Internet-Draft
Expiration Date: September 1997

Steven Blake
Anoop Ghanwani
Wayne Pace
Vijay Srinivasan

IBM Corporation

March 1997

ARIS Support for LAN Media Switching

<draft-blake-aris-lan-00.txt>

Status of This Memo

Abstract

ARIS (Aggregate Route-based IP Switching) [ARIS] is a protocol which, in coordination with network-layer routing protocols, establishes link-layer switched paths through a network of Integrated Switch Routers (ISR). This memo describes ARIS protocol mechanisms which enable LAN media switching of IP packets. In addition, this memo describes the functional behavior of ISRs which are interconnected via LAN media (e.g., ethernet, token ring, FDDI). The proposed mechanisms are designed to permit easy implementation using emerging LAN switching technology.

1. Applicability Statement

Several proposals that deal with improving the performance of IP forwarding by carrying labels in the packets have recently been submitted to the MPLS working group [ARIS, TAG, FANP]. The labels are used for indexing tables which enable fast IP forwarding at close to media speeds by minimizing the need for network-layer packet processing. The labels may be carried in different ways depending on the underlying link-layer technology. For instance, in ATM networks, the label may be represented by a particular VPI/VCI value. Since ATM is a label swapping technology, it is possible for label allocation to be a local choice for each node participating in the protocol. This is not possible for LAN switching technologies such as ethernet, token ring, and FDDI, which are not inherently label swapping technologies. As a consequence, a shim consisting of one or more 32-bit label stack entries inserted between the link-layer and the network-layer headers has been proposed as a means to convey the label information [LABEL]. The main drawback of using such an approach is that accessing the labels requires that the frames be processed by software, reducing the benefit offered by label switching. Alternatively, hardware technology specific to label switching may be developed. However, devices incorporating this technology are likely to be more expensive that traditional LAN switches and bridges.

This memo proposes a different approach for label switching on LAN media which uses the ARIS protocol for distribution of labels. The label is carried in the destination address portion of the frame and, for unicast, is usually the MAC address of the egress point from the network as identified by ARIS. With this approach, an implementation using emerging bridge/switch hardware capable of supporting the IEEE 802.1p forwarding and filtering rules is possible [802.1P]. However, the labels must now have global significance and are required to be unique. The focus of this memo is to describe a label distribution and switching mechanism which can be applied among ISRs which are interconnected via point-to-point LAN media links. Such a mechanism can provide significant benefit in the backbone of campus networks, for example.

2. Introduction

An Integrated Switch Router (ISR) is a link-layer switch which has been augmented with IP routing capability, in addition to the ARIS protocol [ARIS]. Virtual circuits (VCs) which are established by application of the ARIS protocol enable switching of IP packets across a network of ISRs. Here the term "virtual circuit" is used loosely to imply a switched path in any switching technology.

ARIS switched path establishment is coupled to IP routing by means of the "egress identifier". An egress identifier may refer to an egress

ISR which forwards traffic either to a foreign routing domain, or across an area boundary within the same network. Alternatively, an egress identifier may refer to a particular (S,G) multicast pair. ARIS supports a wide variety of egress identifier semantics, each providing a different level of traffic aggregation.

In the unicast traffic case, ARIS establishes a switched path for each egress identifier advertised by an ISR by forwarding an Establish message to each of that ISR's upstream neighbors. After ensuring that the downstream ISR is on the routed path associated with the egress identifier, and that the switched path is loop-free, the upstream ISRs continue to forward Establish messages further upstream until they reach all ingress ISRs in the ARIS network. The resulting switched path resembles a multipoint-to-point tree terminating at the egress ISR. The direction of path establishment is reversed for multicast traffic, and the resulting switched path forms a point-to-multipoint tree.

Each ARIS switched path for an egress identifier is associated with a unique VC between adjacent ISRs. ISRs typically will swap the VPI/VCI field of a cell (ATM) or the DLCI of a frame (Frame Relay) with a new label value before forwarding to a downstream ISR on the switched path. This operation is commonly referred to as "label swapping". ISRs can merge multiple inbound VCs of a switched path onto a single outbound VC if the underlying hardware supports this capability. This reduces VC consumption and affords greater network scalability.

Unlike ATM and Frame Relay, traditional LAN switching technology is not based on label swapping. LAN switches forward a LAN frame based on the 6-byte destination IEEE MAC address (DA) encoded in the frame header [802.1D]. LAN switching hardware typically is not capable of swapping the DA in the frame prior to forwarding. This style of forwarding is referred to here as "label switching". LAN switches are only able to forward frames unambiguously if each (individual) DA is associated with only one network end-point. This requirement does not present a significant limitation on network scalability since the DA field is large enough to represent 2^46 unique end-points.

