Uç Öğrenme Makineleri Kullanılarak Internet Trafik Bilgisinin Sınıflandırılması

Fatih Ertam[1], Engin Avcı[2]

[1] Fırat Üniversitesi, Enformatik Bölümü, Elazığ

[2] Fırat Üniversitesi, Yazılım Mühendisliği Bölümü, Elazığ

fatih.ertam@firat.edu.tr, enginavci@firat.edu.tr

Özet: Son zamanlarda nesnelerin interneti (internet of things, IoT) kavramı ile internetin kullanımının çok yüksek düzeylere ulaşması sebebiyle internet ile beraber sunulan servis kalitesinin artırılması, ağın verimli kullanılması ve farklı hizmet paketlerinin oluşturulabilmesi gibi konuların önemi daha da artmıştır. İnternet üzerinden akan trafik verisinin sınıflandırılması, özellikle büyük ağlarda güvenliğin sağlanması, trafik yönetiminin etkin bir şekilde yapılabilmesi için oldukça önemli bir hale gelmiştir. Internet trafiğini hızlı bir şekilde artması ve kullanılan uygulamaların çeşitliliği ağın kontrol edilebilmesi için ağ yöneticileri tarafından bu bilginin bilinmesi neredeyse bir zorunluluk olmaya başlamıştır. Sınıflandırma için yaygın olarak port, yük ve istatistik bilgileri kullanılmıştır. Port ve yük tabanlı yaklaşımlar ile yapılabilecek sınıflandırma seçenekleri sınırlı olduğu için özellikle denetimli makine öğrenme (ML) algoritmaları internet trafik sınıflandırılmasında sıklıkla uygulanmaya başlamıştır. Destek vektör makinesi (DVM) ve yapay sinir ağları (YSA) tabanlı sınıflandırıcılar önceki çalışmalarda oldukça fazla kullanılmıştır. Yapılan çalışmada klasik sınıflandırıcı yöntemler yerine Uç öğrenme makinesi (UÖM) algoritması kullanılmıştır. UÖM ile yapılan sınıflandırma başarımının daha yüksek olduğu görülmüştür.

Anahtar Sözcükler: Makine Öğrenmesi, Denetimli Öğrenme, Sınıflandırma Algoritmaları, Internet Trafik Sınıflandırma, YSA, DVM, UÖM

Classification of Internet Traffic Information by Extreme Learning Machine

Abstract: Together with the term ‘internet of things’, the use of internet has recently reached a very high level. As a result, issues such as development in the service quality, efficient use of network and the availability of different service packets have gained more importance. In order to perform traffic management effectively, the classification of traffic data flowing over the internet -especially providing the security on major networks- has become more important. This information is required to be known by the network administrators to manage the network properly because the internet traffic and variety in applications has rapidly increased. Port, payload and statistical information have been commonly used for the classification. Because there are a limited number of options in the classification made according to the port and payload based approaches, the supervised machine learning (ML) algorithms have started to be frequently used in the internet traffic classification. Support vector machine (SVM) and Neural networks (NN) based classifiers were used so many times in the previous studies. In the performed study, extreme learning machine (ELM) algorithm has been used instead of traditional classifying techniques; and the success of the classification made with ELM has been found to be higher than the others.

Keywords: Machine Learning Algorithms, supervised learning, classification algorithms, internet traffic classification, NN, SVM, ELM

1. Giriş

Internet trafiğinin sınıflandırılması özellikle kurumsal ağlarda hizmet kalitesinin artırılabilmesi, ağın performanslı kullanılabilmesi, yeni internet hizmet paketlerinin oluşturulabilmesi, bant genişliği için kaynaklarının paylaştırılabilmesi, trafik analizlerinin yapılabilmesi gibi amaçları karşılayabilmesi için son zamanlarda sıklıkla kullanılmaktadır.

Ağ trafiğinin sınıflandırılması için genel olarak port tabanlı, yük tabanlı ve istatistik tabanlı sınıflandırma yöntemleri kullanılmaktadır.

