INTERNET DRAFT
draft-ohta-sun-01.txt

M. Ohta
Tokyo Institute of Technology
March 1997

Simple Unified Networking

Status of this Memo

Abstract

The concept of LIS for IP over ATM causes a topology mismatch between the link and the internetworking layer. While it introduces some inefficiency with CATENET based operation, it is not so much a problem unless we try to solve this minor problem.

Short-cutting attempts such as NHRP can't solve the inefficiency issue at all even though, or, just because, it utterly destroys the CATENET model, which resulted in inelegant modifications of existing protocols, which, in turn, causes scalability problems.

Moreover, the creation of short-cut VCs itself suffers a scalability issue.

But, CSRs (Cell Switching Routers), or RSVP-signaled ATM switches, make it possible to have end-to-end cell-by-cell relaying over IP routers. That is, there is no reason to have LISes and there is no inefficiency

The way to go for the Internet is Simple Unified Networking with the CATENET model.

1. Introduction

See RFC1620 [RFC1620].

2. Inefficiency Remains

On the Internet today, routing metric roughly approximate the real distance between networks. As a result, semi-optimal routing over the Internet is possible, though some policy restriction may impose some additional detour.

But, with the LIS model mentioned in RFC1620, routing metric has nothing to do with the real distance. This is not a problem within an NBMA with link layer shortcutting where link layer metric is the approximated metric.

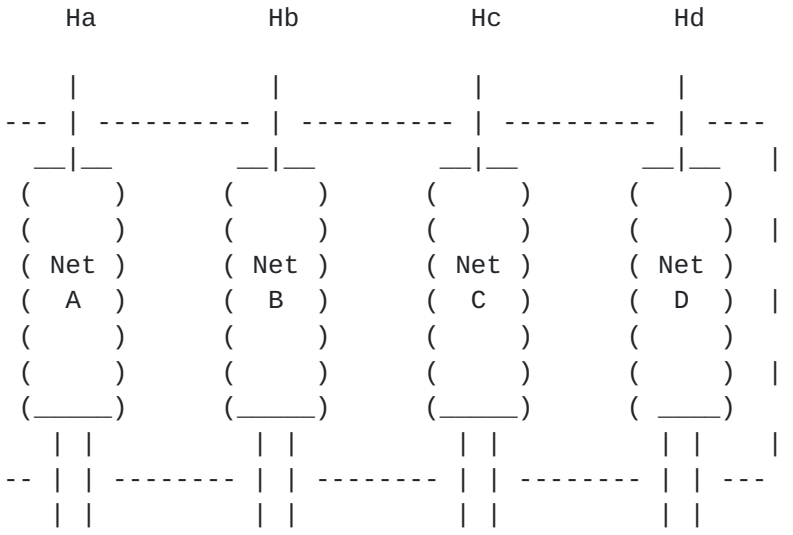
That is, when an NBMA is a leaf of the Internet with only a single entry router to the rest of the Internet, there is no inefficiency problem.

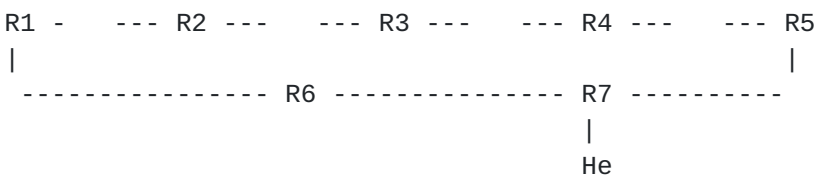
But, in such a case, shortcutting in NBMA is a local optimization issue unrelated to the Internet architecture.

Otherwise, when the leaf NBMA has multiple entry routers or when a host in the NBMA is multihomed, the distortion causes inefficient routing.

Finally, when the NBMA is not leaf but a transit network, the distorted metric can pollute the rest of the Internet to be a serious inefficiency problem.

For example, consider the following configuration:





where Nets A, B, C and D are highly logical LIS in a large shared medium network.

For example, suppose R1 is located at Munich, Ha Sunnyvale, R2 Montreal, R3 Memphis, R4 Kuala Lumpur, Hd San Jose, R5 Mountain View, R6 Danvers, R7 Menlo Park, and He Palo Alto.

Then, without shortcutting, Ha and Hd may communicate hop-by-hop from Sunnyvale, Montreal, Memphis, Kuala Lumpur and finally to San Jose. Not an inefficient path.

The problem is that routing metric at the Internetworking layer does not reflect the real world metric at all.

But, if we can somehow make use of the fact that Ha and Hd are placed in a single shared medium, Ha and Hd can communicate locally within Silicon Valley between Sunnyvale and San Jose.

That's the inefficiency issue that mechanisms in RFC 1620 wanted to resolve.

The problem is that though the inefficiency may be removed within the shared medium, it's not the only inefficiency.

When Ha and He communicate over a path Ha-R6-R7-He, the traffic will pass from Sunnyvale, Munich, Danvers, Menlo Park and finally to Palo Alto, even though the path Ha-R5-R7-He exists within Silicon Valley.

