# A Study of Email Usage and Performance Over Cellular Technology

Aymen Hafsaoui
EURECOM
France
Hafsaoui@eurecom.fr

Guillaume Urvoy-Keller
EURECOM
France
Urvoy@eurecom.fr

Denis Collange
Orange Labs
France
Denis.collange@orange-ftgroup.com

Matti Siekkinen
Aalto University School of Science
and Technology, Finland
Siekkine@cc.hut.fi

***Abstract*—In this paper we investigate the performance of Cellular networks, focusing on the key mail and webmail services. Specifically, the primary motivation of this research is to compare mail and webmail performance and to discuss different factors that affect performance as perceived by end users. We discuss key factors related to the network configurations, e.g., proxies, that can bias the computation of performance metrics, e.g., RTT. For Webmail, we are able to further characterize webmail servers and client devices including the operating systems (OS) of the devices. In our trace, Hotmail is the most popular service and represents 23% of all webmail connections. As for users' device, the Iphone of Apple drives the market and thus appears by far the most commonly used end user device. While mail generates more TCP connections than webmail, it represents less bytes. We show that webmail achieves higher throughputs than legacy mail. We investigate this difference of throughputs and uncover several key factors that explain this difference. In particular, we demonstrate that while losses can have a detrimental impact for both mail and webmail, it impacts more the former. Also, the mail application (mail client and server) significantly impacts the way data is exchanged and thus the throughput.**

***Index Terms*—Mail, webmail, passive measurement, Cellular.**

## I. Introduction

Electronic mail (e-mail) is one of the most used Internet services, both in terms of popularity and amount of traffic generated. There are two services to check or send e-mail: mail (based on POP, IMAP, SMTP, etc.) and webmail services.

In this work, we present observations from a passive packet level trace with more than 1.7M TCP connections, collected at the access network of a major European ISP. Our study includes different classes of access: 3G, EDGE and 2.5G connections. A given user can be observed using any of these technologies as Cellular contracts work in a best effort manner: the client is granted a 3G access whenever it is available at the base station to which it is connected; or downgraded to former technologies, EDGE or 2.5G, if 3G is not available. We further observe a diversity related to users devices, e.g., mobile phones and USB pluggable 3G modems.

The main aim of this paper is to study the performance of Cellular access network and to bring to light phenomena introduced by Cellular core[1] network equipments, which can bias measurements. For instance, we noticed that the use of a active devices in Cellular Network affects end-to-end RTT estimation. In this initial exploration, we cast a first look at mail and webmail traffic performance, in terms of throughput and identify several factors that affect this metric.

We noticed that mail and webmail performance is affected mainly by the way application delivers data to TCP and also by losses, whose occurrence tends to be higher than on wired technologies [2]. Through a careful examination of webmail traffic, e.g., by mining HTTP exchanges, we present statistics about webmail servers, browsers, end users devices and operating systems used by Cellular clients.

The rest of this paper is structured as follows. We provide an overview of related works in Section 2. In Section 3, we present the main characteristics of the trace and we explore the time stability of our trace, a prerequisite before deriving additional performance metrics, e.g., throughput distribution. Section 4 reports on the impact of core network equipments in Cellular networks. In Section 5, we focus on the performance of mail and webmail. Finally, Section 6 concludes the paper.

## II. Related Works

A number of previous studies have examined the performance of modern Cellular networks.

In a recent work [1], the authors cast a first look at mobile hand-held device (MHD) usage by profiling the services users are interested in when they are at home, connected with their MHD to their residential DSL access thanks to a Wifi connection. They base their study on anonymized packet level data representing more than 20,000 residential DSL customers, spanning over a period of 11 months. They find that IPhones and IPods are by far the most commonly observed MHDs. This observation has an impact on the most popular mobile applications, which turn out to be Safari, ITunes, and Weather. Also, they observed that the largest fraction of volume of MHD HTTP content is multimedia.

In [2], the authors consider the use of the RTTs as a possible signal for detecting network anomalies. They present large-scale measurements from an operational 3G network and investigate the stability of RTT distribution. Their results

[1]Core relates here to the wired part of the ISP network that enables access of 2G/3G clients to the Internet.

[2]Though it does not constitute a complete proof per se, we observed that loss rate of mail traffic over Cellular access was consistently higher (for the traces with analyzed) than for Fiber-To-The-Home (FTTH) and ADSL accesses.

978-1-4244-8840-7/10/26.002010$IEEE$978 − 1 − 4244 − 8840 − 7/10/26.00 2010 IEEE

confirm that modern 3G networks yield considerably lower RTT values than initial GPRS deployments. However, they observed that in the network they monitor, RTTs over UMTS/HSxPA do not vary during the day, which means that they are poorly correlated with the network load. This limits the usage of RTTs to detect anomalies in the network they consider.

