z/TPF Networking on IBM z17

Communications Subcommittee

Jamie Farmer

2025 TPF Users Group Conference
May 05-07, Austin, TX

IBM Z

IBM

Disclaimer

Any reference to future plans are for planning purposes only. IBM reserves the right to change those plans at its discretion. Any reliance on such a disclosure is solely at your own risk. IBM makes no commitment to provide additional information in the future.

Background

- The OSA-Express adapter on IBM Z is the network interface card (NIC) for an IBM Z server.

  - The OSA-Express adapter is the mechanism operating systems use to communicate across TCP/IP.

- The protocol z/TPF uses to communicate with the OSA-Express adapter is **queued-direct I/O (QDIO)**

  - One OSA-Express adapter can be shared by multiple LPARs or test systems.

Pain Point

- When the QDIO protocol was developed the top network speed was 100 MbE (100 megabits per second)

  - Today, network speeds are 25 GbE (25 gigabits per second)

    - **256 times the throughput of 100 MbE!**

- The QDIO architecture supports 8 megabytes of buffer space for inbound packets

  - Network line speeds have increased, and in turn, more traffic volume

  - Moving to 10 GbE and 25 GbE networks you are more likely to occasionally see...

    TTCP0323W ALL READ BUFFERS FULL FOR OSA-OSA1, PACKETS MIGHT BE LOST
    MILLISECONDS SINCE OSA POLLING WAS CALLED: 11 MS

Introducing the Network Express Adapter

- The Network Express adapter is the new network interface card for IBM z17.

    - Redesigned protocol called **enhanced QDIO (EQDIO)**

        - Architected for the high-speed networks of today and for the future.

        - **32 times** more memory available to buffer data between the Network Express adapter and z/TPF.

    - 10 GbE or 25 GbE Network Express adapters are available

        - Fiber optic connector only (long wave or short wave)

OSA-Express Adapter is Still Available

- The OSA-Express adapters are still available on the IBM z17.

  - Can carry forward or purchase OSA-Express7S adapters for the IBM z17

- For console connectivity through TPF Operation Server you must carry forward or purchase the 1 GB OSA-Express adapter (OSA-ICC)

  - Fiber optic connector only (long wave or short wave)

    - If using 1000BASE-T copper for console connectivity, network updates might be required to move to IBM z17.

z/TPF Terminology for OSA-Express and Network Express

- The term **OSA-Express connection** will still exist.

- A z/TPF **OSA-Express connection** can be a ...

  - Connection to an **OSA-Express adapter** that uses traditional QDIO

  - Connection to a **Network Express adapter** that uses enhanced QDIO.
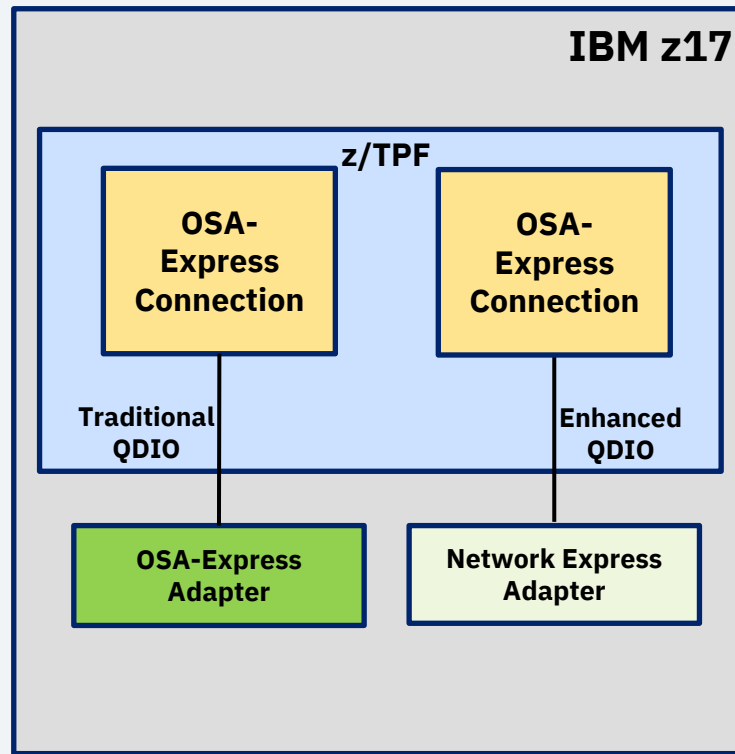
Using Network Express for TCP/IP on IBM z17

- To move your z/TPF TCP/IP workload to Network Express you have two options:

  1. Migrate to IBM z17 and move directly to Network Express adapters

     - Can set up z/TPF definitions in advance of migration.

  2. Migrate to IBM z17 with OSA-Express adapters, and then upgrade to Network Express adapters

     - When using VIPAs, you can transition from OSA-Express to Network Express without having to take a network outage!

Defining Network Express Adapters to IBM z17

- A Network Express adapter IOCDS definition uses a new CHIPID type – OSH

```
CHPID PATH=F9,TYPE=OSH,SHARED,PARTITION=(TPFTEST,TPFPROD,MVS1)
CNTLUNIT CUNUMBR=F900,UNIT=OSA,PATH=(F9)
IODEVICE UNIT=OSA,ADDRESS=(E00,016),CUNUMBR=F900,UNITADD=00,
         PARTITION=(TPFTEST,TPFPROD,MVS1)
```

- In the previous example, E00-E0F is defined to 3 partitions (TPFTEST, TPFPROD, MVS1). This is a total of 48 devices for this definition.

- The maximum number of devices for a Network Express adapter (OSH CHIPID) is 256.

  - Unlike connections to OSA-Express that require 3 devices, connections to Network Express require only a single device.

As-Is: Enabling z/TPF to use Connections to OSA-Express

- You define the maximum number of OSA-Express connections to OSA-Express adapters on z/TPF with the MAXOSA parameter in keypoint 2 (CTK2)

  - For example, MAXOSA = 2

OSA Control Block Table

OSA-Express #1
OSA-Express #2

To-Be: Enabling z/TPF to use Connections to Network Express

- You define the number of OSA-Express connections to OSA-Express adapters on z/TPF with the MAXOSA parameter in keypoint 2 (CTK2)

  - For example, MAXOSA = 2

- You define the maximum number of OSA-Express connections to Network Express adapters on z/TPF with the new MAXEOSA parameter in keypoint 2 (CTK2)

  - For example, MAXEOSA = 2

OSA Control Block Table

OSA-Express #1
OSA-Express #2
Network Express #1
Network Express #2

Both MAXOSA and MAXEOSA can be non-zero. The OSA control block table is allocated based on both definitions.
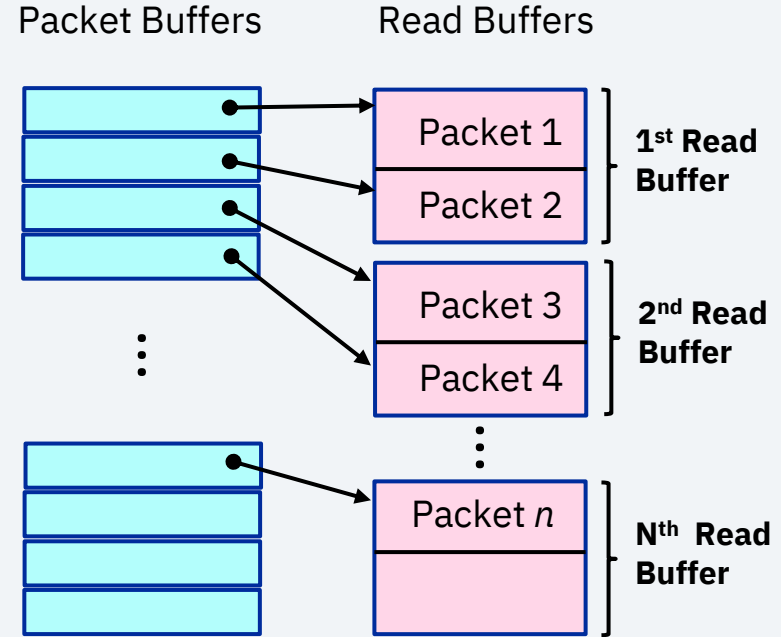
Defining Network Express Read Buffers

- With the new EOSABUFF parameter in keypoint 2, you can configure the number of allocated read buffers for each Network Express connection.

