Subscription Protocol and Admission Control in Residential Ethernet

Felix Feifei Feng
Samsung Electronics

IEEE 802.3 RESG, Sept. 2005
San Jose, CA, USA

feng.fei@samsung.com

Background

- Deterministic low latency and low jitter for data delivery can be provided only if the availability of network resources is guaranteed and intermediate bridges are appropriately configured along the entire transmission path.
  - This requires a subscription protocol for explicit negotiation (admission control) of network resources and configuration of bridges.
  - Such a subscription protocol provides the function of establishing end-to-end streams in the layer 2 Residential Ethernet.
    - The subscription protocol could be further interfaced with the signaling protocol of upper layer applications.
- Subscription protocol and admission control
  - Subscription protocol is the signaling to carry the traffic specification and negotiation result in an “end-to-end” sense
  - Admission control is the local operation in each node that judges whether local performance bounds (typically, the delay performance) can be guaranteed in this single node

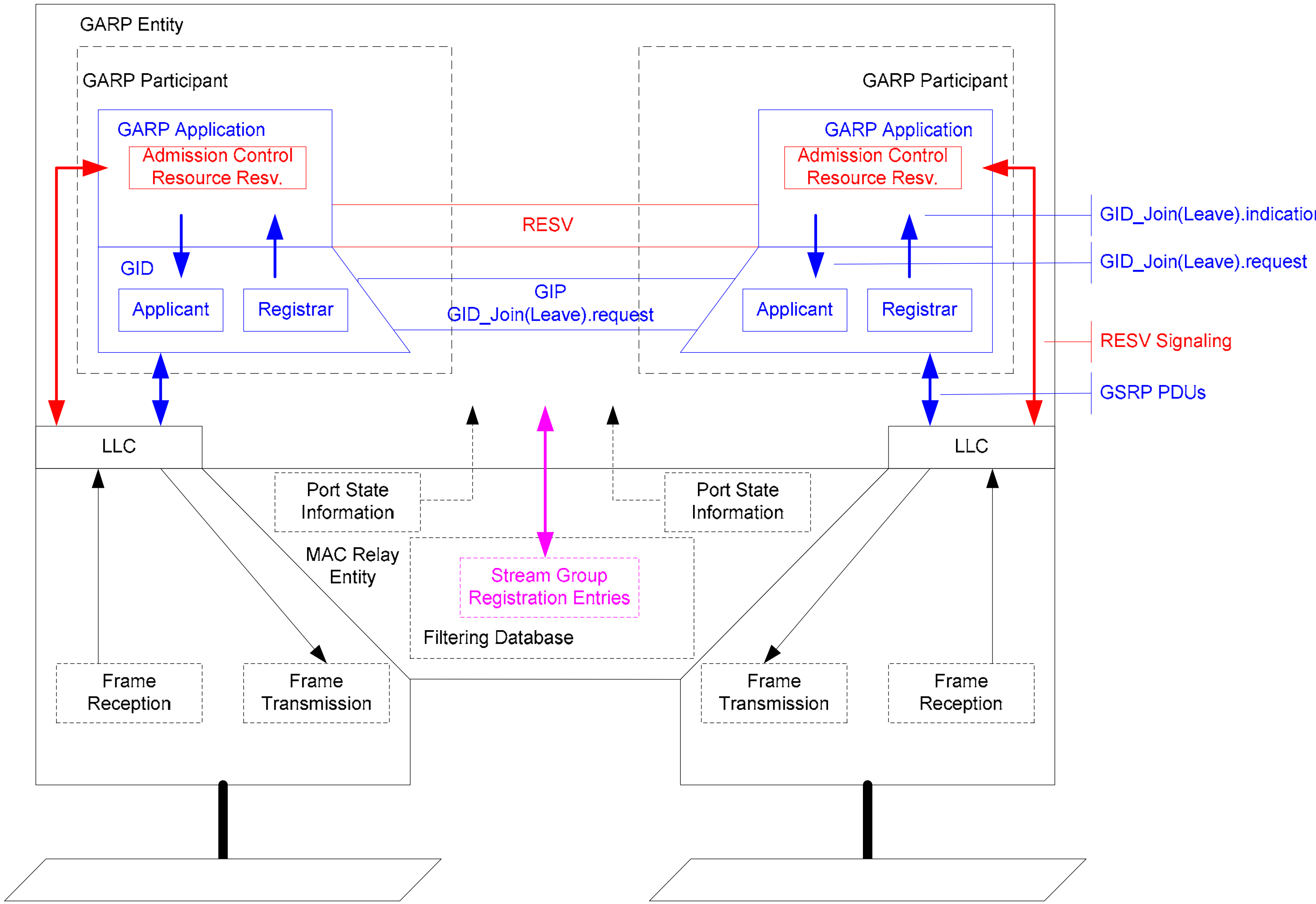
Network Architecture Overview

- Identifying and forwarding of isochronous stream data frames
  - Each isochronous stream will be assigned a locally unique identifier
    - The identifier may use MAC addresses, VLAN tags, or other formats.
      - Choosing among these formats is another discussing topic out of this document.
  - Bridge forwards isochronous stream data frames according to its isochronous filtering database (reservation state database)
    - Isochronous filtering database defines whether a certain port should forward or filter a certain isochronous stream data frames. The default value is "filtering".
      - Isochronous data frames will be forwarded to outbound ports that are explicitly enabled in the isochronous filtering database
      - Isochronous data frames will never be aimlessly flooded
- Subscription protocol (or called Simple Reservation Protocol – SRP in ResE) is employed to manage isochronous streams by updating isochronous filtering database of each bridge along the isochronous stream paths.

