MODELING E-MAIL TRAFFIC FOR 3G MOBILE NETWORKS

Philipp Svoboda, Wolfgang Karner, Markus Rupp
Institute of Communications and Radio-Frequency Engineering
Vienna University of Technology, Austria
Gusshausstrasse 25/389, A-1040 Vienna, Austria
Email: {psvoboda, wkarner, mrupp}@nt.tuwien.ac.at

ABSTRACT

Traffic models are to provision the increasing load of packet switched traffic in mobile 3G networks. However, most traffic models were designed to fit the need of low delay wired connections. In this paper we propose a new traffic model for POP3 e-mail traffic generation in 3G mobile networks scenarios. The model is based on traces which were extracted in a live 3G network of a mobile operator. The main improvement of the model is the explicit modeling of the login process that takes place between the client and the server. We implemented the model in the network simulation tool ns-2 and based on this we simulated various scenarios to visualize the impact of a leading initialisation phase in high RTT environments. The results proof that the extension can reproduce the effect from the login in a more precise way. Finally a comparison between the measured and the simulated service footprints were compared. The new model performs better than common models in this scenario.

I INTRODUCTION

In the recent past many users migrated from a wired internet connection to a mobile internet access. The growing popularity of mobile internet access technologies has brought mobile services into the focus of recent research. Due to the fundamental differences between the wireless and the traditional wireline systems and protocols, services encounter a different real-world performance. Today the mobile world includes many different radio technologies like GPRS, EDGE, UMTS and HSDPA, each featuring different bandwidth and RTTs (Round Trip Time). GPRS, the very first real packet switched wireless access, performs rather poorly while HSDPA is a fast access competing even with ADSL lines. Traffic models have to be sensitive to the properties encountered in such systems.

In the internet the following three protocols are used to transmit most of the e-mail traffic: SMTP (Simple Mail Transfer Protocol) [1], POP3 (Post Office Protocol v.3) [2] and IMAP (Internet Message Access Protocol) [3].

SMTP is used mainly to send e-mails between local systems. The protocols POP3 and IMAP were designed for clients who only sporadically connect to the e-mail server. E-mails are stored on the server and transfered to the client on demand. All protocols rely on the TCP (Transmission Control Protocol) for an error free end to end delivery. In this paper we introduce a traffic model for the POP3 e-mail service in a mobile environment. As we have shown in [4], this service is within the top three of the service share. Although it is smaller than web traffic in terms of volume, most business customers rely on the proper function of this service. In contrast to wire line connections 3G networks impose a high delay in their links. Therefore, a proper modeling of this service needs a different approach than in the low delay case of wired connections. In a low delay scenario a login process can be omitted and the e-mail transfer is reduced to some kind of push service only.

The rest of this paper has been organized in the following way. In Section II we show the measurement setup, present the monitoring interfaces and explain the collected data. Section III presents and discusses the numbers extracted from the monitoring system, introduces a model for e-mail services and analyzes the performance compared to other models. Related work is cited in Section IV Several conclusions are drawn in Section V

II THE MEASUREMENT

The measurement setup in the reference network scenario is presented in Figure 1. As with most access networks, the 3G mobile network has a hierarchical tree-like deployment. The MSs (Mobile Station) or UEs (User Equipment) and the base stations are geographically distributed. Going up in the hierarchy (first BSC/RNC, then SGSN, ultimately GGSN) the level of concentration increases, involving progressively smaller number of equipments and physical sites. In a typical network there are relatively few SGSNs and even fewer GGSNs. Therefore, it is possible to capture the whole data traffic from home subscribers on a small number of Gn/Gi links. For further details on the structure of a 3G mobile network refer to [5]. In order to meet privacy requirements traces are anonymized by hashing all fields related to user identity at the lower 3G layers (IMSI, MSISDN, etc.), and removing the user payload above the TCP/IP layer.

The input traces were captured on a live GPRS/UMTS network with the METAWIN monitoring system [6]. It is a monitoring tool designed to record traffic in a mobile core network capable of decoding all intermediate protocols used in such networks (see [5]).

