Performance Evaluation of WAP-based Applications over GPRS

Peter Stuckmann, Christian Hoymann
Aachen University of Technology
Chair of Communication Networks
Kopernikusstr. 16, 52074 Aachen, Germany

***Abstract*— The Wireless Application Protocol (WAP) as one application of mobile data services is examined with the aim to provide guidelines for the dimensioning of next generation mobile radio networks. The performance of WAP-based applications over GPRS is evaluated in scenarios where the GPRS radio resources are shared with conventional Internet applications like WWW and e-mail. Simulation results for quality of service measures for the different applications and GPRS system measures are produced with the simulation tool GPRSim that models the application and user behavior, the TCP/IP and WAP protocol architecture, the GPRS protocol architecture and the radio channel. Especially the effect of traffic generated by conventional Internet applications on the WAP performance and vice versa are determined. These results give an estimation of the GPRS network capacity needed for a traffic mix of WAP-based and conventional Internet applications.**

I. Introduction

Driven by the exponential growth of the Internet market the growing demand of data services is also expected for mobile users. Besides conventional Internet applications like WWW and E-Mail, specific service platforms and applications have been developed for future mobile data services like the General Packet Radio Service (GPRS) and the Universal Mobile Telecommunication System (UMTS). Reasons for this are conditions different to fixed networks namely limited bandwidth, higher delays and error rates on the radio interface and mobile terminals with limited user interfaces, lower processing power and memory capacity compared to PC platforms.

The European Telecommunications Standardization Institute (ETSI) as well as the Wireless Application Protocol (WAP) Forum have addressed these specific requirements and characteristics by some of their work. While ETSI has developed three so-called classmarks of the Mobile Station Application Execution Environment (MExE), the WAP Forum has compiled a set of specifications targeting at mobile devices with limited capabilities. Because of such an optimized design, the generated traffic patterns cannot be mapped onto today's Internet models. Extra work is required to model and simulate WAP traffic.

For GPRS networks both conventional Internet applications running on laptop computers or enhanced PDAs and WAP-based applications running on smart phones and PDAs are predicted. For the introduction and evolution of GPRS networks dimensioning guidelines are needed for operators, equipment manufactures and system integrators. They should describe the relationship between the needed radio resources, the offered traffic, and the desired quality of service for different applications. This paper aims at presenting simulation results for a predicted traffic mix of WAP, WWW and e-mail applications that are usable for capacity estimation and radio network dimensioning [1], [2], [3].

In Section II the WAP specification is introduced. Next the simulation tool GPRSim is presented in Section III followed by the simulation scenarios and simulation results in Section IV.

II. Wireless Application Protocol (WAP)

The WAP specifications, which are the basis of the implementation in today's mobile terminals, including the June 2000 Conformance Release, also known as WAP 1.2.1, aim at optimizing operation in 2G networks. Therefore WAP 1.2.1 defines a distinct technology comprising protocols and content representation. WAP is a suite of specifications that defines an architecture framework containing optimized protocols (e.g., WDP, WTP, WSP), a compact XML-based content representation (WML, WBXML) and other mobile-specific features like Wireless Telephony Applications (WTA) [4].

A. WAP Release 1.x

In addition to the goal of optimized operation in 2G networks, WAP has been developed because today's graphics-enhanced web services cannot be brought to and displayed on thin clients, e.g., GSM mobile phones, and IP as network layer may not be applicable in some environments, e.g., WAP over Short Message Service (SMS) or Unstructured Supplementary Service Data (USSD). Because of the optimizations and different protocols it is not possible to run WAP end-to-end to a regular Internet site. Instead, a WAP Gateway must be used. The main services a WAP Gateway provides is protocol conversion between WAP stack and Internet stack. In addition to this standardized functionality, many gateway vendors provide a variety of value-added services that allow for personalization, for example.

B. WAP Release 2.0

In the specification WAP 2.0 [5] some existing WAP protocols have been extended by new capabilities. WAP 2.0

converges with widely used Internet protocols like the Transmission Control Protocol (TCP) and the Hypertext Transfer Protocol (HTTP). Internet Engineering Task Force (IETF) work in the Performance Implications of Link Characteristics (PILC) Working Group has been leveraged to develop a mobile profile of TCP for wireless links. This profile is fully interoperable with the common TCP that operates over the Internet today.

