Gateway design for computer network interconnection

David C. Walden and Randall D. Rettberg

Bolt Beranek and Newman Inc.
Cambridge, Massachusetts USA

Abstract

Issues associated with the interconnection of packet-switching networks via entities called gateways are discussed. A gateway virtual network is proposed, and a prototypical implementation is described.

1. Introduction

The work done to date on the interconnection of computer networks has usually assumed a configuration such as

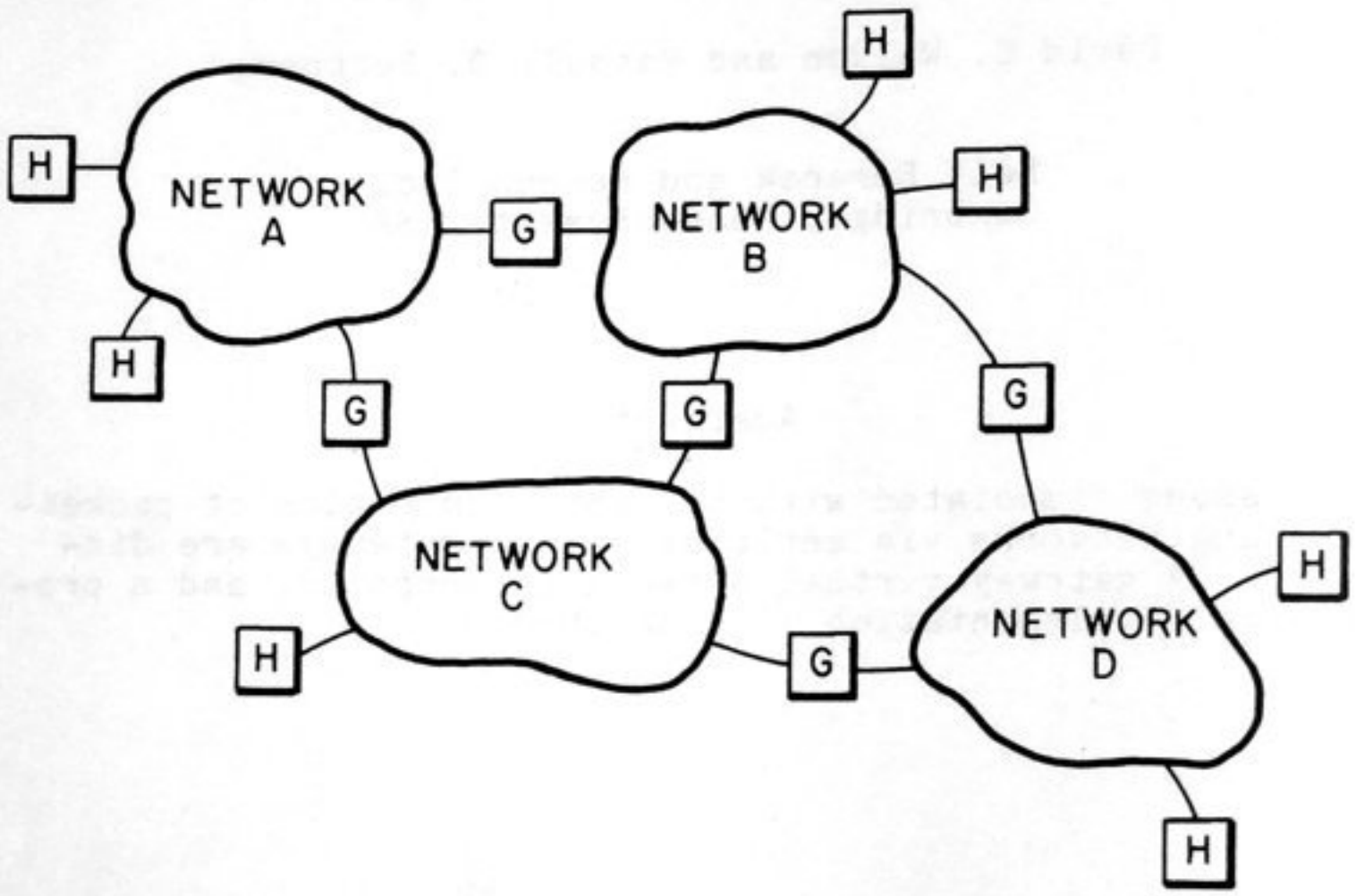

Figure 1 -- Networks Connected by Gateways

that shown in figure 1. (In this paper, our discussion refers only to networks of the the packet-switching variety, and assumes the reader to be familiar with packet-switching terminology.) On each network there are hosts (denoted by H in the figure) which desire to communicate with hosts on other networks. The networks are connected together by units (denoted by G) called "gateways."* The gateways must in some way convert traffic in the format of one network into traffic in the format of another network.