  - Valid values for EOSABUFF are 4, 16 or 64 (small, medium and large)

EOSABUFF Value	Number of Read Buffers	Storage Allocated for each MAXEOSA defined	Usage
4	4096	16 MB	Unit Test System (VPARS)
16	16,384	64 MB	Small Native Test Systems
64	65,536	256 MB	Production Systems

- Storage is allocated above the 2 GB bar

Network Express Packet vs Read Buffers

- The Network Express read buffers contain inbound packet data.

- A separate set of packet buffers are also allocated

  - Packet buffers are 8-bytes in length

  - 4 Packet buffers allocated for every one read buffer

- Packet buffers point to the beginning of the packet in the read buffer.

- Packet buffers exist for efficiency

  - Hosts, like z/TPF, do not need to parse packets in read buffers

Packet Buffers   Read Buffers

Packet 1
Packet 2
} 1st Read Buffer

Packet 3
Packet 4
} 2nd Read Buffer

Packet *n*
} Nth Read Buffer

Network Express Read Buffer Usage Display

```
ZOSAE BUFFER OSA-EOSA1
CSMP0097I 22.01.15 CPU-B SS-BSS   SSU-HPN   IS-01
OSAE0018I 22.01.15 READ AND PACKET BUFFER USAGE FOR ENHANCED QDIO OSA-EOSA1
   READ                          PACKET
  BUFFERS    INSTANCES    TIME   BUFFERS    INSTANCES    TIME
  -------  -------------  ----   -------  -------------  ----
   1-2%              51     1     1-2%              45     1
   2-6%              37     2     2-6%              26     2
   6-11%              9     4     6-11%              6     4
  11-31%              0     0    11-31%              0     0
  31-51%              0     0    31-51%              0     0
  51-71%              0     0    51-71%              0     0
  71-99%              0     0    71-99%              0     0

 Number of read buffers defined: 65 536
 Number of packet buffers defined: 262 144
 Number of read buffer full conditions: 0
 Number of read buffer full conditions caused by system error: 0
 _
 Number of packet buffer full conditions: 0
 Number of packet buffer full conditions caused by system error: 0
 Average time between polling during read buffer full condition: 0
 Average time between polling during packet buffer full condition: 0
 TOD of last buffer full condition: N/A _
  TOD of last buffer statistics reset: E0847E1047AD4B78
```

Number of times in-use buffers reached a percentage threshold and the average time (in milliseconds) between OSA polling calls when that threshold was reached.

Number of buffers defined

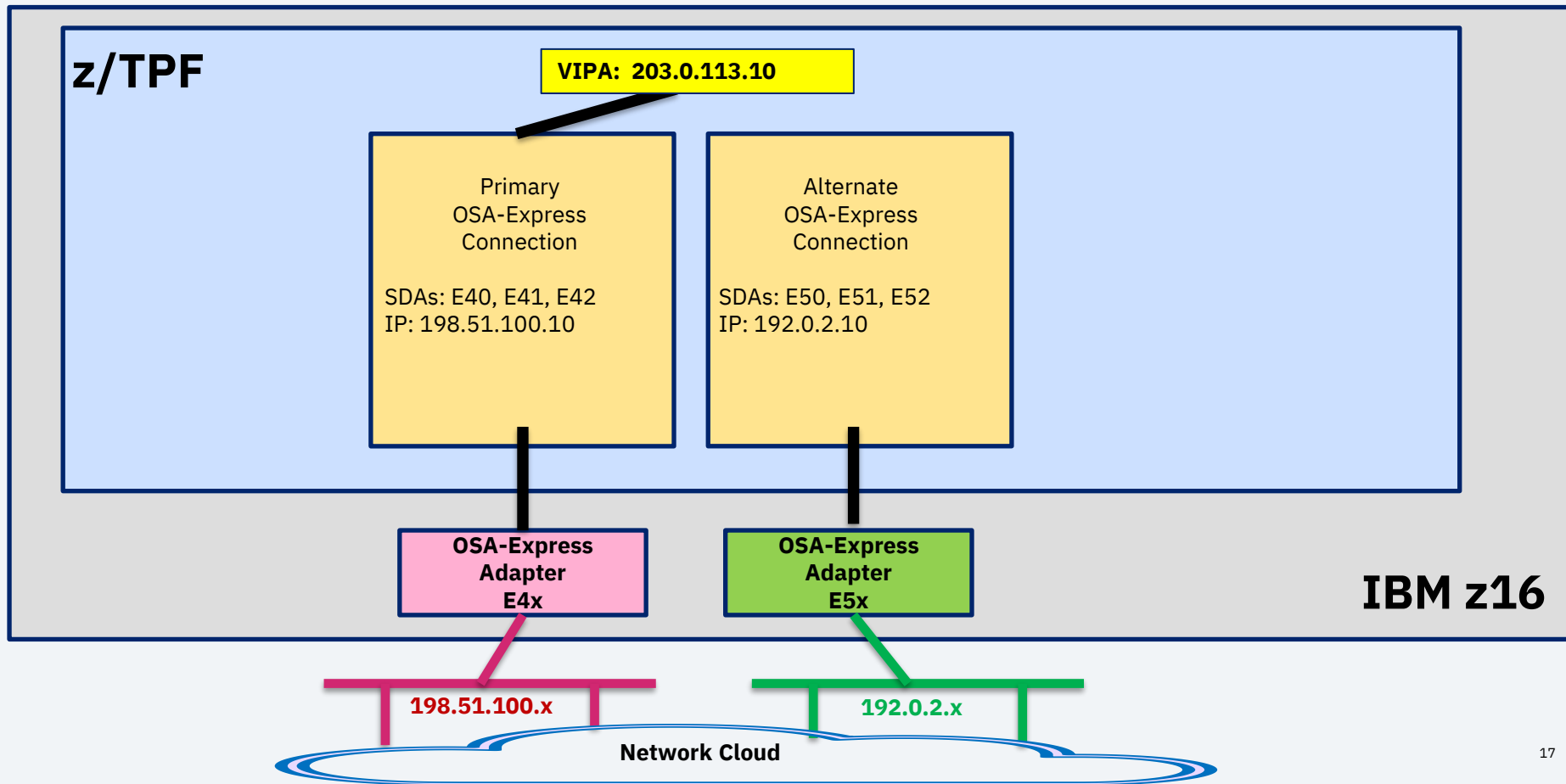**Statistics for buffer full conditions**

Defining an OSA-Express Connection to Network Express

- Use the same ZOSAE command to define OSA-Express connections to OSA-Express adapters or Network Express adapters

  - Can modify existing definition to connect to either adapter

- New parameters on ZOSAE for connections to Network Express

  - **EOSADEV** : Device used to connect to the Network Express adapter

    - Only one device is required for Network Express

  - **MODE** :