The core dataset used in this work consists of four different one-day periods in April 2007. The measurements were taken from the live network of a mobile provider in the EU. We extracted our data from Gn interface traces. Monitoring at the Gn interface is a good trade-off between the available signaling information and the number of links to be tapped. At the Gn interface we are able to correlate the data traffic with specific users. Therefore, we can extract values such as the number of e-mails, the inter service time and the volume on a per user base. This information gives us the opportunity to model the service the way it is accessed by the customer, rather

1-4244-1144-0/07/$25.00

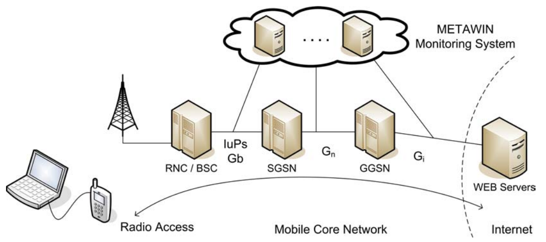


Figure 1: Measurement Setup in a 3G Network

than aggregating all the service calls to derive a standard profile per customer. We applied our analysis at the TCP/IP level extracting port numbers to identify service calls. With this information we cannot parse the application level.

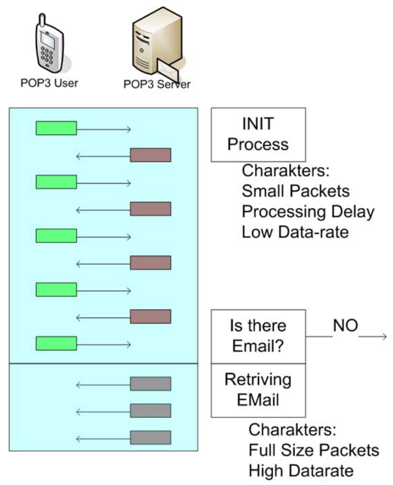
Measurements in a live network often encounter a bias generated from some heavy hitters. In this case we found out that basically more than 20% of all e-mail requests originated from fewer than 0.25% of the users. We filtered out this group in a preprocessing step. The detection was implemented by marking all IMSIs as “bad” which had more than 100 requests within six hours in any sub part of the traces (Note: More than 99% of the users had fewer than 80 requests within the full week). This rule excluded fewer than 0.32% of the population from the traces.

III Traffic Model

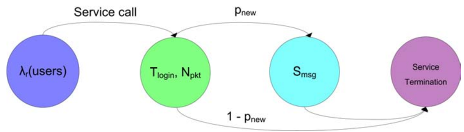
In the literature e-mail models cover the e-mail size [7, 8]. However, in order to receive his e-mail the user first has to authenticate himself with his password at the server. This login process transfers only a small amount of data. We decided to use our traces to set up a model for e-mail traffic that is suited for mobile internet access technologies. Figure 2(a) gives the principle idea behind these thoughts. The goal is to separate the login procedure from the e-mail download itself. This has two benefits: first we can model the different payloads on the IP-level separately (login = small payload, download = large payload) and second we can visualize the impact of high RTTs at the link. In this model a high RTT, such as GPRS, will increase the service time significantly.

A Extracted Parameters

In a first step of the extraction process we evaluated the number of users connecting to the e-mail services. Due to the NDA restrictions we cannot disclose the absolute numbers. However, we will present normalized values. In the measurement period 22.1% of the customers used e-mail. This number is surprisingly low, which comes from the fact that also mobile terminals, only operating on WAP, are included in the population. However, these users will neither use e-mail nor TCP services. Note, that in this context we calculated the fraction of active users to users downloading an e-mail message. An active user has to have at least one PDP-context, meaning that only data active users were counted. Calculating the share with respect to datacard customers we obtained a fraction of 86.8%, indicating that almost all of them access e-mail servers. Due to the



(a) Service call of a single user



(b) Model states for traffic generation

Figure 2: Simulation model for POP3 service

lack of IMEI tracking the determination of the terminal type has to rely on a service based rule set. We consider a terminal as a datacard if no WAP traffic was recorded in the dataset similar to [4]. In the following sections we will only focus on datacard users.

