The Spatial Reuse Protocol (SRP)

Tutorial

IFIP Workshop on IP and ATM Traffic Management WATM 2001/
Seventh EUNICE Open European Summer School

Ecole Nationale Supérieure des Télécommunications/
Institut National des Télécommunications Paris, France

September 3, 2001

Dominic Schupke
Munich University of Technology
Institute of Communication Networks
80290 Munich, Germany
Tel.: +49 89 289-23511
Fax: +49 89 289-63511
Email: Schupke@lkn.ei.tum.de

Outline

Introduction
SRP Packets
Link Aspects
Protection
Topology Discovery
SRP Nodes
Packet Delay Performance
SRP Network Availability
Outlook IEEE 802.17
SRP Characteristics

Ring-based solution to interconnect routers:

- Dual counter-rotating rings
- Point-to-point links possible (two-node ring)
- Traffic is bidirectional and can be prioritized
- Multicasting possible
- Control information carried in opposite direction from data
- Protection switching
- Spatial Reuse Protocol (SRP) as media independent MAC layer with objective packet over optical networks

SRP: RFC 2892 (Informational) by Cisco with pending patents
Bandwidth Efficiency

Efficient use of bandwidth by spatial reuse
- Sources transmit concurrently
- Destination stripping of unicast traffic
- Bandwidth consumed only on traversed segment

Local reuse of bandwidth
- Fair share of bandwidth for all connections
- Additional access for less than fully utilized segments of the ring (left path)

Statistical multiplexing gain
- Ring instead of star topology
- Ring protection on packet level (unlike SDH)

Minimal protocol overhead
- May hold in comparison to ATM

SRP Packets
Packet Format

SRP payload encapsulation:

<table>
<thead>
<tr>
<th>SRP Header</th>
<th>Destination Address</th>
<th>Source Address</th>
<th>Protocol Type</th>
<th>Payload</th>
<th>FCS</th>
</tr>
</thead>
</table>

- Fixed-sized header for SRP control
- Destination/Source Address: 48 bit address assigned by the IEEE
- Protocol Type: defined values for IPv4 packets, Address Resolution Protocol (ARP) packets and SRP control packets; new version supports ATM cells as well
- Payload supported with upper bound on maximum transfer unit (MTU) of 9194 octets
- Frame Check Sequence (FCS): 32-bit Cyclic Redundancy Check (CRC) over destination address, source address, protocol type and payload
- Erroneous packets are discarded

Header Format

Generic SRP header without addresses and protocol type field:

<table>
<thead>
<tr>
<th>TTL</th>
<th>R</th>
<th>D</th>
<th>PRI</th>
<th>MOD</th>
<th>Usage</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Bits: 11 1 1 3 3 12 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Time To Live (TTL): hop-count, if expired \(\Rightarrow\) packet will be stripped
- Ring Identifier (R): differentiation between inner and outer ring
- Destination Strip (D): indicator for destination to strip the packet
- Priority Field (PRI): eight priority levels
- Mode (MOD): differentiation between data packets, idle packets and several control messages
- Usable Bandwidth Field (Usage): indicator for the currently usable bandwidth this node sees
- Parity Bit (P): parity bit over above fields
Control Packets

- SRP control packet format similar to payload packets but with additional header before payload
- Priority should be highest
- Control packets are transmitted only to adjacent nodes
- Control message sinks:
  - Host (similar to payload interface)
  - Local buffer at node (usually protection messages; implementor configures handling)
- Currently defined messages:
  - Protection switching messages
  - Topology discovery (TD) packet
- Control packets have own TTL (control layer hop-count)
- Control payload with variable length

Link Aspects
Keepalive Indications and Clock Synchronization

Keepalive packets:

- Control packets generated periodically when no data packets going upstream
- Propagation of messages contained in header (usage, etc.) when link is idle
- Together with data packets indication that a valid data link exists

SRP Clock Synchronization:

- Receive clock derived from incoming receive stream
- Transmit clock derived from a local oscillator
- Differences in clock frequency accommodated by inserting small amount of idle bandwidth at each node’s output

Media for SRP

SRP packets can be sent over any point-to-point layer 1 link:

- SONET/SDH using octet stuffing
  - Incase of full SONET/SDH protection: hold off timer for SRP protection switching to avoid multilayer protection interactions
  - Synchronization: external or transmit clock locked to receive clock
- SONET/SDH *framing* using octet stuffing for employment of SRP directly over an optical channel
- ATM (delineation by AAL5)
- Point-to-point Ethernet (delineation by Ethernet header and length field?)

SRP over SONET/SDH is the first media type perceived for SRP applications.

SRP together with information of layer 1 ⇒ support of SONET MIB (RFC 1595)
Protection

Intelligent Protection Switching (IPS)

- Bidirectional protection switching
- First step: loopback of traffic by the nodes adjacent to the failure (ring wrap)
- Second step: a new ring topology is discovered and reconfiguration may take place
- Ring wrap controlled similar to SDH/SONET bidirectional line-switched rings
- Objective: wrapping as fast as SONET/SDH or faster
IPS: Protection Switching Requests

- Request types for protection switching in the order of priority:
  1. Forced Switch (FS): by operator
  2. Signal Fail (SF): detection through media signal or SRP keepalive failure
  3. Signal Degrade (SD): detection through e.g., excessive bit error rate
  4. Manual Switch (MS): by operator
  5. Wait to Restore (WTR): automatic switch back after leaving SF or SD state plus a WTR period

- Higher priority requests rule out lower priority requests potentially present in the ring (exception: multiple SF and FS switches)

- Protection request of a node transmitted over both the short and the long path to peer node
Topology Discovery (TD)

- Each node performs TD by sending out TD packets on one or both rings
- New TD packet contains egressing ring ID and MAC binding
- Each node receiving a TD packet appends its MAC address binding and sends it to the next hop on the ring
- Wrap state on the ring:
  - Wrapped node indicates a wrap when appending its MAC
  - MAC address bindings are not added on the wrapped section
- The node that generated the TD packet takes the packet from the ring if first MAC binding matches and ingress and egress ring ID are equal
- Topology map is changed only after receiving two TD packets which indicate the same new topology
Receiver Architecture

Receiver packet handling:

- Destination address (DA) match \(\Rightarrow\) packets are copied to the control and data receive-buffers
- DA match and also unicast packet and destination strip bit set \(\Rightarrow\) packet will be stripped
- No DA match or multicast packet without source address (SA) match \(\Rightarrow\) packet placed into the transit buffers depending on priority
- Pass through to transit obviating layer 3 possible (no DA match)

(Picture source: Cisco)
Fairness Algorithm (SRP-fa)

- Congestion detection if limit of depth of the low priority transit buffer is exceeded
- Congestion at a node $\Rightarrow$ advertisement of node’s transmit usage counter to upstream nodes via the opposite ring
- Upstream nodes adjust their transmit rates so as not to exceed the advertised values.
- Nodes also propagate the advertised value received to their immediate upstream neighbor.
- Nodes receiving advertised values who are also congested propagate the minimum of their transmit usage and the advertised usage.

Due to limitations in the SRP-fa:
Practical limit to the number of nodes of a ring is currently 32.
Transmission Flowchart

Transmission decisions at each node:

**Congestion/unfairness branch**
- TB, Hi has packet:
  - yes: Send packet from TB, Hi
- TB, Low full:
  - yes: TX, Hi has packet
    - yes: Send packet from TX, Hi
    - no: Send packet from TX, Low
- TX, Low has packet:
  - yes: Send packet from TX, Low
  - no: Send packet from TB, Low

**Normal operation branch**
- TB, Hi has packet:
  - yes: Send packet from TB, Hi
- TB, Low full:
  - no: TX, Hi has packet
    - yes: Send packet from TX, Hi
    - no: Send packet from TX, Low
- TX, Low has packet:
  - yes: Send packet from TX, Low
  - no: Send packet from TB, Low