Port tabanlı sınıflandırma için akış bilgisinden alınan port bilgileri IANA (Internet Assigned Numbers Authority) tarafından belirlenen protokollere ait port numaraları ile karşılaştırma yapılarak sınıflandırılır [1]. Port tabanlı sınıflandırma yöntemleri hızlı çalışmalarına rağmen aynı portu kullanabilecek farklı uygulamaların olması, port yönlendirme yapılabilmesi gibi nedenlerle doğru bir sınıflandırma yapılabilmesi için yeterli olmamaktadır [2,3]. Son zamanlarda geliştirilen uygulamaların bir kısmı kullanmış oldukları port numaralarını gizleyerek veya dinamik bir şekilde port numaralarını değiştirme yöntemiyle firewall tarafından tanınmamaya çalışmaları da bu yöntemin yetersizliğini ortaya koymaktadır.

Ağ üzerinden geçen paketlerin sınıflandırılabilmesi amacıyla kullanılan diğer bir yöntem yük tabanlı sınıflandırma yöntemidir. Bu yöntemle incelenen paketlerin sadece port, ip adresi gibi bilgileri yerine uygulama katmanında yer alan bilgileri de incelenir. Derin paket analizi olarak da bahsedilebilen bu teknoloji ile amaçlanan paket yük içerisinde bulunan imzalara karşılık gelen uygulama ve protokollerin tespit edilebilmesidir. Özellikle şifreleme kullanılmış uygulamalarda bu yöntemin kullanılabilmesi zorlaşmaktadır, protokollere ait imzaların zamanla değişebilmesi, protokol bilgilerinin çok iyi bir şekilde raporlanmamış olması, imza arama işleminin yüksek işlem gücüne sahip sistemler ile yapılması ihtiyacı gibi nedenler bu yöntemin kullanılarak sınıflandırma yapılmasında problem oluşturmaktadır.

Sınıflandırma için kullanılan port tabanlı ve yük tabanlı yaklaşımlarda karşılaşılan problemlerden dolayı akış davranışlarını istatistik tabanlı olarak inceleyerek sınıflandırma yapan yaklaşımlar son zamanlarda sıklıkla kullanılmaktadır. Bu yaklaşım ile akış üzerinden ortalama paket büyüklüğü, sunucu ile istemci arasında ve istemci ile sunucu arasında paketlerin varış zamanları gibi öznitelikler belirlenmeye çalışılır. İstatistik tabanlı sınıflandırma için makine öğrenmesi algoritmaları kullanılmaktadır.

Genel olarak makine öğrenmesi tabanlı yöntemler; özellik vektörünün çıkarılması, özellik vektörünün istatistiksel dağılımının tahmin edilmesi ve örneğin tanınması şeklinde üç adımdan oluşur [4,5].

Özellik vektörünün çıkarılabilmesi amacıyla ağ üzerindeki trafik bilgisini dinleyerek kayıt altına alabilen çeşitli yazılımlar bulunmaktadır. Wireshark, tcpdump gibi yazılımlar ile ağ trafiği dinlenerek akış bilgisi alınabilir. Bu akış bilgisi üzerinden özniteliklerin belirlenebilmesi için de kullanılan yazılımlar bulunmaktadır [6]. Öğrenme için denetimli, denetimsiz veya yarı denetimli yöntemler kullanılmaktadır.