Framework of GARP-based SRP

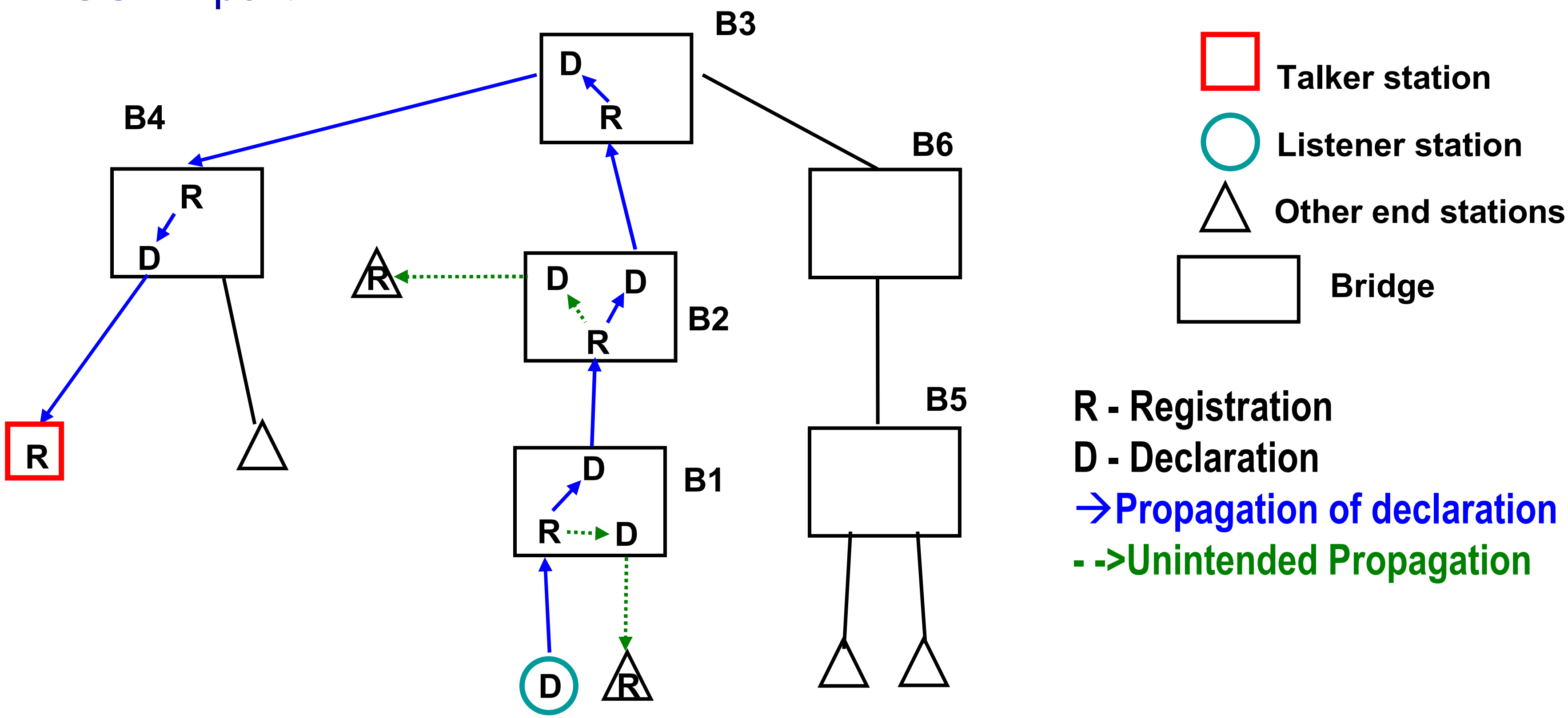
- SRP signaling can be divided in to two parts→ a GARP based signaling part (GSRP) and a RESV signaling part:
  - Listener uses GARP to show its intention of joining specified isochronous stream
    - With GARP registration, the talker and intermediate bridges know where are potential listeners and how to get to them
  - Talker sends out RESV (reservation) signaling towards its listeners
    - RESV signaling triggers admission control operations in intermediate bridges, and locks resources if the admission control is successful.
    - RESV signaling servers as the end-to-end explicit ACK/NACK to listeners

GARP-based SRP Architecture



Example: first stream

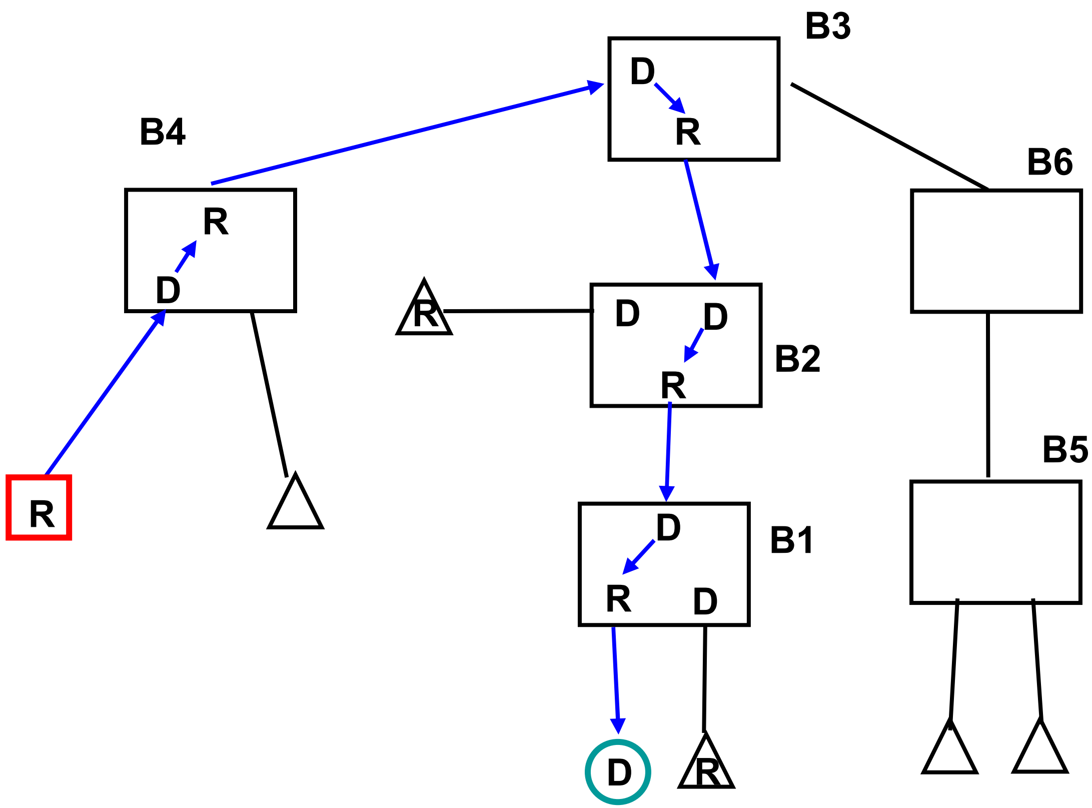
❑GSRP part



- With GSRP registration, the talker and intermediate bridges know where are potential listeners and how to get to them
- Assume in the above figure, B3/B4 have learnt the talker's address, and B1/B2 haven't, then:
  - GSRP floods the registration if the talker's address is not in the bridge FDB (eg. B1, B2)
  - GSRP relays the registration through specific outbound port if the talker's address is known by the bridge FDB (eg. B3, B4)

Example: first stream (cont.)

