WORKLOAD CHARACTERIZATION OF MAIL SERVERS*

Laura Bertolotti – Maria Carla Calzarossa
Dipartimento di Informatica e Sistemistica
Università di Pavia
I-27100 Pavia, Italy
E-mail: {bertolot, mcc}@alice.unipv.it
http://mafalda.unipv.it

Keywords: mail server, workload characterization, performance evaluation, benchmarking.

ABSTRACT

Mail servers have to cope with the explosive growth of Internet users by providing an acceptable quality of service. Performance evaluation and capacity planning of mail servers are crucial issues to be addressed. Workload characterization represents the basis of all these studies. This paper deals with a detailed characterization of the workload of mail servers. Our approach is experimental, that is, based on the analysis of measurements collected on various mail servers. By applying statistical and numerical techniques, we obtain models able to capture and reproduce the behavior of mail server workload. Some of the results of our studies will be used for the definition of the workload of SPECmail2000, a SPEC standardized benchmark aimed at testing the ability of a system to act as a mail server.

1 INTRODUCTION

Electronic mail, one of the oldest Internet applications, is becoming more and more popular for business and personal communications. This is due to the increased number of Internet users who benefit of the enhanced features associated to email services where messages are no longer limited to text files, but can embed sounds, graphics, and video files, as well as Java applets and desktop applications. Devices, such as, smart phones, let the users easily connect to the Internet remotely. Moreover, many applications, such as e-commerce applications, rely on email to deliver documents, statements and bills.

In this scenario, mail servers play a fundamental role in that they have to provide a fast, highly available, reliable and secure service by delivering incoming and outgoing email messages and by allowing users to access their mailboxes. To meet these requirements and to guarantee an acceptable quality of service by keeping up with the explosive growth of number of email users, performance evaluation and capacity planning of mail servers are crucial issues to be addressed. The basis of all these studies is an accurate characterization of their workload. Despite of its importance, this topic has received little attention.

Workload characterization can be addressed from different perspectives [CS93], [CMT00]. Many papers (see e.g., [LTWW94], [Pax94], [RVH94], [RJH95], [CB96]) deal with workload characterization of client-server environments, computer networks and web servers. In [AW97], ten invariants, which describe the workload of web servers, are identified. In [BC98], the essential elements required in the generation of a representative Web workload are presented. Recent papers (see e.g., [MAFM99]) deal with workload characterization of e-commerce sites. The workload models are based on Customer Behavior Model Graphs, which capture and reproduce the navigational patterns of the customers.

The purpose of this paper is to present a detailed study of mail server workload. We will follow an experimental approach based on the analysis of measurements collected on various mail servers. The workload considered in our study consists of the requests associated to various protocols, namely, the open standard Internet protocols SMTP [RFCa] and POP3 [RFCb]. We propose executable models able to characterize and to reproduce the behavior of the mail server workloads. The results of our studies will be part of SPECmail2000, a standardized benchmark currently under development within SPEC (Standard Performance Evaluation Corporation) [SPE], whose goal is to measure a system's ability to act as a mail server.

The paper is organized as follows. Section 2 describes the characteristics of mail server workloads and the data used in our study. The methodology adopted for the analysis of the workload associated to SMTP and POP protocols is presented in Section 3. The models obtained for each workload type are described in Sections 3.1 and 3.2. Future directions in the field of workload characterization of mail servers and Internet applications are outlined in Section 4.

*This work was supported in part by the University of Pavia under the grant FAR–99

This paper will be published in the Proceedings of SPECTS'2000, July 16–20, 2000, Vancouver, Canada.

2 DATA COLLECTION

The workload of a mail server is the superposition of requests generated by the interactions between different mail servers as well as by client–server interactions. A mail server has to process requests coming from or addressed to other mail servers to deliver incoming and outgoing messages. These requests are typically managed by application software, e.g., sendmail, which implements the SMTP protocol. A mail server has also to process requests issued by the clients, i.e., users, who are willing to access their mailboxes and read their messages. Various protocols, e.g., POP3, IMAP4 [RFCc], implemented both at the server side and at the client side, work for this purpose.