Denetimli makine öğrenmesi en yaygın olarak kullanılan öğrenme yöntemidir. Bu yöntemin kullanılabilmesi için akışlara ait sınıfların bilinmesi gereklidir. Destek vektör makinesi, karar ağaçları, rastgele orman, k-NN, naive bayes, yapay sinir ağları, sıkça kullanılan denetimli makine öğrenmesi yöntemleridir. Yarı denetimli öğrenme hem sınıfların bilindiği verilerin hem de bilinmediği verilerin bir arada olduğu öğrenme yöntemidir. Denetimsiz öğrenme ise sınıfların belli olmadığı durumlarda kullanılan öğrenme yöntemidir. K-means, AutoClass, expectation maximization, dbscan yöntemleri denetimsiz

öğrenme için kullanılan kümeleme algoritmalarıdır [7-11].
Uç öğrenme makinesi gelişmekte olan yeni sayılabilecek bir öğrenme algoritmasıdır. İleri beslemeli sinir ağlarının hızı genel olarak yavaştır, bu yavaşlık bu yöntemlerin kullanılması için bir dezavantaj oluşturmaktadır [12]. ELM kullanılarak başarılı bir şekilde sınıflandırma yapılmış araştırmalar bulunmaktadır [13-15].

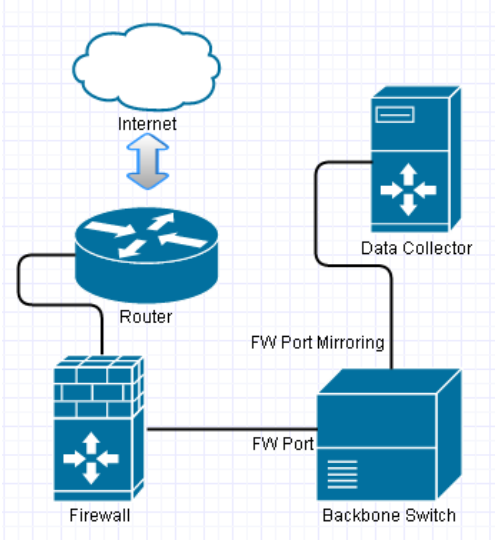
2. Benzer Çalışmalar

Bilinen port numaraları kullanılarak yapılan internet trafik sınıflandırma teknikleri bu yöntemlerin ilk kullanıldığı zamanlarda tercih edilmiştir. Port numaralarının eşleştirilmesi ile yapılan sınıflandırma başarımları yüksek çıkmasına rağmen noktadan noktaya uygulamalarının artması ve kullanılan portların dinamik bir şekilde değişmesi bu yöntem ile yapılan sınıflandırmanın başarısını düşürmüştür. Ağ üzerinde dolaşan paketlerin yakalanarak içeriklerine göre kontrol edilerek sınıflandırma yapılması da farklı bir metot olarak internet sınıflandırma için kullanılmıştır [16,17]. Paket yük içeriği ile yapılan bazı çalışmalar bu alanda bazı yetersizlikler olduğunu ortaya koymuştur [17-20]. Paket içerisindeki imzaların eşleştirilmesi ile sınıflandırma tespiti yapılan bu yaklaşımda özellikle içeriği yeni veya bilinmeyen bir paket olması durumunda sonuç alınamamaktadır. Özellikle internet üzerinden çok fazla miktarda saldırı olabileceği düşünüldüğünde bu saldırıya ait imzanın bilinememesi bu paketin saldırı şeklinde sınıflandırılamamasına sebep olacaktır. Makine öğrenmesi algoritmaları kullanılarak yapılan sınıflandırma yöntemleri ise sınıflandırma yapabilmek için kullanılabilecek bir diğer seçenektir. Bu alanda önemli kabul edilen İlk çalışmalardan birisi olan Moore ve Zuev tarafından yapılmıştır [6]. Moore ve Zuev yapmış oldukları çalışmada akış istatistik öznitelikleri üzerinde denetimli öğrenme algoritması olan naive bayes tekniğini kullanarak network trafik sınıflandırma yapmışlardır. Moore, Zuev ve Crogan yaptıkları çalışma ile istatistiki olarak kayıt altına alınan bir akıp paket dosyasını (packet capture, pcap) bir yazılım aracılığıyla analiz edilerek 248 Adet öznitelik belirleyerek sınıflandırma yapmışlardır. [6]. Daha sonraları bayes sinir ağları ve destek vektör makineleri gibi iyi bilinen algoritmalar ile trafik sınıflandırma yapılmıştır [21,27]. Teja, Cagnazzo ve Castro yaptıkları çalışma ile birden fazla makine öğrenmesi metotlarını bir arada kullanarak ağ trafiğini sınıflandırmaya çalıştılar [29-31]. Farklı ML algoritmalarının karşılaştırılarak sınıflandırma yapıldığı başka çalışmalarda yapılmıştır. Son çalışmalar uygulama akışlarının sunucu ile istemci arasında iki yönlü olacağı var sayılarak istatistiki alınan özellikler ileri ve geri yönde ayrı ayrı hesaplanarak sınıflandırma yapılmaktadır [8,9, 26-32].

