Measuring and Modeling of Application Flow Length in Commercial GPRS Networks

Roger Kalden, Ericsson Research, Aachen, Germany, roger.kalden@ericsson.com
Boudewijn R. Haverkort, University of Twente, The Netherlands, brh@cs.utwente.nl

ABSTRACT

New mobile access networks provide reasonable high bandwidth to allow true internet access. This paper models two dominant applications of those networks. One application, WAP, is novel and specific to mobile networks, the other is HTTP, which is already dominantly present in the internet. However, our measurements reveal that WAP traffic is the dominant application in mobile networks. Therefore, this paper models in particular the flow length of WAP and as comparison the flow length of HTTP. The flow length is a good approximation of the object size structure and hence can be used in further simulation or analytical modeling of mobile networks. We apply different distribution fitting techniques and the KS goodness of fit test to specify exactly the underlying distribution of the empirical data.

I. INTRODUCTION

New mobile networks of the third and evolved second generation have been deployed that enable a mobile accessible, always available wireless Internet. To exploit the full potential from those new mobile access network technologies, and to meet high user expectations, these networks need to be well designed, dimensioned and deployed. Fundamental for executing these tasks are accurate traffic models.

Therefore, the goal of this paper is to provide a sound mathematical description of the data flow length of the two major applications HTTP and WAP[1] in a network of the evolved second generation (GPRS). These results are first of their kind and based on a huge comprehensive measurement study. The models can be deployed in simulative and analytical investigations to further optimize protocols and to plan and tune operational networks.

In [1] we introduced a measurement set-up to collect commercial GPRS IP traffic together with GPRS session information. This set-up is unique as it allows correlating information on the application usage together with GPRS internal user session and mobility events. In the present paper we used that same set-up to measure up-to date GPRS data and to derive actual distribution functions for further modeling.

The dominant mobile web applications in terms of frequency of usage and data volume are HTTP and WAP [1][2][3]. We assume that this will persist in the mid-term future as well. HTTP is the dominantly used protocol for laptop-like access devices, and WAP [3] is dominant for small mobile devices. Currently WAP 1.2 is mainly being deployed, though WAP 2.0 [9] is around the corner and will be largely seen in upcoming terminals and services[2]. Note, that WAP 2.0 is using wHTTP [7], which is a specific configuration of HTTP. But nevertheless, application-wise this is WAP and consequently we will discuss implication thereof in this paper as well.

We presented in [2] first novel results on the length distribution of HTTP and WAP flows, based on the same measurements. However, in [2] our focus had been on the *quantitative* determination of flow length and its implications thereof. Especially the body of the length distribution attracted our interest, as it can have serious implications on the performance of TCP in case of packet loss. The results have been used to reveal typical mobile scenarios for TCP and to suggest optimizations for TCP.

In the present paper we take a different approach and focus on the *distribution* of the flow length. Mathematical models thereof can be deployed in simulative and analytical models.

We use two methods to obtain distribution functions for the measured flow length. On the one hand we use the maximum likelihood estimation (MLE) method [13] together with the Kolmogorov-Smirnov (KS)-Test [14] to determine a general function (e.g. exponential, pareto, etc.) matching the measured data. Knowing such functions provides insight into the properties (e.g. heavy-tailedness) of the measured distribution and allows extrapolation to non-measured regions. On the other hand we use an expectation maximization (EM) algorithm to derive parameters for phase type distributions matching the measured data [15].

The class of phase-type distributions is especially useful in traffic modeling, as these distributions can be used to approximate arbitrarily close all probability distributions having a rational Laplace transformation. Furthermore, traffic modeling problems which have an explicit solution assuming exponential distributions are often also algorithmically tractable when one replaces the exponential distribution with a phase-type distribution. Therefore many intractable queuing problems can be approximated by replacing general distributions by phase type distributions.

[1] WAP can be described as a lightweight web application tailored specifically for mobile devices.

[2] WAP 1.2 uses a specific wireless stack consisting of specially designed WAP protocols. WAP 2.0 uses the widely deployed internet protocols like TCP and HTTP, specifically tuned for mobile devices.

The structure of the paper is as follows. First, we briefly introduce GPRS, and our measurement set-up and data. Next, we define application flow length as used in this paper. Reasons for focusing on WAP and HTTP will be provided in the following section by discussing the dominance of WAP and HTTP flows. Next, we explain the used distribution fitting methods. Subsequently, in the main part, we use the measured data and the present fitting method to derive suitable distributions for the HTTP and WAP flow length. We finish the paper with a conclusion and outlook on further work.

II. GPRS AND MEASUREMENT SET-UP

We briefly introduce the investigated mobile network system and our deployed measurement set-up.

GPRS is a packet switched network embedded in the GSM cellular network. GPRS provides a resource efficient, always online, IP packet switched connection to the Internet, offering favorable characteristics to read Emails, browse the Internet and accessing similar applications over the Internet from a mobile terminal. Over the air interface GPRS offers data rates up to 160 kbit/s.

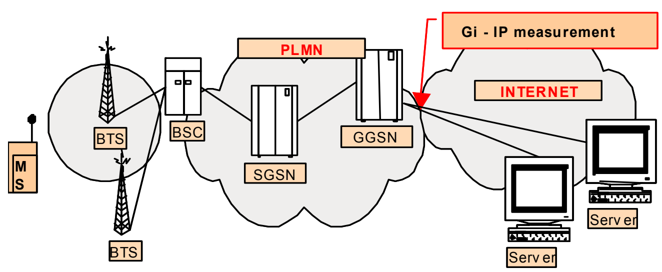

Figure 1: Measurement Set-Up

Fig. 1 depicts the nodes of a GPRS network. Many nodes are inherited from GSM. For us the important node is the Gateway GPRS Support Node (GGSN) which is involved in session management, IP address handling and provides the interface to external IP networks. The Gi interface at the GGSN is the connection between the GPRS PLMN (Public Land Mobile Network) and any IP based network. All servers (Email server, WAP gateway, Web server) and the "Internet" are located outside of the PLMN. Hence all traffic from and to the terminals need to traverse the Gi interface [10].

