ALERT CORRELATION USING SUPPORT VECTOR MACHINE FOR MULTI INTRUSION DETECTION SYSTEMS

XIAOYUN YE· MYUNG-MOOK HAN
Department of Computer Science, Gachon University, Seongnam, South Korea
(E-mail: yxysun@gmail.com)

[ABSTRACT]

This paper presents a new alert correlation model for multiple intrusion detection systems. Based on the analysis of the complex relationship between the alert information of the intrusion detection system, an alert fusion model is proposed and used to alert correlation. The SVM algorithm has an advantage in the multidimensional classification, which can further reduce the influence of false positives and false negatives. The experimental results show that the alert fusion model has high accuracy and low false positive.

[KEYWORDS] *Alert Correlation, Intrusion Detection System(IDS), Support Vector Machine (SVM)*

1. INTRODUCTION

Recently, the Stereo Vision System (SVS) has been developed not only to monitor the wide range of the surrounding but also to measure the distance about the object of interest [1]. Today, the automated highway system (AHS) is being extended to be robust to withstand the influences of dynamic and unstructured environments (e.g. desert). Moreover, the AHS is being developed to adapt the heterogeneous vehicles (i.e. different shapes of vehicles) [2].

The SVS plays an important role in AHS through detecting, tracking, locating, and recognizing the heterogeneous. Measuring distance about the object of interest enables AHS to reduce the cost because the SVS could be used instead of other measuring devices (i.e. RIDAR and LIDAR).

However, the need for high accuracy prevents the SVS to be applied in automated heterogeneous platoon [1].

In SVS, the disparity image has to be accurate enough to generate the depth map associated with the object of interest. Alfraheed et al [3] [4] have introduced the back view of the proceeding vehicle (BVPV) as a reference object for the automated heterogeneous platoon. The problem arisen in their suggestion is that the number of the corresponding points between the left and right images of the SVS (i.e. bilateral images) is not high enough to generate an accurate depth map and disparity images. Starting from this challenge, the extracted features are used to increase the number of corresponding points by using them to match bilateral images. Thus, the corresponding points generated by SURF [5] have to be increased for calibrating SVS successfully.

Furthermore, these points have to be extracted in dynamic environments (e.g. hazy weather, dusty weather or etc). Our significant contribution is to develop a robust and effective approach that is able to extract these points for Automated Highway Systems or Automated Vehicles in these environments.

The novelty introduced here is to improve the current work of the SVS in terms of the automated heterogeneous platoon. The SVS has to be more accurate once it is used to measure the distance of object of interest. In AHS, the reference object is located at different distances which vary from 3 meters to 10 meters. Therefore, the SVS can be installed in AHS instead of high cost distance measurement sensors. In this work, the first stage of improvement introduced aims to efficiently extract the corresponding points in context of feature matching instead of point matching. The extraction has to be also compatible with different distances of the object of interest.

2. RELATED WORK

The complex relationship between intrusion events determines that there will be complex relationships between intrusion detection system alarms, and different fusion methods are adopted according to the different needs of their relationship. There are three types of relationship: temporal relations, concurrency relations, and synergistic relations. (1) Temporal relations: alarms that are triggered by an intrusion event that satisfies a temporal relationship can be considered to satisfy a temporal relationship. (2) Concurrency relations: alarms that are triggered in the same period. That alarms' relationship is called concurrency relationship. (3) Synergistic relations: There are multiple attacks from different attackers or attack sources. There is cooperation between attacks to achieve some attack intention. There are synergies between the alerts triggered by the attacks. These attacks or concurrency for a common goal, or the existence of a sequence and dependencies. According to the above characteristics, we learn from the general process of JDI model shown in Fig.1. But this model is not complete; it can't be applied to the actual information fusion. Such as the process of merge single-to-multiple alerts and merge multiple-to-single alerts, the same source IP may contain multiple attack information, if the attackers use many meaningless attacks to hide the true attack, we can't find them. Intrusion detection systems also have many false positive, so we need to handle that carefully. That disadvantages can make any help for our next analysis. We need to make some change for the model. In machine learning, support vector machines (SVMs, also support vector networks [4]) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked for

belonging to one of two categories; an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. We use some open source intrusion detection system. Such as Snort IDS, Bro IDS, and OSSEC. Snort is a free and open source network intrusion prevention system (NIPS) and network intrusion detection system (NIDS) [5] created by Martin Roesch in 1998 [6]. Snort is now developed by Sourcefire, of which Roesch is the founder and CTO [7], and which has been owned by Cisco since 2013[8][9].

