FUJITSU

shaping tomorrow with you

Automated Migration of Port Profile for Multi-level Switches

September 9, 2011

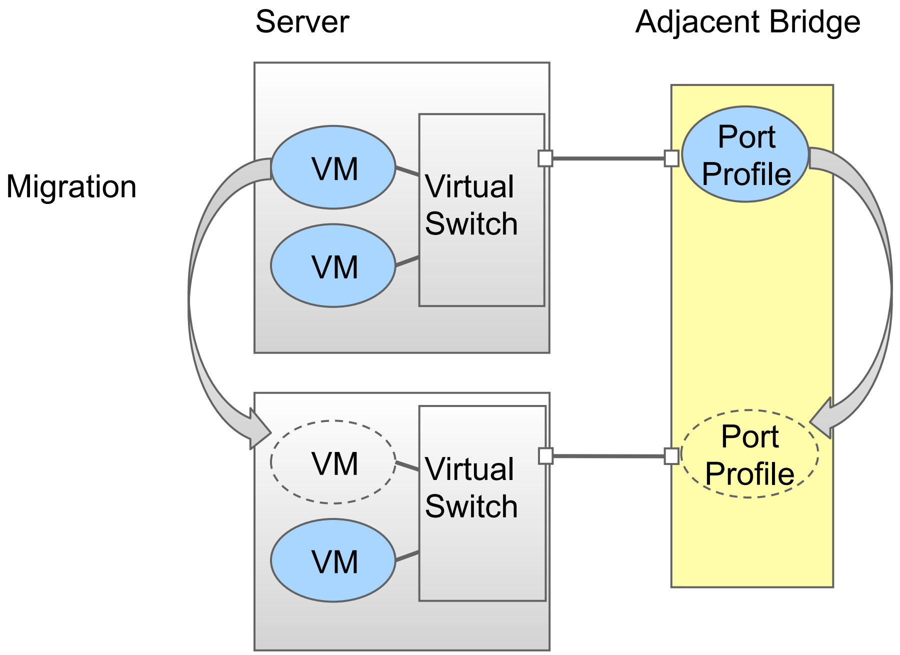
Yukihiro Nakagawa
Server Technologies Laboratory
Fujitsu Laboratories Ltd.

Outline

- Automated Migration of Port Profile (AMPP)
- 802.1Qbg VSI Discovery for AMPP
- AMPP for Multi-level Switches
- Prototype and Evaluation
- Conclusion

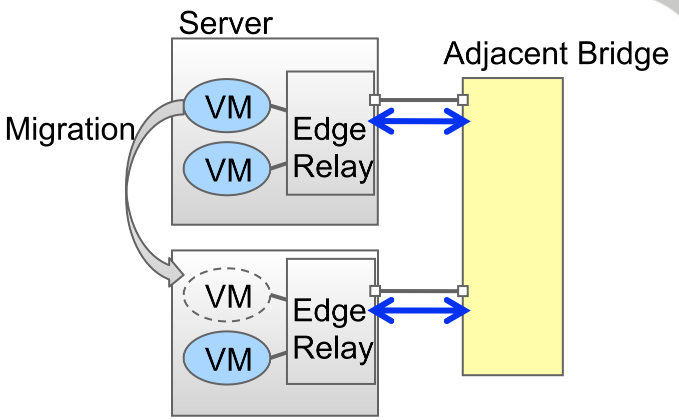
Automated Migration of Port Profile (AMPP)

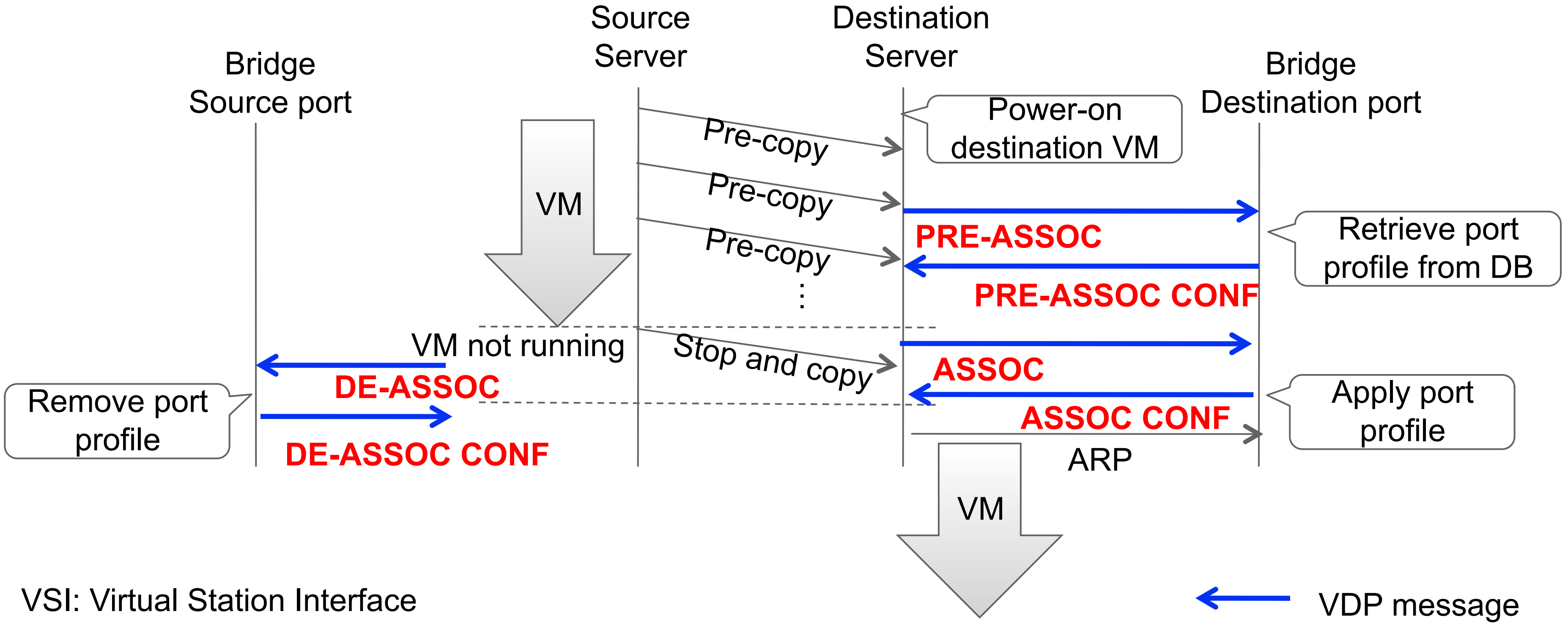
- Dynamic Infrastructure for Cloud Computing
  - A Cloud DC requires dynamic infrastructure and flexible allocation of computing resources on demand.
  - The virtualization technology and virtual machine mobility are important to realize dynamic infrastructure.
  - In this dynamic environment, AMPP automates a task to move port profile in an adjacent bridge along with VM migration.



802.1Qbg VSI Discovery for AMPP

- VSI discovery protocol supports the association of a VSI with a bridge port.
- This protocol enables synchronization between hypervisor and adjacent bridge.
  - An usage example is shown below.