Measurements are conducted by a measurement PC capturing header information of by-passing IP packets on the Gi interface (see Fig. 1, "GI - IP measurement"). The IP packet trace is post-processed by a tool that extracts flow information from the packet traces. The flow information consists of the starting time of the flow (time stamp), length in bytes, length in packets, type of application and some auxiliary data.

The traces used in our analysis were recorded in the Vodafone GPRS network of a European country. The traces comprise 136 Million IP-packets and about 12 Million flows. The majority of the users in our data set are "private" users using GPRS to access specific Email, MMS or WAP services of the operator or the "open" Internet. To a small extend our traffic also includes traffic from corporate users. We cross-checked our results from this paper with measurements conducted in a different Vodafone network. Though the detailed statistics are slightly different, the overall conclusions remain the same.

III. FLOW DEFINITION

Investigating flows is a means to look at application objects at transport protocol level, without dissecting the actual application objects.

A flow is in the most general sense a sequence of packets that are related by the application object they carry. We define the length of flows as the length in the direction of the object data transfer. In particular HTTP and WAP flows are defined as follows.

A. HTTP

A *HTTP flow* is defined as all packets carried in a single TCP connection. The TCP connection is defined by the quadruple IP-address and port number for source and destination and is initiated by a TCP SYN and ends with a TCP FIN. Further, according to our definition, HTTP flows use HTTP ports (e.g. port 80 and port 8080) on the server side and carry *HTTP Get* or *HTTP Post* messages. Depending on the used HTTP version, a single TCP connection might be used to transfer one (HTTP 1.0) or several HTTP objects (HTTP 1.1, persistent connection).

B. WAP

A *WAP 1.2 flow* is defined by a UDP flow (i.e. the quadruple IP-address and UDP-port for source and destination) that carries one WAP object (e.g. a WML page or an embedded image). The download of an object in WAP 1.2 consists of a *WSP Get* command from the client, followed by a *WSP Reply* from the server [5][8]. No *persistent connection*, i.e., the technique whereby multiple application-layer objects may be transferred on a single transport-layer connection, is defined for WAP 1.2. Hence each *WAP 1.2 flow* consists of exactly one WSP Get/Reply Object download.

In our data set, we did not observe any traffic generated by WAP 2.0. However, as we assume that WAP 2.0 will replace WAP 1.2 in the near future, we extend our WAP flow definition and include results on WAP 2.0 as well. Our aim in this paper is to deduce the traffic characteristics of future WAP 2.0 flows from the measurements of WAP 1.2 that have been carried out. WAP 2.0 will use wTCP and wHTTP as transport and transaction protocol.

In order to extrapolate, we have made some assumptions about how WAP 1.2 relates to WAP 2.0. First, we assume that wHTTP in WAP 2.0 will use persistent connections [2][12], i.e., the request and transfer of several application layer objects can be carried out using a single TCP connection. We further assume that the object structure is the same for WAP 2.0 as it is in our measurements for WAP 1.2. When extrapolating from WAP 1.2 to WAP 2.0, we use only the

WAP objects without the transport protocol overhead. This allows hopefully a good prediction about the structure of WAP 2.0 flows. Based on these assumptions, our WAP 2.0 flow is defined by a sequence of WAP 1.2 flows carrying one or several WAP objects (e.g. a WML page or an embedded image) to or from one server (Fig. 2). A new WAP 2.0 flow starts if either an object from a different server is requested or if the inter-arrival time between the last acknowledgement of the previous WAP 1.2 flow and the next WSP Get of the succeeding WAP 1.2 flow is larger than a, heuristically defined, time-out value T_{IAT}. We denote those flows *WAP 2.0 flows (T_{IAT} sec)*. Based on several possible values for T_{IAT} [3], we focus on T_{IAT}=15sec in the following analysis. This is the default value for persistent TCP connections in Apache web servers.

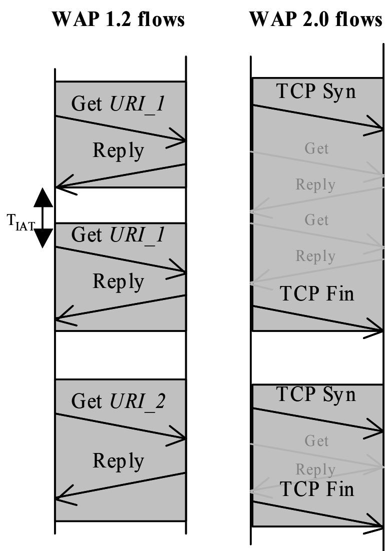

Figure 2: WAP 2.0 flow

IV. Application and Protocol Statistics

Motivation for focusing on HTTP and WAP applications as major GPRS applications is given in this section. Focusing alone on the transport protocol, already the statistics for the captured GPRS traffic looks quite different compared to corresponding statistics from the wireline Internet. In our measurement trace, UDP contributes to 30% of the transferred bytes. Based on our flow definition we have about 80-90% of UDP flows in the same measurement trace, which is considerably higher than what is commonly reported from measurements carried out in the wireline Internet. For comparison, [13] reports 95% of all bytes, 85%-90% of all packets, and 70%-75% of all flows as belonging to TCP.

Figure 3 depicts the distribution of application protocols on all flows in our GPRS trace. WAP currently contributes most of the flows, which accounts mainly for the large number of UDP flows.

