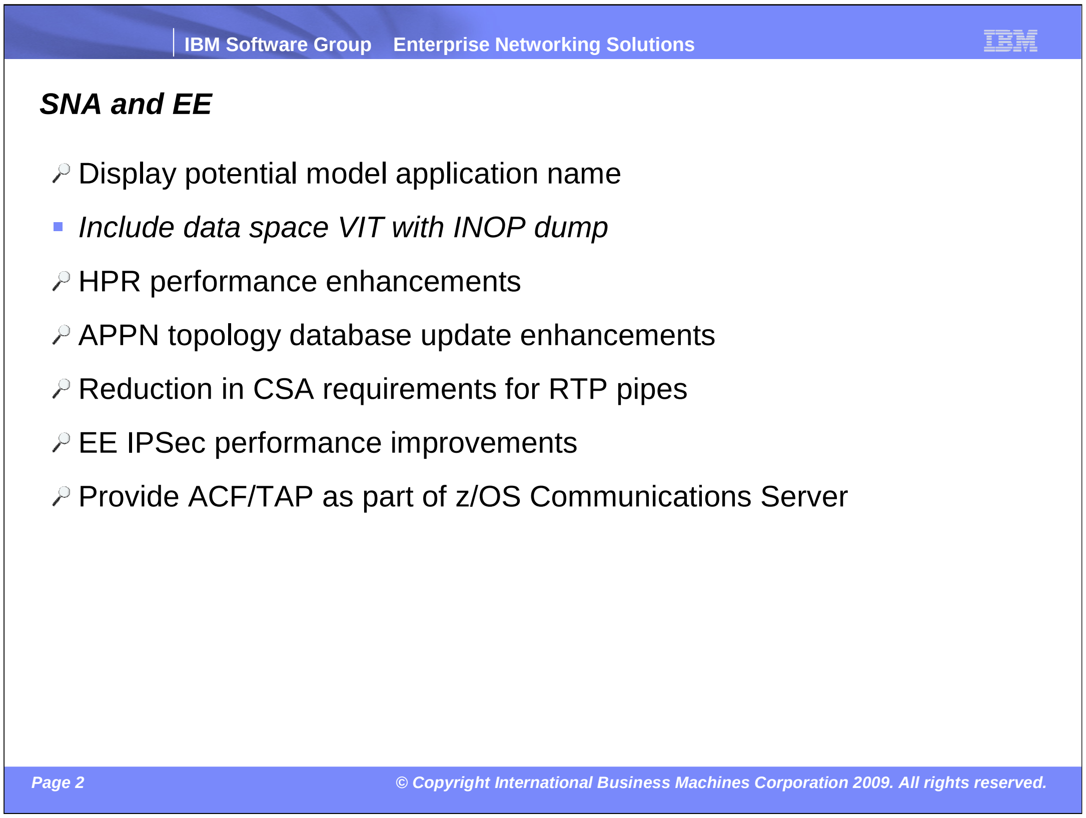
IBM Software Group Enterprise Networking Solutions
z/OS® V1R11 Communications Server

z/OS V1R11 Communications Server – SNA and EE

z/OS Communications Server Development, Raleigh, North Carolina

This presentation will give you an overview of the enhancements to the Communications Server in z/OS V1R11 for the SNA and enterprise extender (EE) theme which includes enhancements to the SNA component of the z/OS Communications Server.