3. Deneysel Çalışmalar

3.1 Veri Setinin Oluşturulması

Veri setinin oluşturulması için üniversite kampüs ağı üzerinden alınan veriler kullanılmıştır. Bu amaçla kampüsde bulunan backbone switch üzerinden firewall cihazına giden verinin olduğu port başka bir porta yönlendirilerek veriler packet capture (pcap) dosyaları şeklinde alınmıştır. Fig.1 de hazırlanan sistem gösterilmektedir. IBM Blade Chasis üzerinde bir host sunucu bilgisayar verilerin alınabilmesi amacıyla kurulmuştur. Sunucu bilgisayar 96 GB RAM, 8 core dan oluşan 2 soketli ve 2.93 GHz çalışabilen Intel Xeon X5577 CPU donanımına sahiptir. Verilerin işlenebilmesi için Matlab R2014b yazılımı kullanılmıştır.



Şekil 1. Kampüs Ağından Verilerin Alınması

Verilerin alınması için tcpdump ve wireshark yazılımlarından faydalanılmıştır. Ticari olmayan bu yazılımlar ile port üzerinden geçen veriler dinlenerek kayıt altına alınmıştır. Kaydedilen veri dosyasının analiz edilerek feature larına ayırmak için Moore, Zuev ve Crogan ın yaptığı çalışmada kullanmış oldukları fullstats yazılımından faydalanılmıştır [6]. Çıkan verilerden 12 Adet öznitelik seçilmiştir. Seçilen öznitelik numaraları ve açıklamaları Tablo-1 de verilmiştir.

Tablo 1 Kullanılan Öznitelikler

Öznitelik	Açıklaması
1	server port
2	client port
45	actual_data_pkts (client --> server)
59	pushed_data_pkts (client --> server)
60	pushed_data_pkts (server -->client)
83	min_segm_size (client-->server)
86	avg_segm_size (server -->client)
95	initial_windows_bytes (client -->server)
96	initial_windows_bytes (server -->client)
113	RTT_samples (client-->server)
162	med_data_ip (client-->server)
180	var_data_wire (server -->client)

Akışlara ait sınıflar içerisinde en fazla kullanılan sınıflardan 6 tanesi seçilerek her bir sınıf için 1000 Adet eğitim, 750 Adet test için seçilmiştir. Toplam 6000 Adet eğitim için, 4500 Adet test için kullanılmıştır. Kullanılan sınıflar ve açıklaması Tablo-2 de verilmiştir.

Tablo 2 Kullanılan Sınıflar

No	Sınıf Tipi	Açıklaması
1	Saldırı	ddos
2	Noktadan Noktaya	bittorrent
3	Mail	Pop3,smtp
4	Web	http, https
5	Servis	DNS
6	Veri tabanı	Mysql, mssql

3.2 Uç Öğrenme Makinesi ile Sınıflandırma

ELM ile eğitim ve test verileri kullanılarak yapılan sınıflandırma için farklı aktivasyon fonksiyonları kullanılarak doğruluk yüzdeleri ve çalışma zamanları karşılaştırılmıştır. Basit UÖM için kullanılan aktivasyon fonksiyonlarına göre doğruluk oranları ve çalışma zamanları tablo-3 de verilmiştir. Gizli nöron sayısı 20 olarak alınmıştır.