When WAP 2.0 is introduced, the UDP flows carrying WAP will turn into TCP flows. However, this does not change our assumption about WAP being a main application. In particular, using our definition of a WAP 2.0 flow, the fraction of WAP flows would be reduced, since one WAP 2.0 flow may contain several WAP 1.2 flows. The fraction of WAP flows in GPRS would change to 52% (T_{IAT} =15 sec), and, accordingly, the fraction of HTTP flows would increase to 30% respectively[4]. Hence, even applying the WAP 2.0 flow definition, WAP still remains the dominant flow type in the future.

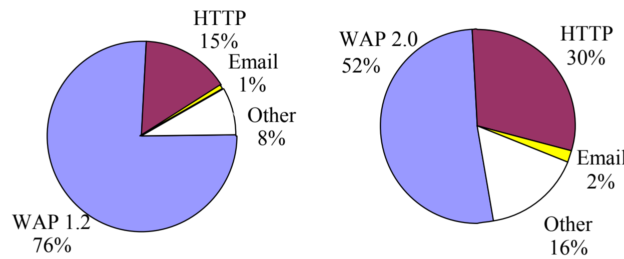

Figure 3: Application flows in GPRS trace

V. Modeling Approach

In this section we present the modeling methods we deploy in our analysis.

Fitting an analytical distribution to empirical data consists of three parts. As prerequisite, we need to assure the *appropriateness* of the sampled data. The data needs to be independent (random), stationary and not showing trends. Next, in order to find the best fitting distribution the following steps need to be done for each potential distribution. The first step is to assume a distribution best describing the random variable from which the samples are and to *estimate the parameters* of this distribution. In the second step the *goodness of the fit*, that is how well the distribution resembles the empirical distribution, is tested. Having applied those steps to all potential distributions, in the end the distribution with the best goodness of fit result is chosen.

In the following paragraph we briefly line out the appropriateness tests, the MLE method for estimating the parameters of a general distribution, and the EM for PH algorithm, which derives the best fitting phase type distribution for a given number of phases. Also, we present the KS test as goodness of fit test. In the subsequent explanation denotes X=(x_1,..,x_N) the set of data samples of size N.

A. Appropriateness tests for data sets

1) Independence

A mathematical notion indicating independence is [17]:

$$\left|ACF(l)\right| \leq \frac{\Phi^{-1}(1-\frac{\alpha}{2})}{\sqrt{N}}, \quad \text{for } l \geq 1, \tag{1}$$

where ACF(l) *is the autocorrelation function*, Φ^{-1} is the percent point function of the standard normal distribution and

[4] Note, the numbers are truncated to the next lower integer value.

α is the level of significance to assume independence. However, for real data the *ACF* stays seldom below the threshold. Therefore in practical circumstances often a visual inspection is done based on the formula.

2) Stationarity

Strict stationarity[5] is also often difficult to prove. Therefore we will only consider stationarity of the first moment (mean). Stationarity is determined by visual inspection of the plots of the cumulative sample mean. The cumulative sample mean is defined as:

$$\overline{X}_{cum}(s) = \frac{1}{s}\sum_{j=1}^{s} x_j \, . \tag{2}$$

For stationarity the cumulative sample mean should converge to a stable value with increased sample size.

Another appropriate visual inspection method is the moving sample mean with a sliding window of size l, over which the average is taken. The definition of the moving sample mean is:

$$\overline{X}^{l}_{mov}(k) = \frac{1}{l}\sum_{j=0}^{l-1} x_{k+j} \, . \tag{3}$$

Plotting the moving sample mean can help to detect short term and long-term trends.

3) Periodicity

A further property for which the sample can be checked is periodicity. The data samples should not show periodicity. Randomness exhibits white noise, that is, all frequencies are equally present in the 'signal'. We can check the periodicity with the help of the discrete Fourier transformation, which is defined as:

$$DFT(k) = \sum_{j=0}^{N-1} x_j \left(e^{\frac{-2\pi i}{N}} \right)^{kj} \tag{4}$$

In case of periodicity strong spikes are visible in the DFT. Therefore, we use visual inspection of the DFT to check for periodicity in the time series.

It is possible to formalize all test cases with stricter boundaries instead of applying visual inspection. However, as mentioned, this leads to a high rejection rate for real world data. We believe that in our circumstances visual inspection, when applied with care, is not less accurate.

B. Maximum Likelihood Estimation

MLE is a method to estimate parameters of a distribution based on sample data. MLE optimizes the probability that a particular data set yields from a chosen probability model. The MLE method works as follows. Let $f(x; p_1,..., p_q)$ be the assumed density function and let $F(x; p_1,..., p_q)$ be the corresponding distribution function, with $P = (p_1,.., p_q)$ the set of q parameters to be estimated. The log-likelihood function L is then defined as

$$L = \prod_{j=1}^{N} f(x_j; p_1,..., p_q) \, . \tag{5}$$

The maximum likelihood estimator $\hat{P}$ of the parameter vector is calculated by maximizing the log-likelihood function. That is, solving

$$\frac{\partial \log L}{\partial \hat{p}_k} = 0, \quad k = 1,..., q \tag{6}$$

yields the MLE.

C. EM for phase type distributions

An alternative method especially suited to fit hyper-exponential[6] distributions to heavy-tailed distributions has recently been proposed in [15]. This method is based on the EM-algorithm and uses only the empirical data. The method is iterative with a complexity of O(*NI*) for each iteration with N the number of measurements and I the number of phases. Note that a hyper-exponential distribution is fully characterised by the number of phases (here denoted as I), the rates λ_i per phase, as well as the probabilities c_i for each phase.

