Development of A Scalable Non-IP/Non-Ethernet Protocol With Learning-based Forwarding Method

Yasusi Kanada (Hitachi, Ltd.)
Akihiro Nakao (University of Tokyo / NICT*)

* National Institute of Information and Communications Technology

Introduction

- **IP was originally intended to *keep it simple and stupid* (*KISS*), but it has become too complex.**
  - Especially, the combination of IP and Ethernet causes complexity.
- **In a NwGN project, we have developed a new protocol called the *IP Ether Chimera* (*IPEC*).**
  - In Japan, several projects towards new-generation networks (NwGN) have been conducted.
  - This research is intended to be the first step toward development of new protocols that will replace the combination of IP and Ethernet.

Complexity of IP/Ethernet

- **IP/Ethernet: A Popular Combination of IP and Ethernet**
  - Originally, this combination was strongly necessary because Ethernet could be used only in LAN.
  - Now, both can be used in WAN.
  - However, this combination is still the most popular.
- **Both IP and Ethernet frames contain addresses: IP and MAC addresses — *redundant!***
  - MAC addresses were originally unchangeable, but now they can be configured by software.
  - Both types of addresses are used for networking (L3 function)!
- **Complexity caused by this combination**
  - ARP is required to find the corresponding MAC address from an IP address.
    - ARP = Address Resolution Protocol.
  - Other similar protocols are required in other situations: RARP, NDP (IPv6).

IPEC — Proposed Protocol

- **We propose a new experimental non-IP protocol called IP Ether Chimera (IPEC).**
  - We intend to remove the complexity caused by IP/Ethernet by IPEC.
- **IPEC (design) consists of *refined components* of IP and Ethernet.**
  - We reuse part of Ethernet (and IP) to build a new protocol.
- **IPEC implementation consists of *refined components* in hardware and software.**
  - We reuse Ethernet LAN cards and drivers (L2 functions) to implement a new protocol.
  - We program networking (L3) functions (currently using slow-path).

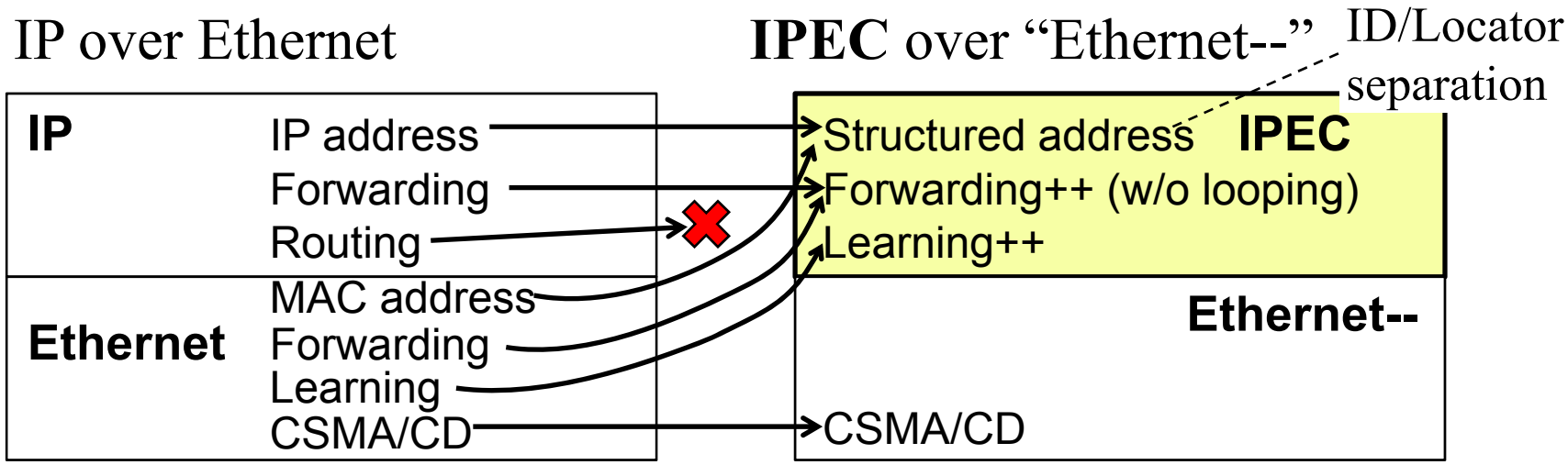
Development Goals of IPEC

- **To implement a simpler *forwarding function* that can handle hierarchical addresses and networks including loops.**
- **To establish a *learning algorithm* that can be used in arbitrarily structured networks including loops.**
  - This algorithm is to be achieved by extending the address learning algorithm of Ethernet switches.
- **To show that a network using virtualization nodes can be used to develop and run non-IP protocols.**

