Yakov Rekhter
Cisco Systems
Dilip Kandlur
T.J. Watson Research Center, IBM Corp.
Curtis Villamizar
ANS
November 1995

"Local/Remote" Forwarding Decision in Switched Data Link Subnetworks
<draft-ietf-rolc-apr-03.txt>

Status of this Memo

Abstract

The IP architecture assumes that each Data Link subnetwork is labeled with a single IP subnet number. A pair of hosts with the same subnet number communicate directly (with no routers); a pair of hosts with different subnet numbers always communicate through one or more routers. As indicated in RFC1620, these assumptions may be too

restrictive for large data networks, and specifically for networks based on switched virtual circuit (SVC) based technologies (e.g. ATM, Frame Relay, X.25), as these assumptions impose constraints on communication among hosts and routers through a network. The restrictions may preclude full utilization of the capabilities provided by the underlying SVC-based Data Link subnetwork. This document describes extensions to the IP architecture that relaxes these constraints, thus enabling the full utilization of the services provided by SVC-based Data Link subnetworks.

1 Background

The following briefly recaptures the concept of the IP Subnet. The topology is assumed to be composed of hosts and routers interconnected via links (Data Link subnetworks). An IP address of a host with an interface attached to a particular link is a tuple <prefix length, address prefix, host number>, where host number is unique within the subnet address prefix. When a host needs to send an IP packet to a destination, the host needs to determine whether the destination address identifies an interface that is connected to one of the links the host is attached to, or not. This referred to as the "local/remote" decision. The outcome of the "local/remote" decision is based on (a) the destination address, and (b) the address and the prefix length associated with the the local interfaces. If the outcome is "local", then the host resolves IP address to Link Layer address (e.g. by using ARP), and then sends the packet directly to that destination (using the Link layer services). If the outcome is "remote", then the host uses one of its first-hop routers (thus relying on the services provided by IP routing).

To summarize, two of the important attributes of the IP subnet model are:

hosts with a common subnet address prefix are assumed to be attached to a common link (subnetwork), and thus communicate with each other directly, without any routers - "local";

hosts with different subnet address prefixes are assumed to be attached to different links (subnetworks), and thus communicate with each other only through routers - "remote".

A typical example of applying the IP subnet architecture to an SVC-based Data Link subnetwork is "Classical IP and ARP over ATM" (RFC1577). RFC1577 provides support for ATM deployment that follows the traditional IP subnet model and introduces the notion of a Logical IP Subnetwork (LIS). The consequence of this model is that a host is required to setup an ATM SVC to any host within its LIS; for destinations outside its LIS the host must forward packets through a router. It is important to stress that this "local/remote" decision is based solely on the information carried by the destination address and the address and prefix lengths associated with the local interfaces.

2 Motivations

The diversity of TCP/IP applications results in a wide range of traffic characteristics. Some applications last for a very short time and generate only a small number of packets between a pair of communicating hosts (e.g. ping, DNS). Other applications have a short lifetime, but generate a relatively large volume of packets (e.g. FTP). There are also applications that have a relatively long lifetime, but generate relatively few packets (e.g. Telnet). Finally, we anticipate the emergence of applications that have a relatively long lifetime and generate a large volume of packets (e.g. video-conferencing).

SVC-based Data Link subnetworks offer certain unique capabilities that are not present in other (non-SVC) subnetworks (e.g. Ethernet, Token Ring). The ability to dynamically establish and tear-down SVCs between communicating entities attached to an SVC-based Data Link subnetwork enables the dynamic dedication and redistribution of certain communication resources (e.g. bandwidth) among the entities. This dedication and redistribution of resources could be accomplished by relying solely on the mechanism(s) provided by the Data Link layer.

The unique capabilities provided by SVC-based Data Link subnetworks do not come "for free". The mechanisms that provide dedication and redistribution of resources have certain overhead (e.g. the time needed to establish an SVC, resources associated with maintaining a state for an SVC). There may also be a monetary cost associated with establishing and maintaining an SVC. Therefore, it is very important to be cognizant of such an overhead and to carefully balance the benefits provided by the mechanisms against the overhead introduced by such mechanisms.

One of the key issues for using SVC-based Data Link subnetworks in the TCP/IP environment is the issue of switched virtual circuit (SVC) management. This includes SVC establishment and tear-down, class of service specification, and SVC sharing. At one end of the spectrum one could require SVC establishment between communicating entities (on a common Data Link subnetwork) for any application. At the other end of the spectrum, one could require communicating entities to always go through a router, regardless of the application. Given the diversity of TCP/IP applications, either extreme is likely to yield a suboptimal solution with respect to the ability to efficiently exploit capabilities provided by the underlying Data Link layer.