ARIS supports LAN media switching by associating each egress identifier with a 6-byte switching label which is unique among all switching labels in use within the ARIS network. The switching label is encoded in the DA field within a LAN frame. ISRs cache the switching label corresponding to each switched path. ISRs can unambiguously identify frames corresponding to any particular egress identifier by the value of the frame's DA field and can forward them directly at the link-layer along the appropriate switched path. This enables packets to be switched at hardware speeds across an entire

network of ISRs.

Unless otherwise specified, the behavior of the ARIS protocol is

identical to that described in [ASPEC].

3. Components for LAN Media Switching

In this memo, a LAN Media ISR (LMISR) refers to a network node which incorporates a LAN Media Forwarding Component (LMFC) along with a network-layer control and forwarding component (IPCC). The LMFC performs label switching based on the 6-byte DA of a received frame. Direct link-layer switching between diverse LAN media types (10/100/1000 ethernet, token ring, FDDI) is possible if supported by the underlying hardware. LMISRs are intended to form the core of a scalable, high-bandwidth campus or enterprise network.

Associated with each LMFC is a LAN Media Forwarding Information Base (LMFIB). The LMFIB specifies the association of switching-labels (DAs) to outgoing interface(s). This table is used to configure the LMFC's filtering database to enable link-layer forwarding. In the default configuration, the 802.1D Spanning Tree protocol is disabled, and every active interface (visible to IP routing) is placed in the forwarding state [802.1D].

To permit IP control traffic to reach the IPCC within a LMISR, and to permit network-layer forwarding of packets on a switched path which has been broken downstream, the IPCC is associated with one or more logical interfaces in the LMFIB. This allows the IPCC to redirect packets on a pre-established switched path through the IPCC. The IPCC implementation SHOULD be capable of simultaneously receiving LAN frames with arbitrary DA values. Note that the LMFIB can be used to filter the addresses which are received by the IPCC.

The LMFC MUST permit the precise specification of the output interface(s) to be associated with each received DA (individual or group address scope). This capability is consistent with the Extended Filtering Mode and Port Filtering Mode C as described in Section 2.6.6 of [802.1P].

The LMFC MUST NOT flood frames with an unknown DA or with the broadcast DA out of every LMFC interface in forwarding state. These rules are necessary to prevent link-layer loops from forming amongst adjacent LMISRs. The LMFC SHOULD support the ability to forward frames with an unknown DA or with the broadcast DA to a particular LMFC interface associated with the IPCC. In addition, the LMFC SHOULD support the ability to drop frames with an unknown DA or with the broadcast DA.

Existing LAN switch implementations typically do not support the capability to swap the DA of a frame. ARIS does not require this capability to function efficiently, but allows LMISRs whose LMFCs are

capable of DA swapping to alter the switching label associated with an egress identifier when forwarding Establish messages upstream

towards an ingress ISR. It is the responsibility of such a LMISR to select a unique 6-byte switching label when transmitting an Establish message for the associated egress identifier, and to perform the correct DA swapping operation to/from the initial DA value when forwarding frames.

4. LAN Media Frame Encapsulation

ARIS support for LAN media switching does not require a new encapsulation format for IP packets. IPv4 and IPv6 packets should be encapsulated according to the appropriate RFC specification for each LAN media [RFC1042, RFC1972, RFC2019]. This includes the default value of the maximum transmission unit (MTU) for each LAN media link.

5. IP Multicast Support

As described in [ARIS], the establishment of point-to-multipoint switched paths for IP multicast traffic is initiated at the root (ingress) node. The switched path tree forwards traffic from the ingress ISR to all egress ISRs on the multicast tree by using multicast switching at the intermediate ISRs.

The ingress LMISR for a multicast switched path tree forwards an Establish message containing the switching label for the associated egress identifier to its downstream LMISRs. The Establish message traverses from the ingress node to the downstream LMISRs in reverse path multicast (RPM) style. The branches of the point-to-multipoint tree that do not lead to receivers are pruned when the multicast routing protocol prunes up by deleting forwarding entries in the LMFIB. The ingress LMISR periodically refreshes the multicast switched path tree by retransmitting an Establish message containing the switching label for the associated egress identifier.

6. Multipath Support

As described in Section 2, a single switching label is associated with an egress identifier in the default configuration. In this case, a LMISR which has received multiple Establish messages for an egress identifier, each associated with an equal-cost path to the corresponding egress LMISR, cannot forward multiple Establish messages with the same switching label to each of its upstream LMISRs, since this will not allow the upstream LMISRs to distinguish the multiple equal-cost paths.

