Internet-Draft
Expiration Date: September 1997

Arun Viswanathan
Nancy Feldman
Rick Boivie
Rich Woundy
IBM Corp.

March 1997

ARIS: Aggregate Route-Based IP Switching

<draft-viswanathan-aris-overview-00.txt>

Status of This Memo

Abstract

IP based networks use a number of routing protocols, including RIP, OSPF, IS-IS, and BGP, to determine how packets ought to be routed. Among these protocols, OSPF and BGP are IETF-recommended standards that have been extensively deployed and exercised in many networks. In this memo, we describe a mechanism which uses these protocols as the basis for switching IP datagrams, by the addition of a simple protocol ("ARIS") that establishes switched paths through a network. The ARIS protocol allows us to leverage the advantages of switching technologies in an internet network.

1. Introduction

In this memo, an Integrated Switch Router (ISR), is a switch that has been augmented with standard IP routing support. The ISR at an entry point to the switching environment performs standard IP forwarding of datagrams, but the "next hop" of the IP forwarding table has been extended to include a reference to a switched path (for example, the VCC in ATM technology). Each switched path may have an endpoint at a neighboring router (comparable to today's IP next hops on conventional routers), or may traverse a series of ISRs along the best IP forwarding path, to an egress ISR endpoint. This allows datagrams to be switched at hardware speeds through an entire ISR network.

The key link between the IP network routing protocols and the ARIS switched path establishment protocol is the "egress identifier", which defines a routed path through a network. The egress identifier may refer to an egress ISR that forwards traffic either to a foreign routing domain, or across an area boundary within the same network. ARIS establishes switched paths towards each unique egress identifier. Since thousands of IP destinations can map to the same egress identifier, ARIS minimizes the number of switch paths required in an ISR network. This allows a large network to switch all of its IP traffic, resulting in improved aggregate IP throughput.

2. ARIS Mechanism

In networks based on destination-based hop-by-hop forwarding, ARIS [ARIS-SPEC] pre-establishes switched paths to "well known" egress nodes. As a result, virtually all best-effort traffic is switched through an ARIS network. These "well known" egress nodes are learned through the routing protocols, such as OSPF and BGP. No routing protocol modification is required for this purpose, as this information is already present within the routing protocols themselves.

Egress ISRs initiate the setup of switched paths by sending Establish messages to their upstream neighbors, typically within the same domain. These upstream neighbors forward the messages to their own upstream neighbors in Reverse Path Multicast style, after ensuring that the switched path is loop-free. Eventually, all ISRs establish switched paths to all egress ISRs, which follow the routed path.

The switched path to an egress point, in general, takes the form of a tree. A tree results because of the "merging" of switched paths that occurs at a node when multiple upstream switched paths for a given egress point are spliced to a single downstream switched path for

that egress point.

2.1. ARIS Messages

ARIS is protocol independent. In the case of IP, ARIS message are transmitted with IP protocol number 104. ARIS uses the following messages to manage the switched paths.

Init
: This is the first message sent by an ISR to each of its neighbors, as notification of its existence. It is periodically transmitted until a positive acknowledgment is received. The Init message may include the neighbor timeout period, acceptable label ranges, and other adjacency information.

KeepAlive
: This message is sent by an ISR to inform its neighbors of its continued existence. It is the first message that is transmitted after an adjacency has been established. In order to prevent the neighbor timeout period from expiring, ARIS messages must be periodically sent to neighbors. The KeepAlive will only be sent when no other ARIS messages have been transmitted within the periodic interval time.

Establish
: This message is initiated by the egress ISR, and is periodically sent to each upstream neighbor to setup or refresh a switched path. It is also sent by any ISR in response to a Trigger message. Each ISR that receives an Establish message for an egress identifier must verify that the path is correct and loop free. If the Establish message changes a previously known switched path to the egress identifier, the ISR unsplices the obsolete switched path. The ISR creates a downstream switched path with the given label for the egress identifier, and replies with an Acknowledgment message. It then allocates a label for each of its upstream neighbors, forwards the Establish message to the upstream neighbors with its unique ISR ID appended to the ISR ID path and the label for the upstream neighbor to use for forwarding, and waits for an Acknowledgment message. This pattern continues until all ISRs are reached.

