Ann Neilson

Internet Experiment Note: 191

C. Sunshine
D. Cohen
J. Postel
ISI
July 1981

Comments on Rosen's Memos

INTRODUCTION

This memo comments on recent IEN's by Eric Rosen of BBN (numbers 182, 183, 184, 187, 188, 189) [1,2,3,4,5,6]. We think these notes raise some important and interesting issues which require further discussion. In the following we focus on the points of disagreement (but don't assume that we agree with something simply because we don't mention it here). After a brief general comment we discuss each note in turn.

There are some good points raised in this series. Unfortunately the presentation is both verbose and incomplete. There is nothing wrong with taking a certain aspect of a topic and exploring it at length, but these memos seemingly present all available alternatives and select the "best" for further development. Our concern is that, in fact, not all alternatives are studied, and not all evaluation criteria are given the proper weight in selecting the "best" alternative. A minor problem is the informality of the references. It is unclear exactly which earlier memos, reports, and papers the author has in mind in some of the discussion, and it is unclear if the author is aware of some very relevant material. In some sections it appears that the author is unfamiliar with much of the relevant material, and hence fails to include important points in his presentation.

IEN 182

This note on "Issues in Buffer Management" is, in the main, a description of buffer management in the ARPANET IMPs. This is quite useful and should be food for thought for gateway designers and implementers since gateways may have some of the same constraints and concerns in buffer management as IMPs. However, the differences that do exist in the goals for gateways and IMPs are not taken into account, so the policies adopted for IMPs are not necessarily appropriate for gateways. Differences in the level of reliability of delivery, and the end-to-end virtual circuit vs. the datagram style of service can lead to substantial differences in the requirements for buffer management.

This is a useful memo in that it exposes a good deal about the buffer management polices used in the ARPANET IMPs, information that is not easily found elsewhere. But it is contains some weakly supported

overly broad conclusions that seem to ignore and sometimes contradict existing results in this area.

IEN 183

This memo presents a proposal for a logical addressing mechanism in the ARPANET, and includes a good deal of discussion of alternatives. Interested readers should see earlier IEN's on the subject from MIT, ISI, and Xerox, plus the classic paper by Shoch, and recent work on "naming authorities" at Xerox, which the author fails to credit or reference [7,8,9].

We prefer the more commonly used term "name" to the phrase "logical address" which the author uses.

The key proposal is to include a name-to-address lookup function in the source switches of a network so that the "user" will not have to supply ("physical") addresses. This seems a worthwhile goal, but the meaning of "user" seems confused between (1) people or application programs using the network, and (2) network access software (such as NCP or TCP) supporting (1) in the hosts. The author seems oblivious of this distinction.

Everyone agrees that category (1) "users" should be able to use names. Of course, most ARPANET hosts' category (2) software already provides this function (the host table) for category (1) "users". The proper discussion should be whether this function is best located in the switches, or in the network support software of the hosts, but this is not explicitly addressed by the author.

The author presents a reasonable approach to implementing a name lookup function without requiring broadcast of dynamic changes to all participants. A basic table of all potentially usable addresses for each name must be distributed to all parties (the "authorized" table), but this is expected to change relatively slowly. Entries in this table are assumed usable ("effective") until an explicit exception message ("destination not accessable") results from using them. The unusable markings are reset after a time interval.

We agree that this is a worthwhile proposal, but the placement of this function in the hosts, the switches, or a separate name lookup service needs further discussion. Since most hosts are already performing this function as noted above, it is clearly within their capabilities. An advantage of placement in the switches seems to be prevention of "spoofing" since hosts can only send/receive messages from/for a specified name if that name is "authorized" for the

addresses they are physically attached to. Of course this requires source and destination switches to check messages in a "trusted" fashion.

There is a small inconsistency in the author's discussion of source-only vs. intermediate ("tandem") node name lookup. At the top of page 11, it is stated that the tandem nodes will be "no more likely" than the source node to have new information during a transient update period. However, on page 12-13, it is pointed out (correctly) that tandem nodes likely WILL produce a "better route selection ... if delay changes or topology changes take place while a packet is in transit."

There will be substantial modification needed to the host software in order to implement this scheme. It is proposed (we think) that both the current scheme and the logical address scheme be available at the same time. The details of the logical address are not very clear, but a 16-bit logical address is suggested, which would require a character string to number lookup in the hosts to make it convenient.