A number of works have focused on mail and webmail performance using measurements. In [3] a technique is proposed to detect webmail traffic from regular HTTP traffic by matching unique webmail keywords in the HTTP payload. Authors of this study found that parsing the first 36 bytes of user payload is sufficient to detect webmail traffic. From a performance viewpoint, they observed that for popular webmail services, the transfer of embedded objects such as advertising banners affects performance.

In [4], the author focused on the performance of SMTP, POP3 and IMAP connections observed on the Internet. Two traces collected on residential sites in Germany (dial-up and DSL) and a last trace from made available by NLANR, have been evaluated thanks to an application level analysis developed by the author. Several key factors for mail performance are considered: DNS lookup latency, TCP connection set-up latency, SMTP command exchange latency and mail upload latency. The author demonstrates that a significant share of the latency both for sending and receiving e-mails is due to serial processing of commands, which can hardly be reduced by increasing bandwidths. In this work, the longest exchanges were observed with IMAP, due to a different mode of operation compared to POP3, as an IMAP client can maintain an open connection to the IMAP server, unlike POP3 clients that establish a new connection each time.

To resolve the problems identified in [4], the authors in [5] investigate the sources of latency in the POP3 protocol and propose enhancements to the POP3 standard. Furthermore, they observed that neither the increase in bandwidth, nor the decrease of the e-mail server response time have the potential to significantly reduce the response time of the POP3 protocol. To improve POP3 standard performance, they propose a new client/server architecture to enhance the e-mail server performance and to reduce the latency in retrieval time.

## III. Data Set

### A. Overview of the Traffic Trace

Our study is based on the analysis of a passive bidirectional Cellular trace collected at an aggregation point within the network of a large European ISP. Table I summarizes the main characteristics of the trace.

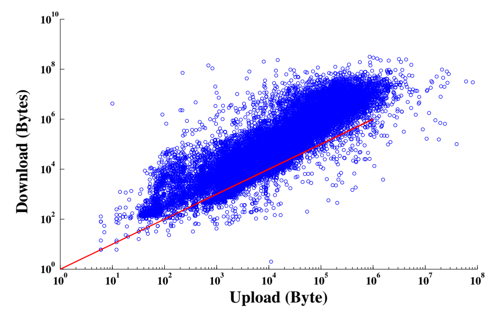
In the present work, our focus is on applications on top of TCP. While analyzing their performance, we restrict our attention to the connections that correspond to presumably valid and complete transfers. We term such a transfer a well-behaved connection. Well-behaved connections must fulfil the following conditions: (i) A complete three-way handshake; (ii) At least one TCP data segment in each direction; (iii) The connection must finish correctly either with a FIN or RESET flag. In our trace, about 70% of the transfers are well-behaved transfers – see Table I.

| Type | Cellular |
|---|---|
| Location | France |
| Date | 2008-11-22 |
| Starting Capture | 13:08:27.38 |
| Duration | 01:39:01.068 |
| NB Connections | 1772683 |
| Well Behaved Connections | 1236253 |
| Volume UP(GB) | 11.2 |
| Volume DOWN(GB) | 50.6 |

TABLE I
TRACE DESCRIPTION

### B. Applications

Concerning traffic breakdown, we observed that, in contrast to typical wired accesses (e.g., FTTH [6] or ADSL [7]) the majority of the transfers target ports 80, 8080 and 443, which strongly suggests that most of the applications used in Cellular networks, flow over HTTP. We would need more sophisticated techniques to fully profile the applications active in our trace [7], [8]. However, it turns out that we got enough side information to accurately identify mail traffic – see Section V-A.




(a) Bytes

Fig. 1. Uplad and Download Traffic per Client

We next turn our attention to per client usage. Figure 1 shows the scatter plot of the traffic volume uploaded and download per user. We observe that Cellular clients tend to download significantly more data than they upload. This is in contrast to wired networks usage profiles where one observes that a significant fraction of users upload large volumes of traffic because of p2p applications [7].

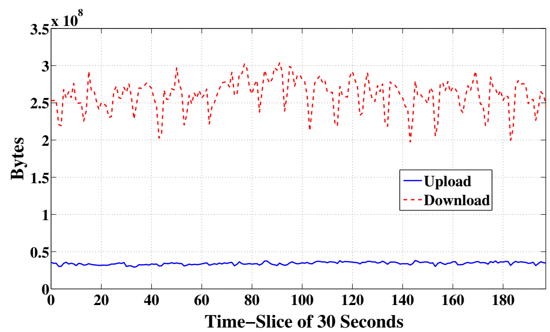
This result is also in line with the findings in [1] where the authors observe that the largest fraction in term of volume of MHD HTTP content is multimedia: watching video from Youtube, listening music from Itunes or downloading applications form Apple Store or Android Market, induce more download than upload traffic .