Let I be the number of phases for the hyper-exponential distribution, and $x_1,\ldots,x_N$ the N independent observations. Select initial values for c_i and λ_i (i=1, …, I). Set ε to be the required precision, then the algorithm proceeds as follows. Let

$$p(x_n \mid \lambda_i) = \lambda_i e^{-\lambda_i x_n} \tag{7}$$

$$p(x_n \mid (c_i, \lambda_i)) = \sum_{i=1}^{I} c_i \, p(x_n \mid \lambda_i) \tag{8}$$

Compute in each iteration:

$$p(i \mid x_n, \lambda_i) = \frac{c_i \, p(x_n \mid \lambda_i)}{p(x_n \mid (c_i, \lambda_i))} \tag{9}$$

$$c_i' = \frac{1}{N}\sum_{n=1}^{N} p(i \mid x_n, \lambda_i) \tag{10}$$

$$\lambda_i' = \frac{\sum_{n=1}^{N} p(i \mid x_n, \lambda_i)}{\sum_{n=1}^{N} p(i \mid x_n, \lambda_i) x_n} \tag{11}$$

The iteration stops when the $| c_i' - c_i | < \varepsilon$ and $| \lambda_i' - \lambda_i | < \varepsilon$. Before starting the next iteration set c_i to c_i' and λ_i to λ_i'. The hyper-exponential distribution is fully specified by the final set of c_i and λ_i values.

D. Kolmogorov-Smirnov Test

The KS test is a commonly used method to assess how well a chosen distribution fits the sample dataset.

The KS test considers the difference between the empirical CDF and the assumedly 'true' distribution function.

Starting with the assumption that the sample data results from the distribution F with the density function $f(x; p_1,..., p_q)$, having parameters p_1 to p_q. $F_e(x)$ is the empirical CDF of the measured data.

The difference measure used by the KS test is:

[5] A random process X_n is *strict stationary* if all statistical properties (that is, all statistical moments) do not change over time.

[6] Hyper exponential is a special phase type distribution.

$$
\begin{aligned}
D &= \sup_x \left| F_e(x) - F(x) \right| \\
&= \max\left(\sup_x \left(F_e(x) - F(x) \right), \sup_x \left(F(x) - F_e(x) \right) \right)
\end{aligned} \tag{12}
$$

D is called the KS statistic. For each true distribution a critical value can be specified to either reject or accept an assumption that the data set is from the chosen distribution. The critical value must be defined for each distribution and depends on the significance level α. The significance level specifies the reliability of the result. Giving the hypotheses H_0: 'The data is from a specific distribution', the significance level α specifies how likely it is that we reject H_0 even if it is true. In practice, often a significance level of 5% is used.

However, empirical data rarely stems from a single assumed theoretical distribution; therefore D is seldom below the critical value for real measurement data. In practice the KS-statistic is used to compare different 'goodness-of-fits'; the distribution with the lowest KS-statistic is concerned the best fitting one.

VI. Application Modeling

In this section we present the results of our flow length analysis. First we checked the appropriateness of our data sets. Next in our analysis we derived parameter sets for various distributions and compare the results to fitted PH-type distributions.

A. Appropriateness test

We discuss here only our inspection for the HTTP flow length sample trace. We tested in the same way the WAP flows, which provided similar results. The discussed plot for HTTP is depicted in Fig. 4. Plot (a) provides a visualization of the time series of the HTTP flow length. It shows a high variance of the data values. Plot (b) shows the *ACF* for lag 0 to 1000. The *ACF* (1) drops immediately to below 0.01, and stays at this level. This indicates highly uncorrelated values and is below the 5% significance level for this data set. Plot (c) shows the results from the Discrete Fourier Transformation (DFT). No distinct dominating frequency parts are visible. Plot (d) depicts the moving average for a window length l=100 and l=1000. At this level still quite some fluctuation is visible, which could be due to a heavy-tailedness of the underlying distribution. However, no clear singled-out trends or level shifts are visible, which is the important conclusion. Finally plot (e) depicts the cumulative mean over the whole data set. After a short transient period in the beginning, the mean quickly stabilizes and shows no visual trends.

Based on the visual inspection, we can conclude that independence and stationarity are a valid assumption. Therefore, all of the data sets are appropriate for our modeling approach.

B. Analysis

In this section we investigate which analytical distribution best fits the empirical CDF for the flow length of HTTP, WAP 1.2 and WAP 2.0. A single analytical distribution is better tractable in further analysis. Therefore we used the MLE method to derive the parameters of a single distribution. But often the empirical distribution does not belong to a single distribution class. Therefore, additionally, we use the EM algorithm. This method allows a close fitting on the empirical distribution, while still yielding a tractable description.

Table 1 summarizes the results from testing different distributions. We show for the different data sets the mean, median and coefficient of variation (CV) statistics, for the empirical as well as the fitted distributions. Table 2 lists the corresponding parameters for the fitted distributions. Fig. 5 to Fig. 10 depict the complementary CDF of the empirical data as well as the relevant analytical function according to the parameters in Table 2.

We used the MLE and KS test methods implemented in the DATAPLOT tool [19] for deriving the analytical distribution. We tested each empirical distribution against the normal, exponential, gamma, logistic, extreme value, Weibull, lognormal, and Pareto distribution. In all cases some distribution from the class of heavy-tailed distributions (Pareto, lognormal, Weibull) scored best. However, none passed the KS test. That is, the KS-statistics is never below the critical value defined for the KS test. Therefore we list the top three ranked distributions in the MLE section in Table 1. In the following we discuss the results for HTTP and WAP more in detail.

For HTTP we took a similar approach as in [18], that is, we have split the empirical CDF in a body and tail part. We investigated the case when the HTTP tail starts at 10^4 bytes and at 10^6 bytes.