1) Service Request Rate

In a next step we evaluated the number of service calls generated per user per day. Fig. 3 shows the empirical CDF of service calls per user. It is interesting to see that both technologies, GPRS and UMTS, represented by the solid and by the dashed line respectively, are very similar to each other. This indicates that the service usage is decoupled from the access technology. Further investigation into this topic showed that also other parameters such as the e-mail size did not differ between the technologies. Therefore, we did not extract the following parameters for GPRS and UMTS separately.

Considering a request generation process, combined for all users in the network, we extracted the inter-arrival time between two consecutive service calls in the trace. The parameter for 1000 active users during the busy hour follows an exponential distribution with the rate $\lambda_r^{-1} = 0.01102$.

We implemented the service generation using an exponential distributed rate. The implementation of a single user as a generation source is something we may realize in a further step of our research.

2) Probability of login-only Session

According to our e-mail model we had to discriminate between a simple login with no e-mail download and a login with an e-mail download. Without a parser at the application layer we had to find a compromise in order to detect service calls with

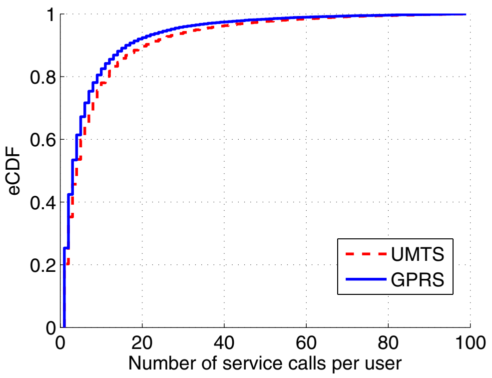


Figure 3: Number of POP3 service calls per user and per day for UMTS and GPRS

no download of e-mails. An analysis of the POP3 protocol (see [9]) revealed that a normal login process should have fewer than 13 packets, including the TCP handshake. We set our arbitrary limit to 15 packets for up- and downlink direction for connections which we consider to be login only. Further, we also set a lower limit of 5 packets in up- and downlink direction for a connection to be counted as valid. This rule filters out most of the port scans and attacks polluting todays internet traces. The final number was 22.4% for the fraction of logins combined an e-mail transfer. In other words, the probability that a user has at least one new e-mail in his mailbox, called p_{new}, is 22.4%. Therefore, simulating the login process separately will be an important part of the model. The volume for all detected login connections is below 300 byte.

3) Size of an E-Mail Message

The e-mail size, called S_{msg} was the next parameter we evaluated. In our case the name E-Mail Size is a bit misleading, as the POP3 service can transfer many e-mails in one connection. The parameter is the volume in bytes a user will transfer in case there is new mail at the server. Figure 4 presents the extracted values for the e-mail size. We also included values from the FUNET experiment [7] and the ones suggested by Paxson [8, 10]. The FUNET model was derived from traces collected in a Finnish university and research network. It uses a Cauchy distribution with a cut-off, S_{max}, for the maximum message size of 10 kbyte.

$$f_{Cauchy}(x, 0.8, 1) = \frac{1}{\pi \cdot (1 + (x - 0.8^2))}, \; x \in \mathbb{N} \qquad (1)$$

Paxson [8] characterized the mail size using two log2 distributions, one for small and one for large e-mails. A fixed quota of 300 byte was added to both distributions in order to model the application overhead, S_{head}, within the login process of every e-mail download. The paper [8] already mentions that it is problematic to model the overhead as it is a function of the number of servers the e-mail has to pass, similar to a hop count, in order to reach its destination. Each server adds several lines to the header of the message, increasing the total size. It is assumed that 20 % of all e-mails follow the larger distribution while the remaining 80 % originate from the smaller distribution. The cross over between the two distributions was set to 2 kbyte. The mail size is limited to 100 kbyte.