The traditional IP subnet model is too restrictive for flexible and adaptive use of SVC-based Data Link subnetworks - the use of a subnetwork is driven by information completely unrelated to the characteristics of individual applications. To illustrate the problem consider "Classical IP and ARP over ATM" (RFC1577). RFC1577 provides support for ATM deployment that follows the traditional IP subnet model, and introduces the notion of a Logical IP Subnetwork (LIS). The consequence of this model is that a host is required to setup an SVC to any host within its LIS, and it must forward packets to destinations outside its LIS through a router. This "local/remote" forwarding decision is based solely on the information carried in the source and destination addresses and the subnet mask associated with the source address, and has no relation to the nature of the applications that generated these packets. This leads to a situation where SVC management is controlled by totally irrelevant factors.

3 QoS/Traffic Driven "Local/Remote" Decision

Consider a host attached to an SVC-based Data Link subnetwork, and assume that the "local/remote" decision the host could make is not constrained by the IP subnet model. When such a host needs to send a packet to a destination, the host might consider any of the following options:

Use a best-effort SVC to the first hop router.

Use a SVC to the first hop router dedicated to a particular type of service (ie: predictive real time).

Use a dedicated SVC to the first hop router.

Use a best-effort SVC to a router closer to the destination than the first hop router.

Use a SVC to a router closer to the destination than the first hop router dedicated to a particular type of service.

Use a dedicated SVC to a router closer to the destination than the first hop router.

Use a best-effort SVC directly to the destination (if the destination is on the same Data Link subnetwork as the host).

Use a SVC directly to the destination dedicated to a particular type of service (if the destination is on the same Data Link subnetwork as the host).

Use a dedicated SVC directly to the destination (if the destination is on the same Data Link subnetwork as the host).

We may observe that the forwarding decision at the host is more flexible than the "local/remote" decision of the IP subnet model. We may also observe that the host's forwarding decision allows to take into account QoS and/or traffic requirements of the applications and/or cost factors associated with establishing and maintaining a VC, and thus improve the overall SVC management. Therefore, removing constraints imposed by the IP subnet model is an important step towards better SVC management.

3.1 Extending the scope of possible "local" outcomes

A source may have a SVC (either dedicated or shared) to a destination if both the source and the destination are on a common Data Link subnetwork. The ability to have the SVC (either dedicated or shared) is completely decoupled from the source and destination IP addresses, but is coupled to the QoS and/or traffic characteristics of the application. In other words, the ability to establish a direct VC (either dedicated or shared) between a pair of hosts on a common Data Link subnetwork has nothing to do with the IP addresses of the hosts. In contrast with the IP subnet model (or the LIS mode), the "local" outcome becomes divorced from the addressing information.

3.2 Allowing the "remote" outcome where applicable

A source may go through one or more routers to reach a destination if either (a) the destination is not on the same Data Link subnetwork as the source, or (b) the destination is on the same Data Link subnetwork as the source, but the QoS and/or traffic requirements of the application on the source do not justify a direct (either dedicated or shared) VC.

When the destination is not on the same Data Link subnetwork as the source, the source could select between either (a) using its first-hop (default) router, or (b) establishing a "shortcut" to a router closer to the destination than the first-hop router. The source should be able to select between these two choices irrespective of the source and destination IP addresses.

When the destination is on the same Data Link subnetwork as the source, but the QoS and/or traffic requirements do not justify a direct VC, the source should be able to go through a router irrespective of the source and destination IP addresses.

In contrast with the IP subnet model (or the LIS model) the "remote" outcome, and its particular option (first-hop router vs router closer to the destination than the first-hop router), becomes divorced from the addressing information.

3.3 Sufficient conditions for direct connectivity

The ability of a host to establish an SVC to a peer on a common switched Data Link subnetwork is predicated on its knowledge of the Link Layer address of the peer or an intermediate point closer to the destination. This document assumes the existence of mechanism(s) that can provide the host with this information. Some of the possible alternatives are NHRP, ARP, or static configuration; other alternatives are not precluded. The ability to acquire the Link Layer address of the peer should not be viewed as an indication that the host and the peer can establish an SVC - the two may be on different Data Link subnetworks, or may be on a common Data Link subnetwork that is partitioned.

3.4 Some of the implications

Since the "local/remote" decision would depend on factors other than the addresses of the source and the destination, a pair of hosts may simultaneously be using two different means to reach each other, forwarding traffic for applications with different QoS/and or traffic characteristics differently.

3.5 Address assignment

It is expected that if the total number of hosts and routers on a common SVC-based Data Link subnetwork is sufficiently large, then the hosts and routers could be partitioned into groups, where each group would have hosts and routers. The routers within a group would act as the first-hop routers for the hosts in the group. If the total number of hosts and routers is not large, then all these hosts and routers could form a single group. Criteria for determining group sizes are outside the scope of this document.

To provide scalable routing each group should be given an IP address prefix, and elements within the group should be assigned addresses out of this prefix. The routers in a group would then advertise (via appropriate routing protocols) routes to the prefix associated with the group. These routes would be advertised as "directly reachable" (with metric 0). Thus, routers within a group would act as the last-hop routers for the hosts within the group.