Trigger
: This message is sent by an ISR when it has detected that a local IP routing change has modified its path to the egress identifier. After unsplicing the obsolete switched path, the ISR sends a Trigger message to its new downstream neighbor requesting an Establish message.

Teardown
This message may be sent when an ISR has lost connectivity to an egress identifier. The message follows the path of the Establish message unsplicing and releasing the obsolete switched path.

Acknowledgment
This message is sent as a response to ARIS messages. When an ISR receives a positive acknowledgment to an Establish message, it splices the upstream label to the downstream label creating a switched path through the ISR.

2.2. Egress Identifiers

The ARIS protocol uses egress identifiers that balance the desire to share the same egress identifier among many IP destination prefixes, with the desire to maximize switching benefits. To provide flexibility, ARIS supports many types of egress identifiers. ISRs choose the type of egress identifier to use based on routing protocol information and local configuration.

The first type of egress identifier is the IP destination prefix. This type results in each IP destination prefix sustaining its own switched path tree, and thus will not scale in large backbone and enterprise networks. However, this is the only information that some routing protocols, such as RIP, can provide. This type of identifier may work well in networks where the number of destination prefixes is limited, such as in campus environments, or even in a wide-area network of a private enterprise.

The second type of egress identifier is the egress IP address. This type is used primarily for BGP protocol updates, which carry this information in the NEXT_HOP attribute [RFC1771]. There are certain types of OSPF routes that also use this type.

The third type of egress identifier is the OSPF Router ID, which allows aggregation of traffic on behalf of multiple datagram protocols routed by OSPF. The latest version of OSPF supports the Router ID for both IP and IPv6 [RFC1583].

The fourth type of egress identifier is the multicast (source, group) pair [RFC1112], used by multicast protocols, such as DVMRP [RFC1075], MOSPF [RFC1584] and PIM ([PIM-SM], [PIM-DM]). The fifth is the (ingress-of-source, group), used for such multicast protocols as MOSPF and PIM-SM.

Other egress identifier types may be defined, including but not limited to IS-IS NSAP addresses, NLSP IPX addresses, IPv6 destination

prefixes, APPN etc.

A hierarchy amongst the egress identifiers may be introduced to allow more flexible control over egress identifier selection. This allows an ISR to autolearn or be configured with non-default egress identifiers, and to select which egress identifiers to use in various routing situations.

2.3. ISR Information Base

The ISR needs three logical information bases to compute routes and forward datagrams: the routing information base, the forwarding information base, and the VC information base. The first, the routing information base (RIB), is used for the computation of best-effort routes by various IP routing protocols. The RIB for the ISR is essentially unchanged from the RIB on a standard router. In the ISR context, the RIB is also used to identify egress points and egress identifiers for the other two information bases.

The forwarding information base (FIB) of the ISR has been extended beyond the content of the FIB on a standard router to include an egress identifier in each next hop entry. The FIB tends to contain many IP destination prefix entries, which point to a small number of next hop entries that describe the hop-by-hop forwarding operation(s). Next hop entries on the ISR consist of an outgoing interface, next hop IP address, egress identifier, and the associated established downstream label for the switched path. The association of the next hops with the egress identifiers is the responsibility of the routing protocols, while the association of the next hop/egress identifiers with the established switched paths is the responsibility of the ARIS protocol.

The VC information base (VCIB), which does not exist on a standard router, maintains for each egress identifier the upstream to downstream label mappings and related states. This mapping is controlled by the ARIS protocol.

2.4. Forwarding

The forwarding ingress ISR performs a conventional longest prefix match lookup in its FIB, which returns the associated switched path label for the particular destination. The ingress ISR may also decrement the TTL by the length of the switched path before the packet is transmitted on the switched path. If no associated switched path is found in the FIB, the ingress node may forward the packet to the next hop via the default hop-by-hop switched path.