If we assume 10^4 bytes as splitting point of the HTTP flow length distribution, the body and the tail is well fitted by a lognormal distribution. If the split point is moved to 10^6 bytes a Pareto distribution is the best choice for the tail. Fig. 5 plots the corresponding empirical and analytical curves. As can be seen from the curves the good lognormal KS test results stem from the upper part of the tail between 10^4 and 10^6 bytes, while the Pareto distribution has a parallel slope at the extreme tail part. This suggests that a Pareto distribution is a good fit for the tail. Therefore, we suggest that HTTP flows for GPRS can be best modeled with a lognormal distribution for the body and a Pareto distribution for the tail. This is in accordance with suggestions for wireline HTTP traffic as presented in [18].

The results for the phase-type distribution for the entire empirical distribution for HTTP are also listed in Table 1 and Table 2. The mean, median and CV statistics, as well as Fig. 6, suggests that 2 phases are not enough, but 4 phases seem to match well for the considered part of the distribution.

In the case of WAP 1.2 again the lognormal distribution is the best-fitting distribution (Table 1). However, the shape parameter of 0.97 indicates 'very little' heavy-tailedness. Also note that the next two best-fitting distributions gamma and exponential are not heavy-tailed. The KS statistic for the exponential distribution was also only 0.1111647 and therefore close to the best KS-statistic of the lognormal distribution. Therefore, we assume no heavy-tailedness for

WAP 1.2 flows. This corresponds also to the fact that the results for the phase type distributions suggest a very good fit for 2 phases and 4 phases.

The fitting results for WAP 2.0 (15 sec) also indicate a lognormal distribution, though with a larger shape parameter, indicating 'stronger' heavy-tailedness. Additionally, the Weibull distribution has a shape parameter less than one indicating heavy-tailedness. However, the Pareto distribution did not score for the first 4 ranks. As Fig. 9 shows a very good fit for the lognormal distribution, we assume that WAP 2.0 should be modeled with a lognormal distribution. Finally, Fig. 10 depicts the results for the 2 and 4 phase type distributions. Again 2 and 4 phases appear to yield a good approximation, allowing also a practical approximation via phase type distributions.

VII. CONCLUSION

Based on a specific method to measure application flows, we have derived distribution functions of flow length. Those results can be deployed in analysis or simulations. We base our results on a comprehensive large measurement base, making our results particular useful. An important contribution is furthermore that we extend our flow definition such that WAP 2.0 can be also modeled. This is done by extrapolating from the measured WAP 1.2 flows to WAP 2.0 flows.

Our main finding is that the majority of flows in GPRS, which can be accounted to WAP flows, is very short, and is showing only weak signs of heavy-tailedness. We found that although WAP 1.2 flows are best modeled by a lognormal distribution, an exponential distribution also comes surprisingly close to the empirical distribution. On the other hand, the length of WAP 2.0 flows is best modeled by a lognormal distribution, indicating some heavy-tailedness by its shape parameter.

Also PH-type distributions appear to be good candidates for modeling WAP traffic. This is especially of interest as PH-type distributions are much better to handle in analytical studies. Therefore we provided PH-type distributions for the investigated application flows. In particular our results suggest that a parsimonious PH-type model with 2 or 4 phases is feasible for WAP flows.

HTTP flows, on the other hand, have strong signs of heavy-tailedness also in GPRS. We found the same model as suggested by [18] in that HTTP flows are best split into a body, modeled by a lognormal distribution, and a tail modeled by a Pareto distribution. Furthermore, phase type distributions with at least 4 phases provide a good approximation of the distribution in the observed data range.

Our previous results [1][2][3] as well as results in this paper show that there are specific characteristics of the traffic in mobile networks that can be attributed to new applications and to the specific mobile environment in which they are used. This should be considered when building networks and designing protocols. However, the trend in mobile networks is currently in the direction of high bandwidth access (e.g. UMTS, HSDPA, HSUPA, etc.), resembling fixed network characteristics. Together with the introduction of flat tariff rates this will change the traffic characteristics again. Still, there is the small form factor of the mobile devices and the mobile situation in which the user uses the devices, which might make the traffic always specific. Those influences and upcoming new applications should therefore be monitored and modeled in a continuous effort.

A. References

[1] R. Kalden, T. Varga, B. Wouters, and B. Sanders. Wireless service usage and traffic characteristics in GPRS networks. In Proceedings of the 18th International Teletraffic Congress (ITC18), vol. 2, pp. 981–990, 2003.

[2] R. Kalden and H. Ekstrom. Searching for mobile mice and elephants in GPRS networks. ACM SIGMOBILE, Mobile Computing and Communications Review, 8(4), pp. 37-46, 2004.

[3] R. Kalden. Mobile Internet Traffic Measurement and Modeling based on Data from commercial GPRS Networks, Dissertation, University of Twente, 2004.

[4] URL: http://www.wapforum.org/

[5] WAP-224-WTP-"Wireless Transaction Protocol" URL: http://www.wapforum.org/

[6] WAP-225-TCP- "Wireless Profiled TCP" URL: http://www.wapforum.org/

[7] WAP-229-HTTP-" Wireless Profiled HTTP" URL: http://www.wapforum.org/

[8] WAP-230-WSP "Wireless Session Protocol Specification" URL: http://www.wapforum.org/

[9] WAP 2.0 "Technical White Paper" URL: http://www.wapforum.org/

[10] 3GPP TS 03.60 General Packet Radio Service (GPRS); Service description; Stage 2 (Release 1998).

[11] R. Fielding, J. Mogul, H. Frystyk, et al., Hypertext Transfer -- Protocol HTTP/1.1. RFC 2616, June 1999.