978-1-4244-8840-7/10/26.002010$IEEE$978 − 1 − 4244 − 8840 − 7/10/26.00 2010 IEEE

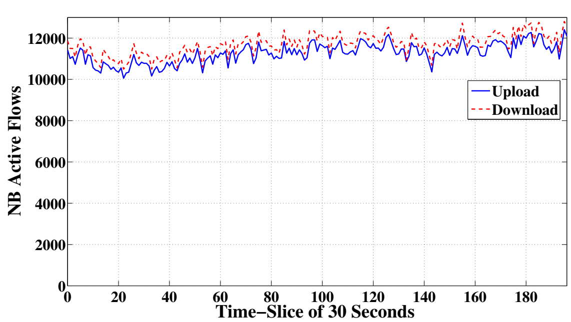
### C. Traffic Stability

In this paragraph, we assess the stability of the traffic within the trace we have collected. To do so, we observe the time series of traffic volume and of the number of active flows. The objective is to assess if several regimes exist in our data, which would then require to analyze performance within each corresponding time interval. As we will see, it is apparently not the case with our trace. This justifies our approach in the next sections where we will look at marginal distributions of throughputs and other metrics where all samples of the trace are grouped to form those distributions.

Figure 2 shows the evolution of traffic volume and the number of active flows for the upload and the download directions. To obtain those figures, we broke up our trace into short time windows of 30 seconds and we computed the number of active flows and the exchanged data volume for each direction in each window. Note that a flow is considered active for a given time slice if it transmit at least one data packet during the slice.




(a) Bytes




(b) Number of Active Flows

Fig. 2. Upload and Download Data Volume: Windows of 30 Seconds

Figure 2(a) shows that traffic is qualitatively more bursty in the download than the uplad direction. This is presumably because bursts are shaped by the limited uplink capacity implemented by the Cellular operator. Another immediate observation from Figure 2(a) is that the ratio between uploaded and downloaded volumes is fairly constant, close to 6. This trend can be explained by the current usage of Cellular network and also by the limited capacity of uplinks as compared to downlinks in current Cellular technology.

Concerning active flows, Figure 2(b) demonstrate that they do not vary drastically over time, further reinforcing the idea that traffic is stable over the time span of our trace.

## IV. Impact of Core Network Equipment

In this section, we highlight that in modern Cellular networks, computing latency turns out to be a complex task. Indeed, we demonstrate that latency can be under estimated due to the use of new mechanisms or services, like proxies for content adaptation or applications acceleration. We investigate how these mechanisms impact our measurements and the performance perceived by end users.

### A. RTT Estimation

Several approaches have been proposed to accurately estimate the RTT from a single measurement point [9], [10], [11], [12], [13]. To estimate RTT, we adopted two techniques. The first method is based on the observation of the TCP 3-way handshake [10]: one first computes the time interval between the SYN and the SYN-ACK segment, and adds to the latter the time interval between the SYN-ACK and its corresponding ACK. It is important to note that we take losses into account in our analysis. The second method is similar but applied to TCP data and acknowledgement segments transfered in each direction [3]. One then takes the minimum over all samples as an estimate of the RTT.

Due to the location of the probe within the network of the ISP, we are able to distinguish between a local and a remote RTT. The local RTT is measured within the access network, including the wireless link, of the ISP, while the remote RTT factors both the latency over the path from inside the network ISP to the first peering link and then to the remote server.

### B. Impact of Active Devices

While analyzing modern Cellular networks, we face a double difficulty: (i) the access technology can vary (from 2G to 3G) from one user to the other and over times and (ii) the capabilities of the device itself varies from one device to the other, which sometimes prevents the user from accessing all types of internet applications. In the network we analyze, devices with limited display capability are serviced by a specific device. Redirection to this specific device is achieved at the mobile client using Access Point Name (APN). It can be seen as the equivalent of a dial-up phone number of an ISP. For convenience, we term those connections APN transfers below. In our trace, we have more than 17% of APN transfers, which can be identified as targeting a specific private IP address and port 8080.

We compared the RTT of APN and non APN transfers. For the latter ones, we restrict our attention to connections targeting port 8080, to have a somewhat comparable basis (though it is still quite arbitrary). In Figure 3, we compare remote RTTs – the two RTT estimation methods gave similar results – for the two types of transfers. We notice a difference

[3] Keep in mind that we focus on well-behaved transfers for which there is at least one data packet in each direction.

of about 100 ms between the two types of traffic, which is explained by the split mode used at the device, which adapts the content for those limited capacity devices.