IEN 184

This memo claims that the previous work on the Internet is deficient due to reliance on an inadequate model of the structure of the Internet. IEN 184 claims to present a new model of the Internet that does provide a basis for future work.

The proposed model of internetwork operation views the gateways more explicitly as switching nodes, with the hosts attached to these nodes. In particular, each host is multi-hommed on all the gateways on the same network as the host.

There is some merit to this model and the questions it raises, but the author is not the first to think of this viewpoint (see for example IEN-135 [10]). There are also some problems with this model that the author seems unaware of.

This new model might be acceptable if one wanted to build a super ARPANET based solely on lines and super-IMPs, but if one is planning to include other technologies such as broadcast satellite and broadcast local networks, the proposed model has serious flaws.

For example, two hosts on the same net may still wish to use Internet protocols to communicate. In the author's model, they would have to do so by going through an intermediate gateway on their net, since by definition, no hosts can communicate directly over a "Pathway" with

no intervening "Switch." This is clearly inefficient in the intranet case, and one way in which it differs from the ARPANET. This would also be true in many single broadcast nets where there are no intervening switches between hosts even at the single network level of "Network Structure."

This memo fails to consider the impact on the host systems. Host will be designed to use a common approach to communication with other hosts whether they be across the room or across the world. With the existing model and Internet Protocol, the same procedures and formats can be used between hosts on the same network and between hosts many networks apart (though different performance parameters may be necessary).

The model developed in the Internet Working Group and described by Cerf (IEN-48 [11]) continues to be the most reasonable basis for developing the Internet.

IEN 187

This memo assumes the model (of IEN 184) of hosts always sending and receiving internet traffic via an "Internet Switch". It goes on to describe the interactions of a host and an internet switch, and then criticizes the existing Internet Protocol for not being a perfect host-switch interface protocol.

We cannot possibly take on all of the topics and "lessons" presented, but Section 2.4 of IEN-187 on fragmentation provides a good example of what is wrong with these reports. Again, the author seems unaware of previous important work on this subject, for example IEN-20 by Shoch (expanded and published in Computer Networks in 1979) [13], or the paper by Sunshine on interconnection of networks published in Computer Networks in 1977 [14]. If the author had read these, he might have avoided several serious deficiencies in his presentation:

1. After a long discussion of the evils of final destination (or internet) fragmentation, the author reveals his preferred approach of hop-by-hop (or intranet) fragmentation as if he invented the idea.

2. There is an important goal that internet fragmentation supports, but intranet fragmentation does not: independent and possibly different routing of each fragment through different exit gateways from a "small packet" net (and subsequently). The author fails to consider this point.

3. In presenting scenarios (page 58) showing the evils of internet fragmentation, the author omits the important scenario of several small packet nets in a row, where repeated intranet fragmentation is just the WRONG thing to do.

4. Packets with the Don't Fragment flag on are not "simply lost in transit" (page 53) if they cannot be forwarded without fragmentation. Specific error packets are returned to the source host, which may try to resend smaller packets.

5. After all his discussion, the author admits in the final paragraph that destination host fragmentation is necessary anyway if the final network gets too large a packet. The author claims this will be necessary only for hosts on nets with "unusually small" maximum packet sizes, but in fact it will be necessary on all nets with less than the maximum maximum packet size of any net in the system if they wish to receive packets from the largest packet size nets.

The net effect of this sort of incomplete presentation is a step backward from the current imperfect level of understanding of this important issue.

The author also attacks the Type of Service (TOS), Time to Live (TTL), Source Routing (SR), Flow Control (FC), and Fault Isolation (FI) features of IP and ICMP.

On Type of Service the author tells us for ten pages all the bad things about the Internet Protocol provision for TOS, while agreeing it is an important concept, but has nothing different to offer, except some vague notion that service catagories should correspond more closely to application types.

On Time to Live the author complains that there is an inconsistency since the TTL is stated to be in seconds, and that the gateways must decrement the TTL by one, and that the gateways are expected to process datagrams faster than one a second. If one assumes that the intention is to guarantee that datagrams stay alive as long as the TTL, he is right. But the intention is really to guarantee that they disappear before TTL. So TTL is an upper bound on how long the datagram may exist. Most reliable transport protocols assume a maximum datagram lifetime (sometimes unknowningly) for the correct operation of their reliability procedures [15].

