

Free-Drop TCP

Yuliy Baryshnikov* Ed Coffman† Jing Feng‡ Vishal Misra§

April 22, 2006

Abstract

To enhance performance, typical TCP variants propose explicit use of traffic measurements or adjustments to the increase/decrease parameters of the AIMD (additive increase, multiplicative decrease) protocol [7, 3, 13, 6]. We introduce here a new class of TCP congestion control algorithms that take quite a different approach: they modify instead the rule for deciding when to cut the congestion window. The class is defined by an additional window with a packet-count parameter w ; the congestion window is reduced by half when a packet loss is detected, at time t say, if and only if there has been at least one dropped packet in the last w packet transmissions prior to time t . An algorithm in the class is called *Free-Drop TCP*, since dropped packets are “free” (they do not cause cuts in the window size) unless they are sufficiently bursty. We propose this new class as a means to achieve higher throughputs in high bandwidth-delay product networks without undue sacrifices in such measures as fairness and coefficients of variation. In terms of the loss probability p , the greater asymptotic throughput for a given window size w is now described by a $\Theta(1/p\sqrt{w})$ law rather than the classical $\Theta(1/\sqrt{p})$ law.

The paper begins with an analysis of a fluid model which leads to explicit estimates of the average throughput for small loss probabilities. We then show the results of several experiments, and compare the response curves under the new algorithm to those of TCP. We establish that a family of ‘shifted’ response functions of the form $O(1/\sqrt{p-\varepsilon})$ can be obtained over a wide range of p by suitably varying w . Increases in throughput under AIMD must be balanced against possible increases in the variance of the congestion window size, higher rates of congestion events and poorer fairness properties. Experiments are described which suggest that the sacrifices made by FD-TCP for greater throughput are modest at worst.

*Bell Labs, Lucent Technologies.

†Electrical Engineering Department, Columbia University. Work funded in part by the National Science Foundation.

‡Electrical Engineering Department, Columbia University

§Computer Science and Electrical Engineering Departments, Columbia University. Work funded in part by the National Science Foundation

1 Introduction

We propose a new TCP congestion control protocol called *Free-Drop TCP*; the dynamics of the congestion window W_t still follow the Additive Increase Multiplicative Decrease (AIMD) rule, but W_t is reduced less reactively. A cut (by half) is made in the congestion window at time t if and only if a congestion event (packet loss) occurs at time t and at least one other occurred during the preceding $w > 0$ packet transmissions; otherwise, W_t remains in its additive-increase mode. The window-size w is a tunable parameter of Free-Drop TCP that controls the slow-down in the response to congestion events; a greater inertia allows one to increase substantially the expected throughput. We call the window introduced by FD-TCP the *packet-loss* window to distinguish it from the standard congestion window (*cwnd*) of TCP.

For purposes of asymptotic analysis, it is convenient to adopt a simpler, more elegant fluid model (see e.g., [10]) as opposed to a discrete model [12]. In this continuous model, packet losses are modeled by a Poisson *loss arrival* process with rate parameter λ . The discrete window of size w is replaced by a time window of duration T . Let $n_T(t)$ denote the number of packet losses in the window $[t - T, t]$. The dynamics of W_t in Free-Drop TCP are

$$W_{t+dt} = \begin{cases} \frac{1}{2}W_t, & \text{if } n_T(t + dt) > n_T(t) > 0; \\ W_t + \frac{dt}{RTT}, & \text{otherwise,} \end{cases} \quad (1)$$

where RTT is the average round trip time. Thus, congestion window growth is linear at rate $1/RTT$ with reductions by half occurring at each loss arrival that follows an earlier loss by less than T time units. This simple model takes no explicit account of the slow-start mechanism, the details of packet acknowledgement, a maximum possible congestion window size, and time-outs as a source of packet loss. Thus, we restrict our analysis to asymptotics largely unaffected by these considerations, except possibly for changes in unspecified multiplicative constants; our focus is on the dependence of throughput on the packet loss rate and on the corresponding changes in higher moments and fairness measures. We adopt common parlance in using the term *congestion event* to refer to packet losses that precipitate reductions in the size of the congestion window.