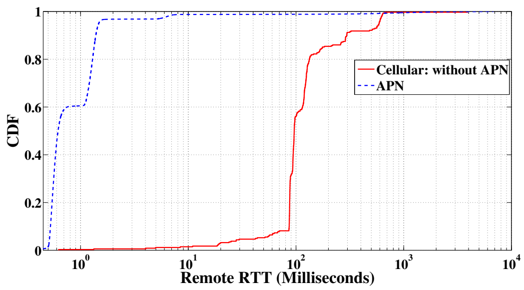



Fig. 3. Remote RTT of APN Transfers

We now restrict our attention to non APN transfers. These transfers are characterized in our trace in a straightforward manner: they have a public remote IP address. Still, the devices that generate this type of traffic do not necessarily communicate directly with the remote server. The ISP is using a set of devices for user authentication (Radius), Network Address Translation (NAT) (as we find also in wired networks), but also, and this is a specificity of Cellular networks, a proxy whose main objective is to boost performance of the initial phases of TCP transfers. This proxy intercepts the first SYN of new connections and responds on behalf of the remote server, with a SYN-ACK, while in parallel, the initial SYN is forwarded to the remote server. The proxy later applies various tricks to (try to) improve the performance of TCP transfers. The way the proxy works at connection establishment leads to a significant discrepancy between the two methods we use to compute the RTT, which often reaches 100 ms, as can be seen from Figure 4.

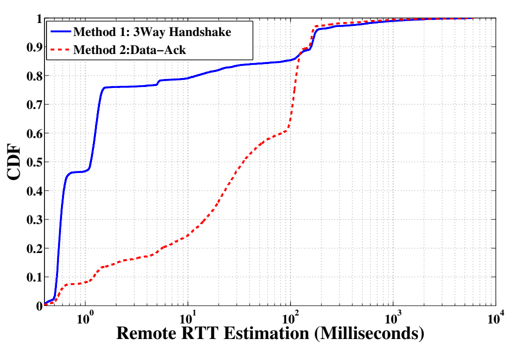



Fig. 4. Proxy Impact for Latency Estimation

The key message from this section is that several specific devices might affect classical performance metrics in Cellular networks, which should be taken into account when performing measurement studies. In the rest of this paper, we focus only on non APN traffic and our estimation for latency will be based on the DATA-ACK method only.

## V. Mail and Webmail: Characteristics and Usage

The E-mail service is often overlooked in traffic analysis studies, even though it represents a key service for end users that use it on a daily basis.

In this section, we detail how mail and webmail traffic is extracted from the trace and evaluate the popularity of mail and webmail usage. We further extract information related to the popularity of the different service providers and end user device for the case of webmail, taking advantage of the fact that the HTTP protocol exposes some key information.

### A. Service Identification

Internet traffic classification is an area that attracted a lot of attention recently [14],[15],[16]. In most cases, mail traffic is classified based on the legacy mail protocols: IMAP, POP3 and SMTP. Concerning webmail, the following identification techniques have been proposed: 1) Map the destination IP address with a list of URLs of popular webmail providers [17] 2) Combine the previous method of URL matching with keyword matching (based on unique keywords that appear in the packets payload that can identify webmail traffic) [3] 3) Use statistical methods [18].

In this paper, we adopted the second approach to detect webmail traffic: we first extract HTTP requests with webmail key words, then we identified the connections corresponding to these requests. To identify mail traffic for the upload and download, we use TCP port numbers and remote address resolution.

### B. Usage and Popularity

Using the detection method presented in the previous paragraph, we extracted mail and webmail traffic from our trace. It turns out that mail and webmail represent about 5% of all flows and 17% of overall traffic volume.

Tables II and III summarize characteristics of mail and webmail connections, including number of connections, volumes uploaded and downloaded in terms of total amount of bytes at the IP layer and in terms of data packets, number of servers, and number of clients.

Concerning mail (see Table II), we observe that POP3/POPS dominates downloads while SMTP/SMTPS (obviously) dominates uploads. IMAP/IMAPS is the most popular service in terms of number of established TCP connections, followed by POP3/POPS and finally SMTP/SMTPS. The smaller number of mails downloaded as compared to mails uploaded is likely to be due to the limited capabilities of MHD (most of them are smart phones and not PC with USB pluggable 3G modems as we will see soon) as compared to legacy wired access with desktops and laptops, which feature convenient displays and also store data that can be used as attachments, as opposed to smart phones in general[4].