To characterize the behavior of this composite workload, we analyze the various types of requests processed by mail servers. In what follows, we focus on the requests associated to SMTP and POP3 protocols and we refer to these requests as SMTP and POP workload, respectively. As already pointed out, our approach is experimental, that is, based on the analysis of measurements collected on mail servers. The measurements that can be collected on a mail server depend on the software running on the server, on the logging options selected and on the request type, e.g., send, connect, retrieve, processed by the server. For each request, the log files collected on the server contain various information. Examples of information recorded for an SMTP request are its time stamp, the size of the message to be sent or being received, the IP address of the sender and of the recipient(s) of the message, the status of the request. In the case of POP requests, the log files contain information such as the identifier of the user who issued the request, a time stamp, the size of the retrieved message.

The measurements considered in our study refer to the workload of the mail servers of an ISP, two enterprises and the University of Pavia. To sanitize the log files, we provided each installation with a set of Perl scripts (see [SPE]) which remove all the confidential information these files might contain. These scripts also extract the information that will be used as characterizing parameters of the mail server workload.
Table 1 presents a summary of the raw data considered in our study. Because of non–disclosure agreements, we will not be able to name the enterprises and the ISP.
Each log file records measurements collected over a day, that is, 24 hours. POP log files have been collected only on the University of Pavia and Enterprise B mail servers. Indeed, collecting measurements, which is always a challenging issue, becomes even more challenging in the case of POP workload. Many installations do not (and are not willing to) log any information about their POP workload because of the overhead introduced on their mail servers. The table lists, whenever available, the number of users of the various mail servers. This number provides a rough characterization of the mail server in terms of its potential load.
For the characterization of the workloads of these mail servers, the measurements of each day will be analyzed separately.

3 WORKLOAD MODELS

To analyze the measurements collected on the various mail servers for SMTP and POP requests, we have applied a typical workload characterization methodology consisting of a number of steps dealing with the formulation of the models, an exploratory analysis of the workload parameters and the application of various types of statistical and numerical techniques.

A key step in the formulation of the models deals with the choice of the parameters to be used to describe the requests processed by the mail server. Because of the objective of our study, that is, to build workload models for benchmarking, the parameters deal with the workload intensity and the functional characteristics of the requests, whereas we did not consider any parameter dealing with the consumptions of the requests on the mail server resources. Indeed, despite of workload models used to describe the input of analytic or simulation system models, executable workload models do not require these kinds of parameters. Moreover, to keep the parameters describing each workload type homogeneous across the various mail servers, we identified a set of parameters common to all the collected log files.

The exploratory analysis, based on descriptive statistics (i.e., basic statistics and distributions) of the parameters, provides preliminary insights into the behavior and the characteristics of the workload. Techniques based on clustering [Har75], numerical fitting [LH74], and probabilistic graphs [Fer84] are applied to build the workload models. As a result of the application of this methodology, we obtain models which are representative of the static and dynamic behavior of the mail server workload.

3.1 STMP Workload Models

In this section, we will focus on the characterization of the SMTP workload, that is, the requests processed by the mail server to send or receive messages. As previously stated, the SMTP requests have to be described in terms of their intensity, i.e., their arrival process, and of their functional characteristics. Hence, the parameters chosen to describe each request are its time stamp together with the size and the number of recipients of the message to be sent or being received. Moreover, from the time stamps, we computed the interarrival times, that is, the time between the arrivals of two consecutive requests.

To study the behavior of SMTP requests, we have initially analyzed their intensity. Figure 1 shows the arrival process of the requests, i.e., the number of arrivals per minute, as measured on the ISP mail server for one of the days considered in our study. The number of requests arrived at the mail server is equal to 23,880, with an average arrival rate of 16.58 requests/minute. As can be seen, this rate exhibits large fluctuations over the 24 hours. The load is light in the early hours of the day, then it gradually increases. Moreover, we have noticed that the load is bursty, that is, it exhibits a self–similar behavior (see e.g., [LTWW94]). The value of the Hurst parameter estimated by applying various tests, such as, the time–variance plot,

Source	Number of days	Number of users	Total SMTP requests	Total POP requests
Enterprise A	2	unknown	35,840	–
Enterprise B	1	unknown	20,790	89,546
ISP	11	8,057	214,225	–
University	11	1,058	138,117	243,684

Table 1: Summary of the collected log files.

