Modeling and Analysis of Web Usage and Experience Based on Link-Level Measurements

Junaid Shaikh, Markus Fiedler, Patrik Arlos
School of Computing
Blekinge Institute of Technology
Karlskrona, 371 79 Sweden
Email: {junaid.junaid, markus.fiedler, patrik.arlos}@bth.se

Denis Collange
R&D Orange Labs
Sophia Antipolis France
Email: denis.collange@orange.com

***Abstract*—Internet traffic monitoring and analysis have been playing a crucial role in understanding and characterizing user behavior on the web. In particular, ON-OFF models capture the essential phases of user communication with web servers. The OFF phases reflect both deliberate and accidental gaps in the traffic flow. In this paper, we present a passive monitoring and analysis method devised to assist in the identification of such traffic gaps that may result in the degradation of Quality of Experience (QoE). Our first contribution consists in a revised ON-OFF model to cater for OFF times reflecting accidental gaps which are induced by the network. Second, a wavelet-based criterion is proposed to differentiate between the network-induced traffic gaps and user think times. The proposed method is intended to be implemented in near-real-time as it does not require any deep packet inspection. Both web service providers and network operators may use this method to obtain objective evidence of the appearance of QoE problems from link-level measurements.**

I. Introduction

There has been an enormous growth in the deployment and usage of wireless networks during the recent years. The performance of these networks is highly varying due to their availability and coverage issues. Particularly, the outages in the traffic in wireless networks are quite frequent and they result in longer waiting times on the web. We use the term ”outage” here to refer to the events causing short-term temporary disruptions in the data transfer due to problems in the network. Outages and competition for resources lead to gaps in the network traffic, and they are perceived badly once their consequences show up in the user interface.

Moreover, the random appearances of such outages result in bursty traffic and sudden degradation in the Quality of Service (QoS). This sudden degradation in the QoS affects the user Quality of Experience (QoE) significantly. On the link level, these outages could be seen in the form of gaps in the traffic flows. For the Internet Service Providers (ISPs), it is important to monitor these traffic gaps resulting from the network outages in order to obtain hints on how to improve their services. However, the outages due to the network problems are not the only reasons for gaps in the traffic. The gaps in traffic may just be due to the inactivity of the user, which we call the user think time between two transactions. Hence, it is important to distinguish between both reasons that lead to gaps in the traffic.

Successful differentiation between these two types of gaps enables the ISPs to identify the network outages by monitoring the traffic flow on the network. We propose a fast approach to keep track of the user think times that doesn’t require any deep inspection of packets for identifying the end of a user transfer.

In this paper, we provide the following contributions. First, we discuss the characteristics of the gaps caused due to the user inactivity on the web. Second, we present an analysis of the features of the gaps induced by the network, and how they could be used to differentiate between a smooth and a disturbed live video streaming transfer on the web. Third, we present a wavelet-based criterion to identify the traffic gaps caused by the problems in the network. We have targeted live video streaming on the web because the consequences of problems in the network can be experienced immediately in the form of freezes in the video. Every freeze results in the loss of information in the case of live video streaming and hence, in user dissatisfaction.

In [1], the authors identified gaps and user-perceived problems, but they didn’t quantify the boundary towards the think times. They decoded the stream and simulated the buffer content afterwards. Moreover, their study was based on Youtube video. According to the best of our knowledge, this study is first in provoking the discussion on the network-based criterion that differentiates between the user think times and the network outages and relates them to user-perceived video delivery issues on web. We also derive a rather simple criterion for discerning network outages from think time that can be evaluated in near-real-time.

The remainder of this paper is structured as follows. Section II describes the methodology of this work, Section III describes the ON-OFF models for the Web and presents the findings on the quantification of the user think times, and Section V quantifies the traffic gaps along with the comparison of smooth and disturbed web-based video transfers. Section VI presents the wavelet analysis of the captures traces and Section VII proposes the criteria for identifying the network-induced traffic gaps. Finally, Section VIII concludes the paper along with a short description of our short-term future work.