Protocol Design Policies

■ Use of structured addresses and learning

- ◆ Ordered and structured addresses similar to IP addresses are used.
- ◆ However, to keep the protocol simple, no routing protocol is to be introduced, and packets are forwarded by an Ethernet-like but extended learning algorithm.

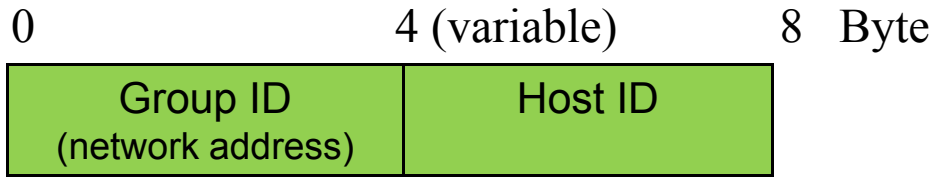


■ Applicability to ID/Locator Separation

- ◆ Addresses with two components, which may be interpreted as an ID and a locator may be used.

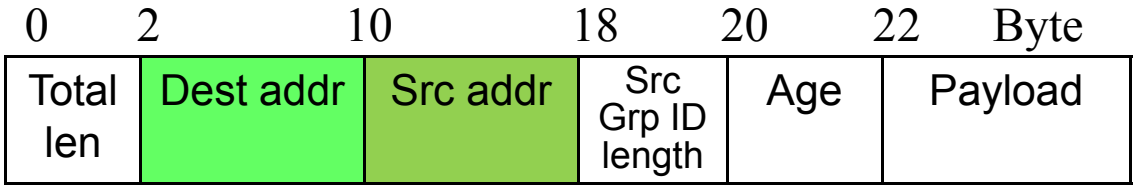
Address- and Protocol-Formats of IPEC

■ Address format



- ◆ **Host ID**: Atomic identifier. 4 bytes in the current implementation.
- ◆ **Group ID**: An identifier of a group of hosts. It may be structured.

■ Packet format



- ◆ **Age**: Something like TTL in an IP packet and is used to avoid packet looping.

Learning Algorithm of IPEC

■ Enhanced version of learning algorithm in Ethernet

Ethernet-like learning, but to learn only groups (thus, more scalable)

if source group of P is not registered in the forwarding table then
Register group, group length, input port, age of P
to the forwarding table (learn the packet);

To learn groups when the new path is shorter than the older one or when the record is too old (should be replaced).

else if **age of E > age of P** or
E is in **registration timeout** status then
age of E = age of P; port of E = port of P;
timestamp of E = current time (ns);

Duplicated packets are discarded (thus a loop is allowed)

else if **age of E < age of P** or
port of E != port of P then
Drop the packet (the forwarding procedure is not applied);
else timestamp of E = current time (ns);

Forwarding algorithm of IPEC

- **Mostly the same forwarding algorithm as Ethernet**

Flud (broadcast) when not yet learned (or forgot)

if destination group of P is not registered in the forwarding table or
E is in **reference timeout** status then
Flood the packet that is a copy of P but the age is incremented;

