Characterizing and Predicting Internet Users Hidden Behaviors with Markovian Models

Paulo Salvador, António Nogueira, Eduardo Rocha
Instituto de Telecomunicações, DETI, University of Aveiro
Aveiro, Portugal
E-mails: salvador@ua.pt, nogueira@ua.pt, erocha@av.it.pt

***Abstract*—The ability to characterize and predict Internet users behaviors in environments where only layer 2 statistics are available can be very important for a network operator. At network entry points, like Wi-Fi or WiMax access points or UMTS or LTE base stations, the operator can perform a low level monitoring of the communications independently of the data encryption level and even without being associated with the network itself. Based on this low level network data, it is possible to infer the user behavior, optimize the access service and offer new security threat detection services. The user behavior inference consists in identifying the underlying web application that is responsible by the layer 2 traffic at different time instants and characterize the usage dynamics of the different web applications. Many identification methodologies have been proposed over the last years to classify/identify IP applications, including port-based analysis, deep packet inspection, behavior-based approaches and learning theory, each one having its own advantages and drawbacks. However, all these methodologies fail when only low level statistics are available or under data encryption restrictions. We propose the use of multiscaling traffic characteristics to differentiate between different web applications and the use of a Markovian model to characterize the dynamics of the user actions over time. By applying this methodology to Wi-Fi layer 2 traffic generated by users accessing different common web services/contents through HTTP (namely social networking, web news and web-mail applications), it was possible to achieve a good matching and prediction of the users behaviors. The results show that the proposed multiscaling traffic Markovian model has the potential to identify, model and predict Internet users behaviors based only on layer 2 traffic statistics.**

Keywords - User profiling; wavelets; multiscaling behavior; multivariate Gaussian distributions; Markovian modeling; behavior prediction.

I. Introduction

Identifying different behaviors of Internet users by analyzing the application types they are running is the key issue of many crucial network monitoring and management tasks. Basic network management functions such as quality of service improvement, network equipment optimization and security threats detection are all based on the ability to accurately classify network traffic into the right corresponding application and describe users behavior over time. Most existing approaches are based on static information about the applications (such as the name and type of the application, its owner, the execution time, or the host on which the application was executed). However, such approaches are not applicable to scenarios involving low level monitoring, traffic encryption or under stringent confidentiality requirements, since they rely on analyzing specific fields of the packet header.

So, this paper proposes the use of multiscaling traffic characteristics to differentiate between different web applications and the use of a Markovian model [1], [2] to characterize the various dynamics of the user actions over time. This methodology will be able to identify and predict the different user behaviors, even if this information is somehow hidden when performing a classical statistical analysis of the generated traffic.

Besides, the proposed methodology can be applied to scenarios where existing identification approaches are not applicable or have limited efficiency, like low level monitoring and service optimization at Wi-Fi [3] or WiMax [4] access points and Universal Mobile Telecommunications System (UMTS) [5], [6] or Long Term Evolution (LTE) [7] base stations.

One of the first and most common forms of traffic classification is port-based classification, which relies on the port numbers employed by the application at the transport layer. However, since many modern applications use dynamic ports negotiation, port-based classification became ineffective [8], [9], with accuracy ranges between 30% and 70%. Chronologically, the next proposed classification technique was deep packet inspection (DPI) or payload-based classification, which requires the inspection of the packets' payload: this classifier extracts the application payload from the layer 4 data unit and searches for a signature that can identify the flow type. Although DPI is widely used by today's traffic classifier vendors, being very accurate [10], [9] for some scenarios, it is unable to deal with low level or encrypted data. Very efficient classification techniques that perform traffic identification without accessing user data were proposed [11], but they also rely on layer 3 and layer 4 traffic statistics that may not be available for the operator at specific network entry points, can be protected by encryption or restricted by confidentiality requirements.

In this work, we propose a methodology for the differentiation of Internet applications based on the multiscaling statistical analysis of low level traffic, together with a modeling approach of the user preferences over time. It is known that several frequency components are introduced by mechanisms operating at different scales of analysis, including user interactions, flow sessions and individual packets dynamics. This creates characteristic multiscaling signatures that can be used to perform an accurate differentiation of the different web applications. A wavelet scalogram [12], which describes the signal energy simultaneously on a frequency and time domain, is constructed based on the wavelet multiscale decomposition of a traffic counting process and can be used to create multiscaling statistical signatures for each web application. The wavelet scalogram communicates the time frequency localization property of the discrete wavelet transform, being possible to capture the correlation that exists between the time variability of the process and the different scales.