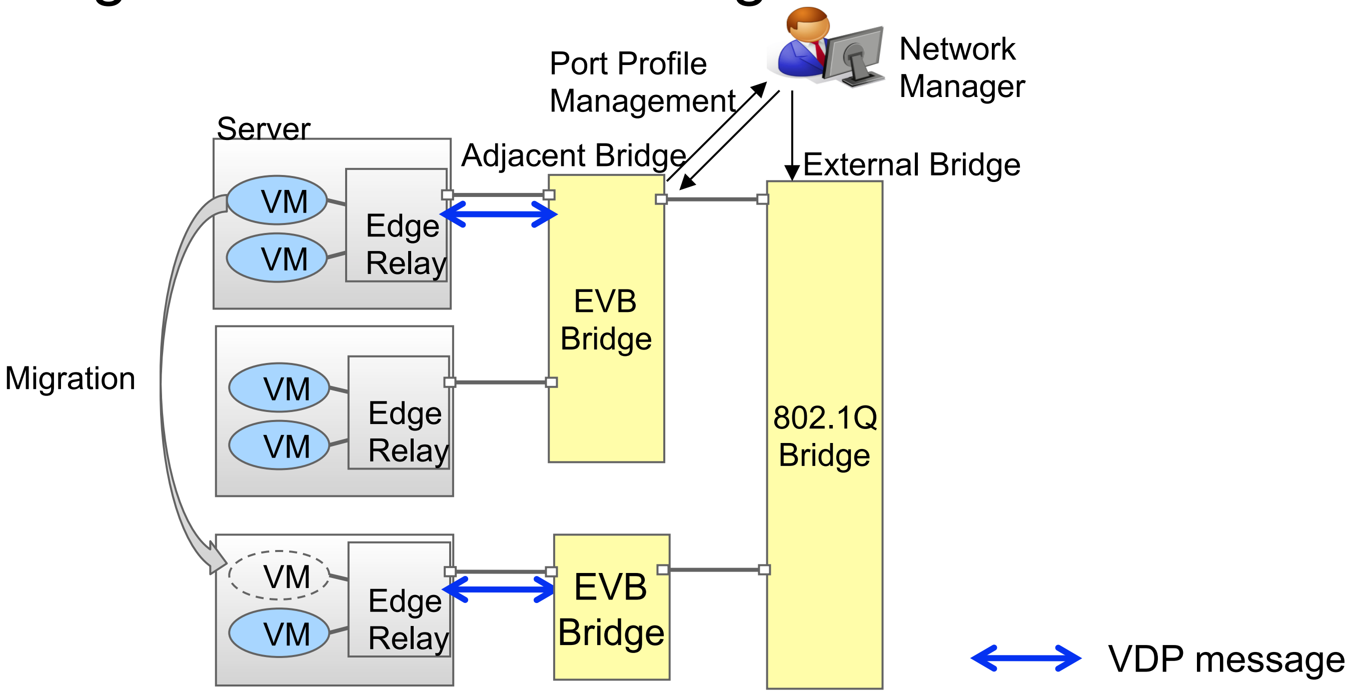


VSI: Virtual Station Interface

Current Definition of VSI Discovery

- The standard protocol is defined between server and adjacent bridge and a network manager needs to configure non-adjacent external bridges.
  - In a blade server configuration, there is a switch blade in the chassis and a VM migration to another chassis always requires network manager.

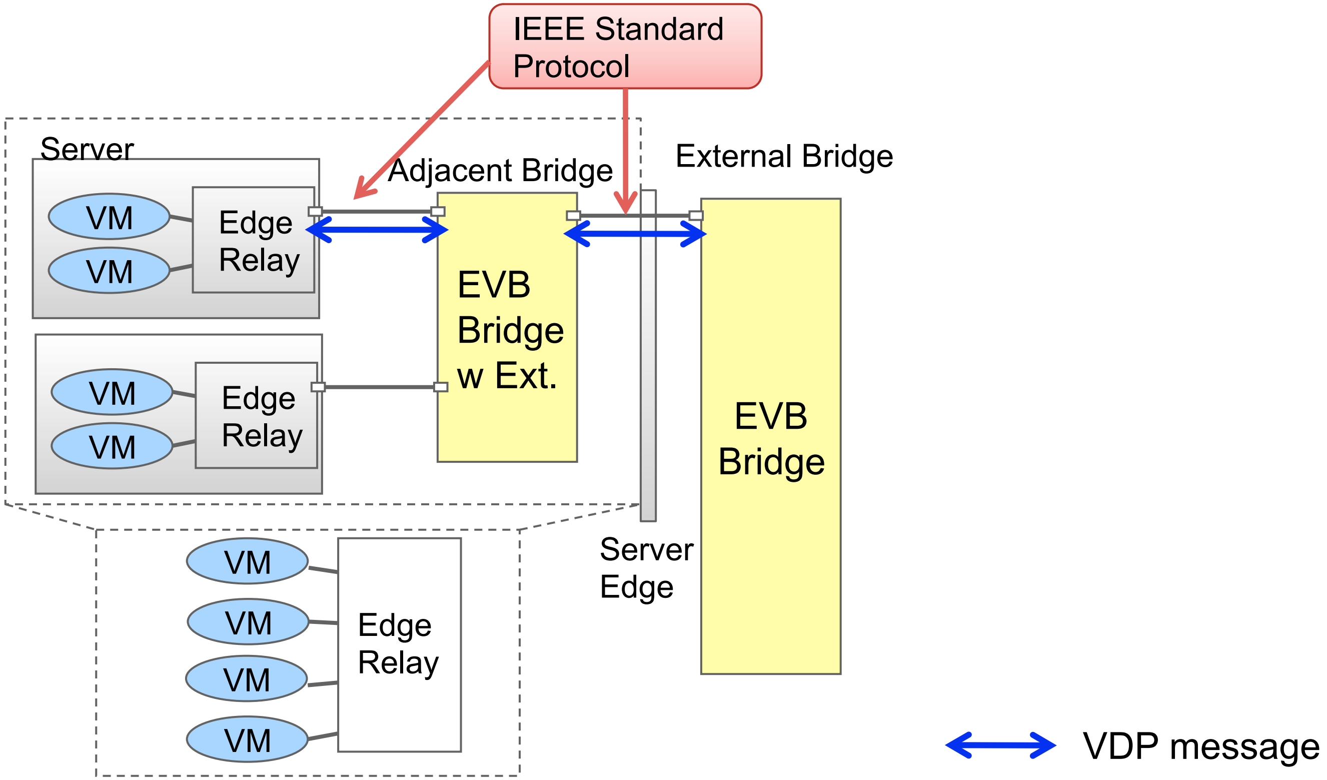
➡ We would like to automatically configure non-adjacent external bridges w/o network manager.



