Large scale SPVC ATM network

Written by Paul Holland and Don Wells

Introduction:

Asynchronous Transfer Mode (ATM) Data Networks differ from Internet Protocol (IP) Data Networks in many ways, but one of the most significant is the connection approach of ATM vs connectionless method of IP. An ATM network is able to guarantee a Quality of Service (QoS) because data flows over defined paths or connections from end-to-end, instead of the self routing packets used by IP. Once an end-to-end connection is created, then all data from the source will flow over the same path through the network to the destination. This allows the network manager the ability to ensure that each physical link is not encountering congestion and consequently packet loss. In an IP network, the individual packets get routed along the way and may take different paths. The problem with that is that they may arrive at their destination in a different order and if they encounter a node that is congested, then the packet may be discarded. This is why voice over IP (VoIP) will not work on a normal IP network. MPLS makes specified paths through an IP network – so it will make the IP network act more like an ATM network than regular IP; however, the paths in an MPLS network are uni-directional, which may cause problems with real-time applications.

The primary method of provisioning connections in an ATM network is through a Network Manager System (NMS), which creates “permanent” virtual connections (PVCs) on each node in the data path. Essentially, a cross-connect from one physical interface to another is created on each node from the source to the destination. Each connection must be provisioned by the NMS and/or by the NMS operator. This can be a time consuming and tedious task, and rerouting around damaged physical links can be quite slow.

In 1997 Alcatel (Newbridge at the time) was working on developing a type of connection called an SPVC (Soft Permanent Virtual Connection) in their ATM networks.

The main idea behind an SPVC is that it would act like a PVC except in how the connection is provisioned and how it reroutes in case of a failure. To provision an SPVC, the source and destination endpoints are identified on their respective nodes, and then the nodes in the network perform the routing of the data path and provision each individual link to match the details of the source and destination endpoints. The routing protocol that calculates the path of the connection is called Private Network to Network Interface (PNNI). PNNI has knowledge of the network (or at least its local portion of the network), and it will select the path with the lowest cost (as assigned by the network administrator) and highest available bandwidth.

There are two primary benefits of SPVCs over PVCs:

- An SPVC is much easier to create, especially in a network with a large number of nodes, and

- If a link fails, then the source endpoint will recreate the link at a rate much faster than can be done with a central network manager can achieve. SPVCs can reroute at a rate of about 250 or 1000 cross connects per second per node (depending on the node type) as compared to a rate of 5 per second for the whole network by a network manager.

It is difficult to give an example indicating the difference in reroute speed that can be realized by using SPVCs instead of PVCs. The reason is that SPVC reroute speed is dependant upon the number of nodes in the network that have source endpoints. If the source endpoints for the connections that cross a failed link are very distributed and there are many paths through the network, then SPVC can realize a reroute speed in the tens of thousands per second. If all the calls must go through a single parallel link of the failed link, then the reroute speed will be limited by that choke point. In a separate simple test, we had 250,000 calls that were rerouted from a failed link on one pair of nodes to a single parallel link. We saw a reroute time of just over 4 minutes (or 1000 calls per second through the single parallel link). In contrast, the network manger would have taken over 13 hours to reroute these calls at 5 calls per second. The SPVC rate would be much less, if there were other routes available through the network (a more realistic customer situation).

The Problem

Alcatel had implemented this great technology (which is also ATM forum standards compliant). The problem is that after 5 years we still had no large customer using SPVCs. We had a few major customers that were interested in the concept of SPVCs but they wanted proof that they could function at the level required by their networks.

We were asked by one of our largest customers, that was very interested in the benefits of SPVCs, to test SPVCs with a large network that had over 500,000 connections, and prove that large scale reroutes worked.

The Switched Data Networks (SDN) division of Alcatel created a special lab called the SDN Scalability Lab (SSL), and developed 3 new Test Plans to do this performance testing.

Summary of the Lab:

The primary objective of the SSL was to get SPVCs deployed in the field by our larger customers and have them act as a catalyst to develop new customers and thus drive more sales of our networking gear.

