2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Multi-Objective Optimization of a Hybrid Model for Network Traffic Classification by combining Machine Learning Techniques

Zuleika Nascimento, Djamel Sadok, Stênio Fernandes and Judith Kelner

***Abstract*— Considerable effort has been made by researchers in the area of network traffic classification, since the Internet is constantly changing. This characteristic makes the task of traffic identification not a straightforward process. Besides that, encrypted data is being widely used by applications and protocols. There are several methods for classifying network traffic such as known ports and Deep Packet Inspection (DPI), but they are not effective since many applications constantly randomize their ports and the payload could be encrypted. This paper proposes a hybrid model that makes use of a classifier based on computational intelligence, the Extreme Learning Machine (ELM), along with Feature Selection (FS) and Multi-objective Genetic Algorithms (MOGA) to classify computer network traffic without making use of the payload or port information. The proposed model presented good results when evaluated against the UNIBS data set, using four performance metrics: Recall, Precision, Flow Accuracy and Byte Accuracy, with most rates exceeding 90%. Besides that, presented the best features and feature selection algorithm for the given problem along with the best ELM parameters.**

I. Introduction

In recent years, the research effort toward network traffics identification has been growing [1] [2] [3] [4] [5] [6] [7]. As the Internet grows exponentially in both traffic volume and number of protocols and applications, it is essential to understand the composition of dynamic traffic characteristics to recognize protocols and applications which are often encrypted.

In this context, identifying traffic that passes over a network is a complex task, since access to the Internet is significantly increasing, bringing with it new users with different goals. To bring to the experts attention what passes through a network is an increasingly important activity.

There are several methods for classifying network traffic as known ports and Deep Packet Inspection (DPI) [8] [9]. The classification method based on ports performs an analysis of port numbers and is employed to identify applications or protocols. This technique proves to be quite ineffective, since most of the applications make use of random ports. The payload inspection technique or DPI, in turn, eliminates the problem of using random port number used for a specific application or protocol. The technique works starting with a classifier that extracts the payload from TCP/UDP packets and scans each packet in search of signatures that can identify the flow type. However, this technique does not work correctly in encrypted traffic data.

Zuleika Nascimento, Djamel Sadok, Stênio Fernandes and Judith Kelner are with the Department of Computing Systems, Informatics Center, Federal University of Pernambuco, Recife, Brazil (email: {ztcn, jamel, sflf, jk}@cin.ufpe.br).

Recently, some methodologies have been investigated as network traffic classification tools. The work presented in [10] demonstrates the use of data mining techniques to classify flow and user behavior profiles. In order to classify the network traffic, the clustering k-means algorithm is used and compared to other model-based clustering methods along with rule-based classification models. Associations were found among flow parameters for several protocols and applications, such as Hypertext Transfer Protocol (HTTP), Mail, Simple Mail Transfer Protocol (SMTP), Domain Name System (DNS) and Internet Relay Chat (IRC). However, the variables used were source port, destination port, source IP address and destination IP address, and they may not be efficient when this technique is used for applications that enable obfuscation techniques or which are constantly changing pairs, IP addresses and random generations of ports number (e.g., eMule, BitTorrent, and Gnutella).

Bar-Yanai et al. [3] proposed a methodology based on a hybrid combination of two machine learning algorithms - K-Nearest Neighbor (KNN) [11] and K-Means [12], but this method works only with prior knowledge of the number of analyzed applications, i.e., the number of formed groups. Some works [1] [2] [4] [8] [9] [10] [13] do traffic classification based on port number, payload, or even the use of machine learning algorithms. Some of these works [1] [10] [9] exhibit signatures or association rules, resulting in extracted patterns. However, as already explained, these methods are not efficient for encrypted data and when applications make use of random ports.

