FUJITSU

shaping tomorrow with you

A Management Method of IP Multicast in Overlay Networks using OpenFlow

August 13, 2012

Yukihiro Nakagawa
Fujitsu Laboratories Ltd.

Outline

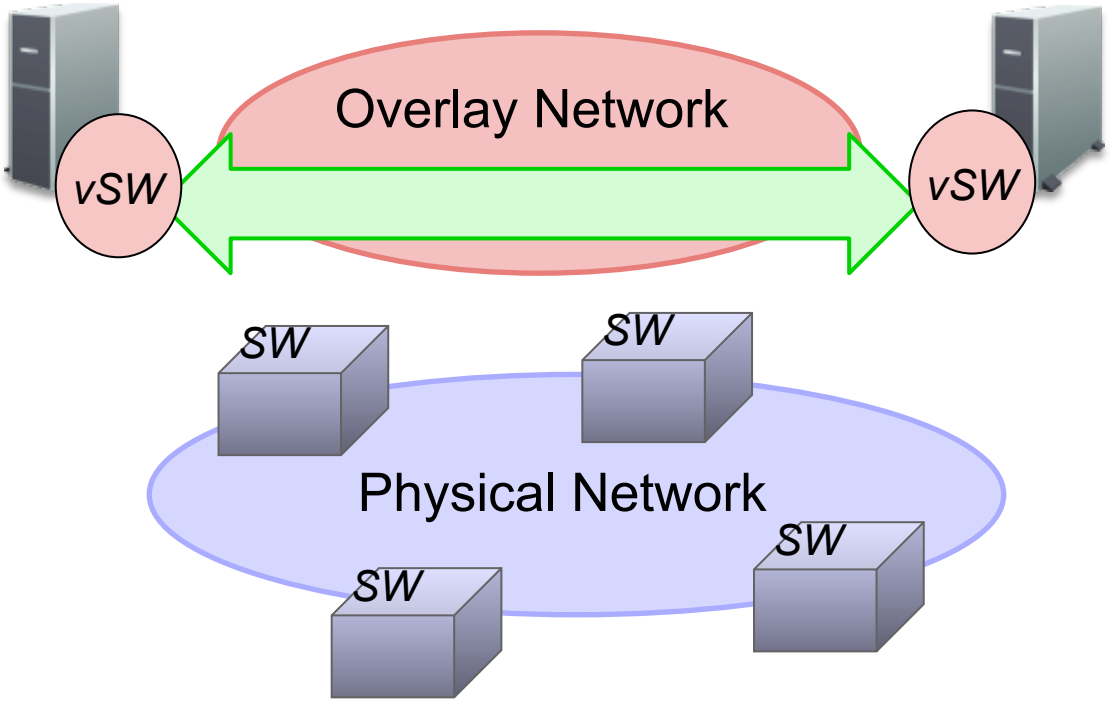
- Overlay Networks and VXLAN
- Challenges in Scalability
- Management of IP Multicast using OpenFlow
- Prototype of VXLAN Environment
- Conclusion

VXLAN: Virtual eXtensible LAN

Overlay Networks and VXLAN

FUJITSU

- Overlay networks
  - Network virtualization is important to realize dynamic infrastructure
  - Overlay networks stretch Layer 2 networks and increase mobility of VMs
- Virtual eXtensible LAN (VXLAN)
  - VXLAN is one of overlay networking technologies being defined in IETF
  - 16 Million overlay networks can be defined for multi-tenancy



Broadcast Domain Mapping in VXLAN

FUJITSU

- VXLAN is one of IP Multicast applications
  - Multiple broadcast domains are mapped into IP Multicast space
  - Multicast trees are dynamically configured by IGMP
  - VXLAN can be used over either Layer 2 or Layer 3

- Our use case is Layer 2 network at server edge
  - Efficient multicast replication can be achieved using IGMP snooping
  - Layer 2 switches prune the multicast trees using IGMP messages

