Chapter 7: Conclusions

7.1. Introduction

Chapter 1 presented a survey of congestion control techniques and a set of desirable characteristics of any congestion control scheme. In this chapter, we review the results from the preceding chapters, and examine the extent to which we have been successful in our design effort. We also examine the weaknesses in our work, and areas for future work.

7.2. Summary of the thesis

We define congestion as the loss of utility to a network client due to an overload in the network. This motivates the design of congestion control schemes, which allow users to gain utility from the network either by reserving resources to prevent overload, or by reacting to increased network loads. While congestion control can, and should, operate at a number of time scales concurrently, we restrict the scope of this thesis to reactive control schemes that operate at time scales ranging from less that one round trip time to a few round trip times. This corresponds to the design of a scheduling discipline that operates at all queueing points, and a transport level flow control protocol that is executed at all the hosts. In the course of the thesis, we describe and analyze the Fair Queueing scheduling discipline (FQ) and the Packet-Pair flow control protocol (PP), and claim that these mechanisms provide the required functionality and performance.

Chapter 2 introduced FQ as a way to provide a fair share of switch resources to competing conversations. We defined the notion of min-max fairness, and showed how this leads naturally to the FQ algorithm. A simple analysis considered the delay distribution of a Telnet conversation competing with FTP conversations at a FQ server. Chapter 3 studied data structures and algorithms for the efficient implementation of FQ, particularly the packet buffering scheme. We concluded that, if packet losses are few, then a simple ordered linked list is the best alternative. If there can be many losses, then a per-conversation linked list data structure is best. In Chapter 4, the traffic delinking property of FQ is used to build a deterministic model for a conversation in a network, and, from a series of lemmas, we derive the paired-packet probe. This is used to design the first version of the PP protocol, which constantly adjusts the data transmission rate to measured changes in the bandwidth-delay product. In Chapter 5 we recognize the inadequacies of the deterministic model, and propose a stochastic extension. This motivates a control theoretic approach to decide how best to use the series of packet pair probes to derive a stable flow control protocol. Practical issues, such as the unavailability of noise variances, motivated the design of a fuzzy prediction scheme. The resulting protocol allows users to obtain a desired delay-throughput tradeoff by choosing a setpoint, and dynamically adjusting the packet transmission rate so that the setpoint is maintained. As in all design problems, the proof of the pudding is in the eating. We showed in Chapter 6 that PP can match, or better, the performance of some widely known flow control schemes in a variety of scenarios, some of which are specifically designed to contradict our assumptions.

We now consider the role of FQ and PP as congestion control algorithms. The FQ scheduling discipline provides a number of advantages. First, it protects well-behaved users from ill-behaved ones, so that well-behaved users can get utility from the network even in the presence of ill-behaved or malicious users (this was the main purpose of the algorithm as proposed by Nagle [101]). Second, because it provides the equivalent of per-channel queueing, users can choose their own delay-throughput tradeoff. As described in §5.3.1, by choosing a setpoint for the queue size at the bottleneck, users can trade low delay for possible loss of throughput. Such tradeoffs are not possible with a FCFS discipline. (Even the selective DECbit protocol can only enforce a global tradeoff, because the FCFS discipline does not allow it to provide different users with different queueing delays.) Third, by partially delinking the traffic characteristics of the sources, FQ allows each user to probe the network state. This allows for sophisticated flow control that can use this information to choose an appropriate sending rate. Finally, FQ switches give incentives for sources to do more sophisticated flow control. As the simulation results for scenario 5 in Chapter 6 showed, sources that change to PP from JK or generic flow control get better performance.

The PP protocol leverages off FQ to do intelligent flow control. It too provides utility to users in several ways. First, it allows users to choose a bandwidth-delay tradeoff that corresponds to their utility function. A user's utility translates to a choice of setpoint, and PP ensures that the flow control tracks the setpoint to the extent allowed by control delays. Second, PP allows FTP sources to maintain their throughput even if the conversation has a large round trip propagation delay, or there is a lot of cross traffic. As we saw in scenario 8, the other popular flow control schemes we investigated do not perform too well under such circumstances. Finally, the protocol adapts quite rapidly to changes in the network state. Thus, even short periods where bandwidth is available at the bottleneck can be utilized.

Thus, both FQ and PP are successful congestion control schemes, in the sense that they allow a user to gain as much utility as possible from the network, even when it is overloaded.

7.3. Requirements re-examined

We now re-examine the seven requirements for a congestion control scheme as presented in Chapter 1, and see, using the analysis of Chapters 2-5 and the simulations of Chapter 6, to what extent these requirements have been met.

7.3.1. Efficiency

We desire a congestion control scheme to be efficient in two ways: not to consume excessive amounts of resources, such as network bandwidth and switch CPU time, and second, to allow the maximum possible utilization of network capacity.