The lab consisted of 24 nodes and a Network Manager (the Alcatel 5620). There were two types of nodes in the SSL lab:

- 20x 7470 (or 7270) - first generation switches (lower capacity than second generation – reroute speed of 250 calls/sec)
- 4x 7670s - second generation switches (higher capacity of connections, physical links , and reroute speed - 1000 calls/sec)

Results

The SSL had only one full time employee, Don Wells, a senior verification designer with about 10 years experience. He authored the Test Plans, and did the majority of the test execution himself. The SSL and its testing lasted for about one and a half years. During that time there were 361 Problem Tickets raised:

- Just over half of those were deemed major or critical.
- 90% could have been found on 3 nodes or less. 75% of these could have been found with a single node and a piece of test equipment.
- 30% were against the diagnostic tools of the node and the NMS.
- VERY few (5) were raised against interoperability between the two generations of switches – this was a huge surprise.
- Most of the critical problems were situations that caused the nodes to reset.

Critical	Major	Minor	Total
23	164	174	361

Table 1: Summary of number of problems found by the SSL project by category.

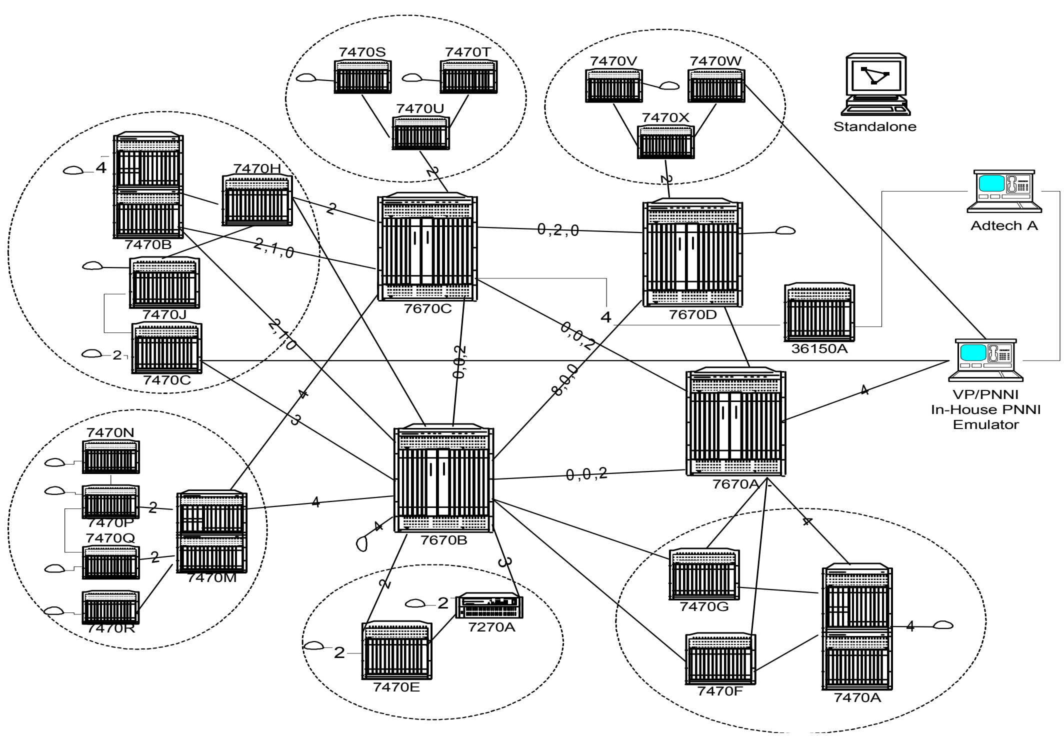

Figure 1 Physical Test Topology of Test Plan 51: see Appendix A for a brief description