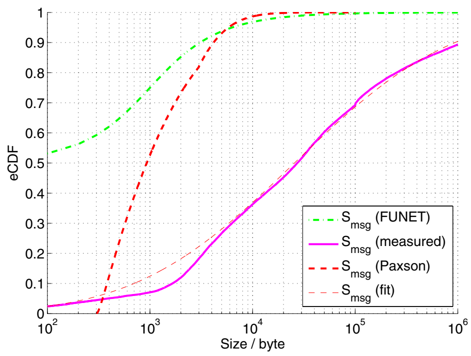


Figure 4: Volume transfered per service call

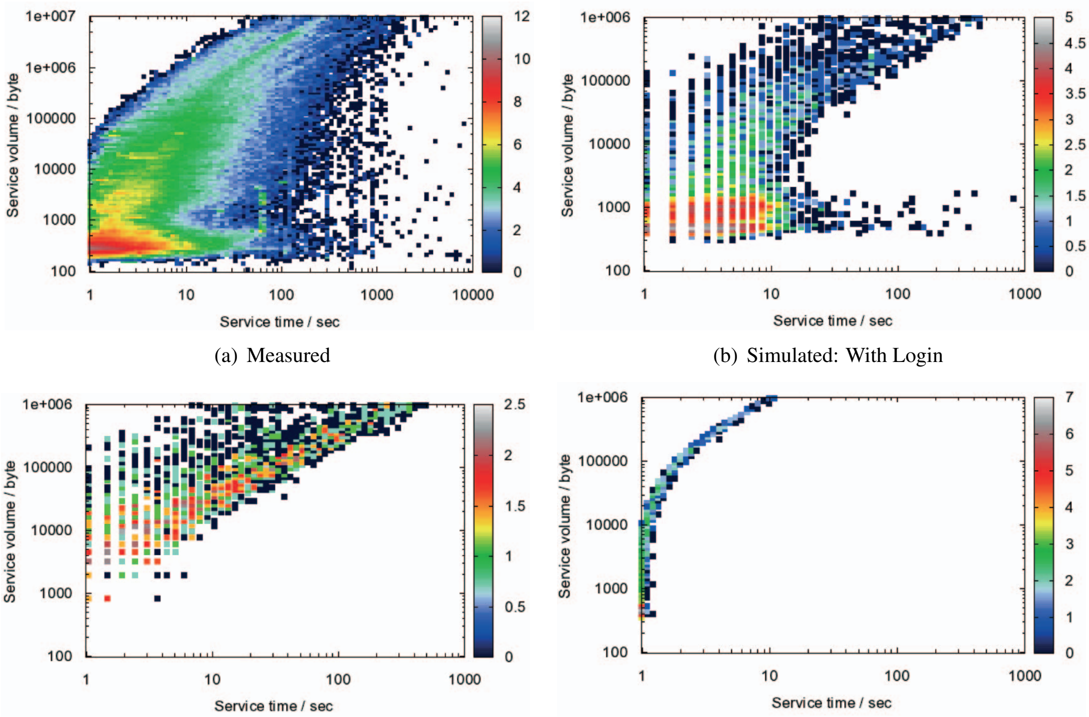
Paxson			
Model Parameter	Distribution	Parameter	
S_{msg} small / byte	$\log_2$-normal	μ =10.0	σ =2.75
S_{msg} large / byte	$\log_2$-normal	μ =9.5	σ =12.8
S_{head} / byte	Constant	x =300	—
S_{max} / byte	Constant	x =100 k	—
FUNET			
Model Parameter	Distribution	Parameter	
S_{msg} / byte	Cauchy	0.8	1.0
S_{max} / byte	Constant	10 kB	—
New model			
Model Parameter	Distribution	Parameter	
S_{msg} / byte	Log-Normal	μ =10.151	σ =2.8096
N_{pkt} / byte	Uniform	min =10	max =12
T_{login} / ms	Uniform	min =100	max =200
λ_r^{-1} / s	Exponential	μ =0.0110	—
p_{new}	Constant	x =22.4 %	—

Table 1: Model Parameter

The values for both models already include the login process. Therefore, the CDF starts at different values compared to our mail size. The median value from Paxson for the e-mail size is more than one order of magnitude smaller than the actual values measured in the live 3G mobile network. The FUNET estimation for the e-mail size is even smaller than in the Paxson model. We conclude that this was due to the age of the models. We fitted the parameter mail size with a log-normal distribution. The login process is modeled as a number of packets, N_{pkt}, with a fixed packet size of 30 bytes payload each. An uniform distribution was used to model the varying number of packets per service call. The processing time at the server side, T_{login}, is also implemented as an uniform distributed variable.