❑RESV part: example of successful reservation

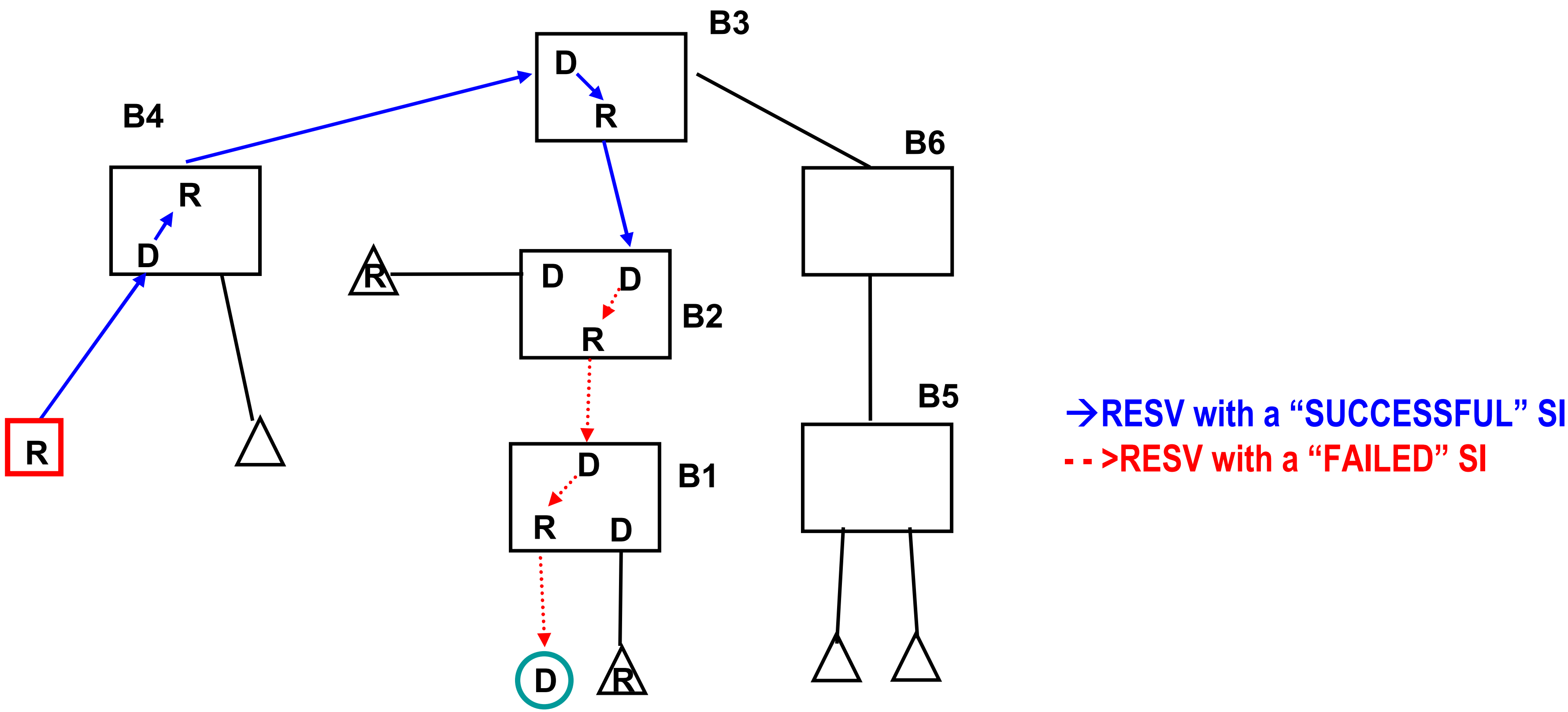


→Propagation of RESV

- RESV signaling triggers admission control operations in intermediate bridges. If the admission control is successful, it then locks resources and updates isochronous filtering database.
- In this example, admission control is successful along the whole path. RESV signaling servers as the end-to-end explicit ACK signaling to listener.

Example: first stream (cont.)

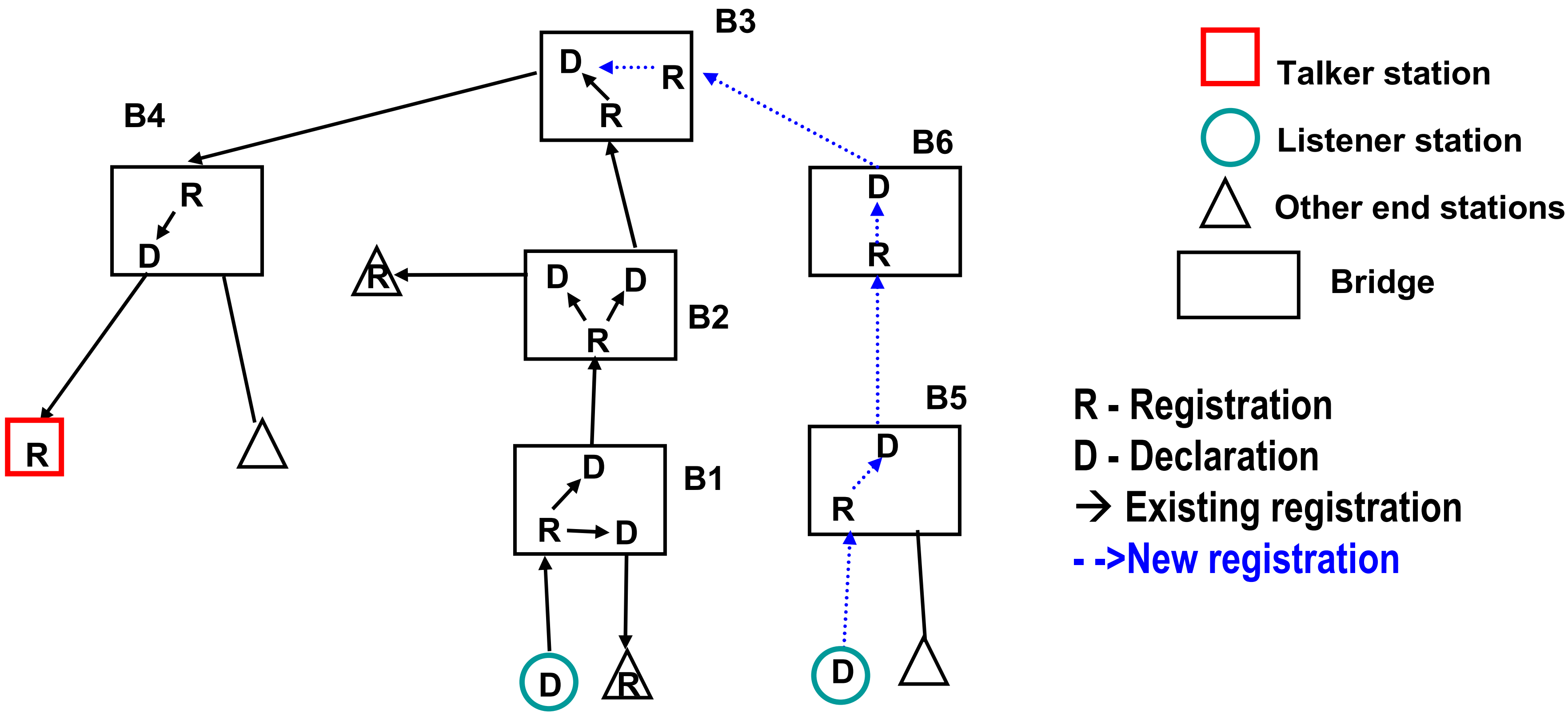
❑RESV part: example of failed reservation



- In this example, admission control is failed at B2. The SI (Status Indication) bits of RESV signaling will be set to FAILED.
- The RESV is still forwarded towards the listener. However, downstream bridges (i.e., B1, B2) will not lock resources for the RESV signaling whose SI is set to FAILED (Admission Control).
- Listener is noted of the failure since RESV with FAILED SI serves as an end-to-end explicit NACK

