

SOURCE ROUTING FOR CAMPUS-WIDE INTERNET TRANSPORT

by

Jerome H. Saltzer

David P. Reed

David D. Clark

Massachusetts Institute of Technology

Laboratory for Computer Science

545 Technology Square

Cambridge, Massachusetts 02139

15 September 1980

A b s t r a c t: For the internet addressing layer of a campus-wide local area network, a source routing mechanism may have several advantages over universal or hierarchical target addresses. The campus environment requires many subnetworks connected by gateways, and it has relatively simple routing in this environment is simplicity of implementation of the gateways that interconnect subnetworks with consequent improvement in location, and overall management effort.

This preprint is of a paper presented at the IFIP Working Group 6.4 Workshop on Local Area Networks in Zurich, August 27-29, 1980. shop.

Introduction

This paper proposes that for the internet addressing layer of a campus-wide local area network, the source routing mechanism suggested may have several advantages over hop-by-hop routing schemes based on universal or hierarchical addresses. The campus environment requires many subnetworks connected by gateways, and it probably has a relatively loose administration. This research was supported by the United States Government and was monitored by the Office of Naval Research under Contract No. N00014-75-C-0661. The primary advantage of implementation of the gateways that interconnect subnetworks with consequent improvement in cost, maintenance effort, recovery time, and effort. Secondary advantages of source routing when applied to the campus environment include: 1) a clearer separation of physical address and protocol design, 2) elimination of stability, oscillation, and packet looping considerations, 3) ability for a source to control precisely a route (response time, reliability, bandwidth, usage policy, or privacy), 4) deferment to a higher protocol level of the detailed design of the fragment of intermediate networks with small maximum packet sizes, and finally, 5) the ability to accommodate both official and unofficial gateways between subnetworks.

Two disadvantages of source routing are: 1) that the route used will tend to be relatively static and therefore cannot optimize use of a dynamic hop-by-hop route selection system, and 2) route selection must be accomplished somehow, and since this protocol level does not have a dynamic route selection system, it must be designed to provide route selection. The argument made here is that the first disadvantage is not serious in an environment such as a campus where communication can make optimization less important. The second disadvantage may be less serious than it appears when one considers that in any case, which service can also provide route selection service. In fact it may be possible to turn this need into an advantage, since one of which is based on simple global or hierarchical network identifiers, while another, perhaps experimental or research service, provides a private route pattern.

This last ability to decouple target identifications and route selection from gateway implementation taken together with the other advantages in using source routes is improved modularity of network implementation.

This paper has three parts. The first explores the nature of the campus environment, especially its administrative properties. The second part discusses work using routing services. The final part discusses the advantages that source routing seems to provide when applied to the campus environment. Reading will find that the second section can quickly be skimmed; potentially novel observations are confined to the first and third sections.

I. What is a Campus Environment?

"The Campus Environment" is a name used here to identify a particular set of physical properties, geographical extents, data communication needs for flexibility that characterize our own university campus. With only minor exceptions they equally apply to a corporate site, a government installation, or a large research center. There will be seven characteristic properties of this campus environment that provide a basis for design decisions for a data communication network. These are quite different from those of a single building, or of a nation-wide, common-carrier-based network. The seven properties are:

- 1) limited geographical extent,
- 2) up to several thousand nodes,
- 3) forces for both commonality and diversity,
- 4) multiple protocols,
- 5) confederated administration,
- 6) independently administered interconnections, and
- 7) gateways to other nets.

The following sections explain and discuss each of these properties in turn.

- 1) Limited geographical extent. The campus environment has a geographical extent beyond a single building, but not so large that it requires transmission media to be installed without resort to a common carrier.

This first property is essential, so as to allow exploitation of low-cost, high-bandwidth communication technology. With current technology, communicating over privately installed equipment and using common carrier facilities can be a factor between 10 and 100.

- 2) Up to several thousand nodes. Within this geographical area, a large number of nodes--that is, computers--will be present. Today the number of such nodes may be in the range of ten to one hundred. Looking ahead to the advent of desktop computers, one might envision by 1990 as many as 100 subnetworks each comprising an average of, say, 50 to 100 nodes, thus linking up to 10,000 nodes by the end of the next decade.