PP does not place any overhead on the net amount of data transferred in the network, since additional probe or state-exchange packets are not used. Concerning the switch CPU time requirement, with PP, switches can be completely passive. Unlike the DECbit scheme, they need not set bits, nor compute averages. FQ implementation looks complicated at first glance, and it seems as if it may take a lot of switch CPU cycles. However, as we showed in Chapter 3, the additional overhead is not exceedingly large. With an efficient implementation scheme, we feel that the benefits of doing FQ outweigh its costs.

With the PP and FQ schemes, the network bandwidth is not underutilized (as it is, for example, with the DECbit schemes in Scenario 6). Since FQ is work conserving, it does not keep bandwidth idle if there are packets waiting to be served. Further, PP adjusts the packet transmission rate so that the bottleneck queue is never emptyl; hence a source obtains all the throughput allocated to it as its fair share. Indeed, we note that in Chapter 6, in all the scenarios, all the available bottleneck capacity is utilized.

Thus, we meet both the efficiency criteria mentioned in Chapter 1.

7.3.2. Heterogeneity

We require that the congestion control scheme accommodate heterogeneity in packet size, transport layer protocols and type of service requirements. FQ clearly allows for heterogeneity in packet sizes: indeed, that is one reason why we modified Round-Robin to obtain FQ. Second, by allowing users to choose their own delay-throughput tradeoffs, PP with FQ can accommodate a variety of service requirements. Finally, the results of scenario 5 indicate that PP can coexist with other protocols, and perform as well as, or better, than they can. Thus, we have satisfied our heterogeneity requirements.

The only caveat to the above is that, for PP to work correctly, every *potential* bottleneck must implement FQ or a similar round-robin-like service discipline. Even if a single bottleneck serves packets using FCFS, then the values reported by the probes will no longer be valid, and PP is no longer feasible. However, if the FCFS bottlenecks enforce some sort of rate control, and can stamp packets with the current service rate, then PP can be salvaged. This restriction is a major hurdle to the implementation of PP in current, FCFS networks.

7.3.3. Ability to deal with ill-behaved sources

Scenario 4 adequately answers this requirement.

7.3.4. Stability

In Chapter 5 we precisely defined the stability of a flow control protocol. Given the state equation that describes the dynamics of the length of the queue at the bottleneck, a flow control protocol is stable if the queue length remains bounded. This is true if the eigenvalues of the discrete time state equation lie in the unit circle, or the eigenvalues of the continuous time state equation lie in the left half plane. In the chapter, we formally proved the stability of PP. We are not aware of stability proofs for any other congestion control scheme in the literature.

7.3.5. Scalability

Scaling is required along two axes: bandwidth and network size. Higher bandwidths translate to larger bandwidth-delay products, so that sources can no longer ignore the propagation delay in doing flow control. Both our deterministic and our stochastic model explicitly model the control delay, and the PP scheme, which is based on these models, is designed to work in environments with large delays. The effectiveness of PP in networks with large propagation delays is seen in the results of scenarios 7 and 8, where, even with large control delays, PP sources behave well.

As the network increases in size, far more conversations are served at each switch. How well do our schemes cope with this? Chapter 3 showed how to implement efficient data structures to buffer data from many conversations. We believe that implementing these data structures in hardware is feasible. Special purpose hardware will allow FQ to serve a large number of conversations at high speeds.

A large number of conversations also means that the service rate fluctuations for any single conversation are smoothed out. PP is specifically designed for this environment, and, in fact, its performance will only improve as the number of users increases. Thus, we claim that our congestion control scheme scales well along both axes.

7.3.6. Simplicity

The simplicity requirement is that a congestion control scheme be easy to specify and implement. Both FQ and PP are conceptually simple, and can each be implemented in under two pages of C code. The control law of Chapter 5 translates to a single line of code, and the entire fuzzy controller takes about 20 lines of code.

One possible complication with PP as we presented it is that is requires one timer per packet. However, this can be changed to a single timer per conversation with little loss of performance. As a packet is transmitted, this timer is set to the packet's timeout value. When the timer expires, the last unacknowledged packet is retransmitted. With this scheme, the timeout interval is a little larger than the correct value, but there is a substantial savings in implementation cost.

FQ conceptually requires per-packet queueing, but, as we saw in Chapter 3, this can be implemented by a single ordered linked list. In any case, with hardware support, we expect that even per-conversation queueing can be implemented at high speeds (we are aware of one implementation that enqueues and dequeues ATM cells at 1.2 Gbps).

These figures, however, ignore the realities of protocol implementation, particularly in the Unix kernel. We are aware of the complexity of implementing a protocol at the kernel level, but there is nothing specific to PP or FQ that makes it any harder to implement in the kernel than any other equivalent protocol.