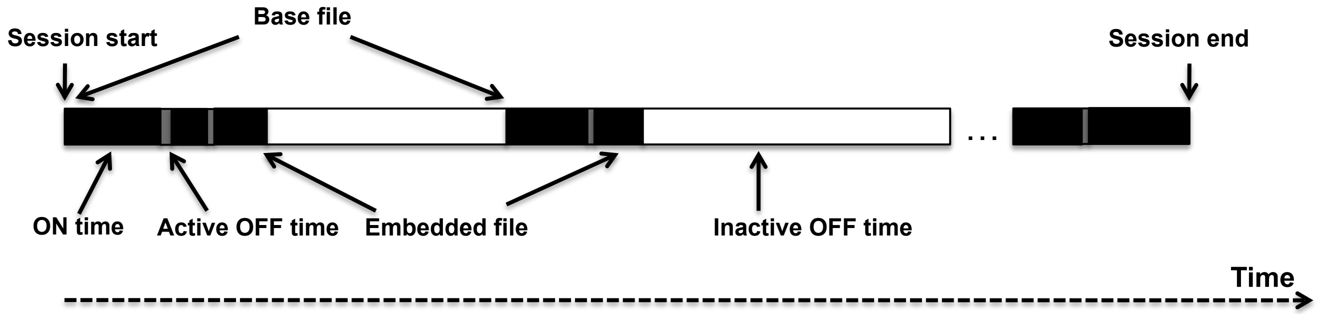

Fig. 1. A web session with serial transfers: One client, one server

II. Methodology

This section presents the methodology used for conducting this study. There are three sides of this study. First, it outlines the user think times on the web based on the previous studies. Second, it presents an analysis on the properties of two live video streaming transfers on web. And third, it proposes a criterion to monitor and detect the presence of the traffic gaps on different timescales.

To outline user think times, a literature review is done to understand how users launch their requests on the web. Our findings are summarized in Section IV.

We then analyzed the inter-packet times in order to be able to differentiate between a smooth and a disturbed web transfer. For this purpose, we did live web-based video streaming from a distant server via two different Wi-Fi networks: one at a home in Sweden and another at a hotel in Germany. The Wi-Fi at hotel was chosen because, usually, networks at hotel show signs of capacity shortage when many uncoordinated users are active at the same time, e.g. during the evenings. Video streaming was done using the Firefox web browser with embedded Flash player. At the same time, the traffic was captured via the Wireshark [2] traffic capturing tool. The Macintosh operating system version 10.6.3 was used on the computer with processor speed 2.53 GHz and 4 GB of memory. Each transfer is 180 s duration long. The captured traffic from the server to the client direction was used for the analysis of inter-packet times.

Subsequently, the user think times were compared to the inter-packet times of smooth and disturbed transfers. This allowed us to draw a borderline between (1) the duration of the gaps generated due to potential user think times between two transfers, and (2) the gaps due to inter-packet times within the same transfer. Finally, the wavelet analysis was performed on both the transfers to visualize the ON and OFF phases along with their frequency and duration on different timescales.

III. The ON-OFF model

The user web session could be characterized by an ON-OFF model as illustrated in Figure 1. Each web session may consist of the transfer of one or more web pages. Similarly, every web page may consist of one or more objects. To retrieve each of these objects, a request is sent from the client to the server. In this section, we will first describe the terms used in the ON-OFF model. Later, we will classify ON-OFF models based on the nature of the web pages and the user behavior.

A. ON times

An ON time during a session is defined by the time elapsed from the arrival of a request from the client side to the end of the corresponding response from the server side. The ON times are illustrated by the black boxes in the ON-OFF model shown in Figure 1.

B. OFF times

An OFF time in the ON-OFF model represents the silent time between two subsequent transactions. For example, the client, after receiving the last object from the server, may take some time referred to as an OFF time before launching the next request. During the OFF time, there is no packet containing the data seen on link-level. The OFF times are further classified into two categories: The active OFF times and the inactive OFF times.

1) Active OFF times: When a user requests a web page, which consists of multiple objects, the client-side web browser may retrieve those objects by sending automatic requests to the server. The time elapsed from the end of the previous response from the server to the arrival of the next automatically-generated request from the client-side web browser is called the active OFF time. The active OFF times are shown as grey boxes in Figure 1.

2) Inactive OFF times: An inactive OFF time is the time a user spends on viewing or reading the contents of the page when one or more objects are already retrieved from the server. Inactive OFF times are also called the user think times and are shown by the white boxes in Figure 1.

C. Base file

When a new webpage is requested, the first request which is generated by the client is for the object which is commonly referred to as the base file of the page.

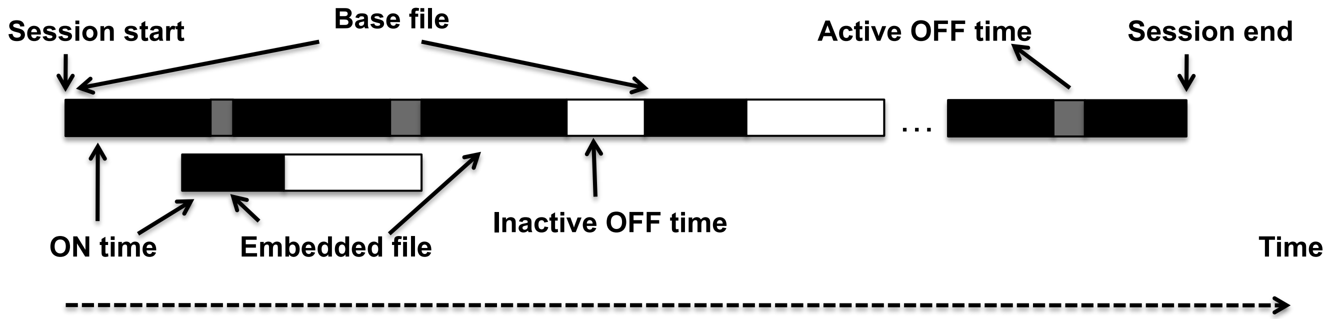

Fig. 2. Parallel transfers in a web session: One client, multiple server

D. Embedded file

The files that are retrieved subsequently after the base file of the page are called embedded files. An embedded file could be an inline image, a link to the another page and a video or a video player etc.

