EuroFGI Workshop on IP QoS and Traffic Control
P. Pereira (Ed.)
Lisbon, Portugal, December 6–7, 2007

OPPORTUNISTIC SCHEDULING OF VOICE AND DATA TRAFFIC IN WIRELESS NETWORKS

Thomas Bonald and Luca Muscariello

Orange Labs, Paris, France
e-mail: thomas.bonald@orange-ftgroup.com, luca.muscariello@orange-ftgroup.com

Abstract. *We compare various scheduling schemes for the downlink channel of wireless systems like HDR or HSDPA serving both voice and data traffic. We first show that scheduling packets of voice traffic with strict priority yields significant wastage of radio resources. We then evaluate the performance of standard opportunistic schedulers that equally handle the packets of voice and data traffic. This solution is shown to be satisfactory for both types of traffic with the limitation that mobility cannot always be guaranteed at the cell border.*

Keywords: Opportunistic scheduling, service integration.

1 INTRODUCTION

While data services contribute most to traffic in 3G wireless networks, the support of VoIP is a critical issue for mobile operators. Unfortunately, the HDR and HSDPA evolutions of cdma2000 and UMTS standards, respectively, have not originally been designed for voice traffic. The primary objective was to offer high data rates by means of a common, time-shared downlink channel, combining link adaptation, hybrid ARQ and opportunistic scheduling [1, 2]. These dynamic schemes take advantage of both the uncorrelated fast variations of channel quality experienced by active users, the so-called multi-user diversity, and the inherent elasticity of data transfers to increase the overall system capacity: a user's packets are delayed until her radio conditions are favourable [3].

The queuing delay introduced by opportunistic scheduling may be unacceptable for VoIP calls. This is why some authors propose to account for the queue length of each user in the scheduling algorithm, see e.g. [4, 5]. However the radio resource allocation is then hardly predictable in a realistic traffic scenario with a mix of voice, video streaming and data flows. It would typically be biased towards those greedy, unresponsive flows like video streaming that feed the buffer independently of the degree of congestion of the wireless link. Such flows may increase packet delay to a degree that is unacceptable for VoIP calls.

Another proposal, which is standardized within 3GPP [2], consists in handling VoIP packets with strict priority. This requires the explicit marking of VoIP packets, which is feasible for those services controlled by the mobile operator only. In practice, a user's traffic is a mix of Web browsing, voice traffic, video streaming and file transfers, whose content is rarely accessible to the mobile operator. Moreover, network neutrality may be imposed in the near future to promote the emergence of new applications and new service providers [6]. This would banish any form of preferential treatment.

In this paper, we compare the performance of various scheduling strategies. We first show by queueing analysis that scheduling VoIP packets with strict priority yields significant wastage of radio resources, which provides yet another argument against this solution. We then focus on two opportunistic schedulers, namely Proportional Fair [3] and Score Based [7], that equally handle packets of voice and data traffic. Using both analysis and simulation, we show that, contrary to common belief, these schedulers may provide satisfactory quality of service for both types of traffic. We shall see that a proper setting of the scheduling algorithm time constant and the buffer size is required in this case.

The rest of the paper is organized as follows. The model is described in the next section. Sections 3 and 4 present the analytical and simulation results, respectively. Section 5 concludes the paper.

2 MODEL

2.1 Radio channel

The radio propagation model corresponds to the Pedestrian A specification set [8] with the parameters summarized in Table 1. This includes both slow fading, due to mobility and shadowing, and fast Rayleigh fading, due to multi-path propagation. Interference is assumed constant.

Table 1: Radio parameters

Parameters	Value
Carrier frequency	1950Mhz
Path loss exponent	3.52
ath loss at 1 km	137.4dB (antenna height of 30m)
Transmission power	38dBm
Antenna Gain	17dBi
Interference	30dBm
Mobile speed	3km/h

Radio propagation is assumed to be isotropic so that the fading pattern depends on the distance to the base station only. Users are uniformly located in the cell, with a cell radius equal to 500m. Figure 1 below shows the distribution of the traffic burst size, defined as the volume that can be transmitted by a user within a single slot, depending on the distance of this user from the base station. Assuming a perfect measure of the radio channel quality, this distribution directly follows from the HSDPA standard [1] with a slot duration of 2 ms and a maximum traffic bust size of 1800B (45 packets of 40B), corresponding to a maximum data rate of 7.2Mbit/s. Figure 1 is obtained from physical layer simulations of the described channel model. The average signal-to-noise ratio is approximately equal to -3dB at cell edge (500m from the base station) and 7dB at cell center (100m from the base station).

