z/OS Communications Server

The Evolution of SNA:
z/OS CS Enterprise Extender
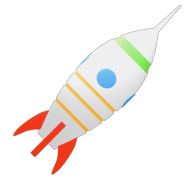# Hints & Tips

z/OS CS Design & Development, Raleigh
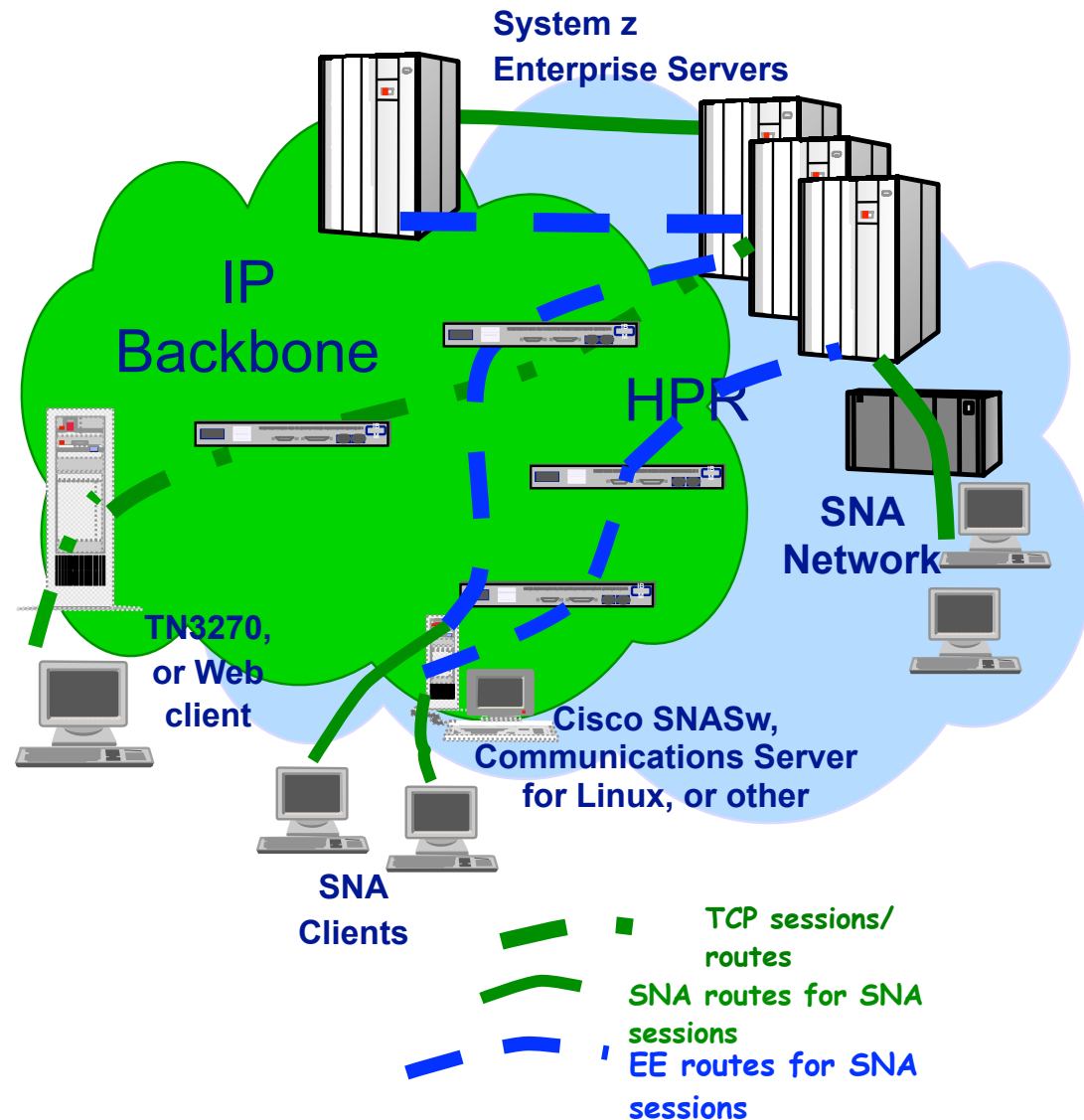
Sam Reynolds - samr@us.ibm.com

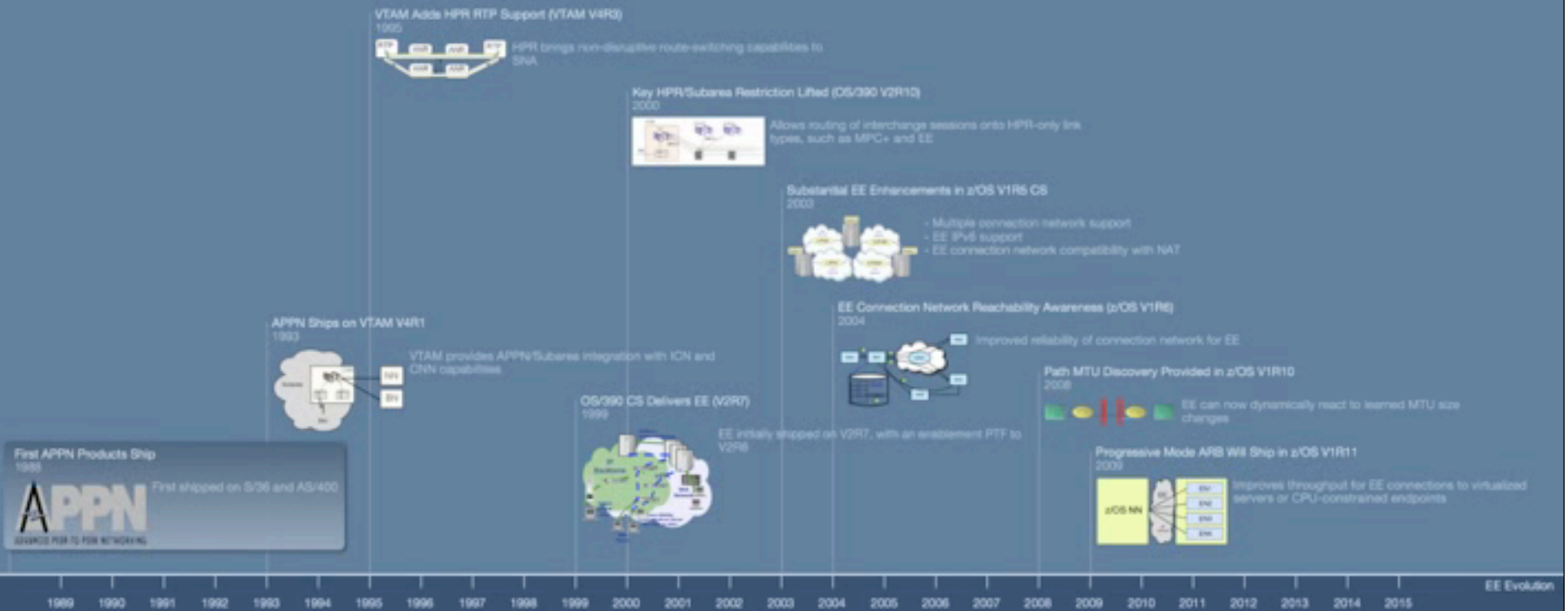SHARE 2010 Summer Technical Conference

What is Enterprise Extender?

- Allows use of IP network for SNA sessions
- EE allows enablement of IP applications and convergence on a single network transport while preserving SNA application and endpoint investment.
- Typically isolates SNA footprints to the "outside" of the network.

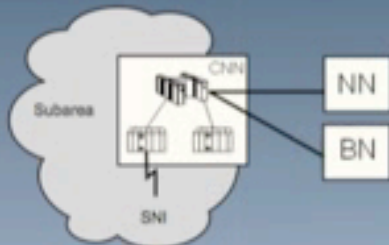**System z Enterprise Servers**

IP Backbone

HPR

SNA Network

TN3270, or Web client

Cisco SNASw, Communications Server for Linux, or other

SNA Clients

TCP sessions/ routes

SNA routes for SNA sessions

EE routes for SNA sessions

The Evolution of EE

VTAM Adds HPR RTP Support (VTAM V4R3)
1995
HPR brings non-disruptive route-switching capabilities to SNA

Key HPR/Subarea Restriction Lifted (CS/390 V2R10)
2000
Allows routing of interchange sessions onto HPR-only link types, such as MPC+ and EE

Substantial EE Enhancements in z/OS V1R5 CS
2003
- Multiple connection network support
- EE IPv6 support
- EE connection network compatibility with NAT

EE Connection Network Reachability Awareness (z/OS V1R6)
2004
Improved reliability of connection network for EE

Path MTU Discovery Provided in z/OS V1R10
2008
EE can now dynamically react to learned MTU size changes
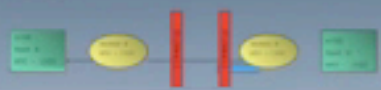
APPN Ships on VTAM V4R1
1993
VTAM provides APPN/Subarea integration with ICN and CNN capabilities

OS/390 CS Delivers EE (V2R7)
1999
EE initially shipped on V2R7, with an enablement PTF to V2R6
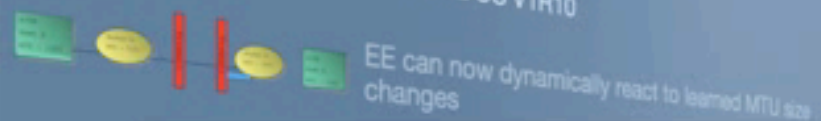
Progressive Mode ARB Will Ship in z/OS V1R11
2009
Improves throughput for EE connections to virtualized servers or CPU-constrained endpoints