IV. A WEB SESSION

A simple web session could be defined as a sequence of requests made by a single client to a particular web server. It starts when a user requests for a page by typing a URL in the address bar, clicking on a hyperlink or clicking a bookmark. Either of these user actions generate a request for the base file from the server. After the base file is retrieved, the subsequent requests for the embedded objects are made automatically by the client-side web browser. The request for every embedded object is qualified by the active OFF time.

After the page is displayed, the user reads or views the contents of the page. For instance, the user may be reading a newspaper, watching a video or filling a form etc. This time is called the inactive OFF time or the user think time.

The structure of websites has changed significantly during the last few years. The emergence of Web 2.0 has fueled the popularity of mashups. A webpage is made up of tens and hundreds of objects, which are hosted by one or more servers. According to [3], on average each webpage is composed of more than 50 objects. When a user requests for a web page, the objects of the pages may get retrieved from several servers. Usually, these objects are not transferred serially, but in parallel to each other.

Figure 2 presents a web page transfer involving a parallel transfer of objects. There can be any of the two possible reasons behind such parallel transfers. First, the client may generate multiple requests to the same server, each from a different TCP port to allow the parallel transfers. Second, when the embedded objects on a page are hosted by more than one servers, the client may generate multiple requests to the multiple servers in parallel to each other. Hence, multiple objects on the same page could be retrieved from multiple servers at the same time.

In both the above cases, multiple TCP connections could be observed on the network level, each carrying a different object at the same time. However, in the latter case, objects retrieved from the different servers are not considered as part of the same web session, as a session is based on the pair of client and server IP addresses. The change in source or destination IP addresses marks a different session.

In [4] authors presented two regions of OFF times, (1) 1 ms to 1 s for the active OFF times and, (2) 30 s to 3000 s for inactive OFF times. In [5], the median value of user think times is given as 15 s based on the silent time threshold between two documents requested from the client side. According to [7], most of the requests from the same user are launched with an inter-request time of less than 64 s. The appropriate value for the silent time threshold for most user sessions were shown to be between 100 s to 1000 s. It means that, on the average, a single user makes a sequence of requests to a particular server with the think times less than 100 s during a single visit. More than half of these requests were generated automatically by the client-side application since their think times were less than 1 s, which we call the active OFF times. User think times during a session were defined based on the inter-request times for the base files. Half of these inter-request times were above 8 s, with a large ratio of these times between 16 s and 64 s. More than 30% of the inter-requests times were less than 1 s and remaining 20% were between 1 and 8 s. The user sessions of the Youtube website were further characterized in another study [9]. There, the user think times were found to be increased to 30 s, possibly because of the video streaming. The users take some time watching the video before making the next request. In the same study, authors found out the value of active OFF times to be between 1 s to 30 s with only 14% of these values exceeding 1 s.

Summarizing the above observations from the literature, we find that the active OFF times are usually less than 1 s, while inactive OFF times vary quite a lot. We observe that their values are generally above 8 s. These observations are further listed in Table I. This draws a borderline that the gaps above 8 s in the traffic between a particular pair of the user and the server potentially shows the user think time between two transfers.

TABLE I
USER THINK TIMES

Literature	Active OFF times	Inactive OFF times	Silent time threshold	Additional information
[4]	< 1 s	30 – 3000 s	Part of Inactive OFF times	-
[5]	-	~ 15 s	-	-
[7]	< 1 s	> 8 s	100 – 1000 s	Football website
[9]	< 1 s	~ 30 s	> 1000 s	Youtube