AMPP for Multi-level Switches

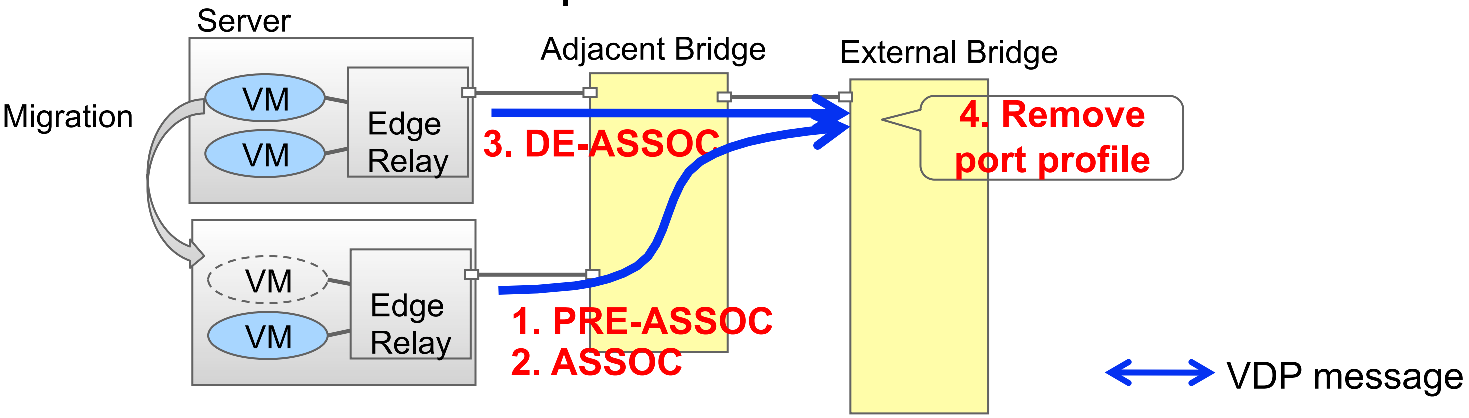
Our Proposal

- Automatically configure non-adjacent external bridges in addition to adjacent bridges using the standard protocol between bridge and bridge.
  - From an upper level switch, server and adjacent bridge can be seen as a server.
- Any standard compliant EVB bridge can be used as an external bridge.



Forwarding of VDP Messages

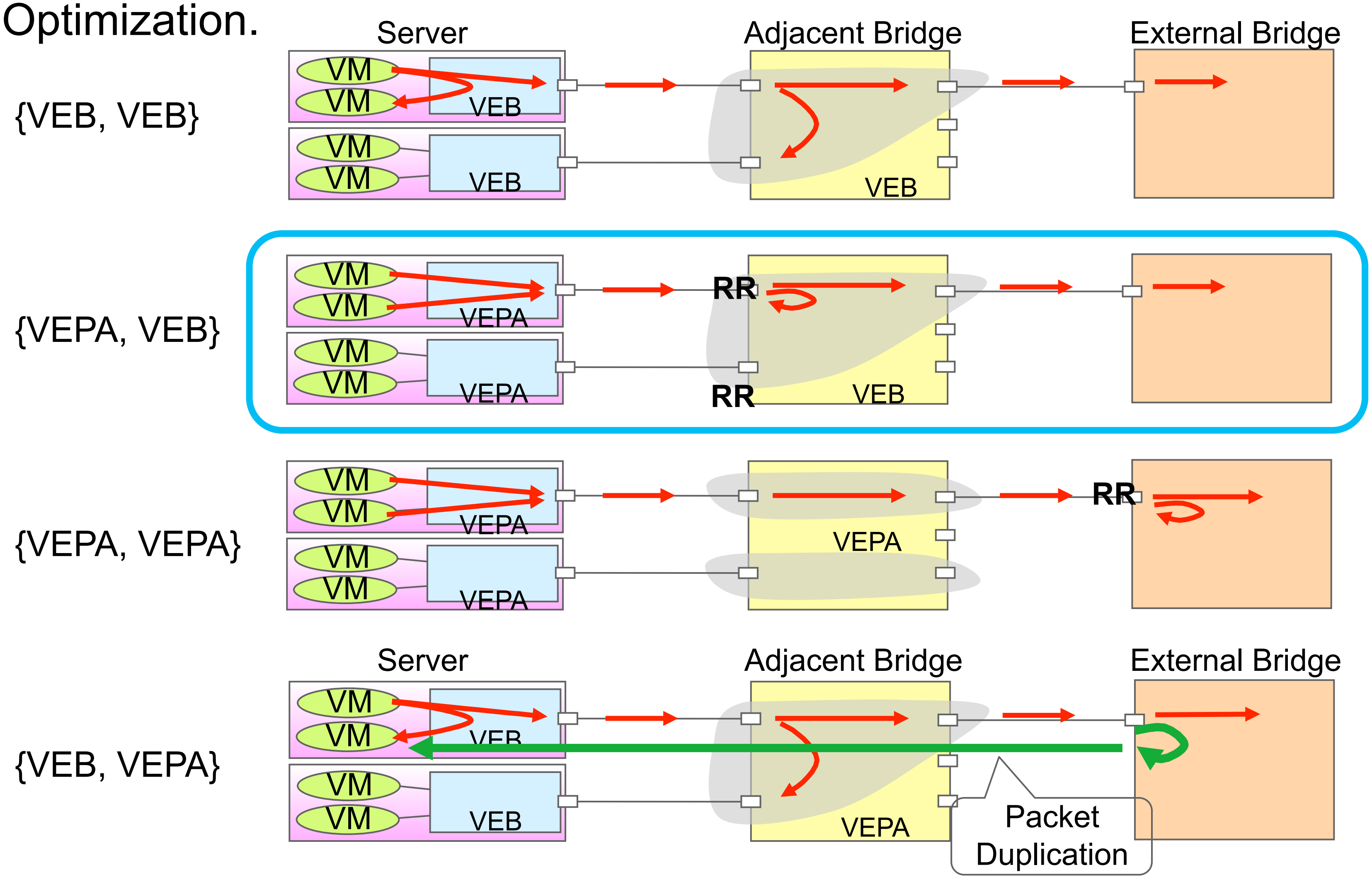
- We selectively forward VDP messages based on internal states that are dynamically configured by processing VDP messages.
  - The forwarding decision of VDP message is made for each VDP TLV type (Pre-Associate, Pre-Associate with Reservation, Associate, and De-associate) . See table II in the proceedings.
  - If we unconditionally forward VDP messages to an upper level switch, AMPP fails and causes unexpected results as shown below.



- To realize VDP forwarding we need to consider:
  - Relationship between Edge Relay Mode of server and that of adjacent bridge
  - Relationship between location of destination and that of source

