IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

IJARSCT

Impact Factor: 6.252

Volume 2, Issue 2, July 2022

Anomaly Detection for Web Log Data Analysis

Madhur Narwan[1] and Ritu Kadyan[2]

Student, Department of Computer Science Engineering[1]

Assistant Professor, Department of Computer Science Engineering[2]

School of Engineering & Technology, Soldha, Bahadurgarh, Haryana, India

Abstract: *Many methods have been developed to protect web servers against attacks. Anomaly detection methods rely on generic user models and application behaviour, which interpret departures as indications of potentially dangerous behaviour from the established pattern. In this report, we conducted the use of a systematic review of the anomaly detection methods to prevent and identify web assaults; in particular, we utilised Kitchen ham's standard approach for conducting a organized analysis of literature in the computer science area. Logs that record system abnormal states (anomaly logs) can be regarded as outliers, and the improved PCA algorithm has relatively high accuracy in outlier detection methods. Therefore, we use improved algorithm to detect anomalies in the log data. However, there are some problems when using the improved PCA algorithm to detect anomalies, three of which are: excessive vector dimension leads to inefficient kNN algorithm, unlabeled log data cannot support the kNN algorithm, and the imbalance of the number of log data distorts the classification decision of kNN algorithm. In order to solve these three problems, we propose an efficient log anomaly detection method based on an improved PCA algorithm with an automatically labeled sample set. This method first proposes a log parsing method based on N-gram and frequent pattern mining (FPM) method, which reduces the dimension of the log vector converted with Term frequency. Inverse Document Frequency (TF-IDF) technology. Then we use clustering and self-training method to get labeled log data sample set from historical logs automatically. Finally, we improve the PCA algorithm using average weighting technology, which improves the accuracy of the PCA algorithm on unbalanced samples. The method in this article is validated on four log datasets with different types. The maximum recall rate & accuracy achieved for BGL dataset is 100 % & 97.62 % respectively. Similarly maximum F1-score achieved for Spirit dataset is 98.19 %. The accuracy, recall rate and F1-Score for Improved PCA Ensemble technique is 97.62 %, 100 % and 96.55 % for BGL/2 Log Set Data. Similarly, the accuracy, recall rate and F1-Score for Improved PCA Ensemble technique is 97.60 %, 98.79 % and 98.19 % respectively for Spirit/2 log set data.*

Keywords: Frequent Pattern Mining, PCA, KNN Algorithm

REFERENCES

[1]. Thang, T.M.; Nguyen, K.V. FDDA: A Framework For Fast Detecting Source Attack In Web Application DDoS Attack. In Proceedings of the Eighth International Symposium on Information and Communication Technology, NhaTrang, Vietnam, 7–8 December 2017; Association for Computing Machinery: New York, NY, USA, 2017; SoICT 2017; pp. 278–285.

[2]. Tripathi, N.; Hubballi, N. Slow Rate Denial of Service Attacks against HTTP/2 and Detection. Comput. Secur. 2018, 72, 255–272.

[3]. Najafabadi, M.M.; Khoshgoftaar, T.M.; Calvert, C.; Kemp, C. User Behavior Anomaly Detection for Application Layer DDoS Attacks. In Proceedings of the 2017 IEEE International Conference on Information Reuse and Integration (IRI), San Diego, CA, USA, 4–6 August 2017; pp. 154–161.

[4]. Zolotukhin, M.; Hämäläinen, T.; Kokkonen, T.; Siltanen, J. Increasing web service availability by detecting application-layer DDoS attacks in encrypted traffic. In Proceedings of the 2016 23rd International Conference on Telecommunications (ICT), Thessaloniki, Greece, 16–18 May 2016; pp. 1–6.

[5]. Shirani, P.; Azgomi, M.A.; Alrabaee, S. A method for intrusion detection in web services based on time series. In Proceedings of the 2015 IEEE 28th Canadian Conference on Electrical and Computer Engineering (CCECE), Halifax, NS, Canada, 3–6 May 2015; pp. 836–841.

DOI: 10.48175/568

[6]. Tripathi, N.; Hubballi, N.; Singh, Y. How SecureareWeb Servers? An Empirical Study of Slow HTTP DoS Attacks and Detection. In Proceedings of the 2016 11th International Conference on Availability, Reliability and Security (ARES), Salzburg, Austria, 31 August–2 September 2016; pp. 454–463.