Thus, this work presents a hybrid model to classify network traffic by using some computational intelligence algorithms, such as Extreme Learning Machine (ELM), Feature Selection (FS) Algorithms and Multi-objective Genetic Algorithms (MOGA). MOGA was used to optimize the ELM classifier and to choose the best feature selection algorithm among seven algorithms, aiming to maximize two important metrics in network traffic classification, Flow Accuracy and Byte Accuracy. To reduce the time taken to train the model, due to the slow optimization process by Genetic Algorithms and the high dimensional search space, ELM was chosen due to its extremely fast learning speed and good generalization [14]. The contributions of this paper include the following:

- Propose a hybrid model by combining three computational intelligence techniques: Multi-objective Genetic Algorithms, Extreme Learning Machine and Feature Selection to tackle with the problem of network traffic classification without the use of port information or payload. Since the payload is not being analyzed, the

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

model is able to deal with encrypted data.

- Propose a method to enhance the results for not only one metric, but two important metrics: Flow Accuracy and Byte Accuracy. In order to accomplish that task, a multi-objective optimizer, the MOGA, is used to optimize two objective functions simultaneously.
- Identify the best feature selection algorithm among seven known algorithms and the most contributing features to the model for a specific popular network traffic ground truth data set, the UNIBS-2009.
- In order to enhance the quality of the model, the ELM parameters are optimized by the MOGA and then detailed in this paper. The parameters were the activation function and number of neurons in the single-hidden layer.

This paper is organized as follows. In Section II, we briefly review the techniques used in this paper. Section III shows the proposed model methodology. Section IV presents the experiments and the analysis of the results. Finally, Section V concludes with final considerations.

II. Fundamentals

The machine learning is a very promising approach for traffic classification, since classification using computational intelligence techniques can be used to identify network traffic data without relying on packet payload. To deal with the analysis of huge network traffic data, machine learning techniques have been used as important tools to create a model to aid computer network analysts.

A. Feature Selection

Feature selection is an important process in machine learning. If not enough features are selected, the predictive power of the model decreases. On the other hand, using all features may reveal impossible since the amount of available training data is usually small with respect to dimensionality (curse of dimensionality) [15]. Besides that, feature selection may also help analysts to understand what features are important in the task of traffic classification for a particular class of protocols or applications. Therefore, feature selection consists of choosing a trade-off between the number of selected features and the adequacy of the learned model [15].

It is not a simple task to select a set of features to enhance the predictive power of a model. For that reason, it has become a research area in machine learning. In this paper, we use the feature selection algorithms presented on Table I. To choose the best algorithm, a multi-objective optimizer is applied, the MOGA (Section II-C). More details on Section III-D.

B. Extreme Learning Machine (ELM)

Extreme Learning Machine is a machine learning algorithm proposed by [14]. It is a very effective training algorithm for Single-hidden Layer Feedforward Neural Networks (SLFNs). The input weights and hidden layer neuron biases are randomly assigned and can produce good generalization performance in most cases and can learn thousands of times faster than conventional popular learning algorithms for feedforward neural networks [14]. ELM avoids problems like local minima, improper learning rate and overfitting commonly faced by iterative learning methods [23]. ELM can be applied as the estimator in regression problem or the classifier for classification tasks. . It has been used in various elds and applications because of better generalization ability, robustness, and controllability and fast learning rate [14]. In this paper, we use ELM due to its extreme fast learning process and good generalization to reduce the computational costs of the full optimization process.

TABLE I
Feature Selection Algorithms.

Feature Selection Algorithms	**Reference**
Fisher Score	[16]
Information Gain	[17]
CFS	[18]
Chi Square	[19]
FCBF	[20]
Kruskal-Wallis	[21]
T-test	[22]