The combination of the previous two properties seems to make it inevitable that local interconnection technologies such as the ETHERNET, PERCHANNEL [6], MITRENET [7], or the Cambridge Ring [8], cannot by themselves completely accomplish the required interconnection. These technologies have demonstrated have limitations on distance on the order of a thousand meters and limitations on node count on the order of a hundred nodes. One must attach clusters of nodes into subnetworks, for example all the nodes in a single building, and then install interconnections (gateways) between subnetworks. On a campus, one might envision by 1990 as many as 100 subnetworks each comprising an average of, say, 50 to 100 nodes, thus linking up to 10,000 nodes. This creates the problem of how to route a message from a source node through a series of subnetworks and gateways, so that it ends up at a desired target.

- 3) Forces for both commonality and diversity. Administratively, there exist forces both for commonality and diversity. The primary force for commonality is a desire to be able easily to set up communications between any pair of nodes on the campus. The primary force for diversity is that one technology, for example, data source, or data sink typically pre-determines the technology of the network to which it must be attached, because off-the-shelf network hardware is available in only one technology. Further, some applications may have special requirements for some connections (e.g., high bandwidth) that can be met only by one technology. Thus the emerging diversity of local networks will continue, and probably will increase.

- 4) Multiple protocols. Although there are several ongoing standardization efforts, the worldwide academic, commercial, and government communities have anything resembling a consensus on how networks should be organized, how protocols should be layered or how functions should be distributed. This is from obscure matters of taste, through fundamental technical disagreements about which requirements should have priority. The communication technology is moving. Many different and competing standards have been proposed, and one can find in the literature a wide variety of proposals. One must anticipate that these arguments will be reflected internally in the campus environment, in the form of a diversity of protocols. One must also anticipate that any mutually consenting* set of nodes be able to carry on communication with one another using a protocol that is common to all.

* Imagery borrowed from a Chaosnet working paper by David Moon.

The diversity of protocols arises for much the same reason as the previously-described diversity of network hardware. The foremost reason is that the network must meet the immediate application requirement. Ability to communicate with other, not-yet-integrated, applications is a low priority consideration. In the absence of a particular supplier's collection of provincial protocols, the purchaser will tend to question them only if it appears that they might hamper the application. One must postpone thinking about interconnection until some real requirement appears, and after the equipment and its protocols have been installed.

This protocol diversity suggests strongly that any network interconnection strategy that must be implemented today should have at the least the capability to accomplish communication between any two nodes while making an absolute minimum number of assumptions about the higher-level network architecture. Some typical assumptions that should be avoided unless an unusual opportunity can be taken are: wide area networking; routing should be optimized; fragmentation/reassembly strategy; flow control requirements; addressing plan; and particular network topology.

- 5) Confederated administration. Because a data communication network is a campus-wide service, there is a strong interest to administer the entire network. This means that network administration will either be done by a haphazard confederation of departments, or by an underfunded central service organization modeled on the one whose role is to minimize telephone costs.

In either case, this property places a requirement on the network interconnection technology that it be robust and self-surviving to every user. The network must be easy to use, easy to accomplish and easy for individual users to participate in if they are so inclined, because trouble isolation and repair may involve multiple users. The network must be important, so that operation can be completely unattended for long stretches of time. Although some central monitoring of network operation is desirable, a design approach that requires close monitoring is undesirable.

- 6) Independently administered interconnections. The topology of subnetwork interconnection will be determined by the needs of the users without central administration.

This property arises from two needs: First, a "dependable" set of gateways that one can expect to exhibit predictable and stable properties. A centrally planned and administered set of gateways would provide this dependability. Second, whenever a node finds that for some reason it is a gateway, some of its applications to serve also as a gateway between the subnetworks; yet it may not want to take on the official responsibility of a gateway that is not centrally administered may arise if some particular application needs, and has purchased the gateway equipment to serve its own needs of delay, reliability, bandwidth, or privacy. The person or organization that has purchased the special gateway equipment may not be prepared to accept the requirement is that a user may wish to avoid use of a sometimes troublesome gateway that is claimed by its owner to be perfectly operating.

- 7) Gateways to other nets. External, public data networks such as TELENET, the ARPANET, TYMNET, X.25, and others will serve as gateways between the campus network and the external networks. In some cases, a "link" in the campus net. In other cases, facilities within the campus net will set up communication paths to services having no direct connection to the campus. Both kinds of cases require careful consideration of the interactions between internal and external network properties.