First APPN Products Ship
1988
First shipped on S/36 and AS/400

APPN
ADVANCED PEER-TO-PEER NETWORKING

1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015

EE Evolution

t (VTAM V4R3)

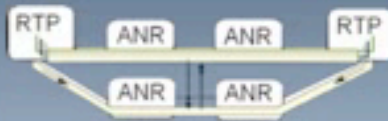brings non-disruptive route-switching capabilities to

Key HPR/Subarea Restriction Lifted (OS/390 V2R10)
2000

Allows routing of interchange sessions onto HPR-only link types, such as MPC+ and EE

Substantial EE Enhancements in z/OS V1R5
2003

- Multiple connection network support
- EE IPv6 support
- EE connection network compatibility with NAT

EE Connection Network Reachability Awareness (z/OS V1R6)
2004

Improved reliability of connection network for EE

Path MTU Discovery Provided in z/OS V1R8
2006

EE can now dynamically react to owned MTU size changes

Progressive Mode ARB Will Ship in z/OS V1R11
2009

ICN and

EE Recommendations

- #1 problem in bringing up EE connections: Firewall issues.
  - Any underlying firewall must allow the transport of UDP ports 12000-12004 (possibly limited to known EE endpoint addresses)

UDP Port 12000
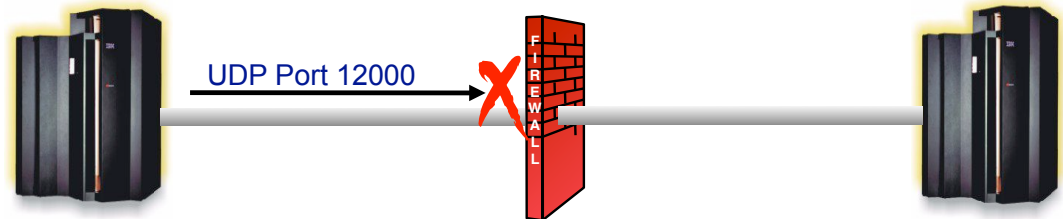
- Consider enabling PSRETRY (off by default) so that HPR pipes will automatically switch to better routes when available.
  - PSRETRY is usually of value only when alternate APPN routes exist.
- Consider enabling HPR Path Switch Summarization (V1R9).
- Configure APPN Link Characteristics
  - TGPs for EE provided with VTAM - Customization of link speed is recommended.
- If coding DWACT=YES and/or DWINOP=YES, it is better to specify these parameters on only one side of the connection (to avoid "PU Busy" type conflicts).
- If defining an EE Connection Network over an IP network which employs Network Address Translation (NAT), you must define the virtual routing node's addressability using the HOSTNAME operand (not the IPADDR operand).
- If using EE with IPSec on V1R10, APAR PK93190 is strongly recommended to improve throughput.

EE Recommendations...

- Consider lengthening the EE LDLC timer parameters (LIVTIME, SRQTIME, SRQRETRY), especially if RIP is being used.
  - It is recommended that the LDLC timer parameters be adjusted on both ends of the connection.
- Lengthen HPR path switch timers (HPRPST) as necessary to ensure that all four timers are longer than the LDLC timeout interval.
  - This will ensure that RTP pipes stay in path switch long enough during IP network instability to allow the EE link to inop, and thereby allow another path to be selected.
  - If multiple LDLC parameter sets are in use (by coding different LIVTIME, SRQTIME, and SRQTIME values on different XCA GROUPs), then the HPRPST values should be adjusted relative to the longest of the LDLC timeout intervals.
- EE MTU processing was improved in V1R9 (local MTU discovery) and V1R10 (path MTU discovery for EE).

RTP PU
CNR00001

EE PU
PU1, CNV00001, etc.

EE LDLC

z/OS Host A MTU - 1500 — Router A MTU - 1500 — Firewall — Firewall — MP(1200) — Router B MTU - 1200 — z/OS Host B MTU - 1500

14
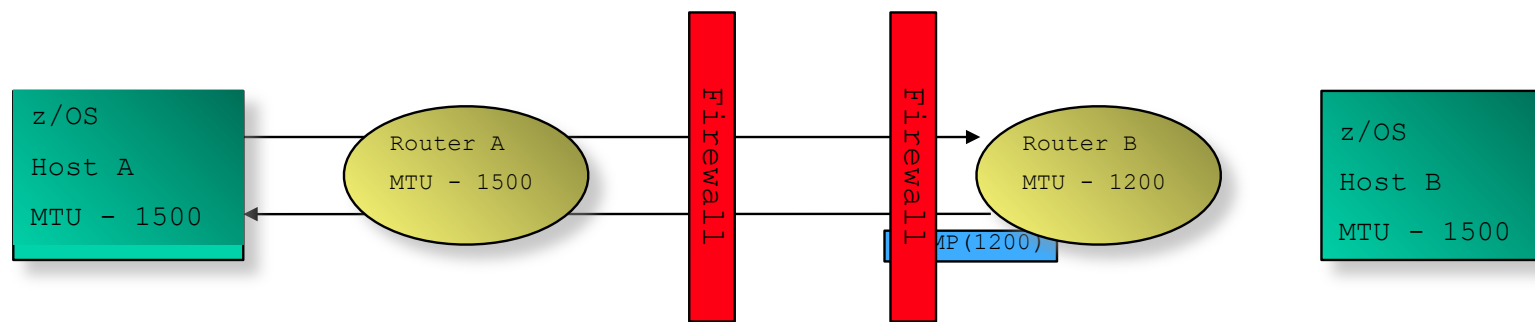
EE/HPR Timer Notes

N O T E S

- The EE LDLC layer monitors the connection, testing it if no activity is detected, and inop'ing the EE link if the tests go unanswered
  - Total LDLC timeout interval = LIVTIME + ((SRQRETRY+1) * SRQTIME)
- The Disconnect timer associated with the EE switched PU is used to trigger an inop if no activity is detected for a specified amount of time
- The RTP layer is responsible for driving status requests frequently enough (in the absence of data traffic) to keep the disconnect timer from tripping. The RTP endpoint will inop itself if its last session goes away, and no new session is queued to it for a period of 10 seconds (or the period specified in the RTP model, if used)
- The path switch timer (set by HPRPST) controls the amount of time that an RTP pipe will stay in path switch state, trying to find an alternate route, before abandoning the attempt and inactivating the pipe.

EE XCA Major Node Recommendations

- Define all of your EE connection networks on the GROUP statements, and not on the XCA PORT. This will provide more consistent definitions as your connection networks expand in the future, and positions you to utilize GROUP-based configuration options.

```
DEMOXCA   VBUILD TYPE=XCA
DEMOPORT PORT    MEDIUM=HPRIP
*
DEMOGP    GROUP  DIAL=YES,CALL=INOUT,                                X
                 AUTOGEN=(5,EV4,P),DYNPU=YES,ISTATUS=ACTIVE
***********************************************************************
* EXAMPLE VRN                                                         *
***********************************************************************
DEMOVRN   GROUP DIAL=YES,CALL=INOUT,VNNAME=NETA.LVRN,                 X
                AUTOGEN=(5,LV01,P),DYNPU=YES,VNTYPE=LOCAL,            X
                HOSTNAME=TUVIPA2.AREA51.SVT390.COM,                   X
                ISTATUS=INACTIVE,TGP=GIGENET
```

  - Having GROUP-based definitions also enables more flexible and less disruptive updating of the EE XCA definitions via the VARY ACT,UPDATE=ALL command

- If IP address availability allows, use a different VIPA for each connection network that you define. This serves to differentiate the VRNs for purposes of manageability, etc.
- Leave the IPPORT operand at its default of 12000. If you change it, you must change it on all EE platforms that you connect to, and some do not support such a change.
- The IPTOS operand on the PORT is used to set the values for the IP Type of Service byte for each of the five EE priorities (low, medium, high, network, and signal).
  - Default: IPTOS=(20,40,80,C0,C0)
  - However: Only the three highest-order bits can be set! Any attempt to set any of the lower five bits in the byte will result in an "invalid parameter" error message
  - The full TOS byte can be customized by using the policy agent

16

Effect of IP Multipath on EE
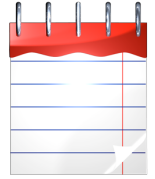
- The IPCONFIG MULTIPATH parameter in the TCP/IP profile enables the multipath routing selection algorithm for outbound IP traffic. (The default is NOMULTIPATH.)
  - When MULTIPATH is enabled, there are two options to choose between: PERCONNECTION and PERPACKET, with PERCONNECTION being the default choice.
- If MULTIPATH is enabled, the choice of PERCONNECTION vs. PERPACKET makes no difference for EE.
  - In either case, IP will use a "per-batch-of-packets-from-VTAM" approach.
  - Since there is no "UDP connection", true PerConnection multipath is really not possible
  - PerPacket is too granular as it leads to too much resequencing overhead at the RTP receiving endpoint.
- **Recommendation**: If you already have MULTIPATH enabled for your TCP applications, there is no requirement to change it. But in general, MULTIPATH is probably not a particularly good idea for EE as it often leads to extra resequencing overhead at the receiver.
- z/OS V1R12 CS will provide a separate VTAM start option which will by default disable multipath for EE
  - MULTPATH = <u>NO</u> | TCPVALUE