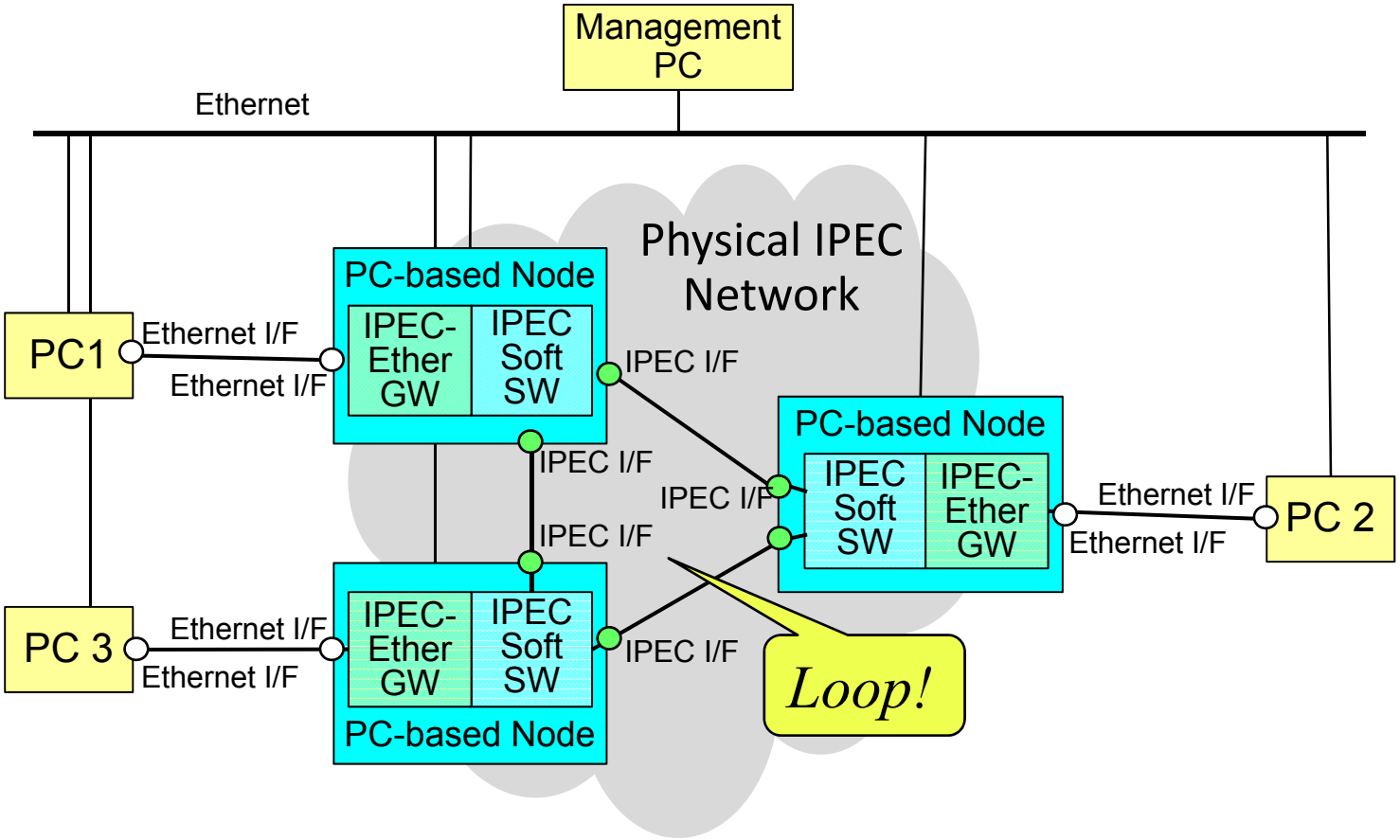
Output to learned port when already learned

else
Output the packet that is a copy of P but the age is
incremented to the port specified in the registered element;