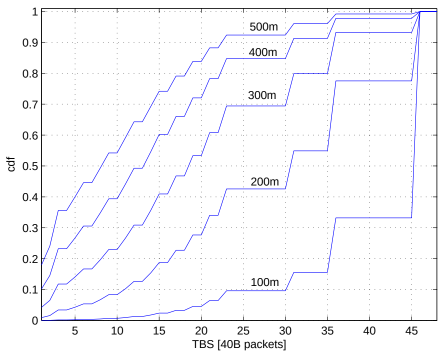

Figure 1: Traffic burst size distribution in number of 40B packets.

2.2 Scheduling schemes

Besides the priority scheduling scheme, where strict priority is applied to packets of voice traffic, we consider the following three fair scheduling schemes, under which the packets of voice and data traffic are equally handled:

- the round robin (RR) scheduler, that serves users in a cyclic way whatever their radio conditions;
- the proportional fair (PF) scheduler, that serves at each slot the user with the highest instantaneous

to average throughput ratio. Specifically, that user with the highest ratio $R(n)/\bar{R}(n)$ is selected at slot n, where $R(n)$ denotes the potential throughput at slot n and $\bar{R}(n)$ is the average throughput at slot n, estimated through a moving average over S slots:

$$\bar{R}(n) = (1/S)R(n) + (1 - 1/S)\bar{R}(n-1).$$

The PF scheduler is described in more details in [3];

- the score-based (SB) scheduler, that serves the user with the largest *score*, defined as the rank of its potential throughput among the past T observations. Specifically, the score of a user at slot n is the rank of her potential throughput at slot n, $R(n)$, among the T values $R(n-T+1), R(n-T+2), \dots, R(n)$. The SB scheduler is described in more details in [7].

For both the proportional fair scheduler and the score-based scheduler, increasing the time constant, S and T respectively, offers the opportunity of waiting a long time for favourable radio conditions before scheduling a user. We then expect the scheduler to better exploit multi-user diversity at the expense of longer packet delays.

2.3 Traffic scenario

We consider a mix of voice and data traffic. Voice flows consist of 40B packets (thanks to robust header compression [9]) sent every 24 ms, which corresponds to a constant bit rate of 13.3kbps. Data flows are permanent TCP flows, which are constrained by the considered wireless link only. We denote by N the total number of data flows.

3 ANALYTICAL RESULTS

In this section, we evaluate the impact of the scheduling policy in the simple scenario of a single voice call, neglecting the effects of time correlation in the fading process. We shall see in particular that handling the packets of this flow with strict priority yields significant resource wastage compared to the fair scheduling schemes described above. In Section 4, we use simulations to evaluate performance in a more realistic scenario of several voice calls, accounting for the time correlation in the fading process.

3.1 Queuing model

Let $Q(n)$ be the number of packets waiting in the queue associated with the voice call at slot n. We denote by $A(n)$ the number of packets arriving at slot n and by $B(n)$ the number of packets served at slot n. The random variables $A(n)$ and $B(n)$, $n = 1, 2, \dots$, are assumed to be i.i.d. with respective generating functions $\hat{A}(z) = \mathbb{E}[z^{A(1)}]$ and $\hat{B}(z) = \mathbb{E}[z^{B(1)}]$. For the numerical applications, packet arrivals are not assumed to be regularly spaced but Bernoulli to account for potential jitter introduced by successive buffering on its path from the encoder to the scheduler. This is represented by the probability generating function $\hat{A}(z) = (1-\lambda) + \lambda z$ where λ is the arrival frequency, here equal to $1/12$ (one packet every 24ms with a slot duration of 2ms).