- - - - - - - - - -

*Note that this is a different use of the term "gateway" than in conventional international telephone system interconnection, where the term is used to refer to an artificial site from which it is convenient to establish tariffs).

Because host-to-host protocols differ from one network to the next, and because these protocols are generally complicated and incompatible, many researchers (including ourselves) believe that hosts on different networks wishing to communicate must do so in a common protocol. Much of the work to date in network interconnection has been to specify such a standard protocol: see for example [Cerf 74a], [Zimmerman 74], [McKenzie 74b]. While there has not yet been agreement on the standard protocol, for the purposes of this paper we assume the terminology and protocol described by Cerf and Kahn [Cerf 74a], [Cerf 74b]. In this protocol the logical entity in the host which performs the protocol functions is called the Transmission Control Program or TCP.

While there has been considerable work on the standard host protocol, there has been less work on the function and structure of the gateway. For the most part it has been assumed that the gateway will forward traffic between networks (and across networks) without specifying how this would be done. In this paper we consider the functions the gateway should perform and how it should perform them.**

2. Gateways as Hosts vs. Gateways as Nodes

One of the outstanding questions of network interconnection is whether the gateways should connect networks at the packet or host level. By packet level, we mean that a portion of the gateway would actually become a node on each of the networks being connected. By host level, we mean that a portion of the gateway would actually be a host on each of the networks being connected. We feel that the gateways should connect at the host level primarily to maintain the sovereignty of the networks involved [Crowther 72]. Furthermore, it is unlikely that a standard packet format can be found, at the present state of ongoing development of all of the packet-switching networks that might be connected, which would permit packet level connection.

By network sovereignty, we mean that connection to the networks must be done at a point where the interface is both well defined and well controlled by the constituent networks. If the point of connection is the host level just mentioned, each network can protect itself against activities of the gateway to the same extent as it may protect itself against the activities of any other host.

- - - - - - - - - -

**[Burchfiel 74], [Lloyd 75], and [Belloni 74] have also addressed this issue. The latter two of these references came to our attention late in the writing of this paper.

As already mentioned, we believe that it will be impossible, in general, for the gateway to convert between the host-to-host protocols of two communicating networks. Thus, rather than communicating in the host-to-host protocol of the network, the gateways should communicate with nodes of the network in the lowest form of host/network protocol supported by the network. Transmissions in the network interconnection protocol of the TCPs should be the text of these host/network protocol messages.

Notice that a host on a given network might find itself having to implement, in addition to the host-to-host protocol of its own network, the standard internetworking host-to-host protocol for communication with hosts in other networks. Of course, one can hope the internetwork standard will eventually prevail throughout the world and the host-to-host protocols of the individual networks will eventually wither away.

To summarize, hosts involved in internetwork communications must adopt a common protocol, and gateways should connect to networks as hosts using the lowest level of host/network protocol. Further, protocols have already been specified [e.g., Cerf 74a] for the former task and the protocol for the latter task is specified by the network for any network to which a gateway is to be attached [e.g., BBN 74].

If the gateways connect to the networks as hosts, then the format of the messages passed to the network is specified by the host/network protocol. This protocol is then used to permit transparent transmission of segments of an internetwork transmission. This can be done by embedding the internetwork segment in the text of a message in the host/network protocol as shown in figure 2. Such a composite message has two leaders and potentially two trailers. The outermost leader and trailer provide information for the network. The leader will specify the address of the gateway host to which the message should be delivered, any allocation or sequencing information which is used by the host/network protocol, and any further information demanded by that protocol. An example of a trailer that might be required by the host/network protocol would be padding and a checksum. Within this outermost leader and trailer is the internetwork data segment with its leader and trailer. The internetwork leader specifies such information as the ultimate destination, sequencing, and reassembly information. The actual data which is being transferred is the text of this message.