EE Connection Network

- A connection network is an APPN technology that reduces the need for predefining APPN links between nodes that are connected to a shared transport
  - With EE, the IP network itself is the shared transport.
- The shared transport (IP network in the EE case) is represented as an APPN Virtual Routing Node (VRN).
- All EE nodes participating in the connection network can send EE packets directly to each other without defining links to all the other participating nodes.
  - Connections are dynamically defined between VRN partners as needed.
- Generally the combination of EE with connection network technology is recommended with the objective of reducing the amount of link definitions that are required and to allow EE endpoint to endpoint communication to flow directly between the associated IP endpoints.
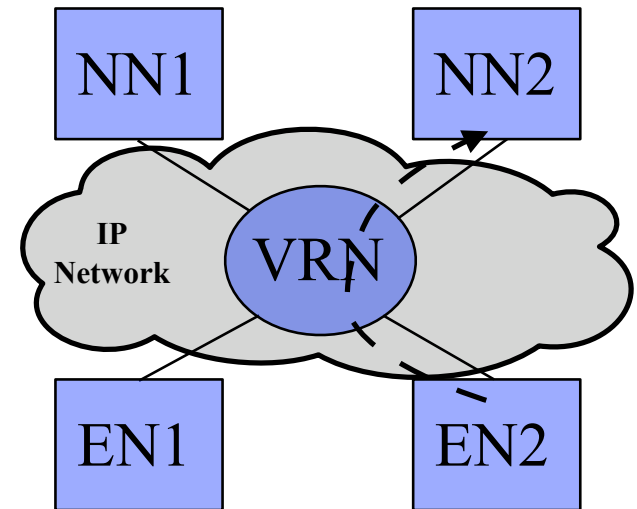
Example: Since EN2 and NN2 both define the VRN, a dynamic connection can be activated between them over the VRN without predefining a static EN2-to-NN2 connection

EE Connection Network Tips

- By default, the connection network dynamic PU name will be CNVxxxxx, where xxxxx is a sequence number. The PU name can be customized by specifying the DYNVNPFX GROUP or start option, resulting in the name yyxxxxxx, where "yy" is the two-character prefix set by DYNVNPFX.
- A limited set of characteristics of the dynamic connection can be customized by coding a DYNTYPE=VN PU entry in the model major node:

```
MODELA1A VBUILD TYPE=MODEL
*
VNMODEL   PU       DYNTYPE=VN,
                   DISCNT=NO
```
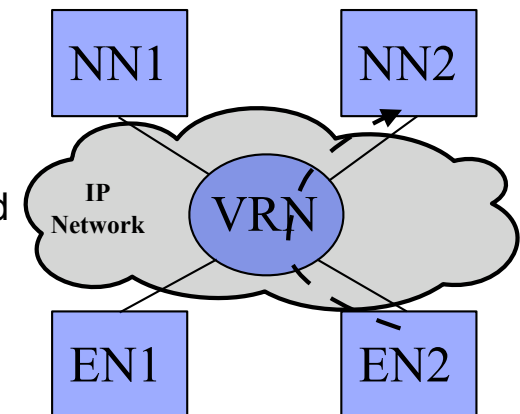
  - The primary use of the DYNTYPE=VN model is to customize the DISCNT value.
  - TG characteristics cannot be specified on a DYNTYPE=VN model PU
    - The DYNTYPE=VN model defines the dynamic PU used for the actual VN connection, not the virtual routing node (VRN) and its associated TGs. The TG created with the dynamic PU is not reported to Topology and Routing Services and is not involved in route calculations. To configure the weight of the TGs associated with the VRN, specify the TG characteristics on the GROUP statements in the XCA major node.
  - Consider coding a DYNTYPE=VN model, with DISCNT=NO, or a delay value of 60+ seconds.
    - **Important note for CICS LU6.2 Users**: Specifying DISCNT=NO will prevent CICS from terminating its sessions at the end of every transaction.
- For any APPN TG to be activated, whether predefined or dynamically-activated (as is usually the case with connections associated with connection networks), an ADJCP entry must exist. If no ADJCP entry is coded for the partner CP, and DYNADJCP=NO,then the APPN connection activation will fail. So, to allow connection network connectivity you must set DYNADJCP=YES, or code ADJCP definitions for the partner CPs.
  - This provides a mechanism to control which partners are allowed to connect to your CP

Connection Network Failures
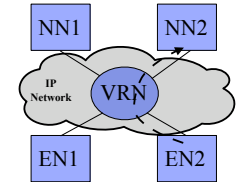
- Typical VRN Configuration
  - ENs predefine links to NNS for CP-CP
  - Dynamic CN links used EN-to-EN

- What happens if IP network fails?
  - Affected RTPs begin path switching
  - VRN path is chosen again
    - Topology component is *not* notified of the failure
    - VRN path still has lowest weight
  - Path switch fails even though a functional alternate path exists (through NN)

- VRN-specific message issued:
  - For VRN dial failures and VRN link INOPs
    - IST1903I - Identifies the VRN and partner node
  - Can be used to drive automation that quiesces the VRN or increases the weight of the link to the VRN

IST1903I FAILURE OVER VRN NETA.VRN1 TO CP NETA.WORKSTN1

NETA.EN

IP Network

VRN1

NETA.NN

NETA.WORKSTN1
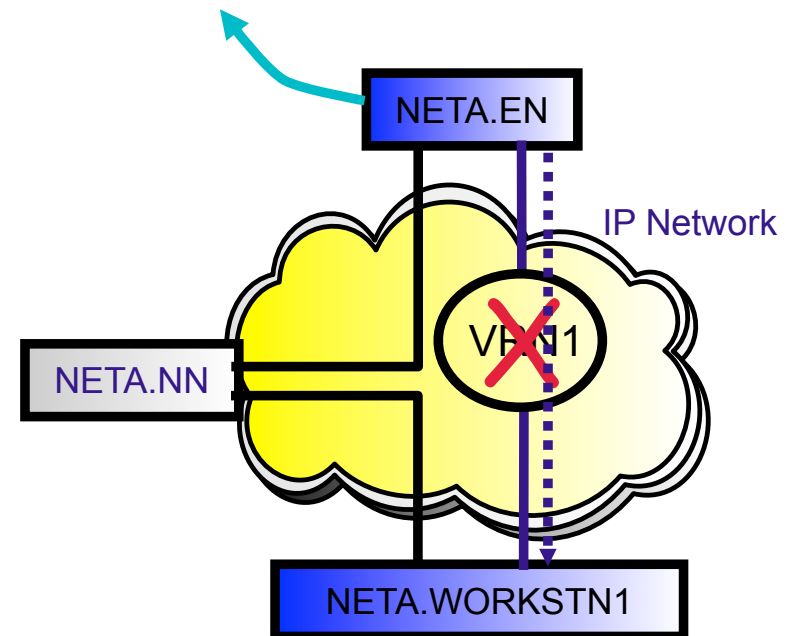
EE Connection Network Reachability Awareness
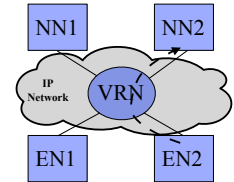
1 NN1 successfully contacts NN2 across VRN1.

2 NN1's attempt to contact NN3 across VRN1 fails.

3 In NN1's topology database, an "unreachability record" is associated with VRN1 for the partner NN3 (with a duration of 60 seconds) and a TDU is sent to partner NNs. The NN1->VRN1->NN3 route will not be used for the next 60 seconds.

4 With the unreachability record in place, a route is chosen from NN1 to NN3 by going through NN5 instead of VRN1.

Topology Database

NN1	
NN2	
NN3	
VRN1	NN3 60s
...	
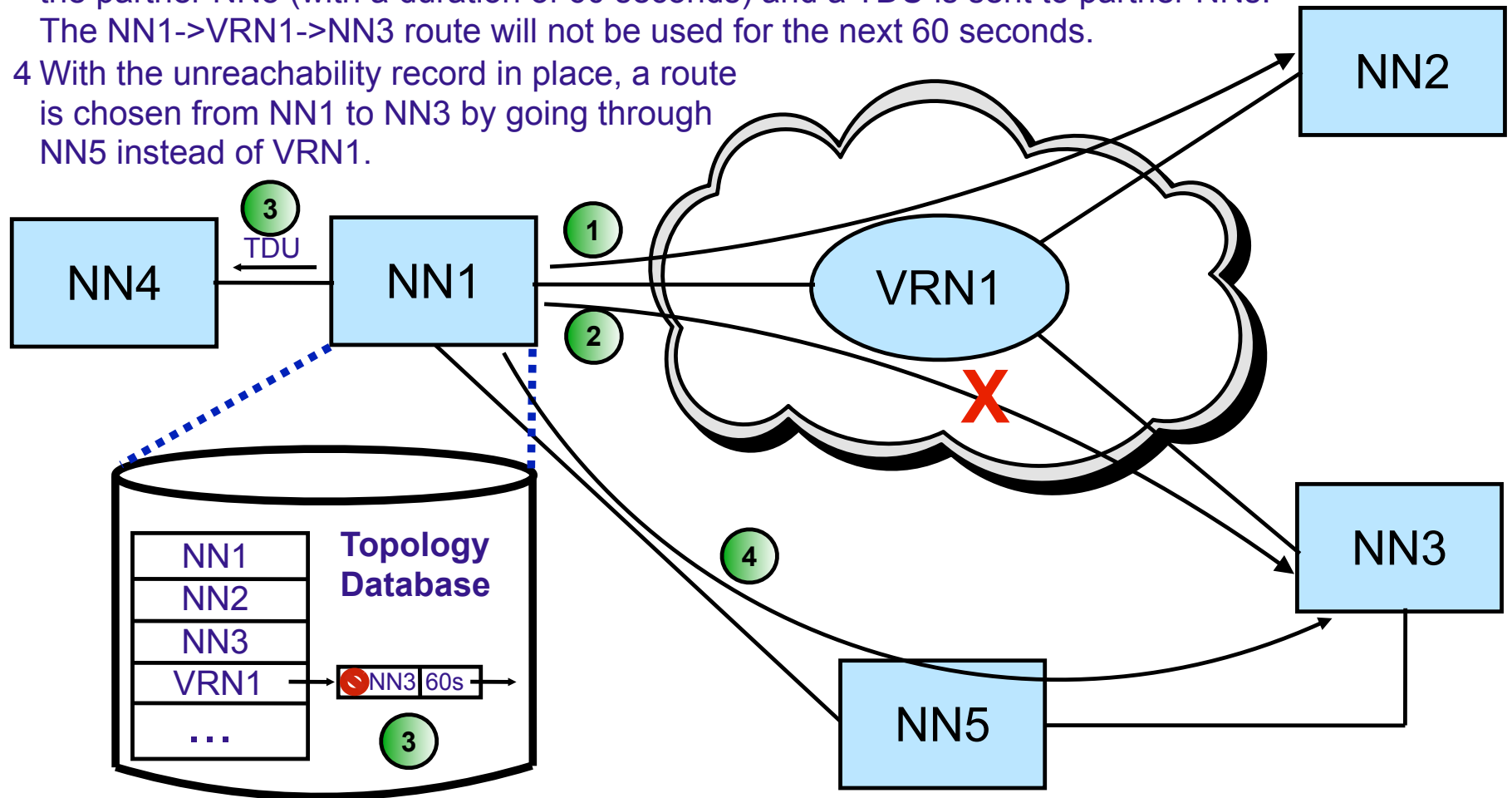
EE Connection Network Reachability Awareness