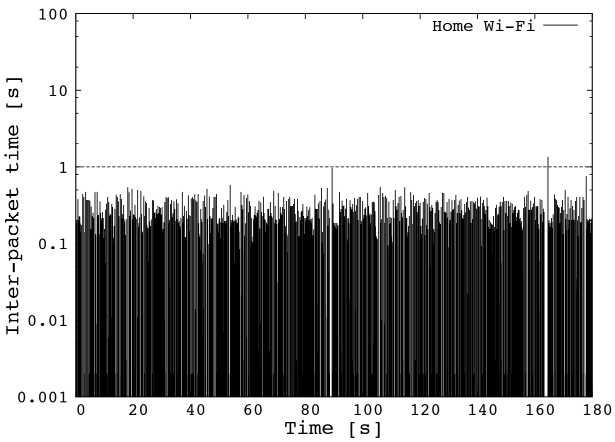

Fig. 3. Inter-packet times of video stream captured via Home Wi-Fi

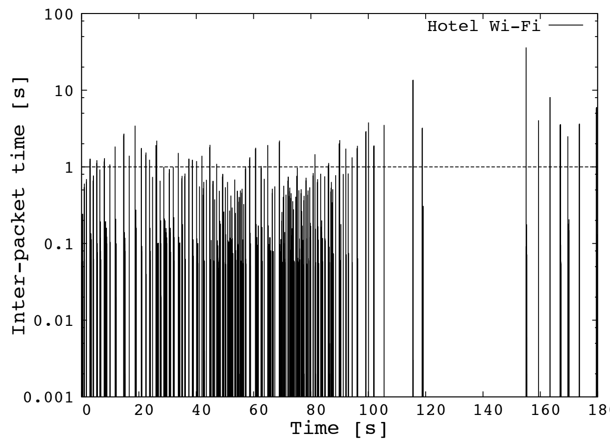

Fig. 4. Inter-packet times of video stream captured via Hotel Wi-Fi

V. NETWORK-INDUCED GAPS

In the previous section, we saw that the user think times are usually above 8 s, as a user normally takes at least this much time to view the content and clicking the button for the next request. In this section, we discuss the gaps that are induced by the network, which can be due to weak signals in the wireless networks, congestion problems or poor performance of network protocols etc, and are the potential candidates for destroying the user experience. For this purpose, we did some experiments of live video streaming in the Web browser via two different networks. Experiments were conducted in Wi-Fi environment at a home in Sweden and a hotel in Germany. The video streaming was quite smooth in the home environment while it was freezing quite frequently in the hotel environment.

Figure 3 shows the inter-packet times that we observed from the transfer performed in the home environment. Obviously, typical inter-packet times are well below 1 s with an average of 72 ms. This characterizes the gaps when the video streaming is working fine without freezes. We see that these inter-packet gaps rise and approach 1 s occasionally. However, such occasional increases in the inter-packet times did not hurt the video streaming and we didn't observe any freezes. This further shows that frequency of such inter-packet times around 1 s also matter. Once the buffer gets empty, then the freezes occur and hence affects the user experience [10].

Figure 4 illustrates the inter-packet times observed from the trace captured in the hotel environment. The difference is quite obvious. There are frequently occurring inter-packet gaps above 1 s and some of them may result in the freezes in video streaming depending on the buffer size, and hence delay the delivery of the transfer. Users are often very sensitive to the delays while watching the live video streaming, which is probably due the nature of content they are watching e.g. sport events or news. They lose a piece of information each time the video is frozen. Hence, such gaps above 1 s are very much threatening the user experience in the live video streaming context. Most of these inter-packet gaps were found between 1 s and 4 s, occurring more or less with the regular intervals of time during the first 100 s of video streaming. We did observe a considerable number of freezes in the video during this period. Finally, after 100 s of video streaming, there were even longer gaps above 4 s. The inter-packet gaps between 1 s and 4 s repeating in cycles are alarming and should be given special attention while monitoring.

Table II summarizes the results from the two transfers. Here, we use the term OFF times for the gaps (i.e. inter-packet times) above 1 s. Clearly, the video streamed on the hotel Wi-Fi illustrates the bad transfer full of OFF times. Most of these OFF times are less than 4 s. These frequently occurring OFF times between 1 s to 4 s characterize very well the existence of bad transfers and should be given special attention. Although, 1 s to 4 s OFF times during a transfer are alarming, it does not mean that the good transfers do not consist of such gaps at all. Therefore, the frequency of such OFF times is also a major point of consideration before declaring a transfer "good"

TABLE II
NETWORK-INDUCED GAPS

Network	Total duration	Total OFF times	1 s – 2 s	2 s – 3 s	3 s – 4 s	Above 4 s
Home Wi-Fi	180 s	1	1	0	0	0
Hotel Wi-Fi	180 s	37	24	4	5	4

or "bad". So we also need to keep track of the duration of the ON times between two subsequent OFF times. For this reason, we present in Figure 5 the CCDF of the durations of OFF times, ON times and ON+OFF times from the bad transfer. ON+OFF times represent the cycle and illustrate the frequency of the occurrence of gaps.