Further, WAP 2.0 does not require a WAP proxy, since the communication between the client and the server can be conducted using HTTP 1.1. However, deploying a WAP proxy can still optimize the communication process and may offer mobile service enhancements, such as location, privacy, and presence based services. In addition, a WAP proxy remains necessary to offer Push functionality.

In addition to protocol work, the WAP Forum has continued its work on service-enabling features for the mobile environment, like the push service or synchronization issues. Although WAP 2.0 has been finished in 2001, WAP 1.x protocol stacks will still be used in the mobile terminals in the next years. In this paper, only WAP 1.2.1 is regarded.

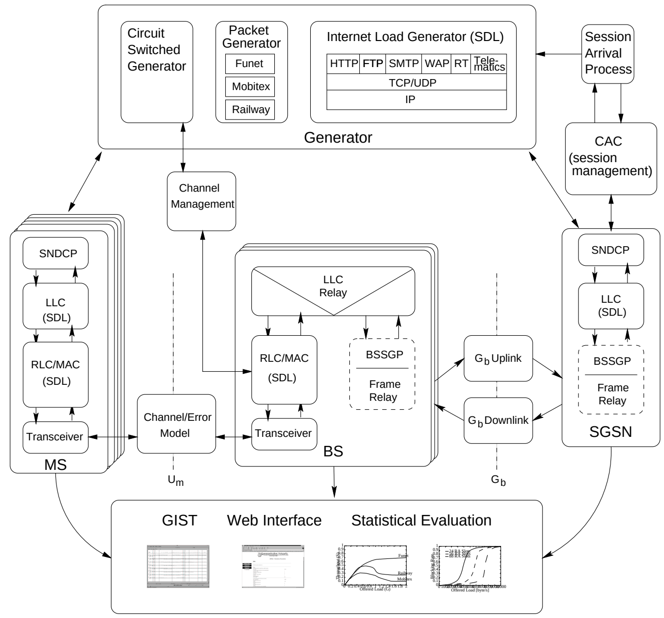

Fig. 1. The GPRS/EGPRS simulator GPRSim

III. Simulation

The capacity model studied in this paper is represented by the (E)GPRS simulator GPRSim that in fact is an emulator for GPRS and EGPRS. It represents a GSM/(E)GPRS network with its protocol stacks at the air interface, the radio channel attributes and application specific traffic sources that are represented by traffic load generators. Based on this model it is possible to create traffic performance results and to derive from there dimensioning graphs so that the network engineering for an expected traffic and the desired Quality of Service (QoS) can be performed.

The (E)GPRS Simulator GPRSim [6] is a pure software solution based on the programming language C++. Up to now models of Mobile Station (MS), Base Station (BS), and Serving GPRS Support Node (SGSN) have been implemented. The simulator offers interfaces to be upgraded by additional modules (see Figure 1).

For the implementation of the simulation model in C++ a Class Library (CNCL) is used, a predecessor to the SDL Performance Evaluation Tool Class Library (SPEETCL) [7] that enforces an object oriented structure of programs and is especially suited for event driven simulations.

Different from usual approaches to establish a simulator, where abstractions of functions and protocols are being implemented, the approach of the GPRSim is based on the detailed implementation of the standardized GSM and (E)GPRS protocols. This enables a realistic study of the behavior of EGPRS and GPRS. The real protocol stacks of (E)GPRS are used during system simulation and are statistically analyzed under a well-defined and reproducable traffic load.

The complex layers of the protocol stack like Subnetwork Dependent Convergence Protocol (SNDCP), Logical Link Control (LLC), Radio Link Control/Medium Access Control (RLC/MAC) based on GPRS/EGPRS release 99, the Internet traffic load generators and TCP/IP itself are specified formally with the Specification and Description Language (SDL) [8] and are translated to C++ code by means of the Code Generator SDL2CNCL [7] and are finally integrated into the simulator.

A. Packet Traffic Generators

The Internet sessions studied consist of the applications World Wide Web (WWW) and e-mail running on top of the TCP/IP protocol stack.

In the following, the parameters of the two applications that specify the characteristic traffic load to the (E)GPRS are presented. Related documents can be found in [9] and [10]. The parameters of these models have been updated by parameters given by ETSI/3GPP propositions for the behavior of mobile Internet users [11] (see Table I). For the log2-normal and log2-Erlang-k distributions the parameters of the transformed distribution functions are given in brackets.