Tablo 3 UÖM Aktivasyon Fonksiyonlarına göre doğruluk yüzdeleri

Aktivasyon Fonksiyonu	Doğruluk Yüzdesi	Öğrenme Zamanı (saniye)
Sin	78.86	7.02
hardlim	71.87	6.91
Tansig	82.99	6.28
Sig	79.94	7.23
Tribas	79.20	7.12
Radbas	82.36	7.45
Rbf	80.23	7.49

Kullanılan aktivasyon fonksiyonları içerisinde en yüksek doğruluk yüzdesi tanjant sigmoid ve radial basis aktivasyon fonksiyonları ile sağlanmıştır. Tanjant sigmoid fonksiyonu kullanılarak gizli nöron sayıları artırılması ile doğruluk oranın doğru orantılı bir şekilde yükseldiği gözlemlenmiştir. Tablo-4 de sadece Tanjant sigmoid fonksiyonunun kullanılması ile farklı gizli nöron sayılarına göre ortaya

çıkan doğruluk oranı ve çalışma süreleri verilmektedir.

Tablo 4 Tanjant Sigmoid ve Radial basis Aktivasyon Fonksiyonlarının Farklı gizli nöron sayılarındaki doğruluk değerleri

Gizli Nöron Sayısı	Doğruluk Yüzdesi	Çalışma Zamanı (saniye)
Tansig - 20	82.99	6.28
Radbas -20	82.36	7.45
Tansig - 40	87.71	7.44
Radbas -40	87.62	7.47
Tansig - 80	89.70	7.05
Radbas -80	89.92	7.13
Tansig - 160	91.50	7.26
Radbas - 160	92.08	7.63
Tansig - 320	92.67	7.87
Radbas - 320	92.84	7.58
Tansig - 640	92.27	8.18
Radbas - 640	92.25	8.52

3.3 YSA, DVM ve UÖM Sınıflandırmalarının Karşılaştırılması

UÖM yöntemleri ile yapılan sınıflandırmanın performansının karşılaştırılabilmesi için klasik sınıflandırma yöntemlerinden yapay sinir ağları ve destek vektör makineleri, UÖM için kullanılan aynı eğitim ve test verileri ile sınıflandırma yapılmıştır. Bu sınıflandırma yöntemi ile ortaya çıkan doğruluk oranları ve zamanları Tablo-5 de verilmiştir.

Tablo 5 UÖM ile Diğer Sınıflandırma Metotlarının Karşılaştırılması

Sınıflandırma Metodu	Doğruluk Yüzdesi	Çalışma Zamanı (saniye)
YSA	69.52	202.23
DVM	58.99	72.52
UÖM (Radial Basis, 320 nöron)	92.84	7.58

4. Sonuç ve Öneriler

Kampüs ağı üzerinden akan ağ trafik bilgisi ile elde edilen verilerin sınıflandırılması işlemi için hem YSA, DVM gibi klasik sınıflandırma yöntemleri kullanılmış hem de UÖM gibi yeni sayılabilecek sınıflandırma yöntemleri kullanılarak karşılaştırılmıştır. Klasik sınıflandırma yöntemleri yerine kullanılan UÖM algoritmaları ile yapılan sınıflandırmada yüksek oranda başarım elde edildiği görülmüştür. Basit UÖM içerisinde kullanılan aktivasyon fonksiyonları da kendi aralarında karşılaştırılmıştır. Bu aktivasyon fonksiyonlarından radial basis ve tanjant sigmoid fonksiyonları ile elde edilen başarım oranının daha yüksek olduğu görülmüştür. Radial basis ve tanjant sigmoid aktivasyon fonksiyonlarının kullanıldığı UÖM yönteminde gizli nöron sayıları değiştirilerek karşılaştırma yapılmıştır. Gizli nöron sayılarının artması ile sınıflandırma başarım oranının yükseldiği gözlenmiştir. Gizli nöron sayısı 20 den başlanılarak her seferinde iki kat artırılmıştır. 640 Adet gizli nöron sayısından sonra doğruluk oranının çok fazla değişmediği ve yaklaşık aynı değerlerde çıktığı gözlenmiştir. Verilerin analizi yapılırken sınıflandırma algoritmaları 10 kez çalıştırılarak ortalamalarının alınmasıyla elde edilmiştir.