[4]Our experience with wired traces of DSL and FTTH accesses shows that email traffic is also asymmetric in wired accesses, e.g., because of mailing lists and wanted or unwanted advertisements; but the extent of asymmetry is far smaller than in Cellular networks.

| | SMTP/SMTPS | POP3/POPS | IMAP/IMAPS |
|---|---|---|---|
| Nb cnxs | 7330 | 51202 | 64493 |
| Upload (MB) | 116.1 | 5.7 | 59.8 |
| Download (MB) | 1.2 | 1741.6 | 853.8 |
| Upload (Data Pkts) | 78631 | 261828 | 731616 |
| Download (Data Pkts) | 10130 | 1523961 | 1270844 |
| Nb Servers | 360 | 1578 | 883 |

TABLE II
MAIL TRAFFIC

Table III shows general information about webmail traffic in our trace. We observe a similar results as for mail: users to download more than they upload. We hypothesize again that it is a result of the limited capacities of MHD in general.

| | Cellular |
|---|---|
| Nb Cnxs | 16275 |
| Upload (MB) | 1364.4 |
| Download (MB) | 7169.8 |
| Upload (Data Pkts) | 1712270 |
| Download (Data Pkts) | 2022705 |
| Nb Servers | 528 |

TABLE III
WEBMAIL TRAFFIC

Comparing mail and webmail volume statistics, we observe that webmail is much more popular than classical mail. Concerning connection sizes, a number of webmail connections are smaller in size as compared to mail transfers.

At this stage a natural question is: Why Cellular users tend to use more webmail than classical mail? We can envisage several options: 1) Cellular users prefer webmail at the expense of legacy mail 2) Webmail services offer in general better performance than mail 3) Cellular devices are more adapted to webmail usage.

Option 1 stems from two intuitions. First, the natural intuition that the complexity of configuring a POP/IMAP client as compared to using a Webmail access is a barrier for a lot of users. Second, the intuition that users prefer to have a mail account from a mail service provider that is not their network provider in order to keep the account even if the network provider changes. Though mail service providers offer in general POP/IMAP interfaces, Web based interfaces, i.e., webmail, is by far the most popular way to reach those services.

We gathered also statistics on webmail servers, client devices and their OSs, and clients browsers, taking advantage of the presence of many key information in the HTTP fields. Figure 5 reports the percentage of transfers per webmail service providers (for the most popular ones). We observe a dominance of Hotmail, Gmail and Yahoo. Only after we find webmail services offered by network providers like Orange, Tele2 and Alice. These results show that webmail service providers that propose free mail boxes are much more popular than the corresponding services offered by network providers. The latter means that hypothesis 1) mentioned above plays a role in the higher popularity of webmail at the expense of traditional mail.

Let us now focus on MHDs and their OSs. Figure 6 shows that among the currently popular MHDs and OSs, we find iPhone at the first position followed by Microsoft OSs (Vista, XP and CE). MacOS, Linux and other MHD devices remain marginal in our data set. The above result was obtained for clients using webmail and not for clients using mail, as we have no access to similar information in the latter case. We can however conjecture that the trends (OS shares) for these other clients be similar. More generally, the above observation is in line with current market trends that shows that, at least in France, the Iphone is the dominating smart phone at the moment. The small fraction of OSes of laptops suggests that devices connected with USB pluggable 3G modems are still marginal in the Cellular network we study.

| OS/Device | NB Cnxs | NB Sessions | NB Cnxs/Session |
|---|---|---|---|
| Iphone | 9780 | 2562 | 3.81 |
| Windows Vista | 1283 | 304 | 4.22 |
| Windows XP | 1170 | 376 | 3.11 |
| Windows CE | 290 | 140 | 2.07 |
| Macintoch | 169 | 55 | 3.07 |
| Symbian | 138 | 50 | 2.76 |
| Linux | 22 | 12 | 1.83 |
| Others | 326 | 126 | 2.58 |

TABLE IV
WEBMAIL CONNECTIONS AND SESSIONS

Operating systems and Web browser can impact network performance through several parameters and especially the number of connections established to the Web server they connect to. To assess if there was significant difference in the strategy used by the different OS/device in our data set, we report in Table IV the mean numbers of connections per webmail session. A session consists of all the connections between a specific pair of client and server IP addresses in our trace. The results in Table IV suggest a similar behavior for the dominant OSes/devices we observe in our trace.