1) WWW Model: WWW sessions consist of requests for a number of *pages*. These pages consist of a number of *objects* with a dedicated *object size*. Another characteristic parameter is the delay between two pages depending on the user's behavior to surf around the Web [9], [11]. Table I gives an overview of the WWW traffic parameters. The small number of objects per page (2.5 objects), and the small object size (3700 byte) were chosen, since Web pages with a large number of objects or large objects are not suitable for thin clients such as PDAs or smart phones served by (E)GPRS. The maximum object size in our model is set to 100 kbyte.

TABLE I

TRAFFIC MODEL PARAMETERS

WWW Parameter	Distribution	Mean	Variance
Pages per session	geometric	5.0	20.0
Intervals between pages [s]	negative exponential	12.0	144.0
Objects per page	geometric	2.5	3.75
Object size [byte]	$\log_2$-Erlang-k ($k = 17$)	3700 (transf.: 9.3)	$1.36 \cdot 10^6$ (transf.: 5.4)
e-mail Parameter	**Distribution**	**Mean**	**Variance**
e-mail size (lower 80 %) [byte]	$\log_2$-normal	1700 (transf.: 10.0)	$5.5 \cdot 10^6$ (transf.: 2.13)
e-mail size (upper 20 %) [byte]	$\log_2$-normal	15700 (transf.: 9.5)	$62.9 \cdot 10^9$ (transf.: 12.5)
Base quota [byte]	constant	300	0
WAP Parameter	**Distribution**	**Mean**	**Variance**
Decks per session	geometric	20.0	3800
Intervals between decks [s]	negative exponential	14.1	198.8
Size of 'Get Request' packet [byte]	$\log_2$-normal	108.2 (transf.: 6.34)	$4.1 \cdot 10^3$ (transf.: 0.71)
Size of 'Content' packet [byte]	$\log_2$-normal	511.0 (transf.: 8.6)	$3.63 \cdot 10^5$ (transf.: 1.55)

2) E-mail Model: The e-mail model describes the traffic resulting from the message download from a mail server by an e-mail user. The relevant parameters are the amount of data per e-mail and its distribution (see Table I). A constant base quota of 300 byte has been added per e-mail [10]. The distribution function is defined by two ranges. The lower range models e-mails without attachments and the upper range models e-mails with small attachments. The maximum e-mail size is set to 100 kbyte.

3) Wireless Application Protocol (WAP) Model: A WAP traffic model has been developed and applied in [12], which is characterized by a very small mean packet size (511 byte) modelled by a log2-normal distribution with a limited maximum packet size of 1400 byte (see Table I).

B. Transmission Control Protocol (TCP)

Classical TCP is implemented based on the description in [13] including slow start and congestion avoidance. In our HTTP implementation a TCP connection can be reused to transmit the following HTTP objects.

C. Air Interface Transmission Error Model

Within the channel/error model it is decided whether a received data or control block is error free or not. For this purpose a set of curves is used gained from link level simulations that allow the mapping of an actual C/I value to the corresponding block error rate (BLER) of a given radio block [3]. The TU3 (Typical Urban) channel model of GSM Rec. 05.05 was assumed there. For the Modulation and Coding Schemes (MCS) in EGPRS similar curves are available.

IV. SIMULATION RESULTS

A. Simulation Scenarios

The cell configuration is given by the number of Packet Data Channels (PDCHs) permanently available for GPRS. In this paper 1 and 4 fixed PDCHs have been regarded. A C/I of 12 dB (13.5% BLEP) has been regarded and Coding Scheme 2 (CS-2) has been used. LLC and RLC/MAC are operating in acknowledged mode. The multislot capability is 1 uplink and 4 downlink slots. The MAC protocol instances in the simulations are operating with 3 random access subchannels per 52-frame. All conventional MAC requests have the radio priority level 1 and are scheduled with a FIFO strategy. LLC has a window size of 16 frames. TCP/IP header compression in SNDCP is performed. The maximum IP datagram size is set to 1500 byte for WAP and 536 byte for the TCP-based applications. In the Internet stack for WWW and e-mail TCP is operating with a maximum congestion window size of 8 Kbyte. The transmission delay in the core network and external networks, i.e., the public Internet is neglected, since it is assumed that the servers are located in the operator's domain and the core network is well dimensioned.