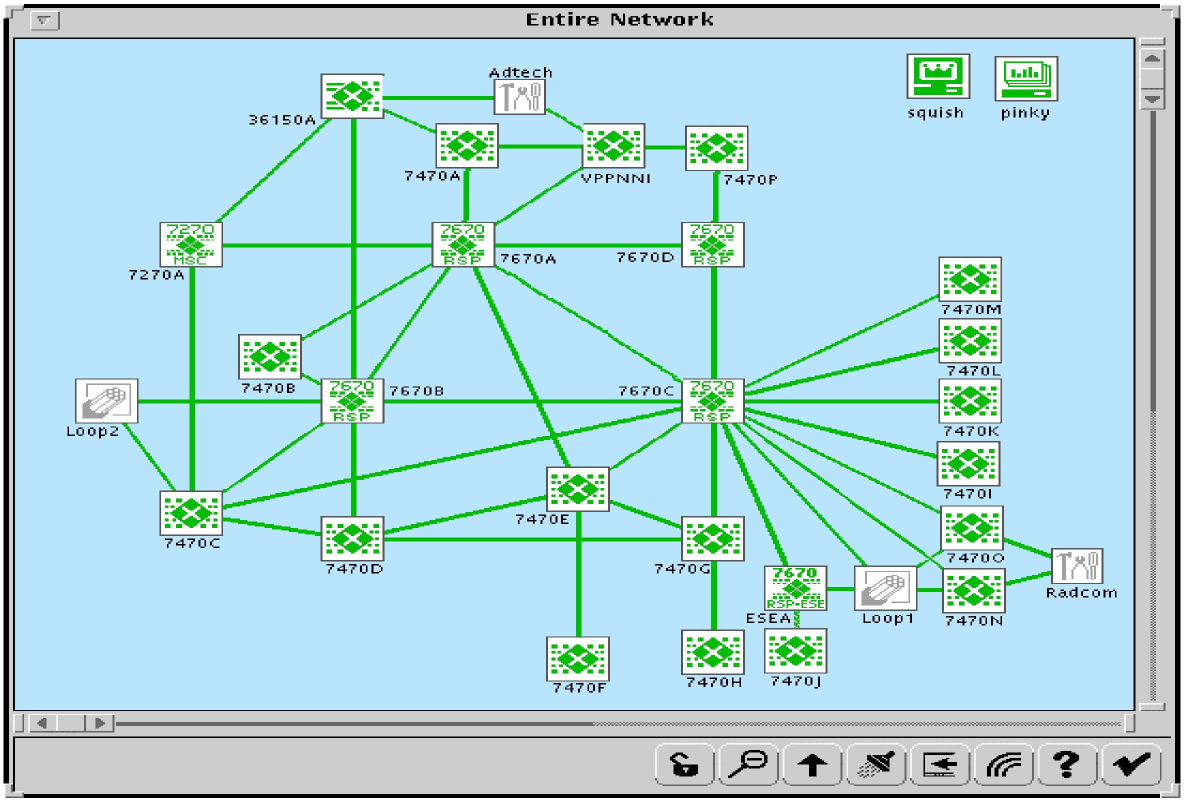

Figure 2: Test Topology of Test Plan 52 as viewed from the Network Manager. See Appendix A for a brief description.

Results (continued)

The test plans were designed to allow many different variables to come into play. The reason these TPs were so successful at discovering problems was because of the way features were allowed to interoperate. For example, the SPVCs that were being rerouted were “groomed” to follow specific paths through the network. All other SPVCs were allowed to route in whatever direction they chose. This was by design intent and it caused an element of randomness that helped find some interoperability problems but also meant that tests were not always reproducible. In a real-time system, even with the same external stimuli, the internal states are unknown and thus cause different results.

Overall the SSL was a major success. The primary reason for its success was a commitment from the design team to have people who knew the system and were VERY knowledgeable addressing the problems found. In a normal testing situation, a problem report would have been raised by the tester, and after a few hours (or days) a random designer would show up to fix the problem. In the SSL lab, we had dedicated top designers whose highest priority was investigating and fixing SSL found issues.

The Call Control recovery (reroute speed) needed to be measured accurately for the testing to succeed. The primary criterion for a test passing was to have all paths recovering in the expected time frame. The secondary criterion is that the paths recovered with a linear (or close to linear) rate. We had a problem of being able to determine the rate of recovery with the tools we had at our disposal. We could not use a data snake (a single path through the network where the output from one connection becomes the input of the next connection), because data would not flow at all until the last path in the snake was connected. No data how the recovery proceeded could be gathered. To solve this problem we created a VP switch out of one of our old ATM switches, which gave us a huge number of parallel data sources. We then set the rate of this data to be one cell per second. We now had a way of accurately measuring the number of paths that are currently up and able to pass data. This was crucial to the success of the SSL project.

In addition to collecting data on the number of paths passing traffic, the CPU utilization on the network manager was also gathered using the Unix "top" command.