2.5. TTL Decrement

In order to comply with the requirements for IPv4 routers, the IP datagram Time-To-Live (TTL) field must be decremented on each hop it traverses [RFC1812]. Currently, switched packets within an ATM like networks cannot decrement the TTL. However, ARIS can decrement TTLs appropriately by maintaining a hop-count per egress identifier. This hop-count is calculated by including a hop-count field in the Establish message, which is incremented at each ISR as it traverses the upstream path. Before forwarding a packet on a switched path, an ingress ISR decrements the TTL by the hop-count plus one. If the decrement value is greater than or equal to the TTL of the packet, the packet may be forwarded hop-by-hop.

3. Loop Prevention

The ARIS protocol guarantees that switched path loops are prevented, even in the presence of transient IP routing loops. With datagram forwarding loops, each hop decrements the TTL, so traffic is eventually dropped. However, some switching technologies, such as ATM, do not have a counter similar to the TTL, so traffic persists in a switched path loop as long as the switched path loop exists. The same it true for Frame Relay and LAN switches. At best, the traffic in the switched path loop steals bandwidth from other (UBR) switched paths; at worst, the traffic interferes with IP routing traffic, slows down routing convergence, and lengthens the life of the switched path loop.

The ARIS protocol avoids creating switched path loops by the use of an "ISR ID" list, similar in function to the BGP AS_PATH attribute. Each ISR in the establishment path appends its own unique ISR ID to each establishment message it forwards. In this way, an ISR is able to determine the path a message has traversed, and can ensure that no loops are formed.

Further, if an ISR modifies or deletes an egress due to an IP route change, or receives a message that modifies an existing switched path to an egress, the ISR must unsplice any established upstream switched path from the downstream switched path. Hence transient IP routing loops, potentially created by the route change, cannot produce switched path loops. The ISR must then re-establish a new switched path to the modified egress. Note that ARIS does not attempt to suppress transient IP routing protocol loops; it only avoids establishing switched path loops with this information.

4. Egress ISRs

In the ARIS protocol, Establishment messages are originated from the egress ISR. An ISR is considered an egress ISR, with respect to a particular egress identifier, under any of the following conditions:

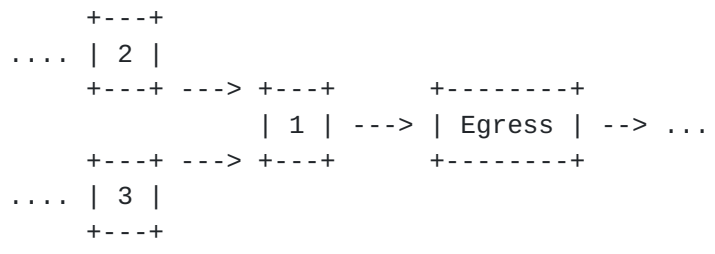
1. The egress identifier refers to the ISR itself (including one of its directly attached interfaces).

2. The egress identifier is reachable via a next hop router that is outside the ISR switching infrastructure.

3. The egress identifier is reachable by crossing a routing domain boundary, such as another area for OSPF summary networks, or another autonomous system for OSPF AS externals and BGP routes.

5. Examples

5.1. Establish Initiation Example



Example: Egress initiates Establish

a) The Egress ISR learns of an egress identifier that indicates the egress is itself (see "Egress ISRs"). It creates a FIB entry for its next hop and egress identifier (itself).

b) The Egress creates a VCIB entry with an allocated upstream label to ISR1, and initiates an Establish message with the upstream label, and itself in the ISR ID path.

c) ISR1 verifies that the Establish message was received from the expected next hop (Egress) by matching its FIB entry, and verifies that the ISR ID path is loop free. It then creates a VCIB entry and a switched path with the downstream label to the Egress, replaces the default switched path label in the FIB with

this new label, and replies to the Egress with an Acknowledgment message.