Example: second stream

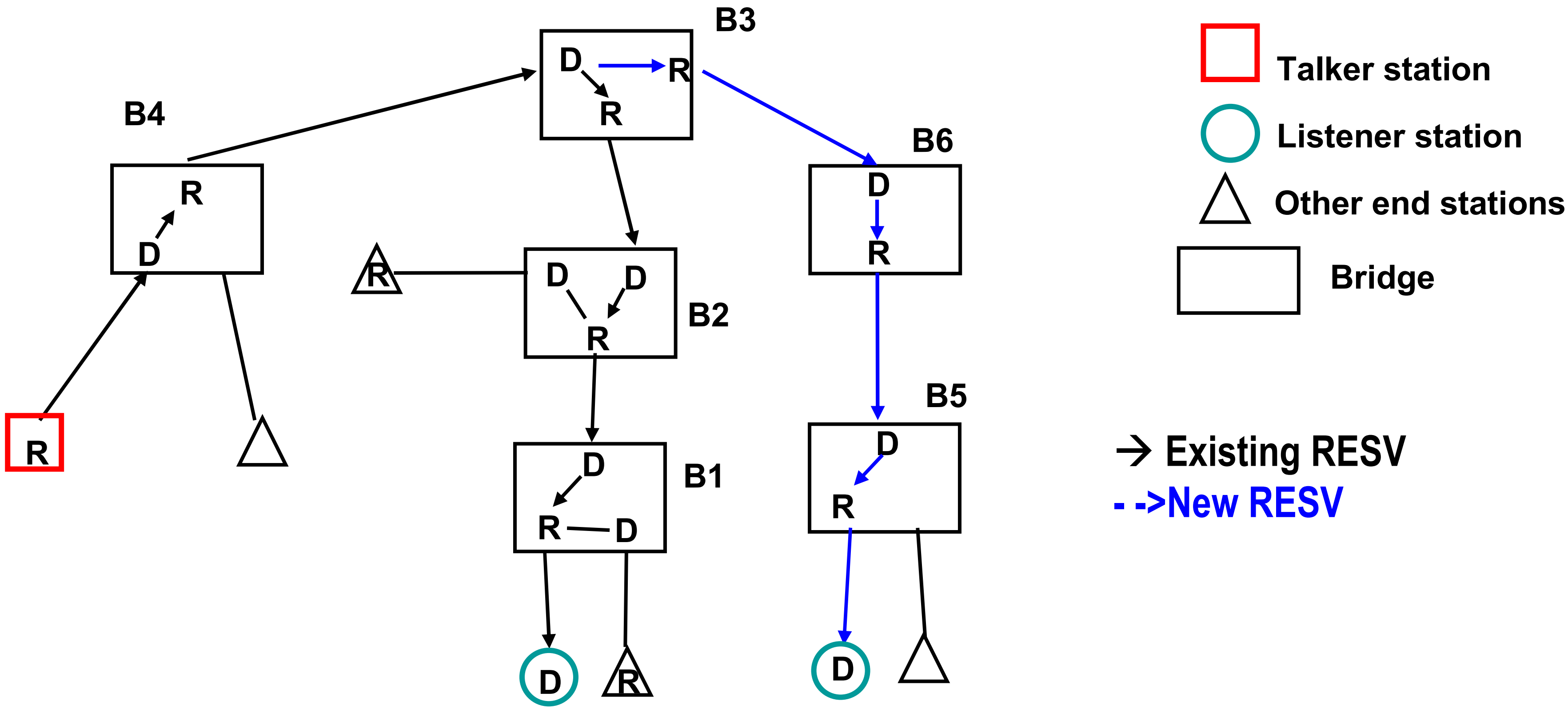
❑The second listener joins stream using GSRP



- As per GARP, registration from different listeners in a stream session is merged in the intermediate bridges according to the multicast tree topology.

Example: second stream (cont.)

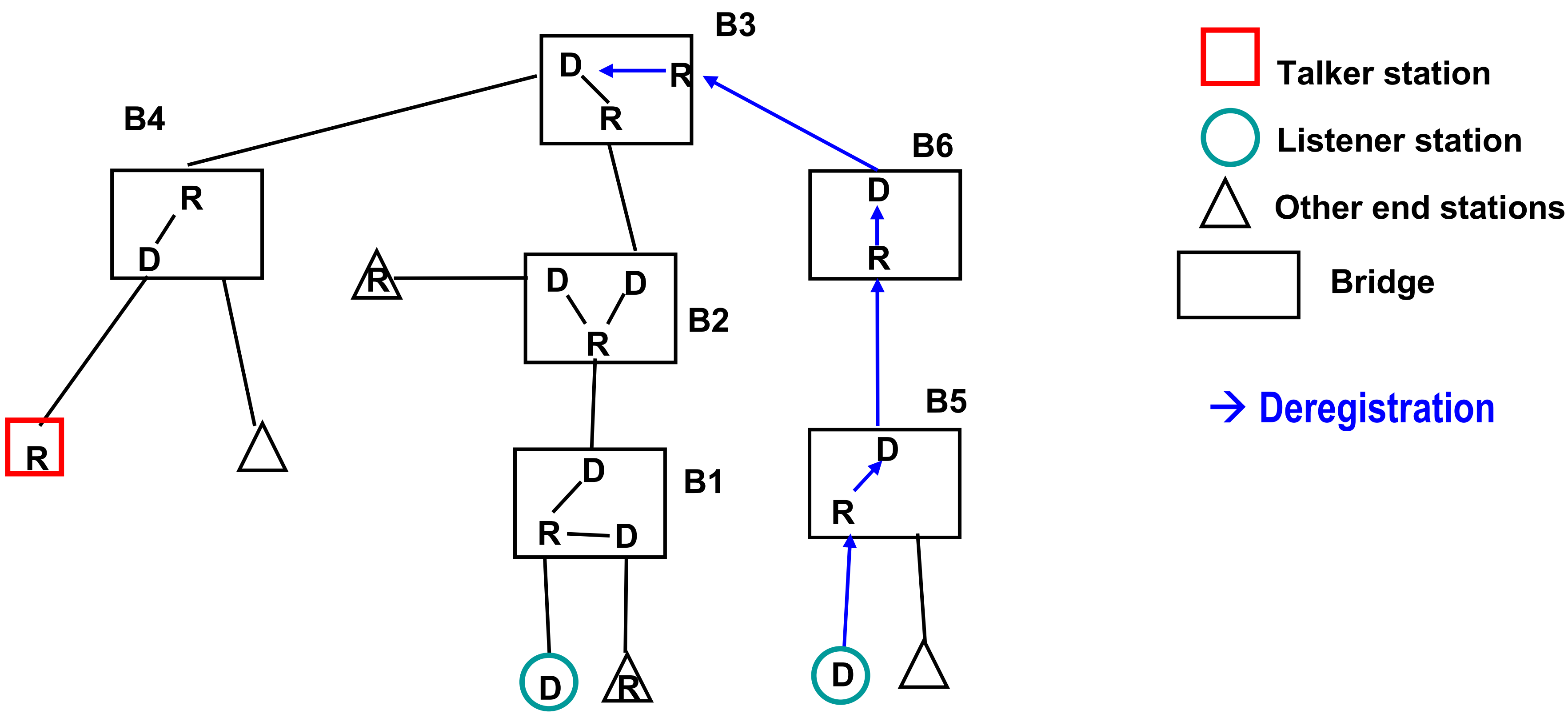
❑Intermediate bridge issues new RESV



- The reservation response delay is disassociated from the RESV refresh timer value
- SI (status indication) bits of this new RESV might be set to either SUCCESSFUL or FAILED according to the upstream reservation state