Local Switching at Adjacent Bridge

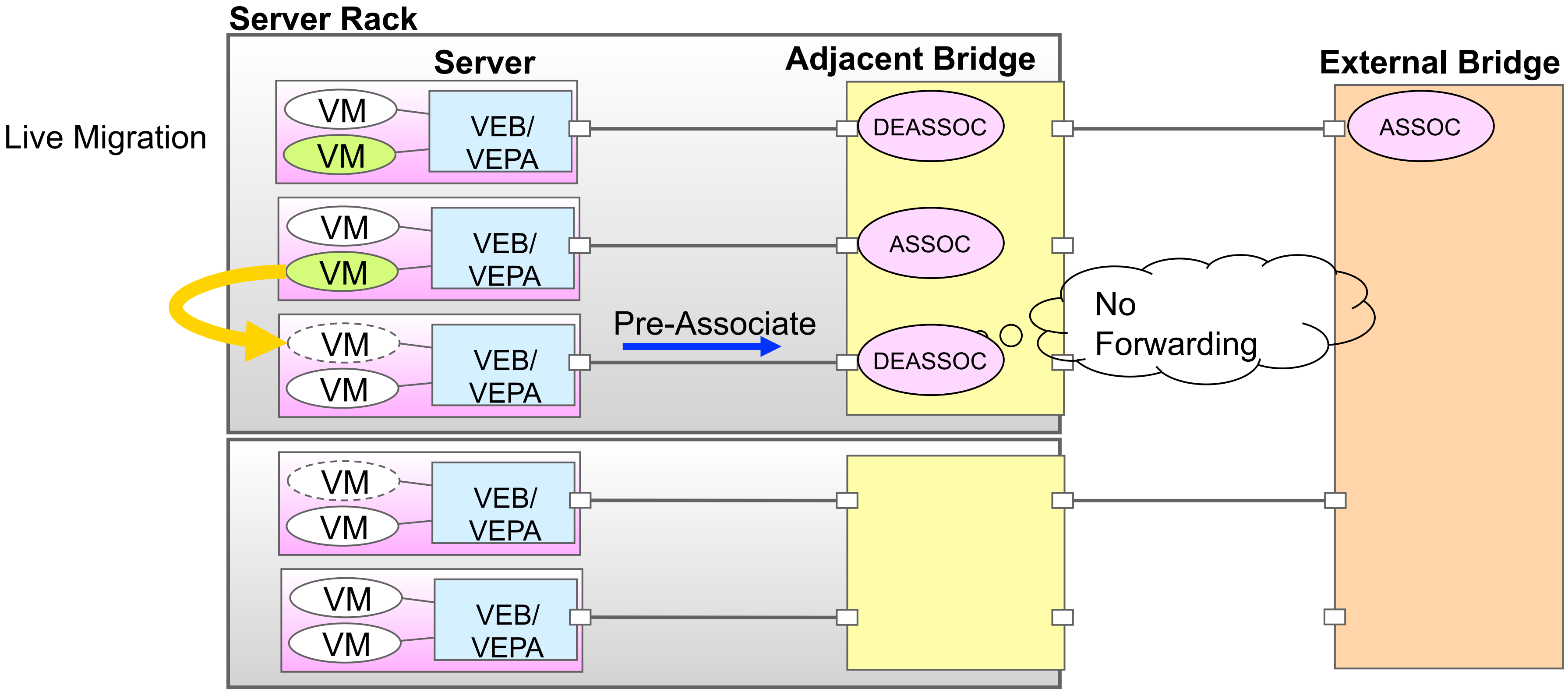
- Combination of Edge Relay Mode
  - Edge Relay Mode {VEPA, VEB} for local switching and performance Optimization.



VDP Forwarding Case 1 (1/3)

Pre-Associate

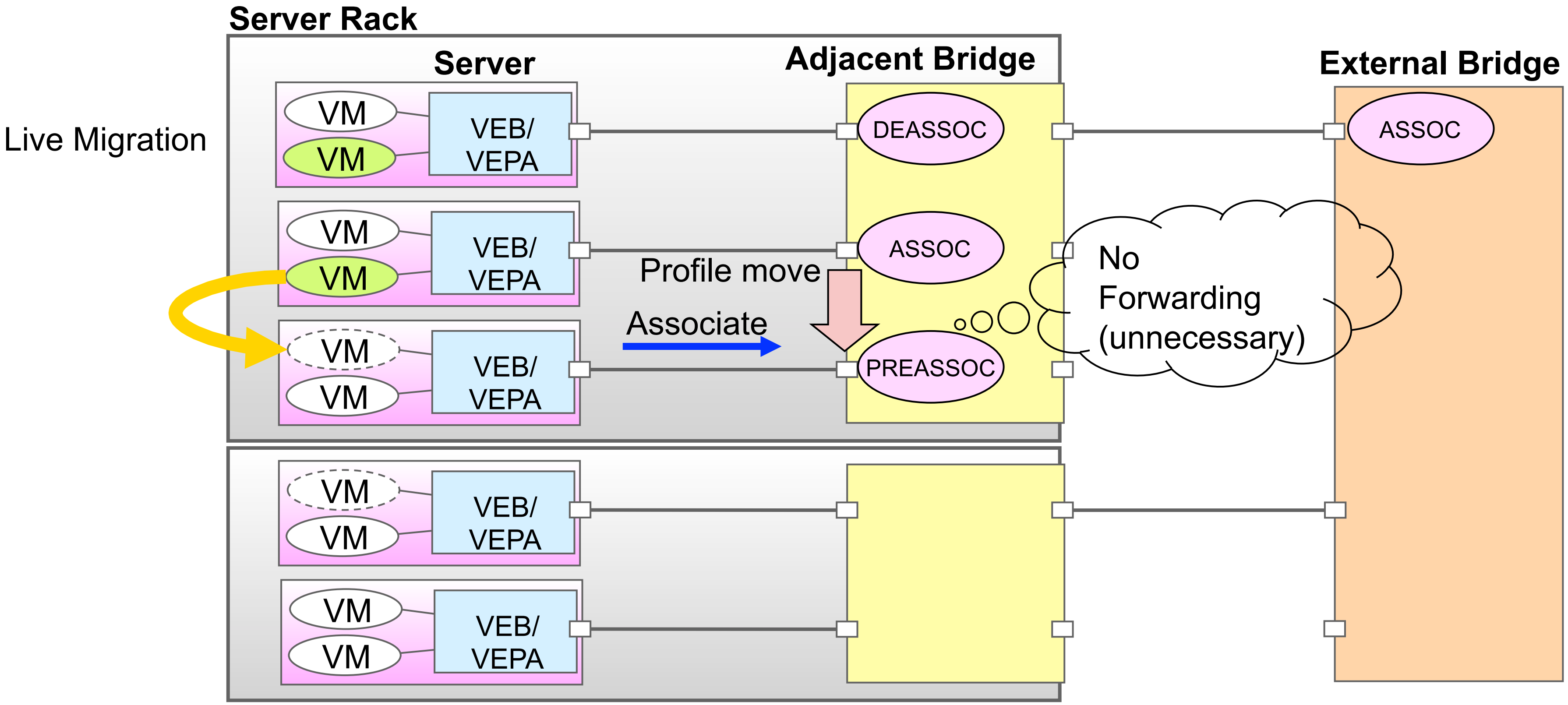
- When a migration is initiated, Pre-Associate message is sent from the destination server to the adjacent bridge at the destination port.
- The vsiState of the reception port (destination port) is DEASSOC and vsiState of another port (source port) is ASSOC, and the bridge does not forward Pre-Associate TLV.