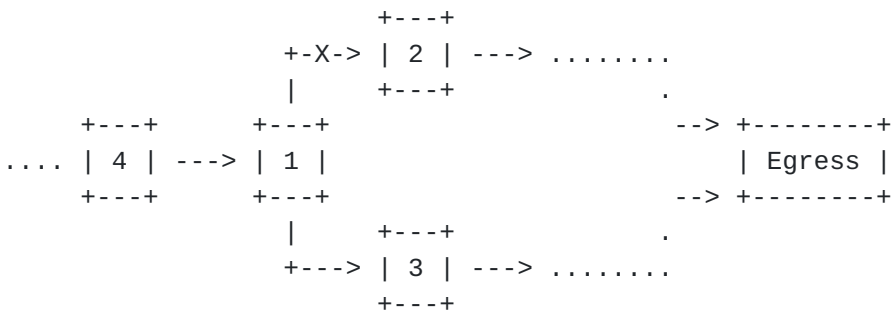
d) ISR1 allocates an upstream label to each of its upstream neighbors, ISR2 and ISR3, and updates the corresponding VCIB entry. It forwards the Establish message to each upstream neighbor, with its own ISR ID appended to the ISR ID path and with the label to use.

e) When ISR1 receives each acknowledgment from each upstream neighbor, it updates the VCIB and splices the corresponding upstream label to its Egress downstream label.

All upstream ISRs recursively follow the same procedures as ISR1, until all Ingress ISRs have been added to the switched path to the Egress.

The Egress ISR is responsible for periodically sending refresh Establish messages, to prevent switched path timeouts. If a refresh is not received in the allotted time, switched paths are unspliced and associated labels are released.

5.2. Trigger Example



Example: ISR1 changes routes from ISR2 to ISR3

a) ISR1 learns of a new path to the Egress via ISR3 from the routing protocols. It removes the FIB and VCIB entries for the next hop ISR2/Egress. ISR1 creates a new FIB entry for the next hop ISR3/Egress with the default switched path to the next hop.

b) ISR1 sends a Trigger message to new downstream node ISR3 requesting an Establish message for the switched path to the Egress.

c) ISR3 allocates an upstream label, updates its corresponding VCIB

entry, and replies with an Establish message to ISR1, containing the full ISR ID path and the label.

d) ISR1 verifies that the Establish message was received from the expected next hop (ISR3), and that the ISR ID path is loop free. It then creates a new VCIB entry and a switched path with the given downstream label to ISR3, and replaces the default switched path label in the FIB with this new label.

e) ISR1 sends an Acknowledgment message to ISR3.

f) ISR3 receives the Acknowledgment, updates the VCIB and splices its ISR1 upstream label to its downstream label.

g) ISR1 appends its ISR ID to the Establish message, and forwards the message to ISR4 with the upstream label.

h) ISR4 verifies the Establish message, updates the VCIB, and unsplices the current switched path to ISR1/Egress from its upstream node(s), and sends an Acknowledgment to ISR1.

i) ISR1 receives the Acknowledgment, updates the VCIB and splices the ISR4 upstream label to the ISR3 downstream label.

j) ISR4 appends its ISR ID to the path, and forwards the establishment message to its upstream neighbors with a label. When ISR4 receives an Acknowledgment from an upstream neighbor, it updates the VCIB and splices the upstream label to the ISR1 downstream label.

All upstream ISRs recursively follow the same procedure as ISR4, until all ingress ISRs have been updated.

6. Explicit Routes

Today's Internet is predominantly based on the destination-based hop-by-hop forwarding paradigm. However, other routing and forwarding paradigms, such as strict source routing, may be useful to provide specialized and customized services. For this reason, the ARIS protocol supports the building of switched paths through explicit routes.

This is enabled by introducing the explicit source route path information in the Establish message. The Establish message is forwarded along the explicit path as identified by the source route information. ARIS supports building of point-to-point, point-to-multipoint and multipoint-to-point switched paths either from the

egress node or the ingress node. Note that the switched paths built by source routing are guaranteed to be loop-free. It's also possible to set up bi-directional switched paths or switched paths with QoS using this approach.