[7]. Wang, C.; Miu, T.T.N.; Luo, X.;Wang, J. SkyShield: A Sketch-Based Defense System Against Application Layer DDoS Attacks. IEEE Trans. Inf. Forensics Secur. 2018, 13, 559–573.

[8]. Wang, Y.; Liu, L.; Si, C.; Sun, B. A novel approach for countering application layer DDoS attacks. In Proceedings of the 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, 25–26 March 2017; pp. 1814–1817.

[9]. Xie, Y.; Tang, S. Online Anomaly Detection Based on Web Usage Mining. In Proceedings of the 2012 IEEE 26th International Parallel and Distributed Processing SymposiumWorkshops PhD Forum, Shanghai, China, 21–25 May 2012; pp. 1177–1182.

[10]. Lin, H.; Cao, S.; Wu, J.; Cao, Z.; Wang, F. Identifying Application-Layer DDoS Attacks Based on Request Rhythm Matrices. IEEE Access 2019, 7, 164480–164491.

[11]. Xiao, R.; Su, J.; Du, X.; Jiang, J.; Lin, X.; Lin, L. SFAD: Toward effective anomaly detection based on session feature similarity. Knowl.-Based Syst. 2019, 165, 149–156.

[12]. Kozik, R.; Chora´s, M.; Hołubowicz,W. Evolutionary-based packets classification for anomaly detection in web layer. Secur. Commun. Netw. 2016, 9, 2901–2910.

[13]. Wang, L.; Cao, S.; Wan, L.; Wang, F. Web Anomaly Detection Based on Frequent Closed Episode Rules. In Proceedings of the 2017 IEEE Trustcom/BigDataSE/ICESS, Sydney, NSW, Australia, 1–4 August 2017; pp. 967–972.

[14]. Yuan, G.; Li, B.; Yao, Y.; Zhang, S. A deep learning enabled subspace spectral ensemble clustering approach for web anomaly detection. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 3896–3903.

[15]. Bronte, R.; Shahriar, H.; Haddad, H. Information Theoretic Anomaly Detection Framework for Web Application. In Proceedings of the 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC), Atlanta, GA, USA, 10–14 June 2016; Volume 2, pp. 394–399. [CrossRef]

[16]. Luo, Y.; Cheng, S.; Liu, C.; Jiang, F. PU Learning in Payload-based Web Anomaly Detection. In Proceedings of the 2018 Third International Conference on Security of Smart Cities, Industrial Control System and Communications (SSIC), Shanghai, China, 18–19 October 2018; pp. 1–5. [CrossRef]

[17]. Ren, X.; Hu, Y.; Kuang, W.; Souleymanou, M.B. A Web Attack Detection Technology Based on Bag of Words and Hidden Markov Model. In Proceedings of the 2018 IEEE 15th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), Chengdu, China, 9–12 October 2018; pp. 526–531.

[18]. Kozik, R.; Chora´s, M.; Hołubowicz,W. HardeningWeb Applications against SQL Injection Attacks Using Anomaly Detection Approach. In Image Processing & Communications Challenges 6; Chora´s, R.S., Ed.; Springer International Publishing: Cham, Switzerland, 2015; pp. 285–292.

[19]. Maggi, F.; Robertson,W.; Kruegel, C.; Vigna,G. Protecting aMoving Target: Addressing Web Application Concept Drift. In Recent Advances in Intrusion Detection; Kirda, E., Jha, S., Balzarotti, D., Eds.; Springer:Berlin/Heidelberg, Germany, 2009; pp. 21–40.

[20]. Valeur, F.; Vigna, G.; Kruegel, C.; Kirda, E. An Anomaly-Driven Reverse Proxy for Web Applications. In Proceedings of the 2006 ACM Symposium on Applied Computing, Dijon, France, 23–27 April 2006; Association for Computing Machinery: New York, NY, USA, 2006; pp. 361–368.

[21]. Guangmin, L. Modeling Unknown Web Attacks in Network Anomaly Detection. In Proceedings of the 2008 Third International Conference on Convergence and Hybrid Information Technology, Busan, Korea, 11–13 November 2008; Volume 2, pp. 112–116.

[22]. Yu, S.; Guo, S.; Stojmenovic, I. Fool Me If You Can: Mimicking Attacks and Anti-Attacks in Cyberspace. IEEE Trans. Comput. 2015, 64, 139–151.

[23]. Sakib, M.N.; Huang, C. Using anomaly detection based techniques to detect HTTP-based botnet C traffic. In Proceedings of the 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, Malaysia, 22–27 May 2016; pp. 1–6.