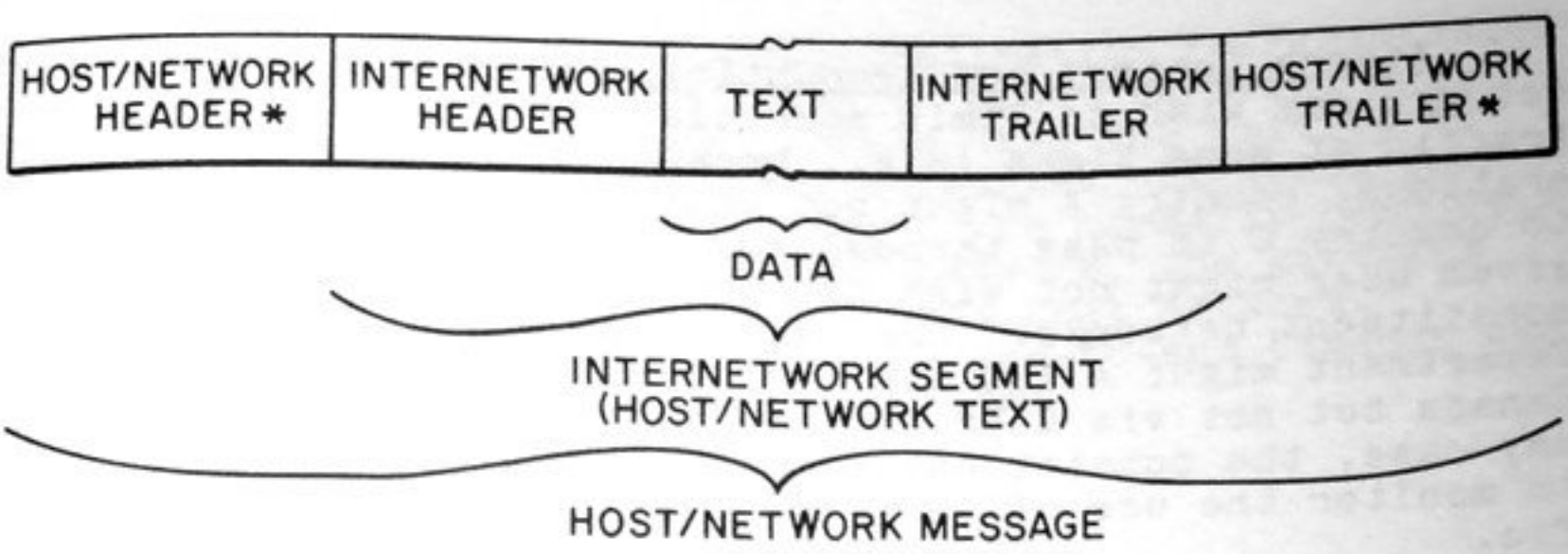

Figure 2 -- An Internetwork Segment Embedded in a Network Message

3. Gateway Characteristics

We will now examine some of the characteristics a gateway must have in addition to being able to pass messages between two networks. The gateway must have:

- -a capability for inter-gateway routing
- -access control and accounting mechanisms
- -a capability for fragmentation
- -control of congestion at the gateways
- -in some cases, a capability for inter-gateway retransmission

In the following paragraphs we elaborate on each of these points.

Routing. Inter-gateway routing is desirable for all of the standard reasons one desires routing. For example, for reliability one must have alternate paths over which traffic may be routed; for achieving higher bandwidth than is available over any single path, one wants to be able to route traffic over parallel paths; different classes of traffic should be able to follow different routes (e.g., traffic requiring low delay should be routed around networks which insert large delays).

Access Control and Accounting. A given constituent network may wish to limit some classes of traffic or all traffic at some times (e.g., because of regulatory considerations, country A might not want traffic from country B to country C to pass through country A's network). Also, a given user might not wish his traffic to pass through some constituent networks. For example, the U.S. Defense Department might allow its traffic to go to England via Canada but not via Cuba for obvious political reasons. In any case, the constituent networks are very likely to want to monitor the use of their network by internetwork traffic.

For efficiency of routing and access control, in large networks with hundreds of hosts and gateways, the routing algorithm will probably need to have a hierarchical structure knowing about logical and/or physical areas.

Fragmentation. Because of the differences in message size of the constituent networks connected by a gateway, the gateway must have the ability to fragment a larger message arriving from one network into smaller messages which are acceptable by the next network. When such fragmentation occurs, the message stream must eventually be reassembled into its original structure. The protocol proposed in [Cerf 74a] provides the reassembly function at the destination host.

Congestion Control. Congestion will inevitably occur at the gateway unless specific measures are taken to prevent it. This congestion can occur as a result of speed mismatches between the networks connected by a gateway, because several gateways on a network may simultaneously transmit traffic to the same other gateway, because traffic may have to be held during a period of recovery from a failure, and so on. One specific kind of congestion results from deadlocks, such as when gateway A is full of traffic for gateway B which is full of traffic for gateway A.

Retransmission. When a message is lost in the network between two gateways, one can either retransmit the message between the two gateways or assume that the message will be retransmitted from the source host to the destination host. It has been shown [NAC 73] that hop-to-hop retransmission is more efficient than source-to-destination retransmission if the possibility of message loss is appreciable; and even when there is little possibility of message loss, the variance of retransmission delays is less with hop-to-hop retransmission than with source-to-destination retransmission. While some networks deliver messages very reliably,

other networks rely on source-to-destination retransmission and in some cases are quite cavalier about throwing away messages. Thus, the gateways should have the ability to retransmit messages across lossy networks. It seems that at least when the hosts on a network are normally responsible for retransmission across that network, the gateways ought to provide retransmission across that network. ([Mader 74] supports the notion that end-to-end timeout and retransmission can be unduly inefficient.)