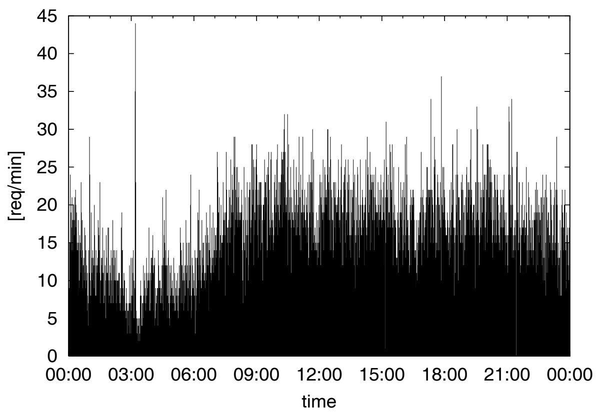

Figure 1: Arrival rate of the SMTP requests as measured over a time period of 24 hours.

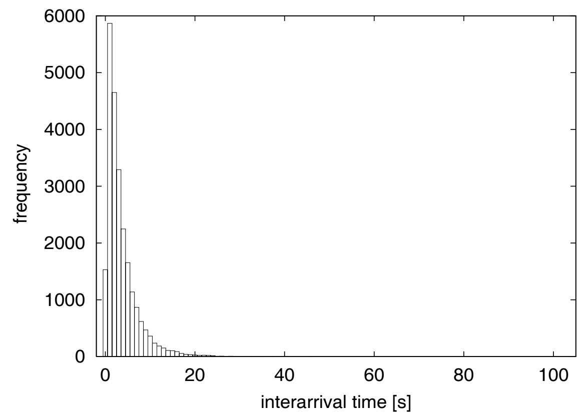

Figure 2: Interarrival time distribution.

is equal to 0.9186. Values of this parameter close to one denote a high degree of self–similarity. On the contrary, the number of bytes exchanged over time is not characterized by a bursty nature, as denoted by the value of the Hurst parameter, which is equal to 0.573.
These results provide preliminary insights in the behavior and the nature of the workload and will drive the choice of its distributional models.

Note that, unless otherwise stated, all the results presented in this section, will refer to the requests processed during one day only.
Table 2 presents the basic statistics of parameters used in our study. The parameters are highly dispersed around their mean, as denoted by the large values of their standard deviation. The interarrival times are characterized by a standard deviation which is approximately equal to the mean, with the corresponding coefficient of variation slightly larger than one. Figure 2 shows the interarrival time distribution. The distribution is highly positively skewed. About 5% of the requests are characterized by interarrival times in the range between 10 and 105 seconds.
Figure 3 shows the message size the distribution. As in the case of interarrival times, the distribution is highly positively skewed. Its 90–th percentile corresponds to 11 kbytes, and its 99–th percentile is 67 kbytes. Moreover, the distribution is characterized by a standard deviation which is more than six times larger than the corresponding mean. About 99.35% of the messages are characterized by a small size, namely, less than 100 kbytes.
Figure 4 shows the distribution of the number of recipients of the messages. The distribution is centered on one, which corresponds to the mode value of the distribution. Approximately 94% of the messages have one recipient. Moreover, the 99.8% of the messages have less than 20 recipients. There are also messages which have as many as 406 recipients. The standard deviation of this parameter is approximately three times larger than the corresponding mean.

To further investigate the behavior of the workload and to develop the appropriate distributional models, we applied numerical fitting techniques, based on the least squares method. These techniques identify the parameters of the distribution which best fits the empirical data.
The method used to select the most appropriate analytic distributional model is based on the analysis of the properties of the empirical distribution together with the analysis of an error function associated to the fitting algorithm. For the message size distribution, we have discovered that a log–normal distribution whose analytic expression is given by:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-(\log x - \mu)^2 / 2\sigma^2}$$

is the most appropriate because of most of its mass is concentrated on small values. The shape and location parameters σ and μ identified by fitting techniques are equal to 0.739 and 0.87, respectively.
In the case of the interarrival times, we could not fit the experimental data with one function only. Because of the bursty nature of the arrivals, the interarrival time distribution is characterized by a heavy tail. Hence, we applied right censoring techniques with the objective of finding a thresh-

Parameter	mean	std. dev.	min	max
Interarrival time [s]	3.62	3.84	0	105
Message size [bytes]	7,776.83	52,419.40	176	4,998,263
Number of recipients	1.24	3.87	1	406

Table 2: Basic statistics of the parameters of the SMTP workload.