The inefficiency can be avoided by reducing metric within the shared medium. But, it causes other type of inefficiency. That is, the shared medium will be used for transit even though the physically shortest path exists outside of it.

The problem is that routing metric at the Internetworking layer within the shared media does not reflect the real world metric at all.

When a LIS contains hosts at room 1035, Fairmont Hotel, San Jose; room 1036, Fairmont Hotel, San Jose; Holidy Inn San Jose; Palo Alto; Los Angels and Munich; there is no meaningful metric for the LIS to be used outside of the LIS.

Also, It is obvious that no intra-shared-media protocol can solve the route selection problem outside of the medium at R7.

That is, routing metric should be mostly proportional to the physical distance. Then, the least metric path will be almost optimal.

It means that LISes should not be so logical and mostly contiguous.

As a result, the CATENET model with no extension works efficiently over the shared media.

3. Inscalability Problems

3.1 RSVP Inscalability

As the Internet protocols are designed with the CATENET model, modification to the model naturally makes some protocol not work and other protocol not to scale.

For example, RSVP scales to the number of recipients because RESV messages are merged on routers upstream toward the sender.

But, in a large shared medium with no intermediate entity to recognize IP, merger of the RESV messages is impossible.

As it is essential to merge RESV at the data branch point, RESV merging servers external to the shared medium does not work.

That is, all the RESV messages concentrate and implode at the upstream most router or the sender on the shared medium, which means not so many recipients can be supported.

Note that, in the worst case when most of the hosts in the shared medium are the recipients, the amount of imploding packets is almost equal to the amount of ATMARP packets for a single ATMARP server receives, if the entire shared medium is served by a single ATMARP server.

That is, on multicast-aware shared medium, it's enough to make the entire medium a single subnet, maybe with SCSP.

3.2 VC Shortage at the Egress Router

It is unlikely that the Internet mostly consists of a large single shared medium.

Thus, when hosts in a shared medium wants to communicate to the Internet outside of the medium, the egress routers must be directly

connected to each such host through a dedicated VC.

But, shared medium can support only a limited number of VCs for a single node.

On existing commercial shared medium service such as X.25 or framerelay, it is typical that the number of supported VCs is less than 100.

It is typical that ATM switches can support only several thousands of VCs for each port.

Thus, not so many hosts can communicate with the external Internet efficiently

Other hosts can still communicate hop-by-hop. But, as the size of the shared medium glows, the efficiency as a whole approaches that of hop-by-hop.

That is, it is necessary to make the hop-by-hop communication efficient by not making LISes logical, which means that no inefficiency exist to be removed by shortcutting attempt.

4. Cell Switching Routers

It seems to the Author that some people thought that cell-by-cell relaying was impossible over IP routers, which, seemingly, motivated them to support shortcutting over ATM shared medium.

While it was understandable, cell-by-cell relaying over IP routers is possible.

The point is that it is possible to signal ATM switches with RSVP [RSVP], ST2 [ST2], IFMP [IFMP] or some other IP-based signaling protocol.

Then, switch-local traffic control module sets up the cell switching fabric appropriately.

The ATM switch signaled by IP is, in general, called CSR (Cell Switching Routers) [CSR1, CSR2].

CSR is merely one of several ways to build a router and this memo does not recommend nor discourage to deploy the technology.

5. Conclusion.

RFC 1620 was wrong.

While it suggests several ways to have shortcuts, note that the discussions in section 2 and 3 does not depend on how the shortcuts are created. That is, modifications on the way to have shortcuts does not affect the conclusion that they are no good.

It is not necessary nor possible to modify the CATENET model, the architecture of the Internet, to have efficient and scalable Internet to accommodate shared medium such as ATM.

Shortcutting attempt, such as NHRP, may still be used in LAN or WAN NBMA environment with a small number of hosts. But, if the number of hosts is small, it is often, if not always, possible to make the entire NBMA a single LIS. Anyway, these local optimization does not affect the global architecture of the Internet.

The Simple Unified Networking with the CATENET model is the way to go.

6. Acknowledgements

Thank you Joel Halpern Sam Wilson and other members of ION working group for constructive comments to improve the quality of the memo.

7. References

[CSR1] Hiroshi ESAKI, Ken-ichi NAGAMI, Masataka OHTA, "High Speed Datagram Delivery over Internet using ATM Technology", Networld+Interop '95 Engineer Conference, E12-1~E12-9, (1995).

[CSR2] Yukinori GOTO, Masataka OHTA, Masaki HIRABARU, "Design of Internet Resource Reservation on ATM Network", Proceedings of The 10th International Conference on Information Networking (ICOIN-10), pp.510-516, 1996.

[IFMP]

[RFP1620]

[RSVP]

[ST2]

8. Security Considerations

(to be filled)

9. Author's Address

Masataka Ohta
Computer Center
Tokyo Institute of Technology
2-12-1, O-okayama, Meguro-ku
Tokyo 152, JAPAN

Phone: +81-3-5734-3299
Fax: +81-3-5734-3415
EMail: mohta@necom830.hpcl.titech.ac.jp