A Methodology for building a Dataset to Assess Intrusion Detection Systems in Wireless Networks

ED' WILSON TAVARES FERREIRA[1, 2], AILTON AKIRA SHINODA[1], RUY DE OLIVEIRA[2], VALTEMIR EMERENCIO NASCIMENTO[2], NELCILENO VIRGÍLIO DE SOUZA ARAÚJO[3]

[1] Departamento de Engenharia Elétrica, Universidade Estadual Paulista "Júlio de Mesquita Filho", Ilha Solteira, BRASIL
[2] Departamento de Área de Informática, Instituto Federal de Mato Grosso, Cuiabá, BRASIL
[3] Instituto de Computação, Universidade Federal de Mato Grosso, Cuiabá, BRASIL
edwilson.ferreira@ifmt.edu.br, shinoda@dee.feis.unesp.br, ruy@cba.ifmt.edu.br, valtemir.nascimento@cba.ifmt.edu.br, nelcileno@ic.ufmt.br

Abstract— this paper proposes building a dataset to be used in evaluation of Intrusion Detection Systems (IDS). We collected traffic in a real wireless network, processed such data and then evaluated IDS classification techniques with our processed data. Actually, we built a dataset to assess classification and pattern recognition standards. The outcome confirms that the built dataset may be deployed satisfactorily in evaluations IDS in wireless scenarios.

Keywords — Wireless LAN, Intrusion Detection, Dataset.

1 Introduction

People are getting used to technological gadgets such as smart phones and tablets with Internet access. Most of these devices are equipped with wireless capabilities based on the IEEE 802.11 standard. Using such wireless networks, users are usually able to get Internet access much cheaper than they would by using cell phone networks.

The ever increasing number of users carrying on financial transactions through the Internet to either access bank systems or conduct online shopping has attract bad guys attention toward attacking the global network.

In Brazil, the Center for Studies, response and handling of security incidents (CERT.br) registered 352,925 cases of security incidents in 2013. This represents a reduction of 24% compared to the previous year [1]. Figure 1 shows that from 1999 to 2014 the number of registered incidents has increased significantly, despite drops in a few years in between.

These security problems include many sort of incidents, like fraud attempts and brute force attacks on both SSH and content servers.

Confidentiality, integrity and availability are essential features for information security. Any action that compromise such features in a given system is called intrusion. The Intrusion Detection System (IDS) must be able to identify bad actions inside the network without impacting the normal system operation. Like antivirus and firewall, an IDS is a security tool toward strengthening the information security in communication systems [2].

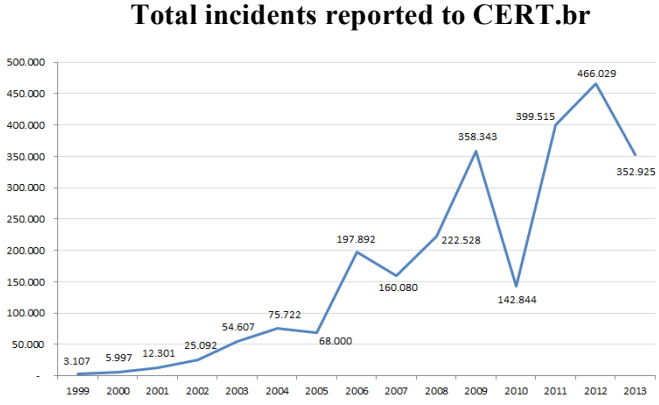

Fig. 1. Statistic Incidents Reported to CERT.br

Depending on the approach used to detect suspicious activities, the IDS may be classified in two categories: anomaly-based detection and signature-based detection. The former keeps track of the activities in the network to detect effective deviation from a considered normal behavior. The latter consist of searching known attack profiles.

Comparing these both categories, one can say that a disadvantage of the anomaly-based approach is the high number of false positive alarms, and that the signature-base one demands prior knowledge of the attacks profiles. Concerning the advantages, the former approach is able to detect unknown attacks, and the latter is a low computing-intensive method.

IDS are used to monitor, evaluate and inform security violations that may be intentional or not. Yet, detection and prevention techniques do not advance in the same pace, which makes it difficult to bring them together. This arises confusion in understanding the detection methodologies in recent systems [3].