Assume the queue is stable:

$$\rho = \frac{\mathbb{E}[A(1)]}{\mathbb{E}[B(1)]} < 1.$$

The probability generating function of the queue size $Q(n)$ is then given by [10]

$$\hat{Q}(z) = \frac{\hat{A}(z)(\mathbb{E}[B(1)] - \mathbb{E}[A(1)])(z-1)}{z^m - z^m \hat{A}(z)\hat{B}(1/z)} \prod_{k=1}^{m-1} \frac{z - z_k}{1 - z_k}, \tag{1}$$

where $z_0 \equiv 1, z_1, \dots, z_{m-1}$ are the m zeros of the polynomial

$$z^m - z^m \hat{A}(z)\hat{B}(1/z) \tag{2}$$

in the unit circle $|z| < 1$. If (2) is a polynomial of degree $n > m$, we have the simpler expression:

$$\hat{Q}(z) = \hat{A}(z) \prod_{k=m}^{n-1} \frac{1 - z_k}{z - z_k}, \tag{3}$$

where $z_m, \ldots, z_{n-1}$ are the $n - m$ zeros of (2) outside the unit circle. We deduce the probability that the queue is non-empty in steady state:

$$\mathbb{P}(Q > 0) = 1 - \hat{Q}(0), \tag{4}$$

as well as the mean queue size:

$$\lim_{n\to\infty} \mathbb{E}[Q(n)] = \mathbb{E}[A(1)] + \sum_{k=m}^{n-1} \frac{1}{z_k - 1}, \tag{5}$$

3.2 Priority scheduling

We first assume that voice traffic has strict priority over data traffic. VoIP packets do not "see" data packets so that the probability generating function $\hat{B}(z)$ coincides with the traffic burst size distribution of Figure 1. We measure the channel utilization as the fraction of slots consumed by the voice call, which corresponds to the probability that the considered queue is non-empty, given by (4). The results are shown in Figure 2. Under priority scheduling, the channel utilization of the voice call is independent of the

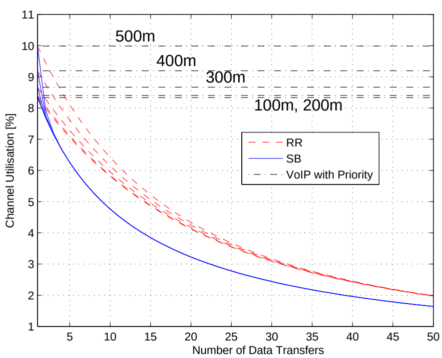

Figure 2: Channel utilization of the voice call as a function of the number of competing data flows (distance from 100m to 500m, from bottom to top).

number of competing data flows. It is equal to 8% at cell center (one packet every 12 slots) and increases slightly at cell edge due to packet losses and retransmissions. The channel utilization of the voice call is much less under fair scheduling, as explained in the following.

3.3 Fair scheduling

Under the RR scheduler, voice packets are served once every $N + 1$ slots, where N is the number of competing data flows. The corresponding sequence $B(n)$, $n = 1, 2, \ldots$, is not i.i.d. Previous results apply, however, by considering the queue size every $N + 1$ slots. The probability generating function associated with arrivals over $N + 1$ slots is equal to $\hat{A}(z)^{N+1}$, while the probability generating function associated with the number of packets served every $N + 1$ slots, which depends on the distance to the base station, follows from the traffic burst size distribution of Figure 1.

The PF scheduler is hard to analyse due to the time correlation introduced by the moving average. In particular, the service sequence $B(n)$, $n = 1, 2, \ldots$, is not i.i.d. even in the absence of time correlation in the fading process. The SB scheduler, on the other hand, is designed in such a way that each user is scheduled with probability $1/(N+1)$ at each slot. Moreover, the traffic burst size of the scheduled user has the same distribution as the maximum of $N+1$ independent samples of the traffic burst sizes of this user. The corresponding probability generating function $\hat{B}(z)$ easily follows. For large values of N, the random variable $B(1)$ tends in distribution to a bimodal random variable equal to the maximum burst size with probability $1/(N+1)$ and null otherwise. The simulation results of Section 4 suggest that the PF scheduler and the SB scheduler have similar performance.

Under fair scheduling, the voice call must share the slots with data traffic: when the number of data flows increases, VoIP packets are delayed and served in bursts, which decreases the channel utilization of the voice call, as illustrated by Figure 2. This in turn improves the total throughput of data flows. For $N = 20$ competing data flows for instance, the voice call consumes 50% and 60% less slots under RR and SB scheduling, respectively, than under priority scheduling. The price to pay is a significantly higher packet delay. This is illustrated by Figure 3 showing the numerical evaluation of (5) for the three considered schedulers. We shall see in the following section that a proper setting of the buffer size and the algorithm time constants is required to maintain reasonably low queuing delays.