[12] J. C. Mogul, The Case for Persistent-Connection HTTP. DIGITIAL Western Research Laboratory, Research Report 95/4

[13] J. Aldrich, R.A. Fisher and the making of maximum likelihood 1912-1922, Statistical Science, 12(3), pp. 162–176, 1997

[14] I. M. Chakravart, R. G. Laha, and J. Roy, Handbook of Methods of Applied Statistics, Volume I. John Wiley, 1967, pp. 392-394

[15] R. El Abdouni Khayari, R. Sadre, and B.R. Haverkort. Fitting world-wide web request traces with the EM-algorithm. Performance Evaluation, 52:175–191, 2003.

[16] K. Thompson, G. Miller, and R. Wilder. Wide-area internet traffic patterns and characteristics (extended version). IEEE Network, 11(6):10–23, 1997.

[17] V. Paxson and S. Floyd. Wide-area traffic: the failure of poisson modeling. In Proceedings of the ACM conference on communications architectures, protocols and applications, pp. 257–268, 1994.

[18] M. E. Crovella, M. S. Taqqu and A. Bestavros. Heavy-Tailed Probability Distributions in the World Wide Web. In A Practical Guide To Heavy Tails, pp. 3–26, Chapman & Hall, 1998.

[19] DATAPLOT - Tool for Exploratory Data Analysis URL: http://www.itl.nist.gov/div898/software/dataplot/homepage.htm