Wireless networks are susceptive to various types of attacks. Because of that, several extensions have been proposed to IEEE 802.11, aiming at reducing or eliminating such deficiencies [4]. And distinct approaches have been proposed to IDS [4]–[6]. Nonetheless, as there are many diverse possibilities for the sort of topology, number of users and kind of interferences in the wireless signal, it is not trivial to compare the existing IDS mechanisms.

In order to compare IDS, one can use either a dataset, built from data captured from a given network, or data coming from simulations. When a dataset is used in the evaluation it plays a key role in the validation of the methodology employed in the proposed IDS. A set of data with high quality allows us to assess the proposed ISD approach and its efficiency in the evaluated scenario. However, due to a lack of proper dataset, a great part of the researches on intrusion detection makes use of simulation data [7].

The various techniques developed in recent years have evolved substantially, leading typical IDS to reach high detection rates, up to 98%, with false positive rates as low as 1%. On the other hand, it has been hard to compare these new techniques, as stated in [8].

To compare the several existing IDS approaches, it is important to employ the same scenario for all evaluations. Nevertheless, this is not trivial as factors like user profiles, network topology, channel interference, obstacles, number of users, among others are really difficult to reproduce. Simulation might be an option [9], but it is only an approximation of the real scenario, as approached in [10], [11], thereby complicating the comparisons.

This paper proposes a dataset, generated by collecting data in a real wireless network, to be used for comparing wireless based IDS approaches. The key idea is to provide a methodology for doing this comparison in an as much accurate manner as possible.

The remainder of this paper is organized as follows. Section II discusses the main related work. Section III explains the methodology used in building the proposed dataset. In section IV, the dataset evaluation scenario is presented. Section VI shows the effectiveness of the dataset. And section VII brings the concluding remarks and potential future work.

2 Related Work

This section presents key related work in terms of construction and validation of datasets, as well as those related to detection systems for wireless networks.

Most existing intrusion detection approaches has been developed for wired networks, and these approaches uses several classifying mechanisms such as neural artificial networks [12]–[14], clustering [15]–[17] and genetic algorithms [18]–[20].

A hybrid approach in [6] makes use of information from MAC layer and upper layers to intrusion detection in wireless networks. This approach is used in the feature selection process. For this, the authors used the information gain measure and the well-known *k-means* classifier. They also used neural networks, based on the MLP (multilayer perceptron) in the IDS learning and test processes. The proposed system was projected to reduce the number of features needed for the correct IDS operation in a wireless environment.

Similarly, the work in [4] also uses the feature selection for IDS in wireless networks. The purpose in their work was also to create a self-learning mechanism to diminish the number of features needed for the correct IDS operation in a wireless environment. They used clustering through *k-means*, and for the detection they used neural artificial networks.

In a previous work [21], we proposed creating a hybrid IDS, in which we first conducted the feature selection and then the intrusion detection. For both mechanisms we used the Kappa-Fuzzy ARTMAP approach. Even though the evaluation results were good, the dataset in place had been collected in a wired network.

3 Methodology in Creating the dataset

The dataset was built with real data collected from the network traffic. As a result, the data represent properly typical wireless users behavior, that in this case were students and staff of the institution utilized in this experiment.

Aiming at improving IDS tests possibilities; we used two distinct scenarios, each one with its own configuration and topology, i.e, one represents a typical domestic application and the other, a little bit more complex, represents a corporative environment. For each of these scenarios, we got a dataset.

3.1 Scenario 1 – Dataset creation with WEP/WPA criptography

Even though WEP is an outdated protocol, due to mainly its security vulnerabilities, there are still a lot of networks using WEP [22]. Because of that, we decided to use this encryption protocol in the comparisons we conducted in evaluation different IDS techniques.

The topology for scenario 1 is shown in Fig. 2. It is a simple topology that represents typical domestic environments.

For creating the dataset, we made use of different forms of Denial of Service (DoS) attacks. Generally we worked with popular DoS attacks, and so even non expert people, using widely available tools, may perpetrate such disturbs. This sort of attack exploits vulnerabilities in the management frames to render the IEEE 802.11 services, using pre-RSN (Robust Security Network), unavailable.

In order to generate the ChopChop, Deauthentication, fragmentation and duration forms of attack, we used the Airplay [23] application in station 1. To collect the data we used the Wireshark [24] application in station 2, and the normal (without attack) data were generate by station 3 using applications based both HTTP and HTTPS protocols.