C. Multi-objective Genetic Algorithms (MOGA)

This paper proposes a method to not only maximize one metric, but K metrics, that is, optimize K objective functions. Given a n-dimensional vector $\mathbf{x} = \{x_1, ..., x_n\}$ in a search space containing all possible solutions $\mathbf{X}$, where $\mathbf{x}$ represents one possible solution, the main objective is to find a vector $\mathbf{x}^* \in \mathbf{X}$ that maximizes K objective functions $\phi(\mathbf{x}^*) = \{\phi_1(\mathbf{x}^*), ..., \phi_K(\mathbf{x}^*)\}$, that is:

$$\mathbf{x}^* = \arg\max_{\mathbf{x}} \phi(\mathbf{x}) \tag{1}$$

In this paper we propose to maximize two objective functions, that is, $K = 2$, for both Flow Accuracy and Byte Accuracy, forming then a multi-objective optimization problem. To reach that, we focus on the optimization of Feature Selection and ELM training phases by using a Multi-objective Genetic Algorithms [24]. MOGA are based on Genetic algorithms (GA), which are computer based search procedures based on the natural selection principle. Such procedures were first used by [25]. GAs are capable of finding a global (best) solution for a problem, with high probability; it is also applied to solve complex problems of many practical applications (e.g., function optimization). Since the objective is to maximize multiple objective functions, a trade-off is considered. The Fig. 1 shows a graphical illustration of a multi-objective optimization task. All feasible solutions are presented on the grey area and the optimal surface is called pareto front, pareto-optimal or non-dominated solutions. In this illustration, x_1 and x_2 are possible optimal solutions and $f_1(x)$ and $f_2(x)$ are two objective functions. One may decide to choose x_1 or x_2 as the best solution, since there is a trade-off between them, that is, if the objective is to minimize the functions and x_1 is chosen, then one is deciding to have better rates of $f_1(x)$ instead of $f_2(x)$.

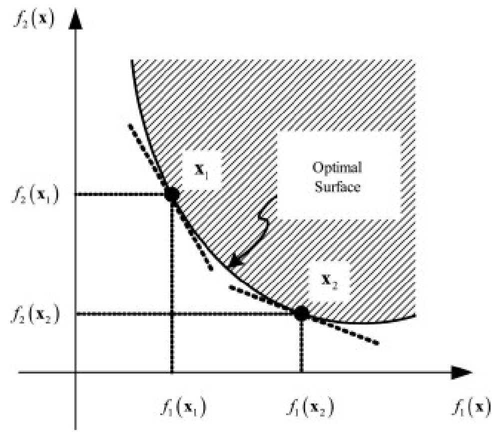

Fig. 1. Pareto-optimal solutions in the space of multi-objective functions. (Figure obtained from [26])

III. Proposed Model

The proposed techniques in the literature cover several models to deal with traffic identification and pattern extraction. However, the number of new Internet applications increases at a high speed, and the classification of such a changing scenario is a complex task, especially when it comes to new applications done without analysing the payload. To deal with this problem, this paper presents a hybrid model for traffic classification based on ELM, MOGA and Feature Selection process.

A. Architecture

The proposed model makes use of a single ELM model to classify network traffic data, nevertheless the training process is what turns it into a hybrid methodology to enhance the model quality. The Fig. 2 presents overview of the hybrid model. It is divided into four main modules: Pre-processing, Feature Selection, a Classifier (ELM) and an Optimizer (MOGA). In a nutshell, the network trace is pre-processed so that soft computing techniques could be used. The optimization process is conducted by a multi-objective genetic algorithm (MOGA) which selects the best feature selection algorithm, the number of features to be selected for training and the extreme learning machine parameters (number of neurons and activation function). Later, the results are analyzed. The assembly of the hybrid proposed model is detailed in the next subsections.

B. Data Set

The network trace used in the experiments were generated at the University of Brescia (UNIBS) in Italy in September and October 2009, during three days. The traces [27] were captured by using *tcpdump* on the Faculty's router. The traffic was generated by a series of workstations running the *gt* client daemon [28], assuring the ground truth information. The original traffic classes were grouped into four classes of protocols: Web, P2P (Edonkey, BitTorrent), Mail (SSL mainly) and VoIP (Skype (TCP)).