VIII. APPENDIX

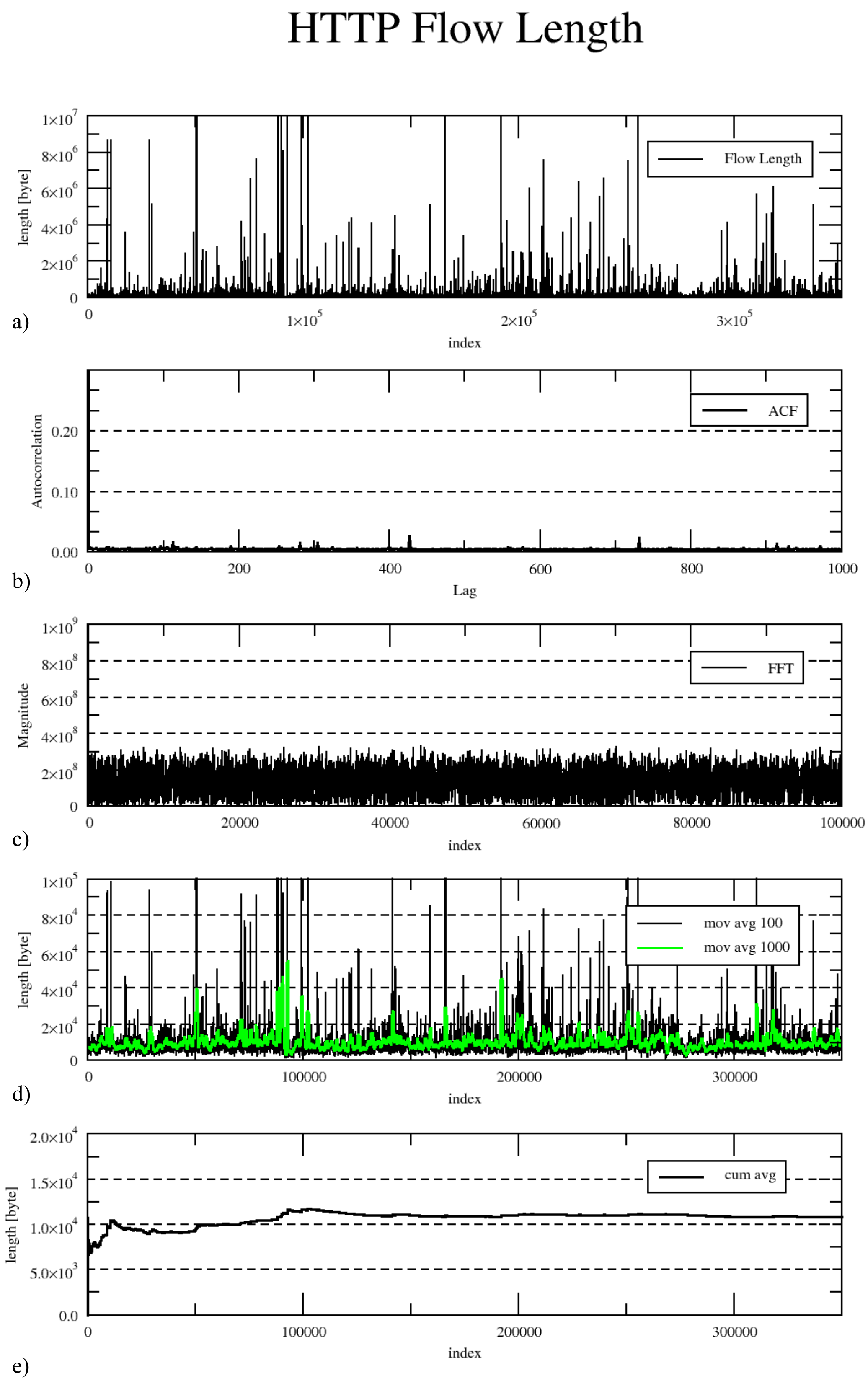

Figure 4: HTTP flow length – appropriateness tests

Table 2: Flow length – distribution fitting results

Flow length		Distribution	Mean	Median	CV	KS-statistic
HTTP body						
Empirical			2067.647	1133	0.991764	
MLE	1st	lognormal	2108.478	1364.436	1.178128	0.093395
	2nd	Gamma	2067.647	NA	0.88706	0.123352
	3rd	Weibull	2078.149	581.0078	0.891748	0.124916
HTTP tail (starts at 10^4 byte)						
Empirical			40489.89	14109	8.86301	
MLE	1st	lognormal	36825.57	11881.79	2.933571	0.052849
	2nd	Weibull	33067.29	11218.63	1.578251	0.067133
	3rd	Gamma	40489.89	NA	1.401365	0.137995
	4th	Pareto	73051.7	12818.46	NA	0.14798
(tail starts at 10^6 byte)		Pareto	390141	123855.3	NA	0.0473
HTTP Total						
Empirical			10821.22	1757	14.20517	
EM		PH 2-phase	9473.915	1912.137	2.439723	
		PH 4-phase	10821.22	1865.381	8.261301	
WAP 1.2						
Empirical			1263.716	854	2.599148	
MLE	1st	Lognormal	1200.212	748.5217	1.253407	0.090788
	2nd	Gamma	1263.716	444.4164	0.951797	0.104647
	3rd	Exponential	1327.716	NA	0.982386	0.111165
EM		PH 2-phase	1263.716	769.9679	2.320188	
		PH 4-phase	1263.716	764.1593	2.838391	
WAP 2.0						
Empirical			3699.885	1888	2.260354	
MLE	1st	Lognormal	3651.368	1739.848	1.845107	0.066968
	2nd	Weibull	3574.891	1182.613	1.24251	0.078783
	3rd	Gamma	3699.885	NA	1.010861	0.123792
EM		PH 2-phase	3699.885	1856.392	1.900573	
		PH 4-phase	3699.885	1843.729	2.513669	

Table 3: Flow length – distribution parameters

Flow length		Distribution	Parameter			
HTTP body						
MLE	1st	Lognormal	Shape	0.932979	Scale	748.5217
	2nd	Gamma	Shape	1.270849	Scale	1219.585
	3rd	Weibull	Shape	1.123434	Scale	2168.302
HTTP tail						
MLE	1st	Lognormal	Shape	1.504118	Scale	11881.79
	2nd	Weibull	Shape	0.655713	Scale	24436.78
	3rd	Gamma	Shape	0.509211	Scale	79515.02
	4th	Pareto	Shape	1.1586	Location	10000
(tail starts at 10^6 byte)		Pareto	Shape	1.34466	Location	100000
HTTP Total						
EM	Total	PH 2-phase	λ_0	5.42E-04	c_0	7.53E-01
			λ_1	3.06E-05	c_1	2.47E-01
	Total	PH 4-phase	λ_0	0.000604	c_0	0.696531
			λ_1	6.14E-05	c_1	0.260152
			λ_2	1.51E-05	c_2	0.041389
			λ_3	7.18E-07	c_3	0.001929
WAP 1.2						
MLE	1st	Lognormal	Shape	0.971754	Scale	748.5217
	2nd	Gamma	Shape	1.036182	Scale	1219.585
	3rd	Exponential	Scale	1263.716	Location	64
EM	Total	PH 2-phase	λ_0	0.000913	c_0	0.989655
			λ_1	5.77E-05	c_1	0.010345
	Total	PH 4-phase	λ_0	0.000937	c_0	0.964116
			λ_1	0.000307	c_1	0.032526
			λ_2	2.6E-05	c_2	0.003305
			λ_3	2.36E-05	c_3	5.28E-05
WAP 2.0						
MLE	1st	Lognormal	Shape	1.217624	Scale	1739.848
	2nd	Weibull	Shape	0.810749	Scale	3185.072
	3rd	Gamma	Shape	0.978627	Scale	3780.69
EM	Total	PH 2-phase	λ_0	0.000456	c_0	0.829553
			λ_1	9.07E-05	c_1	0.170447
	Total	PH 4-phase	λ_0	0.000468	c_0	0.779992
			λ_1	0.000144	c_1	0.20788
			λ_2	2.1E-05	c_2	0.01196
			λ_3	8.06E-06	c_3	0.000168

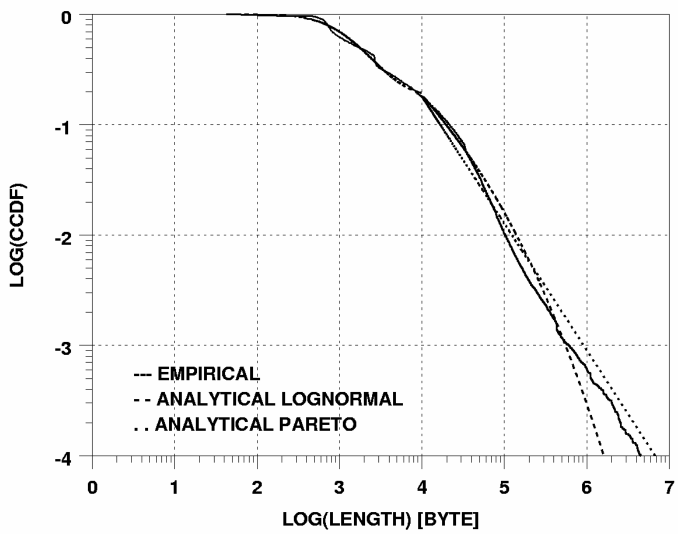

Figure 5: HTTP flows – lognormal and Pareto (partial)