N O T E S

- EE Connection Network Reachability Awareness detects a dial failure or connection INOP for a connection over an Enterprise Extender connection network and prevents that specific path to the partner node from being used for a period of time. If alternate paths are available, APPN Topology and Routing services will select the optimal alternate session path for session establishment or an HPR path switch.
- When the time expires, if the path through the EE virtual routing node (VRN) still has the lowest weight of any available path to the partner node, the path over this particular VRN will be selected on the next attempt to redial the partner node.
- Unreachable partner information is maintained in the Topology Database and is associated with an EE VRN or with an end node that is on the origin side of the VRN.
- Unreachable partner information is sent to an end node's NNS or broadcast to a network node's adjacent network nodes in Topology Database Updates (TDUs).
- The period of time that a path through the EE VRN to the unreachable partner will remain unavailable is configurable. The UNRCHTIM start option allows the specification of the default number of seconds that a partner node for a session path through an EE connection network is considered unreachable after connection network failures.
- During this UNRCHTIM period, the path through the EE VRN to this partner node will not be considered for new sessions or HPR path switches.
- UNRCHTIM can also be specified on the PORT and/or GROUP statements in the EE XCA major node. This provides the capability of specifying different unreachability durations for connection networks of different characteristics, or to different business partners, etc.
- The range for UNRCHTIM is 0, or 10-65535 seconds.
- UNRCHTIM=0 indicates that paths through EE connection networks will always be considered for routing. This is the default value.
- Current unreachability information can be displayed using the DISPLAY TOPO command.
- Unreachability records may be cleared using the MODIFY TOPO command.
- To prevent the performance impact of unreachability lists growing without bound, once a certain number of records is associated with a connection network, that VRN will be treated as quiesced. No further attempts will be made to use it (and no more records will be added) until the record count drops below a threshold.
- A second parameter specifiable with UNRCHTIM allows control of the number of records at which the VRN will be quiesced. The VRN will be considered usable again when the record count drops below 80% of that value. So, for example, UNRCHTIM=(60,20) would set the unreachable time to 60 seconds, with the record count allowed to grow until 20 before the VRN is quiesced. The quiesced status would be cleared when the record count dropped below 16 (80% of 20).
- If UNRCHTIM is enabled, the default value for the record limit is 10.

Topology Garbage Collection
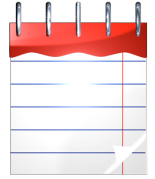
- An integral part of the maintenance of the APPN Topology Database is the garbage collection process.

- Each topology resource has a "days left" field:
  - Every time a topology update is received about the resource, the "days left" field is set to 15 days.
  - Once per day, every topology resource in the database is examined, with each "days left" field decremented. If a resource's "days left" field reaches zero, that resource is marked for garbage collection.

- The garbage collection activity occurs once per day, and by default it will occur at 24 hour intervals after VTAM initialization.

- For large topology databases, garbage collection has the potential to be a long-running process that could affect performance.

- If VTAM was restarted at a non-optimal time (that is, during a time of day that typically has high activity), note that garbage collection will run at that same non-optimal time every day.

APAR OA21513 (V1R9) provided a new start option to specify the time at which topology database garbage collection should occur each day:

  TOPOTIME = *VTAM-init-time* | *preferred-gc-time*

where *preferred-gc-time* is in the range 00:01 - 24:00

NN1

Topology Database

Resource	Days Left
NNA	14
TG1	14
TG2	14
NNB	11
TG3	11
NNC	1
...	...
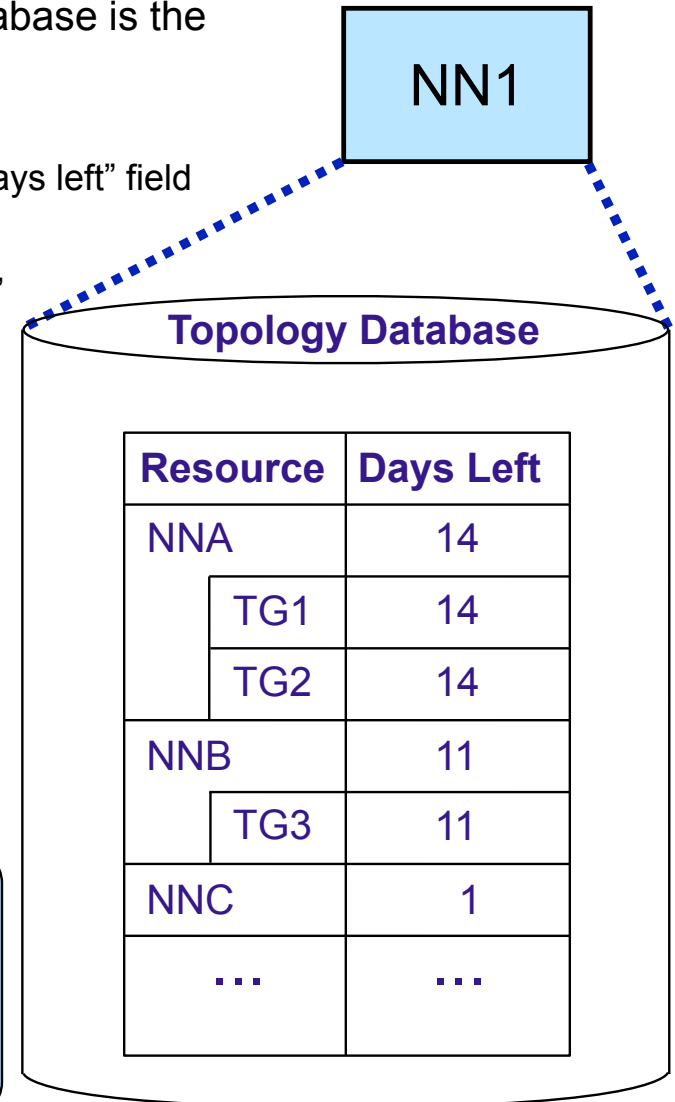
Preventing Unwanted Searches from EBN Partners

- V1R8: Predefine Non-Native ADJCPs With ALIASRCH=NO

  - Prevents ALIAS searches from entering your network through one or more adjacent CPs
    - Forces adjacent customers to ensure that searches entering your network are network-qualified
  - Default value of ALIASRCH=YES can still be used for "trusted" or "test" networks

- V1R10: AUTHNETS=(NET1,NET2,...) On ADJCPs

  - Only allow EBN partner to search for targets with Net IDs specified in the AUTHNETS list

```
************************************************************
* Adjacent CP Major Node for NETA.A1                      *
************************************************************
*
ADJCPA1    VBUILD  TYPE=ADJCP
B1         ADJCP   NETID=NETB,ALIASRCH=NO,AUTHNETS=NETA
B2         ADJCP   NETID=NETB,ALIASRCH=NO,AUTHNETS=(NETA,NETC)
C1         ADJCP   NETID=NETC,ALIASRCH=YES
D1         ADJCP   NETID=NETD,ALIASRCH=NO,AUTHNETS=
```
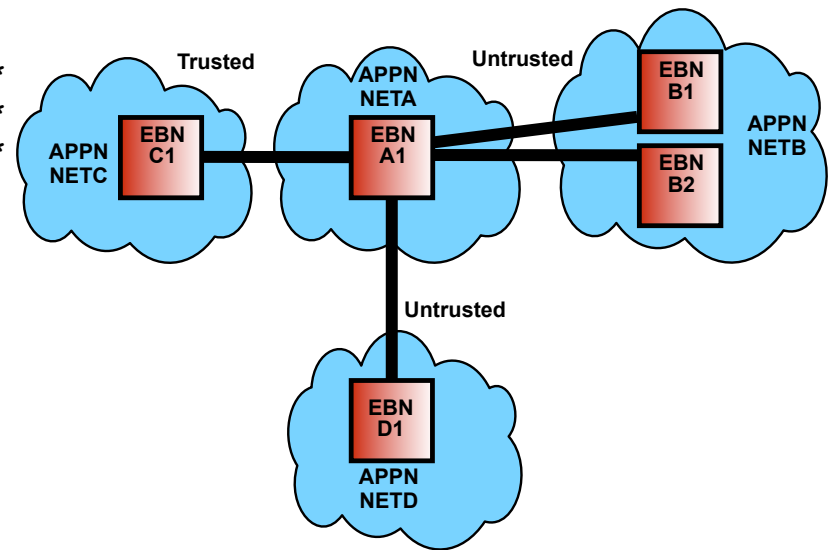
Diagnostic Tools

EE Tool Box

EE Connectivity Test Command

- The Enterprise Extender connectivity test command is useful in debugging various network problems. This command can be used to test an existing Enterprise Extender connection, or it can be used to assist in diagnosing why an EE connection cannot be established.

