Chapter 1

Introduction

The control of congestion in connectionless packet-switched wide-area networks is a major problem, especially in such networks as the Internet, which is experiencing an exponential growth in users and network traffic. This thesis outlines a rate-based framework for congestion control in these networks, examines the requirements of the framework, and describes a number of control mechanisms which meet the framework's requirements.

The first section (Chapters 1 to 3) describes the problem of congestion in these networks, and reviews the current methods employed to deal with congestion, plus other solutions that are described in the literature. The second section (Chapters 4 to 7) outlines my proposed rated-based congestion control framework, and describes its ramifications and parameters. Finally, the framework is tested in several network simulations where it is compared against some existing methods, and the third section (Chapters 8 to 11) gives experimental results and summarises them.

1.1 Connectionless Packet-switched Networks

Most computer networks suffer from network congestion to some extent. This thesis limits its scope to congestion in *connectionless packet-switched wide-area networks*, and in particular, to networks with the same architecture as the Internet.

The network architecture under consideration has the following characteristics:

1. The network is a *wide-area network* with nodes in the network passing data to other nodes along links. The nodes are *arbitrarily* connected; the network does not have a particular topology (e.g spanning tree, hypercube etc.).
2. The network is *connectionless*. There is no reservation of network bandwidth or resources between a source of data transmission and its intended destination. End-to-end "connections" may be set up by the Transport Layers in the source and destination, but the other nodes within the network are oblivious to these Transport Layer operations.
3. The network is *packet-switched*. Packets are individually routed from source to destination. They may take different routes, and may arrive at the destination out-of-order.
4. Variable-sized packets are routed from one link to the next via nodes, often known as *packet switches*, *gateways*, or more commonly, *routers*. Routers usually buffer in-

coming packets before they are transmitted on an outgoing link. This is described in greater detail in the next section.

5. The network uses *semi-static routing*. Routes can change slowly over time, but most routes are static.
6. The network sets *no bandwidth constraints* on data sources. A source host may attempt to transmit data at a rate which will exceed the bandwidth or resources on the links and nodes between it and the destination host.
7. Links are assumed to have fixed bandwidths with no service guarantees. There is no expectation of error- or flow-control on any link.
8. The network makes *no transmission guarantee*. The links on the network have finite bandwidth, and the nodes have finite buffer space for packets waiting to be routed. If network components are unable to forward data on to the destination — for any reason — the data may be discarded. This type of network is also known as a *best effort* network.
9. The network is an *internetwork*. The nodes in the network are connected by links that vary greatly in type. There may be large variations in:
   - The physical characteristics of the links,
   - The Medium Access Control mechanism of the links,
   - The bit rate on the links,
   - The bit error rate on the links, and
   - The maximum packet size on the links.
10. The bandwidth on each link is generally static. It usually changes only when upgrades to the network infrastructure are performed.

1.2 Router Operation

All the links on a network are joined together by routers. These forward packets arriving on incoming links to the appropriate outgoing links, to ensure that the packet is routed to its intended destination. Figure 1 shows the basic architecture of a router.

The router is connected to I incoming links and O outgoing links. I and O may be different, although I usually equals O. In most situations, input and output links are paired to form either a full-duplex channel where data can flow in both directions simultaneously, or a half-duplex channel where data can flow in only one direction at a time.

Incoming packets are buffered in the input buffers. Once in the buffer, they are selected by the Packet Selection Function to be passed to the Routing Function. The Routing Function determines on to which outgoing link a packet must be placed for it to get closer to its destination: this is done with a routing table which, as described earlier, is semi-static. When the correct link for the packet is found, the Routing Function passes the packet to the Packet Dropping Function for queueing on the output buffer for the link. When the packet reaches the other end of the queue, it is transmitted over the link to the next router, or to the final destination.

The Packet Selection Function can choose any of the packets queued in the input buffers for routing by the Routing Function. Normally this would be done in a First-In First-Out method, but other selection methods may be useful in a congested environment.