The conventional Internet traffic regarded for comparison is composed of pure WWW traffic and pure e-mail traffic, respectively. The traffic mix in the second scenario is characterized by 60 % WAP, 28 % e-mail and 12 % WWW sessions (see Table I).

B. Performance Measures

To characterize the traffic performance of GPRS several performance and system measures are defined in the following.

The *Mean IP throughput per user* is the downlink IP throughput measured during transmission periods, e. g., the download period of a single object of a Web page. This is an important QoS parameter from a user's point of view. The statistical evaluation of this measure is done by counting the amount of IP bytes transmitted in each TDMA frame period for each user, if a packet train is running. Thus, the throughput

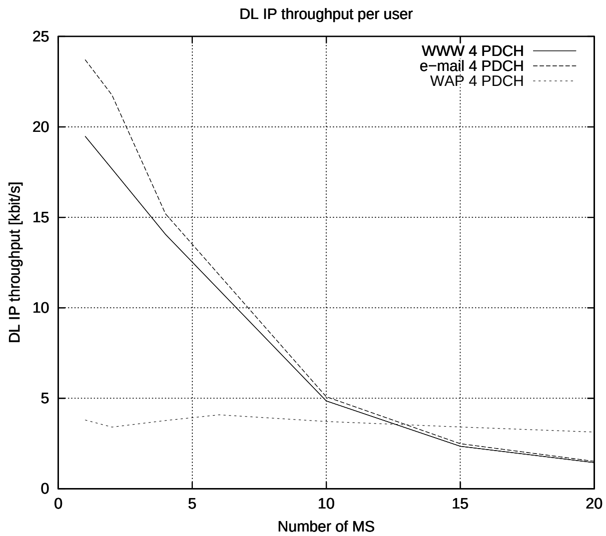

Fig. 2. Mean downlink IP throughput per user for WWW, e-mail and WAP traffic

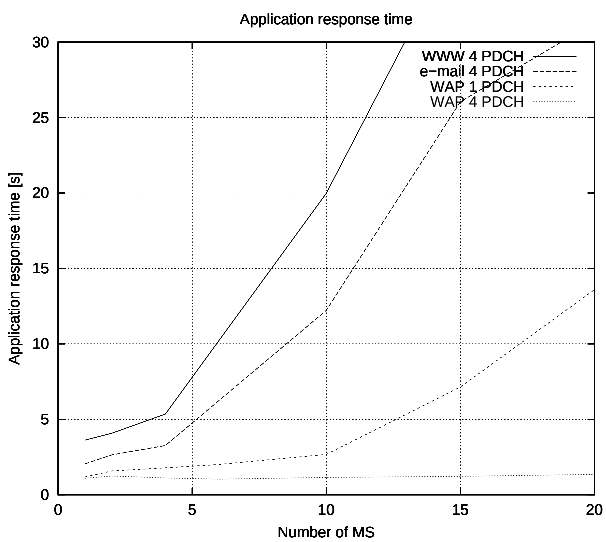

Fig. 3. Mean application response time for pure WWW, e-mail and WAP traffic

is not averaged over inactive periods. The number of IP bytes transmitted divided by the TDMA frame duration represents a simulation sample value in the evaluation sequence. At the end of the simulation the mean throughput is calculated from this evaluation sequence.

The *Mean Application Response Time* is the difference between the time when a user is requesting a page, a WAP deck or an e-mail and the time when it is completely received.

C. Simulation Results for WAP in Comparison to Internet Applications

To be able to compare the user-perceived performance of WAP in comparison to conventional Internet applications, the application response time is shown in Figure 3 for pure WWW and pure WAP traffic.

In situations with low traffic load the response time for a WAP deck is below 2 seconds, while the response time for a Web page is around 4 seconds. The reason is that a Web page has a larger content size and is transmitted over TCP. In load situations with higher traffic load the response time for a WAP deck remains nearly constant for up to 20 MS. If only 1 PDCH is available, the WAP response time increases to more than 10 seconds for 20 MS in the radio cell. Because of the larger content size the response time for Web pages passes 20 seconds already with 10 active MS in the radio cell even if 4 PDCHs are available. The reason for the strong increase in response time for WWW and e-mail could be seen in other evaluated measures like the downlink PDCH utilization that is not plotted here. 100 % PDCH utilization is reached for WWW traffic with 15 MS and even less for pure e-mail traffic, while 15 WAP users are only utilizing the PDCHs with 30 % for the same PDCH configuration.