4. The Gateway Virtual Network

Notice that the characteristics of a gateway described above are very similar to the characteristics of a node on a packet-switching network [Crowther 75]. This leads one to the notion of a gateway virtual network wherein the gateways act as nodes and the network spanned by the gateways acts as virtual lines fully connecting all the gateways on that network. This concept is illustrated in figure 3, where network C has been replaced by the virtual lines it provides. Further, notice in the figure that logically there is a gateway associated with each host attempting internetwork communication.

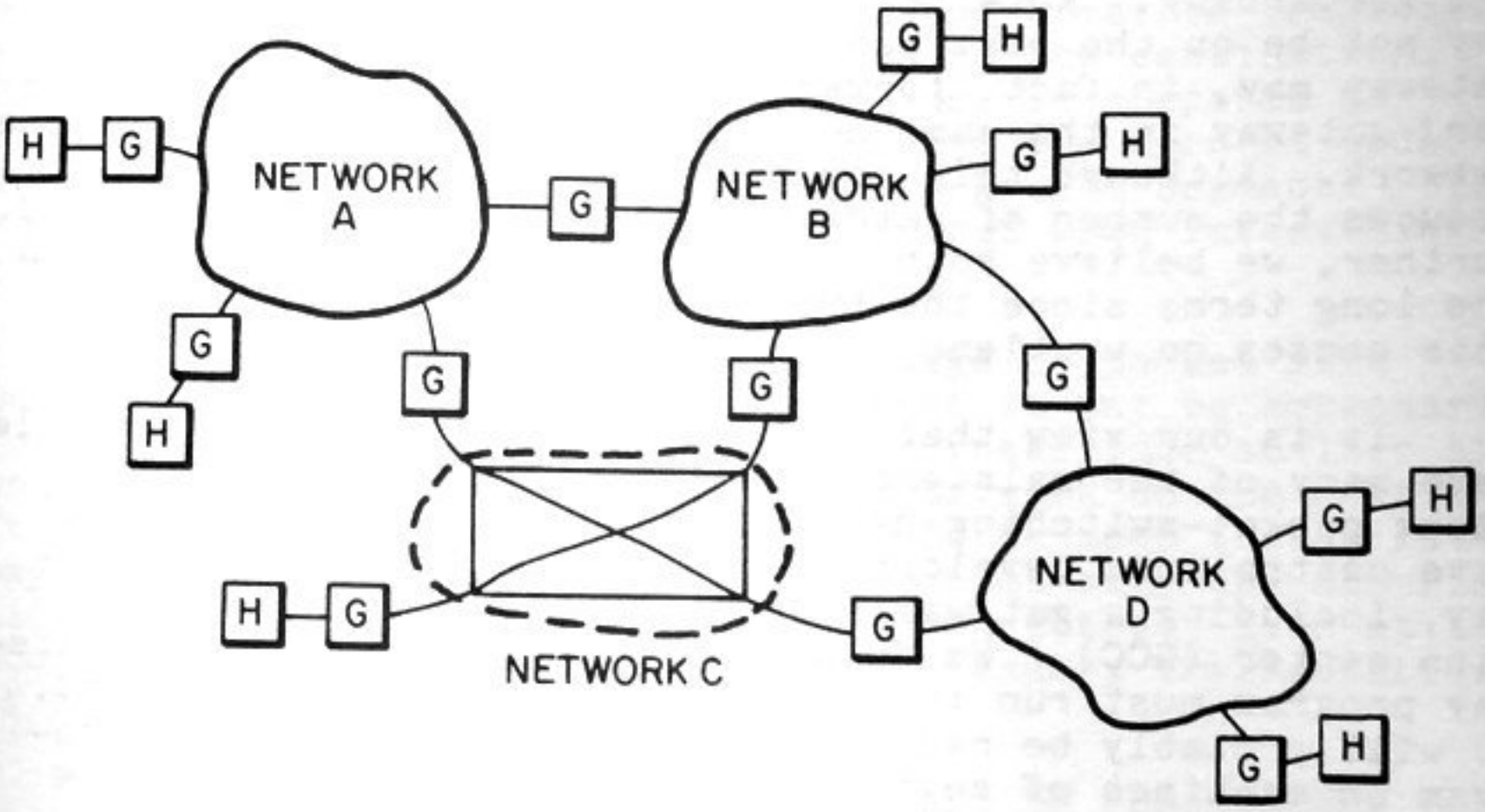

Figure 3 -- Gateway Virtual Network

Although the figure shows a separate gateway for each host, we do not mean to imply that the gateways must be physically separate machines or that there need necessarily be a one-to-one correspondence between hosts and gateways. For instance, the logical entity that is the gateway may take the form of a program running in a host computer. Alternately, the gateway could be in a stand-alone machine connecting two networks or serving one or more hosts. On the other hand the gateway connecting two networks could even take the form of a program running in a host which is connected to both networks. In general, a gateway should be able to connect any combination of hosts and networks. (The political and economic issues relating to whether the gateway should reside in a host or a separate gateway machine are discussed in [Kuo 75].) Furthermore, it is entirely possible for a host not to have its own gateway at all, preferring to use a gateway elsewhere in its network to perform the gateway functions for the host. In this case, the host would simply know the address of a couple of gateways in its network and would send its internetwork traffic arbitrarily to one of these gateways for routing and forwarding. Note that the gateway arbitrarily chosen may not be on the best path to the destination, and this gateway may, in fact, forward the traffic to another (better) gateway in the same network for forwarding outside the network. Although this approach may be inefficient, it reduces the number of gateways that have to be constructed. Further, we believe both approaches should be supported in the long term; since the two approaches are compatible, this causes no problems.

It is our view that the gateway virtual network should have many of the maintenance characteristics of a stand-alone packet-switching network [McKenzie 72]. It should have centralized development and maintenance responsibility, including a gateway network monitoring and coordination center (GCC). We do not, however, feel that the gateway program must run in any one brand of machine; indeed, it will probably be necessary to support the gateway program on machines of several nationalities because of the international extent of these networks.

The gateway virtual network is a general solution to the problem of interconnecting networks which is not highly dependent on the nature of the networks being connected. Because of this we expect that networks which are connected in the future will not require modification to facilitate connection. While one can conceive that a simpler but less general form of interconnection may be possible through modification of the current networks, future networks may be much less amenable to such an attachment and therefore