The results obtained by applying the proposed methodology to layer 2 traffic promiscuously captured in the vicinity of a Wi-Fi network access point (without authenticating) show that it is able to achieve a good identification accuracy. It was possible to identify, model and predict the behavior of users accessing three common web applications: social networking (without chatting and game interactions), news web journals and web-mail.

For validation purposes, the ground-truth of the data was created by asking a pre-determined set of users to replicate their traditional Internet behavior using a controlled environment (user terminals and network).

The remaining part of this paper is organized as follows: Section II presents some of the most relevant related work on statistical classification of web applications and user behavior modeling; Section III presents some important background on multiscaling analysis; Section IV presents the details of the proposed identification methodology and behavior model; Section V presents the results of a proof-of-concept of the methodology and, finally, Section VI presents some brief conclusions about the presented model and identification methodologies.

II. Related work

The statistical approach to classification is based on collecting statistical data of the network flow, such as the mean packet size, flow duration, number of bytes per time interval, number of packets per time interval, etc. The statistical paradigm relies on the assumption that each application has a unique distribution of properties that represents it and can be used to univocally identify it. This approach has been the subject of intensive research in recent years.

First of all, Paxson *et al.* [13] established a relationship between flow application type and flow properties (such as the number of bytes and the flow duration). In [14], the authors proposed a methodology for separating chat traffic from other Internet traffic using statistical properties such as packet sizes, number of bytes, duration and packets inter arrival times. In [15], Mcgregor *et al.* explored the possibility of forming clusters of flows based on flow properties such as packet size statistics (e.g., minimum and maximum), byte count, idle times, etc., using an expectation maximization (EM) algorithm to find the clusters' distribution density functions. A study focusing on identifying flow application categories rather than specific individual applications was presented in [16]. Although it was limited by a small dataset, the authors have been able to show that the k-nearest neighbor algorithm and other techniques can achieve good results, correctly identifying around 95% of the flows. In reference [17], the authors were able to obtain an average success rate of 87% in the separation of individual applications using an EM based clustering algorithm. In [18], Moore *et al.* studied the basic Navie Bayes algorithm, enhanced by certain refinements, showing that it is able to achieve an accuracy level of 95%.

In [19], realtime classification was addressed by studying the feasibility of application identification at the beginning of a TCP connection: based on an analysis of packet traces collected on eight different networks, the authors found that it is possible to distinguish the behavior of an application from the observation of the size and the direction of the first few packets of the TCP connection. Three techniques were applied to cluster TCP connections: K-Means, Gaussian Mixture Model and spectral clustering. Crotti *et al.* [20] presented a realtime classification mechanism based on three simple properties of the captured IP packets: their size, inter-arrival time and arrival order. Based on new structures called protocol fingerprints, which express these quantities in a compact way, and on a simple classification algorithm based on normalized thresholds, the proposed technique showed promising results on classifying of a reduced set of protocols. In [21], a traffic classification approach based on Support Vector Machines (SVM) was proposed: using a simple optimization algorithm, a statistical traffic classifier was able to perform correctly with only a few hundred samples for training. Note that these algorithms were tested only against basic application protocols. Encrypted applications communications add additional constraints to the detection problem by making the traffic packet headers and data inaccessible to network based monitoring systems. Therefore, the detection methods that rely on packets headers/data information are completely inappropriate in encrypted communications scenarios [8], [22].

Bar-Yanai *et al.* [23] introduces a hybrid statistical algorithm that integrates the k-nearest neighbors and k-means machine learning algorithms. The proposed algorithm is fast, accurate and is insensitive to encrypted traffic, overcoming several weaknesses of the DPI approach (like asymmetric routing and packet ordering). The strength of the algorithm was demonstrated on encrypted BitTorrent, which is known to use packet encryption, port alternation and packet padding (on initial flow packets) to avoid detection.