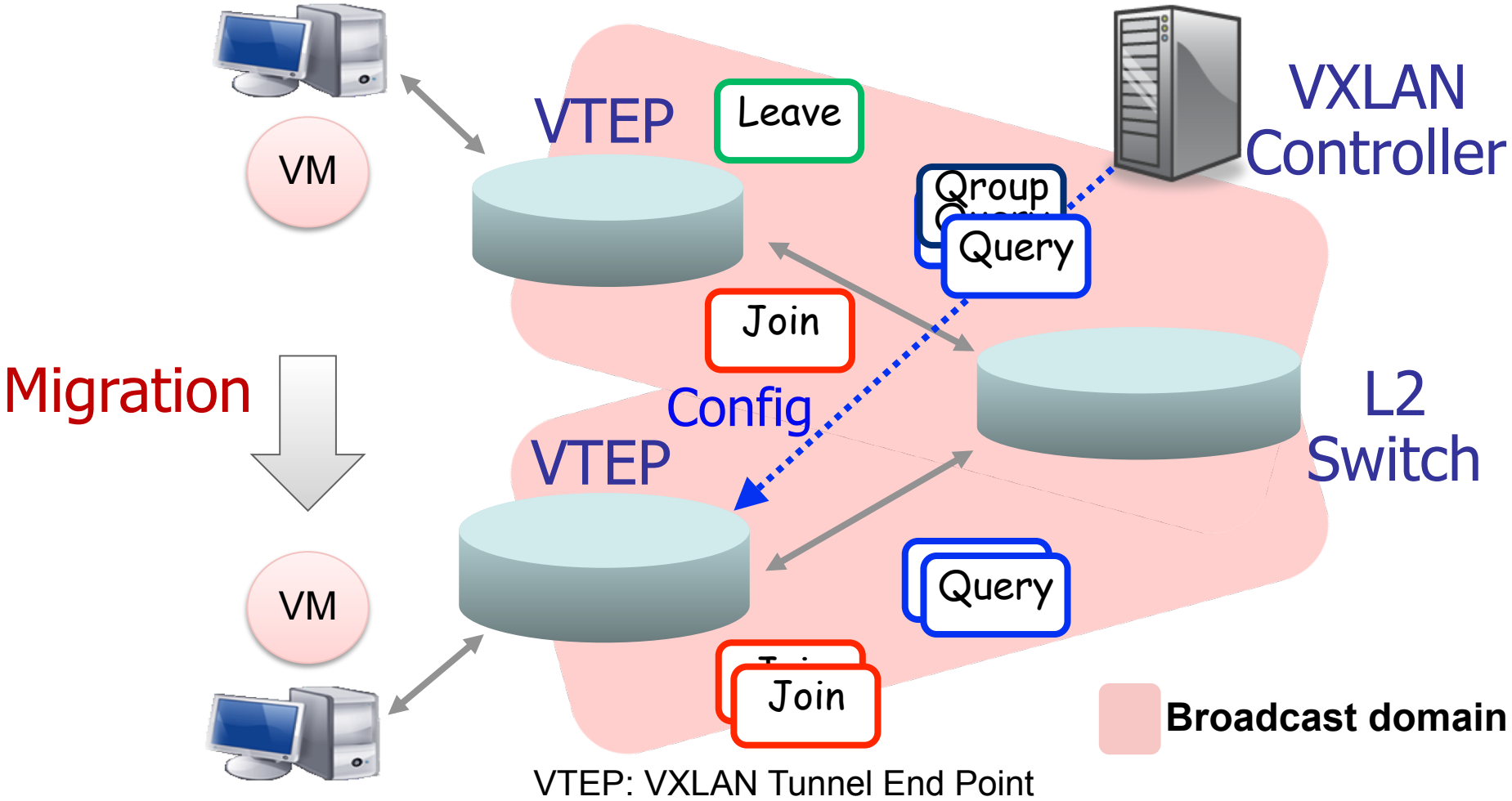
Layer 2 switch requirements for multiple broadcast domains