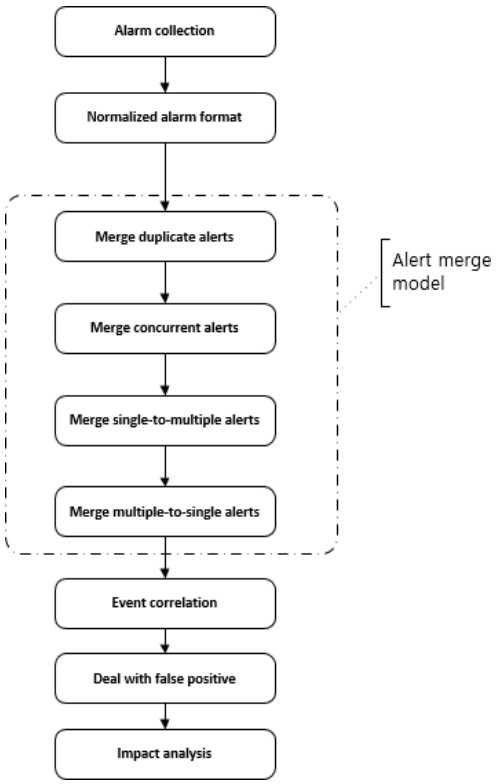

Figure 1: General Process Of JDI Model

Bro While focusing on network security monitoring, Bro provides a comprehensive platform for more general network traffic analysis as well. Well ground in more than 15 years of research, Bro has successfully bridged the traditional gap between academia and operations since its inception. Today, it is relied upon operationally in particular by many scientific environments for securing their cyberinfrastructure. Bro's user community includes major universities, research labs, supercomputing centers, and open-science communities [13]. OSSEC is a free, open-source host-based intrusion detection system (HIDS). It performs log analysis, integrity checking, Windows registry monitoring, rootkit detection, time-based alerting, and active response. It provides intrusion detection for most operating systems, including Linux, OpenBSD, FreeBSD, OS X, Solaris, and Windows. OSSEC has a centralized, cross-platform architecture allowing multiple systems to be easily monitored and managed [14]. For example, when we handle the process of alert fusion, if the false positive is present in the alert data, the result of our fusion will be poor, and the result will be unreliable. We need a mechanism to reduce this condition. According to the characteristics of SVM, we use it to filter the false positive in the alert database. The data set will be divided into two parts: training data set and test dataset. Training data set is an attack-free data set. In training data set we only need four features: SIP (source IP address), SPT (source port), DIP (destination IP address), DPT (destination port). Training dataset's structure is shown in below: T = {SIP, SPT, DIP, DPT} When we replay the test dataset, we got an alert database. Table.1 gives the attributes in the alert database and their meanings. Our proposed model uses SVM algorithm to make up for the disadvantage of the model. We will talk about that in next chapter. Every record in the alert database is shown in the following format: R = {ST, ET, IID, AID, AC, SIP, SPT, DIP, DPT, TS}

3. PROPOSED METHOD

The following describes the implementation of the main components of our alert fusion system based on our model shown in figure 2. The processing flow is as follows: (1) Replay the DARPA dataset with the multi intrusion detection systems (Snort, Bro, OSSEC). (2) Normalized alert's format using Table.1. (3) Training SVM Classifier with the attack-free dataset. (4) Merge duplicate alerts. (5) Merge concurrent alerts. (6) Merge single-to-multiple alerts using the SVM classifier to filter the false positive alerts. (7) Merge multiple-to-single alerts using the SVM classifier to filter the false positive alerts. (8) Put all data into alert correlation database The alert merge model process in our proposed model is shown below: Merge

duplicate alerts: In this step, we focused on processing of duplicate alerts from the same intrusion detection systems.

