Internet Experiment Note No. 200

INTERNET PROJECT RESEARCH PLANNING REPORT

David D. Clark
MIT Laboratory for Computer Science
Computer Systems and Communications Group

1. Introduction

This paper represents a first attempt to provide a structure to the current and planned activities in the Internet project. It is particularly concerned with the evolution of the protocol architecture, and the identification of new functions which the protocol is to support. It attempts to identify the basic decisions which must be made as part of the planning process.

-N.B.: This document is a draft. It is distributed at this time for comment and discussion only. Anyone using this document as a basis for project planning at this stage is extremely misguided.

-The changes and extensions described here are not intended to have an impact on current host implementations of TCP and IP, except for hosts specifically involved in future research and demonstrations. Implementors only involved in realizing the currently defined level of service need not be concerned with this report.

The Internet project has reached a point where it must provide a stable and usable service to its users. This is a painful but crucial stage in any system project which hopes to prove its point in the real world. Thus, our planning must now take into account two requirements, first that we provide a functional service, and second that we develop and integrate new functions into the existing architecture. There are several reasons why new functions must be added to the existing architecture. First, there are additional service requirements, which are not currently supported but which will be required before the service meet the currently forseen needs. Second, there are explicit research goals which have been established for the project, with announced demonstration dates. Third, there are additional research goals which are of interest to the participants in the project, and which are generally perceived as being appropriate and compatible with the internet philosophy. This last collection of topics can easily grow without bound, leading to a project so large as to be intractible. One of the first planning issues to be addressed is precisely what problems we choose to solve in this next research cycle, and what problems we choose to ignore.

2. Possible Research Topics

The following list catalogues a number of areas in which future evolution of the architecture has been proposed. The comments provided with each section attempt to give some idea of the amount of thought which has already gone into this topic, and the range of options available for approaching each topic.

2.1. Types of Transport Services:

The issue here is what sort of services will the transport layers of the architecture provide to the applications above. Currently, TCP provides the only wide-spread service, a bidirectional, reliable data stream. It is unclear whether TCP is the most suitable protocol for high bandwidth, high delay occasions, such as bulk transfer of data, and it is certainly clear that TCP is inappropriate for speech and other low delay, loss-tolerant applications. Other functions which might be useful include multicast rather than point-to-point data transmission. Speech is a clearly defined goal for internet support, but it is currently being supported only outside the internet architecture, using alternative protocols. To provide serious speech support, IP may have to be evolved to provide the necessary support.

Other services such as bulk, high bandwidth transfer and multi-casting are not specifically required for demonstration or service. Thus, it may be possible to ignore these issues. On the other hand, it is reasonable to consider exploring these topics because it is probable that they can be examined somewhat in isolation, without a global upheaval to the entire architecture.

2.2. Addressing and Routing:

As we proceed from an assumption of an internet with 50 nets to an internet with 1,000 nets, substantial upheavals will occur in the architecture and its implementations. The simple routing algorithm of sending to every gateway the location of every net, would require exchanging tables with 1,000 entries, and not only would this cause an overflow of limited gateway storage space, but it would also use up a substantial part of the network bandwidth if the system were at all responsive. Currently, the internet has a vaguely hierarchical structure, in which at the top there are nets and beneath this level is a substructure which is understood only by the net itself. Presumably, it will

be necessary to generalize this structure by grouping nets into areas, which physically encompass several nets. However, this idea of areas is counter to an original internet goal, which is that individual networks should be connectable together in any configuration, with all possible physical paths being used as realizable routes through the internet. If there is a structure to the net, than only routes that conform to the structure will be realizable.

I feel that a restriction of this sort on the internet is quite reasonable. An obvious structure for the internet is to imagine that there is a central cluster of nets which provide the function of long-haul transport. These transport nets would have connected to them additional sets of nets which provide an access function to the internet. Typical transport nets would be the ARPANET and SATNET: typical access nets would be local area nets and packet radio. The implication of this structure would be that the transport nets would never rely on the access nets for transport function.

An additional complication of routing is the TYPE OF SERVICE field of IP. The TOS field is presumably used to affect the service provided by the individual nets: however, it also should serve the function of selecting among routes at the internet level. Currently, this function is completely missing. Adding it can only make the routing problem much more complicated.