	Broadcast Domain	L2 Switch Requirement
Traditional Layer 2 Network	VLAN	802.1Q VLAN
VXLAN Overlay Network	IP Multicast	IGMP Snooping

Example of Multicast Path Mgmt

FUJITSU

- IGMP Join/Leave messages to configure broadcast domains



➡ Switch's IP Multicast capability is important

Limitation of VXLAN Scalability

FUJITSU

- Issue in applying VXLAN as it is to Layer 2 network
  - Number of BC domains is limited by number of IP Multicast groups
  - IGMP protocol is processed by local CPU on the switch
    - # of IGMP messages becomes bottleneck

- Examples of 10GbE switch implementation

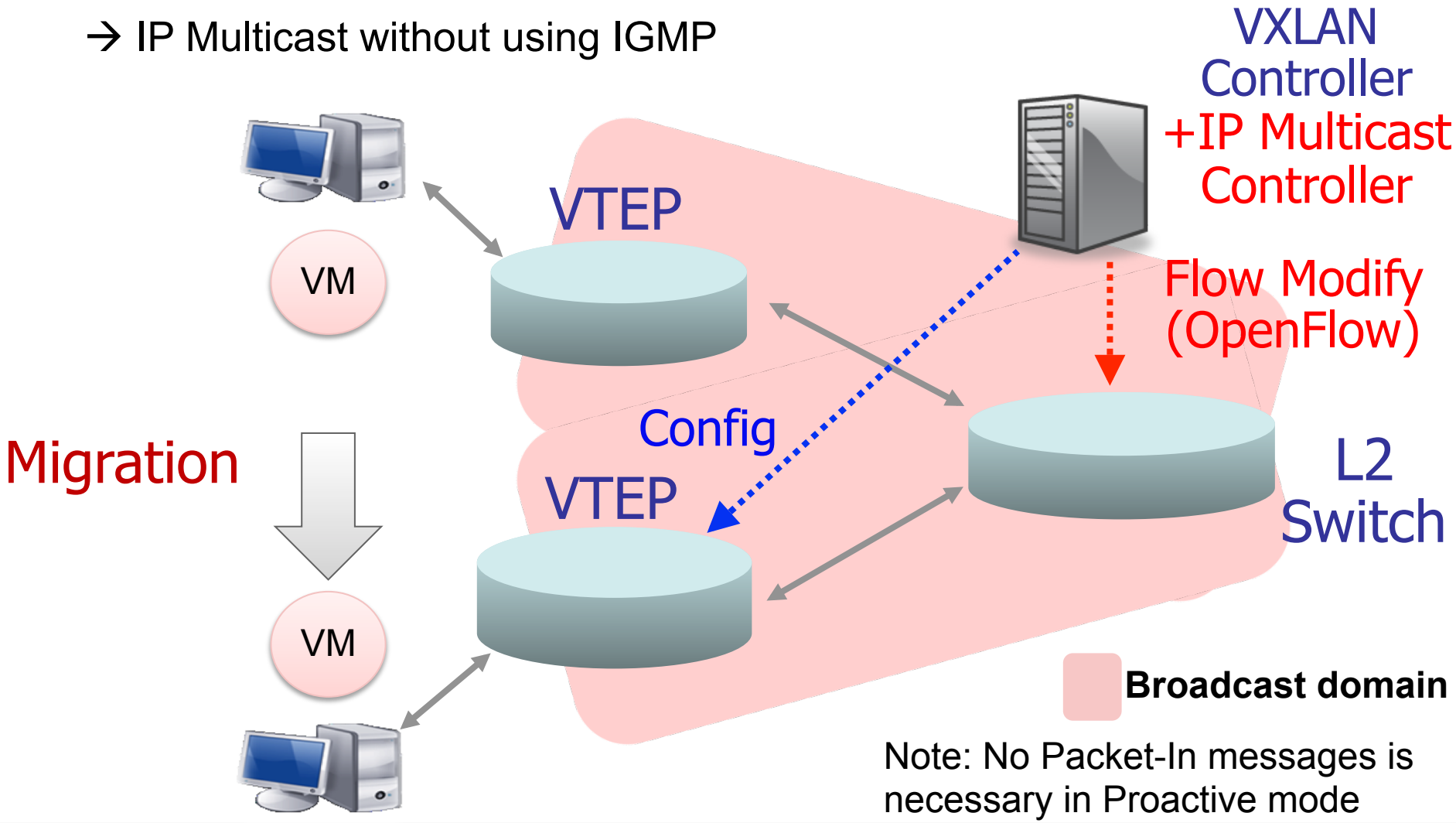
	Vender A	Vender B	Vender C	Vender D
Switch capability	1.28Tbps	1.28Tbps	1.28Tbps	1.28Tbps
VLAN table	4k	4k	4k	4k
Mac table	128K	128k	128K	120K
L3 routing table	16K	8K	16K	8k
Multicast table	4K	8K	8K	3.5K

