Towards Reliable Intrusion Detection in High Speed Networks

Eduardo K. Viegas[1], Altair O. Santin[1]

[1] PPGIA - Pontifical Catholic University of Parana, Curitiba (PUCPR)
Curitiba - Parana 80215-901 - Brazil

{e.viegas, santin}@ppgia.pucpr.br

Abstract. *Intrusion detection schemes must be able to detect intrusion attempts at a high network bandwidth, besides having to deal with the lack of realistic training/testing data, changes in network traffic behavior, unreliable classifications over time and adversarial settings. In this work a new intrusion detection model, namely reliable intrusion detection, is introduced, whose main characteristic is the usage of both batch and stream learning algorithms coupled together. The proposed model advances the state-of-the-art in intrusion detection, providing reliable detection even in the presence of network traffic behavior changes and lack of model updates. The work relevance was recognized in the publication of 5 top-tier journals, 10 international and national conference papers, and 1 registered patent, being cited almost 200 times in current works in the literature.*

1. Introduction

Over the last years, several works have applied machine learning (ML) techniques, mostly through pattern recognition schemes, for the detection of network-based attacks. In a pattern recognition scheme, the classification of intrusion attempts is, in general, achieved through a two-phase process: training and testing [R. Sommer and V. Paxson, 2010]. In the training phase, the classifier learns the environment behavior, as present in the training dataset, producing a model. Afterwards, in the testing phase, the model is evaluated regarding its accuracy using a test dataset, which is expected to represent the production environment behavior [C. Gates and C. Taylor, 2010].

However, on the other side, the network traffic behavior changes in a daily-basis, either due to the discovery of new attacks [R. Sommer and V. Paxson, 2010], or due to the offering of new services [E. K. Viegas et al. 2017-1]. In such context, due to the evolving behavior of such environment [E. K. Viegas et al. 2017-2], and the high network throughput [E. K. Viegas et al. 2019], the identification of network attacks becomes a challenging task, in which a designed detection mechanism can become out-of-date before they are even deployed in real-world (production) environments [E. K. Viegas et al. 2019]. This because network-based intrusion detection field presents several challenges to ML techniques when compared to other fields [E. K. Viegas et al. 2017-1]. Thereby, when a new ML-based approach is under development it must undergo through a more comprehensive evaluation. However, in general, the majority of works employs a traditional evaluation approach [E. K. Viegas et al. 2017-3], in which the accuracy rates measured in a single test dataset are assumed to be evidenced in production [E. K. Viegas et al. 2017-1], despite the challenges that networked environments present. In such a case, a ML-based scheme must be able to detect

intrusion attempts at a ***high network bandwidth***, besides having to deal with the ***lack of realistic training/testing data, not generalization capable models, changes in network behavior, unreliable classifications over time, and adversarial attack setting.***

1.1. Objective and Contributions

This work was motivated by the need of a reliable intrusion detection model able to deal with the aforementioned challenges of production environments. To tackle these challenges, this work introduced a new intrusion detection model, namely reliable intrusion detection, whose main characteristic is the usage of both batch and stream learning algorithms coupled together. The model exploits the characteristics of each type of learner in a cascade pipeline to overcome the challenges of high-speed networks. Batch learning schemes are designed in such a way, that they provide reliable classifications over time and are able to generalize the behavior from the training dataset in the model. On the other hand, the used stream learning detection schemes are built to be resilient to adversarial attacks to hinder attacks over the designed system. Finally, batch and stream learning algorithm are coupled together to provide classification reliability over time, while also reliably adapting to network traffic behavior changes. The research provided the following contributions:

i. An approach named BigFlow, which performs reliable and near real-time network traffic measurement and classification in the Big Data context;
ii. A tool-based method that produces real and valid network traffic in a controlled and reproducible environment for creating intrusion datasets to evaluate both batch and stream learning intrusion detection schemes;
iii. An intrusion dataset with real and labeled network traffic, based on MAWI database, built by analyzing over 10 years of real network traces, composed by more than 30 TB of data and 30 billion network flows;
iv. A new and fine-grained evaluation method for ML intrusion detection;
v. A new multi-objective feature selection method that improves the generalization capacity of batch learning schemes, by considering the network properties;
vi. A new rejection method that provides classification reliability even when facing unknown network traffic behavior;
vii. A new approach to reliably deal with evolving network data streams to perform anomaly-based intrusion detection, in the presence of an adversary;
viii. A classification reliability assessment method through a conformal evaluator module, that provides a reliability degree while facing new network traffic behavior even in the absence of model updates. The model assesses the classifier confidence according to the behavior seen in the training dataset;
ix. A new reliable intrusion detection mechanism made of both batch and stream learning algorithms, providing classification reliability and ongoing updated classification models with minimal human assistance.