require major modifications to themselves or even to the previously interconnected networks. Furthermore, the gateway virtual network, with its adaptive routing, congestion control, and gateway-to-gateway retransmission capability provides for high performance host-to-host communication across multiple networks (note that this approach is in contrast to [Opderbeck 74], which advocates sacrificing such efficiency for simplicity).

5. Division of Functions Between the Gateway and TCP

Just as in a stand-alone network there is a question of the division of responsibility between the hosts and the nodes (i.e., should reassembly be done by the hosts or should the nodes deliver traffic in order), in the gateway virtual network there is a question of the division of responsibility between the TCP and the gateway. It is best to implement these functions in either the gateway or the TCP, rather than blurring the implementation across the boundary between them. Some points are quite clear. For instance, reassembly must be done by the destination TCP, as traffic traversing the virtual network may be fragmented and pieces routed on alternate paths to the destination. Thus, the ultimate destination TCP must be prepared to rearrange the communication stream into the correct order and reassemble it into the internetwork data segment. (The reassembly problem is further aggravated when intervening networks do not maintain ordering.)

Second, the differences in message or packet size between constituent networks mean that it may be necessary for messages to be successively fragmented into smaller and smaller units as they pass from one network to the next. It seems natural that this task should be done in the gateway since it knows the message and packet size characteristics of the networks to which it is connected. Even at the source host, the TCP can leave any necessary fragmentation of the message stream to the gateway.

Just as clearly, the gateways must be responsible for the routing calculation, since only the gateways have the global knowledge necessary to make a sensible routing decision.

The access control and accounting functions should also be in the gateway since these functions are desired between the networks.

It is less clear whether the gateways or TCPs should perform the congestion control and retransmission func-

tions. One alternative would be for the gateways not to worry about retransmission and to solve any congestion problems simply by discarding traffic, in each case relying on the source and destination TCPs to provide the necessary recovery mechanisms ([Belsnes] advocates this). In fact, the TCP of [Cerf 74a] does provide such recovery mechanisms. On the other hand, if much traffic is discarded by the gateways to control congestion, or if there is even one network which loses traffic frequently, then we believe relying on source-to-destination TCP retransmission will be prohibitively inefficient and will also be expensive, both in direct (network-imposed accounting) and indirect (TCP overhead) costs. For this reason we think it is incumbent on the gateways to shoulder the burden of controlling their own congestion and for the gateways to provide the option of retransmission across a lossy network. Of course, even though the gateways provide these functions, the TCPs should retain the end-to-end retransmission capability at their level for reliability, since retransmissions performed by the gateway level are for efficiency rather than for complete reliability.

The remainder of the functions performed by the [Cerf 74a] TCP are properly the functions of the TCP and not the gateways, as these functions are concerned with end-to-end issues, user process level issues, etc., while the gateways are properly concerned only with traffic switching issues.