7. Multicast

The ARIS protocol can be used to setup switched paths for IP multicast traffic. The establishment of a point-to-multipoint switched path tree is initiated at the root (ingress) node. The switched path tree carries traffic from the ingress ISR to all egress ISRs, using multicast switching at intermediate ISRs.

The choice of egress identifier for multicast routing protocols such as DVMRP and PIM-DM is the (S,G) pair itself. This egress identifier creates one ingress routed point-to-multipoint switched path tree per source address and group pair. The creation of a switched path is initiated by an ingress node on receipt of traffic from a certain sender for a particular multicast group. The Establish message traverses from the ingress node to the downstream ISRs in the Reverse Path Multicast (RPM) style. The branches of the point-to-multipoint switched path tree that do not lead to receivers are pruned when the multicast routing protocol prunes up by deleting forwarding entries in the multicast FIB.

Having multicast switched paths set up on the basis of (S,G) works well with the IGMPv3 Group-Source messages, since these IGMP messages can create unique trees for each sender within the same group [11].

For multicast routing protocols, such as PIM-SM, that use a shared tree, an appropriate choice of egress identifier is (*, G) or (RP, G) (where RP is the PIM-SM Rendezvous Point for the group). The switched path establishment for the shared-tree works exactly as explained above, except that the Establish message is initiated when the PIM Join/Prune message from the receiver's DR (Designated Router) reaches the RP node. The Establish message for a source-specific tree is also originated at the ingress node. This again is initiated by the receipt of a PIM Join/Prune message. The Establish message for a source-specific tree uses the (S,G) egress identifier. In both cases, the Establish message is forwarded according to the states created by the PIM protocol.

All multicast switched path trees are periodically refreshed by retransmitting an Establish message. The periodic refreshes may also be used to keep the multicast forwarding states active since the intermediate ISRs may not forward packets at network layer. When a new receiver is explicitly grafted in the multicast distribution

tree, the ISR into which the new branch is spliced may issue an Establish message downstream or wait for the next refresh cycle to create the switched path branch along the newly grafted branch to the multicast distribution tree.

The loop prevention mechanism for multicast works in the exact same manner as for the unicast case expounded previously. Each ISR appends it's ISR ID to the path in the Establish message before forwarding it to the downstream ISRs. ISRs which receive an Establish message verify a loop-free message via the ISR ID path.

8. Host-to-Host Connectivity

Dedicated switched paths for host-to-host connectivity may be established with either RSVP [Rsvp] or ARIS. Since this may pose scalability problems in networks that support a large number of active hosts, it is desirable to provide complete host-to-host switched path connectivity using the pre-established aggregated ARIS connections in a network. This maintains good scaling properties in the backbone of the network by conserving labels for premium services, and at the same time provides end-to-end switching for hosts directly attached to the ARIS network. In this approach, a dedicated switched path is created between the host and the ingress (or egress) and this in turn is spliced to the aggregated switched path. The creation of the switched path can be either be initiated by the host or by an ISR by thresholding on the flow.

9. Label Conservation

An important goal of the ARIS protocol is to minimize the number of switched paths required by ISRs to switch all IP traffic in a network. Since switches may support only a limited label space, ARIS restrains its label consumption so that labels are available as needed for its own use, as well as for other services, such as RSVP. Further benefits include simplification of network management, both for automated tools and for human comprehension and analysis, and switched path setup overhead.

The consumption of labels is minimized:

- by the use of egress routers that may map thousands of IP destinations to the same switched, and
- by enabling the merging of switched paths.

This combination can provide O(n) switched paths, where n is the number of egress nodes.

9.1. Aggregation

The network routing domain has the greatest performance and label conservation when all routers in the domain are ISRs. Maximum ARIS benefits are also tied closely to an IP network routing topology with a high ratio of IP destinations to egresses, as exists in a typical IP backbone. However, ARIS is flexible enough to be highly beneficial even in networks with partial ISR deployments or arbitrary network routing topologies.