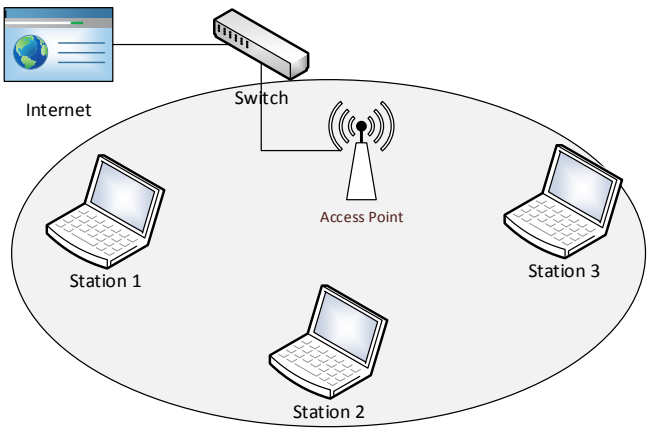

Fig. 2. Topology applied in WLAN scenario 1

The ChopChop attack was first implemented in C programming language, in 2004. This kind of attack can decrypt a WEP frame regardless of the unavailability of the cryptography keys. For that, this algorithm works with exclusive OR logic operations, used in both the RC4 protocol and the CRC32 algorithm, for computing the Integrity Check Value (ICV), as presented in [22].

The deauthentication attack takes place when the attacker broadcasts false frames, whose address is "FF: FF: FF: FF: FF: FF", in the network. A given station receiving such a frame gets disconnected from the network. This process is then repeated continuously [22].

In the fragmentation attack, the intruder sends a frame as a successive set of fragments. The access point will assemble them into a new frame and send it back to the wireless network. Since the attacker knows the clear text of the frame, he can recover the key stream used to encrypt the frame. This process is repeated until he/she gets a 1,500 byte long key stream. The attacker can use the key stream to encrypt new frames or decrypt a frame that uses the same three byte initialization vector IV [25].

And the duration attack exploits vulnerabilities of the Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA) algorithm, in which the compromised station reserves a communication channel for a given timeframe. In order to capture the channel for long periods of time, the attacker injects frames with a large reservation time parameter into the network (large value for the NAV parameter). This prevents the other stations from using the network during such intervals. Like the previous type of attack, the intrusion is continuous by the attacking station sending new reservation frames before the expiration of the previously sent frame [22].

Concerning the preparation of the collected data to be useful in evaluating IDS, after the collection of the raw data, a pre-processing operation was performed. The resulting dataset contains the following fields: *protocol version, type, subtype, to DS, from DS, more fragment, retry, power management, more data, WEP, order, duration, address1, address2, address 3 and sequence control*.

Similarly to what was carried out in [25], we worked only with samples of the whole data set collected. This allows for this approach to be useful in situations where computational resources are limited. The exact number of samples we used are shown in Table 1.

3.2 Scenario 2 - Dataset creation with WPA2 criptography

In the WPA cryptography, the IEEE 802.1x [26] authentication mechanism permits secure users association into the network. This is the sort of cryptography commonly used in enterprise networks,

as illustrated in Fig. 3. We have here a more complex scenario, in comparison with the previous one, that was implemented in the campus of our educational institution.

Table 1 – Distribution of the sampled dataset for scenario 1

Type	**Training**	**Validation**	**Test**
Normal	6000	4000	5000
ChopChop	900	600	800
Deautenticação	900	600	800
Duration	900	600	800
Fragmentation	900	600	800
Total Samples	9600	6400	8200

Source: Adapted from [25]

The real implementation contains severals wireless stations, two access point (AP) and a RADIUS authentication server. Three stations (client 1, client 2 and client 3) were used to generate normal traffic, based on HTTP and HTTPS web applications. Another station with Airplay [23] was in charge of generation the attacks. Yet a fourth station was configured with Wireshark [24] to capture the whole traffic in the network.