There does have to be some communication between the TCP level and the gateway level. Most obviously, the TCP must specify the address of the destination TCP to which the gateways are to route the traffic. This particular communication can be effected simply by having the gateways understand the TCP traffic formats which include such addressing information. Another area of communication required between the TCPs and gateways is to specify certain transmission characteristics for the traffic (e.g., networks through which the traffic must not be routed, maximum acceptable delays, and average throughput required over a period of time). This area has not been explicitly addressed previously and requires further study.

At present we have no opinion on whether the security function should go in the gateway or the TCP. Current U.S. military communications security standards appear to require pairwise encryption between source host and destination host. If such standards are maintained, then the security function must obviously reside with the TCPs. Alternatively, there has been much discussion of so-called "link encryption." Use of link encryption would lead one to place the security function in the gateways. A third

alternative would be to place the security function between the source and destination TCPs and their gateways. In this case the TCP "header" must be in the clear, permitting the gateways to access the information. There are elaborations of all the above schemes and other schemes are possible.

6. A Prototypical Implementation

A prototypical implementation of the gateway is illustrated below. It consists of modular structures which carry on communications at a message level with various networks, modules which transform a particular network message format into a TCP type packet and vice versa, and modules which perform message routing and other centralized message processing functions such as flow and access control.

The purpose of each of these elements of the gateway will be clear if we follow a message as it is processed by

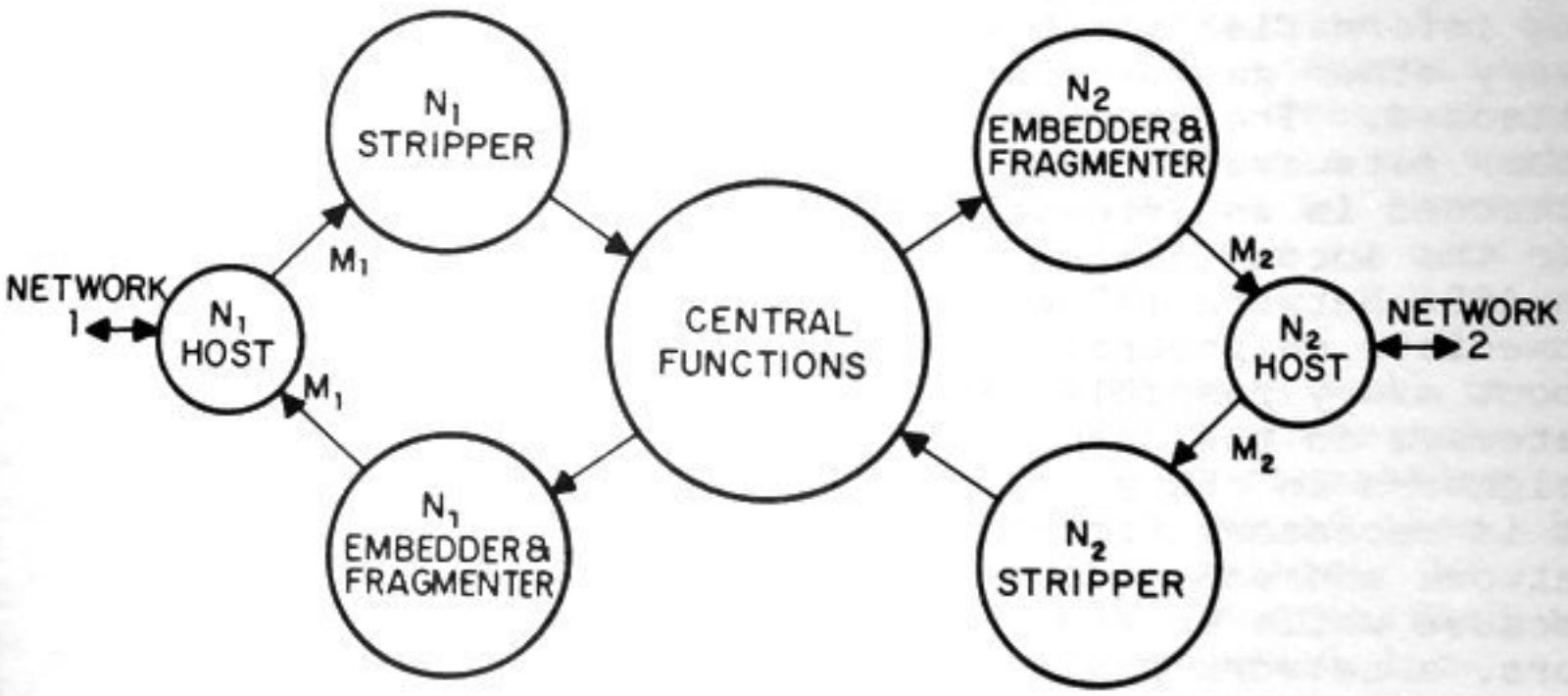

Figure 4 -- Components of a Gateway

the gateway from Network 1 to Network 2 in figure 4. The line to Network 1 is operated by the code labelled "N1 Host". This code acts as a host on that network, passing the required control information and transferring messages to and from other hosts on that network according to the lowest level message protocols of that network. In the case of a connection to the ARPA Network, this code would