9.2. Switched Path Merging

The merging of switched paths enables ARIS to create switched path trees, each of which connects all of the ingresses to a given egress. This results in n trees, where n is the number of egresses in a network, while still providing the benefits of full mesh connectivity (without O(n**2) switched paths).

10. ARIS on Specific Switching Technologies

10.1. Asynchronous Transfer Mode (ATM)

The ability of the ARIS protocol to conserve the number of switched paths depends on the hardware capabilities of the ISR. Some ATM switching components can "merge" multiple inbound VCs onto one outbound VC at close to standard switching rates. These merge-capable components are able to buffer cells from the inbound VCs till all cells of a frame arrive, and inject the frames into the outbound VC, without interleaving cells from different frames.

The ARIS usage of "merged" VC flows requires that ATM switching hardware have the capability of preventing cell interleaving (see "VC Conservation"). Unfortunately, much of the existing ATM switching hardware cannot support VC merging. One solution to this problem is to use virtual paths (VPs) to egress points, rather than virtual circuits (VCs). The virtual path extension merges VPs, creating trees of VPs to the egress points, instead of merging VCs. Cell interleaving is prevented by the assignment of unique VC identifiers (VCIs) within each VP.

The ISRs within a network are assigned unique VCIs to prevent VCI

collisions when paths from different ISRs are merged. Each ISR requires a block of VCIs as labels to distinguish between cell paths to the same egress identifier. By assigning a unique block of VCIs to each ISR, ARIS guarantees that an ISR at a network merge point can safely merge upstream VP flows for an egress identifier to a single downstream VP without VCI collisions.

Although the virtual path extension uses VCs much less efficiently than a VC merging implementation, it reduces network latency and hardware requirements because frame reassembly and re-segmentation is not required on intermediate ISRs. In addition, although this variation uses more VC space, the work involved in establishing and maintaining switched paths is still O(n).

An alternative approach to the VC merging problem is to use an end-to-end VC label upstream allocation. This allows the ingress node to choose the downstream VC. In this approach, ISRs acknowledge the Establishment message with a label only after they receive an Acknowledgment message from their own upstream neighbor. Thus, the Establishment message traverses fully to the ingress node before being acknowledged. Ingress ISRs immediately acknowledge the Establishment message with the VC label. These acknowledgements may be merged as they travel downstream to the egress node. This method adds latency in the VC set up, and removes the benefits of ARIS VC aggregation (O(n**2) versus O(n) VCs). However, it adds the flexibility of performing VC-switching instead of VP-switching, which also makes switching possible at the routing boundaries.

10.2. Frame Switching Technology

While ARIS solves the problem of cell interleaving in the case of ATM by Virtual Path switching, it naturally and easily maps to a frame switching environment. This is due to the fact that multiple upstream flows can be merged into a single downstream flow without the problems of cell interleaving.

10.3. LAN Switching Technology

LAN switches are different than other frame switches, in that they typically do not perform label swapping, and instead switch frames based on their 6-byte IEEE MAC destination address. The label in this case can be considered as the 6-byte MAC address, which has global significance within the ARIS network. The advantages of this approach are that it augments LAN switches with routing functionality and helps achieve media speed switching between LAN segments [ARIS-LAN] without requiring hardware enhancements.

11. Layer-2 Tunneling

Like IP-in-IP tunnels, the L2-in-L2 tunnels can be useful in several different scenarios. In this, a L2 PDU is encapsulated into another L2 PDU. The outer shell carries the PDU to a predetermined termination point, at which the outer shell is removed and the PDU is switched based on the inner shell (now the outer shell after the de-encapsulation). Note that in a L2-tunnel, the label switching and swapping happens only on the outer shell. The L2 header of the inner shell is not examined until the tunnel termination point.