1.2. Publications

The impact of this work can be evidenced in the achieved publications, resulted in the following journals publications:

1. Eduardo Viegas; Altair Santin; Alysson Bessani; Nuno Neves. *BigFlow: Real-time and reliable anomaly-based intrusion detection for high-speed networks*. ***Elsevier Future Generation Computer Systems***, 2019. Qualis A2. IF 6.125

2. Eduardo Viegas; Altair Santin; Luiz Oliveira; André França; Ricardo Jasinski; Volnei Pedroni. *A reliable and energy-efficient classifier combination scheme for intrusion detection in embedded systems*. ***Elsevier Computers & Security***, 2018. Qualis A2. IF 3.579;
3. Eduardo Viegas; Altair Santin; Luiz Oliveira. *Toward a reliable anomaly-based intrusion detection in real-world environments*. **Elsevier Computer Networks**, 2017. Qualis A1. IF 3.311;
4. Eduardo Viegas; Altair S.; André F.; Ricardo J.; Volnei P.; Luiz O.. *Towards an Energy-Efficient Anomaly-Based Intrusion Detection Engine for Embedded System*s. **IEEE Transactions on Computers**, 2017. Qualis A1. IF 3.052;
5. Cleverton V.; Altair S.; Eduardo Viegas; Vilmar A. *SDN-based and multitenant-aware resource provisioning mechanism for cloud-based big data streaming*. ***Journal of Network and Computer Applications***, 2019. Qualis A2. IF 5.570;

Besides the aforementioned journals, this work has also resulted in the following conferences publications related to the work:

1. Eduardo Viegas; Santin, A. O. ; Cogo, V. ; Abreu, V. . A Reliable Semi-Supervised Intrusion Detection Model: One Year of Network Traffic Anomalies. **International Conference on Communications (ICC), 2020**, Qualis A1
2. Eduardo Viegas.; Santin, A. O. ; Cogo, V.; Abreu V. . Facing the Unknown: A Stream Learning Intrusion Detection System for Reliable Model Updates. **Advanced Information Networking and Applications**, **2020**, Qualis A2;
3. Santos, R. ; Viegas, Eduardo K. ; Santin, A. O. ; Cogo, V. V. . A Long-Lasting Reinforcement Learning Intrusion Detection Model. **Advanced Information Networking and Applications (AINA)**, **2020**, Qualis A2;
4. Sanz, I. J. ; Lopez, M. A. ; Viegas, Eduardo Kugler ; Sanches, V. . A Lightweight Network-based Android Malware Detection System. **IFIP Networking, 2020**, Qualis A2
5. Eduardo Viegas; Altair Santin; Vilmar Abreu; Luiz Oliveira. *Enabling Anomaly-based Intrusion Detection Through Model Generalization*. ***IEEE Symposium on Computers and Communications***, 2018. Qualis A2;
6. Eduardo Viegas; Altair S.; Nuno N.; Alysson B.; Vilmar A.. *A Resilient Stream Learning Intrusion Detection Mechanism for Real-Time Analysis of Network Traffic*. ***Global Communications Conference (GLOBECOM),*** 2017. Qualis A1;
7. Eduardo Viegas; Altair S.; Vilmar A.; Luiz O.. *Detecção de Intrusão Através de Aprendizagem de Fluxo no Ambiente do Adversário*. ***SBSeg***, 2017. Qualis B3;
8. Eduardo Viegas; Altair Santin; Vilmar Abreu; Luiz Oliveira. *Stream learning and anomaly-based intrusion detection in the adversarial settings*. ***IEEE Symposium on Computers and Communications***, 2017. Qualis A2;
9. Vilmar A.; Altair S.; Eduardo Viegas; Maicon S.. *A multi-domain role activation model*. ***International Conference on Communications (ICC)***, 2017. Qualis A1;
10. Cleverton V.; Altair S.; Eduardo Viegas; Vilmar A.. *A Machine Learning Auditing Model for Detection of Multi-Tenancy Issues Within Tenant Domain*. CCGRID, 2018. Qualis A1;