In this paper, we use the flow-based classification, therefore the data were converted to TCP network flows by using *Tcptrace* [29]. A network flow is represented by a 5-tuple: Source IP, Destination IP, Source Port, Destination Port and Protocol, nevertheless the proposed method does not make use of these information due to the port-based classification and obfuscation issues. Instead, *Tcptrace* generated statistical information are used, which more than 100 new features.

C. Pre-processing

The pre-processing phase consists of preparing the data to enhance the model results. Therefore, the following steps below were performed:

- Scale the trace in the range [0,1], since this process improves the effectiveness and performance of computational intelligence algorithms.
- Since the trace used to validate the model is imbalanced, the technique of random oversampling [30] was employed instead of undersampling due to its losing information issue.
- Only TCP traces were taken into account, since UDP represents less then 4% of flows of the entire trace and almost 0% when the these flows are converted to bytes. Unknown traffic class were not used due to the same reason.
- The data set was divided into training and test sets, with a proportion of 75% and 25% respectively. The sets were shuffled and then sampled. The flow and bytes composition are detailed in Table II and Table III right after the oversampling process, changing its original composition [27].

TABLE II
Training Data Set

Class of protocols	Flows (%)	Bytes (%)
Web	25.00	7.12
P2P	25.01	70.97
Mail	24.99	1.04
VoIP	25.00	20.87

TABLE III
Test Data Set

Class of protocols	Flows (%)	Bytes (%)
Web	25.01	7.33
P2P	24.97	72.77
Mail	25.02	1.04
VoIP	25.00	18.86

D. Feature Selection

As already mentioned, *Tcptrace* extracts many flow-based features that form a high-dimensional feature space. In this work, we use 127 features. To select the set of features to use in our model, the algorithms of Table I are tested and the one with best results (in two metrics, Flow Accuracy and Byte

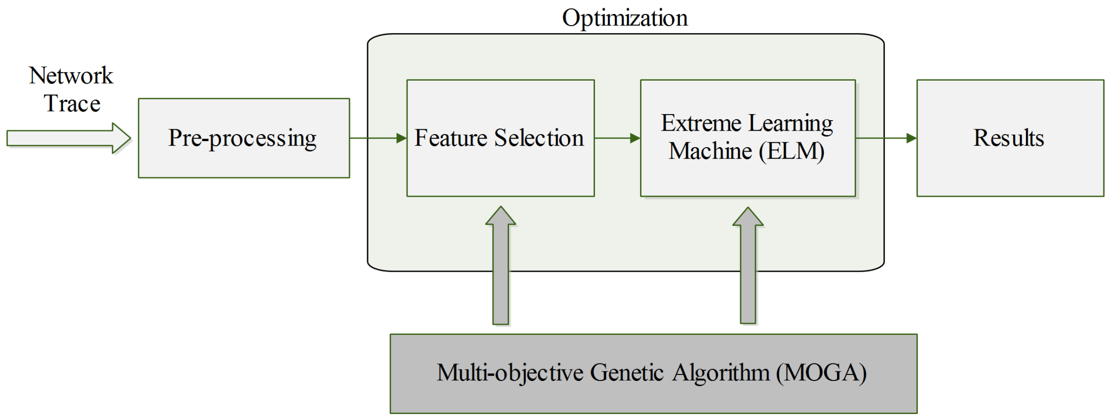

Fig. 2. Proposed Network Traffic Classification Model.

Accuracy (see Section IV-A)) are selected by optimization (MOGA). These algorithms lists the features in sorted by ranking, that is, the higher the ranking the better it contributes to classification. Therefore, MOGA are also used to select the number of features in this ranking list.

E. Classification (ELM)

The ELM was chosen as the unique classifier of the proposed model. This machine learning algorithm was used due to its extremely fast learning speed and good generalization [14]. Besides that, it usually has only one parameter (number of neurons) to be changed, which helps reduce the MOGA dimensional feature space. In this paper, we aim to optimize not only the number of neurons, but to also select the best activation function among five: Sigmoid, Sine, Hard Limit, Radial Basis and Triangular Basis. By using ELM, the whole process of optimization time is reduced.