implement the standard Host/IMP protocol [BBN 74]. Incoming messages received from Network 1 are passed to the "Network 1 Stripper," where the Network 1 message leader is removed. The result is an internetwork protocol segment. This segment is passed to the central portion of the gateway, which examines the address specified in the segment and queues it for output to that network. The "Network 2 Embedder and Fragmenter" takes these internetwork segments and converts them into messages in the Network 2 format. In this case, once again, the format will be ARPA Network messages. If the internetwork segment does not fit within the text of a network message, it will be fragmented and re-formatted until all parts fit into network messages. Finally, the message is presented to the Network 2 message interface for delivery to the appropriate gateway host.

The control function will be responsible both for determining where to route an individual internetwork segment and for performing the distributed routing functions in cooperation with the other gateway nodes. In operation, this distributed routing can be accomplished in the following way. Each gateway exchanges routing packets, containing information about all gateways in all networks, with every other gateway on each network to which it is attached. The gateway is given the host addresses of all other gateways on each of the networks to which it is attached in an internal table which will be used eventually for the access control logic. This is similar to the way an ARPA Network IMP exchanges routing packets with its immediate neighbors, where the packets contain information about every possible destination [McQuillan 74]; the other gateways on the same network are logically the immediate neighbors in the gateway network. In this case, however, it is necessary for each gateway to contain internally the network address of the other gateways in the same network, because while an ARPA Network IMP has at most five neighbors, a network has potentially many hundreds of hosts, a large fraction of which may be gateways.

The information passed in the routing messages can contain at least three types of measurement about the network: delay, bandwidth, and delay variability. For instance, gateways can accumulate delay and estimate bandwidth for the network connection between each pair of gateways in the same network. The gateways can then maintain a structure similar to that used by the ARPA IMP for routing. For each potential destination gateway, the intermediate gateway to which messages should be addressed for minimum delay and maximum bandwidth can be determined along with the expected delay and bandwidth. Periodically, the gateway reports to each of the other gateways on the same net-

work the expected delay and bandwidth via this gateway. The other gateway then adds the delay to that gateway and determines the minimum of the bandwidth through that gateway and the bandwidth on the path to that gateway, to determine the delay and bandwidth which may be expected along this path to the destination. Through this mechanism, each gateway can have the necessary information to route packets. Priority letters could be routed for delay, while others could be routed for bandwidth.

We recognize that this sample routing scheme suffers from an inability to accurately determine the expected delay and bandwidth along a path, particularly as the actual path across a network may change unbeknownst to the gateways. This is intrinsic to the implementation of gateways as hosts rather than as nodes, since only the nodes of a network normally have the necessary state information (e.g., line speeds between node pairs or the number of packets queued for a line), and so far, no network passes this information to its hosts. We believe that the reasons mentioned previously in this note for connecting gateways at the host level are very important, and that the way to improve the routing efficiently between gateways is for the networks to pass expected delay and bandwidth information, for example, to their hosts. Even without the issue of gateway routing, it might be useful for hosts on a network to receive such delay and bandwidth information to improve their use of the network (e.g., deciding whether to send a large file now or later).

We have suggested that in general a gateway will be required in association with each TCP. Since these gateways form a network of their own, routing and formatting packets, and since there may eventually be many of them, it is advisable to keep them as identical as possible. We therefore suggest that the gateway routines be specified in some reasonably universally available higher level language. If possible, the gateway code should be machine compiled from that specification. Otherwise, the routines can be hand compiled from the higher level language specification. We hope that in this way, the implementation effort and variability will be minimized in implementations at the various hosts.

7. Backwards Compatibility

Although gateways permit the interconnection of networks and communication across network boundaries, the ability of a host on a foreign network to use the facilities of a local network is limited by the form of network

interconnection used. Since all communication across network boundaries, as described in previous sections of this paper, must be in the form of TCP segments, it is impossible for a foreign host to participate in the same levels of communication in which a local host might participate. There are some instances where it is desirable to permit a foreign host to communicate at (or near) the host/network protocol level. For example, a local host may communicate with other, less powerful hosts in the lowest level of host/network protocol, but a foreign host cannot communicate with those hosts no matter how hard it tries or no matter how willing it is to accept inefficiencies in the communications. In this case, the less powerful host may be provided by the network itself (such as the statistics hosts in an ARPA Network IMP [BBN 74]), and it may be impossible to have it understand the internetwork protocol.