An LMISR which wishes to utilize multiple equal-cost paths to an egress has the following alternatives:

o Forward only one Establish message for an egress identifier to

each upstream LMISR, and forward traffic on that switched path at the IP layer,

- o Forward multiple Establish messages for an egress identifier to each upstream LMISR, where each Establish message contains a distinct switching label (all but one of which must be generated dynamically by the LMISR). The LMISR must be capable of DA swapping between the dynamically generated label(s) and the original label selected by the egress LMISR.

7. Explicit Route Support

Explicit routes for point-to-point, point-to-multipoint, and multipoint-to-point forwarding are established as described in [ARIS]. In the case of point-to-point explicit routes, either the ingress or the egress may initiate the path establishment, and may select the switching label. In the case of multipoint-to-point explicit routes, the egress initiates the switched path establishment and selects the switching label. In the case of point-to-multipoint explicit routes, the ingress initiates the switched path establishment and selects the switching label.

8. Security Considerations

An analysis of security considerations will be provided in a future revision of this memo.

9. Intellectual Property Considerations

International Business Machines Corporation may seek patent or other intellectual property protection for some or all of the aspects discussed in the forgoing document.

10. Acknowledgements

The authors wish to acknowledge the following individuals for their input and assistance: Rick Boivie, Ed Bowen, Brian Carpenter, Allen Carriker, Gene Cox, Ed Ellesson, Jim Ervin, Nancy Feldman, John Linville, Sanjeev Rampal, Norm Strole, Arun Viswanathan, and Jeff Warren.

11. References

[ARIS] A. Viswanathan, N. Feldman, R. Boivie, R. Woundy,

"ARIS: Aggregate Route-Based IP Switching", Internet Draft
<draft-viswanathan-aris-overview-00.txt>, March 1997.

[TAG] Y. Rekhter, B. Davie, D. Katz, E. Rosen, G. Swallow, D. Farinacci, "Tag Switching Architecture - Overview", Internet Draft <draft-rekhter-tagswitch-arch-00.txt>, January 1997.

[FANP] K. Nagami, Y. Katsube, Y. Shobatake, A. Mogi, S. Matsuzawa, T. Jinmei, H. Esaki, "Flow Attribute Notification Protocol (FANP) Specification", Internet Draft <draft-rfced-info-nagami-00.txt>, February 1997.

[LABEL] E. Rosen, Y. Rekhter, D. Tappan, D. Farinacci, G. Fedorkow, "Label Switching: Label Stack Encodings", Internet Draft <draft-rosen-tag-stack-01.txt>, March 1997.

[802.1P] "P802.1p Standard for Local and Metropolitan Area Networks- Supplement to Media Access Control (MAC) Bridges: Traffic Class Expediting and Dynamic Multicast Filtering", P802.1p/D5, LAN MAN Standards Committee, IEEE Computer Society, February 1997.

[802.1D] ISO/IEC 10038, ANSI/IEEE Std 802.1D-1993 "MAC Bridges".

[ASPEC] N. Feldman, A. Viswanathan, "ARIS Specification", Internet Draft <draft-feldman-aris-spec-00.txt>, March 1997.

[RFC1042] J. Postel, J. Reynolds, "A Standard for the Transmission of IP Datagrams over IEEE 802 Networks, Internet RFC 1042, February 1988.

[RFC1972] M. Crawford, "A Method for the Transmission of IPv6 Packets over Ethernet Networks", Internet RFC 1972, August 1996.

[RFC2019] M. Crawford, "A Method for the Transmission of IPv6 Packets over FDDI Networks", Internet RFC 2019, October 1996.

12. Authors' Addresses

Steven Blake
IBM Corporation
P.O. Box 12195
Research Triangle Park, NC 27709
Phone: +1-919-254-2030
Fax: +1-919-254-5483
E-mail: slblake@vnet.ibm.com

Anoop Ghanwani
IBM Corporation
P.O. Box 12195

Research Triangle Park, NC 27709
Phone: +1-919-254-0260

Fax: +1-919-254-5410
E-mail: anoop@raleigh.ibm.com

Wayne Pace
IBM Corporation
P.O. Box 12195
Research Triangle Park, NC 27709
Phone: +1-919-254-4930
Fax: +1-919-254-5410
E-mail: pacew@raleigh.ibm.com

Vijay Srinivasan
IBM Corporation
P.O. Box 12195
Research Triangle Park, NC 27709
Phone: +1-919-254-2730
Fax: +1-919-254-5410
E-mail: vijay@raleigh.ibm.com