F. Multi-objective Optimization (MOGA)

In a nutshell, the MOGA is used to enhance the model quality, optimizing the ELM parameters and to choose the best feature selection algorithm and number of features to be used. The goal is to maximize two different metrics, the Flow Accuracy and Byte Accuracy presented in Section IV-A, by finding the best pareto point. The parameters are presented in Table IV. The discreet crossover and discreet mutation functions were implemented according to [31], since the fitness function must return integer values only. The upper and lower bound constraints for each gene are described in Table V.

TABLE IV

MULTI-OBJECTIVE GENETIC ALGORITHMS PARAMETERS.

MOGA Parameters	Value
Crossover Function	Crossover Definition by [31]
Crossover Fraction	0.8
Mutation Function	Mutation Definition by [31]
Elite	2
Population	40
Stall Generation Limit	50
Generations	100

TABLE V

BOUND CONSTRAINTS PER GENE.

Parameter	Lower Bound	Upper Bound
ELM Neurons	10	1000
ELM Activation Function	1	5
FS Algorithm	1	7
FS Number of Features	1	127

IV. EXPERIMENTAL RESULTS

This section introduces the experimental results when applying the model against the UNIBS-2009 test set (Table III). The experiments were executed in MATLAB (R2012a).

A. Metrics

The proposed method performance was evaluated using the metrics described below:

- Flow Accuracy: it is the number of correctly classified flows divided by the total number of flows.
- Byte Accuracy: it is the percentage of correctly classified flow bytes over the total amount of bytes in all flows.
- Recall: measures the per-class accuracy. It is defined by (2). Recall R_i is the number of flows from class $i = 1, ..., M$ correctly classified (TP_i), divided by the number of flows in class i (n_i).

$$R_i = \frac{TP_i}{n_i} \tag{2}$$

- Precision: permits to measure the fidelity of the classification model regarding each particular class. It is defined by (3) and corresponds to the percentage of flows correctly classified as belonging to class i (TP_i) among all the flows classified as belonging to class i, including false positives FP_i.

$$P_i = \frac{TP_i}{TP_i + FP_i} \tag{3}$$

B. Model Performance

After the training process, the best chromosome found is detailed in Table VI for ELM Neurons, ELM Activation Function, FS Algorithm and FS Number of Features. Therefore, these parameters were used to create and analyze the model performance in the test data set, which presented Flow Accuracy of 91.28% and Byte Accuracy of 96.76% as described in Table VI. That was the best solution (chromosome) found by MOGA for both metrics simultaneously, that is, the best pareto point. The 26 selected features are presented in Table VII ordered by ranking.

TABLE VI
EXPERIMENTAL RESULTS AND PARAMETERS.

Experimental Variables	Results
ELM Neurons	856
ELM Activation Function	Sigmoide
FS Algorithm	Information Gain
FS Number of Features	26
Flow Accuracy	91.28%
Byte Accuracy	96.76%

TABLE VII
SELECTED FEATURES FOR UNIBS-2009 DATA SET.

Ranking	Feature
1	sacks_sent_b2a
2	avg_win_adv_b2a
3	RTT_full_sz_min_a2b
4	initial_window_bytes_b2a
5	adv_wind_scale_b2a
6	avg_win_adv_a2b
7	min_win_adv_a2b
8	avg_owin_b2a
9	RTT_full_sz_stdev_b2a
10	ambiguous_acks_b2a
11	avg_retr_time_a2b
12	segs_cum_acked_a2b
13	RTT_sdv_(last)_a2b
14	RTT_avg_(last)_b2a
15	actual_data_pkts_b2a
16	sack_pkts_sent_a2b
17	min_segm_size_a2b
18	avg_segm_size_a2b
19	actual_data_bytes_b2a
20	max_sack_blks/ack_a2b
21	RTT_min_b2a
22	RTT_full_sz_smpls_b2a
23	truncated_packets_b2a
24	RTT_samples_b2a
25	max_retr_time_a2b
26	triple_dupacks_a2b