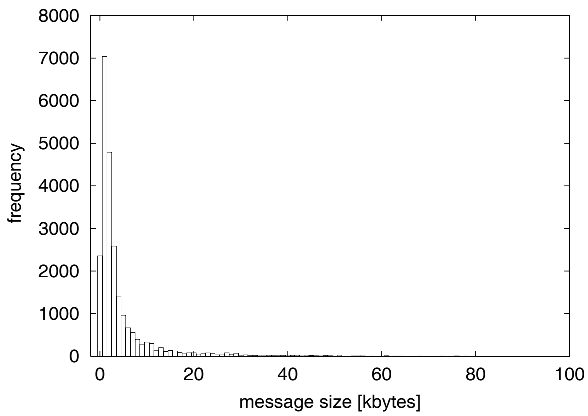

Figure 3: Distribution of the message size.

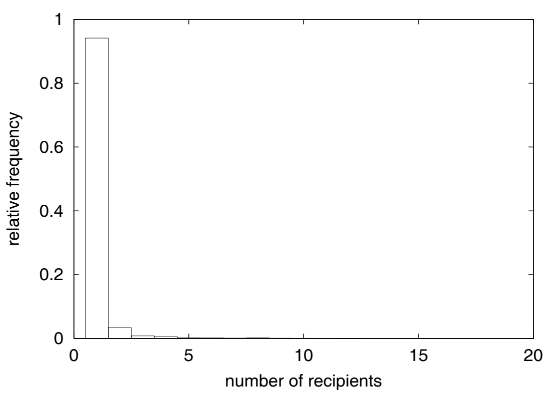

Figure 4: Distribution of the number of message recipients.

old between the body and the heavy tail of the distribution. A Weibull distribution, whose analytic expression is given by:

$$f(x) = \frac{b}{a}\left(\frac{x}{a}\right)^{b-1} e^{-\left(\frac{x}{a}^{b}\right)}$$

with shape parameter a equal to 3.027 and scale parameter b equal to 1.124, best fits the body of the distribution. A Pareto distribution, whose analytic expression is given by:

$$f(x) = \alpha k x^{-(\alpha\ +1)}$$

with shape parameter α equal to 2.459 and location parameter k equal to 3.958, reveals very appropriate for the tail. Figure 5 shows the empirical interarrival time distribution together with the two distributions identified by applying fitting techniques. The dotted curve represents the experimental data, the solid curves represent the models. The dashed vertical line denotes the value of the threshold identified by censoring the empirical distribution. This value is approximately equal to 7 seconds. Hence, the Weibull distribution models interarrival times whose values are smaller than the threshold, whereas the Pareto distribution is used to model interarrival times larger than the threshold.

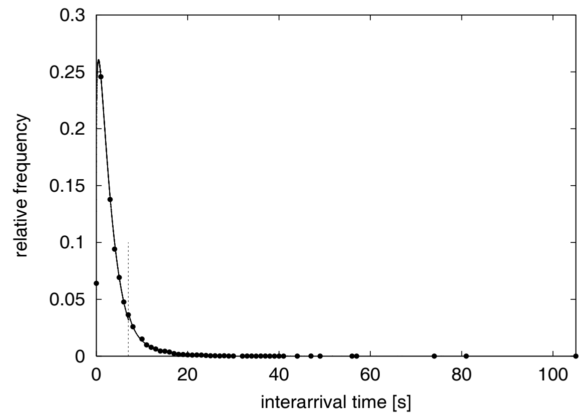

Figure 5: Empirical interarrival time distribution (dotted curve) and corresponding models (solid curves).

To model the distribution of the number of recipients, we did not apply any fitting technique. Instead, we subdivided the distribution into 20 buckets, each characterized by a size and a probability. These buckets, of size one each, represent the distribution with number of recipients from 1 up to 20. We limited our model to 20 recipients because the probability for a message to have more than 20 recipients is negligible, namely, less than 0.002.

