How Path MTU Discovery not Work

Masataka Ohta
Tokyo Institute of Technology
mohta@necom830.hpcl.titech.ac.jp

Abstract

- Multicast path MTU discovery (PMTUD) is a new feature of IPv6. However, ICMP implosion with multicast PMTUD can be serious when most MTU bottlenecks are located near individual receivers. ICMP Packet Too Big, at least those generated against multicast packets, will be filtered, which is a standard violation, which means there is no reason not to filter unicast ones. Thus, unicast PMTUD is not expected to work. We should not send packet >1280B, except for IP over IP tunnels.

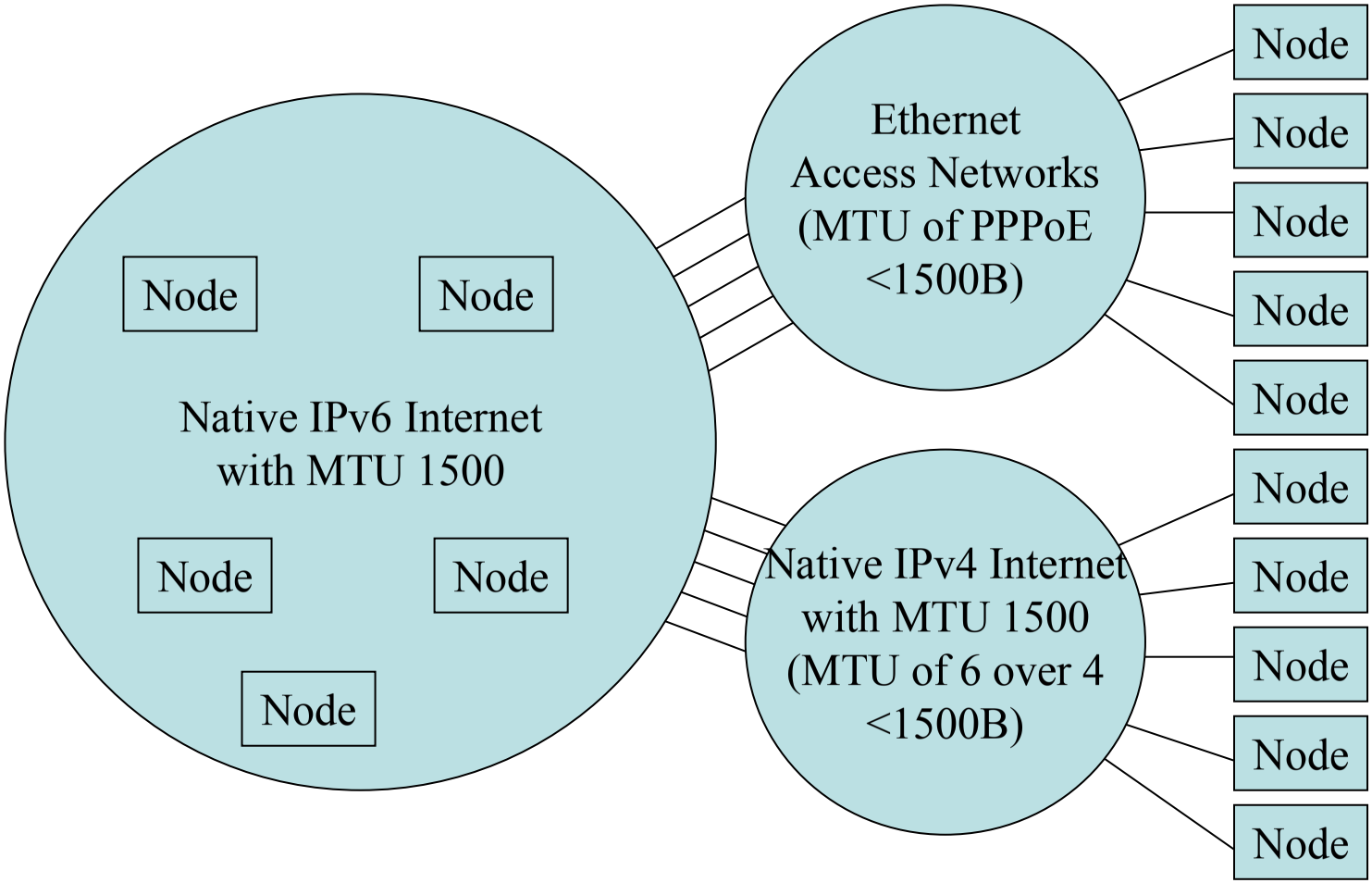
PATH MTU Discovery

- Measure Path MTU by ICMP Packet Too Big
  - Path MTU is set to the value contained in the ICMP packet
  - does not work if ICMP Packet Too Big is filtered or not generated
- Periodically send larger packet to detect MTU increase by path change
- “SHOULD be supported” (node requirement)
  - ISPs SHOULD NOT filter ICMP Packet Too Big?