Note that the campus environment has all these properties only if we assume the technological opportunity mentioned in point one: that there be high-speed communication paths in the range from 1 to 10 Mbits/sec. between any two points within the campus. Availability of interconnect media and subnetworks in these forms. Gateways that operate with such bandwidths may be harder to construct, and that concern is one of the considerations involved in the design of the network. Networks that can sustain these data rates for very long are likely to be rare; software often limits the rate at which a node can act as either a data source or a data sink. The network must be exploited in two ways:

- 1) to provide enough capacity to handle the aggregate demand of many lower-bandwidth sources and sinks of
- 2) non-optimal strategies that are relatively simple to implement or administer can be considered; it is not a maximally utilized.

The availability of high bandwidth, together with lack of a requirement to use that bandwidth efficiently, is probably the most fundamental difference between the "local area network" and the commercial long-haul data communication network, a difference that can lead to significantly different design decisions.

II. How Source Routing Works

1. The basic mechanism. Source routing among a collection of subnetworks is a mechanism that comes into play at a relatively high layer.* Figure one illustrates this layer arrangement. If one tried to interpret a collection of interconnected subnetworks according to the "transport" layer. The lower layers, which we may collectively call the "local transport" layers, constitute a protocol stack within each subnetwork. A single ETHERNET or ring net. Routing within the local transport protocol is usually accomplished by physically broadcasting the packet. The node that recognizes its own local transport address at the front of the packet will receive it.

The intermediate, internet layer is a protocol for delivery of a packet between any pair of nodes on the campus. One starts a packet with a local transport address field, and some form of identification of the target node in what may be called the "target identification" field. The local transport protocol, which examines the target identification and determines what local transport address to use to get to the next gateway. In turn, the target node's local transport protocol examines the target identification to determine which local transport address should be used for the next step of this packet's journey. This series of local transport protocols carries the packet on its journey to the target.

There have been suggested several alternatives for the interpretation of target identifiers by gateways ranging, on the one hand, from the simple identification of some route from the source to the target. Three possibilities along this spectrum are:

- 1) Unstructured unique identifier. Every node on the campus-wide net has as its target identifier a permanent unique identifier. This allows it to determine the appropriate next step in the route to every possible identified node. (Thus this approach is somewhat general.) In the most general form, the unique identifier provides no routing information whatsoever. The unique identifier may be interpreted as the identification of the point on the network to which the target node is attached, depending on the network's convention on how to reattach at a different place.
- 2) Hierarchical identifier. In this alternate form of hop-by-hop routing, the target identifier of each node is a multi-part field. It consists of an identifier of the subnetwork to which the node is attached and a node number (usually the local transport address). For each identifier, each gateway has a set of tables or rules that allow it to determine the appropriate next step in the route to every possible identified node. Since there are fewer subnetworks than nodes, these tables should be much smaller than in the case of the unstructured unique identifier. Reducing the number of parts of the identifier, and the argument can be extended to identifiers of more than two parts, network groups, and still smaller tables. When the identifier identifies parts of the network, this kind of network identifier is almost always thought of as identifying the network attached to that part.
- 3) Source route. The internet transport layer contains, in the place of the target identifier, a variable-length string of local transport addresses. The gateway merely takes the next local transport address from the string, moves that address to the local transport protocol address field, and forwards the packet. The gateway needs no knowledge of network topology, so the tables required for hop-by-hop routing vanish. A source route unambiguously identifies the route to the target, independently of what node is attached to that point. Any attempt to make an interpretation that a

III. Advantages of source routes in the campus environment

1. Separation of routing from target identification. The main difference between source routing and its alternative of target identification are moved from the internet gateways to some other agent. In turn, this responsibility change allows the internet to be implemented without freezing a particular form of network-wide identification of nodes or services. A commitment to a particular form of identity resolution part of a routing service, and since it doesn't matter to a gateway where a route comes from (the gateway cares only of identity resolution going on at the same time, perhaps implemented by different routing services. In practical situations there might be a routing method implemented by standard routing services, and in addition some experimental or special-purpose routing services developed with interactive resolution of catalogued service identities, or protocols that allow sending one packet to more than one target node. This is inappropriate to embed now in the internet transport protocol layer on grounds of inexperience. But they can be tried in the environment of disruption and without change to the gateways. It is even possible for one routing service to have a different view of the extent of the network. Routing virtual networks are thus implementable using multiple source-route services. This feature might be used, for example, to segregate routes that involve routes through external, tariffed, networks.* * Note that this separation of identity resolution from routing applies in both the case where points and the case where internet identifiers label nodes or services. In the latter case, one can imagine also an additional layer of binding (between addresses) and the internet node and service identifiers; this additional layer of binding could be the function of a service similar to the routing service. If this feature might have value in certain situations, one should understand that it is distinct from the modularity here imposed between routes and