By looking at the distributional models identified for the various days considered in our study, we have noticed that the behaviors over different days are very similar. As an example, Figure 6 shows the envelope obtained from all the curves corresponding to the interarrival time distributions of the various days considered in our study. The solid curves in the envelope correspond to the centroids of the models, namely, these curves represent a Weibull and a Pareto distribution whose parameters are obtained by averaging the corresponding parameters of the distributions identified for the various days. As the narrow envelope denotes, the distributions of the various days are very close to each other. Hence, the models identified for one day seem perfectly representative of the load of the other days considered in our study.

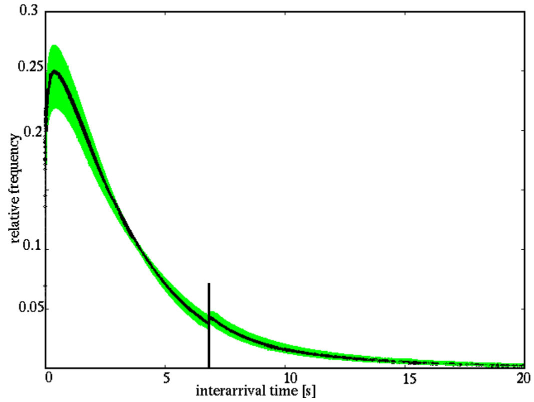

Figure 6: Envelope of the interarrival time distributions obtained for the 25 analyzed log files.

3.2 POP Workload Models

The POP workload is generated by the users who access their mailboxes on the mail server. The options adopted in the configuration of the POP clients determine the operations performed on the mail server. For example, when a user retrieves a message from a server, a simultaneous deletion of the message from the mailbox stored on the server can occur. As an alternative, the messages can be deleted from the server mailbox when they are deleted on the mailbox stored on the client.

Our study focuses on the analysis of the POP sessions with the aim of characterizing these sessions according to their load on the mail server. The characterization of the POP workload is then two-folds. We analyze the mailbox accesses and how the mailboxes vary as a consequence of the user accesses. The parameters collected for each POP session are a time stamp, the user identifier, the number and size of messages deleted from the user mailbox stored on the server and the number and size of the messages stored in the user mailbox on the server.

As already pointed out, the POP log files were collected on two different mail servers for a total of 12 days. From the analysis of the arrival rate of each day, we have noticed that, because of the peculiarity of these installations, most of the load towards the mail servers is issued during the working hours. The servers process approximately 90% of their daily POP load in the time period from 8:00am to 7:30pm. Hence, within each day, we analyzed the POP sessions issued within this time period only.

We initially characterized the POP workload by analyzing the basic statistics and the distributions of parameters, such as the interarrival time between two consecutive sessions and the mailbox size. Table 3 presents the statistics of the POP sessions issued during one of the days considered in our study. The number of requests processed over this time period is equal to 21,116. As can be seen, all the parameters are characterized by a large variability. Each user mailbox contains on the average approximately 14 messages, and there are mailboxes which contain as many as 590 messages. The average size of a mailbox is about 437 kbytes, and there are mailboxes whose size exceeds 12 Mbytes.

As in the case of SMTP workload, we have derived analytic models of the distributions of these parameters. These models describe the overall behavior of the POP workload. For example, a Weibull distribution best fits the interarrival time distribution. Its characterizing parameters a and b are equal to 2.184 and 1.078, respectively.

To obtain a more detailed description of the POP requests, we have analyzed the behavior of users as seen at the server side. This characterization is based on the analysis of what a user does in terms of the mailbox manipulations. For this purpose, we have associated to each session a tuple of two attributes (D, F), which describe the status and the actions performed on the mailbox. In particular, the attribute D ("Delete") denotes whether, within a session, any message was deleted ($D = 1$) or not ($D = 0$). The attribute F ("Full") denotes whether the mailbox is full, that is, it contains messages ($F = 1$), or not ($F = 0$). By using these attributes, each user session can be described by one of the following four states:

- empty session, i.e., empty mailbox (state 00);
- no delete on retrieval, that is, no messages are deleted from a mailbox which contains messages (state 01);
- delete on retrieval, that is, all messages are deleted from the user mailbox stored on the server (state 10);
- retrieve and delete, that is, a few messages are deleted from the user mailbox stored on the server; the mailbox still contains messages (state 11).