    - Indicates the desired mode of the OSA-Express connection: EQDIO (Network Express) or QDIO (OSA-Express)

- When MODE-EQDIO, the following parameters are ignored on the OSA definition

  - READ, DATA, OLM, PORT, PORTNUM, NET

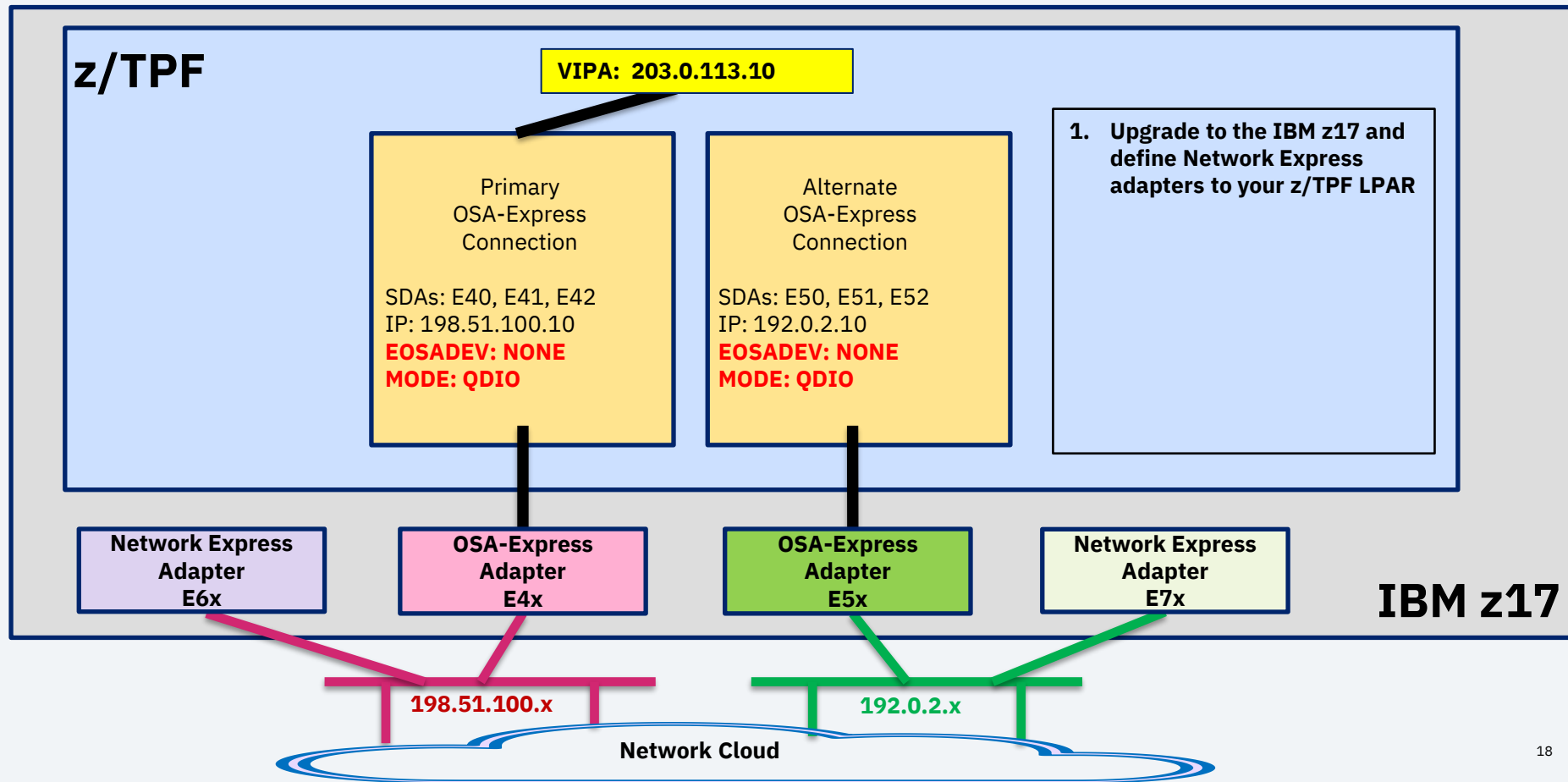   - NET (network type) is automatically determined during activation

Migrating to Network Express

- You can choose to move forward with an IBM z17 and maintain connectivity through OSA-Express adapters (as you do today)

    - When using virtual IP addresses (VIPAs), you can migrate to Network Express without having to take a network outage.

    - With your VIPAs assigned to one OSA-Express connection, you can transition the other OSA-Express connection to use a Network Express adapter.

Typical z/TPF OSA-Express Architecture

z/TPF

VIPA: 203.0.113.10
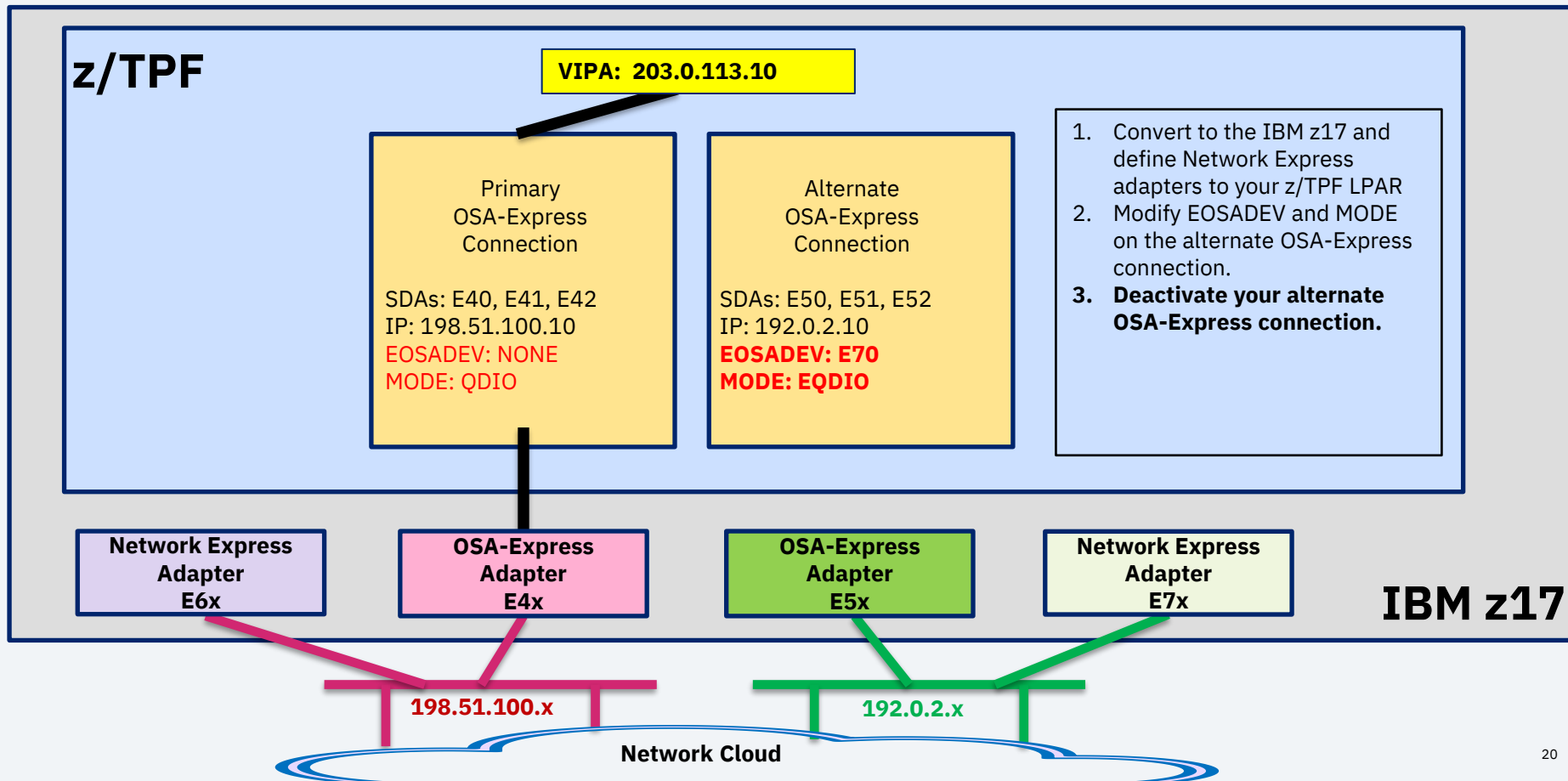
Primary
OSA-Express
Connection