At the same time, the source route field format places little constraint on the format of the local transport addresses for any particular subnetwork. Octets whose number is known by the gateway that moves the packet to that subnetwork. This flexibility means that paths can go almost anywhere in the network works no matter what their addressing or internal routing strategy, so long as at the far end of the outside network is a gateway that understands the format.

Separation of the mechanics of routing from the functions implemented by a labeling or addressing system has the advantage of clarifying the design down to how much naming function should be embedded in the lowest protocol layers. For example, it is usually proposed that an extra part of the internet address. This field is known as a "link" field in the ARPANET [10], the "channel" in X.25 [11], and the "socket" in the Xerox PUP [13]. One argument develops over how big this field should be--just large enough to distinguish among the activities or large enough to distinguish among all activities or connections the host will ever carry on. The former choice takes the view that the field choice makes the socket number a unique identifier, which is handling a labeling function for the host, perhaps allowing it to distinguish between connections. A superficial concern whether this field can be interpreted in different ways by different higher-level protocols. This argument is really about how efficiently perform the fan-out mechanics required is the same one that should be interpreting the labeling properties of the field.

The source routing strategy finesses both these arguments in that it allows the design of the packet format at the level of the internet transport layer without concern about socket number size or position in the protocol layering. As many octets of route as the target host needs to distinguish among the source route and learned as part of the initial negotiation with the target host using the initially obtained route to its negotiator. A unique identifier for the negotiation, and it can be included in a connection identifier field of the next higher level of protocol, to insure that packets arriving over the network are correctly routed.

2. Gateway simplicity and network maintenance. With the source routing scheme just described, a gateway must remember the route (if the route octet count hasn't been exceeded) and it remembers nothing after the packet goes by. This simplicity of operation and lack of state in the gateway with a small amount of random logic and a pair of packet buffers interconnecting two local network hardware interfaces. Such a program, has an exceptionally simple recovery strategy: a hardware reset to a standard starting state will always suffice. In practice, at least in the laboratory, statistics and respond to trouble diagnosis requests, but the basic principle that recovery is trivial remains intact.

(There is one way in which a source-routing gateway is more complex than its hop-by-hop counterpart. Every packet that arrives must be checked for a valid offset, so a small amount of lookup is needed to perform the forwarding operation. A related consequence is that higher-level protocols must be able to perform routing within the packet.)

To create a gateway that can sustain a through transmission rate comparable to that of the subnetworks involved requires careful budgeting. A bandwidth of 10 Mbits/sec. requires being able to pass twelve hundred fifty 1000-octet packets/second, leaving a time budget of only 800 nanoseconds used for the gateway, there must therefore be fewer than 400 instructions executed for each packet, with the implication that whatever routing strategy the source routing approach makes meeting such a budget a realistic possibility.

Maintenance is directly aided by having such a simple gateway mechanism. With little to do, there is little to go wrong, failures should be straightforward. Even in the case where a gateway is actually implemented by software in a node attached to two local transport networks that the program required is short, the cycles required are few, and that therefore the program is not only likely to be trouble-free but also easy to maintain. In a supervisor, where it is less likely to fail because of interference by other programs in the same node. Perhaps even more important in the source routing approach means that the software required can be quick to implement.

3. Route Control. One of the more interesting opportunities that arises when source routing is used is that the node that is the source of the route through the internet that outgoing packets follow. This control can be applied to solve several problems, as follows:

a) Trouble location. If trouble develops in a network gateway, it will be noticed first as failure of packets routed through that gateway. If a node that notices such a problem, one can route a test packet "out and back", through some set of gateways and back to the source. The steps in the route that failed, should quickly locate the troublesome gateway. One can also imagine extending this idea to routing a test packet to check on the operation of the lower levels of that node's operating system. An interesting aspect of this approach to trouble location is that to undertake network diagnosis; trouble location is not restricted to a network maintenance center that has some particular address.

b) Policy implementation: Some local networks may be paid for by a supporting organization that wants to have a say in the routing of traffic through its network.