To reproduce the dynamic behavior of the users, that is, the sequences of actions performed on their mailboxes stored on the mail server, we have introduced models based on probabilistic graphs. The nodes of these graphs, one for each user, represent the various states as defined above, and the arcs, with their associated probabilities, represent the transitions among these states. A user graph is then described by its states and the corresponding transition probability matrix $P = [p_{s_i,s_j}]$, $(s_i, s_j \in S)$, where $S = \{00, 01, 10, 11\}$ denotes the state space. From the transition probability matrix P, the steady state probabilities, π_{s_i}, $(s_i \in S)$, that is, the limiting state probabilities of being in a given state, are obtained by solving the system of linear equations:

$$\pi_{s_j} = \sum_{s_i \in S} \pi_{s_i} \, p_{s_i,s_j} \qquad s_j \in S$$

with the condition $\sum_{s_j \in S} \pi_{s_j} = 1$.

For each of the 636 users who checked their mailbox more than once, that is, opened more than one session, during the analyzed day, we have built a graph. By analyzing the steady state probabilities for the various graphs, we have noticed that the most popular state is the state 00, which corresponds to an empty session. On the average, the probability for a user of opening an empty POP session is approximately 0.71. This is a consequence of the options set by most users on the client side to periodically check for mail on the server.

Parameter	mean	std. dev.	min	max
Interarrival time [s]	1.96	2.32	0	40
Number of messages per mailbox	13.98	59.04	0	590
Mailbox size [bytes]	437,643.25	1,789,244.25	0	12,569,464
Number of deleted messages	1.71	6.61	0	244
Size of deleted messages [bytes]	54,857.38	314,329.51	0	9,304,370

Table 3: Basic statistics of the parameters describing POP sessions.

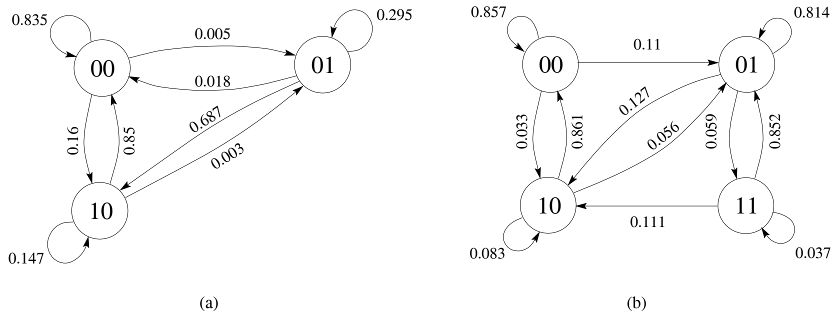

Figure 7: Probabilistic graphs belonging to group 1 (a) and group 2 (b).

To characterize users with similar behavior, we have applied clustering techniques to the graphs built for each user. We used the steady state probabilities π_{s_i} as parameters describing each graph. The k–means algorithm adopted for our analysis classified the graphs into an optimal partition consisting of three groups. Table 4 presents the centroids, i.e., the geometric centers, of the these groups. The

Parameter	group 1	group 2	group 3
π_{00}	0.747	0.070	0.
π_{01}	0.014	0.855	0.361
π_{10}	0.239	0.045	0.111
π_{11}	0.	0.030	0.528

Table 4: Steady state probabilities corresponding to the centroids of the three groups.

first group is a big group containing about 93.87% of the user graphs. The graphs belonging to this group are characterized by a high probability of being in state 00. The graphs of the second group, which are about the 5.66% of the user graphs, are characterized by high probability of being in state 01. Finally, the graphs of the third group, which account for the remaining 0.47%, are characterized by a large value of π_{11}.
Figure 7 shows two representative graphs chosen from group 1 and group 2, respectively.
We have noticed that the user graphs identified by applying clustering techniques for the users of one day, are valid for more than 90% of the users of the mail servers considered in our study.

4 CONCLUSIONS

Performance of mail servers is a challenging issue because of the explosive growth of the number of users accessing email services. To understand and evaluate the performance of any type of system, a detailed analysis of its workload is a fundamental step.