➡ We would like to make VXLAN more scalable

Proposal: IP Multicast with OpenFlow

FUJITSU

- Configure IP Multicast path by Central controller
  - Use MAC Table and set it up proactively using OpenFlow
    → IP Multicast without using IGMP



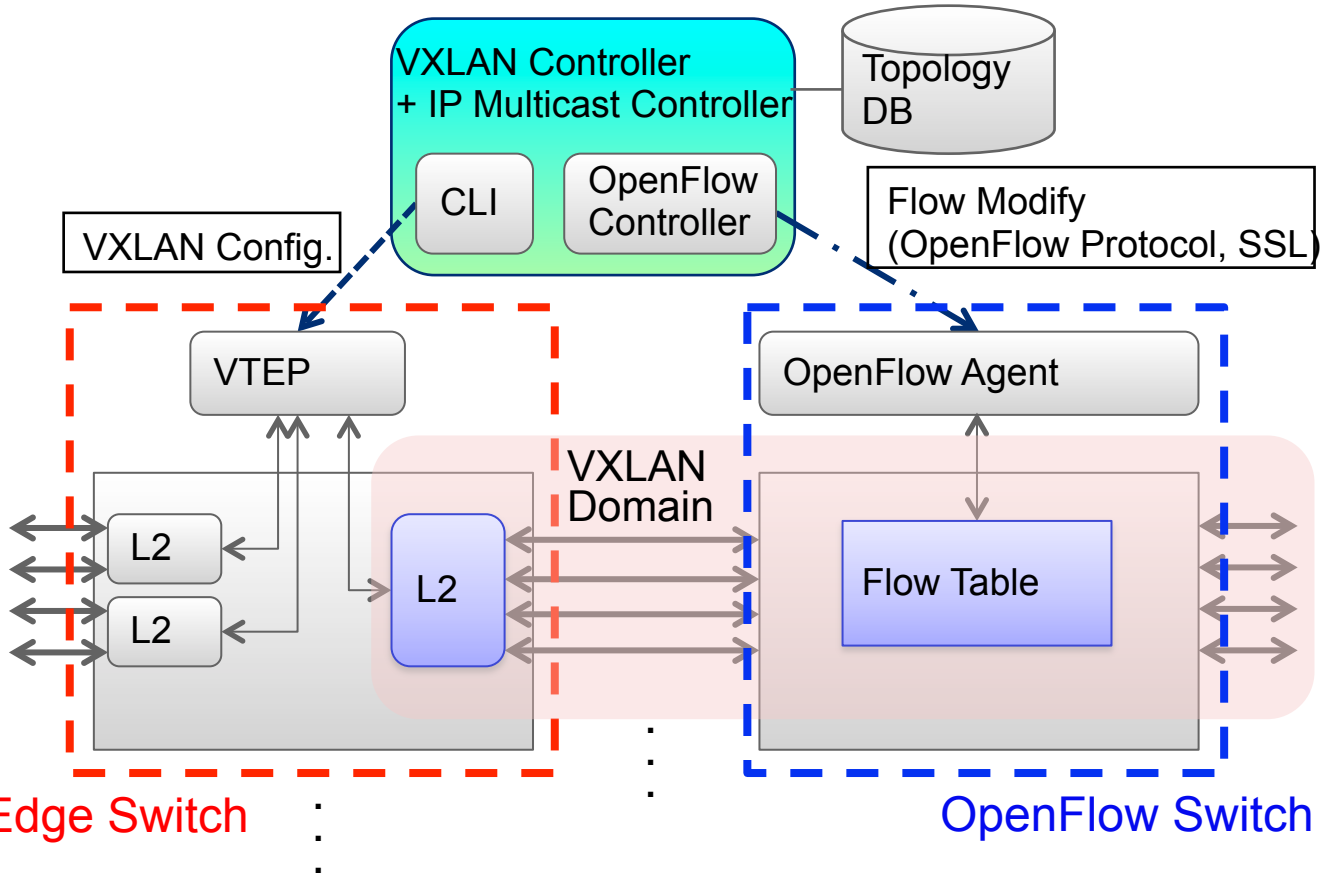
Note: No Packet-In messages is necessary in Proactive mode