Figure 2 shows the mean downlink IP throughput per user during transmission periods. While the throughput performance for pure WAP traffic remains relatively constant with an increasing number of mobile stations and 4 PDCHs, it decreases dramatically for pure WWW and pure e-mail traffic because of the higher offered traffic and the higher utilization. The poor throughput performance for WAP traffic can be explained by the low WAP deck size. Such transaction-oriented applications are more influenced by the high round-trip-time, which is mainly caused by the high delay over the air interface, than by the available bit rate. Since the response time for a WAP deck is less than 1.5 seconds, which should be acceptable for a wireless application, the user is not aware to this low throughput performance. Since WWW and e-mail applications comprise higher file sizes to download than WAP-based applications do, the throughput performance perceived by a user in situations with low traffic load ranges from 14 to 24 kbit/s. These performance values are mainly influenced by the characteristics of the offered traffic. Since the e-mail traffic model has higher file sizes than WWW, the throughput performance is better. With an increasing number of mobile stations up to 15 the saturation is reached and the performance for WWW and e-mail users gets unacceptable and even gets worse than the low throughput for pure WAP traffic. In this situation with high traffic load WWW and e-mail traffic performance is less influenced by the characteristics of the traffic model like the file size, but by the load at the air interface.

D. Simulation Results for a Traffic Mix of WAP and WWW/e-mail

Since the predicted traffic mix for GPRS networks will be composed of WAP traffic and conventional Internet applications like WWW and e-mail, the GPRS traffic performance for a traffic mix of 60 % WAP, 28 % e-mail and 12 % WWW sessions will be regarded, here.