The workload models presented in this paper have been obtained by analyzing a large set of measurements collected on various mail servers. These models are representative of both static and dynamic characteristics of the mail server workloads.
Our workload models will be implemented into SPECmail2000, a SPEC benchmark aimed at testing the ability of a system to act as a mail server. These models will describe the intensity and the characteristics of the requests to be issued against the mail server under test during the benchmark experiment. Because of the confidentiality of the project, we are not able to provide any further details about the benchmark design and specifications and the performance metrics.

Future work will be dedicated to study the impact on the mail server performance of requests using different protocols, e.g. IMAP4. Moreover, we will apply workload characterization techniques to model the new emerging Internet services, such as, e–trade.

Acknowledgements

The authors wish to thank the Computation Center of the University of Pavia, the Enterprises and the ISP for providing the log files of their mail servers. A special thank goes to the SPEC Mailserver sub–committee for the stimulating discussions.

References

[AW97] M.F. Arlitt and C.L. Williamson. Internet Web Servers: Workload Characterization and Performance Implication. *IEEE/ACM Transactions on Networking*, 5(5):631–645, 1997.

[BC98] P. Barford and M.E. Crovella. Generating Representative Web Workloads for Network and Server Performance Evaluation. In *Proc. ACM SIGMETRICS Conf.*, pages 151–160, 1998.

[CB96] M.E. Crovella and A. Bestavros. Self–similarity in World Wide Web Traffic: Evidence and Possible Causes. In *Proc. ACM SIGMETRICS Conf.*, pages 160–169, 1996.

[CMT00] M. Calzarossa, L. Massari, and D. Tessera. Workload Characterization - Issues and Methodologies. In G. Haring, C. Lindemann, and M. Reiser, editors, *Performance Evaluation – Origins and Directions*, volume 1769 of *Lecture Notes in Computer Science*, pages 459–484. Springer–Verlag, 2000.

[CS93] M. Calzarossa and G. Serazzi. Workload Characterization: A Survey. *Proceeding of the IEEE*, 81(8):1136–1150, 1993.

[Fer84] D. Ferrari. On the Foundations of Artificial Workload Design. In *Proc. ACM SIGMETRICS Conf.*, pages 8–14, 1984.

[Har75] J.A. Hartigan. *Clustering Algorithms*. Wiley, 1975.

[LH74] C.L. Lawson and R.J. Hanson. *Solving Least Squares Problems*. Prentice–Hall, 1974.

[LTWW94] W.E. Leland, M.S. Taqqu, W. Willinger, and D.V. Wilson. On the Self–similar Nature of Ethernet Traffic (Extended Version). *IEEE/ACM Trans. on Networking*, 2(1):1–15, 1994.

[MAFM99] D. Menascé, V.A.F. Almeida, R. Fonseca, and M.A. Mendes. A Methodology for Workload Characterization of E–commerce Sites. In *Proc. ACM Electronic Commerce Conf.*, 1999.

[Pax94] V. Paxson. Growth Trends in Wide–area TCP Connections. *IEEE Network*, 8(4):8–17, 1994.

[RFCa] J. Postel. Simple Mail Transfer Protocol. RFC 821, August 1982. http://www.cis.ohio-state.edu/htbin/rfc/rfc821.html.

[RFCb] J. Myers and M. Rose. Post Office Protocol - Version 3. RFC 1939, May 1996. http://www.cis.ohio-state.edu/htbin/rfc/rfc1939.html.

[RFCc] M. Crispin. Internet Message Access Protocol - Version 4. RFC 2060, December 1996. http://www.cis.ohio-state.edu/htbin/rfc/rfc 2060.html.

[RJH95] S.V. Raghavan, P.J. Joseph, and G. Haring. Workload Models for Multiwindow Distributed Environments. In H. Beilner and F. Bause, editors, *Quantitative Evaluation of Computing and Communication Systems*, pages 314–326. Springer, 1995.

[RVH94] S.V. Raghavan, D. Vasukiammaiyar, and G. Haring. Generative Networkload Models for a Single Server Environment. In *Proc. ACM SIGMETRICS Conf.*, pages 118–127, 1994.

[SPE] Standard Performance Evaluation Corporation. http://www.spec.org.