There are two fundamental bottlenecks in the given router architecture. Firstly, there is a minimum time needed for the router to decode the network header of the incoming

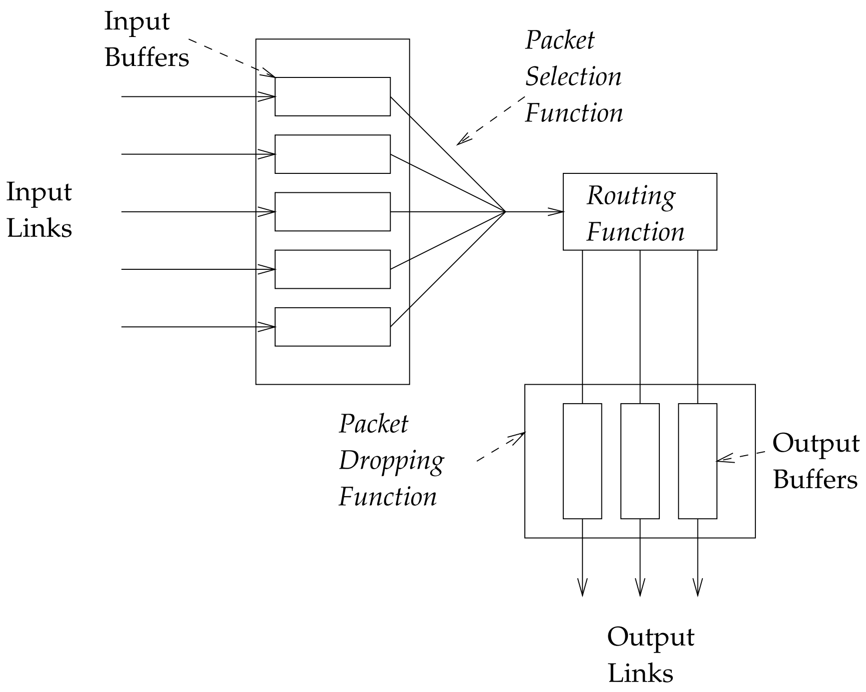

Figure 1: Basic Router Architecture

packet, determine the route for the packet, and pass the packet to an outgoing link for transmission. There is also a delay for the packet to be transmitted on the outgoing link: this may just be the packet's transmission time for a full-duplex link, or there may be an extra delay for the link to become clear on a half-duplex link. These delays form one bottleneck.

The second bottleneck indicates that the router must be prepared to buffer output packets, to prevent them from being lost while the router waits for an outgoing link to become clear. The two bottlenecks together indicate that the router must buffer incoming packets, to prevent them from being lost if they arrive too quickly for the router to process.

By definition, the router's input and output buffers are finite. If a buffer becomes full, then no more packets can be queued in the buffer, and the router has no choice but to discard them. This causes data loss in the data flow between a source and destination, and usually causes the source to retransmit the data.

Although a router cannot queue a packet if the corresponding output buffer is full, it can choose either to discard the unqueued packet, or to discard a packet already in the output queue, and then queue the unqueued packet. This choice is performed by the Packet Dropping Function. This cannot be done for the input buffers, as the packet has not been placed in the router's internal memory until it has been queued in the input buffers. Thus, packets may be lost on input, and the router has no control over which packets are lost.

Finally, the router's buffers may share a common memory space, or they may have individual memory spaces. With the former, no buffer can become full until the router's entire memory space is allocated to buffers, at which point *all* buffers become full. With the latter, any buffer's utilisation has no influence on any other buffer's utilisation.

1.3 Congestion in Connectionless Packet-switched Networks

A network is congested when one or more network components must discard packets due to lack of buffer space. Given the above architecture, it is possible to see how network congestion can occur. A source of data flow on the network cannot reserve bandwidth across the network to its data's destination. It, therefore, is unable to determine what rate of data flow can be sustained between it and the destination.

If a source transmits data at a rate too high to be sustained between it and the destination, one or more routers will begin to queue the packets in their buffers. If the queueing continues, the buffers will become full and packets from the source will be discarded, causing data loss. If the source is attempting to guarantee transmission reliability, retransmission of data and increased transmission time between the source and the destination is the result. Figure 2 from [Jain & Ramakrishnan 88] demonstrates the problem of congestion.