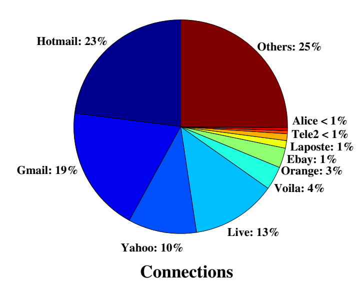



Fig. 5. Webmail Service Provider

978-1-4244-8840-7/10/26.002010 $IEEE978 - 1 - 4244 - 8840 - 7/10/26.00$ 2010 IEEE

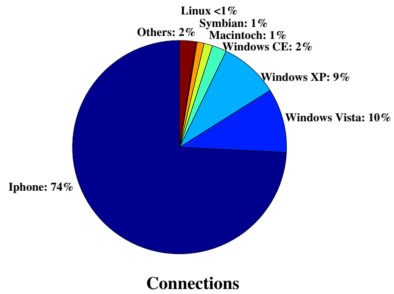



Fig. 6. OS and Devices for Webmail Traffic

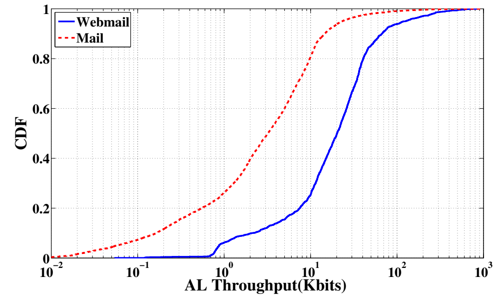



Fig. 7. Application Level Throughput

## VI. Mail and Webmail: Performance

In this section, we analyze in details the performance of mail and webmail traffic. We contrast the performance of those two ways for users to access their mailbox and try to understand if the higher popularity of webmail is due to its better performance.

### A. Application Level Throughput

Throughput is an important metric for a lot of applications. Common practice is to use throughput for applications generating bulk transfers, while response time is used for interactive applications. Mail traffic in general appears to be a mixed application, generating interactive and bulk transfers. Bulk transfers are generated by large mails (with attachments), while interactive transfers are due to mailbox checking and the sending/reception of small mails. In this section, we use the throughput to compare the performance of mail and webmail. Our purpose here is to show that the access technology influences the throughput but is not the only factor. Congestion, transport layer details or the application on top (e.g., rate limiters in p2p applications) can also impact the observed throughput.

Before presenting any result, we need to specify the way we compute throughput. We have shown in [19] that a straightforward estimation of throughputs where the amount of bytes transfered at the TCP layer is divided by the total duration between the first packet (first SYN) and last packet of the connection (e.g., FIN) provides a biased view of the throughput perceived at the user side. The tear down of a connection, that we define as the time between reception of the last data packet and the last control packet can be extremely high due to numerous reasons: the application, the server implementation or the operating system. We thus introduced in [19] the notion of Application-Level (AL) throughput where the amount of bytes transfered at the TCP layer is divided by the total duration between the first packet (first SYN) and last *data* packet of the transfer.

In Figure 7, we report the AL throughput for mail and webmail connections. A first striking observation is that webmail offers significantly higher throughputs than mail. More than 85% of webmail connections achieve a throughput higher than 10 kb/s, unlike mail where the equivalent portion is only 20%. Several factors can explain this discrepancy. In the following paragraph, we explore in more details mail and webmail traffic characteristics, in order to find which parameters degrade mail performance. We focus on volumes of data exchanged, application impact, and time spent to recover from losses.

### B. Detailed Performance Comparison

*1) Connections Size:* Figure 8 depicts the cumulative distribution of well-behaved mail and webmail connection size in bytes. It appears that mail transfers are clearly smaller than webmail transfers. This observation is in line with the results in Tables II and III where we noticed the smaller number of webmail connections but the larger amount of data exchanged. We believe that two factors explain this observation: (1) webmail applications not only convey data related to the mailbox of the user but also data related to the HTTP frame of the Web page in which the content of the mailbox is displayed, (2) web(mail) applications use persistent connections unlike legacy mail protocols (POP, SMTP - but not IMAP), which results in longer transfers. A smaller amount of data to transfer leads inevitably to a smaller throughput with TCP on average, which is a first explanation behind the observation of mail achieving smaller throughputs than webmail. As we will see later in this section, different connection size is not the only factor that explains the lower throughput of mail as compared to webmail.

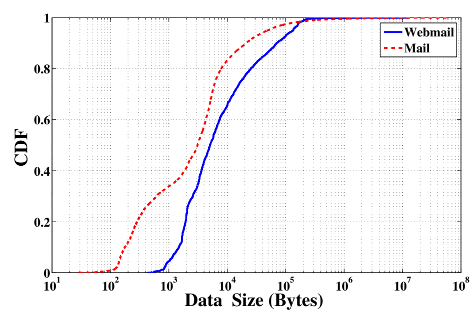



Fig. 8. Connections Size

*2) Impact of Application on Top:* For client/server applications, one generally observes that even if the server is sending a large amount of bytes/packets, the actual data exchange is fragmented: the server sends a few packets (hereafter called train), then waits for the client to post another request and then sends its next answer [19]. If such a behavior is predominant, it can have a detrimental impact to TCP if the train size is too small, as it prevents TCP from performing FR/R in the case of losses.

We evaluate here the distribution of train sizes for mail and webmail transfers. We distinguish between the initiator of the connection, which is in our case the Cellular client and the remote party, which is the mail or webmail server.