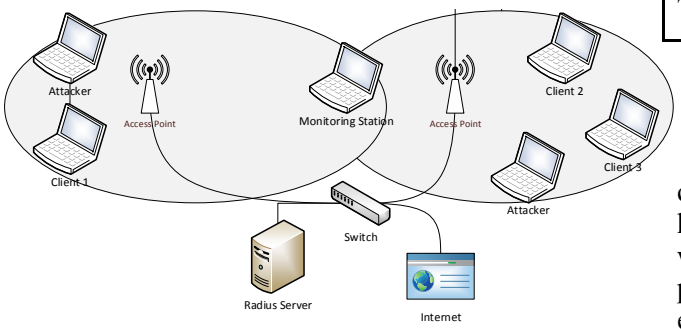

Fig. 3. Topology WLAN applied in generating the data set with WPA2 enabled

The attacks deployed in this scenario are common in wireless networks: deauthentication, fake authentication, fake AP and synflooding. The first attack is identical to the one generated in the scenario 1. The fake authentication occurs when faked frames are injected into the network aiming at including a station that is not a legitimate client of the network. This is done by first capturing frames that contain Initialization Vectors. The fake AP attack establish an access point that is not legitimate in the network, and lastly the synflooding attack aims at generating a large amount of frames into the network to block the network devices that are not prepared to handle such an overload

As was done in scenario 1, the data were collected and pre-processed, toward the dataset, with the following MAC layer fields: *protocol version, type, subtype, to DS, from DS, more fragment, retry, power management, more data, WEP, order, duration, address1, address2, address3 and sequence control*.

The collected data were organized on the basis of the holdout approach proposed in [27]. Specifically, we divided the data in 75% and 25% for training and testing data set, respectively, as illustrated in Table 2.

Table 2 – Distribution of the sampled dataset for scenario 2

Type	**Training**	**Test**
Normal	4500	1500
Deauthentication	750	250
Fake authentication	750	250
Fake AP	750	250
Synflooding	750	250
Total samples	7500	2500

4 Dataset evaluation

The dataset was evaluated by using well-known classification techniques found in the IDSs compared here. In the comparisons, the following parameters were used: the error parameter, during the training phase, the percentage of classification, during the evaluation itself and the Kappa coefficient, as explained later.

The Mean Absolute Error (MAE) is defined as the average of the difference between and computed and measured results. The closer to zero the better the classification is. On the other hand, the Root Mean Square Error (RMSE) is computed as the average of the error square root. A minimum MAE does not imply necessarily in a minimal variation. Thus, it is more effective to use both MAE and RMSE in the evaluations [28].

Both parameters MAE and RMSE provide a simple way to quantify the effectiveness of the classifiers used here in the evaluation of our proposed dataset. They are, however, incipient and so more advanced metrics are encouraged.

Regarding the Kappa coefficient, it was initially used by observers in the psychology field as an induced agreement metric [29]. This metric gives us the degree of acceptance or of agreement responses among a group of judges. Equation 1 provides us with the Kappa outcome, once we have the observed agreement (Po) and the agreement by chance (Pa).

An outcome of k=1 means the classification was correct, while k=0 indicates the classification was totally by chance. Therefore, results close to one are associated to the best classifiers.

$$k = \frac{Po - Pa}{1 - Pa} \quad (1)$$

The dataset evaluation relied on the following classifiers: Bayesian networks, decision tables, IBk J48, MLP and NaiveBayes. The main criteria used here was the popularity of such classifiers.

The Bayesian networks have been used in many approaches for IDS, such as [2]. These networks are directed acyclic graphics for representing a probability distribution on a set of random variables. Each vertex represents as random variable and each node represents a correlation among the variables [30].

The decision table classifier works as follows. It represents a set of conditions needed to determine the occurrence of a group of actions by means of a table format [31]. This technique has also been used in IDS approaches [32].

The IBk algorithm refers to a way of implementing the kNN (k-nearest neighbor) clustering method, which is used for classification and regression toward finding the closest neighbors of a given instance. In the IBk, three neighbors, the ones closest to the search standard neighbors, are used. This is a relatively simple technique that has been used in IDS approaches as well [33].

The J48 algorithm relies on decision tree classifications. By this technique, the classification of a new item depends on the prior creation of a decision tree which uses attributes obtained from the training data. By computing the information gain of each of these attributes, J48 can optimize classification mechanisms in IDS [34].

The MLP is an artificial neural network that maps input parameters to proper outputs. It consists of many layers of nodes in a directed graphic. Several IDS approaches have used MLP [12].

Finally, the NaiveBayes classifier refers to a probabilistic algorithm based on the Bayes theorem with independence hypothesis among the predictors. This is a relatively simple to implement algorithm, as it does not need complex iterative parameters. Thus, NaiveBayes is also useful in evaluating IDS approaches [35].