[24]. Medvet, E.; Bartoli, A. On the Effects of Learning Set Corruption in Anomaly-Based Detection of Web Defacements. In Detection of Intrusions and Malware, and Vulnerability Assessment; Hämmerli, M.B., Sommer, R., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; pp. 60–78.

[25]. Davanzo, G.; Medvet, E.; Bartoli, A. Anomaly detection techniques for a web defacement monitoring service. Expert Syst. Appl. 2011, 38, 12521–12530.

[26]. Juvonen, A.; Sipola, T.; Hämäläinen, T. Online anomaly detection using dimensionality reduction techniques for HTTP log analysis. Comput. Netw. 2015, 91, 46–56.

[27]. Wang,W.; Guyet, T.; Quiniou, R.; Cordier, M.O.; Masseglia, F.; Zhang, X. Autonomic Intrusion Detection. Know.-Based Syst. 2014, 70, 103–117.

[28]. Vartouni, A.M.; Kashi, S.S.; Teshnehlab, M. An anomaly detection method to detect web attacks using Stacked Auto-Encoder. In Proceedings of the 2018 6th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS), Kerman, Iran, 28 February–2 March 2018; pp. 131–134.

[29]. Zolotukhin, M.; Hämäläinen, T.; Kokkonen, T.; Siltanen, J. Analysis of HTTP requests for anomaly detection of web attacks. In Proceedings of the 2014 World Ubiquitous Science Congress: 2014 IEEE 12th International Conference on Dependable, Autonomic and Secure Computing, DASC 2014, Dalian, China, 24–27 August 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 406–411.

[30]. Asselin, E.; Aguilar-Melchor, C.; Jakllari, G. Anomaly detection for web server log reduction: A simple yet efficient crawling based approach. In Proceedings of the 2016 IEEE Conference on Communications and Network Security (CNS), Philadelphia, PA, USA, 17–19 October 2016; pp. 586–590.

[31]. Zhang, S.; Li, B.; Li, J.; Zhang, M.; Chen, Y. A Novel Anomaly Detection Approach for Mitigating Web-Based Attacks Against Clouds. In Proceedings of the 2015 IEEE 2nd International Conference on Cyber Security and Cloud Computing (CSCloud), New York, NY, USA, 3–5 November 2015; IEEE Computer Society: Piscataway, NJ, USA, 2015; pp. 289–294.

[32]. Zhang, M.; Lu, S.; Xu, B. An Anomaly Detection Method Based on Multi-models to Detect Web Attacks. In Proceedings of the 2017 10th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China, 9–10 December 2017; Volume 2, pp. 404–409.

[33]. Parhizkar, E.; Abadi, M. OC-WAD: A one-class classifier ensemble approach for anomaly detection in web traffic. In Proceedings of the 2015 23rd Iranian Conference on Electrical Engineering, Tehran, Iran, 10–14 May 2015; pp. 631–636.

[34]. Kozik, R.; Choras, M. Adapting an Ensemble of One-Class Classifiers for aWeb-Layer Anomaly Detection System. In Proceedings of the 2015 10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, 3PGCIC, Krakow, Poland, 4–6 November 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 724–729.

[35]. Cao, Q.; Qiao, Y.; Lyu, Z. Machine learning to detect anomalies in web log analysis. In Proceedings of the 2017 3rd IEEE International Conference on Computer and Communications (ICCC), Chengdu, China, 13–16 December 2017; pp. 519–523.

[36]. Yu, J.; Tao, D.; Lin, Z. A hybrid web log based intrusion detection model. In Proceedings of the 2016 4th IEEE International Conference on Cloud Computing and Intelligence Systems, CCIS 2016, Beijing, China, 17–19 August 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 356–360.

[37]. Threepak, T.;Watcharapupong, A. Web attack detection using entropy-based analysis. In Proceedings of the International Conference on Information Networking, Phuket, Thailand, 10–12 February 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 244–247.

[38]. Swarnkar, M.; Hubballi, N. Rangegram: A novel payload based anomaly detection technique against web traffic. In Proceedings of the 2015 IEEE International Conference on Advanced Networks and Telecommuncations Systems (ANTS), Kolkata, India, 15–18 Decemnber 2015; pp. 1–6.

[39]. Xu, H.; Tao, L.; Lin, W.; Wu, Y.; Liu, J.; Wang, C. A model for website anomaly detection based on log analysis. In Proceedings of the 2014 IEEE 3rd International Conference on Cloud Computing andIntelligence Systems, Shenzhen, China, 27–29 November 2014; pp. 604–608.