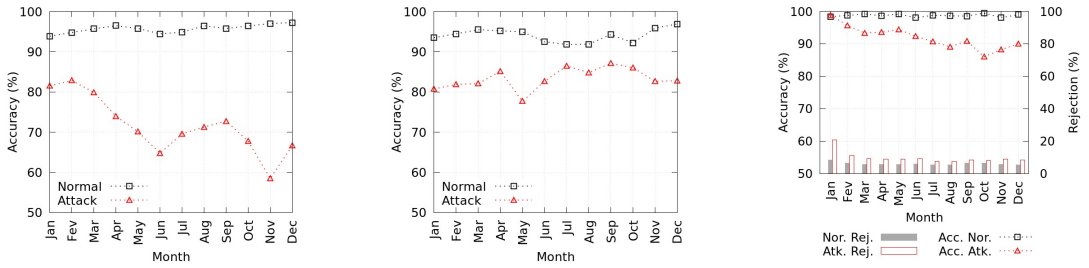
A patent was also registered regarding the reliable intrusion detection model as a product, including all of the listed contributions registered as:

- Eduardo Viegas; Altair Santin. *MECANISMO DE DETECÇÃO DE INTRUSÃO CONFIÁVEL BASEADA EM MACHINE LEARNING EM REDES DE ALTA VELOCIDADE*. 2018, Brazil. Patent. Register Number: BR10201801101;

The main impact of this work is the in-depth evaluation and design of novel intrusion detection models aiming the reliability of classifications. In this work, the reliability of current intrusion detection techniques is extensively evaluated, and the results shows that the state-of-the-art is unable to provide reliable intrusion detection. In the following sections, the main results obtained in this work is presented, regarding each of the problem it solves in the intrusion detection field.

2. Dealing with network traffic behavior changes in high-speed networks

The behavior of network traffic changes over time, either due to new types of malicious actions or alterations in the transmitted content (e.g., due to the offering of new services [E. K. Viegas et al. 2017-1]), the attack models require constant revision. Consequently, the model's accuracy observed on the training dataset might not be evidenced on unseen data. In such a case, the intrusion detection engine will no longer be trusted by the operator given that the alarms are not generated as expected [E. K. Viegas et al. 2018]. In this work, we have assessed this accuracy loss experimentally, using a real network traffic dataset spanning a year and several ML classifiers. Figure 1-a shows that the accuracy of a Random Forest classifier trained in the beginning of the year can decrease by up to 23% during the year. In addition, when model updates are performed weekly using the same classifier, the accuracy does not significantly drops, as shown in Figure 1-b. However, to perform such updates periodically is not a feasible task in high-speed networks.



(a) *No-update* random forest classifier

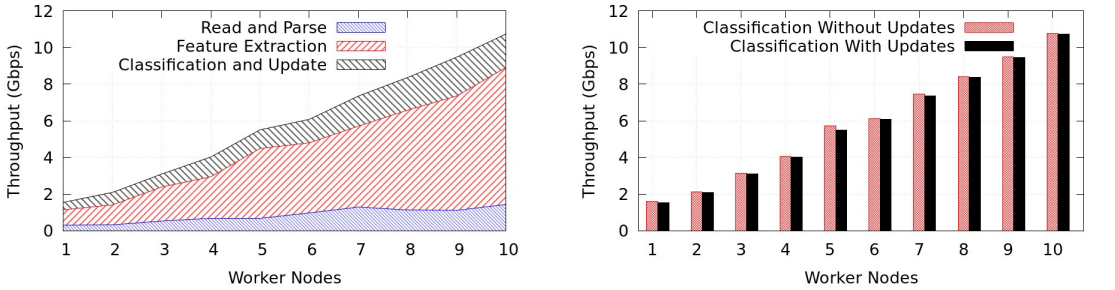
(b) *Weekly-update* random forest classifier

(c) BigFlow with OzaBoosting with verifier module

Figure 1. (a) Random Forest classifier accuracy behavior over time without model updates; (b) Random Forest classifier accuracy behavior over time with weekly model updates; (c) Proposed technique accuracy behavior over time [E. K. Viegas et al. 2019].