The BLINC [11] approach is based on observing and identifying patterns of host behavior at the transport layer, analyzing the social, functional and application level patterns. The fact that this approach relies on layer 3 and layer 4 traffic statistics makes it impossible to be used by an operator in certain entry points of the network where only low level data is available.

Rocha *et al.* [24] presented a methodology for the detection of security attacks and the classification of Internet flows that relies on multidimensional Gaussian distributions [25]. In this way, it is possible to account for the correlation between the values that are obtained for the different dimensions, allowing to infer even more accurate probability distributions. The proposed approach starts by performing a multiscale analysis to the sampled IP data-streams, obtaining multiscale estimators for all streams; the estimators are subsequently processed by mapping a dimension to each timescale, so that the multivariate distributions (for each protocol) can be inferred; an algorithm will then find the dimensions where the separation between the several distributions is most noticeable and each of the traffic streams is then classified according to the probability of belonging to each one of the inferred distributions.

III. Multiscaling analysis

The inability of conventional Fourier analysis to preserve the time dependence and describe the evolutionary spectral characteristics of non-stationary processes requires tools that allow time and frequency localization. Wavelet transforms can provide information concerning both time and frequency, which allows local, transient or intermittent components to be elucidated [12]. Such components are often obscured due to the averaging inherent within spectral only methods, like Fast Fourier Transform (FFT) [26], for example.

Wavelets are mathematical functions that are used to divide a given signal into its different frequency components. They consist of a short duration wave that has limited energy. Wavelets enable the analysis of each one of the signal components in an appropriate scale. Starting with a mother wavelet $\psi(t)$, a family $\psi_{\tau,s}(t)$ of "wavelet daughters" can be obtained by simply scaling and translating $\psi(t)$:

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{|s|}}\psi(\frac{t-\tau}{s}) \qquad (1)$$

where s is a scaling or dilation factor that controls the width of the wavelet (the factor $\frac{1}{\sqrt{|s|}}$ being introduced to guarantee preservation of the energy, $\|\psi_{\tau,s}\| = |\psi|$) and τ is a translation parameter controlling the location of the wavelet. Scaling a wavelet simply means stretching it (if $|s| > 1$) or compressing it (if $|s| < 1$), while translating it simply means shifting its position in time.

Given a signal $x(t) \in L^2(\Re)$ (the set of square integrable functions), its Continuous Wavelet Transform (CWT) with respect to the wavelet ψ is a function of time (τ) and scale (s), $W_{x;\psi}(\tau, s)$, obtained by projecting $x(t)$ onto the wavelet family $\{\psi_{\tau,s}\}$:

$$W_{x;\psi}(\tau, s) = \int_{+\infty}^{-\infty} x(t)\frac{1}{\sqrt{|s|}}\psi(\frac{t-\tau}{s})dt \qquad (2)$$

By analogy with the terminology used in the Fourier case, the energy components of the signal are given by the square of the CWT components of the signal and the (local) Wavelet Power Spectrum (sometimes called Scalogram or Wavelet Periodogram) is defined as the normalized energy over time and scales:

$$E_x(\tau, s) = 100\frac{|W_{x;\psi}(\tau, s)|^2}{\sum_{\tau'}\sum_{s'}|W_{x;\psi}(\tau', s')|^2} \qquad (3)$$

Scalograms reveal much information about the nature of non-stationary processes that was previously hidden, so they are applied to a lot of different scientific areas: diagnosis of special events in structural behavior during earthquake excitation, ground motion analysis, transient building response to wind storms, analysis of bridge response due to vortex shedding, among others [27].

IV. Multiscaling Behavior Modeling

A. Multiscale traffic data

Let us assume that process $x(t)$ represents a counting statistic of a layer 2 traffic trace to and from a specif user terminal (e.g., number of frames on the upload direction, number of bytes in the download direction, etc.). The user is identified by a layer 2 address depending on the underlying communications technology. It is possible to apply a multiscaling analysis to process $x(t)$ by calculating the scalogram using equation (3). We characterize the multiscale user behavior by the estimator of the standard deviation of that user's traffic energy within a time window for a set of timescales. Therefore, a traffic process energy standard deviation at time interval k and time scale s using a sliding time window of width W can be defined as:

$$\hat{D}_x(k, s) = \sqrt{\frac{1}{W-1}\sum_{\tau\in[k-W,k]}\left(E_x(\tau, s) - \overline{E_x(k, s)}\right)^2} \qquad (4)$$

with $k = \{W, W+1, W+2, \ldots\}$ and

$$\overline{E_x(k, s)} = \frac{1}{W}\sum_{\tau'\in[k-W,k]} E_x(\tau', s) \qquad (5)$$