Background: A full accounting of the literature on TCP/AIMD and variants would exceed our space constraints, so we confine ourselves to useful references that also have large bibliographies. For these, a discussion of AIMD variants can be found [8], and [6, 4] have recent lists of references to the many TCP variants. References to the early stochastic modeling and analysis of TCP can be found in [10], which introduced fluid models in the analysis of TCP (as well as an application of stochastic differential equations). Fluid models were subsequently pursued by many authors; see e.g., [1] for a recent list of references. Modeling within the framework of dynamical systems has had substantial coverage, with the primary focus being issues of stability and convergence properties; see e.g., [8] for a collection of numerous references.

2 Free-Drop TCP Analysis

To find the asymptotic throughput B for small loss rates, we apply TCP's inverse-square-root law [11], which has the form

$$B \sim \frac{\alpha}{RTT\sqrt{p}} \quad (2)$$

as $p \rightarrow 0$,* where, depending on model details, α is usually a constant in the range $\sqrt{3/2} \leq \alpha \leq \sqrt{2}$. The fact that this result survives in so many of the more or less sophisticated studies of TCP variants is an artifact of the AIMD congestion control that underlies each variant. An elementary derivation of the inverse-square-root law in [9], while heuristic, brings out clearly the basic observations leading to its form in (2). Experiments show that the range of loss rates p for which (2) is a useful approximation includes those in practical TCP environments. In applying the approximation, however, one must bear in mind that, if p is too large, the missing second-order terms can no longer be ignored, and if p is taken too small, the system begins to degenerate; the moments of the throughput either tend to ∞ if no maximum is placed on W_i , or, if there is such a maximum, throughput will tend to a constant.

As shown in [10], when the condition $n_T(t) > 0$ is removed from (1), one has

$$B = \frac{2}{RTT^2\lambda} \quad (3)$$

so with the loss-rate substitution $Bp = \lambda$, one again obtains (2). With the condition $n_T(t) > 0$ retained in (1), the probability that a loss arrival causes a reduction in the congestion window is just the probability, $1 - e^{-\lambda T} = \lambda T + O(\lambda^2)$, that there were one or more loss arrivals in the preceding interval of duration T . One is tempted to replace the congestion control rate λ of TCP by a new rate $\lambda \cdot \lambda T$ of Free-Drop TCP, replace T and λ according to $w = BT$, $\lambda = Bp$, and then write

$$B \sim \frac{\alpha}{RTT p \sqrt{w}} \quad (4)$$

This formula in fact holds, but additional argument is needed to justify it, since the congestion control events are not independent under Free-Drop TCP. This follows directly from the fact that congestion events in overlapping packet-loss windows are not independent. Thus, an underlying Poisson process of congestion events does not exist, which was needed to establish (3). A similar independence assumption required by (2) is also violated in discrete TCP/AIMD models.

However, the desired Poisson property is obtained asymptotically for small λ , and to see this, an informal argument using routine observations suffices. Note that it is just those loss interarrival times less than T which end in a reduction of the congestion window. Define a sequence of Bernoulli trials on the sequence of interarrival periods, with the i -th success interpreted as the event that the i -th interarrival period is less than T time units. The probability of success is $1 - e^{-\lambda T}$, so the number of trials between consecutive successes tends to ∞ , i.e., successes become rare events, as $\lambda \rightarrow 0$. As λ shrinks, the times between congestion events asymptotically approach sums of geometrically distributed numbers of rate- λ exponential interarrival times, and hence an exponential distribution with rate parameter $\lambda(1 - e^{-\lambda T}) \sim \lambda^2 T$. This is the desired asymptotic Poisson law which justifies the asymptotic result in (4).

Clearly, the above discussion suggests that, in principle, a greater premium is placed on taking p small when applying (4) as an approximation. However, our experiments in Section 3 show that the excellent quality of the approximation remains virtually unchanged in the parameter space of interest.

If we fix w , then the dependence on p (i.e., p^{-1}) in (4) shows an improvement over the dependence on p in the corresponding result for HS-TCP [2], $B \sim \alpha/p^{0.82}$. However, even if we stick to TCP, the key question is whether $p^2 w < p$, and no analysis currently exists which answers this question. There are two balancing effects at work in FD-TCP: by slowing down window cuts it strives more for greater allocations; in the aggregate, the number of flows will tend toward higher utilization. But in doing so, FD-TCP produces more packet losses; as p increases for fixed w more and more packets are dropped. This points up another difficulty in comparing TCP with FD-TCP on the basis of (2) and (4): holding other parameters fixed, the two values of p , the drop probability, are not the same in general. The next section resorts to experiments to make more definitive comparisons.

