Issues in Automatic Topology Discovery for RPR

July 2001, 802.17 Meeting
Portland OR

Brian Holden   PMC-Sierra, Inc.

Outline

- Requirement Questions
- Taxonomy of solutions
- An example algorithm
- Recommendations

Scalability Requirement

- **How Scalable must the solution be?**
  - Handful of nodes
    - » This allows manual solutions
  - 10 to 100 nodes
    - » This allows solutions where nodes have complete knowledge
  - 1000's of nodes
    - » This would rule out any full knowledge solution

Integration Requirement

- **Is there a requirement that the topology discovery mechanism be integrated with other ring mechanisms?**
  - **With protection mechanisms (i.e. Steering/Wrapping)?**
  - **With the congestion management/avoidance mechanism?**
  - **With other OAM&P mechanisms (i.e. MPLS OAM ITU proposals)?**
  - **With 802.1D topology discovery?**
- **It is usually bad to represent the same information twice**
  - **What if the representations disagree?**

Topology Knowledge Requirement

- **Is there a requirement that a station know the topology of the ring?**
  - Most of the Steering Proposals that have been presented require accurate, real-time knowledge of the ring topology
  - Some of the Congestion Management Proposals that have been presented require accurate, real-time knowledge of the ring topology to deliver fairness and optimal ring utilization
- **Is there a requirement that a station know the capabilities of stations and links?**
  - There have been presentations about heterogeneous link speeds as well as link upgrade scenarios that could benefit by a means of determining the capabilities of the elements of the ring
  - There may be benefit in distributing information on the support of optional behaviors

Link State Knowledge Requirement

- **Is there a requirement that a station know the state of all links in the ring?**
  - **Most of the Steering Proposals that have been presented require accurate, real-time knowledge of the ring topology including the state of each link, so that the frames can be sent in the optimal direction**
- **Is there a requirement to detect MAC-layer failures?**
  - **As distinct from PHY-layer failures**

Ownership Requirement

- Are stations on a given ring owned by different entities?
  - Models of ownership on one ring:
    - » Enterprise, SiteA, SiteB (typical)
    - » Network, CustomerA, CustomerB (carrier owns all)
    - » Network, CustomerA, CustomerB (customers own box at site)
    - » Network1, CustomerA, CustomerB, Network2 (regular + peering)
    - » Network1, Network2, Network3 (when used as peering point)
  - This may affect the desirability of a given solution

Security Requirements

- Is there a requirement that any of the knowledge of the ring topology and/or capability be held securely?
  - Must any knowledge of a portion of a given ring be withheld from another portion of a given ring (from say a competitor)?
  - Are authentication fields required in topology messaging?
    - Does this extend to congestion management messaging?
    - Perhaps having a VLAN for the ring stations would help here.
  - Are there interactions with other security mechanisms?
- Is there a requirement that the ring have mechanisms to defeat Denial-of-Service Attacks?
  - Who decides?
  - What actions are taken?
  - Should it be robust to simple spoofing or to coordinated action attacks?

Outline

- Requirement Questions
- Taxonomy of solutions
- An example algorithm
- Recommendations

Station Numbering

- How are the stations given a number?
  - IEEE-48 MAC Address
    - » In most 802 standards including token ring, no further number is used beyond the IEEE-48 MAC address which is globally unique
  - Auto-Numbered
    - » In some protocols such as in DHCP and USB, the stations are automatically given a number by the network upon addition
  - Manually numbered
    - » In SCSI, in IP at the sub-network level, and in SONET rings, the stations or groups of stations are manually given a number

Ring Mastership

- **Which station is the master of the ring?**
  - Pre-defined
    - » In USB, the station which is associated with the CPU is the master of the bus. This is also true of SCSI and Fiber Channel.
  - Auto-Selected
    - » In Token Ring, an "Active Monitor" station is automatically selected by the ring. Its job is to watch for lost tokens and to kick off the neighbor discovery algorithm
  - No Master
    - » In plain Ethernet as well as in many routing algorithms like BGP, there is no entity which acts as the master. Each acts independently

Topology Knowledge

- How much does each node know of the topology?
  - No Knowledge
    - » In basic Ethernet, a station is only told how to resolve addresses
  - Knowledge of Neighbors
    - » In Token Ring, a station only knows the IEEE-48 address of its upstream neighbor, although a mechanism does exist for a station to query other members of the ring
  - Partial Topology
    - » In some router networks, the members know what is advertised. This may be the local topology and a summarized version of the rest of the network. 802.1D has the Generic Attribute Registration Protocol (GARP) for this purpose.
  - Full Topology
    - » In USB, IEEE 1394, and Fiber Channel the master station knows the full topology of the bus

Capability Knowledge

- How much does each node know of the capabilities (such as the link rate) of the other members of the Ring?
  - No Knowledge
    - » In basic Ethernet, a station has no knowledge of the capabilities of any other station
      - (hubs/switches actually know the rate, half/full duplex, and if flow control is on)
  - Partial Capability Knowledge
    - » In some router networks, the members know the capabilities that are advertised. This may be the capabilities of its local region and a summarized version of the rest of the network. This summarization is needed to allow scalability. 802.1D has the Generic Attribute Registration Protocol (GARP) for this purpose.
  - Full Capability Knowledge
    - » In USB and IEEE 1394, the master station knows the relevant capabilities of all of the members.