IBM Software Group Enterprise Networking Solutions

SNA and EE

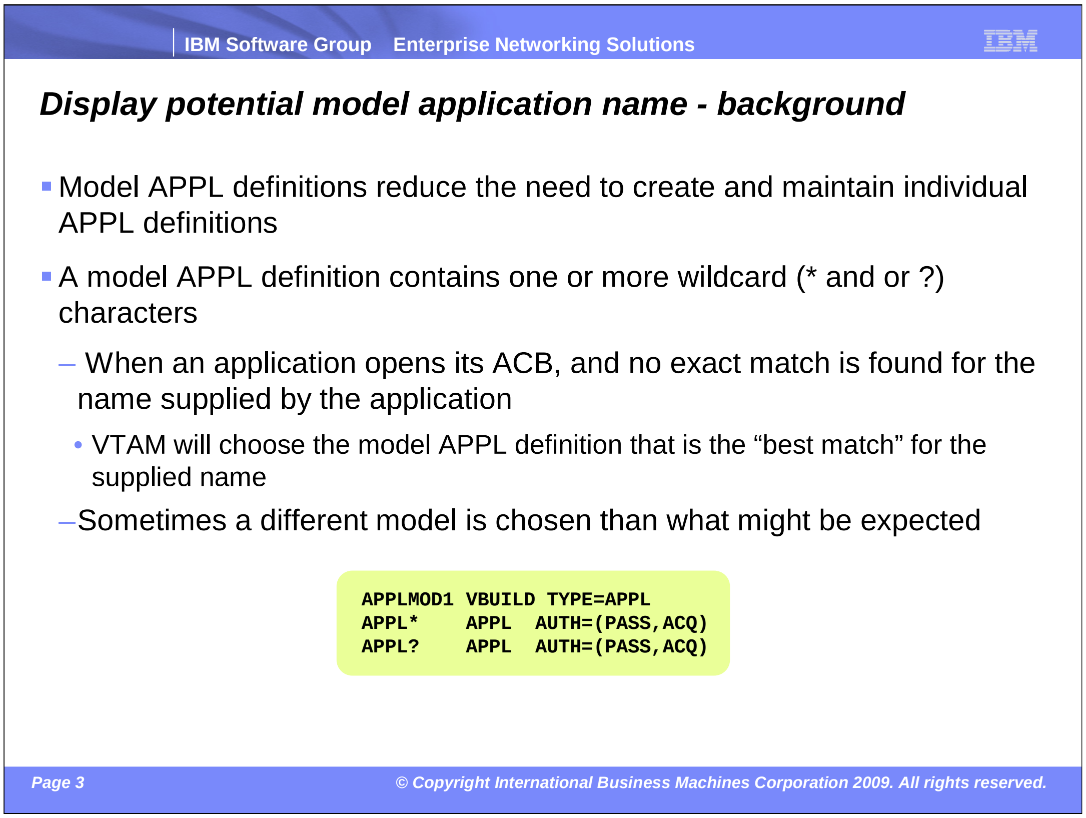
- Display potential model application name
- *Include data space VIT with INOP dump*
- HPR performance enhancements
- APPN topology database update enhancements
- Reduction in CSA requirements for RTP pipes
- EE IPSec performance improvements
- Provide ACF/TAP as part of z/OS Communications Server

As always, there are enhancements to **SNA and EE**. The main enhancement in this release is a new high performance routing (HPR) adaptive rate-based (ARB) protocol that addresses performance issues seen in environments where distributed platforms are implemented in virtualized systems.

z/OS V1R11 Communications Server enhances the DISPLAY MODELS command to identify which application model definition is used to build a dynamic application definition. Customers can use this capability to prevent problems with dynamic application definitions being built incorrectly.

VTAM® INOP Dump processing is enhanced to automatically capture the VIT data space (ISTITDS1) in the dump when the VIT data space is in use.

The presentation will briefly discuss the remaining items in this list on slides that follow.



Display potential model application name - background

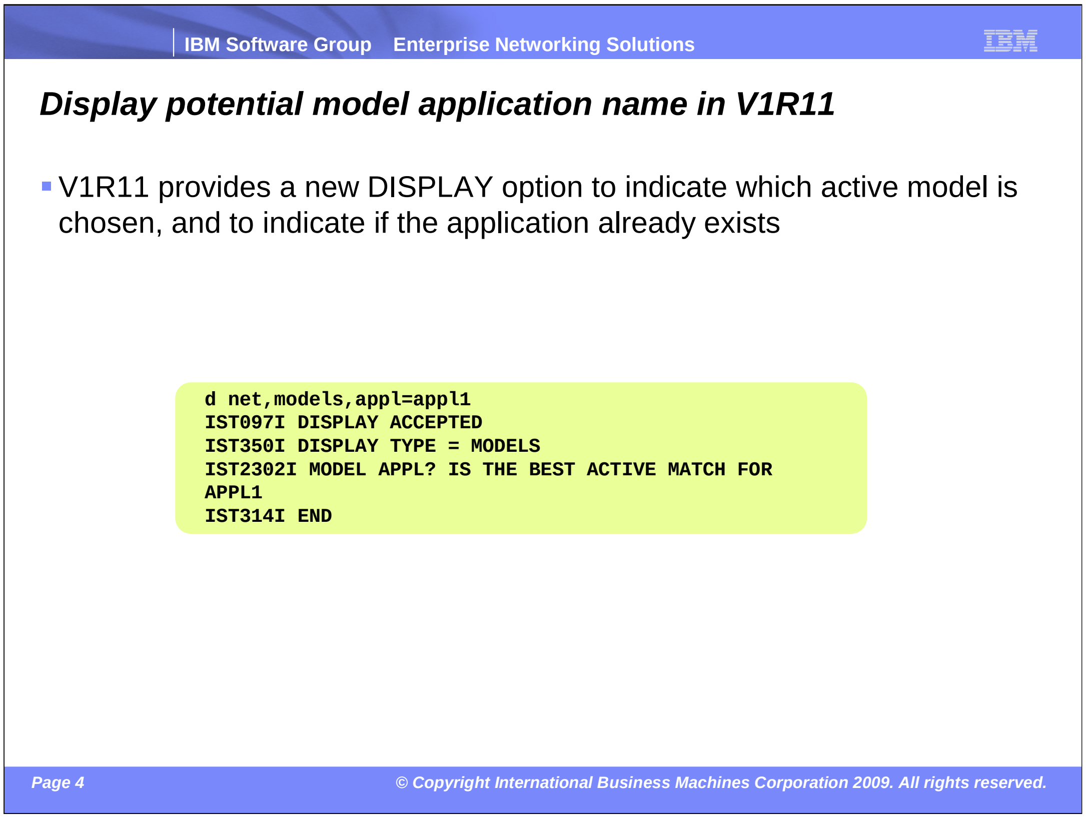
- Model APPL definitions reduce the need to create and maintain individual APPL definitions
- A model APPL definition contains one or more wildcard (* and or ?) characters
  - When an application opens its ACB, and no exact match is found for the name supplied by the application
    - VTAM will choose the model APPL definition that is the “best match” for the supplied name
  - Sometimes a different model is chosen than what might be expected

```
APPLMOD1 VBUILD TYPE=APPL
APPL*    APPL  AUTH=(PASS,ACQ)
APPL?    APPL  AUTH=(PASS,ACQ)
```