Table 1 compares the fittings of the parameters for the different models. Note that the Paxson values are based on a log-normal distribution with base two. The parameter λ_r^{-1} given in Table 1 is based on 1000 active users, within the busy hour, which lasts from 7.45 to 8.45 p.m., averaged over all four traces. Figure 2(b) presents the state model for the e-mail model using the parameters from Table 1.

In this paragraph we will summarize the model. If a service



(a) Measured

(b) Simulated: With Login

(c) Simulated: No Login

(d) Simulated: Login, ADSL

Figure 5: POP3 Scatter Plot: Service Time vs. Service Volume (log binning, linear coloring)

request is generated the model enters the login phase. This phase is executed for every user. At the end of the login process the service call will start downloading an e-mail with the probability p_{new} and terminates the session after the successful transfer of the e-mail. If the model decides not to download an e-mail message the service call will terminate right after the login phase. The number of users in the network can be changed by increasing the request rate.

B Simulation Setup

In this section we will describe the simulation setup that we used to verify our new model. We used ns-2 to verify our e-mail model. A basic setup with two nodes was used to simulate the underlying network conditions. The link parameters were set accordingly to the RANs (Radio Access Network) which we have been monitoring. The bandwidth was set to discrete values of 40, 64, 128, and 384 kbit. We simulated the RTT as an uniform distribution from 200 to 500 ms, accordingly to values we have measured for the TCP round trip time in [11].

Into this artificial scenario the new e-mail agent was implemented. The agent supports e-mail transfers optional with and without login. It is therefore possible to extract the impact of the login session separated from the e-mail download process. The application uses the e-mail size according to our measurements in the network.

C Measured Service Footprint

As a first result we extracted a scatter plot of the service time over the data volume transfered. Each point in the scatter plot is equal to one POP3 service call of a client in the network. Hence a single user can be represented with several points in these scatter plots. In this scatter plot both axis were binned in a logarithmic way. The color of each dot indicates the number of occurrences in this specific [x,y] bin in the plot. Figure 5(a) presents the service footprint recorded in the mobile core network at the Gn interface of the provider.

In this figure there are two main clusters visible. The lower cluster is due to users who only access the system and do not find any new e-mail at their account. They complete the login procedure but do not transmit any e-mails, this is visible from the fact that they only have a very limited traffic volume per service call. The wide spread in time reveals the large variation in the RTT for the different terminal connections. This is consistent with our finding in [11].

The second cluster above 1 kbyte is generated by terminals which have new e-mails waiting at the server for delivery. This cluster is modeled by the mail size parameter we extracted in the previous section. The datarate of these terminals is limited to the parameters of the used RAN. Therefore, there is an upper boundary at a net datarate of approximately 1 Mbit.

D Simulation Results

In this section we present the results of the simulation. To compare the output of our model with the measured results presented in Fig. 5(a) we simulated different setups with login and without login process. The second scenario is similar to the Paxson model, where in fact only the size of the mails is distributed differently. Both scenarios can be used to analyze the impact of a separate login process emulation. While the third setup is an ADSL like internet access.

The result of the first scenario is depicted in Fig. 5(b). The figure is very similar to the measured result in Fig. 5(a). The 2D correlation coefficient of this result with the measured data is 0.91, which indicates a high similarity. However, the variation in the direction of the x-axis is smaller than for the recorded results. This is due to the fact that we only use a simple emulation of the underlying RAN system, introducing limited variation.

The result of the second scenario in Fig. 5(c) differs strongly from the recorded trace in Fig. 5(a). Here the correlation is only 0.3, which indicates the large difference between the results. As the login emulation is not active there is no cluster and no additional delay from the network. The data is directly pushed to the mobile terminal.

Finally we reran the first scenario with the link parameters set accordingly to an ADSL line (bw = 1 Mbit, RTT = 10 ms). Figure 5(d) presents the results for this setup. The link delay is two order of magnitudes smaller than in the mobile RAN case. Therefore, the result is similar to a scenario where the login process was neglected. We conclude that this is also the reason why common e-mail models, designed for fast LAN connections, skipped the login process.