Relation to 802.1D GARP

- **How similar is this to the Generic Attribute Registration Protocol in 802.1D?**
  - Different
    - » The problems being solved are somewhat different
  - Derived from it
    - » Find a way of adapting GARP
  - Very similar
    - » Find a way of adapting GARP with minimal changes

Link State Knowledge

- **How much does each station know of the state of each link in the ring?**
  - **No Link State Knowledge**
    - » **Some Wrapping solutions can insulate the stations from the need to know the state of each link**
  - **Bi-directional Link (Segment) State Knowledge**
    - » **Most Steering solutions require at least knowledge of the state of a given segment**
  - **Link State Knowledge**
    - » **Enhanced behaviors are possible if each node knows the state of each link in each segment**
  - **Complex Link State Knowledge**
    - » **Support for a 802.3ad aggregated link (or other composite links) which supported failure of some of the links would require more detailed knowledge**

Link State to Topology Mapping

- **When is a link state change turned into a topology change?**
  - Immediately
  - After a long timeout
  - When conflicting evidence arrives
    - » Update when the ring is rejoined with a different pairing of addresses

Drop/Insert Port State Knowledge

- **How much does each station know of the state of each drop and insert port connected to the ring?**
  - No Drop/Insert Port State Knowledge
  - Single Drop/Insert Port State Knowledge
    - » There may be benefits of knowing the state of the Drop/Insert Port
  - Complex Drop/Insert Port State Knowledge
    - » There may be benefits of knowing the state of N Drop/Insert Ports

MAC-Layer Failure Detection

- There may be failures that cannot be detected at the PHY layer. These include:
  - A non-responding MAC layer device
  - A non-responding MAC layer control processor
  - A mis-configured MAC layer
- What is the Mechanism for detecting MAC layer failures?
  - None
  - Use of topology discovery messaging
  - Other
- What is the speed requirement for this detection?
  - Is it slower than for PHY level failures?

Periodicity of the Algorithm

- Does the algorithm run continuously or only upon change?
  - Change-Based Algorithm
    - » The algorithm runs only on startup or when a station detects a change. These are typically good for achieving fast response time and low overhead.
  - Continuous Algorithm
    - » The algorithm runs continuously, continually testing and reporting the current state. These are typically good for revealing undiscovered faults and keeping a database accurate.
  - Continuous plus send on changes
    - » This may give the benefits of both

Neighbor Discovery Mechanism

- **Most algorithms need a way of unambiguously querying for the IEEE-48 address of their neighbors**
- **Possibilities are:**
  - **TTL=1 broadcast**
    - » **There are several possible uses for TTL based messaging. It is possible to determine the full topology based on messages with a known starting TTL**
  - **Well-known IEEE-48 address**
    - » **Possibly use spare 802.1D non-forwarding addresses**
  - **Bit in RPR header**
  - **Other?**

All-Ring-Stations Multicast Mechanism

- Many algorithms need a multicast mechanism which can send a message to all stations on the ring
- Possibilities are:
  - Well-known IEEE-48 address
    - » Token ring has an "All Stations MAC address" of 0xC000FFFFFFFF
  - Defined multicast group
  - Bit in RPR header
  - Other?

Control Message Mechanism

- Most algorithms need way of distinguishing RPR control plane messages from data plane messages
- Possibilities are:
  - Bit in RPR shim
  - Other?

Extensibility

- **How friendly is the topology algorithm to extensions?**
  - Unfriendly – (lack Protocol ID & other allowances)
  - Friendly to standards based extensions
  - Friendly to both standards based extensions and proprietary extensions - (proprietary Protocol ID's and other allowances made)

Outline

- Requirement Questions
- Taxonomy of solutions
- An example algorithm
- Recommendations

Example Algorithm

- Here is an example algorithm:
  - Each station periodically asks its neighbors for its IEEE-48 address, and state of its links
    - » Each station expects to be asked periodically
      - declares that an incoming link is bad if it is not asked
    - » Each station expects to hear a response
      - Declares that an incoming link is bad if it does not hear a response
      - Response includes backward direction link state
  - Each station periodically (and immediately on link failure) does a ring-only broadcast of:
    - » Its own & both of its neighbor's IEEE-48 addresses
    - » Its capabilities
    - » The state of itself and each of its links

Example Algorithm (cont.)

- The stations match up the IEEE-48 addresses in the ring-only broadcast to determine the topology of the ring
- The stations can then assemble a complete database of:
  - The topology of all stations and links on the ring
  - The IEEE-48 address of all stations
  - The capabilities of each station and link
  - The state of all links and stations
- This database is kept accurate in real-time by the continual flow of the ring-only broadcasts

Outline

- Requirement Questions
- Taxonomy of solutions
- An example algorithm
- Recommendations

Recommendations

- **Our recommended requirements of the solution:**
  - » Is scalable from 10-100 stations
  - » Is tightly integrated with the other ring mechanisms
  - » Provides a topology database which is accurate in real-time for the use of protection and congestion management algorithms and which knows the relevant capabilities of the ring elements
  - » Maintains accurate link state information
  - » Allows more than one network to own stations
  - » Takes a few sensible security/DOS precautions such as authentication

Recommendations (cont.)

- **Our Recommended Answers to the Taxonomy List Questions:**
  - » Uses only IEEE-48 addresses
  - » Has no master station
  - » Has full topology knowledge
  - » Has full capability knowledge
  - » Explore the GARP question more
  - » Has full link state knowledge
  - » Updates topology when it receives conflicting information
  - » Knows about the state of a single drop/insert port

Recommendations (cont.)

- **Our Recommended Answers to the Taxonomy List Questions (cont.):**
  - » Runs continuously with change-based link state update
  - » Uses the topology discovery messages to detect MAC layer failure
  - » Uses a well-known IEEE-48 address for neighbor query
  - » Uses a reserved ring-only multicast address
  - » Uses bit in RPR shim to distinguish control plane messages
  - » Is friendly to both standards-based and proprietary extensions