SRP Packet Delay Performance
Modeling

- Approximation for the transfer delay on a SRP ring
- SRP ring: two counter-rotating buffer (or register) insertion rings with two client traffic priorities
- Consider only one buffer insertion ring with $N$ nodes

Clients’ send and receive: in both ring directions by independent access units

High priority (Hi) as well as low priority (Low): appropriate transmit buffer (TB)

Transmit switch (TS): scheduling according to flowchart
Assumptions on SRP

Multicast and control packets negligible

Node states (according to flowchart):

<table>
<thead>
<tr>
<th>Priority</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>States</td>
<td>TB,Hi</td>
<td>Tx,Hi</td>
<td>Tx,Low</td>
<td>TB,Low</td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

⇒ Normal operation: state A

No further effects through external flow control

Assumptions on SRP

Multicast and control packets negligible

Node states (according to flowchart):

<table>
<thead>
<tr>
<th>Priority</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>States</td>
<td>TB,Hi</td>
<td>Tx,Hi</td>
<td>Tx,Low</td>
<td>TB,Low</td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

⇒ Normal operation: state A

No further effects through external flow control

General Assumptions

For all traffic:

- The latency for each node is constant
- The capacity of all buffers is infinite
- The interarrival times and service times are independent (thus we can use independent submodels for each node)

Independently for high and low priority traffic:

- The arrival processes to each Tx,Hi/Low are Poisson, with same mean arrival rate $\lambda_{Tx,Hi/Low}$ for all nodes
- All nodes have the same distribution of packet lengths, with first and second moments $\bar{X}_{Hi/Low}$ and $\bar{X}^2_{Hi/Low}$, respectively
- All nodes have the same transmission pattern:
  - node $i$ transmits to node $(i + j) \mod N$ with probability $q_{j,Hi/Low}, j = 1, 2, \ldots N - 1$ (balanced ring)
Approach

Use the virtual cut-through approach [Hammond and O’Reilly, 1988]

Input processes of buffers TB,Hi and TB,Low are Poisson
⇒ at TS: head-of-the-line nonpreemptive priority M/G/1 queuing system

Delay Contributions:

\[ T_{Hi} = W_{Tx,Hi} + \frac{X_{Hi}}{B} + \alpha_{Hi} W_{TB,Hi} + T_{phys} \]

\[ T_{Low} = W_{Tx,Low} + \frac{X_{Low}}{B} + \alpha_{Low} W_{TB,Low} + T_{phys} \]

\( W_{Tx/TB,Hi/Low} \): Waiting time in buffer
\( B \): Ring bitrate
\( \alpha_{Hi/Low} = f(q_{j,Hi/Low}) \): Average number of TBs a packet experiences
\( T_{phys} \): Average latency time between two nodes

Normalized Delay over Number of Nodes

- Min. hop routing
- Max. hops: \( \frac{N-1}{2} \) (1/2 of ring)
- Equal traffic to each destination

\[ q_1 = \frac{N-1}{2} \]

\[ q_2 = \frac{N-1}{2} \]

⇒ \( \alpha_{Hi/Low} = \frac{N-3}{4} \) (N odd)

Packet length X fix, \( \rho_{Hi, max} = 0.1 \rho_{bd} \) and \( \rho_{Low, max} = 0.85 \rho_{bd} \) (\( \rho_{bd} = \frac{N}{1+\alpha} \))
Normalized Delay over Utilization

Hi priority traffic:

Low priority traffic:

\[ \rho_{Low} = q \times \rho_{bd} \]

\[ \rho_{Hi} = q \times \rho_{bd} \]

\( N = 31 \)
Why Availability Considerations?

Each network element is subject to failures or disruptions.