Model APPL definitions were created to reduce the need for creating and maintaining many APPL definitions. One model APPL definition can be used to generate thousands of dynamic definitions for applications

The model APPL definition is defined within the APPL major node. The definition name contains one or more wildcard characters in it. The ‘?’ wildcard character represents exactly one character. The ‘*’ wildcard character represents 0 or more characters.

When an application issues OPEN ACB, VTAM first checks to see if there is an APPL statement defined that is an exact match. If not, VTAM will try to find the model APPL definition statement that is the best match for the name supplied by the application. The search for the best name is a character by character comparison, from left to right. An actual character match is considered to be a better match than a wildcard and the single character wildcard (‘?’) is considered to be a better match than the multi-character wildcard (‘*’). When a match is found, a dynamic application RDTE is created with the model APPL definition statement.

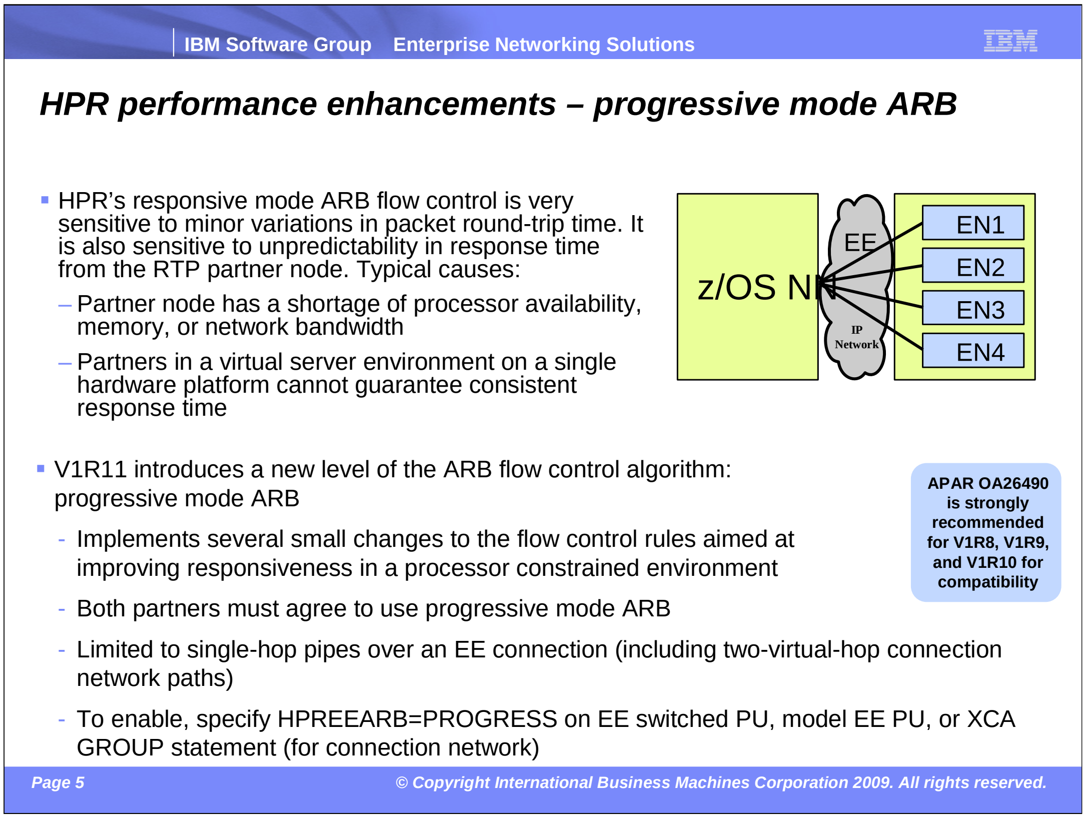


Display potential model application name in V1R11

- V1R11 provides a new DISPLAY option to indicate which active model is chosen, and to indicate if the application already exists

```
d net,models,appl=appl1
IST097I DISPLAY ACCEPTED
IST350I DISPLAY TYPE = MODELS
IST2302I MODEL APPL? IS THE BEST ACTIVE MATCH FOR
APPL1
IST314I END
```