Choosing J timescales ($\{s_1, s_2, \ldots, s_J\}$) of interest, it is possible to define a vector $B_{x,k}$ that describes the inferred localized multiscaling characteristics (at time interval k) of the traffic process x:

$$B_{x,k} = \{\hat{D}_x(k, s_j), j = 1, \ldots, J\} \qquad (6)$$

B. Markov Modulated multivariate Gaussian Processes Model

The proposed discrete time Markov Modulated multivariate Gaussian Process (dMMGP) model characterizes position and mobility of a subject based on the following assumptions: (i) the multiscaling behavioral metrics for the use of a specific web application can be described by a multivariate Gaussian distribution, (ii) the time scales of importance can be predetermined, (iii) a ground truth for the web applications usage multiscaling characteristics can be pre-established and (iv) the transition between applications can be described by an underlying (homogeneous) Markov chain where each state maps the multiscaling behavior characteristics of a specific web application usage.

The dMMGP can then be described as a J-dimensional random process (B) with a multivariate Gaussian distribution that characterizes the behavior of a user in an universe of A possible applications in a J-dimensional environment (for J time scales of importance), whose parameters are a function of the state (S) of the modulator Markov chain (B, S) with A states. The dMMGP model states will map the applications multiscale characteristics and the dMMGP model transitions will define the user behavior/dynamics on the usage of the different applications. The former will be inferred based on pre-established ground truth (set of known flows) for the web applications multiscaling characteristics and the later will be inferred based on the dynamics of the mapping of a set of flows of specific users to the application multiscale characteristics (i.e. model states).

More precisely, the (homogeneous) Markov chain

$$(B, S) = \{(B_k, S_k),\, k = 0, 1, \ldots\}$$

with state space $\mathbb{R}^J \times U$, with $U = \{1, 2, \ldots, A+1\}$, is a dMMGP if and only if for $k = 0, 1, \ldots,$

$$P(B_{k+1} = \mathbf{b},\, S_{k+1} = n | S_k = m) = p_{mn}\Gamma_n(\mathbf{b}) \qquad (7)$$

where $\mathbf{b} \in \mathbb{R}^J$ is a generic multiscale component in a J-dimensional environment, p_{mn} represents the probability of a transition from state m to state n of the underlying Markov chain in time interval $[k, k+1]$, and

$$\Gamma_a(\mathbf{b}) = (2\pi)^{-\frac{J}{2}}\mathbf{\Sigma}_a^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{b}-\mathbf{m}_a)^T \Sigma_a^{-1}(\mathbf{b}-\mathbf{m}_a)} \qquad (8)$$

is the multivariate Gaussian distribution of the multiscaling characteristics of application a flows, it is centered in m_a and has covariance matrix Σ_a.

Whenever (7) holds, we say that (B, S) is a dMMGP with a set of modulating states with size A and parameter matrices $\mathbf{P}$, $\mathbf{M}$ and $\mathbf{S}$. Matrix $\mathbf{P}$ is the transition probability matrix of the modulating Markov chain S,

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1A} \\ p_{21} & p_{22} & \cdots & p_{2A} \\ \cdots & \cdots & \cdots & \cdots \\ p_{A1} & p_{A2} & \cdots & p_{AA} \end{bmatrix} \qquad (9)$$

while matrix $\mathbf{M}$ defines the mean values of each multiscaling Gaussian distribution:

$$\mathbf{M} = \begin{bmatrix} \mathbf{m}_1 & \mathbf{m}_2 & \ldots & \mathbf{m}_A \end{bmatrix} \qquad (10)$$

where $\mathbf{m}_a$ is a $J \times 1$ vector. Matrix $\mathbf{S}$ contains the covariance (sub-)matrices of each multiscaling Gaussian distribution:

$$\mathbf{S} = \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{\Sigma}_2 & \ldots & \mathbf{\Sigma}_A \end{bmatrix} \qquad (11)$$

where $\mathbf{\Sigma}_a$ is a $J \times J$ matrix. Moreover, we denote by $\mathbf{\Pi} = [\pi_1, \pi_2, \ldots, \pi_A]$ the stationary distribution of the underlying Markov chain.