Example: teardown stream

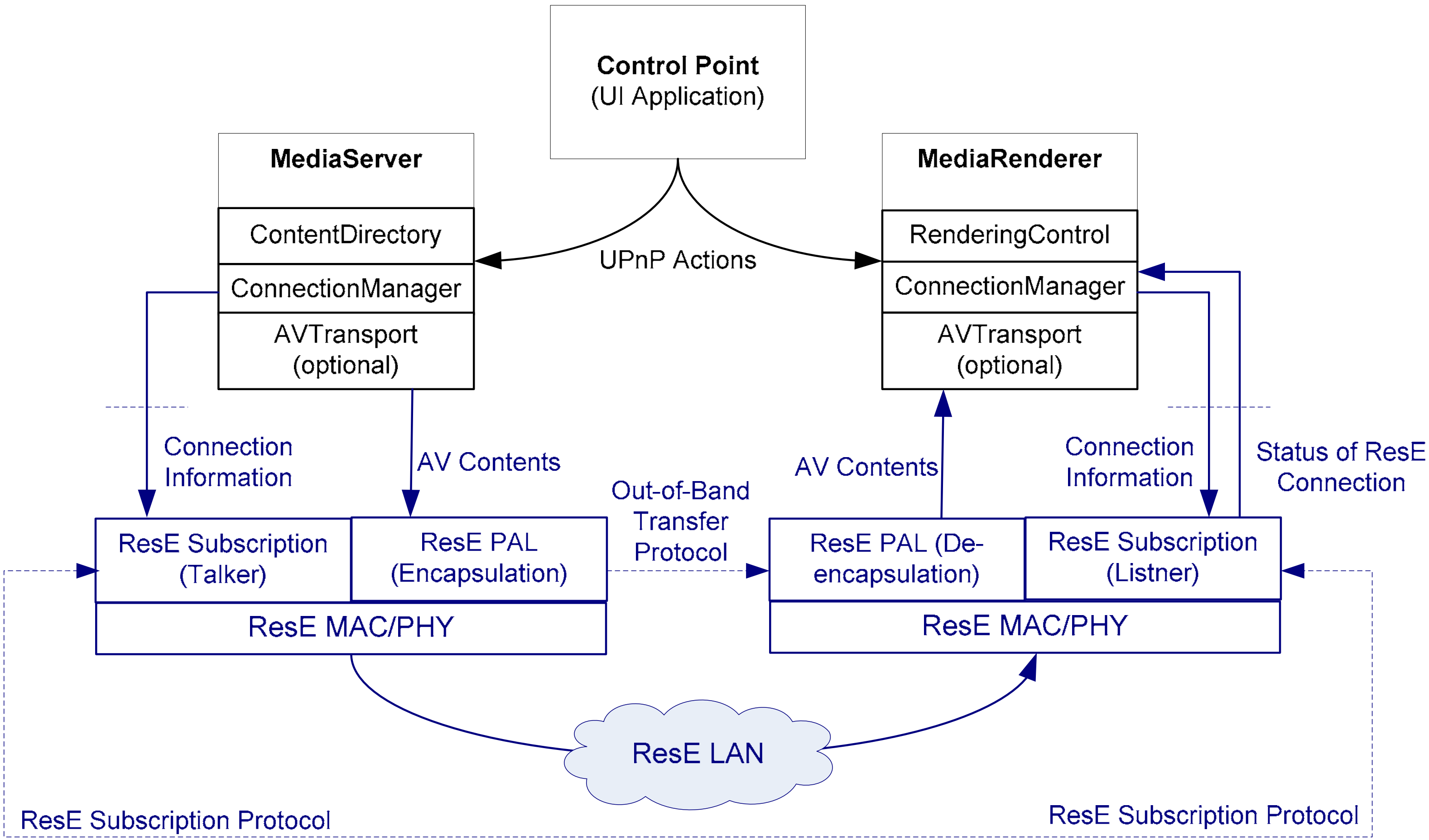
❑Listener uses GARP de-registration to leave isochronous stream



- Upon GSRP deregistration, intermediate bridges release resources that were previously locked for this listener. Isochronous filtering (reservation state) database will be updated correspondingly.
- As per GARP, other listeners will not be affected.

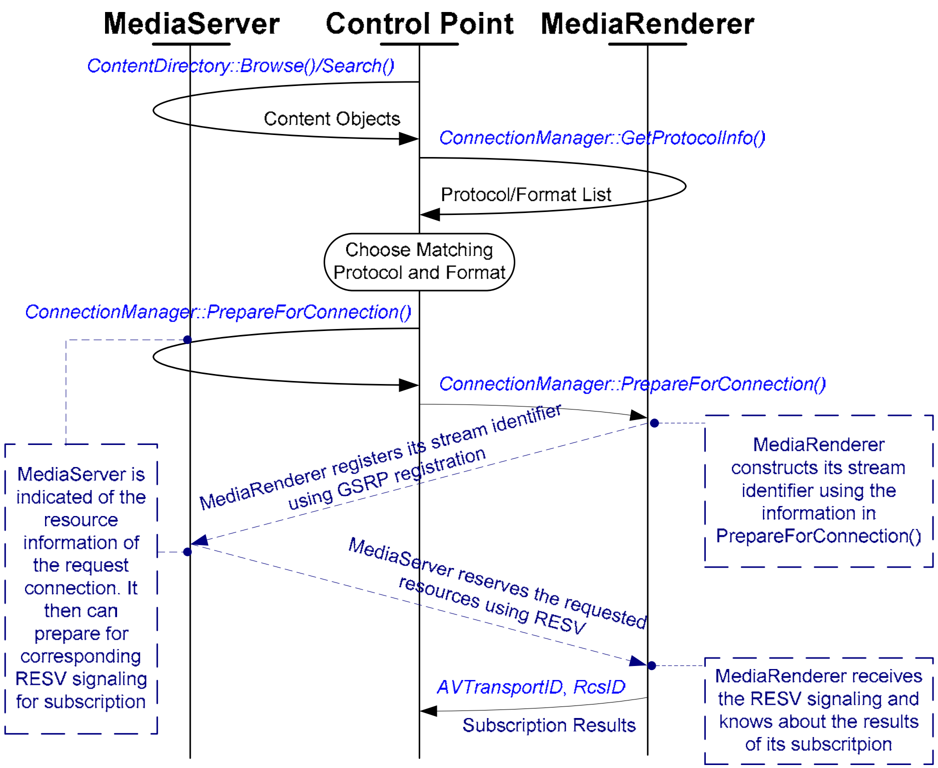
Integration With Upper Layer Applications [5]