Therefore, to address network traffic behavior changes in high-speed networks, this work have designed BigFlow, a system for reliable real-time network traffic classification in high-speed networks. The proposal is based on two main insights. First, BigFlow determines whether the classification outcome should be accepted or not, in contrast to traditional approaches, which always classify events as normal or attack. The purpose is to make the administrator aware that a possible change has occurred in the network traffic behavior. In this sense, when an event is rejected, there is a high probability that a new network traffic behavior is taking place. Although classification

rejection has been used in other areas (e.g., for optical character recognition (OCR) or medical diagnosis), in these areas contextual information can help to identify pattern deviations; however, in the high-speed network traffic field, such a task is challenging. The main challenge that is not present in other areas relates to rejections based on the classifier confidence. This is because a classifier may become unreliable when facing unseen network traffic behavior, thereby committing classification mistakes with high confidence. The second insight relates to the fact that BigFlow employs stream learning techniques to analyze traffic in near real time. Such techniques support incremental model updates based on the rejected instances. The expectation is that after a period (e.g., within one week), the rejected event is properly classified by an expert or a tool (e.g. signature-based network-based intrusion detection system - NIDS) based on public information (e.g., new indicators of compromise). A major advantage of this approach is that the incremental model updates, that incorporates new knowledge into the model, is based only on correctly classified events. This decreases the risk of inaccurate detections, which may lead to a high rate of false positives when processing further packets. Moreover, incremental model updates significantly decrease training time because the current model is not discarded, which is advantageous for high-speed networks. Rejecting low-confidence classifications in an NIDS – the key idea of BigFlow – has lead to two important benefits: better detection accuracy (i.e., fewer misclassifications) and the identification of new characteristics of the evolving traffic, which are then used to incrementally update the classifier model. These benefits improve BigFlow reliability over time, even if the network's traffic behavior changes, as shown in Figure 1-c. In addition, at the same time BigFlow significantly decreases the amount of computational and storage resources needed to operate the system. In combination, these techniques make BigFlow scalable with the number of nodes employed in the system (with a network traffic processing capacity of up to 10 Gbps in our experiments), without losing accuracy over time, as shown in Figure 2.



(a) *BigFlow* throughput performance according to the number of worker nodes.

(b) *BigFlow* throughput comparison with and without updates.

Figure 2. BigFlow scalability tests [E. K. Viegas et al. 2019].

3. The building of generalization capable and reliable intrusion detection models

Regardless of providing an updated intrusion detection model, the reliability of classification must be provided in the present time. In other words, despite the model being updated, unknown traffic behavior may occur in a period of time. This because it is not possible to train an intrusion detection model with all possible network traffic behavior variations. Therefore, the intrusion model must be able to generalize the behavior from the training dataset, as well as ensure the reliability when a classification is made.

3.1. Generalization capable models

The detection system must be able to generalize the behavior from the training dataset to other environments. In such a case, the generalization must take into account the classification of known, similar and new attacks; known, similar and new services, and their contents. In this manner, it becomes possible to ensure that the system can correctly classify the events regarding the time of its building, even from a limited training dataset.