Matrix $\mathbf{P}$ will be unique for each user, and will characterize his/her behavior on the usage of the applications characterized by matrices $\mathbf{M}$ and $\mathbf{S}$. The overall multiscaling behavior of a user can be statistically described by a stationary probability density defined by a weighted sum of A multivariate Gaussian distributions:

$$f(\mathbf{b}) = \sum_{a=1}^{A} \pi_a \Gamma_a(\mathbf{b}), \mathbf{b} \in \mathbb{R}^J \qquad (12)$$

where $\mathbf{b}$ is a multiscale component that belongs to the J-dimensional domain of chosen timescales.

C. Model Inference Procedure

Assuming that we have a ground-truth for a set of A web applications, analyzed over F flows, over K time windows in J timescales of interest, we can define the multiscale profile of an application $a(a = 1, \ldots, A)$ as $G_{a,f,k}$, inferred using equation (6) considering that process $x(t)$ is the f-th flow of application a, with $a = 1, \ldots, A$, $f = 1, \ldots, F$ and $k = 1, \ldots, K$, i.e.:

$$G_{a,f,k} = B_{x,k}, x \leftrightarrow \text{flow } f\text{of application } a \qquad (13)$$

The $\mathbf{M}$ and $\mathbf{S}$ matrices of the dMMPGP model can then be inferred as

$$\mathbf{m}_a = \frac{1}{KF}\sum_{f=1}^{F}\sum_{k=1}^{K} G_{a,f,k} \qquad (14)$$

$$\mathbf{\Sigma}_a = \frac{1}{KF-1}\sum_{f=1}^{F}\sum_{k=1}^{K}\Big(\left(G_{a,f,k} - \mathbf{m}_a\right)\left(G_{a,f,k} - \mathbf{m}_a\right)^T\Big) \qquad (15)$$

The final step of the inference procedure is to infer matrix $\mathbf{P}$, i.e. the transition probabilities between the states defined in the first step. This task is achieved by probabilistically mapping each multiscaling behavior of each unknown flow trace $x(t)$ $B_{x,k}, k = 1, \ldots, K$ to one state/application and then averaging the probabilistic transitions between states, according to a probability vector:

$$\mathbf{q}_k = \{\Gamma_1(B_{x,k}), \ldots, \Gamma_A(B_{x,k})\}, k = 0, 1, \ldots, K \qquad (16)$$

D. Behavior Prediction

Defining $\mathbf{c}_k = \{c_{k,a} : a = 0, 1, \ldots, A\}, k = 0, 1, \ldots, K$, where $\mathbf{c}_k$ is the probability vector defining that within time-window $[k - W, k]$ the user is using application a, and based on equation (12) we can define the multivariate distribution of the predicted multiscaling behavior of the user in a future time-window (z observations in the future) as:

$$\sum_{a=1}^{A} \mathbf{c}_{k+z}\Gamma_a \qquad (17)$$

with

$$\mathbf{c}_{k+z} = \mathbf{c}_k\mathbf{P}^z \qquad (18)$$

where $\mathbf{c}_{k+z}$ represents the probabilistic vector that quantifies the probability of a web application to be in use k time windows in the future.