5. Kaynaklar

[1] IANA, Internet Assigned Numbers Authority [Online] http://www.iana.org/protocols

[2] Madhukar, A., & Williamson, C. A longitudinal study of P2P traffic classification. In Modeling, Analysis, and Simulation of Computer and Telecommunication Systems. MASCOTS 2006. 14th IEEE International Symposium on (pp. 179-188). IEEE. 2006, doi:10.1109/MASCOTS.2006.6

[3] Nguyen, T. T., & Armitage, G. A survey of techniques for internet traffic classification using machine learning. Communications Surveys & Tutorials, IEEE, 10(4), 56-76. 2008, doi: 10.1109/SURV.2008.080406

[4] Yin, C., Li, S., & Li, Q. Network traffic classification via HMM under the guidance of syntactic structure. Computer Networks, 56(6), 1814-1825. 2012, doi:10.1016/j.comnet.2012.01.021

[5] Crotti, M., Gringoli, F., Pelosato, P., & Salgarelli, L. A statistical approach to IP-level classification of network traffic. In Communications, 2006. ICC'06. IEEE International Conference on (Vol. 1, pp. 170-176). IEEE. 2006, doi:10.1109/ICC.2006.254723

[6] Moore, A. W., & Zuev, D. Internet traffic classification using bayesian analysis techniques. In ACM SIGMETRICS Performance Evaluation Review (Vol. 33, No. 1, pp. 50-60). ACM. 2005, doi: 10.1145/1071690.1064220

[7] Erman, J., Arlitt, M., & Mahanti, A. Traffic classification using clustering algorithms. In Proceedings of the 2006 SIGCOMM workshop on Mining network data (pp. 281-286). ACM.2006, doi: 10.1145/1162678.1162679

[8] McGregor, A., Hall, M., Lorier, P., & Brunskill, J. Flow clustering using machine learning techniques. In Passive and Active Network Measurement (pp. 205-214). Springer Berlin Heidelberg. 2004.

[9] Zander, S., Nguyen, T., & Armitage, G. Automated traffic classification and application identification using machine learning. In Local Computer Networks, 2005. 30th Anniversary. The IEEE Conference on (pp. 250-257). IEEE. 2005.

[10] Ma, J., Levchenko, K., Kreibich, C., Savage, S., & Voelker, G. M. Unexpected means of protocol inference. In Proceedings of the 6th ACM SIGCOMM conference on Internet measurement (pp. 313-326). ACM. 2006.

[11] Yang, C., Wang, F., & Huang, B. Internet traffic classification using dbscan. In Information Engineering, 2009. ICIE'09. WASE International Conference on (Vol. 2, pp. 163-166). IEEE. 2009.

[12] Huang, G. B., Zhu, Q. Y., & Siew, C. K. Extreme learning machine: a new learning scheme of feedforward neural networks. In Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on (Vol. 2, pp. 985-990). IEEE. 2004, doi: 10.1109/IJCNN.2004.1380068

[13] Avci, E., Coteli, R. A new automatic target recognition system based on wavelet extreme learning machine. Expert Systems with Applications,39(16), 12340-12348. 2012, doi:10.1016/j.eswa.2012.04.012

[14] Huang, G. B., Zhou, H., Ding, X., & Zhang, R. Extreme learning machine for regression and multiclass classification. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 42(2), 513-529. 2012, doi: 10.1109/TSMCB.2011.2168604

[15] Huang, G. B., Ding, X., & Zhou, H. Optimization method based extreme learning machine for classification. Neurocomputing, 74(1), 155-163. 2010, doi: doi:10.1016/j.neucom.2010.02.019