VDP Forwarding Case 1 (2/3)

Associate

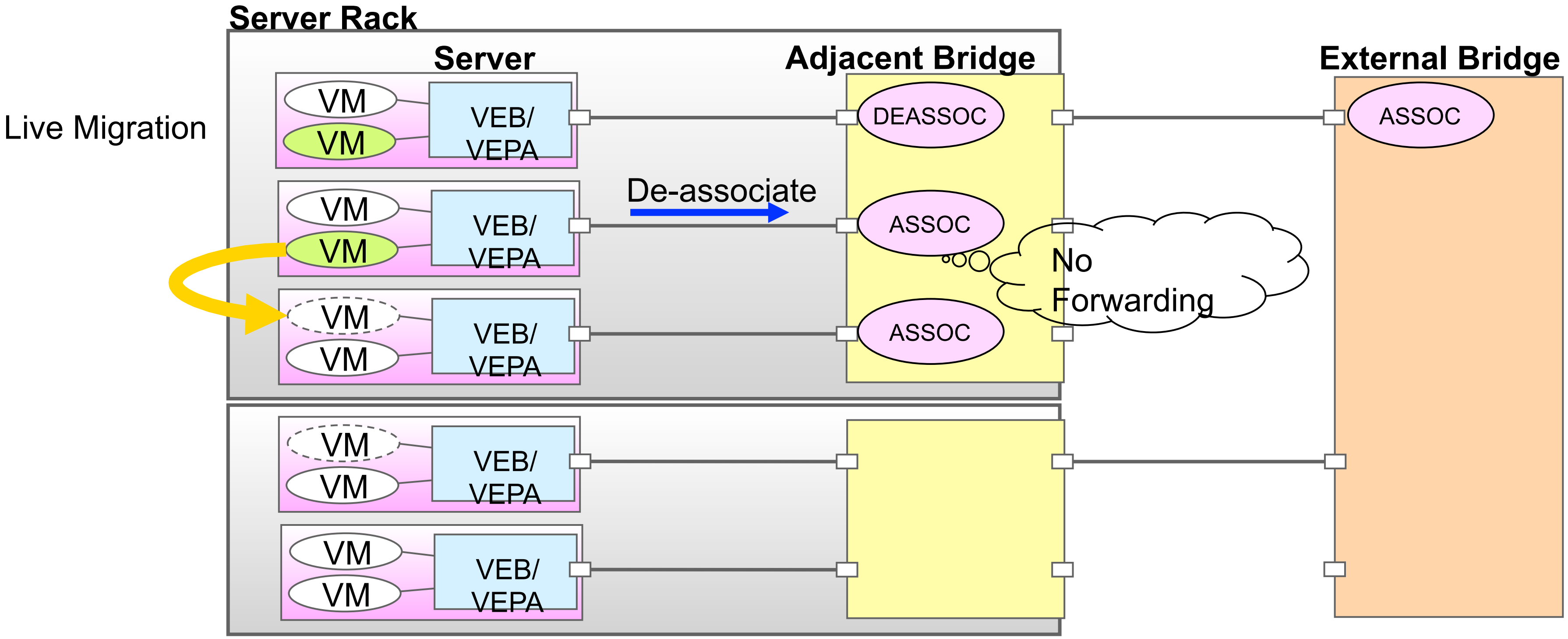
- During the stop and copy phase, Associate message is sent from the destination server to the adjacent bridge at the destination port.
- The adjacent bridge does not forward an unnecessary Associate TLV to the upper ToR switch because vsiState of the upper switch is ASSOC although forwarding of Associate TLV is acceptable.



VDP Forwarding Case 1 (3/3)

De-associate

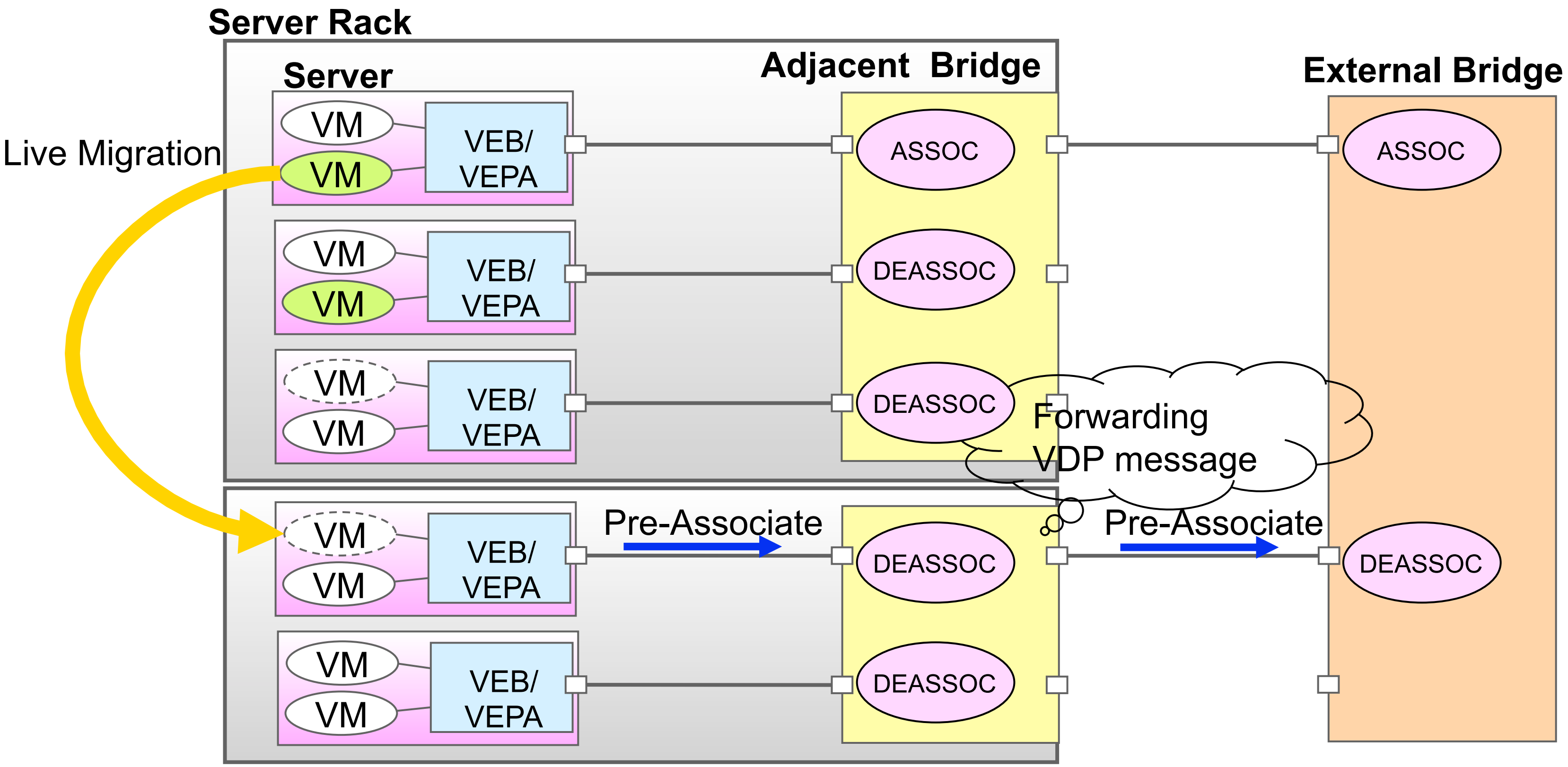
- De-associate message is sent from the source server to the adjacent bridge at the source port.
- The bridge does not forward De-associate message to the upper ToR switch. In this case, forwarding of De-associate TLV is unacceptable because if De-associate message is forwarded, the associate on the destination port is removed while VM is running.