V. Proof of concept

A. Data-set

The test data-set was obtained by capturing, in promiscuous mode, the layer 2 traffic having as source or destination a specific Wi-Fi network access point. The traffic capture was performed without authenticating to the network and consisted only of 802.11 frames. In a controlled environment, where all terminals were using a bare installation of Linux with a daemon that recorded all browse requests, a set of invited users were asked to access and use their usual web applications, maintaining their typical behavior. This approach allowed us to create the ground-truth of a mapping between layer 2 data traces and their originating users and web applications. Within the context of this paper and this proof of concept, we only used the data traces that were created by users accessing three general web applications: social networking, namely Facebook (without chatting and game interactions), news web journals and web-mail access. The total number of data sets was divided in two: the first half was used to infer the underlying dMMGP model of the behavior of each application and user, while the second half of the data sets was used to validate the inferred models by comparing the predicted multiscale behavior (and associated web application usage sequence) of each user. The raw statistical process used was the amount of bytes transmitted from the Wi-Fi access point to each user, sampled every 0.1 seconds. Sampling the raw statistics in 0.1 seconds allows our method to measure and incorporate some of the most characteristic multiscale dynamics of an application: (i) the lower timescales that are strictly related with the way that specific application handles the multiple data sessions, (ii) the medium timescales that are related with the application algorithmic dynamics and (iii) the higher timescales that reflect mainly the user interactions dynamics [28]. For the purpose of the model inference, we use time windows with a width of 120 seconds ($W = 1200$) and considered time windows in 20 seconds interval. The choice of these values is a tradeoff between the amount of (past) data necessary to fully characterize the traffic dynamics and the amount of data that can be process and analyzed in pseudo-real time. The heavier computational task that it is the construction and update of the behavior models which are made off-line and is not an issue. However, to perform the application and user identification the measured data must be matched with previously inferred models in pseudo real-time. The interval between windows of classification was chosen in order to minimize the delay between the moment of an user application change and its effective detection by our methodology. With an appropriate choice of parameters, namely window size and interval of processing, this methodology is fully scalable since the computation power required is proportional to the amount of traffic (number of users) under analysis.

Figures 1 and 2 depicted the 80% and 90% quantile frontiers of the inferred multivariate Gaussian distributions of the multidimensional characteristics of each application (using just 3 timescales) for all users. These distributions reveal that the multiscale characteristics of the three web applications are distinct and have a small overlap in the universe of the three dimensions/scales considered.

After inferring the underlying dMMGP model, we use the test data traces to test the precision of the model in identifying the current web application of an user every 20 seconds. In this test, we were able to obtain a precision of 72.4% of correctly classified windows and the identification results presented in Table V-A. The results show a very good agreement between the identified web application and the real application, considering the reduced amount of information (in terms of raw data and time span of the observation) used for the identification.

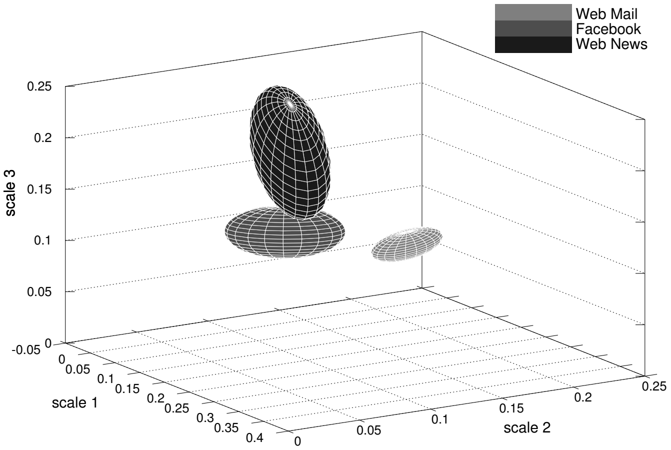

Figure 1. 80% quantile frontiers of the inferred multivariate Gaussian distributions of the multidimensional characteristics of each application.

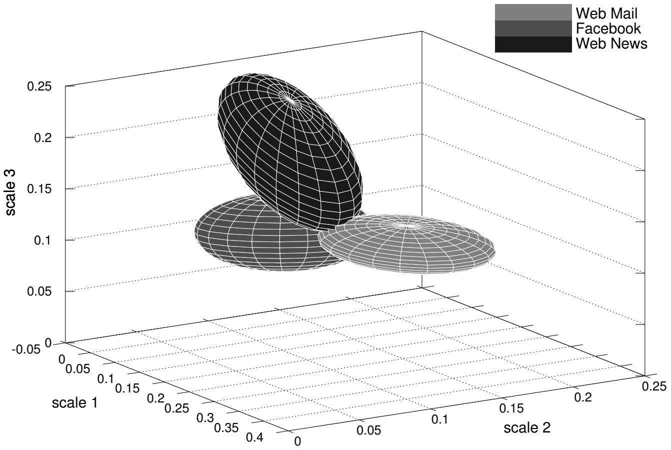

Figure 2. 90% quantile frontiers of the inferred multivariate Gaussian distributions of the multidimensional characteristics of each application.