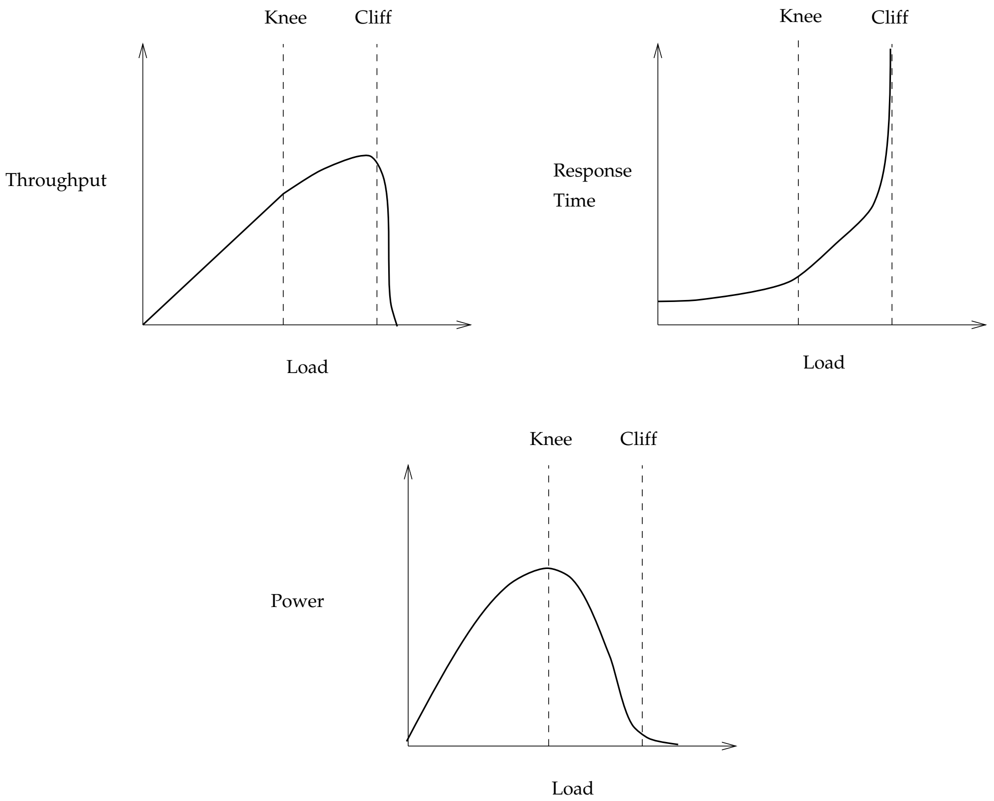

Figure 2: Network Performance as a Function of Load [Jain & Ramakrishnan 88]

As the load (rate of data transmitted) through the network increases, the throughput (rate of data reaching the destination) increases linearly. However, as the load reaches the network's capacity, the buffers in the routers begin to fill. This increases the response time (time for data to traverse the network between source and destination) and lowers the throughput.

Once the routers' buffers begin to overflow, packet loss occurs. Increases in load beyond this point increase the probability of packet loss. Under extreme load, response time approaches infinity and the throughput approaches zero; this is the *point of congestion collapse*. This point is known as the *cliff* due to the extreme drop in throughput. Figure 2 also shows a plot of power, defined as the ratio of throughput to response time. The power peaks at the *knee* of the figure.

1.4 Congestion Control Schemes

Network congestion control schemes can be divided into two categories. *Congestion avoidance* schemes attempt to keep the network operating at the knee of Figure 2. *Congestion recovery* schemes attempt to keep the network operating to the left of the cliff in Figure 2. A congestion avoidance scheme attempts to keep the load on the network to a level where response time is low, and power is optimised. A congestion recovery scheme attempts to keep the network operating by preventing the loss of data due to congestion.

Jain notes in [Jain & Ramakrishnan 88] that "Without congestion recovery a network may cease operating, whereas networks have been operating without congestion avoidance for a long time." Clearly, although congestion recovery prevents network collapse, it is better to operate the network at the point where power is maximised.

1.5 Factors Affecting Network Congestion

Network congestion is affected by the following factors:

1. The amount of traffic generated and placed on the network by its users;
2. The network architecture in terms of its links, their connectivity, and the characteristics of each link;
3. Operating parameters of sources, routers and destinations — such as available buffers, available processing power and system design;
4. Flow control mechanisms used in the Transport Layers of sources of data flow and their respective destinations;
5. Congestion avoidance and congestion recovery schemes in routers and sources;
6. Packet admission mechanisms in the Network and Link layers of sources and destinations;
7. Packet discarding and packet loss mechanisms in routers and destinations.

1.6 Scope of This Thesis

The scope of this thesis is constrained to the protocols and procedures used in the Transport and Network Layers of packet sources, routers and packet destinations of connectionless packet-switched wide-area networks. Therefore, the first three factors fall outside the scope of this thesis, and I will give them little or no coverage.

The next two chapters will review network congestion control schemes which have either been proposed or implemented. Following the thesis scope, only techniques operating at the Transport and Network Layers will be considered. The advantages and disadvantages of each mechanism will be discussed.

Given the deficiencies I will identify with many existing congestion control schemes, and the advantages of rated-based controls schemes that I will highlight, I will then present a rate-based congestion control framework which is designed for connectionless packet-switched networks.