- The EE connectivity test will verify:

    - EE line availability

    - Address resolution capability

    - EE partner reachability

        - The output generated from this request will show the reachability to the remote EE endpoint over all five UDP ports reserved for EE.

        - When MULTIPATH function is enabled in the Enterprise Extender capable TCP/IP stack, the EE connectivity test is repeated for each valid TCP/IP interface which routes EE traffic.

```
D NET,EEDIAG,TEST=YES,LIST=DETAIL,ID=ETU2HO
 ...
IST2067I EEDIAG DISPLAY ISSUED ON 07/11/07 AT 10:41:12
IST1680I LOCAL IP ADDRESS 197.51.125.1
IST1680I REMOTE IP ADDRESS 197.51.153.1
 ...
IST924I -------------------------------------------------------------
IST2133I INTFNAME: LMTU2BR55              INTFTYPE: MPCPTP
IST2134I    CONNECTIVITY SUCCESSFUL                       PORT: 12000
IST2137I      1  197.51.155.14           RTT:    1
IST2137I      2  197.51.153.1            RTT:    4
IST2134I    CONNECTIVITY SUCCESSFUL                       PORT: 12001
IST2137I      1  197.51.155.14           RTT:    2
IST2137I      2  197.51.153.1            RTT:    4
IST2134I    CONNECTIVITY SUCCESSFUL                       PORT: 12002
IST2137I      1  197.51.155.14           RTT:    2
IST2137I      2  197.51.153.1            RTT:    5
IST2134I    CONNECTIVITY SUCCESSFUL                       PORT: 12003
IST2137I      1  197.51.155.14           RTT:    2
IST2137I      2  197.51.153.1            RTT:    6
IST2134I    CONNECTIVITY SUCCESSFUL                       PORT: 12004
IST2137I      1  197.51.155.14           RTT:    3
IST2137I      2  197.51.153.1            RTT:    5
 ...
```

EE Connectivity Test Example

EE Tool Box

IPv4 EE connection is established from HostA to HostB, but HPR performance is poor

IPv4
Cloud1

Router
9.67.1.2

Router
9.67.1.4

HostA
z/OS
9.67.1.1

OSA1

OSA2

HostB
z/OS
9.67.1.6

Router
9.67.1.3
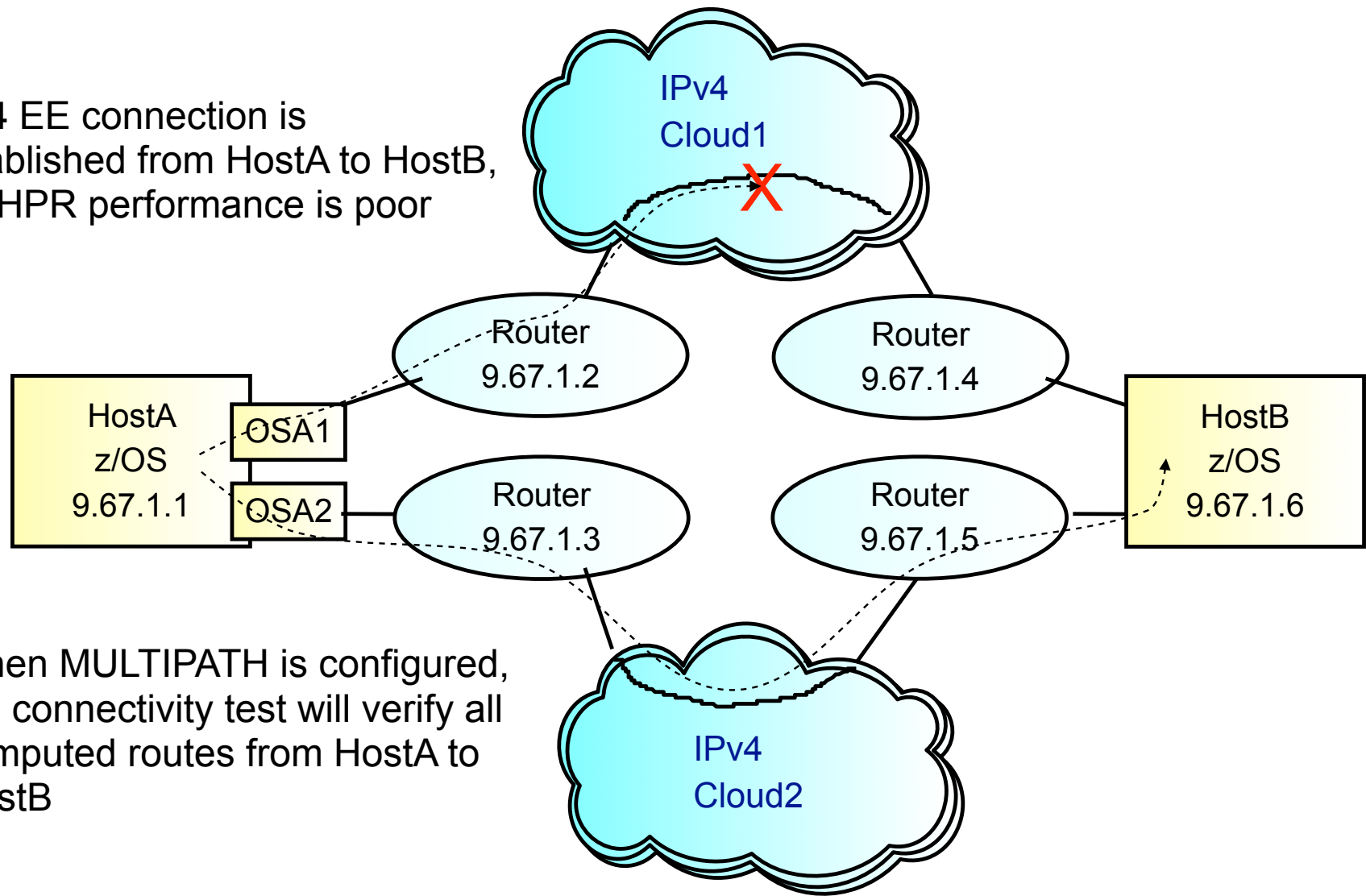
Router
9.67.1.5

When MULTIPATH is configured, the connectivity test will verify all computed routes from HostA to HostB

IPv4
Cloud2

27

EE Connectivity Test Example...

EE Tool Box

D NET,EEDIAG,TEST=YES,IPADDR=(9.67.1.1,9.67.1.6),LIST=DETAIL

```
IST097I DISPLAY ACCEPTED
IST350I DISPLAY TYPE = EEDIAG
IST2119I ENTERPRISE EXTENDER DISPLAY CORRELATOR: EE00000E
IST2067I EEDIAG DISPLAY ISSUED ON 10/04/05 AT 11:05:50
IST1680I LOCAL IP ADDRESS 9.67.1.1
IST1680I REMOTE IP ADDRESS 9.67.1.6
IST2023I CONNECTED TO LINE LN11
IST2126I CONNECTIVITY TEST IN PROGRESS
IST314I END
.

.
```

EE Connectivity Test Example...

```
IST350I DISPLAY TYPE = EEDIAG
IST2130I ENTERPRISE EXTENDER CONNECTIVITY TEST INFORMATION
IST2119I ENTERPRISE EXTENDER DISPLAY CORRELATOR: EE00000E
IST2131I EEDIAG DISPLAY COMPLETED ON 10/04/05 AT 11:05:52
IST2132I LDLC PROBE VERSIONS: VTAM = V1  PARTNER = V1
IST1680I LOCAL IP ADDRESS 9.67.1.1
IST1680I REMOTE IP ADDRESS 9.67.1.6
IST924I -------------------------------------------------------------
IST2133I INTFNAME: OSA1                  INTFTYPE: OSAFDDI
IST2135I    CONNECTIVITY UNSUCCESSFUL    SENSE: ***NA***   PORT: 12000
IST2137I     1  9.67.1.2                   RTT:     2
IST2137I     2  9.67.1.21          D-1     RTT:     3
IST2135I    CONNECTIVITY UNSUCCESSFUL    SENSE: ***NA***   PORT: 12001
IST2137I     1  9.67.1.2                   RTT:     2
IST2137I     2  9.67.1.21          D-1     RTT:     3
IST2135I    CONNECTIVITY UNSUCCESSFUL    SENSE: ***NA***   PORT: 12002
IST2137I     1  9.67.1.2                   RTT:     2
IST2137I     2  9.67.1.21          D-1     RTT:     4
IST2135I    CONNECTIVITY UNSUCCESSFUL    SENSE: ***NA***   PORT: 12003
IST2137I     1  9.67.1.2                   RTT:     2
IST2137I     2  9.67.1.21          D-1     RTT:     4
IST2135I    CONNECTIVITY UNSUCCESSFUL    SENSE: ***NA***   PORT: 12004
IST2137I     1  9.67.1.2                   RTT:     2
IST2137I     2  9.67.1.21          D-1     RTT:     3
 .
 .
```

EE Connectivity Test Example...

```
IST2137I   hop   ipv4address        flags   RTT: time
```

This message displays information gathered during the EE connectivity test over an IPv4 route using the OSA1 adaptor. Take the following IST2137I message from the previous display:

```
IST2137I    2  9.67.1.21            D-1     RTT:    3
```

2 is the TTL *hop* count used in the LDLC probe command.
9.67.1.21 is the source IPv4 address (*ipv4addr)* from the ICMP response.

In this case, the *flags* field (D-1 in this example) has a format of *t-ccc.*
Where t is a representative character of the ICMP message type returned in response to the LDLC probe and *ccc* represents the specific code associated with the ICMP message type.