We also analyzed the results per class, by using the recall and precision metrics, and are detailed in Table VIII. Although the results are mostly above 90%, the classification P2P and VoIP classes could not reach that mark for Precision and Recall respectively. A comparison between Recall (Flow) and Recall (Byte) metrics is also performed as shown in Table IX. Depending on the computer network expert, the byte information tends to be more useful than the flow information, since the specialist could actually know what is consuming the bandwidth, that is, what class of protocol is responsible for part of the traffic. That information could help the experts to change the behavior of their networks by applying restrictions. The byte results in Table IX exceeded the flow results, except for the Mail class. Analyzing the Table VI for Byte Accuracy, it is noted that only 3.24% of the test data set was not detected in terms of byte, with that data set containing encrypted data and without using payload or port information.

TABLE VIII
RECALL AND PRECISION RESULTS PER CLASS.

Class of protocols	Recall	Precision
Web	90.77	96.63
P2P	93.60	88.96
Mail	93.83	91.04
VoIP	86.90	91.68

TABLE IX
RECALL (FLOW VS. BYTE) COMPARISON.

Class of protocols	Recall (Flow)	Recall (Byte)
Web	90.77	98.32
P2P	93.60	94.66
Mail	93.83	81.77
VoIP	86.90	99.79

V. CONCLUSION AND FUTURE WORK

This work proposed a hybrid model based on computational intelligence techniques, optimized by Multi-objective Genetic Algorithms (MOGA) in order to maximize two important metrics in network traffic classification, Flow Accuracy and Byte Accuracy, simultaneously. The model was proposed to deal with the problem of encrypted data classification and the constantly changing ports behavior of some applications. Besides that, presented the most contributing features to the model and the best feature selection algorithm for the data set investigated. The hybrid model makes use of some important machine learning techniques, such as Extreme Learning Machine (ELM), Feature Selection (FS) and Multi-objective Optimization.

The experiments with the proposed model presented good results, exceeding levels of 91% for Flow Accuracy and 96% for Byte Accuracy. Besides that, also showed promising results for Precision and Recall per class, with results of approximately 90%. On the other hand, results of Recall (Byte) for the Mail class can be improved, since it reached a percentage of 81.77%. The experiments showed that within 127 features analyzed, only 26 were responsible to enhance the model predictive power. Also showed that Information Gain was the best Feature Selection algorithm. The training process also optimized the ELM parameters, such as the activation function to be used and the number of neurons in the single-hidden layer. The experiments were evaluated against a well-known network traffic trace, due to its ground truth characteristics, the UNIBS-2009 data set.

Although the model showed to be very promising, reaching high levels on the performance metrics being investigated, it can still be improved, specially for the Mail class. To do so, new experiments could be conducted inspired by the divide-and-conquer strategy [32], where a complex task is divided into simpler (smaller) subtasks that together solve the main problem. It can be performed by dividing the data set into smaller clusters and then create a specific model for each cluster. Growing Hierarchical Self-Organizing Maps [33] could be used to divide the data set and Extreme Learning Machine to conquer each subset. The problem of imbalanced data set can also be better tackled with the Synthetic Minority Over-sampling Technique (SMOTE) [34], technique that shows to be very effective [30]. Besides that, a better investigation could also be performed on different data sets and to compare the results with other network traffic classification models. A deeper analysis of the model should be performed on high throughput networks (e.g., 10Gbps, 100Gbps) with the aid of Apache Hadoop [35], a framework that allows the distributed processing of large data sets across clusters of computers.

ACKNOWLEDGMENT

We thank all the support given by the Point of Presence of Pernambuco (PoP-PE) of the National Network on Teaching and Research (RNP), the Institute of Technology of Pernambuco (ITEP) and the Networking and Telecommunications Research Group (GPRT) that belongs to the Computer Science Center of the Federal University of Pernambuco at Recife, Brazil.