*Hereafter, as our interest is exclusively in small- p asymptotics, we will often omit "as $p \rightarrow \infty$."

3 Experimental results

THROUGHPUT. For any fixed w and small p , (4) shows that the improvement of FD-TCP over TCP is within a constant factor of $1/p$, where the constant is inversely proportional to w . To further investigate how varying the FD-TCP parameter w for given p can improve throughput performance, we conducted experiments on a bottleneck-link model specifically designed to produce a uniformly shifted response curve as shown in Figure 1 along with the TCP response curve, where $\varepsilon < p$ is the amount of the shift. Specifically, with an appropriate variation of w with p , the new protocol shifts the TCP response curve uniformly along the p axis a fixed amount ε . Figure 1(a) illustrates the least-squares fit to the data. Figures 1(b) and 1(c) omit the data points and show response curves for $\varepsilon = 0$ (TCP), 0.0005, 0.001 (Figure 1(b)), and 0.00005 (Figure 1(c)). To produce these shifts w is varied as shown in Figure 4. The functional behavior of $w(p)$ in Figure 4 is easily estimated: just equate (4) to a function $\beta/\sqrt{p-\varepsilon}$ to obtain, for fixed $\varepsilon < p$

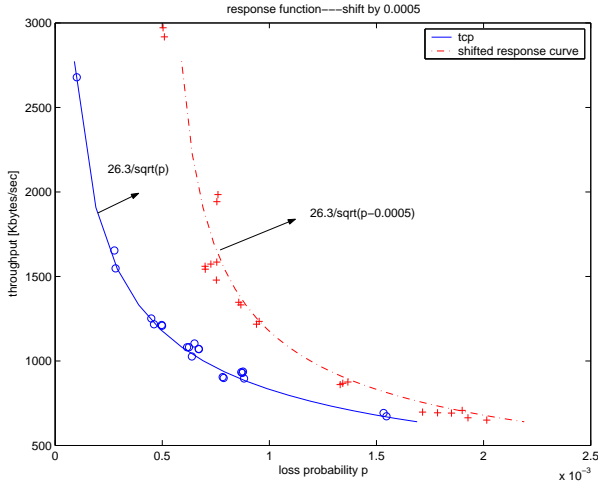
$$w(p) = \Theta\left(\frac{p-\varepsilon}{p^2}\right)$$

FAIRNESS. One measure of the fairness of Free-Drop TCP its coefficient of variation(CoV) of the congestion window(cwnd) size. Intuitively, when window cuts are reduced one expects an increase in variability. For the CoV metric, we sample the congestion window of a flow at periodic time intervals. The measurement starts after a delay of 30 seconds for each 10-minute test. The test topology and settings are the same as the throughput tests. The increased CoV of Free-Drop TCP, about 1.6 times that of TCP as shown Figure 3, can be counted as one of the costs of the higher throughput.

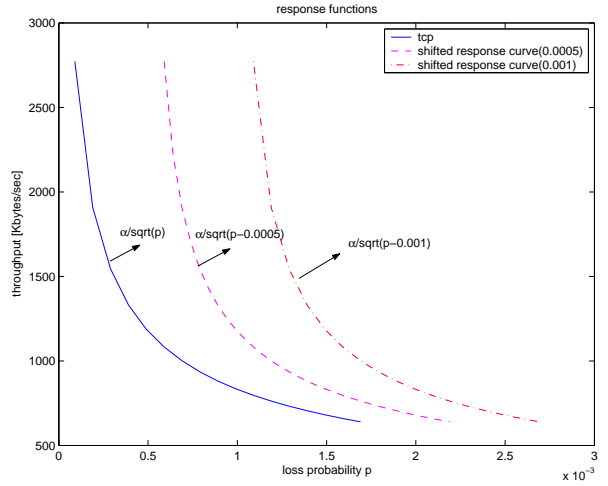
As another measure of fairness to compare TCP and FD-TCP, we use Jain's well-known index [5]. We consider homogeneous sets of flows, and so associate fairness with a roughly equal division of bandwidth to each. To wit, we computed for N competing flows with throughputs B_i ,