The ICMP message type is displayed as one of the following:
D - "Destination Unreachable"   ICMP Type 3
P - "Parameter Problem"   ICMP Type 12
Q - "Source Quench"   ICMP Type 4

Hop 2 returned an ICMP message type 3 with a specific code of 1 in response to the EE probe. Message IST2137I indicates this by displaying the *flags* field as D-1. For a list of the ICMP types and codes, see Appendix E in the z/OS Communications Server: IP System Administrator's Commands, "ICMP/ICMPv6 types and codes".

time is the round trip time for the LDLC probe to be sent to the TTL hop, and for an ICMP response to be received. In this example, the round trip time is 3 milliseconds.

EE Connectivity Test Example...

```
IST924I  -----------------------------------------------------------
IST2133I INTFNAME: OSA2                      INTFTYPE: OSAFDDI
IST2134I    CONNECTIVITY SUCCESSFUL                     PORT: 12000
IST2137I       1 9.67.1.3                    RTT:      9
IST2137I       2 9.67.1.11                   RTT:     14
IST2137I       3 9.67.1.12                   RTT:     19
IST2137I       4 9.67.1.5                    RTT:     23
IST2137I       5 9.67.1.6                    RTT:     27
IST2134I    CONNECTIVITY SUCCESSFUL                     PORT: 12001
IST2137I       1 9.67.1.3                    RTT:      8
IST2137I       2 9.67.1.11                   RTT:     14
IST2137I       3 9.67.1.12                   RTT:     17
IST2137I       4 9.67.1.5                    RTT:     21
IST2137I       5 9.67.1.6                    RTT:     25
.
.
IST2134I    CONNECTIVITY SUCCESSFUL                     PORT: 12004
IST2137I       1 9.67.1.3                    RTT:      7
IST2137I       2 9.67.1.11                   RTT:     11
IST2137I       3 9.67.1.12                   RTT:     12
IST2137I       4 9.67.1.5                    RTT:     17
IST2137I       5 9.67.1.6                    RTT:     23
IST924I  -----------------------------------------------------------
IST2039I CONNECTIVITY TEST INFORMATION DISPLAYED FOR 2 INTERFACES
IST314I END
```

EE Connectivity Test Example...

```
IST2137I   hop   ipv4address          flags    RTT: time
```

This message displays information gathered during the EE connectivity test over an IPv4 route using the OSA2 adaptor. Take the following IST2137I message from the previous display:

```
IST2137I     5  9.67.1.6                        RTT:     27
```

5 is the TTL *hop* count used in the LDLC probe command. In this case, the EE connectivity test was successful over each of the 5 EE ports. Each IST2137I message indicating it was a 5 hop route to reach the EE partner with an IPv4 address of 9.67.1.6.

In this case, the *flags* field is blank as there were not any probe retries or any unexpected ICMP messages returned.

time is the round trip time for the LDLC probe to be sent to the TTL hop, and for the LDLC probe response (UDP datagram) to be received from the EE partner. In this example, the round trip time is 27 milliseconds.

N O T E S

EE Connectivity Test Considerations

- EE connectivity test expects ICMP messages to be returned from intermediate hops
- Firewall configurations may have limited diagnostic output
  - Firewalls may be configured to block ICMP messages
    - Intermediate hops past firewall will appear as unresponsive, but final destination reachability can still be determined (see next foil) assuming destination has EE Test probe "responder" support.
    - Minimally configure firewall to allow ICMP "Time Exceeded"
      - IPv4: ICMP Message Type 11
      - IPv6: ICMPv6 Message Type 3
  - Firewalls may be configured to block UDP traffic
    - Firewalls must allow UDP traffic for EE UDP ports 12000-12004
    - Test probe response may route through firewall
      - Probe response is not an ICMP message. It is a UDP datagram.
- EE Test probe "responder" support
  - APAR PK17858 is available for z/OS Communications Server V1R4, V1R5, V1R6, & V1R7
  - Support available for CS/Windows, CS/AIX, CS/Linux, PComm, & Cisco SNASw
- EE Test requires an available line for each connection to be tested.
  - May need to increase number of defined lines to accommodate the number of concurrent tests
- MAXEETST start option
  - Default is 500
  - May want to limit number of concurrent tests due to line availability

EE Tool Box

EE Connectivity Test w/Firewalls Blocking ICMP

z/OS Host A — Router — Firewall — Router — Firewall — Router — z/OS Host B

1 Probe(TTL=1) ICMP ICMP

2 ICMP Time Exceeded

3 Probe(TTL=2)

4 ICMP Time Exceeded

5 *** Timeout/Retries ***

6 Probe(TTL=3)

7 ICMP Time Exceeded

8 *** Timeout/Retries ***

9 Probe(TTL=4)

10 Probe Reply (UDP)
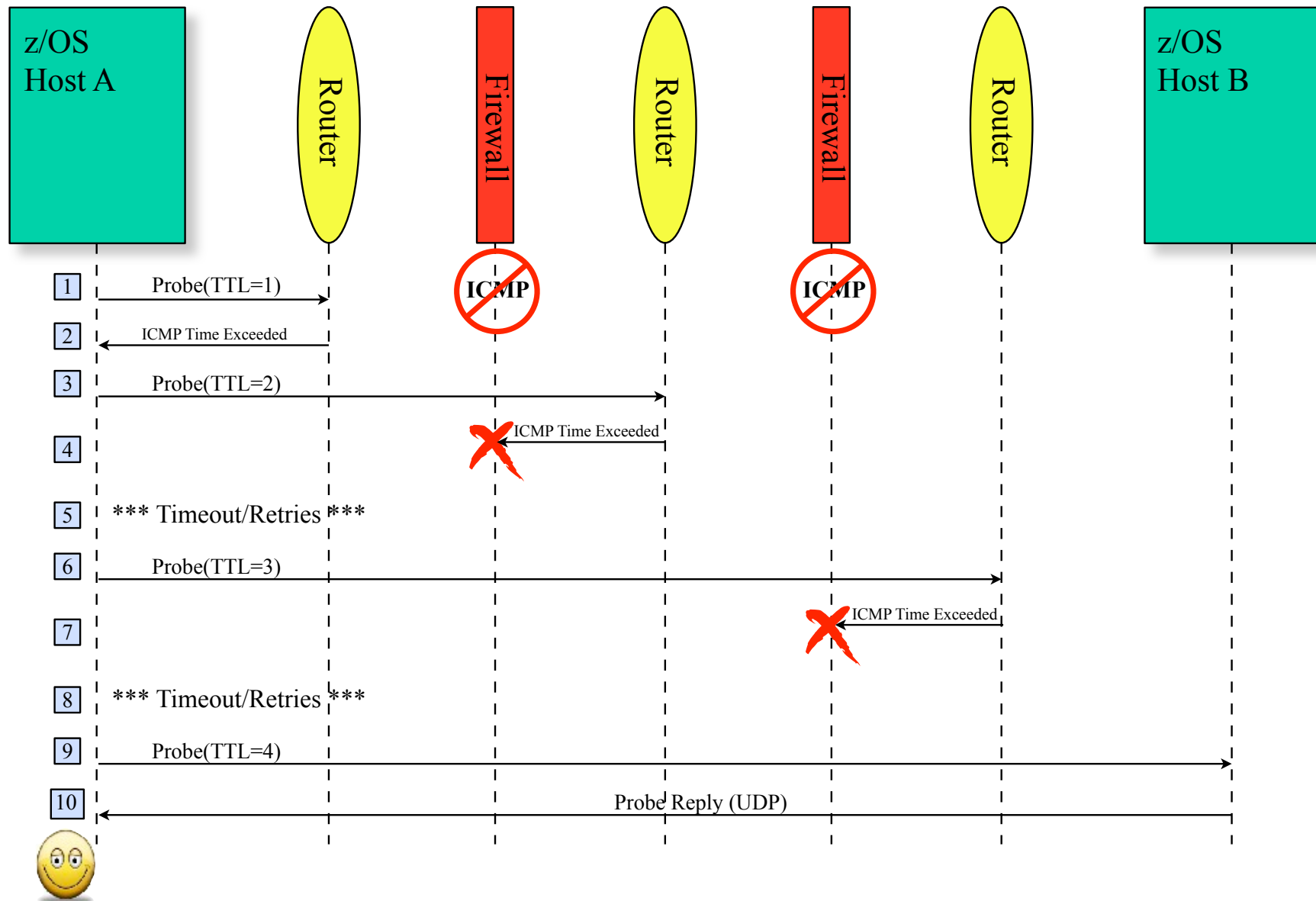
EE Connectivity Test w/Firewalls Blocking ICMP...

The previous chart illustrates what happens when Host A issues an EE connectivity test command with Host B as the partner, but there are intervening firewalls that do not allow the passage of ICMP messages.

(1) Host A begins trying to verify partner reachability by sending an LDLC Probe command inside of a UDP packet. (The same UDP packet would be sent on all 5 EE ports.) For the first flow, the Time-To-Live (TTL) value is set to one.

(2) When the UDP packet reaches the first router in the path, the router decrements the TTL value to zero. Since the TTL is zero, the packet is discarded, and an ICMP "Time Exceeded" message is returned.

(3) Upon receiving the the ICMP response, Host A sends the probe again, but with the TTL value incremented to two.

(4) When the probe reaches the second router on the path, that router decrements TTL to zero, and again the packet is discarded and an ICMP "Time Exceeded" response generated.   However, in this case a firewall between that router and Host A is blocking ICMP, and discards the message.

(5) After three seconds, Host A will timeout and resend the probe, still with a TTL of two. Again, the ICMP response will be discarded. After a second three-second timeout, Host A will retry one last time, and will again timeout after three more seconds.

(6) After the three attempts to send the probe with a TTL of two (and three timeouts spanning a total of nine seconds), Host A will resend the probe with a TTL value of three.

