International Journal of Engineering In Advanced Research Science and Technology
ISSN: 2278-2566
VOLUME-2
ISSUE-2

Alert Aggregation based Pattern Classifier for Effective Network Intrusion Detection system using KDD Dataset

[1]M.L.Manikanta[2]S. Jaganadham

[1]Asst.Professor of CSE, Rajiv Gandhi Institute of Management & Science, Sarparam, Kakinada, Andhra Pradesh
[2]M.Sc student, Rajiv Gandhi Institute of Management & Science, Sarparam, Kakinada, Andhra Pradesh

ABSTRACT:

Alert aggregation is an important subtask of intrusion detection. The goal is to identify and to cluster different alerts—produced by low-level intrusion detection systems, firewalls, etc.—belonging to a specific attack instance which has been initiated by an attacker at a certain point in time. Thus, meta-alerts can be generated for the clusters that contain all the relevant information whereas the amount of data (i.e., alerts) can be reduced substantially. We propose a novel technique for online alert aggregation which is based on a dynamic, probabilistic model of the current attack situation. Basically, it can be regarded as a data stream version of a maximum likelihood approach for the estimation of the model parameters.

Detecting intrusions in networks has become one of the most critical tasks to prevent their misuse by attackers. The rapid increase in network traffic and attacks made the Intrusion Detection Systems to fail in terms of accuracy and efficiency in many situations. We describe the problem of intrusion detection in detail and analyze various well known methods for intrusion detection with respect to two critical requirements viz. our proposed architecture and DARPA Dataset. Present networks and enterprises follow a layered defence approach to ensure security at different access levels by using a variety of tools such as network surveillance, perimeter access control, firewalls, network, host and application intrusion detection systems, data encryption and others. Given this traditional layered defence approach, only a single system is employed at every layer which is expected to detect attacks at that particular location. In this project an efficient way of finding intrusions has been proposed. The main goal of this approach in Intrusion Detection System is to achieve high accuracy and efficiency

INTRODUCTION:

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

How Data Mining Works?

While large-scale information technology has been evolving separate transaction and analytical systems, data mining provides the link between the two. Data mining software analyzes relationships and patterns in stored transaction data based on open-ended user queries. Several types of analytical software are available: statistical, machine learning, and neural networks. Generally, any of four types of relationships are sought:

- Classes: Stored data is used to locate data in predetermined groups. For example, a restaurant chain could mine customer purchase data to determine when customers visit and what they typically order. This information could be used to increase traffic by having daily specials.

- Clusters: Data items are grouped according to logical relationships or consumer preferences. For example, data can be mined to identify market segments or consumer affinities.

- Associations: Data can be mined to identify associations. The beer-diaper example is an example of associative mining.

- Sequential patterns: Data is mined to anticipate behavior patterns and trends. For example, an outdoor equipment retailer could predict the likelihood of a backpack being purchased based on a consumer's purchase of sleeping bags and hiking shoes.

Data mining consists of five major elements:

1) Extract, transform, and load transaction data onto the data warehouse system.
2) Store and manage the data in a multidimensional database system.
3) Provide data access to business analysts and information technology professionals.
4) Analyze the data by application software.
5) Present the data in a useful format, such as a graph or table.

Different levels of analysis are available:

- Artificial neural networks: Non-linear predictive models that learn through training and resemble biological neural networks in structure.
- Genetic algorithms: Optimization techniques that use process such as genetic combination, mutation, and natural selection in a design based on the concepts of natural evolution.
- Decision trees: Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID). CART and CHAID are decision tree techniques used for classification of a dataset. They provide a set of rules that you can apply to a new (unclassified) dataset to predict which records will have a given outcome. CART segments a dataset by creating 2-way splits while CHAID segments using chi square tests to create multi-way splits. CART typically requires less data preparation than CHAID.
- Nearest neighbor method: A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset (where k=1). Sometimes called the k-nearest neighbor technique.
- Rule induction: The extraction of useful if-then rules from data based on statistical significance.
- Data visualization: The visual interpretation of complex relationships in multidimensional data. Graphics tools are used to illustrate data relationships.

IMPLEMENTATION MODULES:

1. Attack Scenario and Model of the Adversary
2. Pattern Classification
3. Adversarial classification:
4. Security modules

MODULES DESCRIPTION:

Attack Scenario and Model of the Adversary:

Although the definition of attack scenarios is ultimately an application-specific issue, it is possible to give general guidelines that can help the designer of a pattern recognition system. Here we propose to specify the attack scenario in terms of a conceptual model of the adversary that encompasses, unifies, and extends different ideas from previous work. Our model is based on the assumption that the adversary acts rationally to attain a given goal, according to her knowledge of the classifier, and her capability of manipulating data. This allows one to derive the corresponding optimal attack strategy.