One simple application of this is private virtual networking, similar in manner to IP-in-IP tunneling. Another important usage is switching through routing hierarchies. At the egress ISR of a switched path that carries aggregated traffic, the packet must be L3 forwarded even if the packets are to continue on a different switched path. This is typical at the egress ISR. Traffic from all ingresses flow towards the egress ISR using the same switched path tree. To avoid L3 forwarding at the egress ISR, the egress can advertise the inner shell label to the ingress ISRs in the Establish message. The ingress ISRs may use this information to build its PDU accordingly. At the tunnel egress ISR, the outer shell is removed and the packet is switched based on the new outer shell. The egress may also introduce a new inner shell for its next egress ISR in the path. In this approach only one inner shell at a time is required. It is possible to envisage multiple levels of inner labels where its operation is similar in concept to loose source routing.

Some other useful applications are RSVP or DVMRP tunnels. With RSVP, multiple sender flows can be "merged" into a L2-tunnel and de-merged later at the end of the tunnel. At the de-merge point, L3 forwarding is avoided by switching PDUs based on the new outer shell. Similarly, in an ISR domain the DVMRP tunnels can be mapped to L2-tunnels. For example, the ATM Virtual Path Switching can be used as a tunneling mechanism for DVMRP tunnels, in that each (S, G) is identified through an unique Virtual Circuit Identifier.

In situations when the ARIS Establish message originates at the egress node, the label to be used at the end of the L2 tunnel may be carried in the Establish message. The ISRs at the start of the tunnel can use this information to build the inner shell. For example, when establishing a multipoint-to-point switched path for an egress BGP node, the establish message can carry the inner shell label for each CIDR prefix. Alternatively, an optimization would be to advertise these labels through an extension to the BGP protocol.

12. Quality of Service

ARIS can be extended to support Quality of Service (QoS) parameters. This will be addressed in a future ARIS revision.

13. ARIS Advantages

This section summarizes the advantages of the ARIS protocol. Several of the advantages listed below come from the egress orientation of the ARIS protocol.

13.1. Single Point of Control

The ARIS protocol is largely root oriented (originating Establish message at the root node of a switched path tree, although not limited to it. For creating multipoint-to-point or point-to-multipoint switched paths this gives the advantage of having a single node, the root node, as the point of control. This provides the convenience of only having to configure a single node to aggregate, deaggregate, switching establishment on/off, or apply QoS etc.

13.2. Aggregation and Merging

As mentioned in a previous section, the switched path conservation in ARIS is derived from the aggressive use of aggregation and switched path merging. With aggregation, several flows are bundled into the same switched path to reach the egress node. The switched path merging provides the multipoint-to-point tree, which is most suitable to carry best-effort traffic. These two features keep the order of switched paths for ALL traffic to O(n), where n is the number of edge nodes.

13.3. Multiple Levels of Aggregation

Multiple levels of aggregation can exists simultaneously in an ARIS network. For example, there can be an aggregated switched path for all networks (CIDRs) behind an egress BGP node, as well as individual nonaggregated switched paths for CIDRs behind the same egress node. This feature can be used to provide special services to a selective set of CIDRs.

13.4. Loop Detection/Prevention

ARIS supports an explicit mechanism to either detect or prevent looped switched paths. This feature can be useful in environments employing switching technology that do not have a TTL equivalent mechanism to contain resource wastage from switched path loops. This mechanism does not require any switch specific hardware implementation and can be effectively used to guarantee loop-free switched paths in networks employing existing, commonly available switches, such as ATM.

13.5. Traceroute Support

Traceroute is a tool commonly used by operators and users of a network to debug, trace and locate network problems. ARIS provides the optional support of making the ARIS switched network visible to the traceroute tool.

13.6. Multicast

ARIS support for multicast, both source-specific and shared tree, is similar in operation to the unicast support. No multicast routing protocol changes are required.