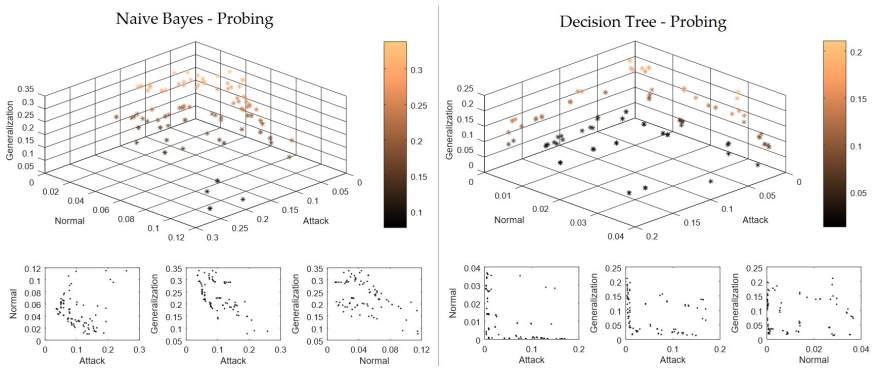


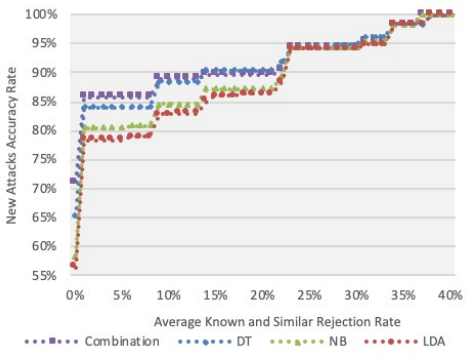
Figure 3. Building of generalization capable models by the means of feature selection technique [E. K. Viegas et al. 2017-1].

To address this situation, we proposed a new tool-based intrusion database creation method that is aimed to produce databases that can easily be updated, reproduce real and valid traffic, are representative, and are publicly available. Through the proposed intrusion database creation method, a new evaluation scheme specific to the machine learning intrusion detection field was presented. This scheme allowed each of the common assumptions in the literature to be validated, such as that new events and new services are detected. Finally, to provide a reliable intrusion detection system, the work presented and evaluated a multiple objective feature selection method. The evaluation approach allows a system administrator to establish the real capacity of a system for detecting each of the common properties in any production environment. Figure 3 shows the obtained results using the proposed multi-objective feature selection technique. The proposed approach improved the detection of all normal and attack variations when compared to traditional detection approaches. The proposed technique improved the detection rates when compared with traditional approaches.

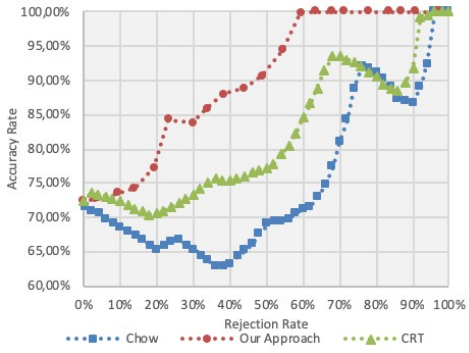
3.2. Reliability in model classifications

Despite having updated and generalization capable models, variations of network traffic behaviors might occur. Therefore, it is expected that an intrusion detection model to be reliable in its classifications. In general, a rejection technique may significantly improve the classification reliability. When the network traffic content changes, the rejection rate could be increased to maintain the classification accuracy stable. Moreover, when the rejection increases, this may indicate that the detection models should be updated. However, if this update is not possible, the intrusion detection alerts will continue to be reliable. The events that are not classified are said to be rejected by the classifier. When a classifier is operating, its accuracy depends on the feature values distribution being similar to that of the training dataset (usually composed of real network traffic). If the distribution changes significantly, the classifier model should be updated, or its accuracy may decrease. This update usually requires expert knowledge to label new

events and to rebuild the model, which may not be practical in real-world environments. To test a classifier designed to operate in such environments, our work assess whether it is still reliable even when the network traffic changes. Therefore, is this work an evaluation scenario is described and an event rejection method that allows the classifier to operate reliably even when it cannot be easily updated is proposed. To overcome the limitations of other works in the literature, in this work a rejection method that takes into account the frequent content changes observed in real-world network traffic was proposed. It was also proposed the usage of several independent classifiers using different machine-learning algorithms. After each classification, it checks whether there are enough similarities between the classifier outputs class (normal or attack) and the class occurrence observed in the training dataset. If there is not a predominant match, the classification is deemed unreliable and the event should be rejected because the features used to build the model and the current event are not similar enough for a reliable classification. An event rejection means that none of the classifiers can reliably assign a class to an input event; in this case, the event is rejected rather than being potentially incorrectly classified. To evaluate the proposed method, three traffic scenarios were used: a baseline scenario, a scenario with network traffic changes but similar to the baseline scenario, and a scenario with new attacks. The baseline scenario was used to obtain the rejection range thresholds and the attack models; the other scenarios were used to evaluate the rejection method. In the real world, it is not possible to choose a different set of thresholds for each event, because the classifier is unable to determine whether an event is a known attack, a similar attack, or a new one. Therefore, the choice of a set of thresholds was made taking into account the tradeoff between accuracy and rejection rate. Figure 4-a shows the accuracy-reject tradeoff between the accuracy in detecting new attacks and the rejection rate for known and similar attacks, using the same set of thresholds during the detection. The graph shows that it is possible to maintain the accuracy for the detection of new attacks, but at the cost of an increased rejection rate for known and similar attacks. For instance, it is possible to maintain the accuracy rate at 95% in a scenario with new attacks, at the cost of rejecting 31% in average of the events in the other two scenarios (known and similar attacks).