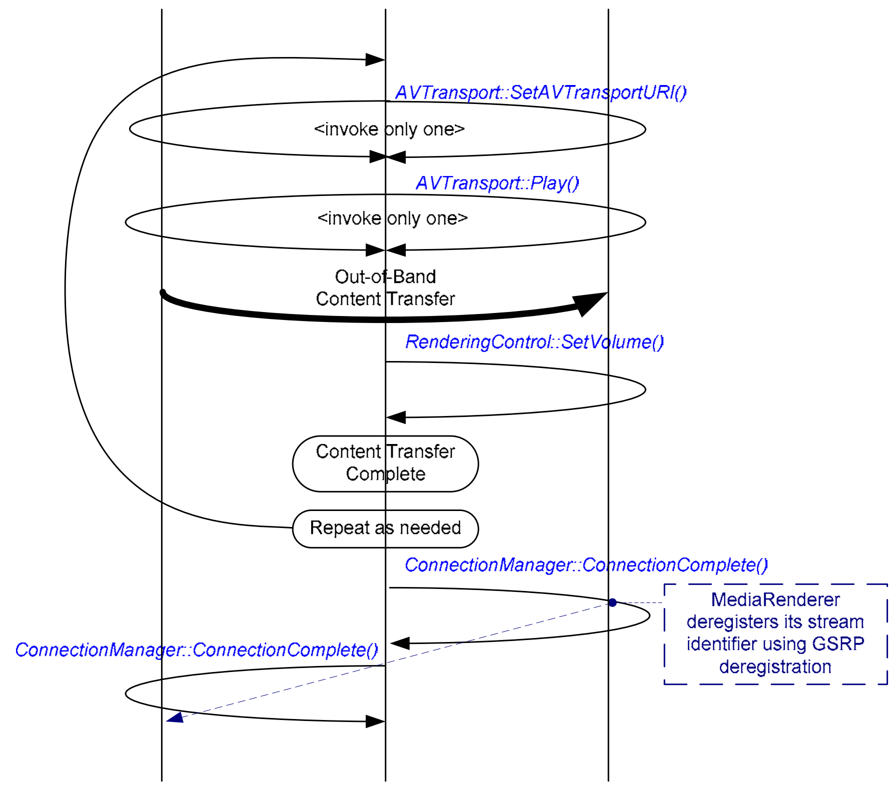
Example: Integration with UPnP-AV



Integration With Upper Layer Applications (cont.)

Example: Integration with UPnP-AV





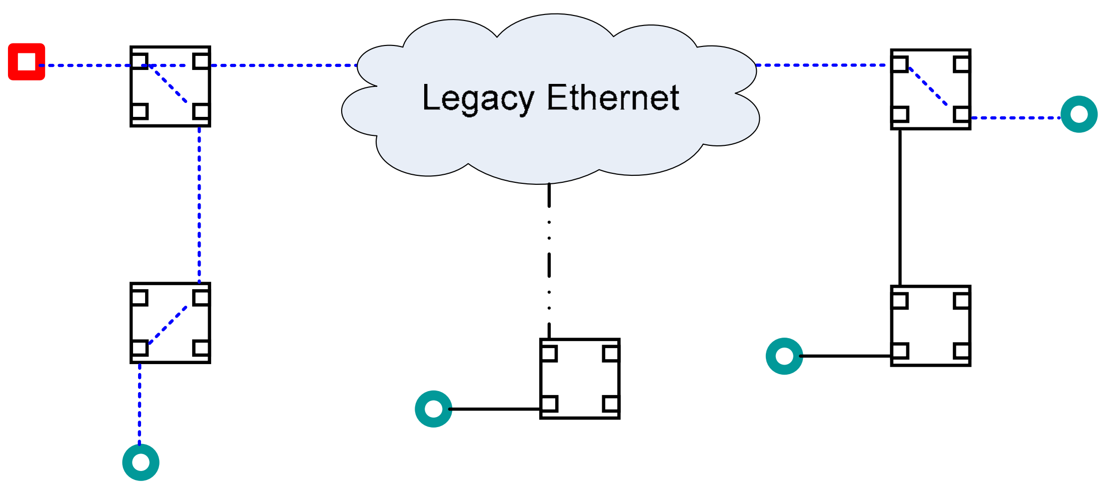
Evaluation of GARP-based SRP

❑Simplicity

- A large part of signaling reuses the established GARP specifications
- Only one new message (RESV) and corresponding processing need to be defined.

❑Functionality

- Explicitly ACK/NACK only using the newly defined RESV message
- Extendable to other scenarios, e.g, flexible reservation.
- Applicable to a hybrid network with ResE bridges and legacy bridges.



Evaluation of GARP-based SRP (cont.)

❑Scalability

- The GSRP signaling is scalable since registration/de-registration messages are merged in intermediate bridges
- The RESV signaling is scalable since it utilizes multicast mechanism.

❑Robustness

- Dead branches are pruned out by GSRP dynamic de-registration or aging of reservation states
- Occasionally packet loss is recovered by setting appropriate aging timer and refreshing timer.
- Signaling always automatically adapts to updated network topology

Admission Control Concept

- Subscription protocol is the signaling to carry the traffic specification and negotiation result in an “end-to-end” sense
- Admission control is the local operation in each node that judges whether local performance bounds (typically, the delay performance) can be guaranteed in this single node

Admission Control Condition [1]

- Assume non-preemptive strict priority schedulers are used
  - The scheduler has n priority levels. Class 1 has the highest priority; class n has the lowest priority
  - Cq is the set of connections at level q. The jth connection in Cq satisfies the traffic specification function (σ_q^j, ρ_q^j).
  - The maximum link speed is l
  - The maximum size of a packet that can be transmitted over the link is S_{max}
  - If $\sum_{q=1}^{n} \sum_{j \in C_q} \rho_j^q < l$,the maximum delay of any packet at priority level m is bounded above by d_m, where

$$d_m = \frac{\sum_{q=1}^{m} \sum_{j \in C_q} \sigma_j^q + S_{\max}}{l - \sum_{q=1}^{m-1} \sum_{j \in C_q} \rho_j^q}$$

Notes on the Admission Control Condition

- The admission control condition given by [1] can be adapted to other forms which are suitable for ResE environment. Different applications of this general condition can be found in Ref [2]
- The admission control condition guarantees a local bounded performance only if the input streams satisfy certain traffic specification functions. However, traffic patterns of these streams could be distorted as they traverse along the network even they satisfy the traffic constraints at the entrance of the network
    - Traffic re-shaping could be necessary in intermediate nodes
    - Ref [3] and [4] give two possible approaches for re-shaping

References

- [1] Hui Zhang, *Service disciplines for packet-switching integrated-services networks*, PhD thesis, Department of Computer Science, The University of California, Berkeley, CA, 1993. 26
- [2] Feifei Feng, *Efficient Transport of Isochronous Streams in Residential Ethernet: With A Generalized Admission Control Approach*, Samsung presentation at RESG conference call, available via http://www.ieee802.org/3/re_study/material/20050725_felix_Efficiency_Improvement.pdf
- [3] Michael J. Teener, *Residential Ethernet: a status report*, available via http://www.teener.com/ResidentialEthernet/Residential%20Ethernet.pdf
- [4] David V. James, *Residential Ethernet (RE) (a DVJ working paper),* available via http://www.ieee802.org/3/re_study/material/DVJ%20whitepapers/dvjRate2005Sep21.pdf
- [5] Feifei Feng, Hyunsurk Ryu, and Kees den Hollander, *End-to-end Stream Establishment in Consumer Home Networks*, submitted to CCNC2006