(7) When the probe reaches the third router on the path, that router decrements TTL to zero, and again the packet is discarded and an ICMP "Time Exceeded" response generated.   Again a firewall between that router and Host A is blocking ICMP, and discards the message.

(8) Once again, Host A goes through a sequence of three-second timeouts and probe retries.

(9) After exhausting its retries with a TTL value of three, Host A sends the probe with a TTL value of four. In this case, the probe makes it all the way to Host B.

(10) Host B does not respond with an ICMP message, but instead sends an LDLC Probe Response message as a UDP datagram. This packet makes it back to Host A, verifying that Host B is a reachable EE partner.

The MAXTIME operand on the EE connectivity test command is used to specify how long VTAM will spend performing the connectivity test before terminating the test, and displaying the available information. The default value of MAXTIME is 60 seconds.

Path MTU Discovery for EE (V1R10)

- z/OS V1R8 CS (and prior):
  - MTU size determined during connection establishment
    - not altered for the duration of the connection.
    - Fragmentation may occur at host and in network

- z/OS V1R9 CS: Local MTU Discovery for EE
  - Local MTU changes are detected
  - Avoids fragmentation at the host, but not in the network

- z/OS V1R10 CS: Path MTU Discovery (PMTU) for EE
  - IPv4: Stacks set "don't fragment" (DF) bit in all outbound IPv4 EE packets
    - IPv6 does not support fragmentation in the network
  - Stack monitors for ICMP/ICMPv6 "Packet too big" messages
  - Stack updates VTAM with learned path MTU
  - Local RTP pipes segment to new size

- PMTU originally architected as TCP-based solution and did not apply to UDP
  - PMTU for EE is an EE-specific adaptation of PMTU, and does not apply to UDP in general

z/OS
Host A
MTU - 1500

Router A
MTU - 1500

Firewall

Firewall

Router B
MTU - 1200

z/OS
Host B
MTU - 1500

ICMP(1200)
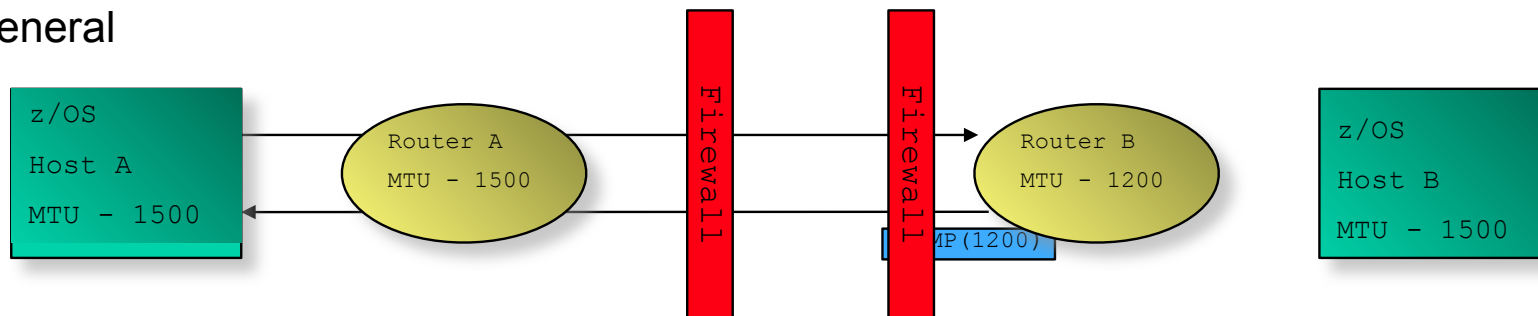
Path MTU Discovery for EE...

- Path MTU Discovery requires ICMP messages to be returned
  - Some firewalls are configured to block ICMP
  - Minimally configure firewalls to permit the following ICMP messages for ports 12000-12004:
    - IPv4: ICMP Message Type 3 - Destination Unreachable
    - IPv6: ICMPv6 Message Type 2 - Packet Too Big
  - Do NOT enable PMTU if ICMP messages are not permitted through the firewalls

```
z/OS            Router A        Firewall  Firewall  Router B        z/OS
Host A          MTU - 1500                          MTU - 1200      Host B
MTU - 1500                                 ICMP(1200)               MTU - 1500
```
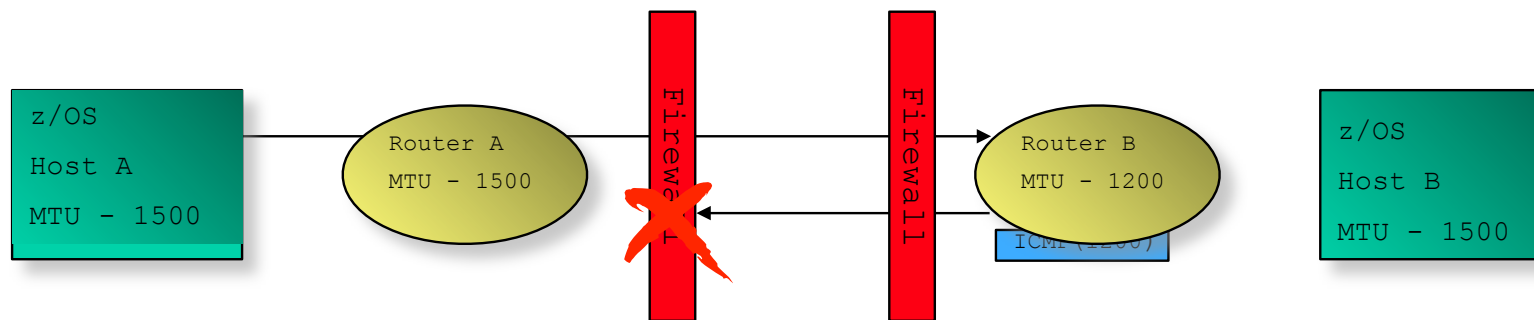
- For situations where enabling PMTU for EE is not viable, or is not an adequate solution, V1R10 provides a new operand to allow static configuration of the MTU size to be used for the EE connection:
  - MTU = *mtu_size*
  - Specifiable on EE switched PU, EE XCA GROUP (connection network), and EE model PU

Improved Recovery from RTP Pipe Stalls (V1R12)

- With the Path MTU Discovery function for EE, MTU size changes over an active EE link are only communicated to the two nodes that act as the endpoint of that EE link (NN2 and NN3 below)
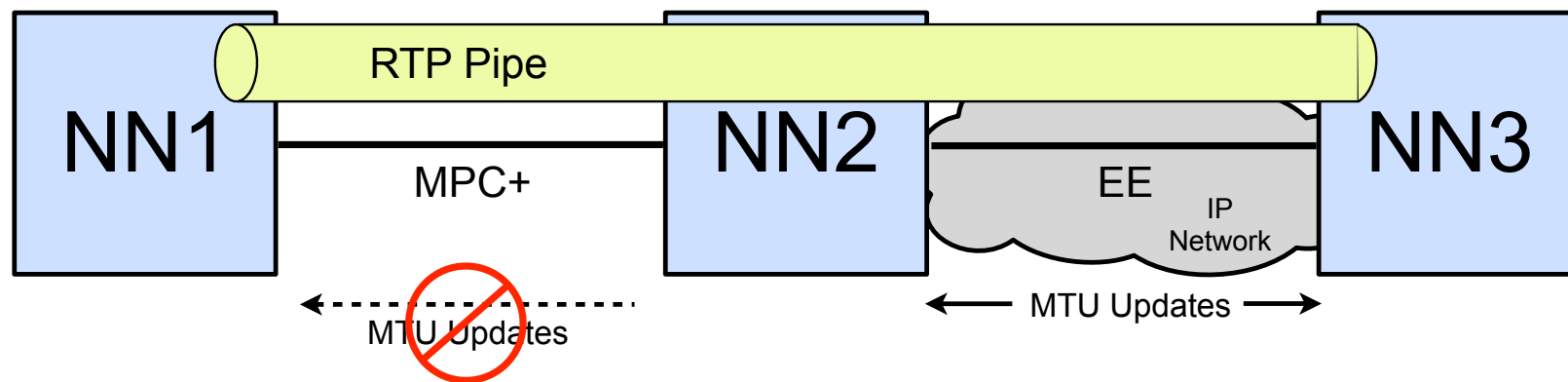
- If an existing RTP pipe begins on a node other than the EE link endpoint, it will not learn the PMTU-discovered MTU size, and will continue to send packets at a non-optimal size, potentially resulting in packet loss and transmission stalls.
- z/OS V1R12 adds logic for VTAM to drive the path switch logic if multiple retransmissions occur (stall detection)
  - Thereby letting NN1 above learn the new current MTU size and adapt to it

```
IST2335I PATH SWITCH REASON: XMIT STALL RECOVERY
```

HPR Path Switch Summarization

- HPR Path Switch Summarization reduces the number of path switch message groups VTAM issues across a 60-second interval

- HPRPSMSG=<u>ALL</u>|count, where count is in the range 10-100

- The HPRPSMSG value specifies the number of IST1494I ("Path Switch Started") messages that will be issued, before supressing the remaining ones across the 60-second interval.