Pattern Classification:

Multimodal biometric systems for personal identity recognition have received great interest in the past few years. It has been shown that combining information coming from different biometric traits can overcome the limits and the weaknesses inherent in every individual biometric, resulting in a higher accuracy. Moreover, it is commonly believed that multimodal systems also improve security against Spoofing attacks, which consist of claiming a false identity and submitting at least one fake biometric trait to the system (e.g., a “gummy” fingerprint or a photograph of a user’s face). The reason is that, to evade multimodal system, one expects that the adversary should spoof all the corresponding biometric traits. In this application example, we show how the designer of a multimodal system can verify if this hypothesis holds, before deploying the system, by simulating spoofing attacks against each of the matchers.

Adversarial classification:

Assume that a classifier has to discriminate between legitimate and spam emails on the basis of their textual content, and that the bag-of-words feature representation has been chosen, with binary features denoting the occurrence of a given set of words

Security modules:

Intrusion detection systems analyze network traffic to prevent and detect malicious activities like intrusion attempts, ROC curves of the considered multimodal biometric system under a simulated spoof attack against the fingerprint or the face matcher. Port scans, and denial-of-service attacks. When suspected malicious traffic is detected, an alarm is raised by the IDS and subsequently handled by the system administrator. Two main kinds of IDSs exist: misuse detectors and anomaly-based ones. Misuse detectors match the analyzed network traffic against a database of signatures of known malicious activities. The main drawback is that they are not able to detect never-before-seen malicious activities, or even variants of known ones. To overcome this issue, anomaly-based detectors have been proposed. They build a statistical model of the normal traffic using machine learning techniques, usually one-class classifiers, and raise an alarm when anomalous traffic is detected. Their training set is constructed, and periodically updated to follow the changes of normal traffic, by collecting unsupervised network traffic during operation, assuming that it is normal (it can be filtered by a misuse detector, and should)

RESULT:

SCREEN SHOTS

REFERENCES

[1] R.N. Rodrigues, L.L. Ling, and V. Govindaraju, "Robustness of Multimodal Biometric Fusion Methods against Spoof Attacks," J. Visual Languages and Computing, vol. 20, no. 3, pp. 169-179, 2009.

[2] P. Johnson, B. Tan, and S. Schuckers, "Multimodal Fusion Vulnerability to Non-Zero Effort (Spoof) Imposters," Proc. IEEE Int'l Workshop Information Forensics and Security, pp. 1-5, 2010.

[3] P. Fogla, M. Sharif, R. Perdisci, O. Kolesnikov, and W. Lee, "Polymorphic Blending Attacks," Proc. 15th Conf. USENIX Security Symp., 2006.

[4] G.L. Wittel and S.F. Wu, "On Attacking Statistical Spam Filters," Proc. First Conf. Email and Anti-Spam, 2004.

[5] D. Lowd and C. Meek, "Good Word Attacks on Statistical Spam Filters," Proc. Second Conf. Email and Anti-Spam, 2005.

[6] A. Kolcz and C.H. Teo, "Feature Weighting for Improved Classifier Robustness," Proc. Sixth Conf. Email and Anti-Spam, 2009.

[7] D.B. Skillicorn, "Adversarial Knowledge Discovery," IEEE Intelligent Systems, vol. 24, no. 6, Nov./Dec. 2009.

[8] D. Fetterly, "Adversarial Information Retrieval: The Manipulation of Web Content," ACM Computing Rev., 2007.

[9] R.O. Duda, P.E. Hart, and D.G. Stork, Pattern Classification. Wiley-Interscience Publication, 2000.

[10] N. Dalvi, P. Domingos, Mausam, S. Sanghai, and D. Verma, "Adversarial Classification," Proc. 10th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 99-108, 2004.

[11] M. Barreno, B. Nelson, R. Sears, A.D. Joseph, and J.D. Tygar, "Can Machine Learning be Secure?" Proc. ACM Symp. Information, Computer and Comm. Security (ASIACCS), pp. 16-25, 2006.

[12] A.A. C_ardenas and J.S. Baras, "Evaluation of Classifiers: Practical Considerations for Security Applications," Proc. AAAI Workshop Evaluation Methods for Machine Learning, 2006.

[13] P. Laskov and R. Lippmann, "Machine Learning in Adversarial Environments," Machine Learning, vol. 81, pp. 115-119, 2010.