$$A_{new} = \{ST_{new}, ET_{new}, IID_{new}, AID_{new}, AC_{new}, SIP_{new}, SPT_{new}, DIP_{new}, DPT_{new}, TS_{new}\}$$

$$A_a = \{ST_a, ET_a, IID_a, AID_a, AC_a, SIP_a, SPT_a, DIP_a, DPT_a, TS_a\}$$

$$A_b = \{ST_b, ET_b, IID_b, AID_b, AC_b, SIP_b, SPT_b, DIP_b, DPT_b, TS_b\}$$

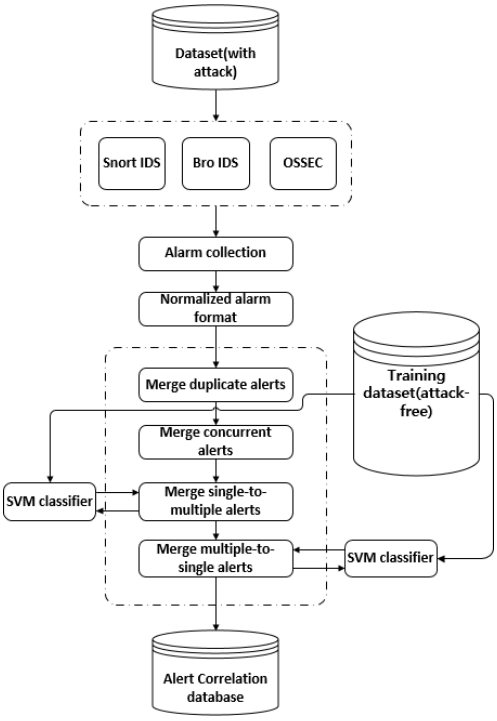

Figure.2 Proposed Model

We set the time window is 120 seconds, then we can get some scenes in this step. If we found another scene with the same scene structure and the same SIP and SPT, we can merge that scene. In this condition, all alerts came from the same intrusion detection system, so they all have the same IID, so we don't need to change that.

Merge concurrent alerts: We focused on processing of concurrent alerts from the different intrusion detection systems. In this step, we need to find out the alerts that alerted by the different intrusion detection systems at the same time. We don't need the SVM algorithm in this step. It can be easily processed with the script program.

Defined new alert Anew, Alert a (Aa) and Alert b (Ab). Use the format we mentioned earlier. Then we can get:

$$A_a = \{ST_a, ET_a, IID_a, AID_a, AC_a, SIP_a, SPT_a, DIP_a, DPT_a, TS_a\}$$

$$A_b = \{ST_b, ET_b, IID_b, AID_b, AC_b, SIP_b, SPT_b, DIP_b, DPT_b, TS_b\}$$

In this condition, all alerts came from different intrusion detection system, so they don't have the same IID, so we need to change that. For example, = can be changed like = {,, but not the sum of value. Merge single-to-multiple alerts: In this part, we focused on processing of single-to-multiple alerts from the intrusion detection systems. First, we set a Time window (120 seconds); the connection for the alerts are at least two connections. This step and next step we use SVM algorithm to get the classification boundary, because, one scene may contain different types of attack. The experimental results will show the advantage for this improvement. Define new alert , Alert a ( ) and Alert b ( ) form database. And define a scene use the format we mentioned earlier. This phase is characterized by some alerts containing the same source IP address and port, but with different destination IP addresses and ports.

The fusion results of the original model and the proposed model are shown in Table. 4 and Table. 5. We can compare the fusion ability of the two models in the Figure. 5 and Figure. 6. Obviously, our model's fusion ability is better than the original model.

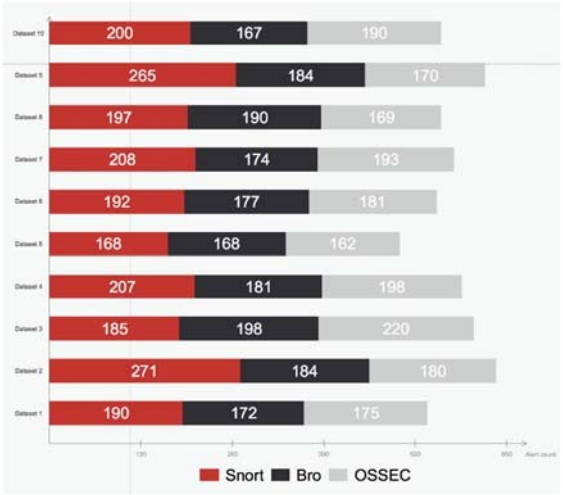

Figure.3 The Alert Collection Results (10 Data Sets)

4. CONCLUSION

This paper presents a new alert correlation structure and uses SVM to filter and optimize the false positive of the IDS systems, which greatly improves the system's fusion ability and the detection accuracy. The simulation was done against DARPA data set and shows the performance of proposed approach with an improvement in false positive rate. We use the average value to measure the ability of the fusion system, which can more accurately reflect the actual situation of the system.