  - If the STARTED message is issued for a pipe, the associated COMPLETED or FAILED message will always be issued as well

- At the end of the 60-second interval, a summarization report is issued on total path switch activity during the interval

  - The report output is limited to 10 Net IDs and 50 CPs, but the "started", "completed", and "failed" counts will accurately reflect all path switch activity

- If HPRPSMSG=ALL is specified, then all path switch message groups will be issued and no summarization provided

- If you have previously specified IST1494I in the message flooding prevention table, you will probably want to remove it when enabling path switch summarization

HPR Path Switch Summarization...

```
IST2191I HPR PATH SWITCH SUMMARY FROM 04/05/06 AT 09:45:14
IST924I -----------------------------------------------------------
IST2192I STARTED   =     12
IST2193I   TGINOP  =     12   SRQTIMER =      0   PSRETRY    =      0
IST2194I   PARTNER =      0   MNPS     =      0   UNAVAILABLE =     0
IST2195I   NETWORK =      3 HIGH =       3 MEDIUM =      3 LOW =      3
IST924I -----------------------------------------------------------
IST2196I COMPLETED =      8
IST2195I   NETWORK =      2 HIGH =       2 MEDIUM =      2 LOW =      2
IST924I -----------------------------------------------------------
IST2197I FAILED    =      4
IST2195I   NETWORK =      1 HIGH =       1 MEDIUM =      1 LOW =      1
IST924I -----------------------------------------------------------
IST2198I NETID          STARTED         COMPLETED        FAILED
IST2199I   CPNAME   NET  HI MED LOW  NET  HI MED LOW  NET  HI MED LOW
IST2205I --------- -------------- -------------- --------------
IST2200I NETA       2   2   2   2    1   1   1   1    1   1   1   1
IST2201I   SSCP2A   1   1   1   1    1   1   1   1    0   0   0   0
IST2201I   SSCP7A   1   1   1   1    0   0   0   0    1   1   1   1
IST2205I --------- -------------- -------------- --------------
IST2200I NETB       1   1   1   1    1   1   1   1    0   0   0   0
IST2201I   SSCP99   1   1   1   1    1   1   1   1    0   0   0   0
IST924I -----------------------------------------------------------
IST2206I 24 PATH SWITCH EVENTS FOR 3 CPS IN 2 NETIDS
IST314I END
```

HPR Path Switch Delay (V1R11)

- If the RTP endpoint suspects a problem with partner communications, it will make several attempts to contact the partner, with a delay between attempts based on a "short request" (SRQ) timer value based on the round-trip time.

```
...
IST1818 PATH SWITCH REASON: SHORT REQUEST RETRY LIMIT EXHAUSTED
...
```

- At times this logic is too sensitive:
  - Transient network or partner conditions can cause temporary swings in round-trip time that can cause unnecessary entry into path switch state
  - In this case, the pipe usually path switches right back onto the same route
  - This wastes cycles and clutters the console with path switch messages

- V1R11 introduces new controls to specify a minimum time period that will be required before entering path switch state
  - Specifies the minimum amount of time a z/OS CS RTP endpoint must wait before initiating a path switch attempt due to an unresponsive partner
  - Does not control path switches initiated due to the PSRETRY function, the MODIFY RTP command, or local TG inops
  - This only affects the path switch logic on the local end of the RTP pipe. The path switch delay value is not negotiated with the RTP partner

HPR Path Switch Delay...

- New start option: HPRPSDLY = 0 | *ps-delay*
  - Specifying a non-zero value for this start option sets the minimum path switch delay value
  - Range: 0 - 240 seconds
  - Default value of zero indicates that the RTP pipe should enter path switch as soon as a predetermined number of retry attempts have been unsuccessful (prior behavior)
  - Can be modified via MODIFY VTAMOPTS

- HPRPSDLY = value-of-HPRPSDLY-start-option | *ps-delay* | EEDELAY   can also be specified:
  - On an EE PU in a switched major node
  - On a connection network GROUP in the EE XCA major node
  - On the EE model PU in a model major node
  - If you set HPRPSDLY to EEDELAY, VTAM will calculate a delay value long enough to allow the EE LDLC mechanism to inop the EE connection in the event the EE partner becomes unreachable
    - Local EE inop will trigger path switch processing

```
    D NET,ID=CNR00004,HPRDIAG=YES
    IST097I DISPLAY ACCEPTED
    IST075I NAME = CNR00004, TYPE = PU_T2.1
     .

     .
    IST924I -------------------------
    IST1984I PATH SWITCH INFORMATION:
    IST2271I PATH SWITCH DELAY = 90
    IST2272I PATH SWITCH DELAYED UNTIL 11/17/08 AT 12:07:34
     .
```

EE Health Verification (V1R12)

- The EE Health Verification function will verify the health of a potential EE connection by sending a probe to the remote partner using all five ports during the connection activation
  - VTAM does not activate the EE connection if the remote partner is not reachable on all ports
  - However, if the remote partner does not support the probe, VTAM will still bring up the EE connection
- EE Health Verification will also optionally verify the health of an active EE connection by sending a probe to the remote partner on all five ports at a user-specified interval
  - VTAM issues a warning message if the remote partner is not reachable on all ports, but will keep the connection active

- EE Health Verification is enabled by the EEVERIFY start option:

- Or the EEVERIFY GROUP/PU parameter:

```
          ___EEVERIFY=_ ACTIVATE _____
 >>__|_____|_____><
      |_EEVERIFY=___ _NEVER_____|
                     |_ACTIVATE_____|
                     |_time_interval_value_|
```

```
 >>_____ _____ _____><
          |_EEVERIFY=___ _NEVER_____|
                         |_ACTIVATE_____|
                         |_time_interval_value_|
```

- The time_interval_value can range from 1-1440 minutes

EE Health Verification...

- When EE Health verification fails for an active EE connection, VTAM issues highlighted warning message IST2323E if it is not already present:

> **IST2323E EE HEALTH VERIFICATION FAILED ON ONE OR MORE CONNECTIONS**

  - This message stays on the console until the condition is cleared or the message is erased by the operator

- DISPLAY EE,LIST=VERIFY is used to determine which connection(s) have failed verification:

```
    d net,ee,list=eeverify
    IST097I DISPLAY ACCEPTED
    IST350I DISPLAY TYPE = EE
    IST2000I ENTERPRISE EXTENDER GENERAL INFORMATION
    IST1685I TCP/IP JOB NAME = TCPCS
    IST2003I ENTERPRISE EXTENDER XCA MAJOR NODE NAME = XCAIP
      ...
    IST924I -------------------------------------------------------------
    IST2324I EE HEALTH VERIFICATION: FAILED CONNECTION INFORMATION
    IST2325I LINE LNIP1 PU SWIP2A1 ON 12/21/09 AT 15:56:39
    IST2326I EE HEALTH VERIFICATION TOTAL CONNECTION FAILURES = 1
    IST2017I TOTAL RTP PIPES =            1      LU-LU SESSIONS =        2
    IST2018I TOTAL ACTIVE PREDEFINED EE CONNECTIONS          =         1
    IST2019I TOTAL ACTIVE LOCAL  VRN EE CONNECTIONS          =         0
    IST2020I TOTAL ACTIVE GLOBAL VRN EE CONNECTIONS          =         0
    IST2021I TOTAL ACTIVE EE CONNECTIONS                     =         1
    IST314I END
```

EE Health Verification...

- Using DISPLAY EE to display an individual EE connection will also provide information on the success or failure of EE Health Verification:

```
d net,ee,id=SWIP2A1
IST097I DISPLAY ACCEPTED
IST350I DISPLAY TYPE = EE
IST2001I ENTERPRISE EXTENDER CONNECTION INFORMATION
IST075I NAME = SWIP2A1, TYPE = PU_T2.1
IST1680I LOCAL IP ADDRESS 9.67.1.1
IST1680I REMOTE IP ADDRESS 9.67.1.2
IST2022I EE CONNECTION ACTIVATED ON 12/21/09 AT 16:21:57
IST2114I LIVTIME:    INITIAL =    10   MAXIMUM =     0   CURRENT =    10
IST2023I CONNECTED TO LINE LNIP1
IST2327I EE HEALTH VERIFICATION OPTION - EEVERIFY = 2 MINUTES
IST2329I EE HEALTH VERIFICATION SUCCESSFUL ON 12/21/09 AT 16:37:21
IST2341I EE HEALTH VERIFICATION HAS NEVER FAILED FOR THIS CONNECTION
IST2025I LDLC SIGNALS RETRANSMITTED AT LEAST ONE TIME    =         0
IST2026I LDLC SIGNALS RETRANSMITTED SRQRETRY TIMES       =         0
  ...
IST314I END
```

- If a connection is failing EE Health Verification, the DISPLAY EEDIAG,TEST=YES command can be used to further diagnose the cause of the failure.

For More Information....

URL	Content
http://www.twitter.com/IBM_Commserver	IBM Communications Server Twitter Feed
http://www.facebook.com/IBMCommserver	IBM Communications Server Facebook Fan Page
http://www.ibm.com/systems/z/	IBM System z
http://www.ibm.com/systems/z/hardware/networking/index.html	IBM System z Networking
http://www.ibm.com/software/network/commserver/zos/	IBM z/OS Communications Server
http://www.ibm.com/software/network/commserver/z_lin/	IBM Communications Server for Linux on zSeries
http://www.ibm.com/software/network/ccl/	IBM Communication Controller for Linux on System z
http://www.ibm.com/software/network/commserver/library	IBM Communications Server Library - white papers, product documentation, etc.
http://www.redbooks.ibm.com	IBM Redbooks
http://www.ibm.com/software/network/commserver/support	IBM Communications Server Technical Support
http://www.ibm.com/support/techdocs/	Technical Support Documentation (techdocs, flashes, presentations, white papers, etc.)
http://www.rfc-editor.org/rfcsearch.html	Request For Comments (RFCs)
http://publib.boulder.ibm.com/infocenter/ieduasst/stgv1r0/index.jsp	IBM Education Assistant

- Recommended White Paper:
  - "Securing an SNA Environment for the 21st Century" - http://www-01.ibm.com/support/docview.wss?rs=852&uid=swg27013237
- Recommended Redbooks:
  - SG24-7359-00 Enterprise Extender Implementation Guide
  - SG24-7334-00 A Structured Approach to Modernizing the SNA Environment
  - SG24-5957-00 Migrating Subarea to an IP Infrastructure