7.3.7. Fairness

FQ, as its name signifies, makes an effort to deliver min-max fairness to all the conversations that it serves. Fairness, however, requires that the flow control protocol be slightly sophisticated about managing its buffers. We saw that the generic flow control protocol does not allow for fair bandwidth allocation even with FQ. However, in all the simulated scenarios, the PP/FQ protocol pair provides fair (or nearly fair) bandwidth allocations to the sources. Thus, we claim that PP/FQ is a fair congestion control scheme.

We conclude that our schemes, with some minor caveats, satisfy the requirements raised in the first chapter. This is not true for a majority of the congestion control schemes in the literature.

7.4. Comments on design methodology

The success of our schemes owes mainly to our design methodology. We believe that, while one should not ignore practical problems, congestion control schemes must be based on a sound theoretical basis. This section presents the mapping from theory to practice that underlies our design effort.

Fair queueing is based on the principle of min-max fairness. The definition of min-max fairness immediately points to a bit-by-bit-round-robin (BR) scheme as a mechanism to obtain it, and FQ can be considered to be a practical implementation of this impractical ideal. This theoretical background makes FQ robust in the face of ill-behaved users.

The similarity of a BR scheme to time-division multiplexing motivates the deterministic model of Chapter 4. Once the model is stated, a little insight yields the packet-pair probe (this approach to the packet-pair scheme was first introduced by Prof. S. Singh and Prof. A. Agrawala [129]). This gives us a clean way to probe network state.

Having recognized that the network state may change, we see the need for a formal basis to express the dynamics of the network state. This basis is provided by control theory, and the control law used by PP is derived from a straightforward application of principles of predictive control.

The need for state estimation to implement control laws is a well known problem [49]. While Kalman estimators, and other related estimators, are theoretically adequate, it is increasingly recognized that, for a large class of practical applications, fuzzy logic has an important role to play. We use some fundamental principles of fuzzy control to build the fuzzy predictor described in Chapter 5. As the simulations in Chapter 6 show, this grounding in theory greatly helps the protocols in practice.

7.5. Contributions

This dissertation has made a number of contributions to the area of reactive congestion control. We review some of them in this section.

Our main contribution lies in the design of the Fair Queueing discipline and the Packet-Pair flow control protocol. (The design of FQ was joint work with S. Shenker and A. Demers.) We believe that both schemes, though conceptually simple, have several interesting features that will make them suitable for networks of the future. The two work together to provide better congestion control than some other schemes which are widely implemented in current networks.

Besides the schemes themselves, our other contributions are in developing deterministic and stochastic models for a conversation in a network, developing a control-theoretic approach to flow control in networks of FQ servers, and design of a fuzzy prediction algorithm.

Our use of deterministic modeling (which is joint work with Prof. S. Singh and Prof. A. Agrawala) makes the exact analysis of transient queueing phenomena possible. Though the model is naive, the stochastic version of the model allows for a formal control theoretic approach to flow control more easily than an equivalent stochastic queueing model.

Control theoretic approaches to flow control have been studied earlier for single M/M/1 servers and for Jacksonian networks. Our contribution lies in carrying out the analysis for a deterministic queueing model with propagation delays. The resulting protocol was shown to be stable. Further, we have implemented the resulting protocol in a realistic network simulator, and have done extensive simulations to study its behavior in a variety of benchmark scenarios.

We recognize the importance of state estimation in a distributed system with propagation delays. While we did derive the optimal Kalman estimator for the system state, we feel that such an approach is impractical for flow control protocols. Instead, our fuzzy prediction technique provides a practical alternative that performs well, and requires no additional information from the system.

7.6. Weaknesses and areas for future work

While we believe that our work has several claims to success, there are some weaknesses as well. Our theoretical models, though adequate for our purposes, are rather naive. We believe that even better results can be achieved by using more sophisticated models for the analysis. Thus, this thesis is only a small step in providing practical solutions to congestion control. What is heartening is that, even with these naive models, much can still be achieved.

Second, the simulations in Chapter 6, though extensive, are still far from a complete study of protocol behavior. While our choice of benchmark scenarios tries to test for several aspects of congestion control schemes, there are surely other aspects that we have overlooked. This is an area where much additional work is needed.

Finally, this thesis ignores two major dimensions of congestion control. To begin with, we do not consider predictive control. However, this is considered in detail in a contemporaneous thesis by D. Verma [141]. Also, we have ignored congestion control on time scales larger than multiple RTTs. Studying issues on larger time scales, and integrating the solutions into a single control scheme, is an avenue for future work. Other areas for future work are mentioned at the end of each chapter.

In conclusion, we feel that the area of congestion control is vast, and still in its infancy. We hope that this thesis makes a contribution to the field.