$$F = \frac{(\sum_{1 \leq i \leq N} B_i)^2}{N \sum_{1 \leq i \leq N} B_i^2}$$

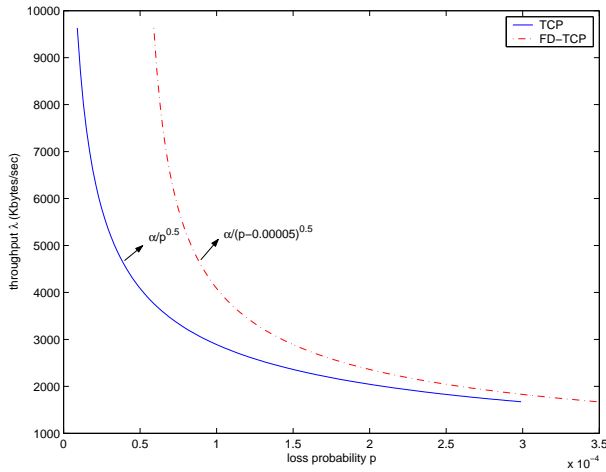
Note that $F = 1$ corresponds to perfect fairness; each of the throughputs is exactly the same. The experimental results are given in Figure 3(a) – 3(d), with those in figures showing that both TCP and FD-TCP are fair protocols with little to choose between them from the standpoint of fairness.



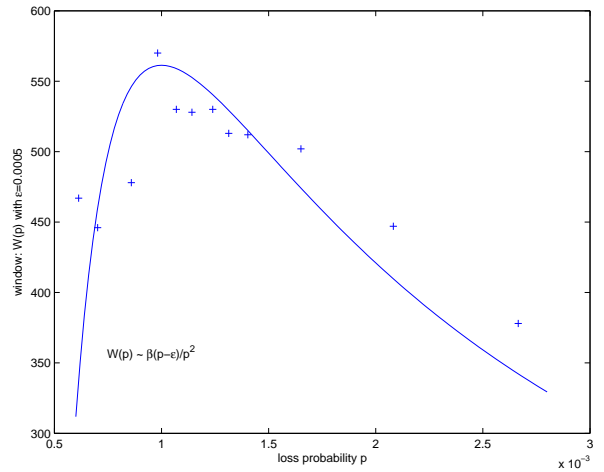
(a) Shifted curve fit to data



(b) $\varepsilon = 0(\text{TCP}), 0.005, 0.001$



(c) $\varepsilon = 0.00005$



(d) $w(p)$ for $\varepsilon = 0.0005$

Figure 1: Shifted throughput curves for $\varepsilon = 0$ (TCP), 0.00005, 0.0005, 0.01 and the function $w(p)$ corresponding to $\varepsilon = 0.0005$. Tests are of 10 minutes duration. All the results are obtained from tests of a single flow traversing over a lossy link with loss probability p . The link bandwidth is 600Mbps and the RTT is 50ms. Scatter points in Figure 1(a) are results from repeated tests with the same setting. Each curve in Figure 1(a), 1(b) and 1(c) is obtained by fixing ε and varying w according to Equation (??), which is shown in Figure 1(d).