Figure 9 reports the distribution of train sizes for webmail and mail transfers. We observe that:

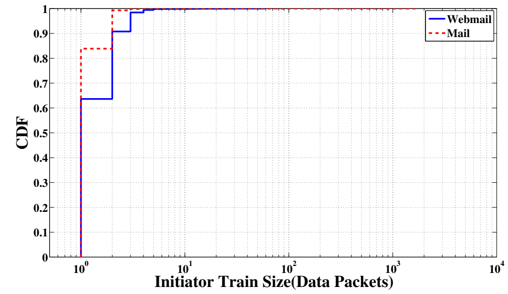
- Trains sent by servers (remote party) are larger than those sent by the initiator (local client);
- Webmail trains are larger than the ones of mail traffic, for both initiator and remote party. In fact, more than 38% of webmail initiator trains are larger than 2 data packets, unlike mail where it is only 16%.
- More than 99% of initiator mail and webmail trains are smaller than 3 data packets, which leaves TCP unable to trigger any Fast Retransmit, even if Limited Transmit is used [20]. This might lead to performance issues during mail uploads.
- More than 92% of remote party trains are also smaller than 3 data packets, compared to only 70% for webmail. This again leaves TCP unable to trigger a fast recovery/retransmit, even if Limited Transmit is used in a lot of case. Mail is more affected than webmail though.

A conclusion of the above analysis is that both mail and webmail throughputs are affected by the behavior of the application on top of TCP with a potentially more detrimental effect for mail than for webmail transfers. Smaller train sizes tend to slow down TCP, as it prevents the protocol from opening its congestion window, but can also lead to longer recovery time during loss events. We turn our attention to this specific issue in the next paragraph.

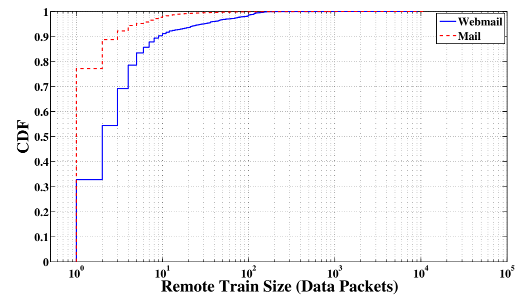
*3) Losses:* To assess the impact of TCP loss retransmission times on the performance of mail and webmail, we developed an algorithm to detect retransmitted data packets, which happen between the capture point and the server or between the capture point and the client. This algorithm is similar to the one developed in [9]. If ever the loss happens after the observation point, we observed the initial packet and its retransmission. In this case, the retransmission time is simply the duration between those two epochs[5]. When the packet is lost before the probe, we infer the epoch at which it should have been observed, based on the sequence numbers of packets. We try to separate real retransmission from network out of sequence events by eliminating durations smaller than the RTT of the connection.

Figure 10 plots the cumulative distribution of retransmission time per each loss event, for mail and webmail traffic. As

[5]Those epochs are computed at the sender side by shifting the time series according to our RTT estimate.




(a) Initiator Side




(b) Remote Side

Fig. 9. Exchanged Trains Size

expected from our study of train times, mail traffic experiences larger recovery times than webmail traffic.

We can further notice two thresholds of common retransmission times at 400 ms and 1 seconds for webmail and mail respectively.

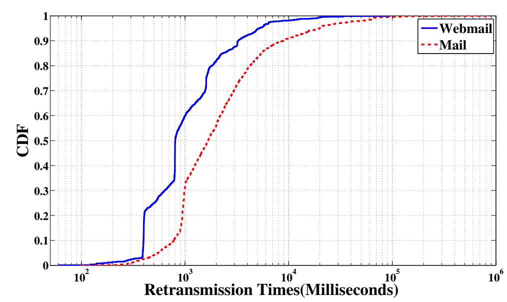



Fig. 10. Retransmission Times per Loss Event

In summary, several factors contribute to the degradation of mail performance as compared to webmail. Some of these factors are driven by clients usage while others are more fundammentally related to the way those different mail implementations work and their interplay with the transport layer.

## VII. Conclusion

In this work we have reported some observations about the Internet traffic of a Cellular network with users connected via handsets or USB pluggable 3G modems. The predominance of Iphone however suggests that the first category of users actually dominates over the second one. Among our findings, we observed that Cellular clients tend to almost consistently download more than they upload, in contrast to ADSL and

Fiber technologies. We have highlighted that measurements from passively collected traces can be biased by specific technologies implemented in Cellular networks to boost performance and control users activity. RTT, which is a key metric, is especially affected by those network appliances.

We cast a first look to mail and webmail traffic in Cellular networks. We found that mail seems to be less popular than webmail as the majority of mail data is transfered using webmail. A first explanation to this difference in usage is the high popularity of free webmail service providers like Google, Yahoo!, Hotmail, etc. We indeed observed that those providers are much more used than the webmail services offered by the network providers. This is presumably because users want an email account that is independent of their network provider, in case they switch to another network provider.