Some of the results received in the early going led to some pretty sad looking patterns:

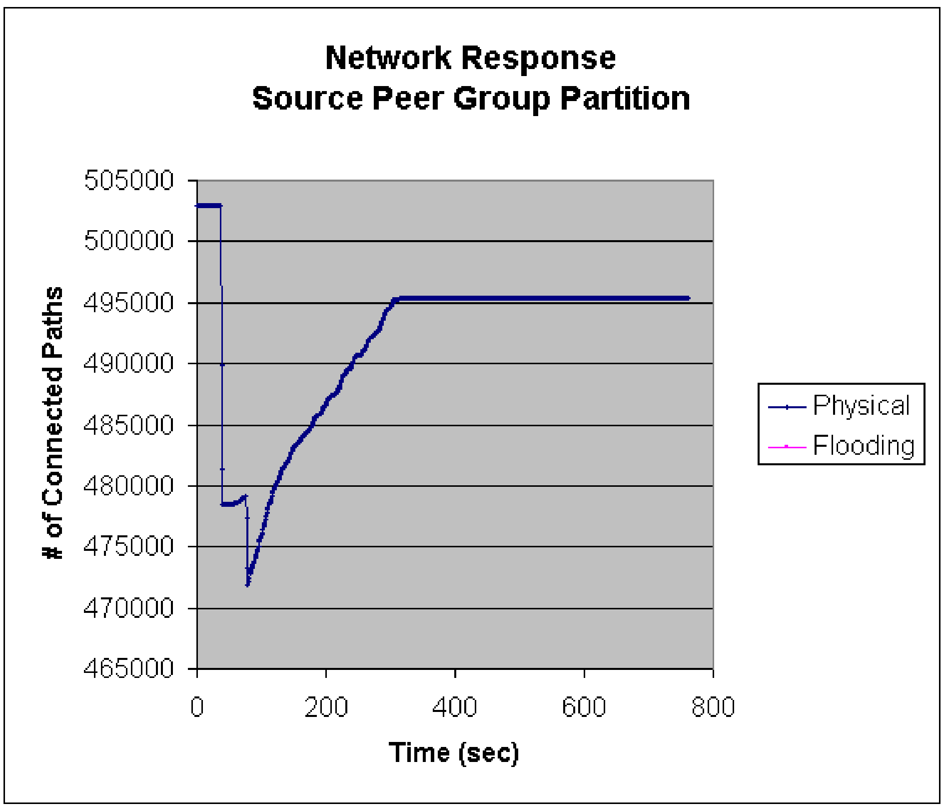

Figure 3: An early graph showing some paths not recovering at all

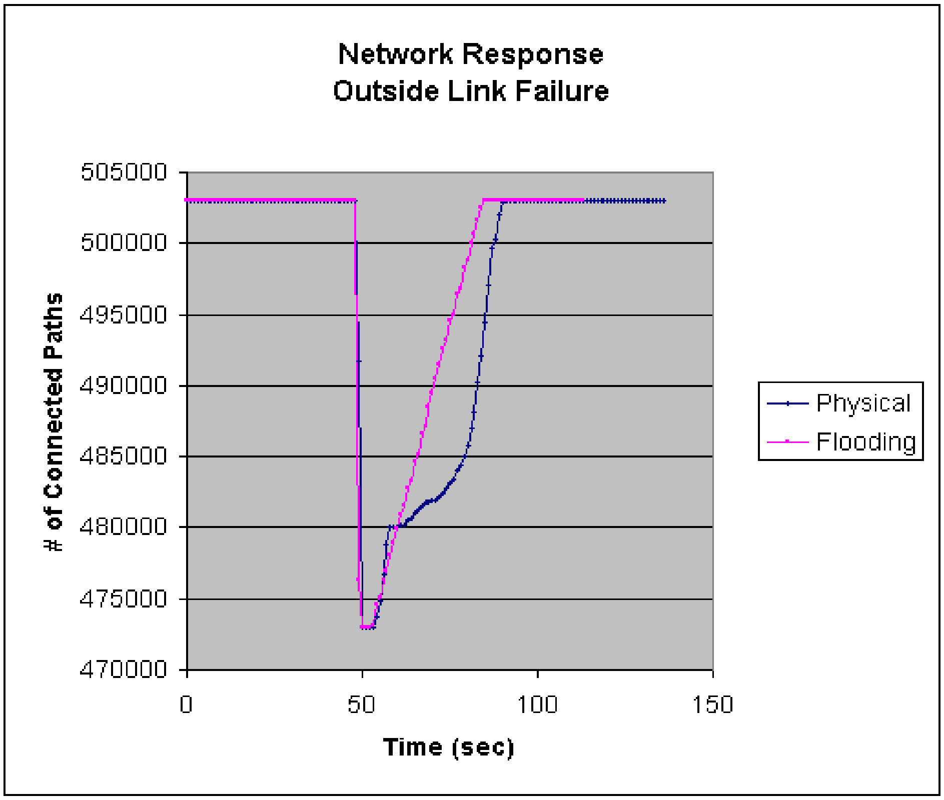

Figure 4: A graph showing an interesting shape of the physical (no emulated nodes) setup

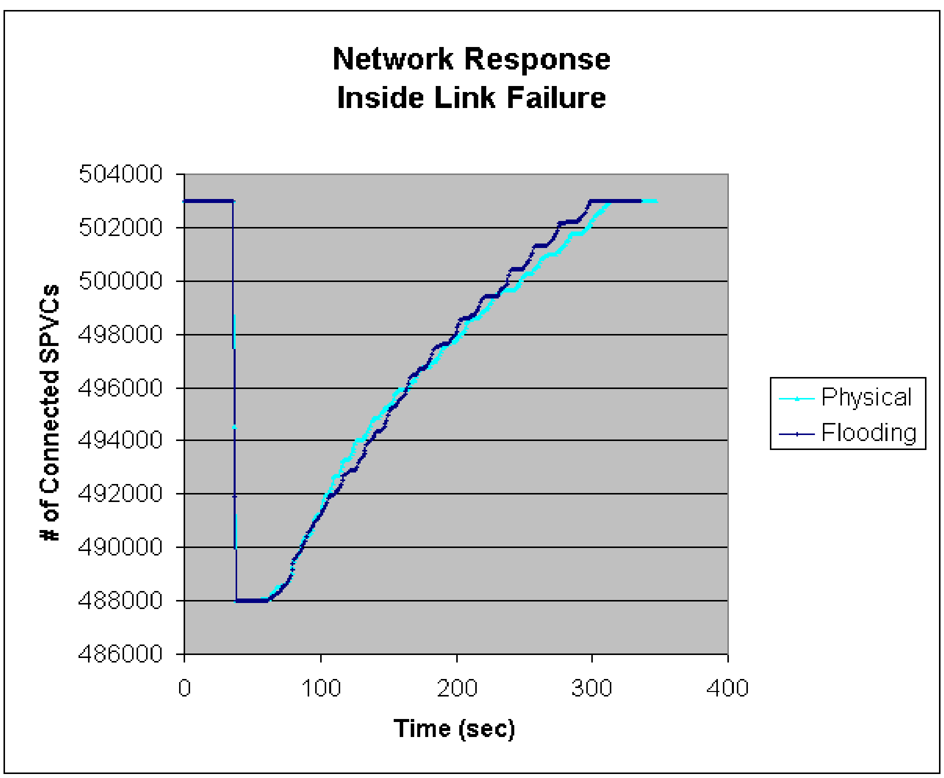

Figure 5: A graph that shows a proper recovery

The stepping in figure 5 is caused by a back off timer. Essentially, if a call attempt is blocked because of congestion or a temporary failure, then the source endpoint will wait for a given period of time before trying to connect again. The steps are caused by all the calls trying to reroute being in the sleep state, waiting for their timers to expire before retrying.

Focusing Questions:

1. As a result of this testing, many problems were fixed in all of the products involved. The customer that initially requested this testing has started to deploy SPVCs, and is happy with the progress. They plan on switching over their entire network from PVCs to SPVCs this year. Many other customers are also starting to deploy SPVCs.
2. The data that was collected was data throughput (of 1 cell per second per path in the network) as a measurement of Call Control Recovery of the network and CPU utilization of the Network Manager.
3. Created charts in Excel. They were very graphic in their ability to show problems in the progress of the rerouting of the SPVCs.
4. We were using and Adtech for the data flow, and the Unix “top” command for the CPU utilization. Both are proven tools that did not need extra validation.

New methods used:
Because we only had one Adtech port to work with, we modified an old Newbridge ATM switch (36150) to be a VP switch. Essentially, it took a single stream of data from the Adtech and it copied it over all the different paths in the network.