REFERENCES

[1] G. Szabó, Z. Turányi, L. Toka, S. Molnár, and A. Santos, "Automatic protocol signature generation framework for deep packet inspection," in *Proceedings of the 5th International ICST Conference on Performance Evaluation Methodologies and Tools*, Brussels, Belgium, May 2011, pp. 291–299.

[2] V. Carela-Español, P. Barlet-Ros, M. Solé-Simó, A. Dainotti, W. de Donato, and A. Pescapé, "K-dimensional trees for continuous traffic classification," in *Proceedings of the Second international conference on Traffic Monitoring and Analysis*, ser. Lecture Notes in Computer Science, F. Ricciato, M. Mellia, and E. Biersack, Eds., vol. 6003. Berlin, Heidelberg: Springer Berlin Heidelberg, Apr. 2010, pp. 141–154.

[3] R. Bar - Yanai, M. Langberg, D. Peleg, and L. Roditty, "Realtime classification for encrypted traffic," in *Proceedings of the 9th international conference on Experimental Algorithms*, ser. Lecture Notes in Computer Science, P. Festa, Ed., vol. 6049. Berlin, Heidelberg: Springer Berlin Heidelberg, May 2010, pp. 373–385.

[4] A. Dainotti, A. Pescape, and K. Claffy, "Issues and future directions in traffic classification," *IEEE Network*, vol. 26, no. 1, pp. 35–40, Jan. 2012.

[5] J. Summers, T. Brecht, D. Eager, and B. Wong, "Methodologies for generating HTTP streaming video workloads to evaluate web server performance," in *Proceedings of the 5th Annual International Systems and Storage Conference on - SYSTOR '12*. New York, New York, USA: ACM Press, Jun. 2012, pp. 1–12.

[6] Z. Nascimento, D. F. Sadok, and S. Fernandes, "Comparative Study of a Hybrid Model for Network Traffic Identification and its Optimization using Firefly Algorithm," in *The Eighteenth IEEE Symposium on Computers and Communications*, Split, Croatia, to be published 2013.

[7] Z. Nascimento, D. Sadok, and S. Fernandes, "A Hybrid Model for Network Traffic Identification Based on Association Rules and Self-Organizing Maps (SOM)," in *The Ninth International Conference on Networking and Services*, to be published 2013.

[8] G. La Mantia, D. Rossi, A. Finamore, M. Mellia, and M. Meo, "Stochastic Packet Inspection for TCP Traffic," in *2010 IEEE International Conference on Communications*. IEEE, May 2010, pp. 1–6.

[9] M. Ye, K. Xu, J. Wu, and H. Po, "AutoSig-Automatically Generating Signatures for Applications," in *2009 Ninth IEEE International Conference on Computer and Information Technology*, vol. 2. IEEE, 2009, pp. 104–109.

[10] U. K. Chaudhary, I. Papapanagiotou, and M. Devetsikiotis, "Flow classification using clustering and association rule mining," in *2010 15th IEEE International Workshop on Computer Aided Modeling, Analysis and Design of Communication Links and Networks (CAMAD)*. IEEE, Dec. 2010, pp. 76–80.

[11] T. Cover and P. Hart, "Nearest neighbor pattern classification," pp. 21–27, 1967.

[12] J. Hartigan and M. Wong, "A k-means clustering algorithm," *1 ACM/IEEE-CS Joint Conference, Applied Statistics*, vol. 28, pp. 100–108, 1979.

[13] M. Soysal and E. G. Schmidt, "Machine learning algorithms for accurate flow-based network traffic classification: Evaluation and comparison," *Performance Evaluation*, vol. 67, no. 6, pp. 451–467, Jun. 2010.

[14] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489–501, Dec. 2006.