We used here the Weka [36] tool to evaluate the above mentioned classifiers. For this, we used the dataset we had collected and sampled to input the classifiers. As this tool has been used successfully in diverse researches, it was used “as is”, without any specific improvement.

The results for the mean errors (absolute and quadratic), computed for both of our datasets, during the classifiers training, are shown in Fig. 4 and Fig. 5, respectively. As the errors were relatively low in both cases, we can deduce that the chosen classifiers worked well with the dataset.

Table 3 and 4 present the simulation outcome, after the training phase, for both scenarios. The results for the Correctly Classified Instances are acceptable, in spite of being slightly lower than the results found in the literature. This an expected result, as we focused here in evaluating our proposed dataset, and so no classifier customization was conducted.

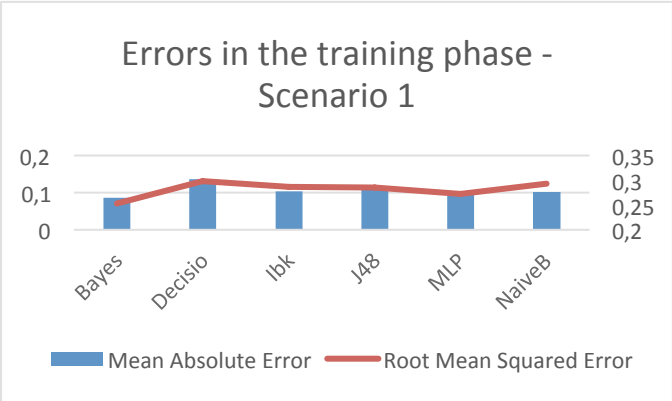

Fig. 4. Errors in the training phase – Scenario 1 (WEP e WPA)

The dataset evaluation represents an important phase of this research, as it allows us to verify the proper response of the classification algorithm commonly used in IDS. The results here did not show any significant discrepancies whatsoever.

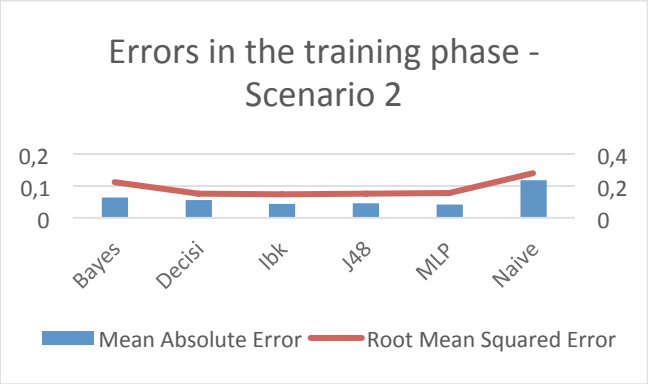

Fig. 5. Errors in the training phase – Scenario 2 (WPA2)

Table 3 – Results for the testing phase of the data set – Scenario 1 (WEP and WPA)

Algorithm	Correctly Classified Instances (%)	Incorrectly Classified Instances (%)	Kappa Coefficient
Bayes Network	82,0610	17,9390	0,6485
Decision Tree	76,3415	23,6585	0,4955
IBk	80,7317	19,2683	0,6390
J48	78,5610	21,4390	0,5906
MLP	81,6463	18,3537	0,6497
NaiveBayes	77,3780	22,6220	0,5860

Table 4 – Results for the testing phase of the data set – Scenario 2 (WPA2)

Algorithm	Correctly Classified Instances (%)	Incorrectly Classified Instances (%)	Kappa Coefficient
Bayes Network	85,8133	14,1867	0,7819
Decision Tree	92,8533	7,1467	0,8778
IBk	92,9067	7,0933	0,8787
J48	92,8533	7,1467	0,8778
MLP	92,4267	7,5733	0,8710
NaiveBayes	66,3333	33,6667	0,5094

5 Results Discussions

The usage of common classification techniques provided good results. The medium errors, computed in the training phase, as illustrated in Fig. 4 and Fig. 5, are relatively low. In addition, the absolute medium error and the mean square error followed the same trend, which confirms the realistic behavior of the data in the collected dataset.

Tables III and IV show that there was no meaningful difference among all obtained results. It is possible to see, however, that the results for the Kappa Coefficient, in scenario 1 (WP/WPA), was the best. Similarly, in scenario 2 (WPA2) the 1Bk classifier performed the best.