IV Related Work

The e-mail service POP3 generates only a very small share of the total traffic volume transfered in the internet. Therefore, only a limited number of models from relatively old studies exist [7, 8]. Due to their age these implementations did not consider the transmission over links with high RTTs. The login process is modeled only as a constant offset in the mail size.

A number of research groups has worked on the impact of mobile access to the TCP layer, identifying several shortcomings and proposing optimizations [12]. However, up to now to our knowledge this is the first model capable of capturing the impact of the login process in high RTT networks.

V Summary and Conclusions

In this paper we proposed a new traffic model for the POP3 e-mail services in 3G mobile networks. Common models do not capture the increase in service time due to the login process [7, 8, 10]. Such an implementation is well suited for LANs with relatively low RTTs. However, they do not reflect the service behavior in high RTT mobile RANs.

We extracted the input parameters for the traffic model from traces taken with the METAWIN system. It is a monitoring system installed at the core network of a 3G mobile operator, capable of capturing statistics on the user level over various network links. In this work we used four one-day period traces that were recorded at the Gn interface in April 2007.

The proposed e-mail model splits up a POP3 e-mail transaction into two parts: the login process and the e-mail download. Every user's request will run through the login process. However, only a small fraction will then download e-mails. Hence, the model is capable of emulating service calls that have no e-mail transfer with the server, e.g., only check for mail. In our traces more than 80% of the service calls did not download any e-mail, which was also implemented in our model.

To benchmark the model we simulated three different scenarios. First we analyzed the impact of the login process compared to common models with no-login. With the login emulation the service footprint is very similar to the measurement results. A second scenario showed that the common models can not generate TCP footprints similar to the measurements. In an ADSL like environment the simulated impact of our model was small.

In our further research we will analyze the impact of bandwidth per user in order to perform detailed network planning operations.

Acknowledgements

This work is supported by the K-Plus project within the Darwin project. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the funding partners.

References

[1] Ed. J. Klensin. RFC 2821: Simple Mail Transfer Protocol. Technical report, Internet Engineering Task Force (IETF), 2001.

[2] J. Myers and M. Rose. RFC1939: Post Office Protocol Version 3. Technical report, Internet Engineering Task Force (IETF), 1996.

[3] M. Crispin. RFC2060: Internet Message Access Protocol Version 4 . Technical report, Internet Engineering Task Force (IETF), 1996.

[4] P. Svoboda and F. Ricciato. "Composition of GPRS and UMTS traffic: snapshots from a live network". *IPS MoMe 2006, Salzburg*, 4:42–44, Feb 2006.

[5] H. Holma and A. Toskala. *WCDMA for UMTS, Radio Access For Third Generation Mobile Communications, Third Edition*. Wiley, 2004.

[6] METAWIN Project. `http://www.ftw.at/ftw/research/projects/`. FTW.

[7] G. Brasche and B. Walke. Concepts, services, and protocols of the new gsm phase 2+ general packet radio service. *IEEE Communications Magazine*, 35:94–104, 1997.

[8] V. Paxson. "Empirically-Derived Analytic Models of Wired-Area TCP Connections". *IEEE/ACM Transactions on Networking*, 3:226–244, June 1995.

[9] B. Lloyd and W. Simpson. RFC1334: PPP Authentication Protocols. Technical report, Internet Engineering Task Force (IETF), 1992.

[10] P. Stuckmann. *Traffic Engineering Concepts for Cellular Packet Radio Networks with Quality of Service Support*. PhD thesis, Technischen Hochschule Aachen, 2003.

[11] F. Ricciato, F. Vacirca, and P. Svoboda. "Diagnosis of Capacity Bottlenecks via Passive Monitoring in 3G Networks: an Empirical Analysis". *Computer Networks*, 57:1205–1231, March 2007.

[12] T. Madsen, P. Schwefel, M. Hansen, J. Bogh, and R. Prasad. "On Traffic Modelling in GPRS Networks". pages 1785–1789, 2005.