Remote use of the host/network protocol can be made possible by including a TCP in the gateway between networks. To send a message in the format of the destination network, the foreign host would embed the destination network message in an internetwork segment addressed to the TCP in a gateway on the destination network. At that gateway, the internetwork leader and trailer would be removed, leaving a message in the proper format for the destination network. The gateway would then insert any leader information necessary to identify the foreign source of the message so that replies to the message may be embedded in internetwork framing and returned to the foreign host.

To communicate with some host on the local network, the foreign host should set up a "mailbox" in a gateway to the local network through communication with the TCP in that gateway. The foreign host would then notify the local host of the address of the mailbox to be used. Messages originating in the local host would then be sent to the mailbox; the gateway would forward them to the foreign host.

This type of connection is much less efficient than the methods described in the previous sections; however, it provides a capability which is not available in any other way.

Acknowledgments

This work has been supported by the Information Processing Techniques Office of the Advanced Research Projects Agency (ARPA) of the U.S. Department of Defense under Contract No. F08606-73-C-0027 and Contract No. F08606-75-C-

0032. The ideas we have presented in this paper have evolved through interactions with a number of other workers in the field, particularly V. Cerf of Stanford, R. Kahn of ARPA, and R. Binder, J. Burchfiel, W. Crowther, N. Liaaen, A. McKenzie, and J. McQuillan, all of Bolt Beranek and Newman (BBN); and to them we are grateful. We are also grateful to R. Brooks of BBN for his help with the preparation of the manuscript for this paper.

References

BBN 74.
Bolt Beranek and Newman Inc., "Specifications for the Interconnection of a Host and an IMP," BBN Report 1822, revised December 1974.

Belloni 74.
A. Belloni, M. Bozzetti, and G. Le Moli, "Routing and Internetworking," International Working Group Protocol Note 10, August 1974 and February 1975.

Belsnes 74.
D. Belsnes, "Flow Control in Packet Switching Networks," INWG Note No. 63, October 1974.

Burchfiel 74.
J. Burchfiel and R. Tomlinson, "An Experimental Simulation of a Satellite Gateway," INWG Experiment Note 2, August 1974.

Cerf 74a.
V.G. Cerf and R.E. Kahn, "A Protocol for Packet Network Intercommunication," IEEE Transactions on Communications, Vol. COM-22 5, May 1974, pp. 637-648.

Cerf 74b.
V. Cerf, Y. Dalal, and C. Sunshine, "Specification of Internet Transmission Control Program," INWG Note 72, revised December 1974.

Crowther 72.
W.R. Crowther and D.C. Walden, "Response to INWG Note 6," INWG Note 10, December 1972.

Crowther 75.
W.R. Crowther, F.E. Heart, A.A. McKenzie, J.M. McQuillan, and D.C. Walden, "Issues in Packet-Switching Network Design," to be presented at the AFIPS 1975 National Computer Conference.

Kuo 75.
F.F. Kuo, "Political and Economic Issues for Internetwork Connections," Computer Communication Review, Vol. 5, No. 1, January 1975, pp. 32-34.

Lloyd 75.
D. Lloyd, M. Galland, and P. Kirstein, "Aim and Objectives of Internetwork Experiments," INWG Experiment Note 3, January 1975.

McKenzie 72.
A.A. McKenzie, B.P. Cosell, J.M. McQuillan, and M.J. Thrope, "The Network Control Center for the ARPA Network," Proceedings of the First International Conference on Computer Communications, pp. 185-191, October 1972.

McKenzie 74a.
A.A. McKenzie, "Some Computer Network Interconnection Issues," Proceedings of the AFIPS 1974 National Computer Conference, pp. 857-859, May 1974.

McKenzie 74b.
A.A. McKenzie, "Internetwork Host-to-Host Protocol," INWG Note 74, December 1974.

McQuillan 74.
J.M. McQuillan, "Adaptive Routing Algorithms for Distributed Computer Networks," BBN Report 2831, May 1974.

Mader 74.
E. Mader, W. Plummer, and R. Tomlinson, "A Protocol Experiment," INWG Experiment Note 1, August 1974.

NAC 73.
Network Analysis Corporation, "Comparison of Hop-by-Hop and End-to-End Acknowledgement Schemes," Packet Radio Temporary Note 7, January 1973.

Opderbeck 74.
H. Opderbeck and L. Kleinrock, "The Influence of Control Procedures on the Performance of Packet-Switching Networks," National Telecommunications Conference Proceedings, San Diego, December 1974.

Zimmerman 74.
H. Zimmerman and M. Elie, "Transport Protocol. Standard Host-Host Protocol for Heterogeneous Computer Networks," INWG Note 61, April 1974.