(a) Tradeoff between the accuracy improvement for new attacks and the rejection of known and similiar attacks.



(b) Accuracy-rejection tradeoff, for the combination classifier in the new attacks dataset, using the evaluated rejection techniques.

Figure 4. Providing reliability in intrusion detection [E. K. Viegas et al. 2018].

Finally, two commonly used rejection approaches that rely on class probabilities, the Chow's rule and the CRT, were compared to our proposed method. The three approaches – CRT, Chow and the proposed approach – were evaluated using the New

Attacks dataset. We used rejection rates from 0% to 100%. Fig. 4-b shows the accuracy-reject tradeoff comparison for the evaluated approaches. The proposed approach outperformed both existing techniques, CRT and Chow's rule. The tradition-al rejection approaches were not able to identify behavior changes and increased the classification confusion; the assigned class probabilities were high even for misclassified instances. In contrast, our approach was able to operate with fewer misclassifications in the presence of traffic behavior changes, reaching 100% accuracy while rejecting 60% of the events. Therefore, the proposed technique improves the current state-of-the-art for providing reliability in classifications.

4. Conclusion

This work has addressed each of the challenges of building reliable intrusion detection schemes by the means of machine learning techniques for production usage. To this end, the approach proposed in this work, namely reliable intrusion detection model, relies in the use of both batch and stream learning algorithms coped together, in which, each learner overcomes a specific challenge. In such a case, batch learning algorithms were designed and evaluated to deal with the ***lack of realistic training/testing data, not generalization capable models,*** and ***unreliable classifications over time.*** On the other hand, stream learning algorithms were used to address ***high network bandwidth***, ***changes in network behavior*** and ***adversarial attack setting***. Therefore, this work significantly advanced the state-of-the-art in intrusion detection. The knowledge produced in this work shows that current approaches for intrusion detection are unreliable. Nonetheless, the datasets created are being openly shared to the scientific community. As a consequence, this work was recognized in the publication of 5 top-tier journals, 10 international and national conference papers, and 1 registered patent, being cited almost 200 times in current works in the literature.

References

C. Gates and C. Taylor (2007). "Challenging the Anomaly Detection Paradigm: A Provocative Discussion," Proc. 2006 Work. New Secur. Paradig., pp. 21–29, 2007.

E. K. Viegas, A. O. Santin, and L. S. Oliveira (2017-1). "Toward a reliable anomaly-based intrusion detection in real-world environments," Comput. Networks, vol. 127.

E. K. Viegas, A. Santin, V. Abreu, and L. S. Oliveira (2017-2), "Stream learning and anomaly-based intrusion detection in the adversarial settings," in Proceedings - IEEE Symposium on Computers and Communications.

E. K. Viegas, A. Santin, N. Neves, and A. Bessani (2019). "BigFlow: Real-time and Reliable Anomaly-based Intrusion Detection for High-speed Networks". in Future Generation Computer System.

E. K. Viegas, A. Santin, N. Neves, A. Bessani, and V. Abreu (2017-3). "A Resilient Stream Learning Intrusion Detection Mechanism for Real-time Analysis of Network Traffic". In. proc. of IEEE GLOBECOM.

E. K. Viegas, A. Santin, L. S. Oliveira, A. França, R. Jasinki, and V. Pedroni (2018), "A reliable and Energy-Efficient Classifier Combination Scheme for Intrusion Detection in Embedded Systems". In: Computers & Security

P802.3cd (2017). P802.3cd Standard for Ethernet Amendment. Available at: http://ieeexplore.ieee.org/document/8115318/

R. Sommer and V. Paxson (2010). "Outside the Closed World: On Using Machine Learning for Network Intrusion Detection," 2010 IEEE Symp. Secur. Priv., vol. 0, no. May, pp. 305–316.