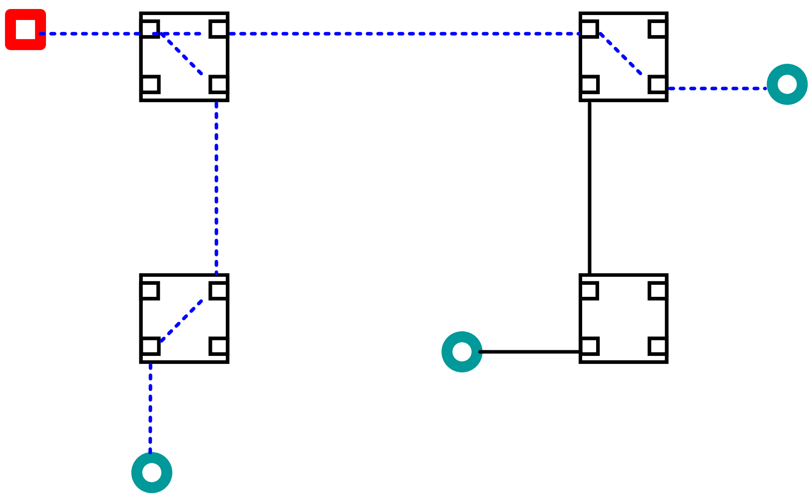
Backup Slides

Stream Group Registration Entries

- Besides the “Static Filtering Entries”, “Dynamic Filtering Entries” and “Group Registration Entries”, a new component, “Stream Group Registration Entries”, is added to the filtering database of each bridge
- Each stream group registration entry specifies:
  - An isochronous stream identifier
  - A RESV (Signaling) Port Map consisting of a control element for each outbound port that specifies forwarding or filtering of RESV signaling frames with the associate isochronous stream. The default value is “filtering”.
  - A Stream (Data) Port Map consisting of a control element for each outbound port that specifies forwarding or filtering of isochronous data frames with the associate isochronous stream. The default value is “filtering”. The filtering state has additional sub-states indicating the reason of filtering: “Initial”, “Admission control failure”, and “Timeout”. The default value is “initial”.
- If there is no refreshing during certain period, stream group registration entries will age out and reservation messages with failure flag will be sent downstream,
- The filtering processing (filtering entries maintenance, filtering decision, etc) for the asynchronous counterpart is uninfluenced by the isochronous group registration entries.
- Stream group registration entries are created, modified and deleted by SRP.

GSRP outline

- GSRP is defined as a GARP application, like GMRP and GVRP
  - The information registered, de-registered, and disseminated via this GARP application are the isochronous stream identifier information, and other optional service requirement information (eg. reservation policy)
  - Registration of stream identifier information forms the corresponding multicast tree and makes bridges/end-stations aware that reservation for this stream should only be forwarded in the direction of the registered members of the stream. Therefore, forwarding of RESV signaling frames for this stream occurs only on ports on which such identifier registration has been received. (Reservation and RESV signaling will be elaborated later)



GSRP forms a multicast tree between talker and listeners. RESV signaling follows this tree.

GSRP outline (cont.)

- Signaling reducing in GSRP (see slide 5 as an example)
  - In the case of ResE, the address of the stream talker has been available before subscription. The signaling overhead can be reduced by disseminating registration/deregistration information towards the talker purposely, other than the aimlessly flooding mechanism. This can be considered as a refinement of GARP Information Propagation Context (GIP context). More specifically, the specification in IEEE802.1D-2004:12.2.3 can be changed as follows:
    - Any GID_Join.indication received by GIP from a given port is propagated as a GID_Join.request to the instance(s) of GID associated with each port who is a “forwarding” port for the corresponding talker address according to the bridge’s FDB.
    - Any GID_Leave.indication received by GIP from a given port is propagated as a GID_Leave.request to the instance(s) of GID associated with each port who is a “forwarding” port for the corresponding talker address according to the bridge’s FDB.

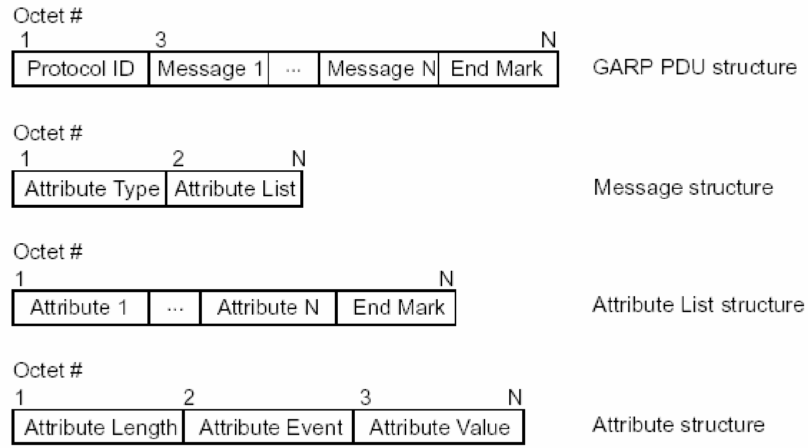
GSRP outline (cont.)

❑GSRP PDU addressing

- As per GARP specification, each GARP application uses a unique group MAC address as the destination address of GARP PDUs, for example:
  - GMRP Address: 01-80-C2-00-00-20
  - GVRP Address : 01-80-C2-00-00-21
  - GSRP will be assigned one such unique group MAC address
- IEEE802.1ak draft 1.0: Each MRP application uses a unique Ethertype value in order to identify the application protocol.
  - GSRP will be conformance to the future version of GARP

❑GSRP PDU structure

- GSRP PDU structure is conformance to GARP. Encoding of GSRP attributes will be defined further:
  - Shall include:
    - –stream identifier, talker MAC address
  - May include other affiliated elements:
    - –reservation style, bandwidth etc.



GSRP outline (cont.)

❑End system registration and de-registration

- The end system GSRP participant issues a GID_Join.request to join a stream
- The end system GSRP participant issues a GID_Leave.request to leave a stream

❑Registration and de-registration events

- On receipt of a GID_Join.indication
  - The GSRP application element specifies the associated port as forwarding in the RESV port map of the Stream Group Registration Entry for the RESV frames of the associated stream. If such a Stream Group Registration Entry does not exist in the filtering database, a new Stream Group Registration Entry is created.
  - If reservation state has been established (successfully or failed) in this bridge, the GSRP application element execute admission control (and resource locking/ database updating if needed) on the associated port for the associated stream, then issues corresponding RESV out of the port.
    - –This RESV can be set with either SUCCESSFUL or FAILED status indication bits according to upstream reservation state
      - »If the admission control is failed, RESV will be set to FAILED, and corresponding stream port map element will be set as “filtering, admission control failure”
    - –It disassociates the reservation response delay from RESV refresh timer value

GSRP outline (cont.)

- Registration and de-registration events (cont.)
  - On receipt of a GID_Leave.indication
    - The GSRP application elements specifies the associated port as filtering in the RESV port map and “filtering, initial” in the stream port map of the Stream Group Registration Entry for the associated stream.
    - If reservation has been successful set up in this bridge for the associated port, the GSRP application element releases the locked resources.
    - If setting that port to filtering results in there being no ports in the RESV port map as forwarding (i.e., all Stream Group members are deregistered), then that Stream Group Registration Entry is removed from the Filtering Database.

GSRP outline (cont.)

- Message relay
  - The “forwarding” of registration/de-registration messages refers only to the logical meaning, since actually messages from different downstream bridges will be merged. Exact operations are defined by GARP state machines.
- Use of GARP in point-to-point LANs
  - Since ResE is restricted to be point-to-point LAN, GSRP can use a simplified GARP state machine (refer to IEEE802.1D-2004 and IEEE802.1ak Draft2.0)

RESV: Actions of talkers

❑ Message format

- RESV signaling PDU carries: Stream identifier, Talker address, Requested_Bandwidth, Reserved_Bandwidth, Status_Indiciation, Error_Code etc.