This release is adding a new display that allows you to specify an application name and find out what model APPL statement it is going to use. The display also shows if the name is already in use so you can determine if there is a potential name conflict.


z/OS NN
EE
IP Network
EN1
EN2
EN3
EN4

HPR performance enhancements – progressive mode ARB

- HPR's responsive mode ARB flow control is very sensitive to minor variations in packet round-trip time. It is also sensitive to unpredictability in response time from the RTP partner node. Typical causes:
  - Partner node has a shortage of processor availability, memory, or network bandwidth
  - Partners in a virtual server environment on a single hardware platform cannot guarantee consistent response time

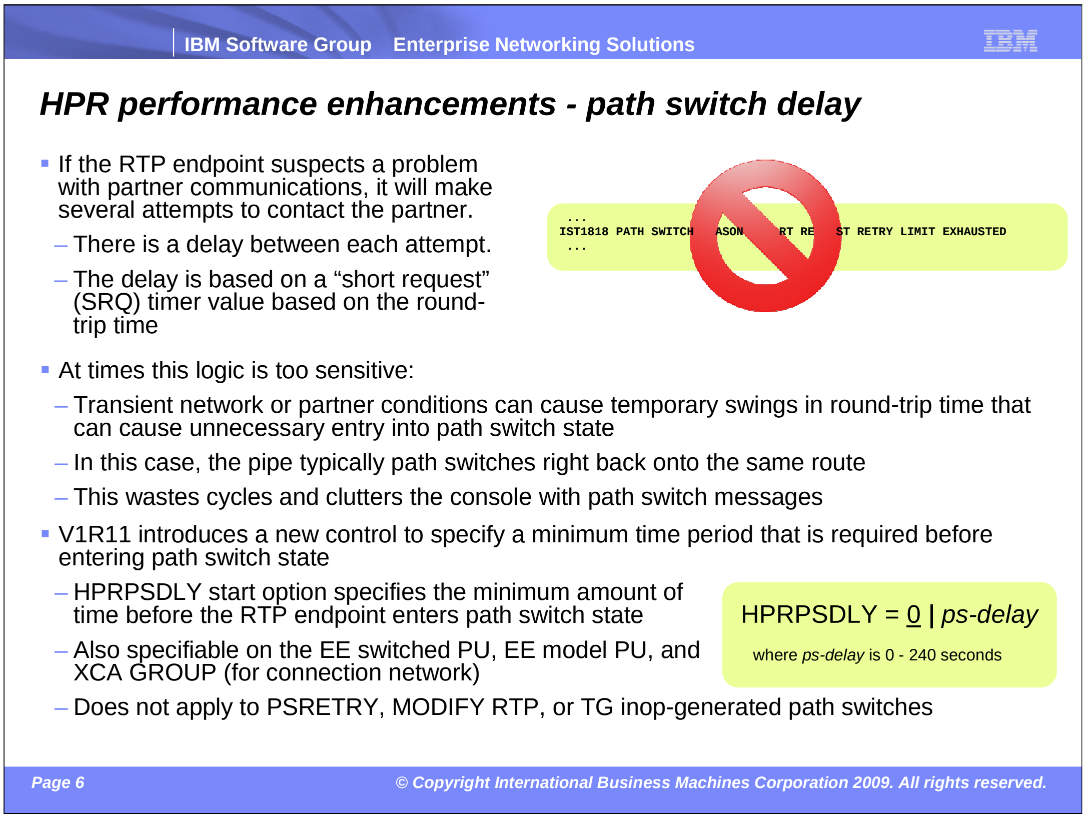
- V1R11 introduces a new level of the ARB flow control algorithm: progressive mode ARB
  - Implements several small changes to the flow control rules aimed at improving responsiveness in a processor constrained environment
  - Both partners must agree to use progressive mode ARB
  - Limited to single-hop pipes over an EE connection (including two-virtual-hop connection network paths)
  - To enable, specify HPREEARB=PROGRESS on EE switched PU, model EE PU, or XCA GROUP statement (for connection network)

APAR OA26490 is strongly recommended for V1R8, V1R9, and V1R10 for compatibility

A new Adaptive Rate-Based (ARB) algorithm is implemented by High Performance Routing (HPR) when used over EE connections. The new algorithm has been designed to improve performance in specific environments, such as distributed virtualized server environments. In such environments, the existing ARB algorithm has proven to be too aggressive in terms of lowering throughput when the EE node is processor constrained.

