ERCIM "Alain Bensoussan" Fellowship Scientific Report

Fellow: Wei Wang
Visited Location : Q2S Centre, Norwegian University of Science and Technology (NTNU)
Duration of Visit: 12 months, 12/01/2009 – 11/01/2010

I - Scientific activity

During the period of the fellowship, I focused on two research topics: anomaly intrusion detection and network traffic measurement and diagnosis.

In the first topic, I extended my past work at INRIA France into an Autonomic Intrusion Detection System (AIDS) at NTNU. The AIDS could provide potential solutions to the following difficulties in the traditional intrusion detection methods. First, a large amount of precisely labeled data is very difficult to obtain in practice. The problem of unavailable labeled training data sets is a key roadblock to the construction of an effective anomaly IDS. Second, data for intrusion detection is typically streaming and dynamic. A practical solution is to keep the detection models always updated by incorporating new incoming labeled data as soon as which is classified during the detection. Quickly and manually labeling the data is difficult and it is thus quite expensive to frequently update or re-train an IDS with new clean labeled data. Third, many current anomaly detection approaches assume that the data distribution is stationary and the model is static accordingly. The static detection models have no ability to adapt to normal behavioral changes. We propose a novel framework of autonomic intrusion detection that fulfills online and adaptive intrusion detection over unlabeled audit data streams. The framework owns ability of self-managing: self-labeling, self-updating and self-adapting. Our framework uses the Affinity Propagation (AP) to learn a subject's behaviors through dynamical clustering of the streaming data. It automatically labels the data and adapts to normal behavior changes while identifies anomalies. The empirical results proved its effectiveness.

During my work at NTNU, for the first topic, I also evaluated the impact of attribute normalization of audit data on the intrusion detection performance. The empirical results on KDD'1999 data show that attribute normalization indeed improves the detection performance, especially for many distance based methods, e.g, k-NN. Statistical normalization is recommended for intrusion detection. The results, we believe, can also give some references for general classification tasks.

I concerned much on the second topic. I visited ETH Zurich and have discussed with several researchers regarding this topic. I have signed a contract with UNINETT to gain access to Netflow data and have done some preliminary experiments. Some more results are expected in our papers in the future.

During the fellowship period, I have also supervised two master students for the semester project (from September to December 2009). One topic is web anomaly detection and the other

is network measurement and diagnosis. I gave lots of instructions and I was happy to see that the two students have gained some results.

II- Publication(s) during your fellowship

1. Wang, Wei; Guyet, Thomas; Knapskog, Svein Johan. Autonomic Intrusion Detection System. In: Recent Advances in Intrusion Detection, *12th International Symposium (RAID 2009)*. Springer 2009 ISBN 978-3-642-04341-3. p. 359-361

   Abstract: We propose a novel framework of autonomic intrusion detection that fulfils online and adaptive intrusion detection in unlabeled audit data streams. The framework owns ability of self-managing: self-labelling, self-updating and self-adapting. Affinity Propagation (AP) uses the framework to learn a subject's behaviour through dynamical clustering of the streaming data. The testing results with a large real HTTP log stream demonstrate the effectiveness and efficiency of the method.

2. Wang, Wei; Zhang, Xiangliang; Gombault, Sylvain; Knapskog, Svein Johan. Attribute Normalization in Network Intrusion Detection. In: *Proceedings of 2009 10th International Symposium on Pervasive Systems, Algorithms, and Networks (ISPAN).* IEEE Computer Society 2009 ISBN 978-1-4244-5403-7. p. 448-453

   Abstract: Anomaly intrusion detection is an important issue in computer network security. As a step of data pre-processing, attribute normalization is essential to detection performance. However, many anomaly detection methods do not normalize attributes before training and detection. Few methods consider normalizing the attributes but the question of which normalization method is more effective still remains. In this paper, we introduce four different schemes of attribute normalization to pre-process the data for anomaly intrusion detection. Three methods, k-NN, PCA as well as SVM, are then employed on the normalized data for comparison of the detection results. KDD Cup 1999 data are used to evaluate the normalization schemes and the detection methods. The systematically evaluation results show that the process of attribute normalization improves a lot the detection performance. The statistical normalization scheme is the best choice for detection if the data set is large.