VDP Forwarding Case 2 (1/3)

Pre-Associate

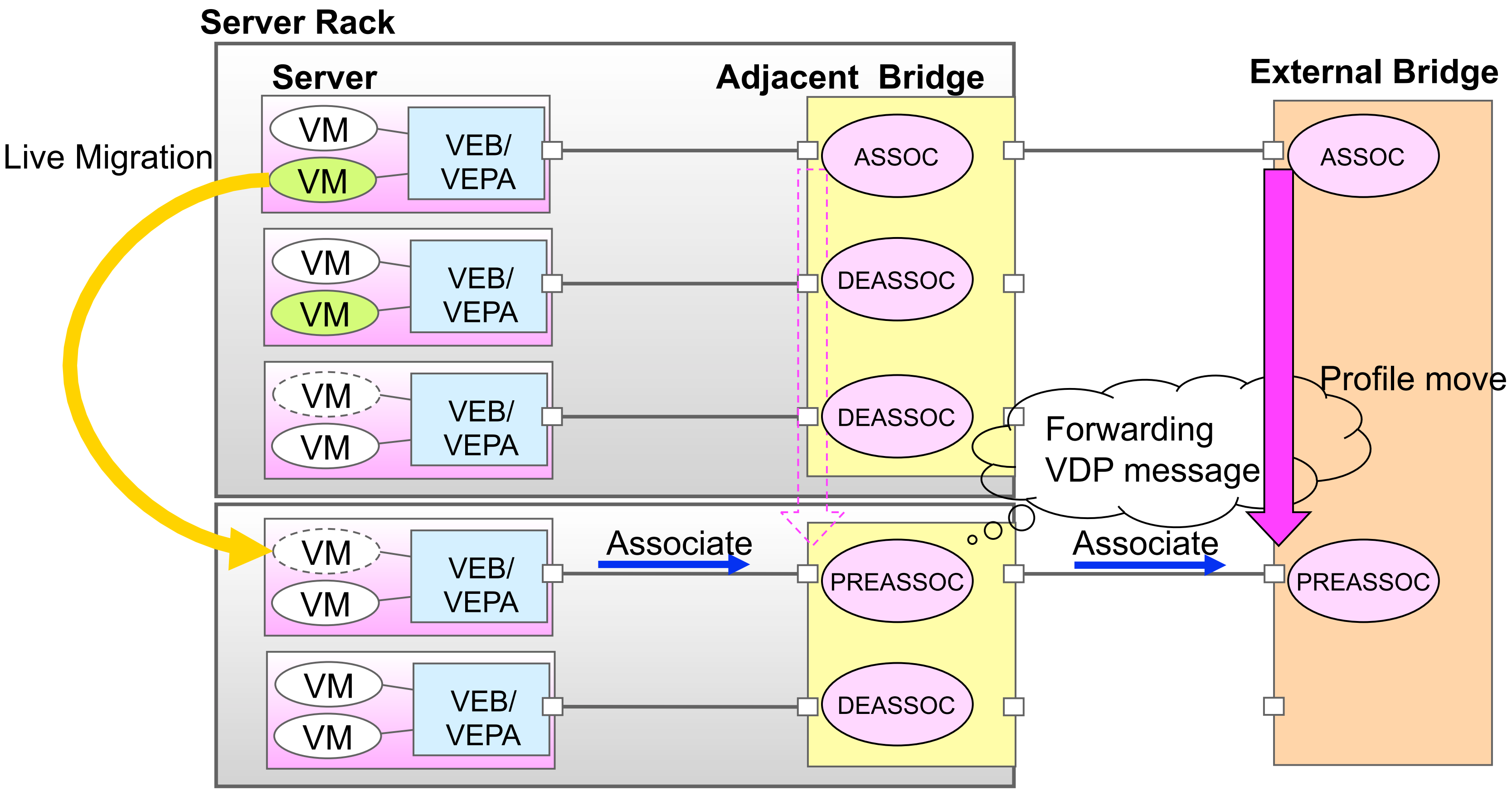
- Pre-Associate message is sent from the destination server to the adjacent bridge at a destination port.
- Pre-Associate message is received when vsiState of the reception port is DEASSOC and vsiState of any other port is DEASSOC, the bridge forwards Pre-Associate TLV.



VDP Forwarding Case 2 (2/3)

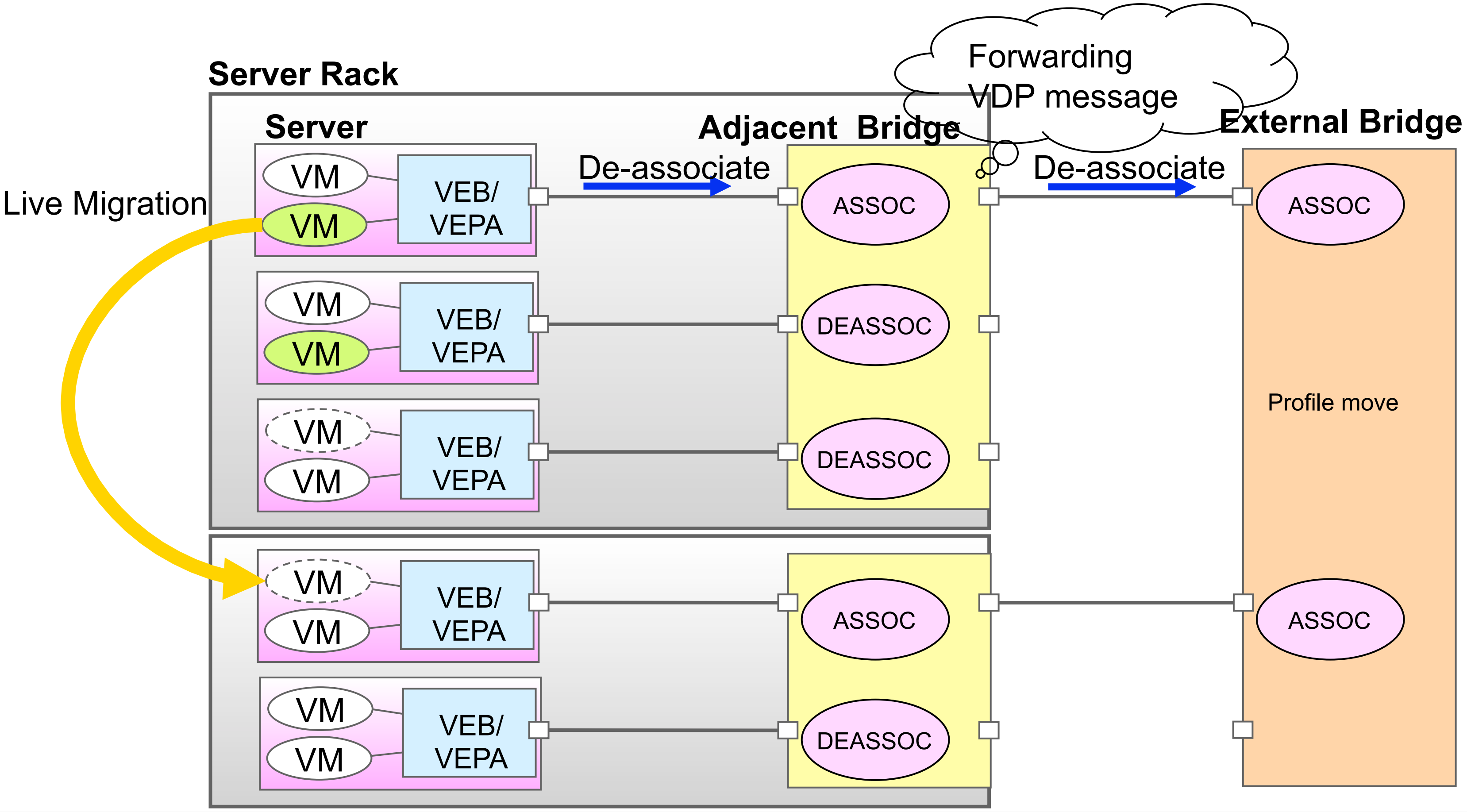
Associate

- During the stop and copy phase, Associate message is sent from the destination server to the adjacent bridge at the destination port.
- The adjacent bridge forwards Associate TLV to the upper ToR switch.