Use of the new ARB algorithm requires that both EE end points support it. Work has been done with distributed EE vendors and support by various distributed platforms is expected to be in place at the time of z/OS V1R11 GA.

There is an APAR for z/OS V1R8 to V1R10. The APAR is needed in this situation. When a down-level VTAM sees the unrecognized ARB level (progressive mode ARB), VTAM compares it against the responsive mode setting. Since it does not match, VTAM drops the ARB level down to Base Mode. So, the pipe will come up, but it will operate under base mode ARB rules, which is certainly not ideal. The APAR changes the down level nodes to just drop back to responsive mode ARB when they see the new ARB level.



IBM Software Group   Enterprise Networking Solutions

HPR performance enhancements - path switch delay

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

```
 ...
IST1818 PATH SWITCH   ASON     RT RE   ST 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 typically path switches right back onto the same route
  - This wastes cycles and clutters the console with path switch messages
- V1R11 introduces a new control to specify a minimum time period that is required before entering path switch state
  - HPRPSDLY start option specifies the minimum amount of time before the RTP endpoint enters path switch state
  - Also specifiable on the EE switched PU, EE model PU, and XCA GROUP (for connection network)
  - Does not apply to PSRETRY, MODIFY RTP, or TG inop-generated path switches

HPRPSDLY = 0 | *ps-delay*

where *ps-delay* is 0 - 240 seconds

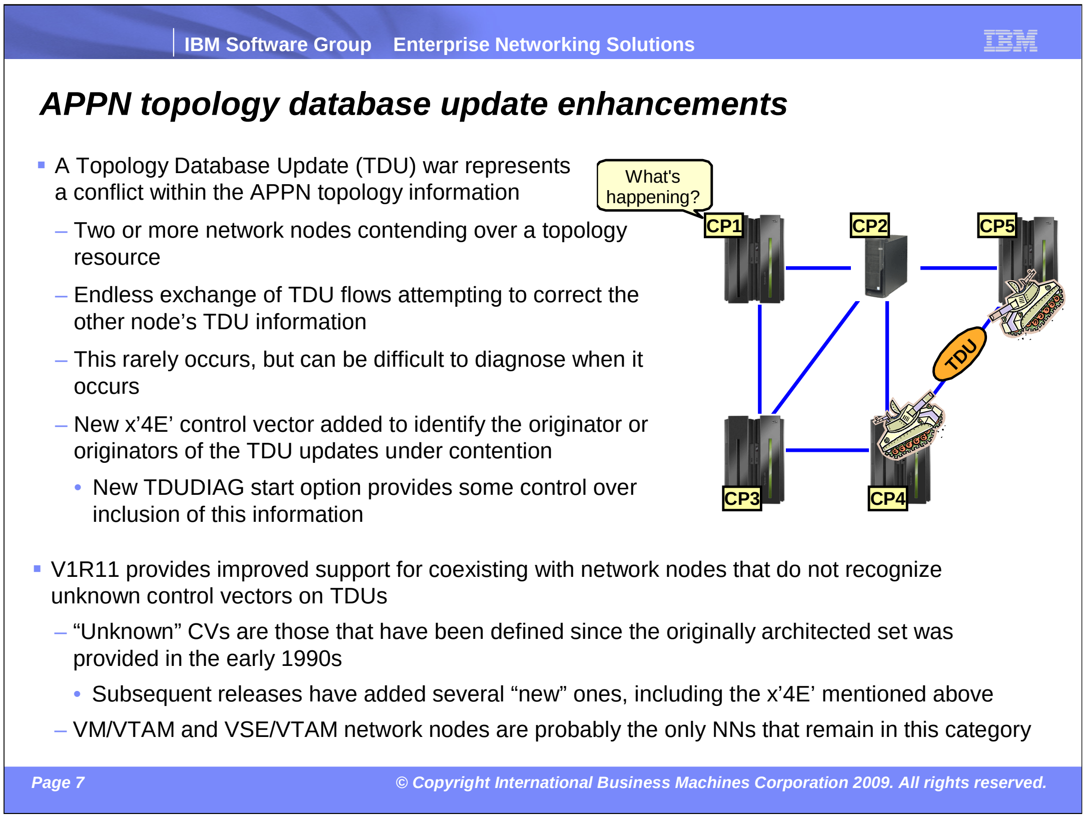
Page 6

One reason an RTP endpoint enters path switch is for an unresponsive partner. These messages are issued after a predetermined number of attempts to contact the partner fail. Each of these attempts is timed out before initiating another retry. The "Short Request Timer" (SRQ) value is the delay between each attempt. This SRQ timer is derived from a formula that relies on the round trip time. In an effort to improve reliability, the HPR path switch logic at times is too reactive to minor variations in round-trip time. Transient network conditions can cause a path switch.

To address these problems, new controls are available to allow you to specify an HPR path switch delay value. This delay will allow for a minimum amount of time that a z/OS Communications Server RTP endpoint must wait before initiating path switch due to an unresponsive partner.

This solution is available for all RTP pipes originating in this z/OS Communications Server node. This solution only controls the path switch logic on this side of the RTP pipe. The path switch delay value is not negotiated with the RTP partner. It is not required to have both RTP endpoints running with the HPR path switch delay function.

A new start option has been introduced to allow you to control the minimum amount of time a z/OS Communications Server RTP endpoint must wait before initiating a path switch due to an unresponsive partner. VTAM also provides the HPRPSDLY parameter on three major nodes. This HPRPSDLY parameter can be specified on the switched PU for predefined connections, and on the XCA GROUP for connection network-based PUs.