REFERENCES:

[1] L. Trujillo and G. Olague, “Automated Design of Image Operators that Detect Interest Points,” Evol. Comput., vol. 16, no. 4, pp. 483–507, Dec. 2008.

[2] Bass T. Multisensor data fusion for next generation distributed intrusion detection systems[J]. 1999.

[3] Chandola, V.; Banerjee, A.; Kumar, V. (2009). "Anomaly detection: A survey". ACM Computing Surveys. 41 (3): 1–58.

[4] J. Haines, D.K. Ryder, L. Tinnel, and S. Taylor, “Validation of Sensor Alert Correlators,” IEEE Security and Privacy Magazine, vol. 1, no. 1, pp. 46-56, Jan. Feb. 2003. [5] Y. Bai, L. Zhuo, B. Cheng, and Y. F. Peng, “Surf feature extraction in encrypted domain,” in 2014 IEEE International Conference on Multimedia and Expo (ICME), 2014, pp. 1–6.

[5] K. Qureshi and M. Irfan, "Usability evaluation of e-learning applications, A case study of It’s Learning from a student’s perspective," Master Thesis, Blekinge Institute of Technology, 2009.

[6] G. D. Hong, C. S. Kim, and C. Lee, “LWIR HgCdTe Infrared Photodetector,” The Journal of Korea Navigation Institute, Vol. 14, No. 5, pp. 668-676, Oct. 2010.

[7] C. Kruegel and W. Robertson, “Alert Verification: Determining the Success of Intrusion Attempts,” Proc. First Workshop the Detection of Intrusions and Malware and Vulnerability Assessment (DIMVA 2004), July 2004.

[8] C. I. Kim and S. Y. Park, “Fast Stereo Matching of Feature Links,” in Visualization and Transmission 2011 International Conference on 3D Imaging, Modeling, Processing, 2011, pp. 268–274.

[9] J. Lin, D. Yan, X. Hu, Q. Xing, and B. Yang, “Dynamic programming algorithm for stereo correspondence of contour,” in 2012 5th International Congress on Image and Signal Processing, 2012, pp. 866–870.

[10] W. Xiaoli, Y. Lei, W. Lirong, and X. Jing, “Characteristic Point Match Algorithm Based on the SURF in Binocular Stereo Vision,” in 2012 Fifth International Conference on Intelligent Networks and Intelligent Systems, 2012, pp. 302–305.

[11] B. Zhang, Y. Jiao, Z. Ma, Y. Li, and J. Zhu, “An efficient image matching method using Speed Up Robust Features,” in 2014 IEEE International Conference on Mechatronics and Automation, 2014, pp. 553–558.

[12] E. Kiperwasser, O. David, and N. S. Netanyahu, “A Hybrid Genetic Approach for Stereo Matching,” in Proceedings of the 15th Annual Conference on Genetic and Evolutionary Computation, New York, NY, USA, 2013, pp. 1325–1332.

[13] L. Trujillo and G. Olague, “Automated Design of Image Operators that Detect Interest Points,” Evol. Comput., vol. 16, no. 4, pp. 483–507, Dec. 2008.

[14] 2018050001, S. Arya, D. M. Mount, “Approximate Nearest Neighbor Queries in Fixed Dimensions”, Open Journal, No. 1, May. 2018.

[15] 2018050002, Mohammad Alfraheed, “An Approach for Features Matching Between Bilateral Images of Streo Vision System Applied for Automated Heterogeneous Platoon”, Open Journal, No. 2, May. 2018.

[16] 2018050003, David G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints”, Open Journal, No. 3, May. 2018.

[17] 2018050004, J. DaiMiklos, A. Vasarhelyi, “Toward Blockchain-Based Accounting and Assurance”, Open Journal, No. 4, May. 2018.

[18] 2018050005, V. Feldman, E. Grigorescu, L. Reyzin, “Statistical Algorithms and a Lower Bound for Detecting Planted Cliques”, Open Journal, No. 5, May. 2018.

[19] 2018050006, C. Williams, A. Vrabie, “country R&D determinants of MNE entry strategy : A study of ownership in the automobile industry”, Open Journal, No. 6, May. 2018.

CONTRIBUTORS:

[1] 20180001, 차민준, 국민대학교
[2] 20170005, 변구훈, 네이버 Embeeded software P.D
[3] 20180015, 구민준, 배달의 민족 Software Engineer
[4] 20180025, 엄형근, 아마존
[5] 20180035, 김용태, 코봇