Prototype of VXLAN Environment

FUJITSU

- We developed prototype which consists of:
  - VXLAN controller + IP Multicast Controller (Network Manager)
  - Edge switch with VTEP (VXLAN gateway)
  - OpenFlow switch

Our 10GbE switch used for Edge and OpenFlow switches



Using OpenFlow in our Environment

FUJITSU

- Flow table forwarding model based on OpenFlow V1.0
  - The flow table for IP multicast is mapped to the MAC address table
- Extension of output action for IP multicast
  - Output port vector is included to specify multiple destinations (see below)

```
struct ofp_action_output_vendor {
 uint16_t type;        /* OFPAT_VENDOR. */
 uint16_t len;         /* Length is 16. */
 uint64_t portvec;     /* Output port vector. */
 uint8_t pad[4];       /* Pad to 64 bits. */
};
OFP_ASSERT(sizeof(struct ofp_action_output_vendor) == 16);
```

➡ This is vender specific. We think multicast packet handlings should be included in OpenFlow spec for interoperability

Summary of Control Message Reduction

FUJITSU

- Bottleneck of control plane was removed in our environment
  - We compared number of messages (See Table 3 in the proceedings)
    - Examples of # VTEP=1K and IGMPv2 are shown below

	Periodical messages			Non-periodical messages
	# of broadcast domains	# of messages per second	Processing time per message	# messages per VM move
IGMP	1K	0.1M Join	10 us	1 Leave + 1 Group-specific query + 1K join
	8K	0.8M Join	1.25 us	
	16M	1.6G Join	0.625 ns	
Our Proposal	16M	None	n/a	2 Flow Modify

Note: Actual # of broadcast domains is limited by hardware resource or table size

In our application, the central controller configures IP Multicast. We achieved more than 4K broadcast domains in our network

Conclusion

FUJITSU

- Proposed IP Multicast with OpenFlow for Overlay Network
  - IP Multicast path is configured by the central controller
  - Use MAC Table for IP Multicast and set it up proactively using Flow Modify
  - Enhance output action to specify multiple output ports for IP Multicast
- Prototype of VXLAN Environment
  - VXLAN controller + IP Multicast Controller (Network manager)
  - Edge switch with VTEP (VXLAN gateway)
  - OpenFlow switch
- Confirmed our proposed method
  - Removed bottleneck of control plane by eliminating IGMP messages
  - Achieved more than 4K broadcast groups in our network at server edge

FUJITSU

shaping tomorrow with you