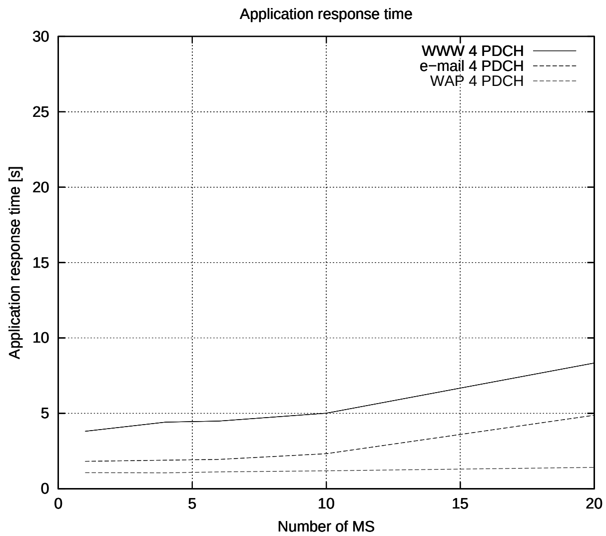

Fig. 4. Mean application response time for a traffic mix WAP/WWW/e-mail

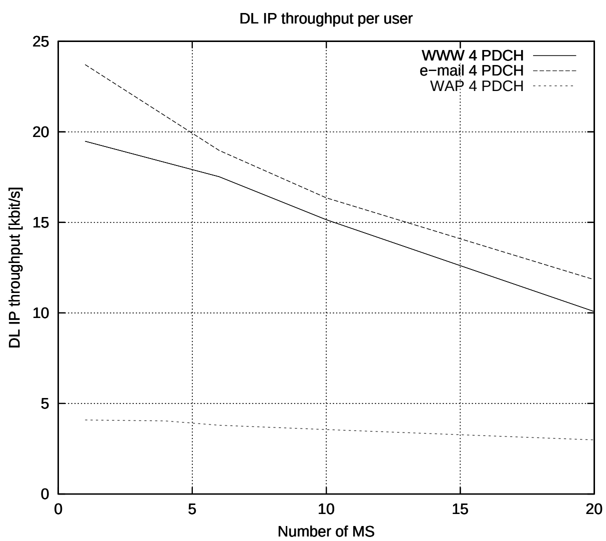

Fig. 5. Mean downlink IP throughput per user for a traffic mix WAP/WWW/e-mail

Figure 4 shows the application response time for WAP decks, e-mails and WWW pages, respectively. Compared to the graphs in Section IV-C the WWW and e-mail performance is not strongly affected by WAP traffic, since small WAP packets can be multiplexed seamlessly with the TCP-based WWW and e-mail traffic. The throughput (see Figure 5) decreases slower with an increasing number of mobile stations than in Figure 2 with pure WWW, e-mail and WAP traffic regarded separately, since here WAP represents the main part of a traffic mix and the total offered traffic per radio cell is increasing much slower. The same applies for the response time. In the scenario with traffic mix WWW pages have a response time of 5 seconds with 10 active stations generating a traffic mix, while 10 stations generating pure WWW traffic have to wait for more than 20 seconds.

On the other hand the WAP response time increases slightly from 1.2 seconds for pure WAP traffic to 2.1 seconds for the traffic mix scenario. The reason is that WWW and e-mail sessions are composed of larger application packets that leave less resources open for WAP users. Nevertheless a response time for WAP decks of 2.1 seconds still should be acceptable.

V. Conclusion

In this paper the performance of WAP and conventional Internet applications over GPRS is presented. First performance characteristics of WAP and Internet applications regarded separately are presented and compared. Furthermore, the effects of coexisting Internet traffic on WAP traffic and vice versa are outlined. It has been shown that WAP traffic can be multiplexed seamlessly with the Internet traffic because of the small and limited WAP deck size, while Internet traffic slightly slows down WAP traffic in situations with high traffic load.

References

[1] B. Walke, *Mobile Radio Networks - Networking, Protocols and Traffic Performance, 2nd Edition*, John Wiley & Sons, Chichester, 2001.

[2] P. Stuckmann and O. Paul, "Dimensioning GSM/GPRS Networks for Circuit- and Packet-Switched Services," in *Proceedings of the 10th Symposium on Wireless Personal Multimedia Communications, ISBN 87-988568-0-4*, Aalborg, Denmark, September 2001, pp. 597–602.

[3] P. Stuckmann, "Quality of Service Management in GPRS-based Radio Access Networks," *Telecommunication Systems 19:3, Kluwer Academic Publishers*, pp. 515–546, March 2002.

[4] Wireless Application Protocol Forum, "Wireless Application Protocol Architecture Specification," Tech. Rep., 1999.

[5] Wireless Application Protocol Forum, "Wireless Application Protocol 2.0 - Wireless Application Protocol Architecture Specification," Tech. Rep., 2001.

[6] P. Stuckmann, "Simulation Environment GPRSim: Tool for Performance Analysis, Capacity Planning and QoS Enhancement in GPRS/EDGE Networks," Technical Report, `http://www.comnets.rwth-aachen.de/~pst`.

[7] M. Steppler, "Performance Analysis of Communication Systems Formally Specified in SDL," in *Proc. of The First International Workshop on Simulation and Performance '98 (WOSP '98)*, 1998, pp. 49–62.

[8] ITU-T SG 10, "Functional Specification and Description Language (SDL)," ITU-T Recommendation Z.100, International Telecommunication Union – Telecommunication Standardization Sector, Geneva, Switzerland, 1993.

[9] M. Arlitt and C. Williamson, "A Synthetic Workload Model for Internet Mosaic Traffic," in *In Proc. of the 1995 Summer Computer Simulation Conference*, Ottawa, Canada, 1995, pp. 24–26.

[10] V. Paxson, "Empirically-Derived Analytic Models of Wide-Area TCP Connections," *IEEE/ACM Transactions on Networking*, vol. 2, no. 4, pp. 316–336, August 1994.

[11] ETSI 3GPP, "Selection Procedures for the Choice of Radio Transmission Technologies of the Universal Mobile Telecommunication System UMTS (UMTS 30.03, 3G TR 101 112)," Technical report, European Telecommunications Standards Institute, Sophia Antipolis, France, April 1998.

[12] P. Stuckmann, H. Finck, and T. Bahls, "A WAP Traffic Model and its Appliance for the Performance Analysis of WAP over GPRS," in *Proc. of the IEEE International Conference on Third Generation Wireless and Beyond (3Gwireless '01)*, San Francisco, USA, June 2001, pp. 338–343.

[13] R. Stevens, *TCP/IP Illustrated*, vol. 1, Addison-Wesley, Massachusetts, October 1996.