Experiment on a LAN

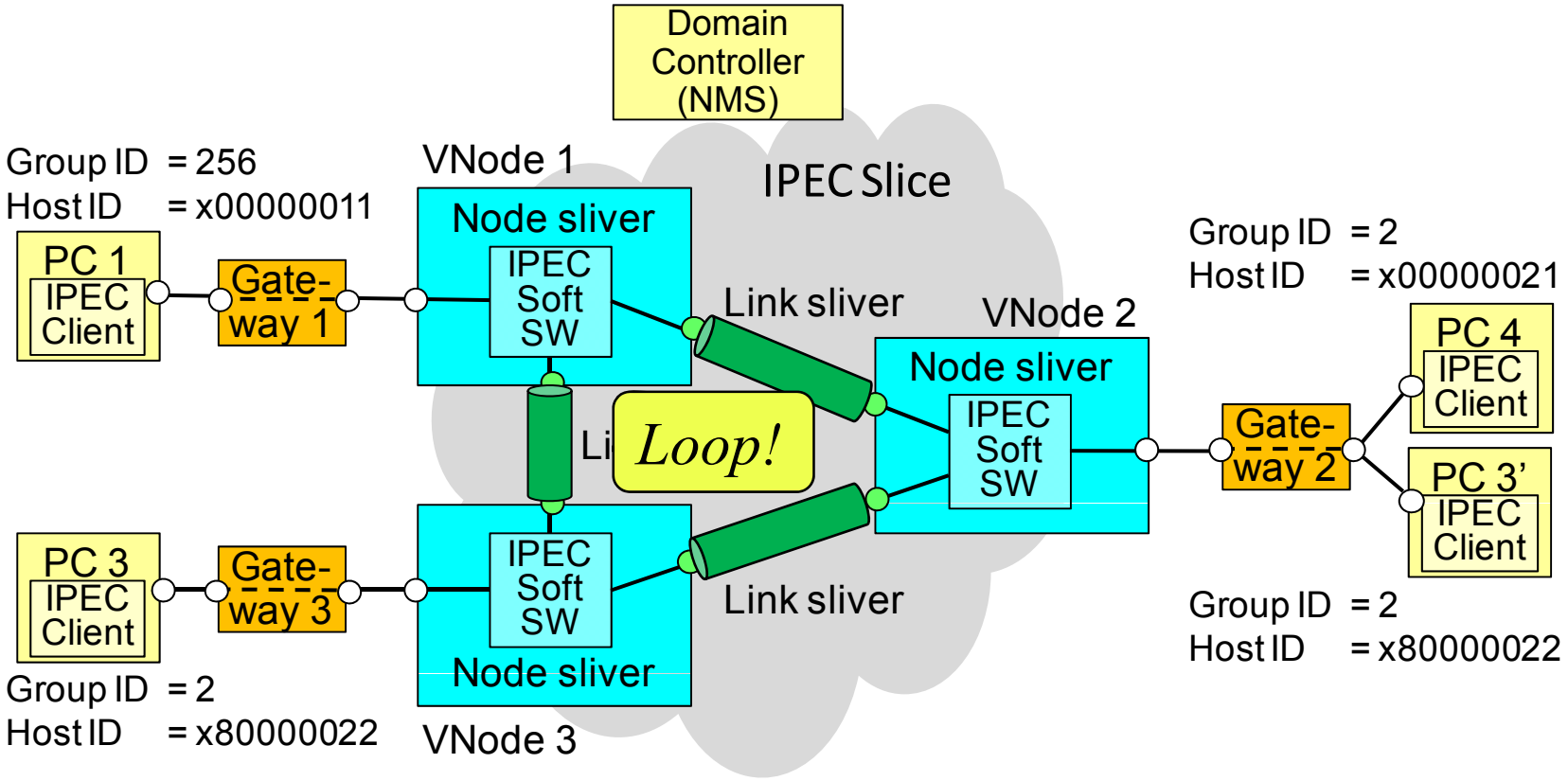
■ IPEC has been implemented on a PC-based network.

◆ LAN cards are used for Ethernet--.



Experiments on A Virtualization Platform

- **IPEC has been implemented to the virtualization platform developed by the Virtualization Node Project (VNP).**
  - VNP has developed “VNode” (the virtualization node).



Measurements and Wide-area Experiments

■ Performance measurement

- ◆ The packet loss rate is less than 0.1% (2-Mbps UDP traffic). The performance is better than in the LAN environment.
- ◆ We have also measured round-trip time using a ping command. It is 2.8 ms on average.

■ Wide-area experiments and demos

- ◆ At Interop Tokyo 2010, two VNodes in Makuhari and one VNode in Hakusan. At Interop Tokyo 2011, three VNodes in Makuhari, Mejirodai, and Hakusan.
- ◆ At the 8th GENI Engineering Conference (GEC8), Nakao have introduced IPEC as an example application of the virtualization platform, and posted the demonstration video on the Web.

Summary and Conclusion

- **An “L3” protocol called IPEC has been developed.**
  - IPEC is a combination of refined protocol components derived from Ethernet and IP.
  - The implementation of IPEC reuses refined hardware and software components of Ethernet.
- **Features of IPEC**
  - IPEC is more scalable than Ethernet, and a mobile group can be more efficiently learned.
  - IPEC can be applied to a network with a loop.
  - Group IDs can be used as locators.
- **IPEC works well both in a native LAN and in a VN environments.**
- **IPEC is less scalable and less universal than IP, but IPEC or its successor may be used in small-scale WAN such as virtual networks.**