IBM Software Group Enterprise Networking Solutions

APPN topology database update enhancements

- A Topology Database Update (TDU) war represents a conflict within the APPN topology information
  - Two or more network nodes contending over a topology resource
  - Endless exchange of TDU flows attempting to correct the other node's TDU information
  - This rarely occurs, but can be difficult to diagnose when it occurs
  - New x'4E' control vector added to identify the originator or originators of the TDU updates under contention
    - New TDUDIAG start option provides some control over inclusion of this information

- V1R11 provides improved support for coexisting with network nodes that do not recognize unknown control vectors on TDUs
  - "Unknown" CVs are those that have been defined since the originally architected set was provided in the early 1990s
    - Subsequent releases have added several "new" ones, including the x'4E' mentioned above
  - VM/VTAM and VSE/VTAM network nodes are probably the only NNs that remain in this category

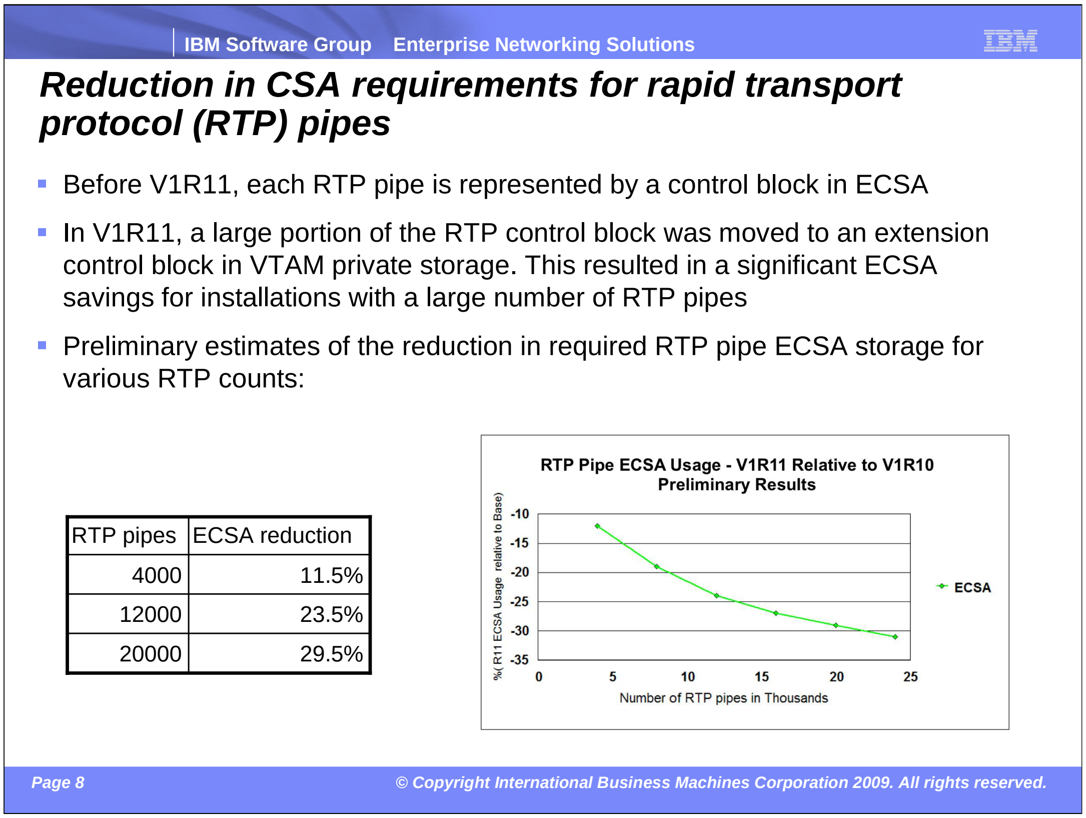
Page 7

TDU Wars were initially a rare occurrence, mostly due to configuration error (duplicate CP names for instance). Recent APPN enhancements increased the possibility of TDU wars, such as the z/OS V1R6 Communications Server connection network reachability awareness.

TDU flows between nodes are controlled by Resource Sequence Numbers (RSNs). The highest RSN value represents the most current topology information for a resource. A node undergoing performance degradation might be an innocent victim, not the perpetrator. It is difficult to determine what APPN node set the RSN value. The TDU only indicates the adjacent partner forwarding information, not necessarily the one setting it.

z/OS V1R11 Communications Server adds a new control vector to TDU exchanges that identify the node that set the RSN. Setting the vector and how often is under configuration control (by means of a new VTAM start option).

Any topology control vectors added since the original APPN architecture are considered to be unknown vectors. Originally, unknown vectors were not included in TDUs sent to any partner nodes if just one partner node did not support unknown vectors. This is changed so VTAM includes unknown vectors in TDUs sent to partner nodes that support unknown vectors, but does not include unknown vectors in TDUs sent to partner nodes that do not have this support.



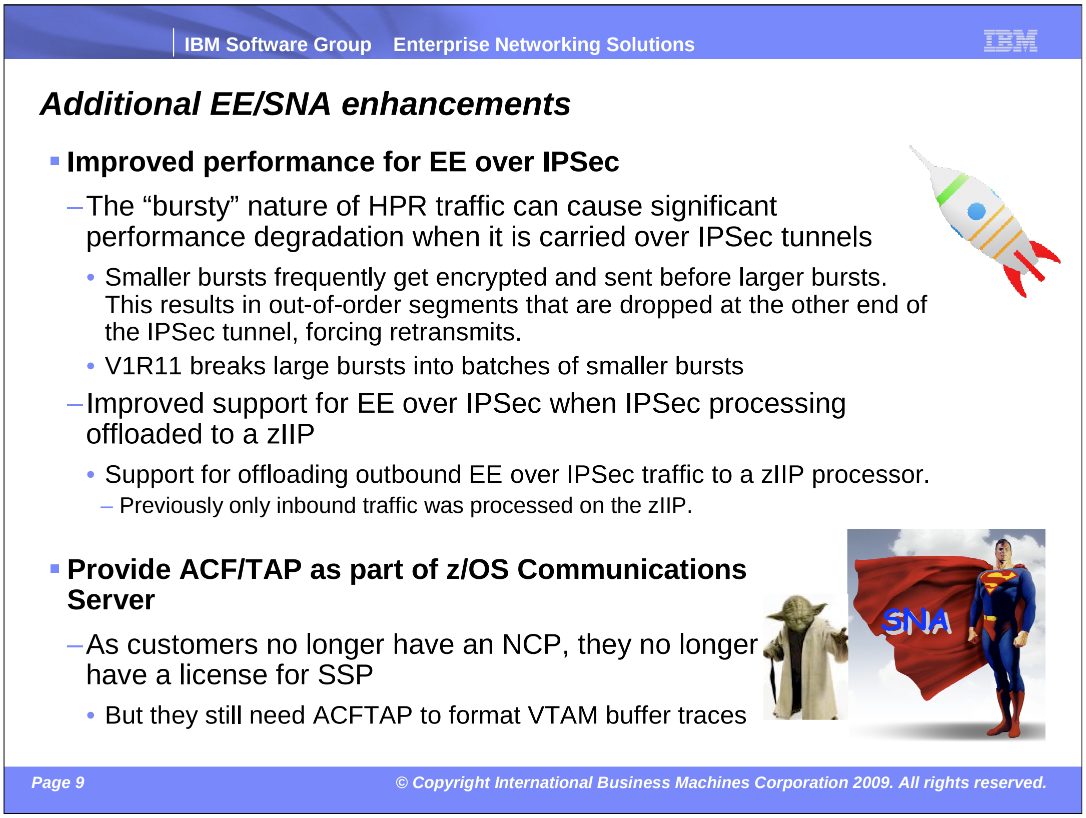
Reduction in CSA requirements for rapid transport protocol (RTP) pipes

- Before V1R11, each RTP pipe is represented by a control block in ECSA
- In V1R11, a large portion of the RTP control block was moved to an extension control block in VTAM private storage. This resulted in a significant ECSA savings for installations with a large number of RTP pipes