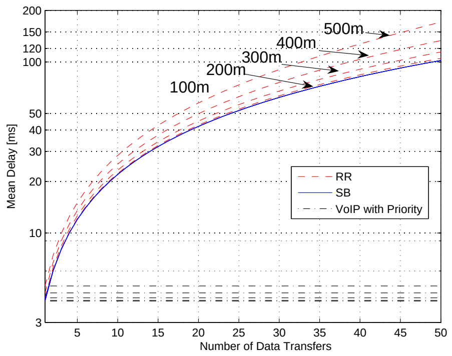

Figure 3: Mean voice packet delay as a function of the number of competing data flows (distance from 100m to 500m, from bottom to top).

4 SIMULATION RESULTS

The simulations are based on the EURANE extension of ns2 [11, 12], using the model described in Section 2. The main functionality additions to ns2 come in the form of the RLC Acknowledged Mode (AM), Unacknowledged Mode (UM), MAC-hs support for HS-DSCH, i.e. HSDPA. Every mobile terminal is supposed to function in AM mode, as we assume that VoIP and data traffic are not differentiated by any mechanism. The UM mode is a buffer management mechanism developed for real time applications that makes use of per packet timer in order to limit channel wastage. Furthermore we developed additional modules to implement the downlink schedulers that we study in the present paper.

Unless otherwise specified, the time constants of the PF scheduler and the SB scheduler are taken equal to 64ms. We first study the impact of the buffer size on the packet delay of voice calls. We then consider call outage and packet loss rate, which are two other key performance metrics. Finally, we analyse the impact of the scheduling algorithm time constant.

4.1 Buffer size setting

Packet delays turn out to be significantly impacted by the time correlation of the fading process: it is not rare that mobiles experience very bad radio conditions for a few seconds, during which packets fill the buffer. This is why it is very important to limit the buffer size dedicated to each user in order to drop packets instead of letting them accumulate. The impact of buffer size on voice packet delay is illustrated by Figure

Table 2: Per user TCP throughput for different values of the number of VoIP calls and the buffer size.

Location	Cell center				Cell edge			
Nb of VoIP calls	5		25		5		25	
Buffer size (pkts)	500	50	500	50	500	50	500	50
RR (kbit/s)	590	587	168	170	157	160	43	42
PF (kbit/s)	710	717	200	197	180	182	74	71
SB (kbit/s)	640	633	203	195	171	173	66	70

4, assuming $N = 5$ ongoing data transfers uniformly located in the cell.

Figure 4 show delays that are clearly unacceptable for a voice call under realistic channel conditions at the cell border (above 300m). This is mainly because of slow fading due to mobility that impacts severely voice quality at the cell border.

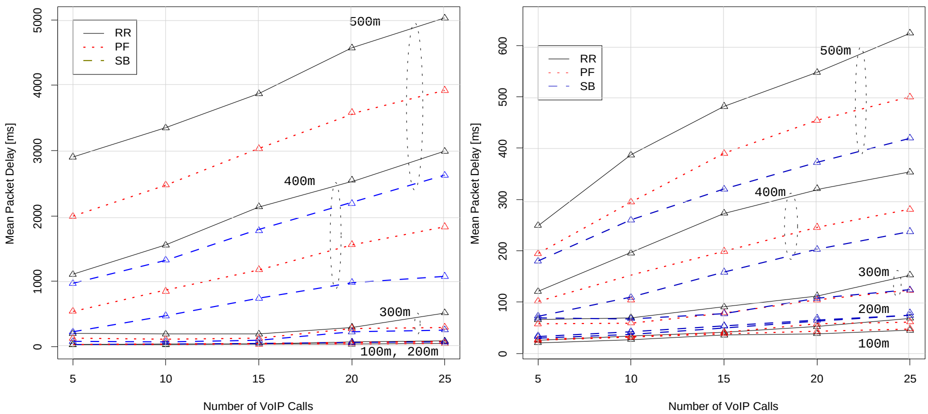

Figure 4: Mean delay as a function of the number of VoIP calls in progress (distance from 100m to 500m, from bottom to top) for two values of user buffer size: 500 packets (left plot) and 50 packets (right plot).