SDAs: E40, E41, E42
IP: 198.51.100.10

Alternate
OSA-Express
Connection

SDAs: E50, E51, E52
IP: 192.0.2.10

IBM z16

OSA-Express
Adapter
E4x

OSA-Express
Adapter
E5x

198.51.100.x

192.0.2.x

Network Cloud

Migrating to the IBM z17

z/TPF

VIPA: 203.0.113.10

Primary
OSA-Express
Connection

SDAs: E40, E41, E42
IP: 198.51.100.10
**EOSADEV: NONE
MODE: QDIO**

Alternate
OSA-Express
Connection

SDAs: E50, E51, E52
IP: 192.0.2.10
**EOSADEV: NONE
MODE: QDIO**

1. **Upgrade to the IBM z17 and define Network Express adapters to your z/TPF LPAR**

Network Express Adapter E6x

OSA-Express Adapter E4x

OSA-Express Adapter E5x

Network Express Adapter E7x

IBM z17

198.51.100.x

192.0.2.x

Network Cloud

Update Your OSA-Express Definition for Network Express

Deactivate Your Backup OSA-Express Connection

z/TPF

VIPA: 203.0.113.10

Primary OSA-Express Connection

SDAs: E40, E41, E42
IP: 198.51.100.10
EOSADEV: NONE
MODE: QDIO

Alternate OSA-Express Connection

SDAs: E50, E51, E52
IP: 192.0.2.10
EOSADEV: E70
MODE: EQDIO

1. Convert to the IBM z17 and define Network Express adapters to your z/TPF LPAR
2. Modify EOSADEV and MODE on the alternate OSA-Express connection.
3. **Deactivate your alternate OSA-Express connection.**