3. Wang, Wei; Zhang, Xiangliang; Sylvain, Gombault. Constructing Attribute Weights from Audit Data for Effective Intrusion Detection. *Journal of Systems and Software* 2009; Volume 82.(12) p. 1974-1981

   Abstract: Attributes construction and selection from audit data is the first and very important step for anomaly intrusion detection. In this paper, we present several cross frequency attribute weights to model user and program behaviours for anomaly intrusion detection. The frequency attribute weights include plain term frequency (TF) and various forms of term frequency-inverse document frequency (tfidf), referred to as Ltfidf, Mtfidf and LOGtfidf. Nearest Neighbor (NN) and k-NN methods with Euclidean and Cosine distance measures as well as principal component analysis (PCA) and Chi-square test method based on these frequency attribute weights are used for anomaly detection. Extensive experiments are performed based on command data from Schonlau et al. The testing results show that the LOGtfidf weight gives better detection performance compared with plain frequency and other types of weights. By using the LOGtfidf weight, the simple NN method and PCA method achieve the better masquerade detection results than the other 7 methods in the literature while the Chi-square test consistently returns the worst results.

The PCA method is suitable for fast intrusion detection because of its capability of reducing data dimensionality while NN and k-NN methods are suitable for detection of a small data set because of its no need of training process. A HTTP log data set collected in a real environment and the sendmail system call data from University of New Mexico (UNM) are used as well and the results also demonstrate the effectiveness of the LOGtfidf weight for anomaly intrusion detection.

4. Wei Wang, Florent Masseglia, Thomas Guyet, Rene Quiniou, Marie-Odile Cordier. A general framework for adaptive and online detection of web attacks. In: *Proceedings of the 18th International Conference on World Wide Web, WWW 2009*, p. 1141-1142, Madrid, Spain, April 20-24, 2009. ACM 2009, ISBN 978-1-60558-487-4.

   Abstract: Detection of web attacks is an important issue in current defence-in-depth security framework. In this paper, we propose a novel general framework for adaptive and online detection of web attacks. The general framework can be based on any online clustering methods. A detection model based on the framework is able to learn online and deal with "concept drift" in web audit data streams. A very large size of real HTTP Log data collected in our institute is used to validate the framework and the model. The preliminary testing results demonstrated its effectiveness.

III -Attended Seminars, Workshops, and Conferences

1. The 18th International Conference on World Wide Web, WWW 2009, Madrid, Spain, April 20-24, 2009
2. 1st International Workshop on Security and Communication Networks, IWSCN 2009, Trondheim, Norway, May 20-22, 2009
3. Norwegian Academy of Technological Sciences (NTVA) Seminar on Internet: Global Market – National Opportunities, Trondheim, Norway, September 9, 2009
4. The 12th International Symposium on Recent Advances in Intrusion Detection, RAID 2009, Saint-Malo, France, September 23-25, 2009
5. Norwegian information security conference - Norsk informasjonssikkerhetskonferanse, NISK 2009, Trondheim, Norway, November 23-25, 2009
6. The 10th International Symposium on Pervasive Systems, Algorithms, and Networks, I-SPAN 2009, Kaohsiung, Taiwan, December 14-16, 2009

IV – Research Exchange Programme (12 month scheme)

First Visit: One week at INRIA Nancy (LORIA), France (14-17 April 2009)
Project Team: MADYNES (Network and Security Management)
Contact Professor: Olivier Festor (firstname.lastname@inria.fr)

During my research exchange program at INRIA Nancy, I visited project team MADYNES. The research topics of the group are very close to mine. I made a presentation titled “Autonomic Intrusion Detection in Computer Security”. I would say the visit is very inspirable. During my visit, I have discussed the related topic with Prof. Festor. He gave me a lot of valuable suggestions regarding my paper. I have also discussed some related topics with team member Jérôme François and have gained much.

Second Visit: Two weeks at ETH Zurich, Switzerland (6 July -16 July 2009)
Project Team: Communication Systems Group
Contact Professor: Prof. Dr. Bernhard Plattner (lastname@tik.ee.ethz.ch)

During my research exchange program at ETH Zurich, I visited CSG group. I selected the group to visit because I would like to exchange some ideas for the second topic – Network traffic measurement and diagnosis. In fact, my goal has been achieved. At least four researchers are working on this topic in the group. I have made a presentation titled "Autonomic intrusion detection for web attacks" in the group seminar. I have deeply discussed the sampling methods for effective network traffic engineering with Bernhard Tellenbach. I have read several papers of the team members and have discussed with them. We also discussed potential collaborations in the future. I would say the visit is very meaningful for my current and future work.