Of course, a small buffer size may reduce the efficiency of TCP transfers. Table 2 above shows the impact of buffer size on the average throughput per user, which turns out to be marginal.

4.2 Call outage and packet loss rate

Voice quality may be measured, besides mean packet delay, by call outage probability and packet loss rate. The call outage probability corresponds to the fraction of time the call is in outage, which is directly perceived by the user. We assume a call is interrupted if more than 5 packets over 10 consecutive packets experience a delay larger than 500ms, following the methodology in [13]. The call is then re-established after an exponential idle period of mean 2 seconds. The packet loss rate, on the other hand, can be compensated for by the encoder through error concealment technique, for loss rates up to 10%. This threshold assumes the use of a highly robust voice encoder as iLBC [14] that can cope with such loss intensity [15]. Encoders of the ITU's family (G.723.1, G.728, G.729), on the contrary, do not work when the loss rate exceeds 5%.

Figure 5 below shows the call outage and packet loss rate for a user buffer size of 50 packets. We observe that the SB scheduler outperforms both the PF and RR schedulers.

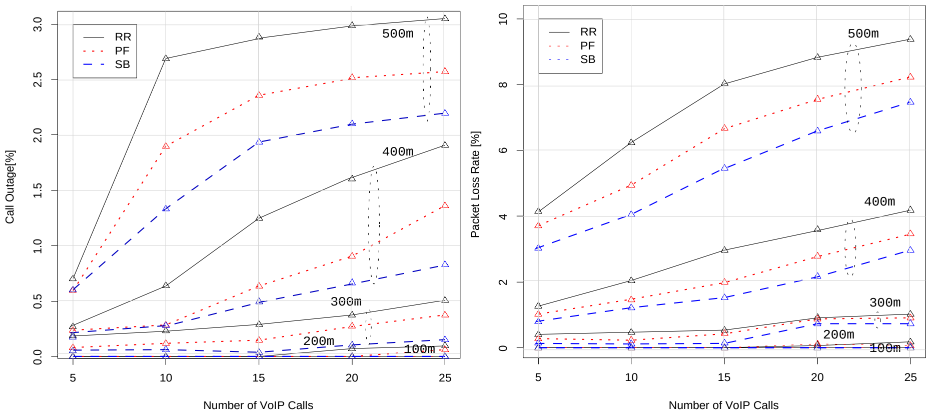

Figure 5: Call outage (left plot) and packet loss rate (right plot) as a function of the number of VoIP calls in progress (distance from 100m to 500m, from bottom to top) for a user buffer size of 50 packets.

4.3 Time constant setting

The time constant of the opportunistic scheduling algorithm may be viewed as the time horizon over which the scheduler tends to be fair and to serve all users. Increasing the time constant offers the opportunity of waiting a long time for favourable radio conditions before scheduling a user, at the expense of longer packet delays. While a time constant of 64ms allows the scheduler to exploit fast fading only, a time constant of 1s allows the scheduler to exploit both fast fading and slow fading.

As shown in Figure 6, the impact of the time constant on voice packet delay in fact depends on the radio propagation environment. In the absence of slow fading, the mean packet delay is larger with a time constant of 1s than with a time constant of 64ms. This is because the scheduler waits longer for favourable radio conditions whereas the scheduling gain is marginal due to the absence of slow fading. In a typical suburban propagation environment with both fast fading and slow fading, the mean packet delay is *less* with a time constant of 1s than with a time constant of 64ms. This is because the scheduler better exploits slow fading fluctuations, which decreases the data traffic load: there are less packets in the buffer on average and the mean packet delay decreases. Thus, provided the buffer is properly sized (as suggested in Section 4.1), a large time constant is beneficial for both voice and data traffic.

The absence of slow fading due to mobility, figure 6 right plot, significantly improves voice quality at the cell border (delays less than 150 ms).

5 CONCLUSION

Using both queuing analysis and simulation, we have shown that opportunistic scheduling is an attractive solution for voice and data traffic integration. This is in contrast with the common belief that priority scheduling is necessary to control voice packet delays. We have seen that strict priority scheduling in fact results in significant wastage of radio resources. It is much more efficient to opportunistically schedule both voice and data packets and to control delays by appropriate buffer sizing.