Network Express Adapter E6x

OSA-Express Adapter E4x

OSA-Express Adapter E5x

Network Express Adapter E7x

IBM z17

198.51.100.x
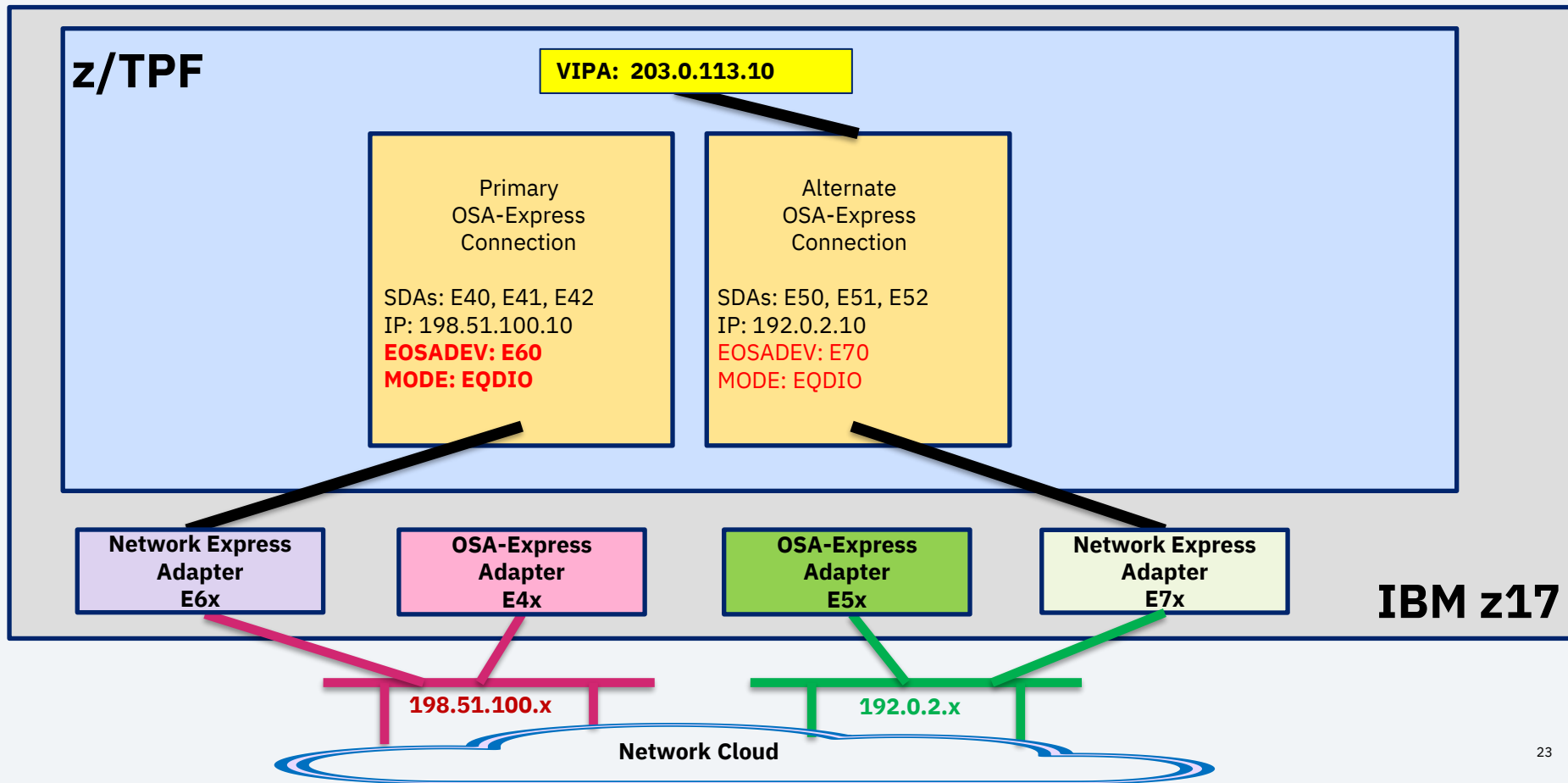
192.0.2.x

Network Cloud

Activate Your Backup OSA-Express Connection

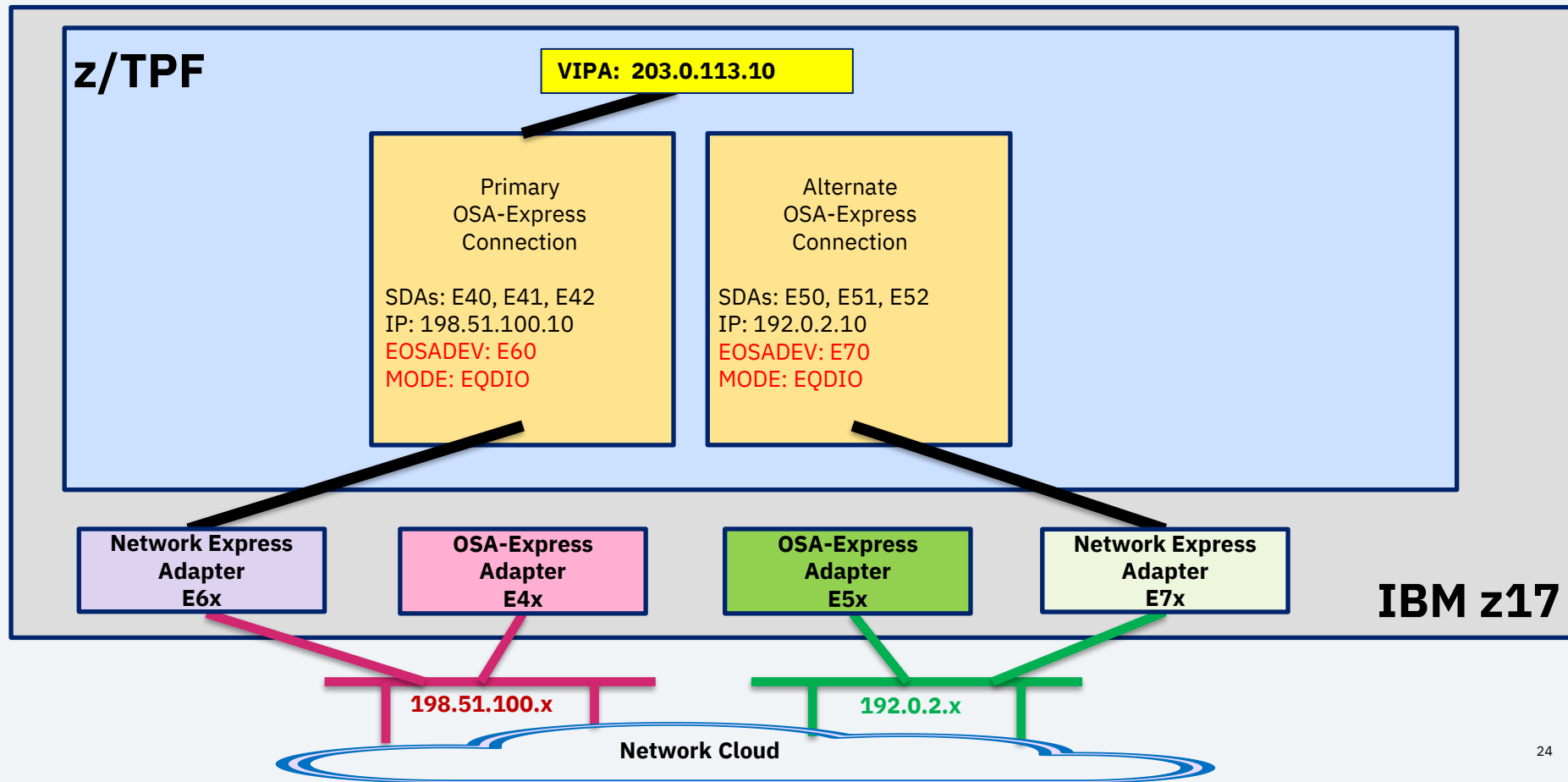# Swing Your VIPAs to the Alternate
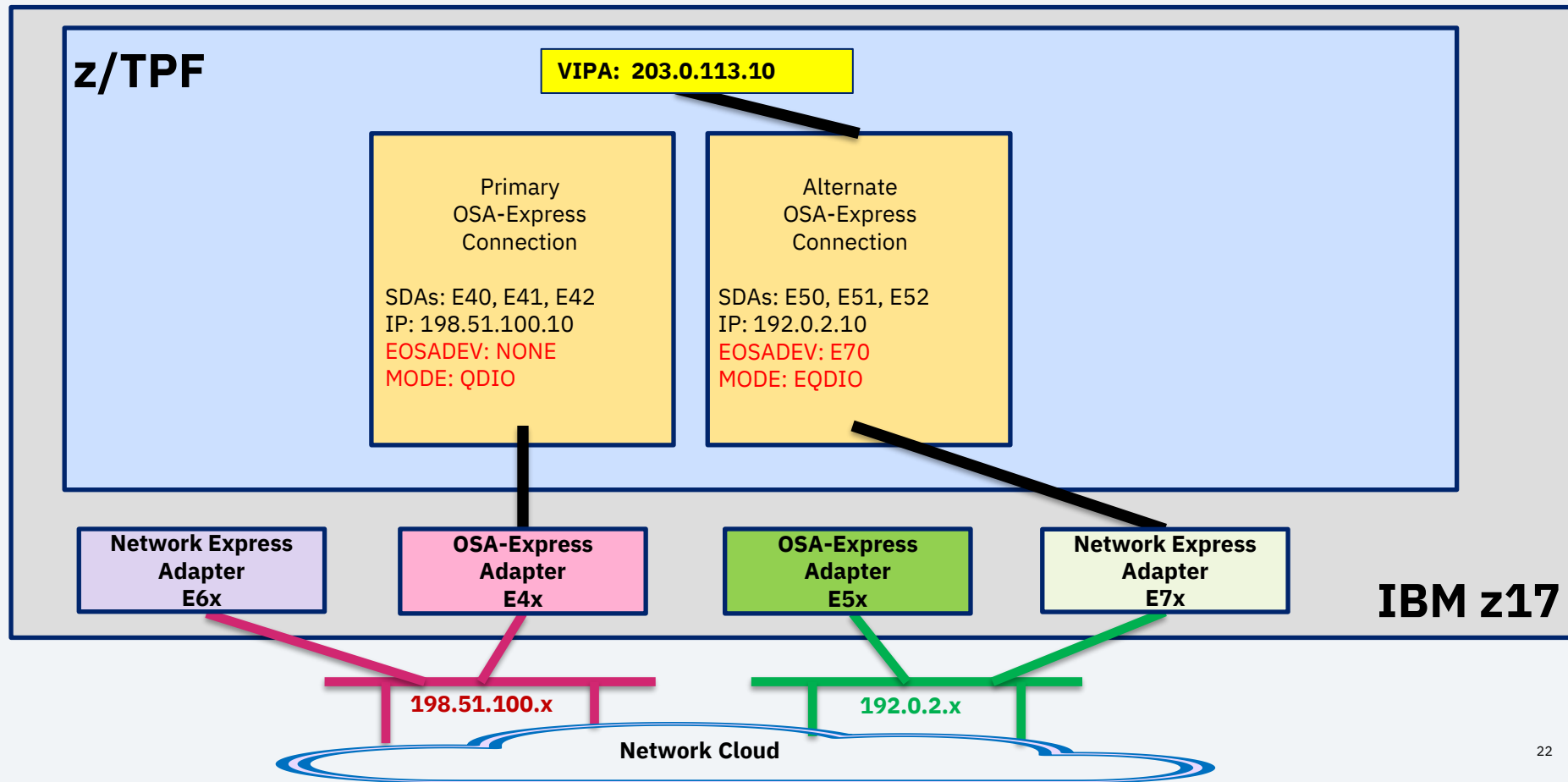
Convert the Primary OSA-Express Connection

Swing Your VIPAs Back to the Primary

z/TPF

VIPA: 203.0.113.10

Primary
OSA-Express
Connection

SDAs: E40, E41, E42
IP: 198.51.100.10
EOSADEV: E60
MODE: EQDIO

Alternate
OSA-Express
Connection

SDAs: E50, E51, E52
IP: 192.0.2.10
EOSADEV: E70
MODE: EQDIO

Network Express
Adapter
E6x

OSA-Express
Adapter
E4x

OSA-Express
Adapter
E5x

Network Express
Adapter
E7x

IBM z17

198.51.100.x

192.0.2.x

Network Cloud

What if My Network Express Adapters are Connected to Separate Network?