3.6 Using Point to Point Links

If RFC1577 is used as an underlying model, the router based overlay is assumed to be comprised of LIS. The LIS model may not be appropriate for some situation. A large group may be served by a single router or a small number of routers to provide some redundancy.

The desired behavior can be accomplished using RFC1577 LIS by allowing the LIS to become very small, containing two hosts each and relaxing the restriction prohibiting shortcut VC. If the all zeros network address is to be reserved, a four address LIS is needed for each host in the group. ATM ARP service is also needed. The LIS model clearly has some inefficiencies for such a case.

An alternate way to model the relation among hosts and routers within each group is by modelling VC as point to point links. Using numbered point to point links, the two addreses on the link need not be adjacent. If each end of the link has a distinct address, the model is a numbered point to point link. A router may allow the use of a single address for the entire router, reusing the same address as the near end of all of the point to point connections. Similarly, a host can use a single address for point to point links to more than one router, if a backup router is used. This model is an unnumbered point to point link, unnumbered since the link endpoints are not uniquely numbered, just the nodes. For SNMP manageability, on unnumbered point to point links, the far end address can be used for identification of a link.

4 Conclusions

Different approaches to SVC-based Data Link subnetworks used by TCP/IP yield substantially different results with respect to the ability of TCP/IP applications to efficiently exploit the functionality provided by such subnetworks. For example, in the case of ATM both LAN Emulation [LANE] and "classical" IP over ATM [RFC1577] localize host changes below the IP layer, and therefore may be good first steps in the ATM deployment. However, these approaches alone are likely to be inadequate for the full utilization of ATM.

It appears that any model that does not allow SVC management based on QoS and/or traffic requirements will preempt the full use of SVC-based Data Link subnetworks. Enabling more direct connectivity for applications that could benefit from the functionality provided by SVC-based Data Link subnetworks, while relying on strict hop by hop paths for other applications, could facilitate exploration of the capabilities provided by the subnetworks.

While this document does not define any specific coupling between various QoS, traffic characteristics and other parameters, and SVC management, it is important to stress that efforts towards standardization of various QoS, traffic characteristics, and other parameters than an application could use (through an appropriate API) to influence SVC management are essential for flexible and adaptive use of SVC-based Data Link subnetworks.

The proposed model utilizes the SVC-based infrastructure for the applications that could benefit from the capabilities supported within such an infrastructure, and take advantage of a based router-based overlay for all other applications. The router based overlay could be based on RFC1577 if the restriction prohibiting subnet shortcut connections is eliminated from RFC1577, replacing it with a statement that any shortcut requires mechanisms beyond what is described in RFC1577. Use of a point to point model is suggested as an alternate to the LIS model for situations where the NBMA LIS model simply does not fit. As such it provides a balanced mix of router-based and switch-based infrastructures, where the balance could be determined by the applications requirements.

The approach proposed in this document combines switch-based infrastructure with router-based overlay and uses each for that which it is best suited: switch-based infrastructure for applications that can justify an SVC establishment; router-based overlay for all other applications.

6 Security Considerations

Security issues are not discussed in this document.

7 Acknowledgements

The authors would like to thank Joel Halpern (NewBridge), Allison Mankin (ISI), Tony Li (cisco Systems), Andrew Smith (BayNetworks) for their review and comments.

References

[LANE] "LAN Emulation over ATM specification- version 1", ATM Forum, Feb.95.

[Postel 81] Postel, J., Sunshine, C., Cohen, D., "The ARPA Internet Protocol", Computer Networks, 5, pp. 261-271, 1983.

[RFC792] Postel, J., "Internet Control Message Protocol- DARPA Internet Program Protocol Specification", STD 5, RFC 792, ISI, September 1981.

[RFC1122] Braden, R., Editor, "Requirements for Internet Hosts - Communication Layers", STD 3, RFC 1122, USC/ISI, October 1989.

[RFC1577] Laubach, M., "Classical IP and ARP over ATM", January 1994.

[RFC1620] Braden, B., Postel, J., Rekhter, Y., Internet Architecture Extensions for Shared Media", May 1994.

[RFC1755] Perez, M., Liaw, F., Grossman, D., Mankin, A., Hoffman, E., Malis, A., "ATM Signalling Support for IP over ATM", January 1995.

14 Authors' Address

Yakov Rekhter
Cisco Systems
170 West Tasman Drive,
San Jose, CA 95134-1706
Phone: (914) 528-0090
email: yakov@cisco.com

Dilip Kandlur
T.J. Watson Research Center IBM Corporation
P.O. Box 704
Yorktown Heights, NY 10598
Phone: (914) 784-7722
email: kandlur@watson.ibm.com

Curtis Villamizar
ANS
100 Clearbrook Road
Elmsford, N.Y. 10523
email: curtis@ans.net