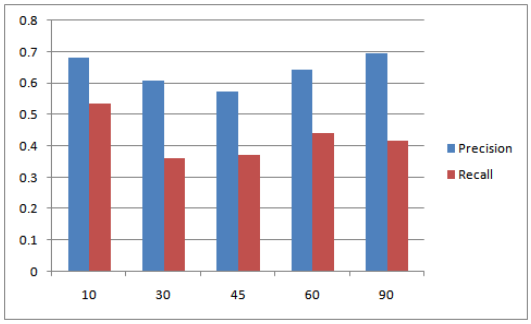


Fig. 4. The comparison result given different time frames .

REFERENCES

[1] A. Arslan, M. Kaya, Determination of fuzzy logic membership functions using genetic algorithm, Fuzzy Sets and Systems, vol. 118, pp. 297-306, 2001.

[2] O. Cordon, F. Herrera, A Three -Stage Evolutionary Process for Learning Descriptive and Approximate Fuzzy-Logic-Controller Knowledge Bases From Examples. International Journal of Approximate Reasoning, vol. 17, pp. 369-407, 1997.

[3] O. Cordon, F. Herrera, F. Hoffmann, L. Magdalena, Genetic Fuzzy Systems. World Scientific Publishing Co. 2001.

[4] P. D'haeseleer, An immunological approach to change detection: Theoretical results". In: Proceedings of the 9th IEEE Computer Security Foundations Workshop, IEEE Computer Society Press, Los Alamitos, 1996.

[5] S. Forrest, S. Hofmeyr, A. Somayaji, Computer immunology, Communications of the ACM, vol. 40(10),pp. 88-96, 1997.

[6] F. Herrera, M. Lozano, J. L. Verdegay, Tuning Fuzzy Logic Controller by Genetic Algorithm," International Journal of Approximate Reasoning, vol. 12, pp. 299-315, 1995.

[7] Intel Lab Data: http://db.lcs.mit.edu/labdata/labdata.html

[8] C.-H. Jung, C.-S. Ham, K.-I. Lee, "A real-time self-tuning fuzzy controller through scaling factor adjustment for the steam generator of NPP," Fuzzy Sets and Systems, vol. 74, pp. 53-60, 1995.

[9] M.-S. Ju, D.-L. Yang, Design of adaptive fuzzy controls based on natural control laws, Fuzzy Sets and Systems, vol. 81, pp. 191-204 , 1996.

[10] H.-M. Lee, Y.-C. Chen, J.-J. Chen, The Intelligent Agent Design of Information Appliance, In: JCIS 2003, Proceeding of the 7th Join Conference on Information Sciences, Cary. NC. USA, pp. 1681-1684, 2003.

[11] H.-M. Lee, J.-H. Huang, The study of IA devices monitoring model. In: The sixth seminar of the research and practices of information management, pp. 430-437,2002.

[12] H.-M. Lee, S.-F. Liao, S.-Y. Lee, An Adaptive Exception Process Model of Information Appliances, Wseas Transactions on Information Science & Applications vol 1, no 3, 778-783 (2004).

[13] H.-M. Lee, S.-F. Liao, S.-Y. Lee, An Emergency Model of Home Network Environment Based on Genetic Algorithm. In: Khosla, R., Howlett, R.J., Jain, L.C. (eds.) KES 2005. LNCS (LNAI), vol. 3682, pp. 1245-1251. Springer, Heidelberg (2005)

[14] H.-M. Lee, C.-H. Mao, S.-Y. Lee, A Fuzzy Neural Network of Information Appliance, International Workshop on Fuzzy System & Innovation Computing 2004 (FIC2004), Kitakyushu, Japan (2004)

[15] H.-M. Lee, C.-H. Mao, S.-Y. Lee, Intelligent Control Model of Information Appliances," In: Negoita, M.G., Howlett, R.J., Jain, L.C. (eds.) KES 2004. LNCS (LNAI), vol. 3215, pp. 123-128. Springer, Heidelberg, 2004.

[16] Mica2Dot: http://www.xbow.com/Products/productsdetails.aspx?sid=73

[17] Z. Michalewicz, Genetic Algorithms + Data Structures = Evolution Programs, 3rd edn. Springer, Berlin Heidelberg New York, 1996.

[18] L.H. Randy, E. H. Sue, Practical Genetic Algorithm, Wiley-Interscience Publication, Chichester, 1998.

[19] G. Eason, B. Noble, and I. N. Sneddon, On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955.

[20] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1992, pp.68-73.

[21] I. S. Jacobs and C. P. Bean, Fine particles, thin films and exchange anisotropy, in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, pp. 271-350, 1963
[22] M. Young, The Technical Writer's Handbook Mill Valley, CA: University Science, 1989.
[23] E. Shi and A. Perig, Designing Secure Sensor Networks, IEEE Wireless Communications, pp. 38-43, 2004.
[24] H.-M. Lee and C.-H. Mao, A Self-adaptive Evolutionary Negative Selection Approach for Home Anomaly Events Detection", Lecture Notes in Computer Science, Volume 4694, pp. 325-332, 2008.
[25] S. Rajasegarar, C. Leckie and M. Palaniswami, Anomaly Detection in Wireless Sensor Networks, IEEE WirelessCommunications, August 2008.