2.3. Flow and Congestion Control:

Currently, the internet has a simple mechanism for congestion control, the Source Quench Packet, which can variously be viewed as a choke packet or as an advisory overload packet. There is no evidence that this mechanism works very well; indeed there are substantial indications that it probably does not. As load builds up in the internet, and especially as we hook together networks whose basic speeds differ by orders of magnitude, it will be necessary to identify and implement a workable congestion mechanism.

One decision currently undecided about the internet architecture is whether the congestion mechanism should include the concept of "enforcement". Most congestion mechanisms push back on their data sources in a manner that is not advisory, but mandatory. For example, networks selectively drop packets, disable communication, or simply refuse certain inputs. The internet can do none of these. If an input host ignores the congestion information currently offered, the increased degradation caused by this is not focused on the offending host, but is randomly distributed on all the traffic in the affected area. It would seem that some mechanism for enforcement would be appropriate. However, enforcement is very difficult, given one of the basic assumptions of the architecture, which is that the gateways have no state. Without state, it is impossible to keep track of whichthe offending host is so that it can be discriminated against. I feel that an important idea in this respect is that of "soft state", in which the gateways attempt to build up a model of the world by observing the traffic passing through them. If they should crash and lose this state information, no transport services is disrupted, but as long as the information exists, it permits the gateway to discard selectively those packets which appear to be violating the congestion control restrictions currently in effect.

2.4. Distributed Control:

An important question about the internet is the extent to which its ownership and management is decentralized. It was originally a design goal that various gateways would be implemented by different organizations, and the individual networks out of which the internet is built were certainly expected to be managed by separate organizations. Realistically, decentralized management is an important goal in the long term. However, it clearly makes development and operation much more difficult. Currently, essentially all of the central gateways as well as the important transport nets are all operated by BBN. A specific decision is required as to whether this situation is to be exploited, in order to simplify the next round of implementations, or whether distributed operation is to be deliberately injected into the system as an additional research goal.

A problem closely related to this is the interaction between the existing service internet and some evolved internet being used as a testbed for these advanced topics. It will, at a minimum, be necessary to construct some sort of firewall between the experimental environment and the service environment, so that they can interact without destroying each other.

2.5. Partitioned Networks

This is a critical research goal, because it has been explicitly identified for demonstration in the two to three year timeframe. The specific demonstration involves attaching high power packet radios to the ARPANET at appropriate locations, providing airborne packet radios which are able to communicate with these ground based radios, and then severing the landlines in the ARPANET and using the packet radios to reconstitute ARPANET service.

There are two ways to approach this problem: at the network level or at the internetwork level. The network level solution, in which the imps are taught about the existence of the packet radio links, is clearly the simpler of the two. Among other things, it requires no modification of the host software. The internetwork solution, in which knowledge of the packet radio links is restricted to the hosts and the gateways, is a more challenging solution, but one which is more in line with the original goals of the current cycle of research. I propose that the problem be solved at the internetwork level, but this requires an upgrade in the sophistication of the host routing algorithm, so that failures to communicate with the host apparently located on the same network trigger the kind of internet rerouting which would normally be invoked only for a pathway known to pass through intermediate gateways.

There are other problems that resemble net partition. The "expressway routing" concept is that a host or gateway may select an internet path to bypass a network, even when the net is fully functional. This sort of action would require that the mechanisms designed to support partition be used under normal operations, not just under crisis conditions.

2.6. Improved Effectiveness:

This somewhat vague title covers a number of important topics. Currently, the TCP specification describes logically correct operations, but makes only limited reference to efficient operation. Issues such as window management and the timing of acknowledgements are left as an exercise for the implementor. Bad design decisions in this area lead to the symptoms sometimes referred to as Silly Window Syndrome. It is important that the specification be expanded to include sufficient information to prevent this sort of inappropriate operation. An area in which this will become immediately visible is in the use of public data networks to carry internet packets. Most tariffs for public nets are based on the number of packets, and the window algorithms and acknowledgment generation algorithms strongly influence the number of packets required to send a given amount of data. There will be a pressing cost requirement to eliminate inappropriate implementations. Even where cost is not measured in dollars, the perceived effectiveness of these protocols is an important component of their acceptability to potential users.