⇒ Availability is an important parameter in mission critical and bandwidth intensive networks:

- Availability for an end-to-end connection is a Quality of Service (QoS) parameter for Service Level Agreements (SLAs).
- The overall network availability facilitates comparison:
  - Different architectural design alternatives
  - Different dimensioning alternatives
- Availability can serve as a parameter for QoS-routing in traffic engineered networks.

Network Model for Availability

- Symmetric concerning availability
- $n$ nodes
- Availability:
  - $R_n$ nodes
  - $R_h$ hosts
  - $R_f$ SRP forwarding device
  - $R_l$ Links (unidirectional)
  - $R_s$ Segments
- The hosts can be bypassed in the “pass-thru mode”.

Dominic Schupke

Technische Universität München
Lehrstuhl für Kommunikationsnetze
Univ.-Prof. Dr.-Ing. Jörg Eberspächer
Availability Models

**All-terminal availability** ($R_{all}$): The availability that all hosts on the ring are operating and can communicate with each other (network operator’s viewpoint).

**Two-terminal availability** ($R_{s-t,ring}$): The availability that two given hosts on the ring can communicate with each other, independent of the states of the other parts of the network (user’s viewpoint).

Assumption: At most one failure at any time.

Models are applicable both on repairable and non-repairable systems. Thus $R$ can represent both “availability” and “reliability.”

---

**All-Terminal Availability of a SRP-Ring**

All parts of the network are operational.

$$r_1 = (R_n R_h R_f R_s R_i^2)^n$$

At least one (unidirectional) link in a segment or the segment itself failed, while all other elements of the ring are operating.

Thus the ring wrap function is performed.

$$r_2 = n(R_n R_f R_h)^n (R_s R_i^2)^{n-1} [R_s (1 - R_i^2) + (1 - R_s)]$$

All-terminal availability:

$$R_{all} = r_1 + r_2$$
Two-Terminal Availability of a SRP-Ring

There are three mutually exclusive events for the two-terminal availability.

The first two are contained in the all-terminal availability excluding all non-terminal hosts.

At one node the ring is disconnected (forwarding part failure or total node failure). All other elements are operating except for the failed node’s adjacent segments and links, and the non-terminal hosts.

Thus the ring wrap is performed.

\[ r_3 = (n - 2)(R_n R_f)^{n-1} R_h^2 (R_s R_f)^{n-2} [R_n (1 - R_f) + (1 - R_n)] \]

Two-terminal availability:

\[ R_{s-t,\text{ring}} = \frac{r_1 + r_2}{R_h^{n-2}} + r_3 \]

Ring Interconnection

Interconnecting node: Failure events are not mutually exclusive.

The individual availability models provide approximations.

Single ring interconnection:

Two-terminal availability for a path routed over \( m \) rings:

\[ R_{s-t} = R_h R_n \prod_{i=1}^{m} \frac{R_{s-t,\text{ring}_i}}{R_h R_n} \]
Two-Terminal Availability Example

Assume: $R_s = 0.9995$ (equivalent to 4.5 hours outage per year),
$R_n = R_h = R_f = R_l = 1$ (no outage)

- 5 min. outage per year
- 6 link-segments

- Intraring: 24 sec. outage per year
- Interring: 47 sec. outage per year
- 9 link-segments

Outlook IEEE 802.17
IEEE 802.17

Specification by IEEE 802.17 Study Group “Resilient Packet Ring (RPR)”

RPR scope:

- Ring operations (forwarding, topology, fairness, protection)

Out of RPR scope:

**Vendor specific:** Service intelligence (adaptation, QoS, protocols)

**PHY specific:** Optical transmission choice (Ethernet, SONET, ...)

⇒ SRP one RPR solution

IETF: Working group for IP over RPR requirements and framework

---

Wrapping vs. Steering

**Wrapping (SRP):**

- a)
- b)
- c)

**Steering:**

- a)  
  - b)
Store and Forward vs. Cut-through

Store and forward (SRP):

Cut-through:

Further Information