Obviously, the OFF times are greater than the ON times. It shows that the time the user can really enjoy the video before the video freezes is less than the duration of freeze time. This is itself a clear indication of the bad transfer. The frequency of freezes sinks as their duration increases. We also observe freezes that are in the range of think times. A buffer of some seconds takes most of the problems away, given the traffic catches up again after the freeze time for which we see evidence in [11]. The curve of ON+OFF times in the figure shows the cycle time i.e. how frequently one ON and OFF times phase finishes. In Figure 5, 80% of the ON+OFF times are still less than 5 s, showing only a slight difference as compared to the OFF times. This further illustrates the frequency of gaps which is high in this case. Here, we have not presented the distribution plot of ON times and OFF times for the transfer done via home Wi-Fi. The reason is that, there is only one OFF time of duration above 1 s during the whole transfer, that follows an ON time of around 170 s. Actually, occasional OFF times of slightly above 1 s after considerably long ON times characterize a good quality transfer, assuming that the buffer size is long enough to keep up with such occasional short outages. Hence, the relative difference between the OFF times and the ON times is also one of the major factors in declaring the quality of a transfer. In the next section, we will further show how this relative difference could be visualized and quantified on the different timescales.

Further, for validation purposes, the direction of the transfer could also be considered. Usually, the user think times are followed by the request in upward direction i.e. from the user to the server. Conversely, the outage in the video streaming download is followed by the data packet (containing a request) in the downward direction. Therefore, if a gap in the traffic is followed by the packet with payload from user side then the gap could be considered as the user think time, while if it is followed by the packet from server side, then the gap was potentially a result of an outage.

VI. WAVELET ANALYSIS

In order to visualize the quality problems at different scales, we have performed the Haar wavelet analysis of both the traffic transfers. It allows us to identify the time scale at which the problem occurs. We can view both the time and the frequency components together, for instance, how long the gaps are and how frequently they occur. Hence, the wavelet analysis performs the localization allow us to locate those time instances where the problems occur. The wavelet analysis also allows us to test the trend and the burstiness of a transfer on the fly, passively from the measurements on the link-level. To check the burstiness at different time-scales, the d coefficients are used; and to view the trend (like moving average) of a transfer, the c coefficients are utilized. The d coefficients are henceforth called the wavelet coefficients and the c coefficients are called the scaling coefficients.

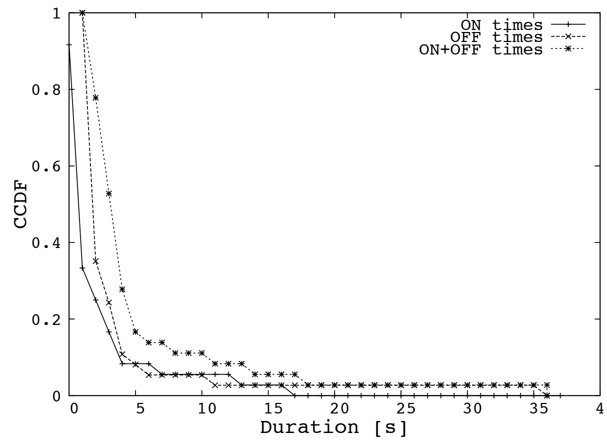

Fig. 5. CCDF distribution of ON and OFF times for the transfer done via hotel Wi-Fi

A. Calculation of Wavelet and Scaling coefficients

The d coefficients extract the detail in the time series (traffic trace) at different scales and different locations. In other words, the d coefficients display the degree of difference between the data points at different locations in a time series.

Let y be the vector that represents data points in a time series: $y = \{y_1, y_2, y_3, y_4, ..., y_n\}$. Let n be the length of the vector y, which must be a power of 2 such that, $n = 2^J$. Thus, on the finest scale $J - 1$, the wavelet coefficients d between the two successive points can be calculated as:

$$d_{j,k} = y_{2k} - y_{2k-1}, \tag{1}$$

where $k = 1, 2, 3, ..., n/2$ and $j = J - 1$.

Let's assume that there are $n = 8$ data points in the vector y, then on the finest scale $J - 1 = 2$, there will be four wavelet coefficients: $d_{2,1} = y_2 - y_1$, $d_{2,2} = y_4 - y_3$, $d_{2,3} = y_6 - y_5$

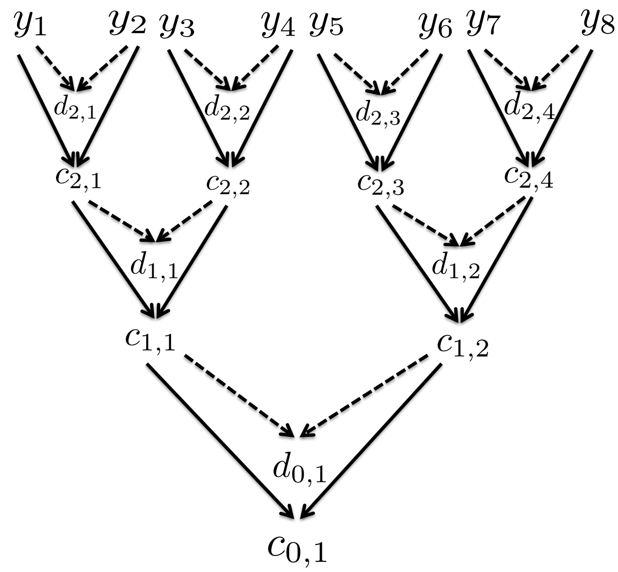

Fig. 6. Wavelet and scaling coefficients

and $d_{2,4} = y_8 - y_7$. Hence, the values of wavelet coefficients demonstrate the variation between the immediate neighbors at a particular scale.