2.7. Access Control:

Plans are already underway in the short term to provide some controls over who can use the internet, in particular by the addition of passwords to the TAC. However, in the long run, much more must be done. The current protection strategy is based on the assumption that individual hosts are sufficiently responsible to restrain their users. As long as we have large time-shared computers, this assumption is still somewhat workable. However, we are moving

as rapidly as possible away from this position to a position in which personal computers are allocated to individual users, which provides no management strategy to ensure that the individual users of these computers, who are presumably attached to local nets over which no access control is imposed, refrain from going through these local nets to reach long haul nets whose use is limited. The most extreme example of this currently in the internet are the personal computers belonging to students at universities such as M.I.T. As part of M.I.T.'s policy toward its local net, these students are being invited to attach to the local net as rapidly as possible. However, the local net is attached to the ARPANET, and there is no mechanism which can prevent these students form utilizing the ARPANET from their personal computers. Clearly, it is ridiculous to expect the students to voluntarily refrain. The only possible mechanism is an access control algorithm in the gateway. But by the basic architectural assumptions of internet, this is almost impossible to provide, because the gateway lacks the information to discriminate between legitimate and undesirable users. The only information available to the gateway are the original and destination addresses, but short of a complete list enumerating all of the used addresses at M.I.T., which might potentially be very long, the address is not a sufficient indicator of validity. A decision must be made whether or not this problem is to be solved. If it is, some substantial work will be required.

Access control and routing interact in an important way in the gateway. The access decision is not a simple yes/no control, but a selection of route based on privilege. For some nets, it will also be necessary to collect billing information, which requires the same sort of unforgable identification as does access control.

2.8. Performance Evaluation:

There are currently a number of activities under way to evaluate the performance of the internet. Some of these are being done as part of operational management of the internet. Others are being done by various users of the internet from their particular vantage point to evaluate the service made available to them. These projects are not especially well coordinated at the moment, and the results they gather are not being used in other than a very general sense to evaluate and identify problems with the internet. Currently, it is difficult to improve the quality of the measurements being made, because there is not space available in the gateways to improve the metering and instrumentation which is installed there. A new version of the gateway code is now being developed by BBN which will have more space for new code.

At a more philosophical level, we lack even a metric against which to compare

our work. What constitutes a "good" internet? Superficially, one thinks of things like low delay, high bandwidth, and reliability. But actually, depending on the class of service being used, these individual goals may or may not be desirable. In fact, a good internet is probably one which has the flexibility to conform to a variety of offered loads, rather than doing a superb job of meeting exactly one application class. In the long run, it would be worthwhile trying to identify what goal we are attempting to meet, before rather than after we meet it.

3. Discussion

As the preceding list suggests, it is possible to put together a wish list of internet features which is unworkably large. One of the first problems we must face is doing some delicate but effective pruning upon this list to bring it to a managable size. Having done this, it will be necessary to make some rather difficult decisions about the general structure into which our future solutions must fit. The above list raises some very important questions about the shape of the internet. For example, the partitioning question most clearly raises the issue of whether or not hosts connected to each other over a single net view themselves as being connected to a network or to an internet. It is silly to proceed until we have an answer to that question. Two other important questions are the extent to which constraints are imposed on the workable topologies of the internet, and the extent to which the internet is assumed to be owned and operated by one or many organizations.

The short time schedule for the announced demonstrations require that progress on this list of functions be fairly rapid. The need for quick progress suggests two conclusions to me. We will not achieve our goals if we attempt to get there by a series of small incremental steps from where we are now. Many of the functional requirements listed above interact with each other in a very strong way, and it will be necessary to take all of these requirements into account as we do a design for the final service. An incremental approach will only trick us into thinking that we could ignore some of these issues and attack others first. I believe that will not work in this case. We must attempt to stabilize the service we have now and then live with that service for some time, perhaps a year or more, during which time we develop the followup, which will ultimately be used for the demonstrations scheduled in the two to three year time frame. Given this conclusion, it then follows we should attempt to minimize those functions which we advertise as part of the current service we are attempting to stabilize, because any effort investigated in expanding the current service is effort which is diverted from the long-term design problem.

4. Milestones

It is my hope that a working version of this document will be produced not later than the meeting of the ICCB in January. Anyone with comments is requested to send them to DClark at MIT-Multics before that time.