On Source Routing the author suggests that this feature exists due only to problems with existing routing procedures and for

experiments, and that any really adequate routing procedure in the gateways will eliminate the need for source routing in normal operations. We suggest that the Internet will be a much more dynamic environment than the author has yet imagined and that source routing will be essential to reach through the Internet to local environments not fully integrated into the main Internet routing world.

On Flow Control and Fault Isolation the author indicates that the current mechanisms are inadequate, but does not suggest workable alternatives. On FC the ICMP "Source Quench" message is cited as a case of "choke packet" flow control which the author does not believe in (page 64). Earlier (page 63) the author complains that "source quench" is only advisory, and later (page 66) the author makes vague suggestions that a better flow control scheme would use advisory messages to suggest that datagrams had been discarded (exactly what source quench does).

All in all this memo comes across as an attack on the Internet Protocol, with few suggestions for improvement. But it is based on an assumption: that the Internet Protocol is a host-switch access protocol. This assumption requires further discussion.

IEN 188

This memo describes logical addressing in the Internet, primarily by recasting the method of IEN 183 in generalized terms. There are a number of inaccuracies and omissions in the discussion. One serious limitation is failure to consider the case of hosts sending Internet datagrams to each other directly on a single net as discussed above.

On page 4 (middle), the author correctly states that IP addresses are hierarchical, but incorrectly states that their second component is necessarily a "physical address." In fact, it may be a name or "logical address" in networks that provide that capability (but must be carried in 24 bits).

On page 7, the author proposes using a "unique name which is meaningful at each level of internet hierarchy." This seems to be a strong violation of layering, and as the author admits, would require the switches in every constituent network to "understand" and be able to lookup the names, probably an intolerable demand on individual network autonomy.

On page 34, the author's claim that hierarchical addressing requires less table space than flat addressing is false. His justification is incomprehensible to us, particularly since he has just finished

proposing an "area" addressing scheme similar to hierarchical schemes in order to reduce table sizes!

In the detailed model of operation given in Section 3.4, an important step is omitted when the first sentence states, "Let's assume that a source Host has given a message to a source Switch ..." How does the source host pick the source switch? In fact, it must pick both a network level (e.g., IMP) and internet level (gateway) switch, assuming it is multi-homed, which at least at the internet level is quite likely. In order to make this selection, the host will have to have a table giving the best switch (at each level) for each possible destination name. But these are precisely the sort of tables the author's scheme is meant to avoid having in the hosts. In light of the comment above about hosts talking to each other directly on the same net, the hosts must at least know the names and addresses of every other host on their own net.

The treatment of mobile hosts is quite brief and offers no improvement over previously proposed solutions.

IEN 189

This memo discusses routing in the Internet, and proposes that the existing gateway routing procedure be replaced by the SFP procedure now used in the ARPANET. This is surely a useful suggestion. The note does however raise a number of issues in its examples of routing problems that indicate an incomplete understanding of the whole area.

The note proposes a "gateway discovery protocol" that could be provided by individual nets. This idea seems worthwhile, although it is not clear how many individual nets would be willing to make such additions. We should like to point out that it is also possible to perform this function directly among gateways in networks which support broadcast or group addressing.

The discussion of routing alternatives makes generally sound if qualitative conclusions, but a few details are confused. The discussion of throughput performance on page 41 assumes TCP will operate with a small enough window over a high delay path so that throughput is reduced, but this is precisely the situation in which proper "tuning" requires a large window, which would allow high throughput.

The analogy with "whole picture" algorithms on pages 44-45 fails to mention that in the whole picture scenario, each person would have to get a piece of paper 100 times bigger than with the local scheme, and

hence this approach has an information distribution requirement that is much higher.

This memo contains several informal citations that could be usefully spelled out for the IEN audience. The author mentions algorithms by Gallager (page 17), Dijkstra (page 20), and Floyd (page 20), all without references. It is safe to say that any list of references containing only the author and his coworkers (as consistently done in this series) cannot be adequate.