Using the test data traces to test the precision of the model in identifying the web applications that are in use 60 seconds in the future we obtained a precision of 55.3% of correctly classified windows. The results show that the identification/predicting results are still significantly above the pure random guess.

The results show that our methodology was able to obtain very good classification and prediction results considering the reduced amount of information (only network layer 2 sampled statistics) and that the web applications under consideration may, in some particular cases, be very similar. Most of the

	Web-Mail	Facebook	Web-news
Web-mail	69.95%	7.84%	22.20%
Facebook	4.12%	83.13%	12.74%
Web-news	7.08%	24.35%	68.55%

Table I
IDENTIFICATION OF THE CURRENT WEB APPLICATION RESULTS.

ISBN: 978-1-61208-209-7

errors can be explained by the fact that some Web-news pages are very similar to social networking applications pages and even incorporate social network features within its own Web-pages. Also, when the Web-news web pages have less content the user dynamics may get similar to Web-mail or Facebook interactions (i.e., small data chunks exchanged at small intervals).

VI. Conclusion and Future Work

We presented a novel approach that uses multiscaling traffic characteristics to differentiate between different web applications and a Markovian model that is able to characterize the dynamics of user actions over time. By applying this methodology to Wi-Fi layer 2 traffic generated by users accessing different common web services/contents through HTTP (namely, social networking, web news and web-mail applications), it was possible to achieve a good matching and prediction of the users behaviors. Our methodology may be applied to preallocate resources in network access points based on past user behavior and pseudo real-time predictions of short term requirements.

As future work, we plan to test our methodology incorporating more applications with completely different behavior (such as video streaming, P2P file transferring, online games, etc.). This will required the improvement of the inner algorithms of the methodology to accommodate multiple and dynamic timescales ranges. Moreover, our short term plans include the developing of a prototype and test in a 3G/4G network base station for optimal dynamic allocation of resources.

References

[1] P. Salvador, R. Valadas, and A. Pacheco, "Multiscale fitting procedure using Markov modulated poisson processes," *Telecommunication Systems Journal, Kluwer Academic Publishers*, vol. 23, no. 1-2, pp. 123–148, 2003.

[2] A. Pacheco, L. C. Tang, and N. U. Prabhu, *Markov-modulated processes & semiregenerative phenomena*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2009.

[3] "IEEE Standard for Information Technology - Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications," *IEEE Std 802.11-2007 (Revision of IEEE Std 802.11-1999)*, pp. 1 –1076, 12 2007.

[4] A. Ghosh, D. Wolter, J. Andrews, and R. Chen, "Broadband wireless access with WiMax/802.16: current performance benchmarks and future potential," *Communications Magazine, IEEE*, vol. 43, no. 2, pp. 129 – 136, feb. 2005.

[5] M. van Nielen, "UMTS: a third generation mobile system," in *Personal, Indoor and Mobile Radio Communications, 1992. Proceedings, PIMRC '92., Third IEEE International Symposium on*, oct 1992, pp. 17 –21.

[6] E. Dahlman, B. Gudmundson, M. Nilsson, and A. Skold, "UMTS/IMT-2000 based on wideband CDMA," *Communications Magazine, IEEE*, vol. 36, no. 9, pp. 70 –80, sep 1998.

[7] D. Astely, E. Dahlman, A. Furuskar, Y. Jading, M. Lindstrom, and S. Parkvall, "Lte: the evolution of mobile broadband," *Communications Magazine, IEEE*, vol. 47, no. 4, pp. 44 –51, april 2009.

[8] H. Kim, K. Claffy, M. Fomenkov, D. Barman, M. Faloutsos, and K. Lee, "Internet traffic classification demystified: myths, caveats, and the best practices," in *Proceedings of the 2008 ACM CoNEXT Conference*, ser. CoNEXT '08. New York, NY, USA: ACM, 2008, pp. 11:1–11:12. [Online]. Available: http://doi.acm.org/10.1145/1544012.1544023

[9] S. Sen, O. Spatscheck, and D. Wang, "Accurate, scalable in-network identification of p2p traffic using application signatures," in *Proceedings of the 13th international conference on World Wide Web*, ser. WWW '04. New York, NY, USA: ACM, 2004, pp. 512–521. [Online]. Available: http://doi.acm.org/10.1145/988672.988742

[10] A. Madhukar and C. L. Williamson, "A longitudinal study of p2p traffic classification," in *Proceedings of the IEEE MASCOTS*, 2006, pp. 179–188.