[14] L. Huang, A.D. Joseph, B. Nelson, B. Rubinstein, and J.D. Tygar, "Adversarial Machine Learning," Proc. Fourth ACM Workshop Artificial Intelligence and Security, pp. 43-57, 2011.

[15] M. Barreno, B. Nelson, A. Joseph, and J. Tygar, "The Security of Machine Learning," Machine Learning, vol. 81, pp. 121-148, 2010.

[16] D. Lowd and C. Meek, "Adversarial Learning," Proc. 11th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 641-647, 2005.

[17] P. Laskov and M. Kloft, "A Framework for Quantitative Security Analysis of Machine Learning," Proc. Second ACM Workshop Security and Artificial Intelligence, pp. 1-4, 2009.

[18] NIPS Workshop Machine Learning in Adversarial Environments for Computer Security, http://mls-nips07.first.fraunhofer.de/, 2007.

[19] Dagstuhl Perspectives Workshop Mach. Learning Methods for Computer Sec., http://www.dagstuhl.de/12371/, 2012.

[20] A.M. Narasimhamurthy and L.I. Kuncheva, "A Framework for Generating Data to Simulate Changing Environments," Proc. 25th Conf. Proc. the 25th IASTED Int'l Multi-Conf.: Artificial Intelligence and Applications, pp. 415-420, 2007.

[21] S. Rizzi, "What-If Analysis," Encyclopedia of Database Systems, pp. 3525-3529, Springer, 2009.

[22] J. Newsome, B. Karp, and D. Song, "Paragraph: Thwarting Signature Learning by Training Maliciously," Proc. Ninth Int'l Conf. Recent Advances in Intrusion Detection, pp. 81-105, 2006.

[23] A. Globerson and S.T. Roweis, "Nightmare at Test Time: Robust Learning by Feature Deletion," Proc. 23rd Int'l Conf. Machine Learning, pp. 353-360, 2006.

[24] R. Perdisci, G. Gu, and W. Lee, "Using an Ensemble of One-Class SVM Classifiers to Harden Payload-Based Anomaly Detection Systems," Proc. Int'l Conf. Data Mining, pp. 488-498, 2006.

[25] S.P. Chung and A.K. Mok, "Advanced Allergy attacks: Does a Corpus Really Help," Proc. 10th Int'l Conf. Recent Advances in Intrusion Detection (RAID '07), pp. 236-255, 2007.

[26] Z. Jorgensen, Y. Zhou, and M. Inge, "A Multiple Instance Learning Strategy for Combating Good Word Attacks on Spam Filters," J. Machine Learning Research, vol. 9, pp. 1115-1146, 2008.

[27] G.F. Cretu, A. Stavrou, M.E. Locasto, S.J. Stolfo, and A.D. Keromytis, "Casting out Demons: Sanitizing Training Data for Anomaly Sensors," Proc. IEEE Symp. Security and Privacy, pp. 81-95, 2008.

[28] B. Nelson, M. Barreno, F.J. Chi, A.D. Joseph, B.I.P. Rubinstein, U. Saini, C. Sutton, J.D. Tygar, and K. Xia, "Exploiting Machine Learning to Subvert Your Spam Filter," Proc. First Workshop Large-Scale Exploits and Emergent Threats, pp. 1-9, 2008.

[29] B.I. Rubinstein, B. Nelson, L. Huang, A.D. Joseph, S.-h. Lau, S. Rao, N. Taft, and J.D. Tygar, "Antidote: Understanding and Defending against Poisoning of Anomaly Detectors," Proc. Ninth ACM SIGCOMM Internet Measurement Conf. (IMC '09), pp. 1-14, 2009.

[30] M. Kloft and P. Laskov, "Online Anomaly Detection under Adversarial Impact," Proc. 13th Int'l Conf. Artificial Intelligence and Statistics, pp. 405-412, 2010.

[31] O. Dekel, O. Shamir, and L. Xiao, "Learning to Classify with Missing and Corrupted Features," Machine Learning, vol. 81, pp. 149- 178, 2010.

[32] B. Biggio, G. Fumera, and F. Roli, "Design of Robust Classifiers for Adversarial Environments," Proc. IEEE Int'l Conf. Systems, Man, and Cybernetics, pp. 977-982, 2011.

[33] B. Biggio, G. Fumera, and F. Roli, "Multiple Classifier Systems for Robust Classifier Design in Adversarial Environments," Int'l J. Machine Learning and Cybernetics, vol. 1, no. 1, pp. 27-41, 2010.