We further observed that webmail performance outperforms the one of mail. We demonstrated that several factors lead mail to offer smaller throughputs than webmail, especially, the size of the transfers, the application semantics which leads to smaller data exchange phases, which slows down TCP in general and prevents fast retransmit if losses are detected.

As future work, we intend to develop a more systematic approach to precisely pinpoint the different factors that affect the performance of applications in Cellular networks. We intend to apply this methodology to other key applications in Cellular networks, e.g., HTTP streaming or Web searches. Ultimately, we would like to contrast the performance of different access technologies by profiling services commonly used by customers over wired and wireless accesses.

## Acknowledgment

This work has been supported by Orange Labs. Authors would like to thanks Orange Labs for their help for capturing the trace.

## References

[1] G. Maier, F. Schneider, and A. Feldmann. A first look at mobile handheld device traffic, zurich. In *PAM*, 7-9 Apr, 2010.

[2] P. Romirer-Maierhofer, F. Ricciato, A. D'Alconzo, R. Franzan, and W. Karner. Network-wide measurements of tcp rtt in 3g. *TMA '09: International Workshop on Traffic Monitoring and Analysis, Aachen*, pages 17–25, May 2009.

[3] S. Pukkawanna, S. Visoottiviseth, and P. Pongpaibool. Classification of web-based email traffic in thailand. *In the Proceedings of IEEE ISCIT2006, Bangkok*, 2006.

[4] J. Charzinski. Observations in e-mail performance. *ITC Specialist Seminar, Wuerzburg*, Jul. 2002.

[5] H.A. Bazar, R. Sureswaran, and O. Abouabdalla. A new approach to enhance e-mail performance through pop3 protocol. *NetApps2008, IProceedings,Sintok*, 21-22 Nov, 2008.

[6] G. Vu-Brugier. Analysis of the impact of early fiber access deployment on residential internet traffic. *International Teletraffic Congress, Paris*, June 2009.

[7] M. Pietrzyk, JL. Costeux, G. Urvoy-Keller, and T. En-Najjary. Challenging statistical classification for operational usage: the adsl case. pages 122–135, New York, NY, USA, 2009. ACM.

[8] H. Kim, KC Claffy, M. Fomenkov, D. Barman, M. Faloutsos, and K. Lee. Internet traffic classification demystified: myths, caveats, and the best practices. pages 1–12, New York, NY, USA, 2008. ACM.

[9] S. Jaiswal, G. Iannaccone, C. Diot, J. Kurose, and D. Towsley. Measurement and classification of out-of-sequence packets in a tier-1 ip backbone. *IEEE/ACM Trans., Piscataway*, 15(1):54–66, 2007.

[10] H. Jiang and C. Dovrolis. Passive estimation of tcp round-trip times. *SIGCOMM Comput. Commun. Rev., Pittsburgh*, 32(3):75–88, 2002.

[11] S. Shakkottai, R. Srikant, N. Brownlee, A. Broido, and KC. Claffy. The rtt distribution of tcp flows in the internet and its impact on tcp-based flow control. 2004.

[12] B. Veal, K. Li, and D. K. Lowenthal. New methods for passive estimation of tcp round-trip times. *PAM, Boston*, pages 121–134, 2005.

[13] Y. Zhang, L. Breslau, V. Paxson, and S. Shenker. On the characteristics and origins of internet flow rates. *SIGCOMM '02, Pittsburgh*, pages 309–322, 2002.

[14] T. Karagiannis, K. Papagiannaki, and M. Faloutsos. Blinc: Multilevel traffic classification in the dark. *ACM SIGCOMM, Philadelphia*, 2005.

[15] L. Bernaille, R. Teixeira, I. Akodkenou, A. Soule, and K. Salamatian. Traffic classification on the fly. *ACM SIGCOMM, Philadelphia*, 2005.

[16] T. Karagiannis, A. Broido, M. Faloutsos, and K. Claffy. Transport layer identification of p2p traffic. *4th ACM SIGCOMM conference on Internet measurement, Sicily*, 2004.

[17] B. Du, M. Demmer, and E. Brewer. Analysis of www traffic in cambodia and ghana. *15th international conference on World Wide Web, Edinburgh*, pages 771–780, 2006.

[18] M. Pietrzyk, G. Urvoy-Keller, and J.L. Costeux. Revealing the unknown adsl traffic using statistical methods. *TMA '09: International Workshop on Traffic Monitoring and Analysis, Aachen*, May 2009.

[19] A. Hafsaoui, D. Collange, and G. Urvoy-Keller. Revisiting the performance of short tcp transfers. *8th International IFIP-TC 6 Networking Conference, Aachen*, pages 260–273, May 2009.

[20] M. Allman, H. Balakrishnan, and S. Floyd. Enhancing tcp's loss recovery using limited transmit. *RFC:3042*, 2001.