z/TPF

VIPA: 203.0.113.10

Primary
OSA-Express
Connection

SDAs: E40, E41, E42
IP: 198.51.100.10

Alternate
OSA-Express
Connection

SDAs: E50, E51, E52
IP: 192.0.2.10
IP: 192.0.3.10
EOSADEV: E70
MODE: EQDIO

OSA-Express
Adapter
E4x

OSA-Express
Adapter
E5x

Network Express
Adapter
E7x

IBM z17

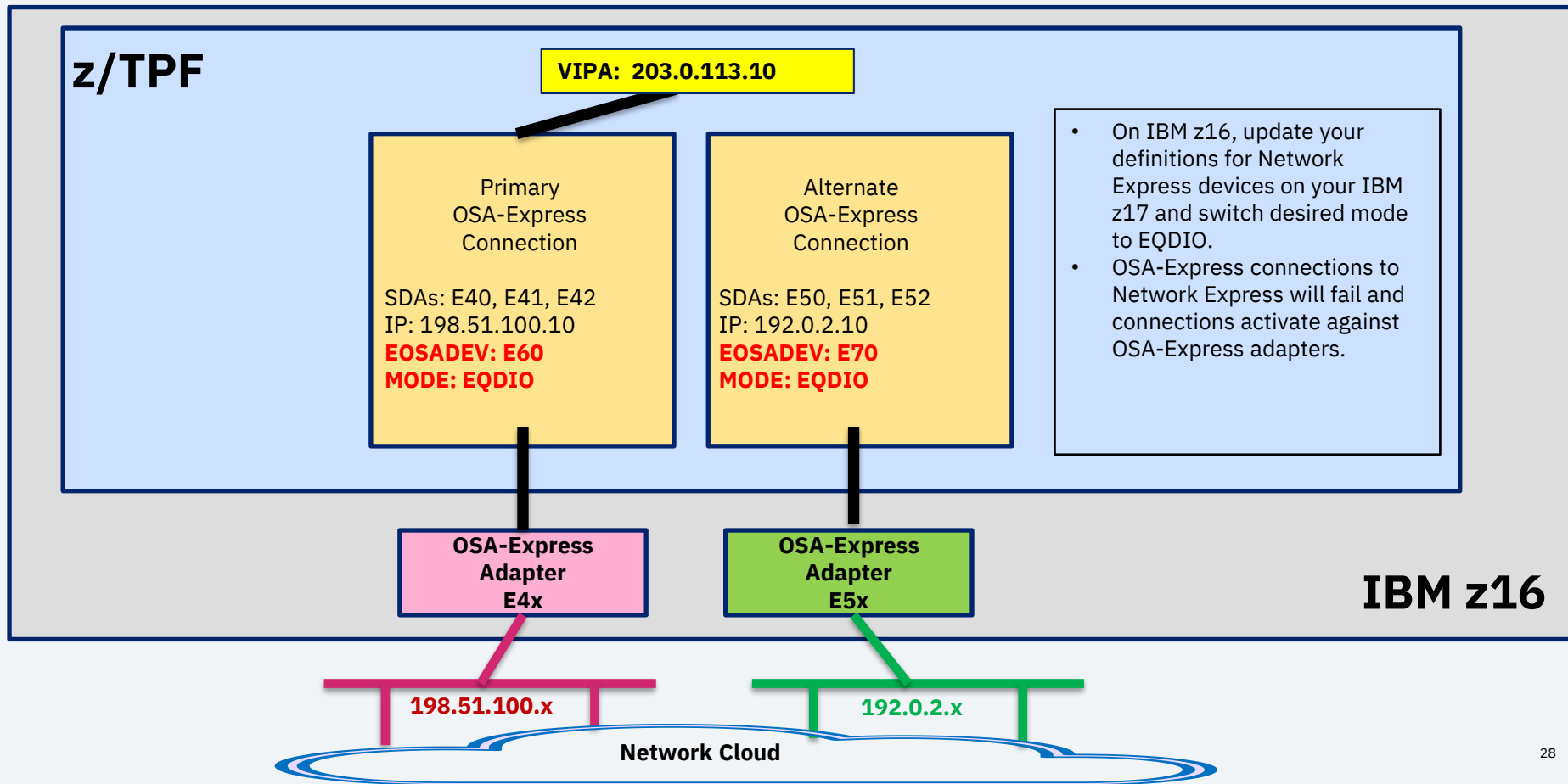198.51.100.x

192.0.2.x

192.0.3.x

Network Cloud
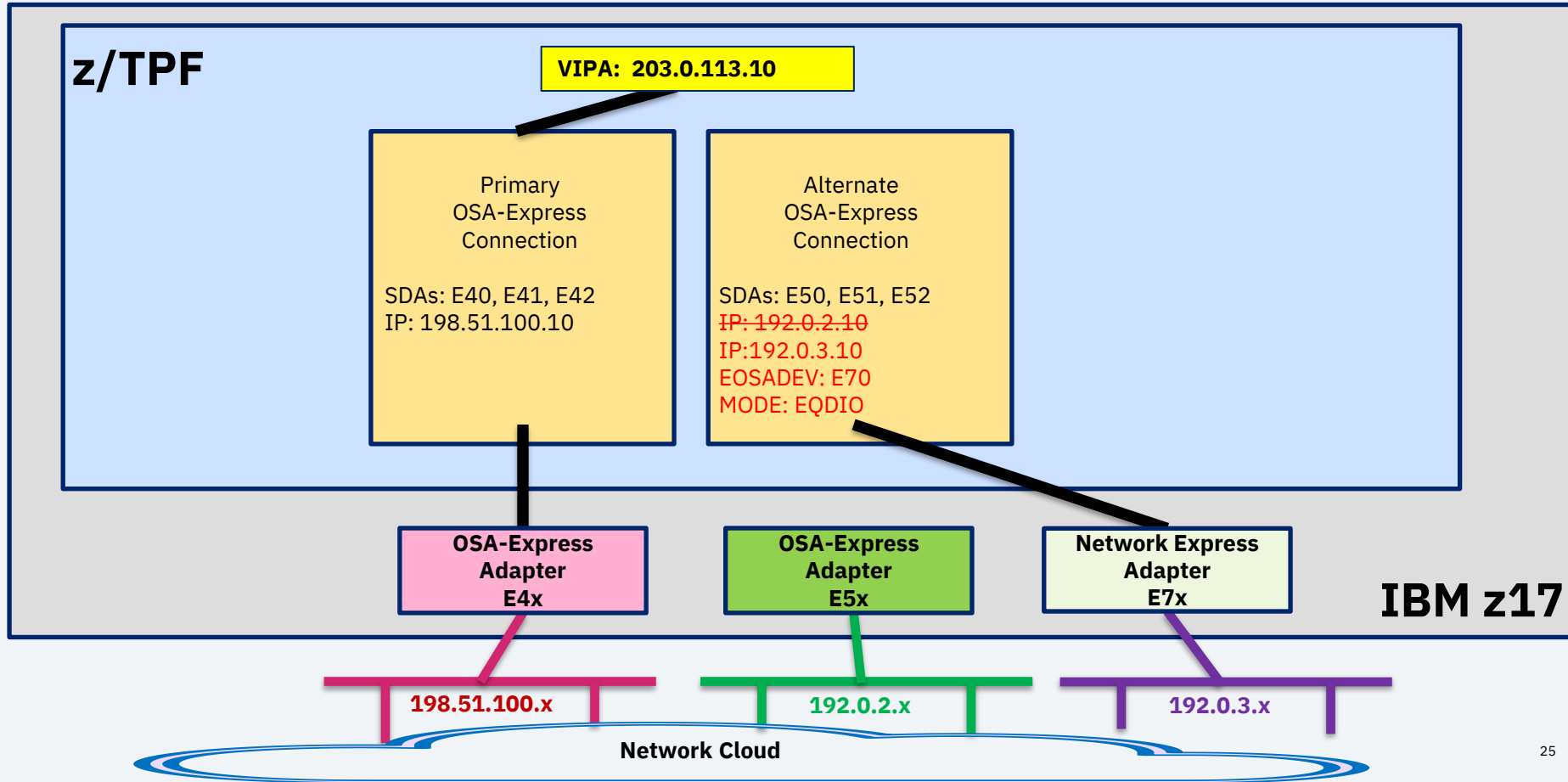
Changing the IP address of the OSA-Express Connection