- Preliminary estimates of the reduction in required RTP pipe ECSA storage for various RTP counts:

RTP pipes	ECSA reduction
4000	11.5%
12000	23.5%
20000	29.5%

RTP Pipe ECSA Usage - V1R11 Relative to V1R10
Preliminary Results
%(R11 ECSA Usage relative to Base)
-10
-15
-20
-25
-30
-35
0
5
10
15
20
25
Number of RTP pipes in Thousands
ECSA

Before z/OS V1R11, each RTP control block used a block of ECSA of over two kilobytes. z/OS V1R11 identified all of the RTP's fields which were required to stay in ECSA, and moved the remaining fields to an extension control block in VTAM private storage (and anchored out of the RTP control block). This reduces the ECSA footprint of the RTP control block to less than one kilobyte. The table and graph on the chart show approximate ECSA savings for various RTP pipe counts. These numbers should be considered preliminary estimates only until the final performance report is available sometime after V1R11 general availability.



IBM Software Group | Enterprise Networking Solutions

Additional EE/SNA enhancements

- **Improved performance for EE over IPSec**
  - The “bursty” nature of HPR traffic can cause significant performance degradation when it is carried over IPSec tunnels
    - Smaller bursts frequently get encrypted and sent before larger bursts. This results in out-of-order segments that are dropped at the other end of the IPSec tunnel, forcing retransmits.
    - V1R11 breaks large bursts into batches of smaller bursts
  - Improved support for EE over IPSec when IPSec processing offloaded to a zIIP
    - Support for offloading outbound EE over IPSec traffic to a zIIP processor.
      - Previously only inbound traffic was processed on the zIIP.
- **Provide ACF/TAP as part of z/OS Communications Server**
  - As customers no longer have an NCP, they no longer have a license for SSP
    - But they still need ACFTAP to format VTAM buffer traces

Page 9

Before V1R11, enabling IPsec for EE resulted in significant reduction in throughput for streaming workloads. Also, zIIP processors, if configured, were used inefficiently. The throughput problem for streaming workloads concerns HPR traffic, which is “bursty” by nature, unlike the “flow-control” nature of TCP traffic. HPR can send megabytes of data within a single burst interval. A large HPR burst followed by a smaller HPR burst from a local node causes out-of-order conditions on the remote node. Filtering and encryption are performed on each data segment in a burst before sending. This algorithm can result in smaller bursts being sent out before larger bursts from the local node.

The second problem is that the zIIP-enabled IPSec provided in V1R8 was designed primarily for TCPIP stack traffic (TCP and UDP). Specific handling for Enterprise Extender was not implemented. Enterprise Extender utilization of zIIPs is dependent on the traffic pattern. The majority of the inbound data is offloaded to zIIPs, all the way through the RTP layer. But none of the outbound data is offloaded to zIIPs. In V1R11, Communications Server exploits zIIPs