APPENDIX A

Summary of the Test Plans:

The testing was broken down into 3 test plans, each with the common theme of performing massive SPVC rerouting.

The following is taken directly from the first test plan:

> *The tests in this document are intended to validate Alcatel's managed SPVC solution, using large-scale re-routes to focus on two key areas:*
>
> 1. *Network survivability and robustness*
> 2. *System and nodal behaviour under stress*
>
> *The nodes in the test network are configured near the connection count limits; while this is not necessarily representative of typical network engineering, it provides the basis for stress conditions and optimizes the number of connections over the available hardware. In addition to proving the stability of the solution, the test results also provide insight into engineering guidelines to ensure network performance by and illustrating the impact of engineering high connection densities.*
>
> *The node-specific results are summarized below.*
>
> - *Alcatel 7670 RSP*
>   - *Power Cycle Test of node with 250,000 SPVCs terminating*
>   - *Outage time of 29 minutes for full recovery of traffic on all connections*
>   - *Power Cycle Test of node with 250,000 SPVCs transiting & 255,000 terminating*
>   - *Outage time of 13 minutes for full recovery of traffic on all connections*
>   - *OC48 reroute soak with 112,000 SPVCs rerouting*
>   - *120 iterations completed in 16.5 hours with 13.4 million SPVCs successfully rerouted*
> - *Alcatel 7470 MSP / 7270 MSC*
>   - *Results suggest optimal performance is achieved when configured as a destination of 7670-sourced SPVCs*
>   - *Power Cycle Test of compact node with 62,000 SPVCs terminating*
>   - *Outage time of 27 minutes for full recovery of traffic on all connections*
>   - *OC3 reroute soak with 16,000 SPVCs rerouting*
>   - *300 iterations completed in 12 hours with 4.8 million SPVCs successfully rerouted*

Summary from the second Test Plans:

This document describes the tests used to validate Alcatel's managed SPVC solution using large scale reroutes across a large, non-hierarchical PNNI network. The focus is on network survivability and robustness, as well as system/nodal behaviour under stress.

This solution test plan has the following objectives:

- b) *To verify the stability of the Alcatel S-PVC solution set during normal and failure conditions.*
- c) *To measure network & NMS performance under various failure scenarios in a large PNNI network namely:*
  - *> 500,000 SPVCs with monitored traffic*
  - *Emulation of 100 PNNI nodes*
  - *High PNNI message flooding rate*
  - *Large network diameter (hop count)*
  - *Alternate path selection during reroute*
  - *Utilize 7470 nodes as SPVC transit nodes*
  - *7470 / 7670 OC12*
  - *25 physical nodes which are arranged as clusters around a core*

The results are summarized below:

- *All tests passed*
  - *All paths passing traffic, as monitored by analyser, after reroute settled*
  - *All paths in 'Connected' state on 5620 after NMS activity settled*
  - *No unexpected alarms on any nodes; all alarms forwarded to 5620.*
- *Network size had no impact on rerouting performance*
  - *Physical (24 nodes) & physical/logical (125 nodes) reroutes identical in all tests.*
- *PNNI messaging traffic had little impact on rerouting performance*
  - *Physical (24 nodes) & physical/logical (125 nodes) reroutes diverge when SPVC back-off sequence enters maximum range.*
- *A simultaneous power cycle of all nodes in the network was executed*
  - *Traffic restored on all 504,000 SPVCs in 23 minutes*
  - *5620 synchronization in 11 hours.*
- *5620 re-synchronization time affected by nodal hold-down timer; engineering guideline needed for timer value which is based on number of connections sourced by the node with following reroute rates.*
  - *7470-7470 & 7470-7670 approximately 100 calls/second*
  - *7670-7670 approximately 2000 calls/second*

The third test plan was similar to the second in the tests executed, but the third had 3 levels of PNNI hierarchy and more nodes being emulated:

- *In-house emulation of 100 PNNI nodes*
- *PNNI Neighbor emulators in the transiting peer group forcing a 7670 node to PNNI flood to 81 neighbours*
- *Radcom Emulator with 21 LGNs at the 2nd level and 24 LGNs at the 3rd level*
- *In-band CPSS through a single gateway node using default bandwidth.*
- *Statistics collection via in-band SNMP polling*
- *High sustained PNNI message flooding rate via the in-house emulator.*