- To ease migration, you can now modify the IP address and subnet of the OSA-Express connection by following these steps...

  1. Deactivate the OSA-Express connection

  2. Remove the gateway IP addresses of the OSA-Express connection (ZOSAE MOD GATEWAY1-NONE)

  3. Delete routing table statements that contain the OSA-Express connection's IP address.

  4. Modify the IP address or subnet mask, or both, of the OSA-Express connection

  5. Define your new gateway IP addresses to the OSA-Express connection

  6. Define routing table statements for this new local IP address.

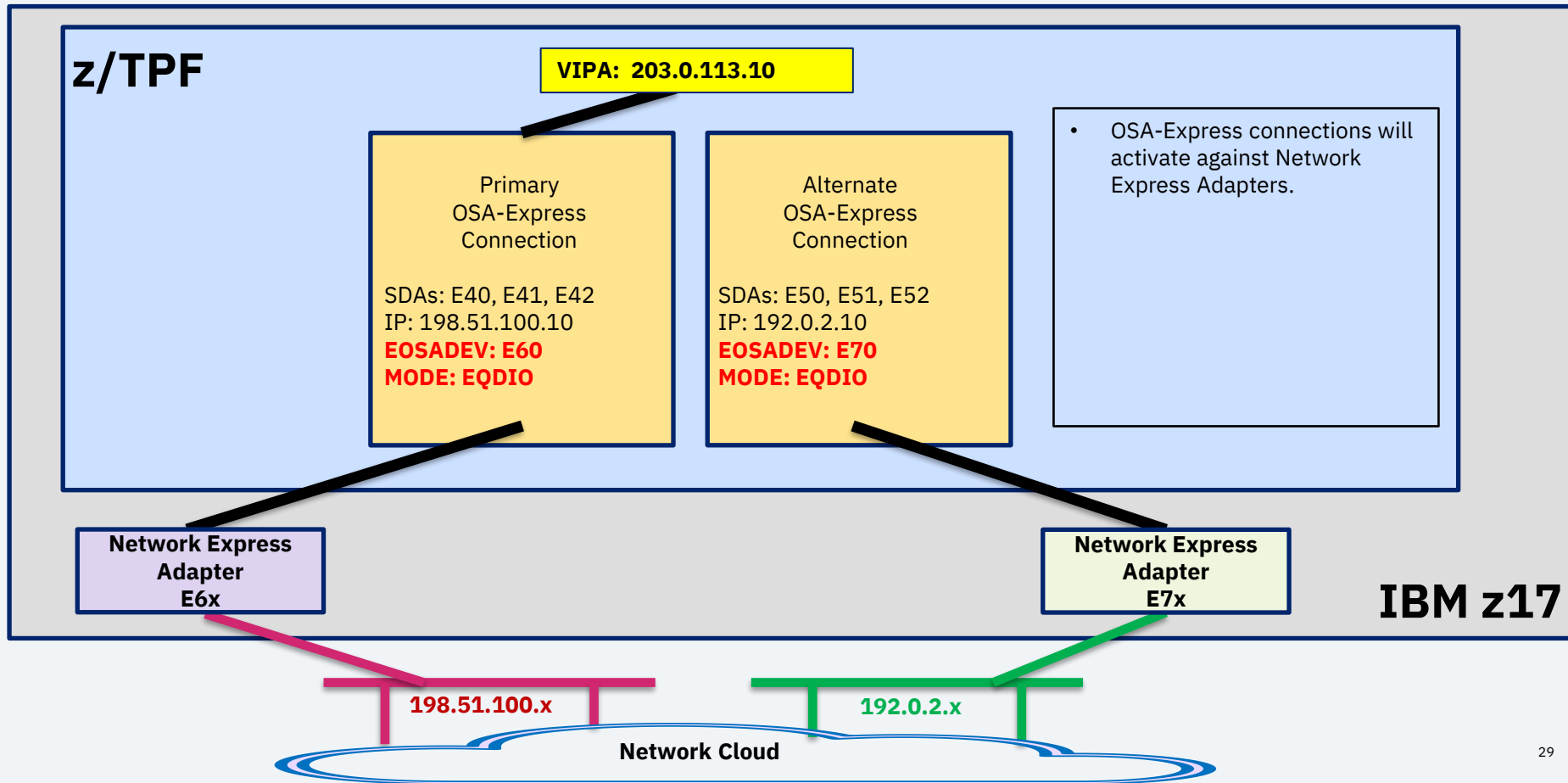 7. Activate the OSA-Express connection.

Moving Directly to Network Express Adapters During IBM z17 Installation

- You can also move directly to Network Express adapters during the IBM z17 installation

- Update your OSA-Express connections for connectivity to Network Express before you migrate to the IBM z17.

- When an EOSADEV is specified, and MODE is set to EQDIO, the z/TPF system tries to bring up OSA-Express connections to Network Express first.

  - If that fails, the z/TPF system will try to activate with existing traditional QDIO definitions.

OSA-Express Connections prior to IBM z17

z/TPF

VIPA: 203.0.113.10

**Primary
OSA-Express
Connection**

SDAs: E40, E41, E42
IP: 198.51.100.10
**EOSADEV: E60
MODE: EQDIO**

**Alternate
OSA-Express
Connection**

SDAs: E50, E51, E52
IP: 192.0.2.10
**EOSADEV: E70
MODE: EQDIO**

- On IBM z16, update your definitions for Network Express devices on your IBM z17 and switch desired mode to EQDIO.
- OSA-Express connections to Network Express will fail and connections activate against OSA-Express adapters.

**OSA-Express
Adapter
E4x**

**OSA-Express
Adapter
E5x**

IBM z16

198.51.100.x

192.0.2.x

Network Cloud

OSA-Express Connections after IBM z17 Installed

Falling Back to a Previous Processor Level

- In this migration scenario, the definitions for connecting to OSA-Express adapters still exist in the OSA-Express connection definition

  - The QDIO devices (read, write, data) are still defined

  - The other definitions like port number and port name are still defined.

- If you fall back to a previous processor, the system will try activating to the Network Express adapter first.

  - If that fails, it will try to activate using the existing QDIO definitions.