for all traffic patterns. For outbound traffic, after processing the first four-segment batch, all subsequent four-segment batches carved from the large burst are redispatched to a zIIP processor. This avoids the latency overhead of redispatch for small, interactive workloads for outbound traffic. For inbound traffic, if processing more than four segments in a single interrupt, and currently executing on a zIIP processor, Communications Server will redispatch onto a general CP before calling HPR. This avoids the latency overhead of redispatch for small, interactive workloads for inbound traffic.

Trademarks, copyrights, and disclaimers

IBM, the IBM logo, ibm.com, and the following terms are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both:

VTAM z/OS

If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of other IBM trademarks is available on the Web at "Copyright and trademark information" at http://www.ibm.com/legal/copytrade.shtml

Other company, product, or service names may be trademarks or service marks of others.

Product data has been reviewed for accuracy as of the date of initial publication. Product data is subject to change without notice. This document could include technical inaccuracies or typographical errors. IBM may make improvements or changes in the products or programs described herein at any time without notice. Any statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business. Any reference to an IBM Program Product in this document is not intended to state or imply that only that program product may be used. Any functionally equivalent program, that does not infringe IBM's intellectual property rights, may be used instead.

THE INFORMATION PROVIDED IN THIS DOCUMENT IS DISTRIBUTED "AS IS" WITHOUT ANY WARRANTY, EITHER EXPRESS OR IMPLIED. IBM EXPRESSLY DISCLAIMS ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NONINFRINGEMENT. IBM shall have no responsibility to update this information. IBM products are warranted, if at all, according to the terms and conditions of the agreements (for example, IBM Customer Agreement, Statement of Limited Warranty, International Program License Agreement, etc.) under which they are provided. Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products in connection with this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products.

IBM makes no representations or warranties, express or implied, regarding non-IBM products and services.

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents or copyrights. Inquiries regarding patent or copyright licenses should be made, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. All customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. The actual throughput or performance that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput or performance improvements equivalent to the ratios stated here.

Note to U.S. Government Users - Documentation related to restricted rights-Use, duplication or disclosure is subject to restrictions set forth in GSA ADP Schedule Contract and IBM Corp.