[16] Moore, A. W., & Papagiannaki, K. Toward the accurate identification of network applications. In Passive and Active Network Measurement (pp. 41-54). Springer Berlin Heidelberg. 2005, doi:10.1007/978-3-540-31966-5_4

[17] Karagiannis, T., Broido, A., & Faloutsos, M. Transport layer identification of P2P traffic. In Proceedings of the 4th ACM SIGCOMM conference on Internet measurement (pp. 121-134). ACM. 2004, doi: 10.1145/1028788.1028804

[18] Auld, T., Moore, A. W., & Gull, S. F. Bayesian neural networks for internet traffic classification. Neural Networks, IEEE Transactions on, 18(1), 223-239. 2007, doi: 10.1109/TNN.2006.883010

[19] Erman, J., Mahanti, A., Arlitt, M., & Williamson, C. Identifying and discriminating between web and peer-to-peer traffic in the network core. In Proceedings of the 16th international conference on World Wide Web (pp. 883-892). ACM. 2007.

[20] Soysal, M., & Schmidt, E. G. Machine learning algorithms for accurate flow-based network traffic classification: Evaluation and comparison. Performance Evaluation, 67(6), 451-467. 2010.

[21] Lee, S., Kim, H., Barman, D., Lee, S., Kim, C. K., Kwon, T., & Choi, Y. Netramark: a network traffic classification benchmark. ACM SIGCOMM Computer Communication Review, 41(1), 22-30. 2011.

[22] Zhang, J., Chen, C., Xiang, Y., Zhou, W., & Xiang, Y. Internet traffic classification by aggregating correlated naive bayes predictions. Information Forensics and Security, IEEE Transactions on, 8(1), 5-15. 2013.

[23] Moore, A., Zuev, D., & Crogan, M. Discriminators for use in flow-based classification. Queen Mary and Westfield College, Department of Computer Science. 2005.

[24] Este, A., Gringoli, F., & Salgarelli, L. Support vector machines for TCP traffic classification. Computer Networks, 53(14), 2476-2490. 2009.

[25] Rodríguez-Teja, F., Martinez-Cagnazzo, C., & Castro, E. G. Bayesian classification: methodology for network traffic classification combination. In Proceedings of the 6th International Wireless Communications and Mobile Computing Conference (pp. 769-773). ACM. 2010, doi:10.1145/1815396.1815572

[26] Williams, N., Zander, S., & Armitage, G. A preliminary performance comparison of five machine learning algorithms for practical IP traffic flow classification. ACM SIGCOMM Computer Communication Review, 36(5), 5-16. 2006.

[27] Kim, H., Claffy, K. C., Fomenkov, M., Barman, D., Faloutsos, M., & Lee, K. Internet traffic classification demystified: myths, caveats, and the best practices. In Proceedings of the 2008 ACM CoNEXT conference (p. 11). ACM. 2008.

[28] Haffner, P., Sen, S., Spatscheck, O., & Wang, D. ACAS: automated construction of application signatures. In Proceedings of the 2005 ACM SIGCOMM workshop on Mining network data (pp. 197-202). ACM. 2005.

[29] Bernaille, L., Teixeira, R., Akodkenou, I., Soule, A., & Salamatian, K. Traffic classification on the fly. ACM SIGCOMM Computer Communication Review, 36(2), 23-26. 2006.

[30] Gómez Sena, G., & Belzarena, P. Early traffic classification using support vector machines. In Proceedings of the 5th International Latin American Networking Conference (pp. 60-66). ACM. 2009.

[31] Li, W., Canini, M., Moore, A. W., & Bolla, R. Efficient application identification and the temporal and spatial stability of classification schema. Computer Networks, 53(6), 790-809.2009.

[32] Palmieri, F., & Fiore, U. A nonlinear, recurrence-based approach to traffic classification. Computer Networks, 53(6), 761-773. 2009.