- This can also be useful in disaster recovery situations where your DR processor is older than IBM z17.
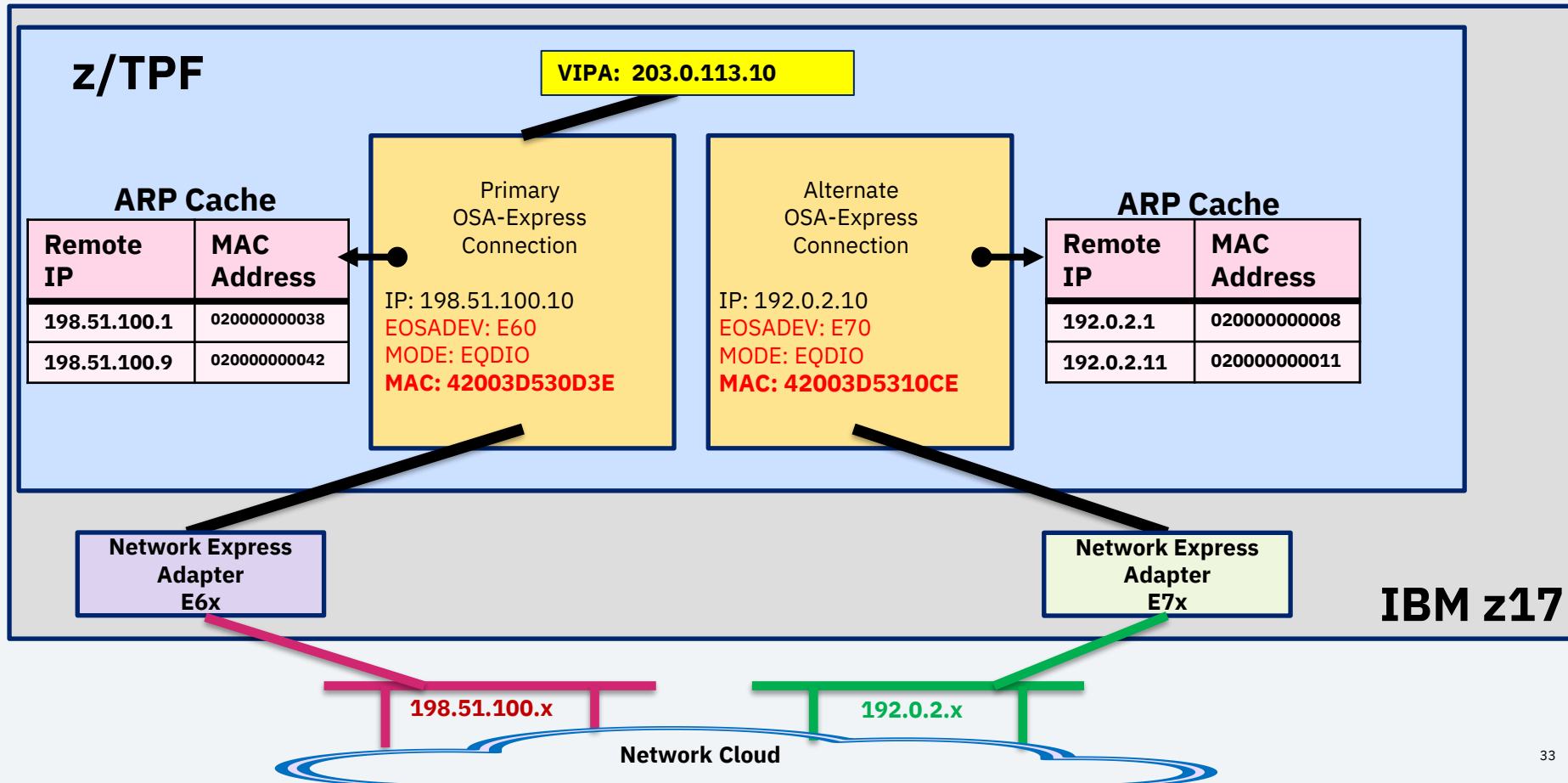
As-Is: With QDIO, z/TPF is a Layer 3 Node

- With the QDIO protocol communicating with an OSA-Express adapter, the z/TPF system acts as a layer 3 network node

  - As a layer 3 networking node, z/TPF communicates at the network layer of the open systems interconnection (OSI) model

    - z/TPF routes packets based on IP addresses.

    - z/TPF does not build ethernet headers

    - z/TPF does not have knowledge of MAC addresses

  - Layer 2 functionality was all handled by the OSA-Express adapter.

To-Be: With Enhanced QDIO, z/TPF is Layer 2 Node

- With enhanced QDIO communicating with Network Express, hosts are required to operate as a layer 2 node.

  - As a layer 2 networking node, z/TPF communicates at the data link layer of the OSI model

  - The z/TPF system needs to build (and parse) ethernet headers and must have knowledge of MAC addresses.

    - Each OSA-Express connection to a Network Express adapter now maintains a local MAC address assigned by the Network Express adapter.

  - The Network Express adapter provides control flows to resolve the MAC addresses for the remote IP address using the ARP protocol

    - Only required for IP addresses on the OSA-Express connection's subnet, for example, first hop gateways

- The z/TPF system now maintains an ARP cache for quickly accessing the MAC address for a remote IP addresses on our subnet.

  - For example, if the OSA-Express connection used a subnet mask of 255.255.255.0, then we would only have at most 256 remote IP addresses in our cache.

The ARP Cache on z/TPF OSA-Express Connections

z/TPF

VIPA: 203.0.113.10

Primary OSA-Express Connection

IP: 198.51.100.10
EOSADEV: E60
MODE: EQDIO
MAC: 42003D530D3E

Alternate OSA-Express Connection

IP: 192.0.2.10
EOSADEV: E70
MODE: EQDIO
MAC: 42003D5310CE

ARP Cache

Remote IP	MAC Address
198.51.100.1	020000000038
198.51.100.9	020000000042

ARP Cache

Remote IP	MAC Address
192.0.2.1	020000000008
192.0.2.11	020000000011

Network Express Adapter E6x

Network Express Adapter E7x

IBM z17

198.51.100.x

192.0.2.x

Network Cloud

Displaying the ARP Cache for an OSA-Express Connection

- You can display the ARP cache for a specific OSA-Express connection that is using enhanced QDIO to communicate with a Network Express adapter

```
User:    ZOSAE ARPC DISPLAY OSA-EOSA2

System: OSAE0017I 08.57.03 ARP CACHE DISPLAY FOR OSA-EOSA2

       REMOTE IP        MAC ADDRESS   STATUS   LAST UPDATED

       --------------- ------------ -------- ---------------

       198.51.100.1     020000000008 ASSIGNED Jun  3 10:25:29

       198.51.100.11    02000000000A ASSIGNED Jun  3 10:24:44

       198.51.100.12    000000000000 IN-PROG  Jun  3 10:24:46

       198.51.100.13    000000000000 UNKNOWN  Jun  3 10:24:48

       END OF DISPLAY
```

Types of Status

ASSIGNED:
  MAC address for remote IP address was resolved

IN-PROG:
  MAC address resolution in progress

UNKNOWN:
  MAC address could not be resolved.

Deleting ARP Cache Entries

- In the event something is wrong with a given MAC address and the ARP cache contains stale data...

    – You can delete the ARP cache entry for a specific remote IP address.

    ```
    User:    ZOSAE ARPC DELETE OSA-EOSA2 RIP-198.51.100.1

    System: OSAE0016I 14.52.45 ARP CACHE ENTRY FOR RIP-198.51.100.1 WAS DELETED
    ```

- The next time the z/TPF system sends a packet to this remote IP address, an ARP request will be sent again to resolve the MAC address.

Summary

- PJ46989 is planned for delivery in June 2025.

  - Provides support for the Network Express adapter on IBM z17.

  - Required to use Network Express adapters.

- No application or middleware changes required!

- No changes to z/TPF utilities, such as IP trace

- Supports migration without a network outage

Thank you