Moreover, a smoothing operation can be performed on the time series by obtaining c coefficients. These coefficients give us information about a time series on the coarser scale. The operation of scaling coefficients is similar to the moving average smoothing operation. Thus, the c coefficients at the finest scale $J-1$ can be calculated as:

$$c_{j,k} = y_{2k} + y_{2k-1}, \tag{2}$$

where $k = 1, 2, 3, ..., n/2$ and $j = J - 1$.

In order to obtain the detail coefficients d at the coarser levels $J-2, J-3, ...0$, the differencing between the two non-overlapping consecutive pairs of c_k is performed at each level, as mentioned in the Figure 6. Hence, the y_{2k} and y_{2k-1} in the equation will be replaced by c_{2k} and c_{2k-1}, respectively:

$$d_{j,k} = c_{2k} - c_{2k-1} \tag{3}$$

Similarly, the c coefficients at coarser scales can be calculated as:

$$c_{j,k} = c_{2k} + c_{2k-1}, \tag{4}$$

where $k = 1, 2, 3, ..., n/2$ and $j = J - 2, ...0$.

Furthermore, for the purpose of normalization, all the obtained c and d coefficients are divided by $\sqrt{2}$ before using them in the spectrum analysis.

We performed the wavelet analysis on the throughput of both the web transfers (collected at home and hotel) mentioned in the previous section. The throughput was calculated as the number of packets received at the client side during each time window. The time window was set to 125 ms. Based on the obtained time series of throughput, we calculated the c and d coefficients from the finest to the coarsest scales. Finally, we computed the power spectrum of c and d coefficients at each scale from j to 0 to observe the scaling behavior. Investigating different series of coefficients allowed us to pinpoint those locations in the transfer, where the change in the perceived performance occurred. The Equations 5 and 6 were used to calculate the power spectrum for c and d coefficients, respectively:

$$\mu_j = \frac{1}{n_j} \sum_{k=1}^{n_j} c(j,k)^2 \tag{5}$$

$$\mu_j = \frac{1}{n_j} \sum_{k=1}^{n_j} d(j,k)^2 \tag{6}$$

where n_j is the number of coefficients at band j. Furthermore, the log_2 of each power spectrum (μ_j) is calculated, which is then plotted against the respective band (j), as depicted in the Figures 7 and 8.

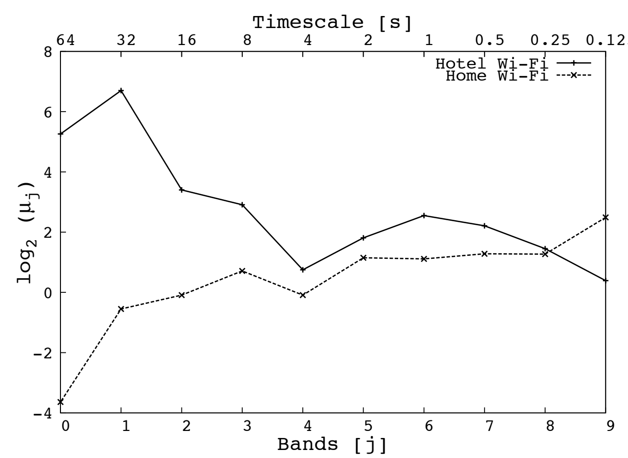

Fig. 7. Spectrum of wavelet coefficients

To perform the wavelet analysis, we considered the throughput of the first 128 seconds of each transfer. As we calculated throughput in the time windows of 125 ms, therefore, the number of data points (n) becomes 1024, yielding scales $J = 10$ (0 to 9). The finest scale is $J - 1 = 9$ (at 0.125 s) and the coarsest scale is 0 (at 64 s).