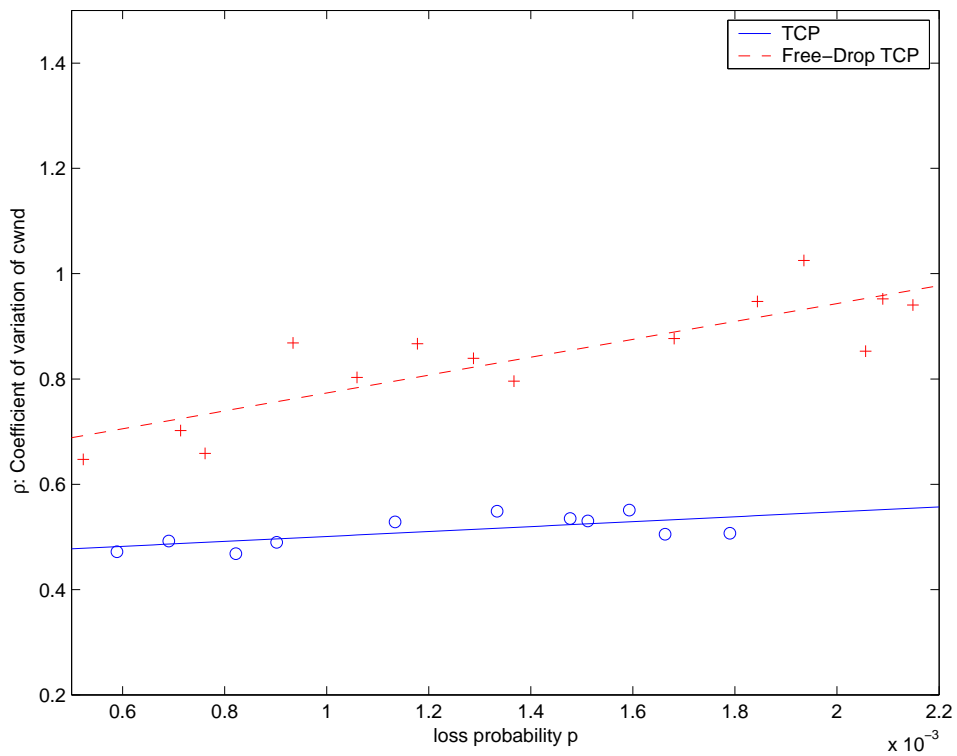
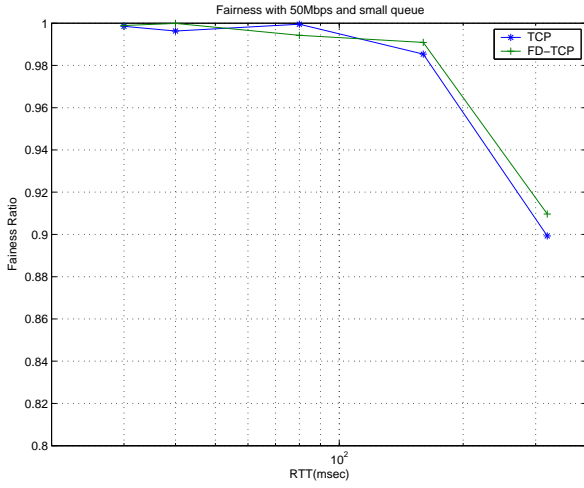
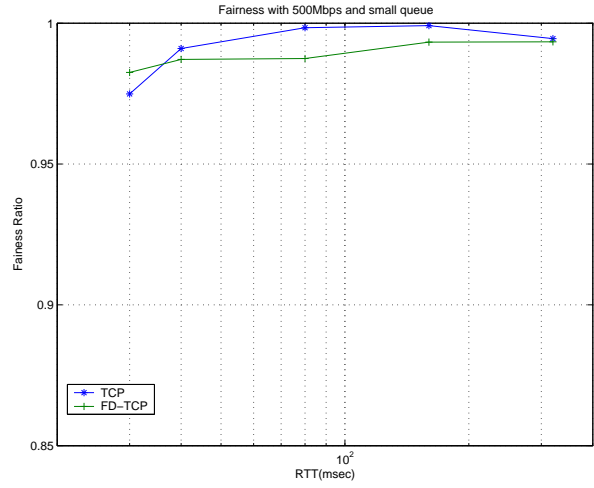


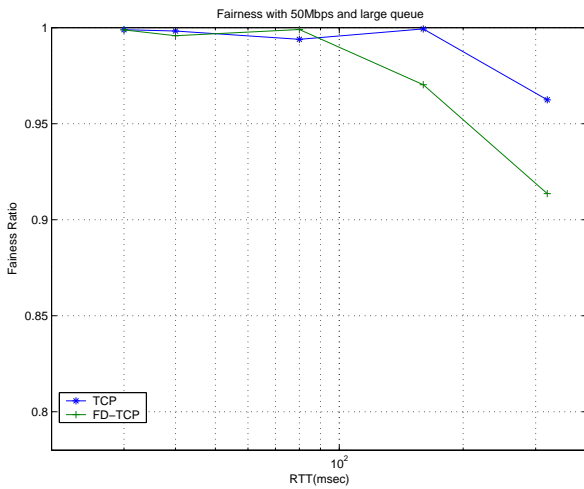
Figure 2: The CoV of cwnd. Bottleneck link bandwidth is 600Mbps and RTT is 50ms. Each point in this figure represents the CoV obtained from a particular test with a fixed link loss probability p . The corresponding throughput of Free-Drop TCP flow is shown in Figure 1(a).