[11] T. Karagiannis, K. Papagiannaki, and M. Faloutsos, "BLINC: multilevel traffic classification in the dark," in *Proceedings of the 2005 conference on Applications, technologies, architectures, and protocols for computer communications*, ser. SIGCOMM '05. New York, NY, USA: ACM, 2005, pp. 229–240. [Online]. Available: http://doi.acm.org/10.1145/1080091.1080119

[12] J. Byrnes, K. A. Hargreaves, and K. Berry, *Wavelets and their Applications*. Springer, 1994.

[13] V. Paxson, "Empirically derived analytic models of wide-area tcp connections," *IEEE/ACM Trans. Netw.*, vol. 2, pp. 316–336, August 1994. [Online]. Available: http://dx.doi.org/10.1109/90.330413

[14] C. Dewes, A. Wichmann, and A. Feldmann, "An analysis of internet chat systems," in *Proceedings of the 3rd ACM SIGCOMM conference on Internet measurement*, ser. IMC '03. New York, NY, USA: ACM, 2003, pp. 51–64. [Online]. Available: http://doi.acm.org/10.1145/948205.948214

[15] A. McGregor, M. Hall, P. Lorier, and J. Brunskill, "Flow clustering using machine learning techniques," *LNCS*, vol. 3015, pp. 205–214, 2004.

[16] M. Roughan, S. Sen, O. Spatscheck, and N. G. Duffield, "Class-of-service mapping for qos: a statistical signature-based approach to ip traffic classification." in *Internet Measurement Conference'04*, 2004, pp. 135–148.

[17] S. Zander, T. T. T. Nguyen, and G. J. Armitage, "Automated traffic classification and application identification using machine learning." in *LCN'05*, 2005, pp. 250–257.

[18] A. Moore and D. Zuev, "Internet traffic classification using bayesian analysis techniques." in *ACM SIGMETRICS*, 2005, pp. 50–60.

[19] L. Bernaille, R. Teixeira, and K. Salamatian, "Early application identification," in *Proceedings of the 2006 ACM CoNEXT conference*, ser. CoNEXT '06. New York, NY, USA: ACM, 2006, pp. 6:1–6:12. [Online]. Available: http://doi.acm.org/10.1145/1368436.1368445

[20] M. Crotti, M. Dusi, F. Gringoli, and L. Salgarelli, "Traffic classification through simple statistical fingerprinting," *SIGCOMM Comput. Commun. Rev.*, vol. 37, pp. 5–16, January 2007. [Online]. Available: http://doi.acm.org/10.1145/1198255.1198257

[21] A. Este, F. Gringoli, and L. Salgarelli, "Support vector machines for tcp traffic classification," *Comput. Netw.*, vol. 53, pp. 2476–2490, September 2009. [Online]. Available: http://portal.acm.org/citation.cfm?id=1576850.1576885

[22] T. T. T. Nguyen and G. J. Armitage, "A survey of techniques for internet traffic classification using machine learning." *IEEE Communications Surveys and Tutorials*, pp. 56–76, 2008.

[23] R. Bar-Yanai, M. Langberg, D. Peleg, and L. Roditty, "Realtime classification for encrypted traffic." in *SEA'10*, 2010, pp. 373–385.

[24] E. Rocha, P. Salvador, and A. Nogueira, "Detection of illicit network activities based on multivariate gaussian fitting of multi-scale traffic characteristics," in *2011 IEEE International Conference on Communications (ICC 2011)*, June 2011, pp. 1–6.

[25] K. S. Miller, *Multidimensional Gaussian Distributions*. John Wiley & Sons Inc, 1964.

[26] E. Brigham, *Fast Fourier Transform and Its Applications*. Prentice Hall, 1988.

[27] K. Gurley and A. Kareem, "Applications of wavelet transforms in earthquake, wind, and ocean engineering," *Engineering Structures*, no. 21, pp. 149–167, 1999.

[28] M. Crovella and B. Krishnamurthy, *Internet Measurement: Infrastructure, Traffic and Applications*. Wiley, 2006.