In Figure 7, we present the plots of the spectrum of the d coefficients from band 0 to 9. Each wavelet spectrum plot illustrates very well the different properties of a transfer and can thus be used to unveil the perceived performance at different timescales. For the hotel transfer, we observe three different scaling behaviors. We divide these three scaling behaviors into three different regimes within one transfer, i.e. less than 1 s (bands above 6), between 1 s to 4 s (bands 4 to 6), and above 4 s (bands below 4). The scaling behavior between 1 s and 4 s is of particular importance, as it characterizes the frequent OFF periods at a scale of 1 s to 4 s. Hence, it results in increased

TABLE III
LINEAR REGRESSIONS BETWEEN BANDS AND THE SPECTRUM VALUES OF WAVELET COEFFICIENTS

Access network	Band 0 to 4	Band 4 to 6	Band 6 to 9
Hotel Wi-Fi	$\log_2(\mu_j)= -1.83j + 8.02$, r = -0.96	$\log_2(\mu_j) = 0.9j - 2.80$, r = 0.99	$\log_2(\mu_j) = -0.72j + 7.08$, r = -0.98
Home Wi-Fi	$\log_2(\mu_j) = 0.22j - 0.55$, r = 0.54	$\log_2(\mu_j) = 0.6j - 2.28$, r = 0.85	$\log_2(\mu_j) = 0.41j - 1.56$, r = 0.83

burstiness while going from band 4 to 6 (timescales: 4 s to 1 s). Since such scaling is absent on the timescales less than 1 s, it indicates that the frequent traffic gaps of around 1 s exist in the network traffic. However, at lower bands (higher timescales), we observe high burstiness indicating a difference in the quality at the different locations in the transfer. For instance, at band 1 (32 s timescale), we observe highest spectrum values, indicating the shift in the quality of the transfer every 32 s. In contrast, we observe a relatively stable behavior in the home transfer where the spectrum values increase with the decreasing timescales. The wavelet spectrum analysis of home transfer indicates usual network traffic behavior without many outages, which implies that the quality of transfer at higher timescales appear smoother, while it appears bursty as we go towards lower time scale, i.e. higher bands.

The above-mentioned behavior is illustrated very well by the linear regressions (fitted on the spectrum data of wavelet coefficients) mentioned in the Table III. For the hotel Wi-Fi network, three regimes are clearly visible. The linear regression of band 4 to 6 (timescales: 4 s to 1 s) shows strong positive correlation proving the existence of scaling on these scales. However, the other two regimes show a negative correlation. It indicates that much of the variation in the hotel transfer is present on the 1 s to 4 s scale. Conversely, from the linear regression, we observe less burstiness in the home transfer on the higher timescales, indicating a rather smooth transfer with much of the scaling at very small timescales. However, the existence of scaling at shorter timescales is the sign of activity (higher ON times).

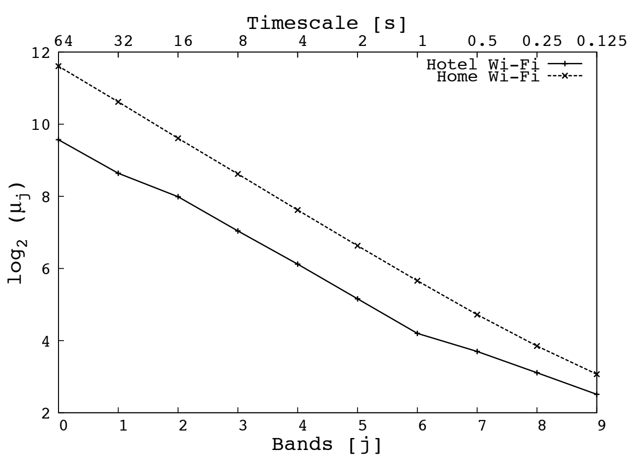

Fig. 8. Spectrum of scaling coefficients

Figure 8 displays the power spectrum of the c coefficients for the home and the hotel transfers. The trend is very clear that the transfer done at home network gives higher values across different scales as compared to the transfer done at the hotel. At higher bands, i.e. shorter time scales, many of the time windows are empty which leads to a slow decay. In contrast, stability of the home transfer is evident from the consistent decay in the spectrum of scaling coefficients. Table IV list the linear regressions for the hotel and the home transfers. There is a strong negative correlation for both transfers. However, the value of α is indicating a faster decay in the case of the home transfer.

TABLE IV
LINEAR REGRESSIONS BETWEEN BANDS AND THE SPECTRUM VALUES OF SCALING COEFFICIENTS

Access Network	Linear regression
Hotel Wi-Fi	$\log_2(\mu_j) = -0.796j + 9.36$, r = -0.99
Home Wi-Fi	$\log_2(\mu_j) = -0.954j + 11.48$, r = -0.99

VII. CRITERIA FOR ALARMING GAPS

In this section, we will propose a criterion to monitor those outages or traffic gaps, which can be helpful in prompting the service providers to take proactive actions for improving the QoE. As mentioned in the previous sections, there could be multiple reasons behind the occurrence of traffic gaps during a web session between a client and a server such as:

- The user think times.
- The server could be heavily loaded and may result in the bursty traffic.
- The client-side web browser causing the active OFF times.
- The problems in the network such as the signal problems in wireless networks, scheduling on the base stations, congestion in the network or the dynamics of the network protocols.