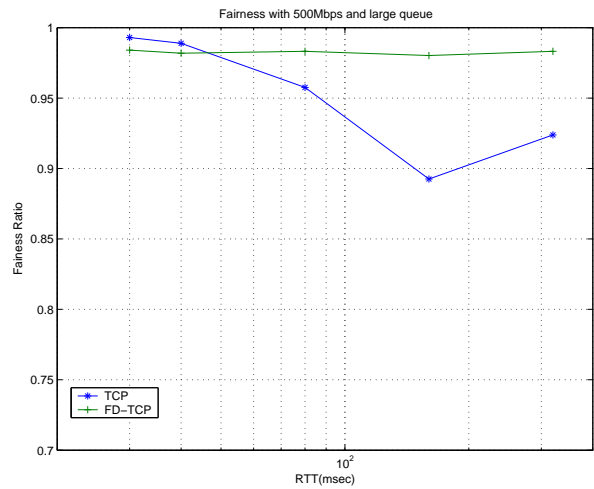
(a) Small bottleneck bandwidth and small queue $\sim 20\%BDP$



(b) Large bottleneck bandwidth and small queue



(c) Small bottleneck bandwidth and large queue $\sim 100\%BDP$



(d) Large bottleneck bandwidth and large queue

Figure 3: Figure 3(a) plots the Fairness Index for bottleneck BW=50Mbps and small queue size ($\sim 20\% BDP$). All the fairness tests are defined on the dumbbell topology with $N=5$ competing flows of the same protocol. Each flow has the same propagation delay and a shared bottleneck link. The average throughput of each flow is measured. Figure 3(b) displays results for bottleneck BW=500Mbps and small queue size. Figure 3(c) plots BW=50Mbps and large queue size ($\sim 100\%BDP$), and figure 3(d) shows results from BW=500Mbps and with large queue size. We observe that Free-Drop TCP is essentially fair under these conditions.

References

- [1] Y. Baryshnikov, E. Coffman, J. Feng, and P. Momcilovic. Asymptotic analysis of a nonlinear AIMD algorithm. *Discrete Mathematics and Theoretical Computer Science*, pages 27–38, 2005. In proceedings, 2005 International Conference on Analysis of Algorithms, Conrado Martinez (ed.).
- [2] S. Floyd, S. Ratnasamy, and S. Shenker. Modifying tcp’s congestion control for high speeds, 2002. citeseer.ist.psu.edu/floyd02modifying.html.
- [3] Sally Floyd. Highspeed tcp. <http://www.icir.org/floyd/hstcp.html>.
- [4] S. Ha, Y. Kim, L. Le, I. Rhee, and L. Xu. A step toward realistic evaluation of high-speed tcp protocols.
- [5] R. Jain, A. Duresi, and G. Babic. Throughput fairness index: An explanation. *ATM Forum/99-0045*, Feb. 1999.
- [6] C. Jin, D. Wei, and S. Low. Fast tcp: Motivation, architecture, algorithms, performance. In *Proceedings of IEEE INFOCOM*, 2004. citeseer.ist.psu.edu/jin04fast.html.
- [7] T. Kelly. Scalable tcp: Improving performance in highspeed wide area networks, 2003.
- [8] D. Leith and R. Shorten. Analysis and design of synchronised communication networks. *Automatica*, 41:725–730, 2005.
- [9] J. Mahdavi and S. Floyd. Tcp-friendly unicast rate-based flow control. January 1997.
- [10] V. Misra, W.B. Gong, and D. Towsley. Stochastic differential equation modeling and analysis of tcp window size behavior. In *IFIP WG7.3 Performance*, November 1999.
- [11] T.J. Ott, J. Kemperman, and M. Mathis. The stationary behavior of ideal TCP congestion avoidance. August 1996.
- [12] J. Padhye, V. Firoiu, and D. Towsley. Modeling TCP Reno performance: A simple model and its empirical validation. *IEEE/ACM Transactions on Networking*, 8(2):133–145, April 2000.
- [13] L. Xu, K. Harfoush, and I. Rhee. Binary Increase Congestion Control (BIC) for Fast Long-Distance Network. In *Proceedings of IEEE/Infocom*, Hong Kong, March 2004.