[15] F. Benoît, M. van Heeswijk, Y. Miche, M. Verleysen, and A. Lendasse, "Feature selection for nonlinear models with extreme learning machines," *Neurocomputing*, vol. 102, no. null, pp. 111–124, Feb. 2013.

[16] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. John Wiley & Sons, New York, 2001.

[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.

[18] L. A. S. Mark A. Hall, "Feature Selection for Machine Learning: Comparing a Correlation-based Filter Approach to the Wrapper."

[19] H. Liu and R. Setiono, "Chi2: Feature selection and discretization of numeric attributes," in *Proceedings of the Seventh IEEE International Conference on Tools with Artificial Intelligence, November 5-8, 1995*, J. Vassilopoulos, Ed. Herndon, Virginia: IEEE Computer Society, 1995, pp. 388–391.

[20] H. Liu and L. Yu, "Feature selection for high-dimensional data: A fast correlation-based filter solution," in *Correlation-Based Filter Solution". In Proceedings of The Twentieth International Conference on Machine Leaning (ICML-03)*. Washington, D.C.: ICM, 2003, pp. 856–863.

[21] L. J. Wei, "Asymptotic conservativeness and efficiency of kruskal-wallis test for k dependent samples," *Journal of the American Statistical Association*, vol. 76, no. 376, pp. 1006–1009, December 1981.

[22] E. Livingston, "Who was student and why do we care so much about his t-test?" *Journal of Surgical Research*, vol. 118, no. 1, pp. 58–65, 2004.

[23] R. Zhang, G.-B. Huang, N. Sundararajan, and P. Saratchandran, "Multi-category classification using an Extreme Learning Machine for microarray gene expression cancer diagnosis." *IEEE/ACM transactions on computational biology and bioinformatics / IEEE, ACM*, vol. 4, no. 3, pp. 485–95, Jan. 2007.

[24] A. Konak, D. Coit, and A. Smith, "Multi-objective optimization using genetic algorithms: A tutorial," *Reliability Engineering & System Safety*, vol. 91, no. 9, pp. 992 – 1007, 2006.

[25] J. H. Holland, *Adaptation in Natural and Artificial Systems*. University of Michigan Press, 1975.

[26] S.-Y. Ok, J. Song, and K.-S. Park, "Optimal design of hysteretic dampers connecting adjacent structures using multi-objective genetic algorithm and stochastic linearization method," *Engineering Structures*, vol. 30, no. 5, pp. 1240–1249, May 2008.

[27] UNIBS, "UNIBS data sets," 2009. [Online]. Available: http://www.ing.unibs.it/ntw/tools/traces/download

[28] F. Gringoli, L. Salgarelli, M. Dusi, N. Cascarano, F. Risso, and k. c. Claffy, "GT: picking up the truth from the ground for internet traffic," *ACM SIGCOMM Computer Communication Review*, vol. 39, no. 5, p. 12, Oct. 2009.

[29] Tcptrace, "Tcptrace - Official Homepage." [Online]. Available: http://www.tcptrace.org/

[30] E. Garcia, "Learning from Imbalanced Data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, pp. 1263–1284, Sep. 2009.

[31] J. Le Besnerais, V. Lanfranchi, M. Hecquet, and P. Brochet, "Multiobjective Optimization of Induction Machines Including Mixed Variables

and Noise Minimization," *IEEE Transactions on Magnetics*, vol. 44, no. 6, pp. 1102–1105, Jun. 2008.

[32] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "Adaptive Mixtures of Local Experts," *Neural Computation*, vol. 3, no. 1, pp. 79–87, Feb. 1991.

[33] A. Rauber, D. Merkl, and M. Dittenbach, "The growing hierarchical self-organizing map: exploratory analysis of high-dimensional data." *IEEE transactions on neural networks*, vol. 13, pp. 1331–41, Jan. 2002.

[34] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, Jan. 2002.

[35] Hadoop, "Apache Hadoop." [Online]. Available: http://hadoop.apache.org/