In this respect, the simulations suggest that the SB scheduler outperforms the PF scheduler in terms of voice quality. Both are much more efficient than a simple RR scheduler. The time constant should be set large (1s, say), especially in suburban propagation environment, to allow the scheduler to exploit slow fading fluctuations. Surprisingly, this tends to decrease voice packet delays: the more efficient utilization of radio resources compensate for the additional scheduling delay. Thus both voice and data traffic take advantage of opportunism. Nevertheless voice quality is unacceptable at the cell border because of slow fading, hence channel conditions do not allow complete mobility in the cell.

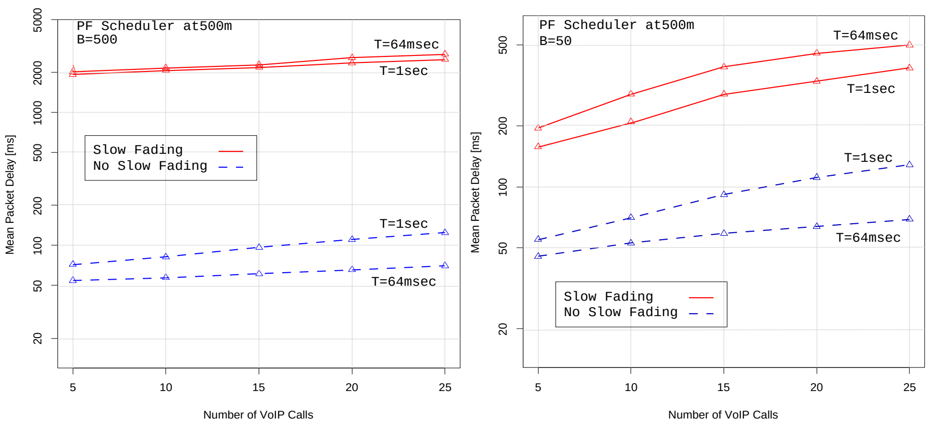

Figure 6: Mean delay as a function of the number of VoIP calls in progress (at distance 500m) for different time constants (64ms and 1s) and user buffer sizes of 500 packets (left plot) and 50 packets (right plot).

REFERENCES

[1] 3GPP TS 25.214, “Physical Layer Procedures (FDD)”, Release 5, 2003.

[2] 3GPP TR 25.308, “HSDPA: UTRAN Overall Description”, Release 5, 2004.

[3] P. Viswanath, D. Tse and R. Laroia, “Opportunistic Beamforming using Dumb Antennas”, IEEE Transactions on Information Theory, vol. 48(6), June, 2002.

[4] S. Shakkottai, A. L. Stolyar, “Scheduling Algorithms for a Mixture of Real-Time and Non-Real-Time Data in HDR”, In Proc. of ITC 16, 2001.

[5] M. Andrews, K. Kumaran, K. Ramanan, A. L. Stolyar, R. Vijayakumar, P. Whiting, “CDMA data QoS scheduling on the forward link with variable channel conditions”, Bell Labs Tech. Memorandum, April 2000.

[6] T. Wu, “Network Neutrality, Broadband Discrimination” Journal of Telecommunications and High Technology Law, Vol. 2, p. 141, 2003.

[7] T. Bonald, “A score-based opportunistic scheduler for fading radio channels”, Proc. of European Wireless 2004.

[8] ITU-R Rec. M. 1225, “Guidelines for Evaluation of Radio Transmission Technologies for IMT-2000”, 1997.

[9] IETF RFC 3095, “Robust Header Compression (ROHC)”, July 2001.

[10] A. Janssen, J. Leeuwaarden, “Analytic Computation Schemes for the Discrete-Time Bulk Service Queue” Queueing Systems, Volume 50, Numbers 2-3, July 2005, pp. 141-163(23).

[11] “The Network Simulator - ns-2”, `http://www.isi.edu/nsnam/ns/`.

[12] “E.U.R.A.N.E.: Enhanced UMTS Radio Access Network Extensions for ns-2”, `http://www.ti-wmc.nl/eurane/`.

[13] 3GPP2 WG5 Evaluation AHG, “1XEV-DV Evaluation methodology - Addendum (V6)”, July 25, 2001.

[14] IETF RFC 3951, “Internet Low Bit Rate Codec (iLBC)”, December 2004.

[15] “iLBC Whitepaper”, available from Global IP sound, October 2004.