❑ Source pruning

- Talkers are able to make use of the stream membership information registered via GSRP to allow them to keep track of the set of streams for which active listeners currently exist. This allow talkers to suppress the transmission of RESV frames if their registered stream membership information indicate that there are no valid recipients of those frames (i.e. Listeners) reachable via the LANs to which they are attached.

❑ Periodically refreshing

- On the basis of soft-state mechanism, talkers periodically refresh their RESV signaling.
  - Previous locked resources (if any) will be released and RESV signaling with FAILED (Talker timeout) SI bits being sent downstream.
  - Corresponding stream port map elements will be set as “filtering, timeout”.

RESV: Actions of intermediate bridges

- On receipt of a RESV signaling
  - Get from the Stream Group Registration Entries the list of outbound ports which are set by GSRP as forwarding for this RESV signaling
    - If the list is empty, the RESV signaling shall be discarded
  - If SI (status indication) bits of the RESV is SUCCESS, which means the reservation is successful in upstream, then the bridge checks on each outbound port whether the port has enough available resources for the this stream (admission control)
    - For ports on which admission control is successful
      - Lock corresponding resource, if it hasn't been locked
      - specifies the port as "forwarding" in the stream port map of the Stream Group Registration Entry for the associated stream, if it wasn't
      - "Forward" the RESV signaling out of this port to the next hop
    - For ports on which admission control is failed
      - Affirm corresponding resource is not locked.
      - Affirm the port is set to "filtering, admission control failure" in the stream port map of the Stream Group Registration Entry for the associated stream
      - Set SI bits of the RESV signaling as FAILED (Admission Control), and "forward" it out of this port to the next hop
  - If SI (status indication) bits of the RESV is FAILED, which means the reservation is failed in upstream
    - For each port in the outbound ports list
      - Affirm corresponding resource is not locked.
      - Affirm the port is set to "filtering, admission control failure" in the stream port map of the Stream Group Registration Entry for the associated stream
      - "Forward" the RESV signaling with a FAILED (Admission Control) SI bits out of this port to the next hop

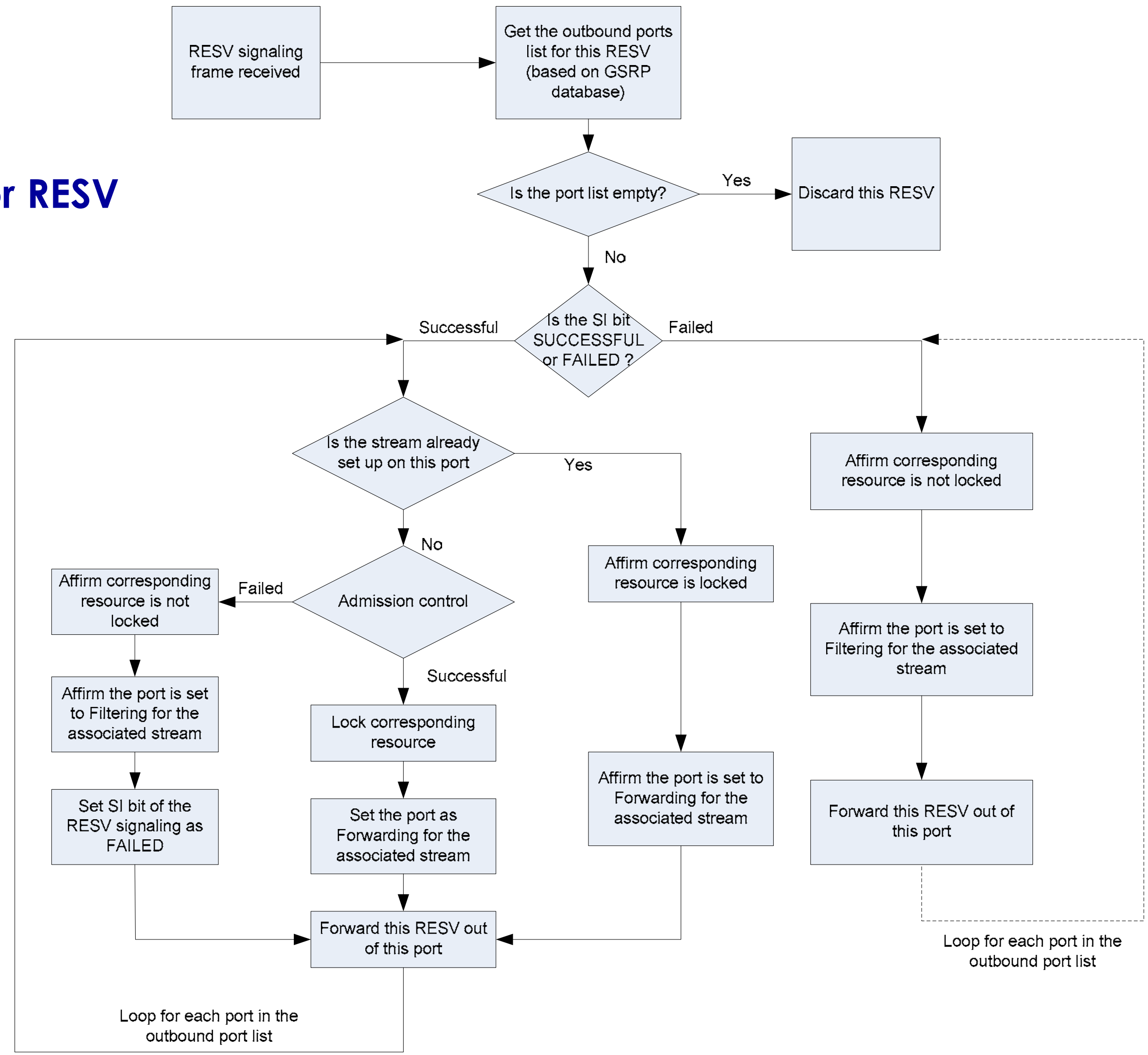
RESV: Actions of intermediate bridges (cont.)

❑Relaying RESV messages

- The “forward” in above description of RESV messages refers only to its logical meaning. Depend on the intermediate bridge’s reservation state database, The actual relay operations can be either immediately or delayed.
  - If the stream port map element state is newly changed, then a corresponding RESV message should be sent immediately out of that port.
  - Otherwise, RESV messages will be sent out of the ports when the bridge’s RESV refreshing timer is fired.

RESV: Flowchart

Example flowchart for RESV signaling processing



Other Considered Alternatives

- Direct mapping of RSVP
  - Talkers multicast PATH messages to pin the signaling and data paths
  - Listeners reply RESV messages backward to lock resources along pinned path.
    - Resource allocation happens in the direction of listeners to talkers
- Ping-Response-Reservation SRP
  - Receivers first send ping messages to desirable talkers
  - Talkers respond to listener ping message
    - Intermediate bridges learn talker addresses for forwarding of reservation signaling
  - Listeners send reservation messages to talkers upon receiving talker ping response
    - Resource allocation happens in the direction of listeners to talkers
- Disadvantages of above alternatives
  - More significant modification or addition to current bridge implementation
  - Explicit ACK function is difficult to be achieved