To confirm if a traffic gap was from the server side and not due the user inactivity, the direction of the data packet after the traffic gap needs to be observed. If the next packet containing the data is from the server side and not from the client side, then the last gap was not due to the user inactivity but due to the network behavior. However, it requires packet inspection to detect the direction.

All the above causes produce scaling in the traffic that might be different on different timescales. In order to identify the traffic gaps induced by a badly behaved network, it is important to observe the duration and the frequency of the gaps. The wavelet analysis of traffic generated in a session is an important tool to visualize on the fly both the duration and the frequency of traffic gaps. We propose the following step-by-step procedure to identify the network-induced traffic gaps:

1) A spectrum analysis of wavelet and scaling coefficients should be performed, and change point separating multiple scaling behaviors in the spectrum plot of wavelet coefficients should be identified.
2) If the time scales between 1 s and 4 s show a different scaling of the wavelet coefficients than their neighboring timescales (above 4 s and below 1 s) – for example if the corresponding slope changes sign – one can deduce that the traffic gaps of 1 s to 4 s are recurring frequently.
3) The scaling behavior on the long (>4 s) and shorter (<1 s) timescales suggest the shift in the quality of transfer at different times and the amount of variation in the traffic during the ON times, respectively. The negative slope is the sign of increasing inactivity.

VIII. Conclusions and future work

This paper proposed a simple wavelet-based criterion that can be useful for the service providers to monitor the user transfers. The criterion is fast as it does not require any deep packet header information and hence enables the service providers to take immediate appropriate measures based on the pure observation of the flow of data associated to the stream.

In this paper, we outlined the difference between the duration of traffic gaps generated due to the user think times and the network outages during a transfer. We found that the network outages that result in freezes in video transfers on web often constitute of duration between 1 s to 4 s, while user think times are usually above 8 s. This implies that, a gap above 8 s after a smooth transfer is likely to characterize the user think time. Therefore, such gaps can be ignored by the service providers. Conversely, the gaps of duration between 1 s to 4 s, occurring with frequent intervals are a sign of poor quality transfer, as shown by the longer durations of OFF times as compared to the ON times and small ON+OFF time durations.

All these properties at different timescales could be visualized with the help of wavelet spectrum analysis. The presence of scaling at timescales below 1 s indicates the ON time and hence, the signs of activity. However, the presence of scaling at timescales between 1 s – 4 s, and the absence of scaling at shorter and longer time scales characterize the frequent OFF times with shorter ON times.

This is an ongoing work. Our short-term future work includes the investigation of traffic gaps with the other types of web traffic along with the investigation of wavelet spectra as a function of time. We further intend to validate this wavelet-based criterion with the help of experiments with real users on our test-bed, in order to differentiate between the user think times and the network-induced traffic gaps.

Acknowledgment

This work had been partly supported by the FP7 Network of Excellence Euro-NF and the CELTIC project QuEEN.

References

[1] S. Rugel, T. M. Knoll, M Eckert, T Bauschert *A Network-based Method for Measurement of Internet Video Streaming Quality*, The First European Teletraffic Seminar, Poland. Feb 2011.

[2] http://www.wireshark.org. Last visited: March 5, 2012.

[3] B. Krishnamurthy, C. Wills, *Cat and Mouse: Content Delivery Tradeoffs in Web Access*, In the proceedings of WWW 2006, Edinburgh, Scotland, May 2006.

[4] Mark E. Crovella, A. Bestavros. *Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes*, IEEE/ACM Transactions on Networking, Volume 5, Issue 6. Dec 1997.

[5] Bruce A. Mah. *An Empirical Model of HTTP Network Traffic*, In Proceedings of IEEE INFOCOM, Kobe, Japan. Apr 1997.

[6] P. Barford, M. Crovella. *Generating Representative Web Workloads for Network and Server Performance Evaluation*, In Proceedings of ACM SIGMETRICS. New York, USA. June 1998.

[7] M. Arlitt. *Characterizing Web User Sessions*, ACM SIGMETRICS Performance Evaluation Review. Volume 28, Issue 2. Sep 2000.

[8] J. Shaikh, M. Fiedler, D. Collange. *Quality of Experience from User and Network Perspectives*, Annals of Telecommunications. Volume 65, Number 1-2. Feb 2010.

[9] P. Gill, M. Arlitt, Z. Li, A. Mahanti. *Characterizing User Sessions on Youtube*, In Proceedings of ACM/SPIE MMCN. San Jose, USA. Jan 2007.

[10] S. Ickin, K. De Vogeleer, M. Fiedler, D. Erman. *The effects of Packet Delay Variation on the Perceptual Quality of Video*, In Proceedings of ON-MOVE. Denver, USA. Oct 2007.

[11] K. Wac, M. Fiedler, R. Bults, H. Hermans. *Estimations of Additional Delays for Mobile Applications Data from Comparative Output-Input Throughput Analysis*, In Proceedings of NOMS 2010. Osaka, Japan. Apr 2010.