One particular example provokes the following response:

> Please replace the second paragraph of page 49 of IEN-189 with the following paragraph:
>
> > "In fact the situation could be even worse. If Switches in Boston know nothing about what happening inside the building on 4676 Admiralty Way then data for the North section of the 11th floor which arrives at the South section of the 11th floor is then sent back from the South section to Boston for alternate routing will just loop back to the South section. The data will be stuck in an infinite loop, never reaching its destination. In IEN 179 [12] Danny Cohen proposed a regional scheme like this, apparently not realizing that it suffers from loops. His proposal also includes a form of hierarchical addressing which is closely bound up with routing, so that a Switch is Boston might not even be able to distinguish data for the South section from data for the North section. That is, in Cohen's scheme, data for the South section and data for the North section would be indistinguishable at the Boston Switches; all such data would appear to be addressed to the South section. Only the Switches at the South section would look further down the address hierarchy to determine whether the data needs further forwarding to the North section. Any such scheme is hopelessly loop-prone, except in a Network Structure whose connectivity is extraordinarily rich, much more so than the Catenet's will ever be."
>
> Since the above suggestion was merely to follow the routing strategy used by the phone companies, TELENET and others, you should warn them immediately about this hopelessly loop-prone situation.

I believe that if the Boston Switch has ALL the information about EVERYthing, EVERYwhere it would be in position to make better decisions, ALWAYS, especially if that information is updated with

absolutely ZERO time delay. If this information is absolutely free (in terms of communication, storage and processing) it may be dumb not to make every Switch always know everything about everything, down to (or "up to"?) the finest granularity (location? site? process? file? register? bit?). However, if this is not absolutely free, some compromises may have to take place.

Oh, one point which I did not quite follow: why if the Nevada/California lines are broken forever, Boston is never told about it - as described by you? By the way, what made you understand that the "The Switch at Nevada would look further down the address hierarchy to determine whether the data needs further forwarding to California" ?

I highly recommend that you get hold of any telephone directory and read the area-codes tables. This may help you understanding that the California area codes are neither above, nor below, nor further on any hierarchy than the Nevada ones, and vice versa.

This is a very subtle point which may escape the casual reader. Mastering this idea may help you understand what IEN-179 is all about. In short, IEN-179 is not an attempt to describe the ideas which you have in mind by using the telephone scenario, but an attempt (which obviously failed, at least in your case) to introduced old well-proven ideas from other communication arenas into ours.

SUMMARY

In summary we are glad to have this information and these opinions presented for discussion in the Internet Working Group, and we hope that others will speak up with their opinions too. We are concerned that too many will be so overwhelmed by the wide ranging arguments to notice that some important considerations were not mentioned.

We especially want to make clear that a fundamentally different model of the Internet architecture is proposed by Rosen, and that we have serious reservations about aspects of that model.

REFERENCES

[1] Rosen, E., "Issues in Buffer Management", IEN 182, Bolt Beranek and Newman, May 1981.

[2] Rosen, E., "Logical Addressing", IEN 183, Bolt Beranek and Newman, May 1981.

[3] Rosen, E., "Issues in Internetting Part 1: Modelling the Internet", IEN 184, Bolt Beranek and Newman, May 1981.

[4] Rosen, E., "Issues in Internetting Part 2: Accessing the Internet", IEN 187, Bolt Beranek and Newman, June 1981.

[5] Rosen, E., "Issues in Internetting Part 3: Addressing", IEN 188, Bolt Beranek and Newman, June 1981.

[6] Rosen, E., "Issues in Internetting Part 4: Routing", IEN 189, Bolt Beranek and Newman, June 1981.

[7] Clark, D., "A Proposal for Addressing and Routing in the Internet", IEN 46, MIT/Laboratory for Computer Science, June 1978.

[8] Cerf, V., "Internet Addressing and Naming in a Tactical Environment", IEN 110, Information Processing Techniques Office, Defense Advanced Research Projects Agency, August 1979.

[9] Shoch, J., "Inter-Network Naming, Addressing, and Routing", Proceedings 17th IEEE Computer Society International Conference, pp72-79, September 1978.

[10] Sunshine, C., "Addressing Mobile Hosts in the ARPA Internet Environment", IEN 135, USC/Information Sciences Institute, March 1980.

[11] Cerf, V., "The Catenet Model for Internetworking", IEN 48, Information Processing Techniques Office, Defense Advanced Research Projects Agency, July 1978.

[12] Cohen, D., "Addressing and Routing", IEN 179, USC/Information Sciences Institute, March 1981.

[13] Shoch, J., "Packet Fragmentation in Inter-Network Protocols", Computer Networks, V.3, N.1, pp3-8, February 1979.

[14] Sunshine, C., "Interconnection of Computer Networks", Computer Networks, V.1, N.3, pp175-195, January 1977.

[15] Watson, R., "Timer-Based Mechanisms in Reliable Transport Protocol Connection Management", Computer Networks, V.5, N.1, pp47-56, February 1981.