[34] B. Biggio, I. Corona, G. Fumera, G. Giacinto, and F. Roli, "Bagging Classifiers for Fighting Poisoning Attacks in Adversarial Environments," Proc. 10th Int'l Workshop Multiple Classifier Systems, pp. 350-359, 2011.

[35] B. Biggio, G. Fumera, F. Roli, and and L. Didaci, "Poisoning Adaptive Biometric Systems," Proc. Joint IAPR Int'l Conf. Structural, Syntactic, and Statistical Pattern Recognition, pp. 417-425, 2012.

[36] B. Biggio, B. Nelson, and P. Laskov, "Poisoning Attacks against Support Vector Machines," Proc. 29th Int'l Conf. Machine Learning, 2012.

[37] M. Kearns and M. Li, "Learning in the Presence of Malicious Errors," SIAM J. Computing, vol. 22, no. 4, pp. 807-837, 1993.

[38] A.A. C_ardenas, J.S. Baras, and K. Seamon, "A Framework for the Evaluation of Intrusion Detection Systems," Proc. IEEE Symp. Security and Privacy, pp. 63-77, 2006.

[39] B. Biggio, G. Fumera, and F. Roli, "Multiple Classifier Systems for Adversarial Classification Tasks," Proc. Eighth Int'l Workshop Multiple Classifier Systems, pp. 132-141, 2009.

[40] M. Br€uckner, C. Kanzow, and T. Scheffer, "Static Prediction Games for Adversarial Learning Problems," J. Machine Learning Research, vol. 13, pp. 2617-2654, 2012.

[41] A. Adler, "Vulnerabilities in Biometric Encryption Systems," Proc. Fifth Int'l Conf. Audio- and Video-Based Biometric Person Authentication, pp. 1100-1109, 2005.

[42] B. Efron and R.J. Tibshirani, An Introduction to the Bootstrap. Chapman & Hall, 1993.

[43] H. Drucker, D. Wu, and V.N. Vapnik, "Support Vector Machines for Spam Categorization," IEEE Trans. Neural Networks, vol. 10, no. 5, pp. 1048-1054, Sept. 1999.

[44] F. Sebastiani, "Machine Learning in Automated Text Categorization," ACM Computing Surveys, vol. 34, pp. 1-47, 2002.

[45] C.-C. Chang, C.-J. Lin, "LibSVM: A Library for Support Vector Machines," http://www.csie.ntu.edu.tw/~cjlin/libsvm/, 2001.

[46] K. Nandakumar, Y. Chen, S.C. Dass, and A. Jain, "Likelihood Ratio-Based Biometric Score Fusion," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 30, no. 2, pp. 342-347, Feb. 2008.

[47] B. Biggio, Z. Akhtar, G. Fumera, G. Marcialis, and F. Roli, "Robustness of Multi-Modal Biometric Verification Systems under Realistic Spoofing Attacks," Proc. Int'l Joint Conf. Biometrics, pp. 1-6, 2011.

[48] B. Biggio, Z. Akhtar, G. Fumera, G.L. Marcialis, and F. Roli, "Security Evaluation of Biometric Authentication Systems under Real Spoofing Attacks," IET Biometrics, vol. 1, no. 1, pp. 11-24, 2012.

[49] K. Wang and S.J. Stolfo, "Anomalous Payload-Based Network Intrusion Detection," Proc. Seventh Symp. Recent Advances in Intrusion Detection (RAID), pp. 203-222, 2004.

[50] B. Sch€olkopf, A.J. Smola, R.C. Williamson, and P.L. Bartlett, "New Support Vector Algorithms," Neural Computation, vol. 12, no. 5, pp. 1207-1245, 2000.

[51] K. Ingham and H. Inoue, "Comparing Anomaly Detection Techniques for http," Proc. 10th Int'l Conf. Recent Advances in Intrusion Detection, pp. 42-62, 2007.

[52] D. Sculley, G. Wachman, and C.E. Brodley, "Spam Filtering Using Inexact String Matching in Explicit Feature Space with on-Line Linear Classifiers," Proc. 15th Text Retrieval Conf., 2006.

[53] Encyclopedia of Biometrics, S.Z. Li, and A.K. Jain, eds., Springer US, 2009.

[54] B. Biggio, G. Fumera, and F. Roli, "Adversarial Pattern Classification Using Multiple Classifiers and Randomisation," Proc. Joint IAPR Int'l Workshop Structural, Syntactic, and Statistical Pattern Recognition, pp. 500-509, 2008.