[40]. Park, S.; Kim, M.; Lee, S. Anomaly Detection for HTTP Using Convolutional Autoencoders. IEEE Access

DOI: 10.48175/568

2018, 6, 70884–70901.

[41]. Chora´s, M.; Kozik, R. Machine learning techniques applied to detect cyberattacks on web applications. Log. J. IGPL 2014, 23, 45–56.

[42]. Tharshini, M.; Ragavinodini, M.; Senthilkumar, R. Access Log Anomaly Detection. In Proceedings of the 2017 Ninth International Conference on Advanced Computing (ICoAC), Chennai, India, 14–16 December 2017; pp. 375–381.

[43]. Kozik, R.; Chora´s, M.; Hołubowicz,W. Packets tokenization methods for web layer cyber security. Log. J. IGPL 2016, 25, 103–113.

[44]. Kamarudin, M.H.; Maple, C.; Watson, T.; Safa, N.S. A LogitBoost-Based Algorithm for Detecting Known and UnknownWeb Attacks. IEEE Access 2017, 5, 26190–26200.

[45]. Yu, Y.; Liu, G.; Yan, H.; Li, H.; Guan, H. Attention-Based Bi-LSTM Model for Anomalous HTTP Traffic Detection. In Proceedings of the 2018 15th International Conference on Service Systems and Service Management (ICSSSM), Hangzhou, China, 21–22 July 2018; pp. 1–6.

[46]. Nguyen, X.N.; Nguyen, D.T.; Vu, L.H. POCAD: A novel pay load-based one-class classifier for anomaly detection. In Proceedings of the 2016 3rd National Foundation for Science and Technology Development Conference on Information and Computer Science (NICS), Danang, Vietnam, 14–16 September 2016; pp. 74–79.

[47]. Lu, L.; Zhu, X.; Zhang, X.; Liu, J.; Bhuiyan, M.Z.A.; Cui, G. One Intrusion Detection Method Based On Uniformed Conditional Dynamic Mutual Information. In Proceedings of the 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), New York, NY, USA , 1–3 August 2018; pp. 1236–1241.

[48]. Moustafa, N.; Misra, G.; Slay, J. Generalized Outlier Gaussian Mixture technique based on Automated Association Features for Simulating and DetectingWeb Application Attacks. IEEE Trans. Sustain. Comput. 2018, 1.

[49]. Alrawashdeh, K.; Purdy, C. Fast Activation Function Approach for Deep Learning Based Online Anomaly Intrusion Detection. In Proceedings of the 2018 IEEE 4th International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing, (HPSC) and IEEE International Conference on Intelligent Data and Security (IDS), Omaha, NE, USA, 3–5 May 2018; pp. 5–13.

[50]. Kaur, R.; Bansal, M. Multidimensional attacks classification based on genetic algorithm and SVM. In Proceedings of the 2016 2nd International Conference on Next Generation Computing Technologies (NGCT), Dehradun, India, 14–16 October 2016; pp. 561–565.

[51]. Angiulli, F.; Argento, L.; Furfaro, A. Exploiting N-Gram Location for Intrusion Detection. In Proceedings of the 2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI), Vietrisul Mare, Italy, 9–11 November 2015; pp. 1093–1098.

[52]. Hiremagalore, S.; Barbará, D.; Fleck, D.; Powell, W.; Stavrou, A. transAD: An Anomaly Detection Network Intrusion Sensor for the Web. In Information Security; Chow, S.S.M., Camenisch, J., Hui, L.C.K., Yiu, S.M., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 477–489.

[53]. Favaretto, M.; Spolaor, R.; Conti, M.; Ferrante, M. You Surf so Strange Today: Anomaly Detection in Web Services via HMM and CTMC. In Green, Pervasive, and Cloud Computing; Au, M.H.A., Castiglione, A., Choo, K.K.R.;,Palmieri, F., Li, K.C., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 426–440.

[54]. Kozik, R.; Chorás, M. The http content segmentation method combined with adaboost classifier for web-layer anomaly detection system. Adv. Intell. Syst. Comput. 2017, 527, 555–563.

[55]. Kozik, R.; Chora´s, M.; Holubowicz, W.; Renk, R. Extreme Learning Machines for Web Layer Anomaly Detection. In Image Processing and Communications Challenges 8; Chora´s, R.S., Ed.; Springer International Publishing: Cham, Switzerland, 2017; pp. 226–233.