We stress that neither customization for the parameters used in these classifiers implementations, nor optimization of procedures were conducted. The reason for that is simply to avoid that such procedures could improperly change the results.

6 Conclusions and future work

The results show that the proposed dataset is viable in evaluating diverse IDS techniques. Even though it is a labeled dataset, in which every registry is identified as being either normal or of a given sort of attack, the dataset is very valuable since it is collected in a real wireless network.

The low errors found in the training phase of the classifiers algorithm confirmed both that the classifiers were properly chosen and that the collected dataset is efficient as far as its purposed is concerned.

Likewise, the results for the Kappa Coefficient followed the same trend concerning the data classified as correct or incorrect, and this confirms the integrity of the generated data.

As future work, we intend to build a third dataset collected out of a wireless network based on the 802.111w standard. This will allow for evaluations of all available scenarios for such a network model.

References

[1] CERT-BR, "Estatísticas do CERT.br -- Incidentes." [Online]. Available: http://www.cert.br/stats/incidentes/. [Accessed: 18-Feb-2014].

[2] P. García-Teodoro, J. Díaz-Verdejo, G. Maciá-Fernández, and E. Vázquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges," *Comput. Secur.*, vol. 28, no. 1–2, pp. 18–28, Feb. 2009.

[3] D. Mudzingwa and R. Agrawal, "A study of methodologies used in intrusion detection and prevention systems (IDPS)," in *2012 Proceedings of IEEE Southeastcon*, 2012, pp. 1–6.

[4] M. Baig and K. Kumar, "Intrusion Detection in Wireless Networks Using Selected Features," *Int. J. Comput. Sci. Inf. Technol.*, vol. 2, pp. 1887–1893, 2011.

[5] I. F. Akyildiz, X. Wang, and W. Wang, "Wireless mesh networks: a survey," *Comput. Networks*, vol. 47, no. 4, pp. 445–487, Mar. 2005.

[6] R. Mohanabharathi, M. T. Kalaikumaran, and S. Karthi, "Feature Selection for Wireless Intrusion Detection System Using Filter and Wrapper Model," *Int. J. Mod. Eng. Res.*, vol. 2, no. 4, pp. 1552–1556, 2012.

[7] A. Shiravi, H. Shiravi, M. Tavallaee, and A. A. Ghorbani, "Toward developing a systematic approach to generate benchmark datasets for intrusion detection," *Comput. Secur.*, vol. 31, no. 3, pp. 357–374, May 2012.

[8] M. Tavallaee, N. Stakhanova, and A. A. Ghorbani, "Toward Credible Evaluation of Anomaly-Based Intrusion-Detection Methods," *IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev.*, vol. 40, no. 5, pp. 516–524, Sep. 2010.

[9] E. Weingartner, H. vom Lehn, and K. Wehrle, "A Performance Comparison of Recent Network Simulators," in *2009 IEEE International Conference on Communications*, 2009, pp. 1–5.

[10] A. M. Law, "How to build valid and credible simulation models," in *Proceedings of the 2009 Winter Simulation Conference (WSC)*, 2009, pp. 24–33.

[11] K. Tan, D. Wu, A. (Jack) Chan, and P. Mohapatra, "Comparing simulation tools and experimental testbeds for wireless mesh networks," *Pervasive Mob. Comput.*, vol. 7, no. 4, pp. 434–448, Aug. 2011.

[12] S. H. Zhong, H. J. Huang, and A. Bin Chen, "An Effective Intrusion Detection Model Based on Random Forest and Neural Networks," *Adv. Mater. Res.*, vol. 267, pp. 308–313, Jun. 2011.

[13] E. Corchado and Á. Herrero, "Neural visualization of network traffic data for intrusion detection," *Appl. Soft Comput.*, vol. 11, no. 2, pp. 2042–2056, Mar. 2011.

[14] S. X. Wu and W. Banzhaf, "The use of computational intelligence in intrusion detection systems: A review," *Appl. Soft Comput.*, vol. 10, no. 1, pp. 1–35, Jan. 2010.

[15] V. Kumar, H. Chauhan, and D. Panwar, ""K-Means Clustering Approach to Analyze NSL-KDD Intrusion Detection Dataset," *Int. J. Soft Comput. Eng.*, vol. 3, no. 4, 2013.

[16] Z. Muda, W. Yassin, M. N. Sulaiman, and N. I. Udzir, "Intrusion detection based on K-Means clustering and Naïve Bayes classification," in *2011 7th International Conference on Information Technology in Asia*, 2011, pp. 1–6.