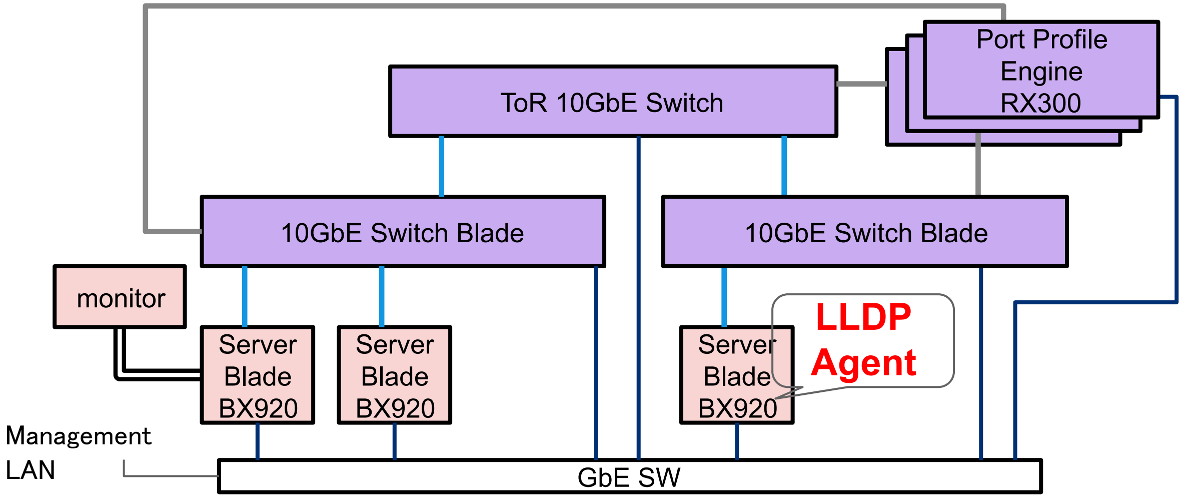
VDP Forwarding Case 2 (3/3)

- De-associate
  - De-associate message is sent from the source server to the adjacent bridge at the source port.
  - The bridge forwards De-associate message to the upper ToR switch.



Prototype of AMPP for Multi-level Switches

- We developed a prototype which consists of
  - Port Profile Engine
    - Standard Protocols: LLDP(EVB TLV), ECP, VDP
    - AMPP for Multi-level switches
  - EVB Packet Analyzer and Visualization tool
- Prototype system
  - Multi-level switch configuration: Switch blades and ToR Switch
  - LLDP Agent with 802.1Qbg patch on the server side

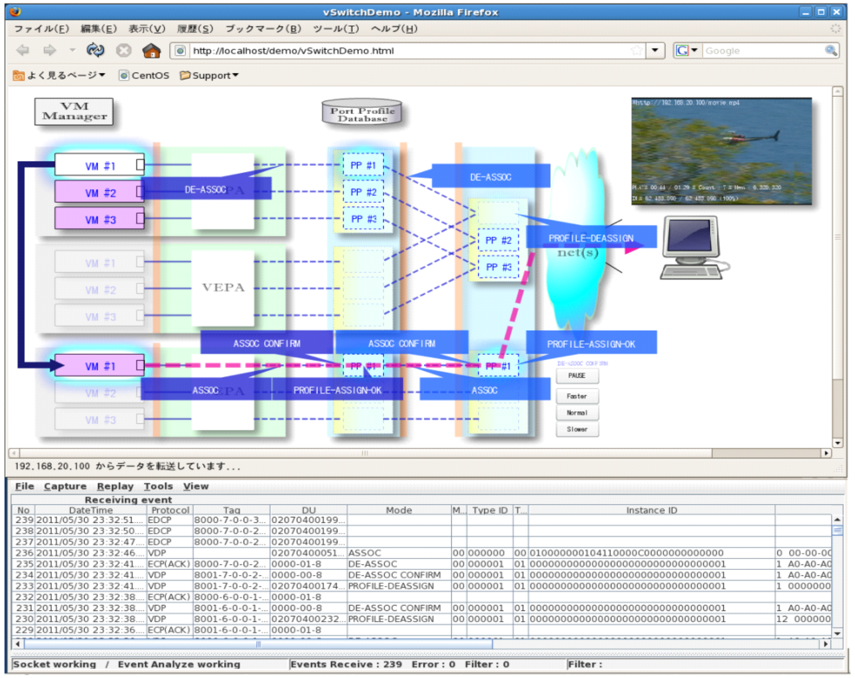


ToR 10GbE Switch

10GbE Switch Blade

Evaluation of AMPP for Multi-level Switches

- Local switching in adjacent bridges for performance optimization
  - As a result of EVB Capability exchanges, Edge Relay Mode {VEPA, VEB} confirmed for local switching and performance Optimization.
- Forwarding of VDP messages
  - Port profiles movement confirmed in a multi-level switch configuration. Visualization of VDP Messages in VM Migration is shown below:



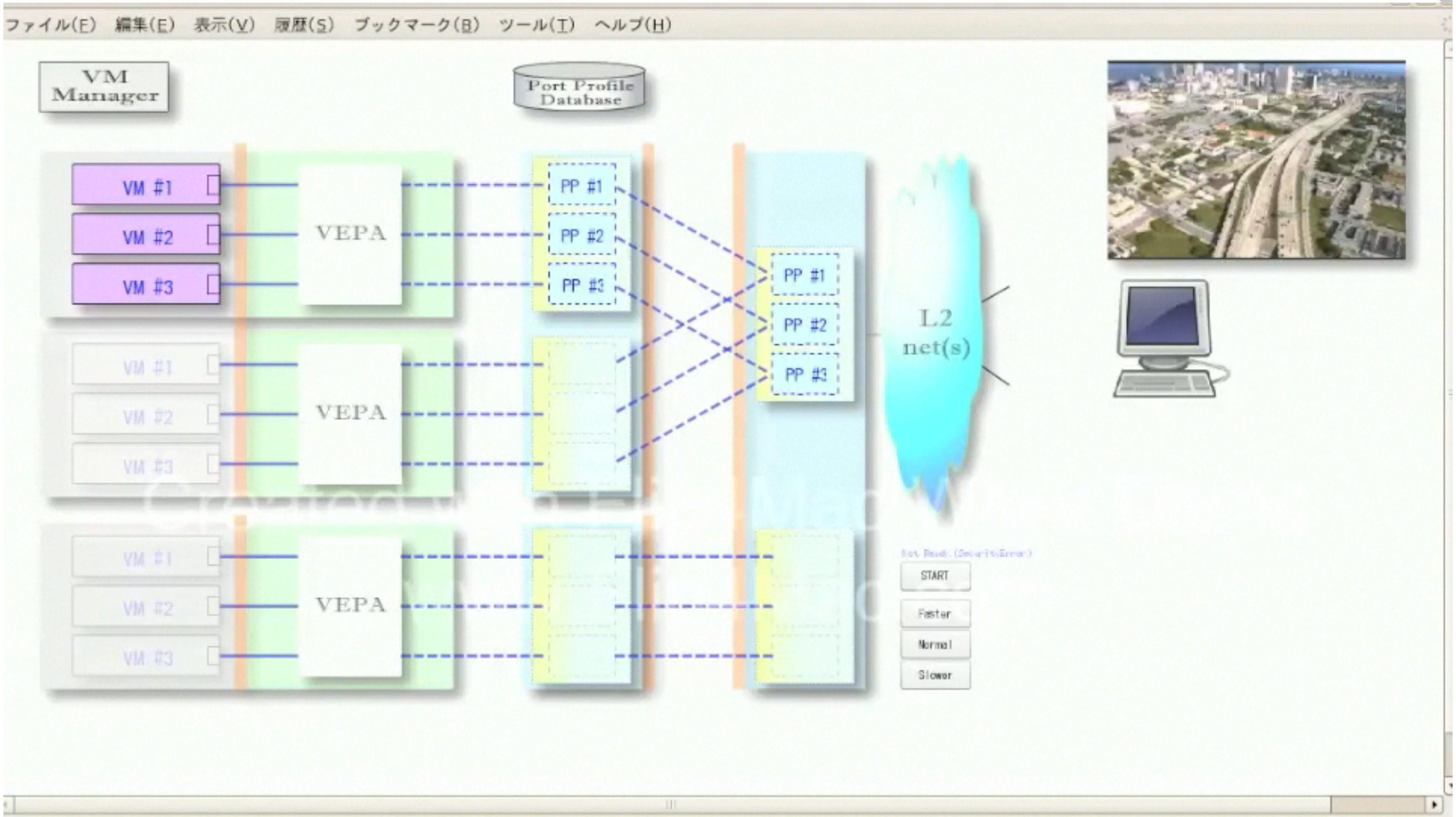
192.168.20.100 からデータを転送しています...

Receiving event

No	DateTime	Protocol	Tag	DU	Mode	M...	Type ID	T...	Instance ID	
239	2011/05/30 23:32:51....	EDCP	8000-7-0-0-3...	02070400199...						
238	2011/05/30 23:32:50....	EDCP	8000-7-0-0-2...	02070400199...						
237	2011/05/30 23:32:47....	EDCP	8000-7-0-0-2...	02070400199...						
236	2011/05/30 23:32:46....	VDP		02070400051...	ASSOC	00	000000	00	0100000001041100000C0000000000000	0 00-00-00
235	2011/05/30 23:32:41....	ECP(ACK)	8000-7-0-0-2...	0000-01-8	DE-ASSOC	00	000001	01	00000000000000000000000000000001	1 A0-A0-A0
234	2011/05/30 23:32:41....	VDP	8001-7-0-0-2-...	0000-00-8	DE-ASSOC CONFIRM	00	000001	01	00000000000000000000000000000001	1 A0-A0-A0
233	2011/05/30 23:32:41....	VDP	8001-7-0-0-2-...	02070400174...	PROFILE-DEASSIGN	00	000001	01	00000000000000000000000000000001	1 0000000
232	2011/05/30 23:32:38....	ECP(ACK)	8000-6-0-0-1-...	0000-01-8						
231	2011/05/30 23:32:38....	VDP	8001-6-0-0-1-...	0000-00-8	DE-ASSOC CONFIRM	00	000001	01	00000000000000000000000000000001	1 A0-A0-A0
230	2011/05/30 23:32:38....	VDP	8001-6-0-0-1-...	02070400232...	PROFILE-DEASSIGN	00	000001	01	00000000000000000000000000000001	12 000000
229	2011/05/30 23:32:36....	ECP(ACK)	8001-6-0-0-1-...	0000-01-8						

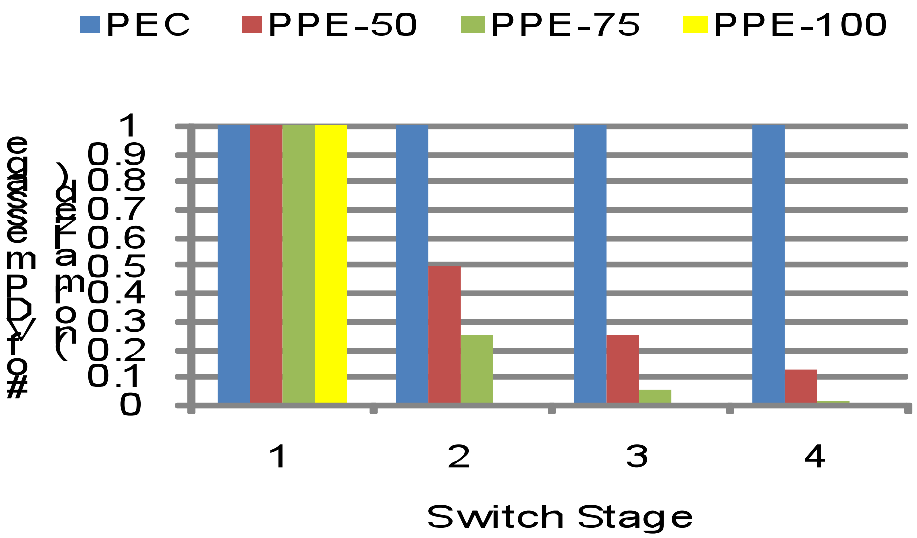
Socket working / Event Analyze working | Events Receive : 239 Error : 0 Filter : 0 | Filter :

Visualization of VDP Messages in VM Migration



Amount of VDP Messages at Root Switch

- Comparison between Port Extension and Port Profile Engines
  - Port Extension in concept (PEC) where all messages are always forwarded and processed in the most upper switch.
  - Port Profile Engine (PPE) with locality as a parameter
    - PPE-100 means 100% of locality where VDP messages are processed in the first stage only. PPE-75 means 75% of VDP messages are processed in the first stage locally and 25 % of messages are forwarded to an upper switch.



Amount of VDP messages at root switch is small and the root switch is not bottleneck.

Conclusion

- Proposed AMPP for multi-level switches
  - Automatically configure non-adjacent external bridges in addition to adjacent bridges using the standard protocol between bridge and bridge.
  - Any standard compliant EVB bridge can be used as an external bridge.
- Developed Prototype of AMPP for multi-level switches
  - Port Profile Engine
    - Standard Protocols: LLDP(EVB TLV), ECP, VDP
    - AMPP Extension for Multi-level switches
  - EVB Packet Analyzer and Visualization tool
- Confirmed AMPP operations in multi-level switch configuration
  - Local switching in adjacent bridges for performance optimization
  - Forwarding of VDP messages

Thank You !

FUJITSU

shaping tomorrow with you