Persistent Reservation
Issue #346

Roger Cummings
Symantec
Basic Situation

Two Initiators

Init A

Init B

Ancient FC-SCSI Bridge

New Tape Drive supports both Reserve & PR
Reserve works well

1. A issues reserve
2. Bridge processes reserve, limits access to LU to A
3. Bridge issues own reserve to LU
4. B access to LU blocked by bridge

Two Initiators

Ancient FC-SCSI Bridge

New Tape Drive supports both Reserve & PR
PR fails horribly

1. A issues PR Register & Reserve which flows thru to device.

2. LU responds & A thinks LU is reserved

3. B can still access device!!!
LU cannot distinguish it from A

New Tape Drive supports both Reserve & PR

Two Initiators

Init A

Init B
And what’s even worse

- We believe there’s no way to detect that the bridge is there, much less what it’s characteristics are!
  - We’d very much like to be proven wrong on this point!
- Need to get the message out ASAP that people deploying PR face a data corruption exposure in systems with old bridge products
  - In short term need all bridges that still have development support to be modified to either discard or reject PR commands if they don’t support PR
  - And here’s an idea for a longer-term solution……
Traceroute example

Tracing route to t11.org [66.155.124.38] over a maximum of 30 hops:

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&lt;1 ms</td>
<td>&lt;1 ms</td>
<td>&lt;1 ms</td>
<td>198.18.0.1</td>
</tr>
<tr>
<td>2</td>
<td>1 ms</td>
<td>1 ms</td>
<td>1 ms</td>
<td>209.226.87.1</td>
</tr>
<tr>
<td>3</td>
<td>27 ms</td>
<td>3 ms</td>
<td>3 ms</td>
<td>10.37.37.201</td>
</tr>
<tr>
<td>4</td>
<td>3 ms</td>
<td>23 ms</td>
<td>3 ms</td>
<td>core1-vancouver-pos1-0.in.bellnexxia.net [206.10.8.101.13]</td>
</tr>
<tr>
<td>5</td>
<td>71 ms</td>
<td>88 ms</td>
<td>97 ms</td>
<td>64.230.229.37</td>
</tr>
<tr>
<td>6</td>
<td>71 ms</td>
<td>75 ms</td>
<td>73 ms</td>
<td>core2-chicago23-pos0-0.in.bellnexxia.net [206.10.8.103.114]</td>
</tr>
<tr>
<td>7</td>
<td>74 ms</td>
<td>71 ms</td>
<td>76 ms</td>
<td>bx1-chicago23-pos11-0.in.bellnexxia.net [206.108.103.125]</td>
</tr>
<tr>
<td>8</td>
<td>71 ms</td>
<td>72 ms</td>
<td>89 ms</td>
<td>p13-0.core01.ord01.atlas.cogentco.com [154.54.11.29]</td>
</tr>
<tr>
<td>9</td>
<td>160 ms</td>
<td>140 ms</td>
<td>127 ms</td>
<td>p15-0.core02.ord01.atlas.cogentco.com [66.28.4.62]</td>
</tr>
<tr>
<td>10</td>
<td>423 ms</td>
<td>381 ms</td>
<td>181 ms</td>
<td>p6-0.core02.jfk02.atlas.cogentco.com [66.28.4.85]</td>
</tr>
<tr>
<td>11</td>
<td>171 ms</td>
<td>201 ms</td>
<td>200 ms</td>
<td>p13-0.core01.phl01.atlas.cogentco.com [66.28.4.2]</td>
</tr>
<tr>
<td>12</td>
<td>244 ms</td>
<td>243 ms</td>
<td>99 ms</td>
<td>p4-0.core01.dca01.atlas.cogentco.com [66.28.4.17]</td>
</tr>
<tr>
<td>13</td>
<td>92 ms</td>
<td>114 ms</td>
<td>108 ms</td>
<td>p15-0.core02.dca01.atlas.cogentco.com [66.28.4.22]</td>
</tr>
<tr>
<td>14</td>
<td>104 ms</td>
<td>224 ms</td>
<td>104 ms</td>
<td>p14-0.core01.atl01.atlas.cogentco.com [66.28.4.161]</td>
</tr>
</tbody>
</table>

- Based on the same ICMP protocol as Ping
- Utilizes a “Time To Live” field in the IP Header that’s decremented by each router
- Provides a simple elegant scheme that allows an application to completely map out an infrastructure
Add Field to INQUIRY

- All current bridges can access byte 0 & 1 of CDB (to detect Reserve)
- Define new Level Count field (LC) containing unsigned integer in byte 1
- Rules as follows
  - If LC=0, existing behavior
  - If LC>0, decrement LC by 1
    - If result=0, respond to Inquiry command
    - Else, relay the command to the next level in the hierarchy as per SAM-3
- Would allow an application to map a hierarchy, and induce a “transparent” bridge to identify itself
  - Even supporting only EVPD=0 would be enough initially to locate the bridges
Feedback requested

• Is the situation with ancient bridges a candidate for an “info letter” sent to all T10 member companies?

• Does pursuing the “traceroute” Inquiry make any sense?