13.7. Multipath

Equal cost multipath is a commonly used paradigm in existing networks to load share traffic across multiple routed paths. ARIS has explicit support for multipath in which multiple switched paths (one corresponding to each routed path) is extended to the ingress node. The ingress node can distribute traffic into these multiple switched path as in conventional routers. Since the Establish message originating at the egress node traverses the multipath nodes on its way to the ingress ISRs, the support for multipath in ARIS is straighforward.

13.8. L2 Tunneling

There is direct support for L2 tunneling in the ARIS protocol. The inner shell labels can be advertised to the upstream ISRs via the ARIS Establish message. This provides a self-contained solution for leveraging L2 tunneling benefits.

13.9. Migration

Since an ISR behaves as a conventional router in addition to a switch, networks can migrate to ARIS on an incremental basis. The ISRs can be simply "dropped" into existing networks employing conventional routers. In addition, due to the superset nature of the ISR with respect to conventional routers, network management tools work as is, with no required learning curve.

14. Security Consideration

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

15. 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.

16. Acknowledgements

The authors wish to acknowledge the following people for their input and support: Brian Carpenter, Steve Blake, Ed Bowen, Jerry Marin, Wayne Pace, Dean Skidmore, Hal Sandick, and Vijay Srinivasan.

17. References

[ARIS-LAN]
S. Blake, A. Ghanwani, W. Pace, V. Srinivasan, "ARIS Support for LAN Media Switching", Internet Draft <draft-blake-aris-lan-00.txt>, March 1997

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

[IGMP-3]
B. Cain, S. Deering, A. Thyagarajan, "Internet Group Management Protocol Version 3", Internet Draft <draft-cain-igmp-00.txt>, University of Delaware, Xerox PARC, August 1995

[PIM-DM]

D. Estrin, D. Farinacci, V. Jacobson, C. Liu, L. Wei, P. Sharma, A. Helmy, "Protocol Independent Multicast-Dense Mode (PIM-DM): Protocol Specification", Internet Draft <draft-ietf-idmr-pim-dm-spec-01.txt>, USC, Cisco Systems, LBL, January 1996

[PIM-SM]
S. Deering, D. Estrin, D. Farinacci, V. Jacobson, C. Liu, L. Wei, P. Sharma, A. Helmy, "Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification", Internet Draft <draft-ietf-idmr-pim-spec-02.txt>, Xerox, Cisco Systems, USC, LBL, September 1995

[RFC1075]
D. Waitzman, C. Partridge, S. Deering, "Distance Vector Multicast Routing Protocol", RFC 1075, BBN, Stanford University, November 1988

[RFC1112]
S. Deering, "Host extensions for IP multicasting", RFC 1112, Stanford University, August 1989

[RFC1519]
V. Fuller, T. Li, J. Yu, K. Varadhan, "Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy", RFC 1519, BARRNET, Cisco Systems, MERIT, OARnet, September, 1993

[RFC1583]
J. Moy, "OSPF Version 2", RFC 1583, Proteon Inc, March 1994

[RFC1584]
J. Moy, "Multicast Extensions to OSPF", RFC 1584, Proteon Inc, March 1994

[RFC1771]
Y. Rekhter, T. Li, "A Border Gateway Protocol 4 (BGP-4)", RFC 1771, IBM Corp, Cisco Systems, March 1995

[RFC1812]
F. Baker (Editor), "Requirements for IP Version 4 Routers", RFC 1812, Cisco Systems, June 1995

Authors' Addresses

Rick Boivie
IBM Corp.
17 Skyline Drive
Hawthorne, NY 10532

Phone: +1 914-784-3251
Email: rboivie@vnet.ibm.com

Nancy Feldman
IBM Corp.
17 Skyline Drive
Hawthorne, NY 10532
Phone: +1 914-784-3254
Email: nkf@vnet.ibm.com

Arun Viswanathan
IBM Corp.
17 Skyline Drive
Hawthorne, NY 10532
Phone: +1 914-784-3273
Email: arunv@vnet.ibm.com

Richard Woundy
Continental Cablevision
The Pilot House - Lewis Wharf
Boston, MA 02110
Phone: +1 617-854-3351
Email: rwoundy@continental.com