[17] S. K. Sharma, P. Pandey, S. K. Tiwari, and M. S. Sisodia, "An improved network intrusion detection technique based on k-means clustering via Naive bayes classification," in *International Conference on Advances in Engineering, Science and Management (ICAESM), 2012*, 2012, pp. 417–422.

[18] S. S. Kandeeban and R. S. Rajesh, "A Genetic Algorithm Based elucidation for improving Intrusion Detection through condensed feature set by KDD 99 data set," *Inf. Knowl. Manag.*, vol. 1, no. 1, pp. 1–9, 2001.

[19] M. Sazzadul Hoque, "An Implementation of Intrusion Detection System Using Genetic Algorithm," *Int. J. Netw. Secur. Its Appl.*, vol. 4, no. 2, pp. 109–120, Mar. 2012.

[20] M. K. Goyal, A. Aggarwal, and N. Jain, "Effect of change in rate of genetic algorithm operator on composition of signatures for misuse intrusion detection system," in *2012 2nd IEEE International Conference on Parallel, Distributed and Grid Computing*, 2012, pp. 669–672.

[21] N. V. de S. Araujo, R. de Oliveira, E. T. Ferreira, V. E. do Nascimento, A. A. Shinoda, and B. Bhargava, "Kappa-Fuzzy ARTMAP: A Feature Selection Based Methodology to Intrusion Detection in Computer Networks," in *2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, 2013, pp. 271–276.

[22] M. Guennoun, A. Lbekkouri, A. Benamrane, M. Ben-Tahir, and K. El-Khatib, "Wireless networks security: Proof of chopchop attack," in *2008 International Symposium on a World of Wireless, Mobile and Multimedia Networks*, 2008, pp. 1–4.

[23] Aircrack, "Aircrack-ng." [Online]. Available: http://www.aircrack-ng.org/. [Accessed: 19-Feb-2014].

[24] G. Combs, "Wireshark · Go Deep.," 1998. [Online]. Available: http://www.wireshark.org/. [Accessed: 19-Feb-2014].

[25] K. El-Khatib, "Impact of Feature Reduction on the Efficiency of Wireless Intrusion Detection Systems," *IEEE Trans. Parallel Distrib. Syst.*, vol. 21, no. 8, pp. 1143–1149, Aug. 2010.

[26] Keun Young Park, Yong Soo Kim, and Juho Kim, "Security enhanced IEEE 802.1x authentication method for WLAN mobile router." PyeongChang, pp. 549–553, 2012.

[27] P. Smith, "Autocorrelation in logistic regression modelling of species' distributions," *Glob. Ecol. Biogeogr. Lett.*, vol. 4, no. 2, pp. 47–61, 1994.

[28] C. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Clim. Res.*, vol. 30, no. 1, pp. 79–82, 2005.

[29] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, Apr. 1960.

[30] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Mach. Learn.*, vol. 29, no. 2–3, pp. 131–163, 1997.

[31] J. Huysmans, K. Dejaeger, C. Mues, J. Vanthienen, and B. Baesens, "An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models,"

Decis. Support Syst., vol. 51, no. 1, pp. 141–154, Apr. 2011.

[32] S. S. Sivatha Sindhu, S. Geetha, and A. Kannan, "Decision tree based light weight intrusion detection using a wrapper approach," *Expert Syst. Appl.*, vol. 39, no. 1, pp. 129–141, Jan. 2012.

[33] H. Om and A. Kundu, "A hybrid system for reducing the false alarm rate of anomaly intrusion detection system," in *2012 1st International Conference on Recent Advances in Information Technology (RAIT)*, 2012, pp. 131–136.

[34] M. K. Nagle and S. K. Chaturvedi, "Feature Extraction Based Classification Technique for Intrusion Detection System," *Int. J. Eng. Res. Dev.*, vol. 8, no. 2, pp. 23–38, 2013.

[35] Z.-G. Chen and S.-R. Kim, "Combining principal component analysis, decision tree and naïve Bayesian algorithm for adaptive intrusion detection," in *Proceedings of the 2013 Research in Adaptive and Convergent Systems on - RACS '13*, 2013, pp. 312–316.

[36] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software," *ACM SIGKDD Explor. Newsl.*, vol. 11, no. 1, p. 10, Nov. 2009.