Thus, from these arguments one can conclude that, at least for the campus-wide internetwork case, source routing is an attractive scheme.

We have concentrated on the application of source routing to the campus environment, without attempting to identify parallel situations that are important. For example, the British Post Office, in its recommended standard end-to-end transport protocol [14], suggests that source routing is useful in the situation of a local network, a public net, and another local net, because of the small likelihood that all of these separately administered networks will be connected.

Finally, the remodularization of network function implied by source routing involves a substantially clever routing service. Although we have not designed such a service, that design has not been sketched here; it remains an area of continuing investigation.

A c k n o w l e d g e m e n t s

This paper records a series of intensive discussions with, among others, Kenneth Pogran and Noel Chiappa. It also borrows ideas and terminology from a project by Danny Cohen, Jon Postel, and John Shoch and from working papers of the M.I.T. Artificial Intelligence Laboratory Chaosnet project. Early drafts were made by Danny Cohen and John Shoch. The basic idea of source routing and the mechanics of source route operation and maintenance were suggested by Vittal in their 1973 paper; the present paper contributes only observations and implications for the special administrative environment by the simplicity of a source routing gateway and the notion of a routing service were suggested by Hopper and Wheeler [15].

References

- [1] Farber, D.J., and Vittal, J.J., "Extendability Considerations in the Design of the Distributed Computer System (DCS)," 1973), Atlanta, Georgia, pp. 15E-1 to 15E-6.
- [2] Sunshine, Carl A., "Source Routing in Computer Networks," C_o_m_p_u_t_e_r_C_o_m_m_u_n_i_c_a_t_i_o_n_R_e_v_i_e_w 1, 7
- [3] Metcalfe, R.M., and Boggs, D.R., "Ethernet: Distributed Packet Switching for Local Computer Networks," C_o_m_m.
- [4] Okuda, N., Kunikyo, T., and Kaji, T., "Ring Century Bus-an Experimental High Speed Channel for Computer Commu" C_o_m_p_u_t_e_r_C_o_m_m_u_n_i_c_a_t_i_o_n_s, September, 1978, pp. 161-166.
- [5] Clark, D.D., Pograd, K.T., and Reed, D.P., "An Introduction to Local Area Networks," P_r_o_c._I_E_E_E_ 66, 11 (Novem
- [6] Thornton, J.E., Christensen, G.S., and Jones, P.D., "A New Approach to Network Storage Management," C_o_m_p_u_t_e
- [7] Hopkins, G.T., "Multimode Communications on the MITRENET," Local Area Communication Network Symposium,
- [8] Wilkes, M.V., and Wheeler, D.J., "The Cambridge Digital Communication Ring," Local Area Communication Network
- [9] International Organization for Standardization, Open Systems Interconnection, "Reference Model of Open Systems Ar" Europe, Paris, France, November, 1978.
- [10] Feinler, E., and Postel, J., Editors. "ARPANET Protocol Handbook," Stanford Research Institute, NIC 7104, January,
- [11] The International Telegraph and Telephone Consultative Committee (CCITT), "Provisional Recommendations X.3, X" Services," International Telecommunication Union, Geneva, 1978.
- [12] Postel, J., Editor. "DOD Standard Transmission Control Protocols," Information Sciences Institute, IEN 129, January
- [13] Boggs, D.R., et al., "PUP: An Internetwork Architecture," I_E_E_E_T_r_a_n_s._o_n_C_o_m_m._C_O_M_-28, 4 (April, 198
- [14] Linington, P.F., editor, "A Network Independent Transport Service," British Post Office PSS User Forum Study Group
- [15] Hopper, A., and Wheeler, D.J., "Binary Routing Networks," I_E_E_E_T_r_a_n_s._o_n_C_o_m_p_u_t_e_r_s_C_-28, 10 (Octo

Figure 1 -- Relation between local transport protocol, internet transport protocol, and other communication

Figure 2 -- Possible implementation of an internet source route.

Notes

Files for figures are located on Alto disk labelled "Muriel's NP Disk".

The files are called "zurich1.draw" and "zurich2.draw". These files must be run through the "redraw" program before use.

Biographical sketches for Saltzer, Reed, and Clark are not on-line.