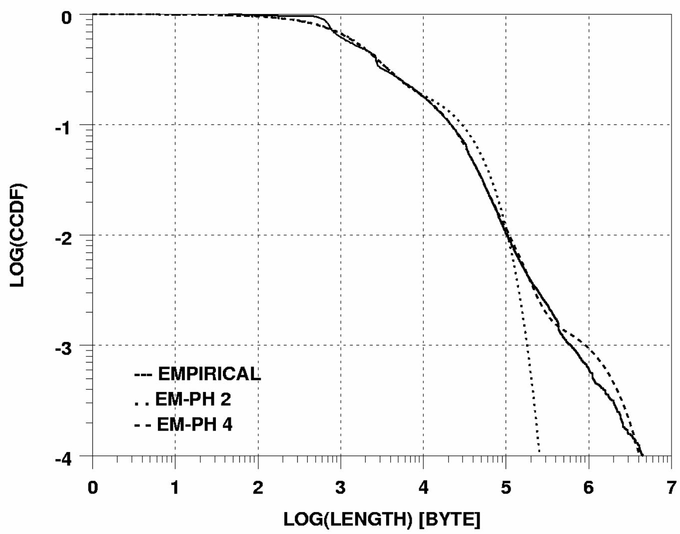

Figure 6: HTTP flows – EM - PH

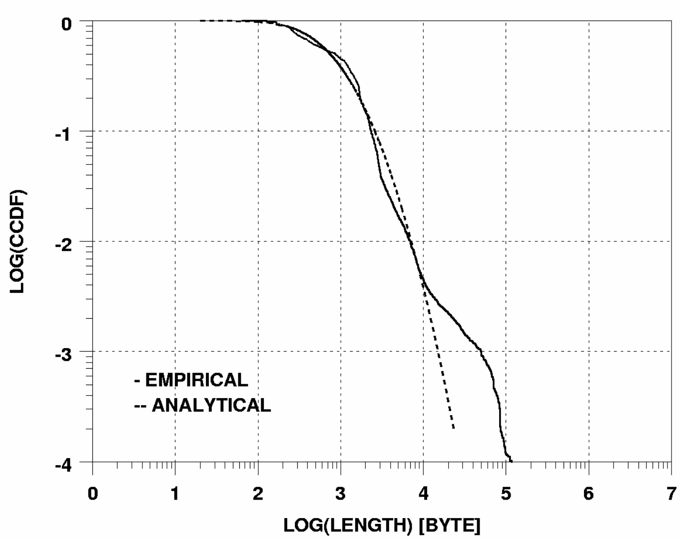

Figure 7: WAP 1.2 flows – lognormal

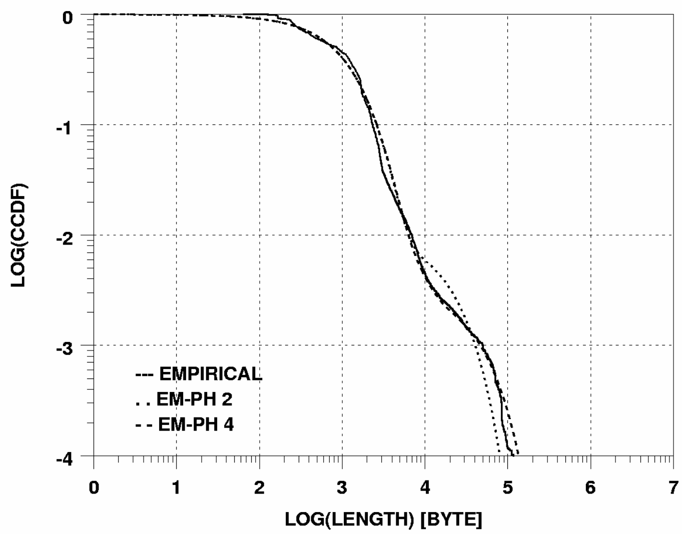

Figure 8: WAP 1.2 flows – EM - PH

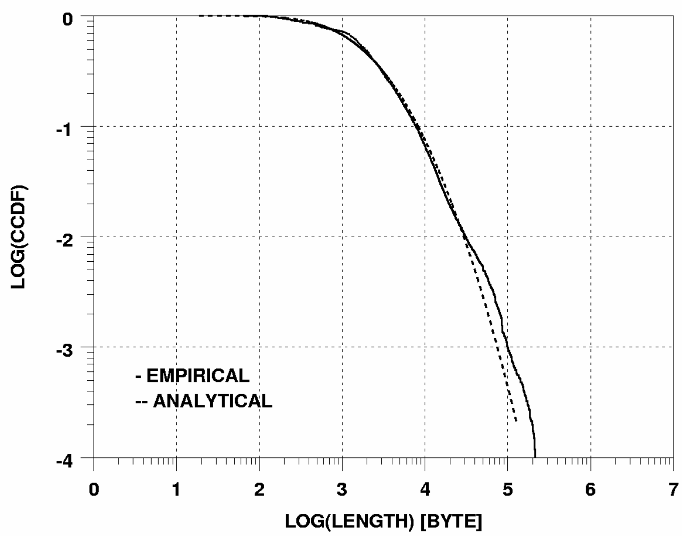

Figure 9: WAP 2.0 (15sec) flows – lognormal

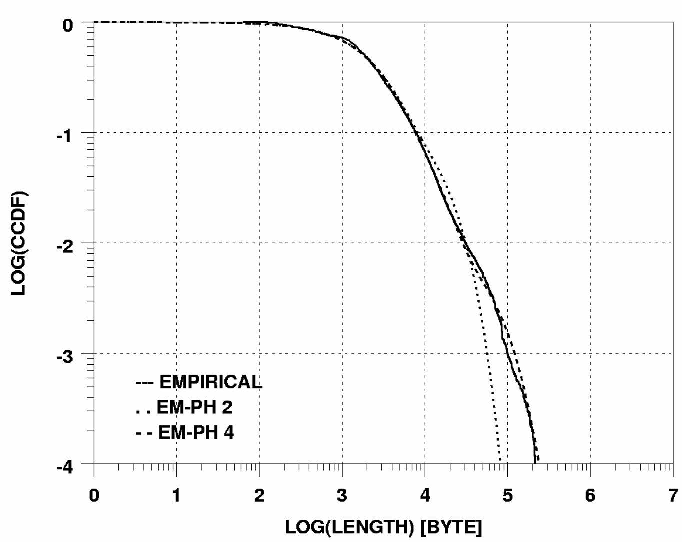

Figure 10: WAP 2.0 (15sec) flows – EM - PH