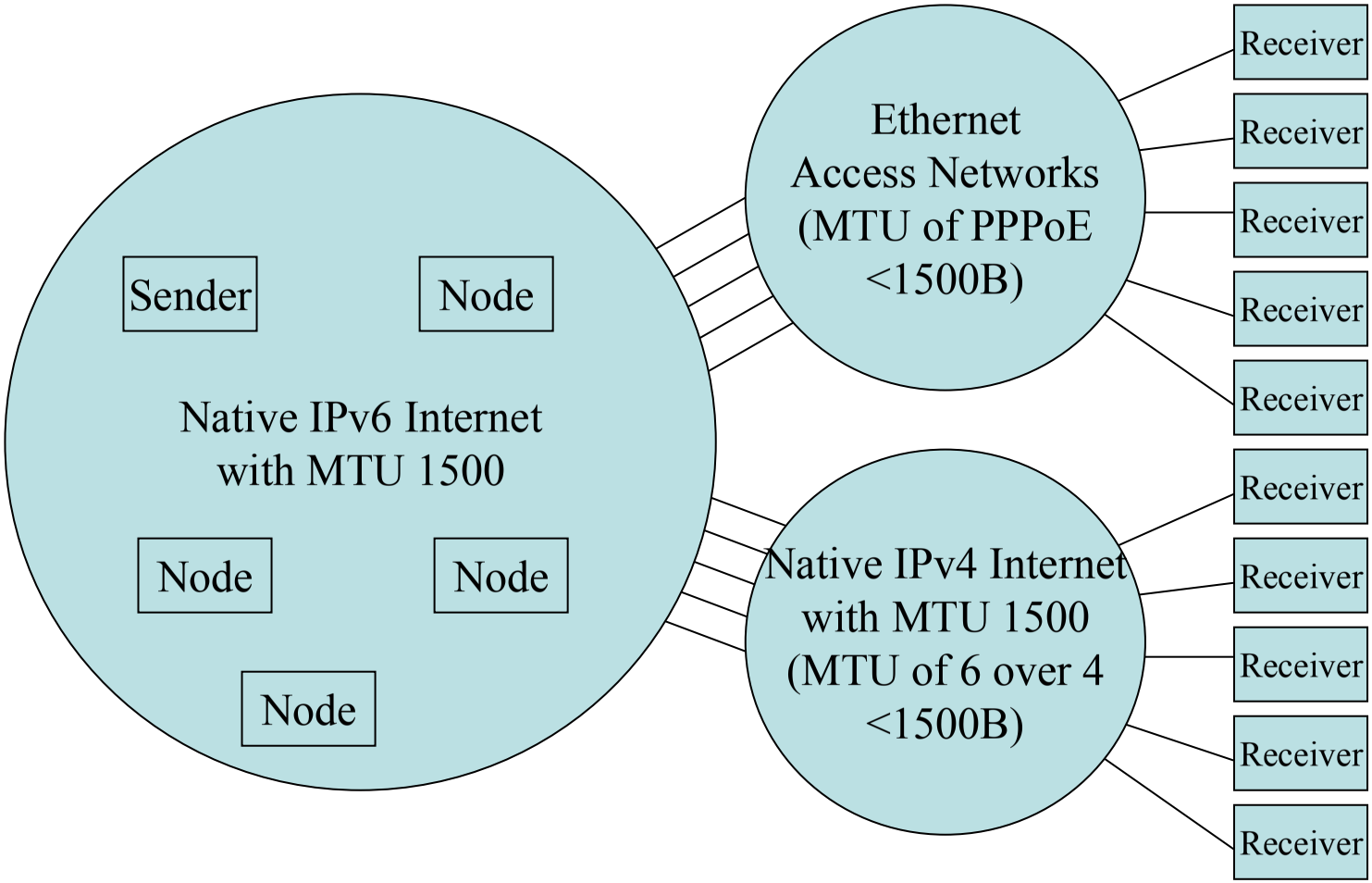
RFC1981 (Path MTU Discovery for IP version 6)

- The Draft Standard Specifies:
  - Path MTU Discovery supports multicast as well as unicast destinations. In the case of a multicast destination, copies of a packet may traverse many different paths to many different nodes. Each path may have a different PMTU, and a single multicast packet may result in multiple Packet Too Big messages, each reporting a different next-hop MTU. The minimum PMTU value across the set of paths in use determines the size of subsequent packets sent to the multicast destination.
  - In the case of a multicast destination address, copies of a packet may traverse many different paths to reach many different nodes. The local representation of the "path" to a multicast destination must in fact represent a potentially large set of paths.
- How large is “a potentially large set of paths”?

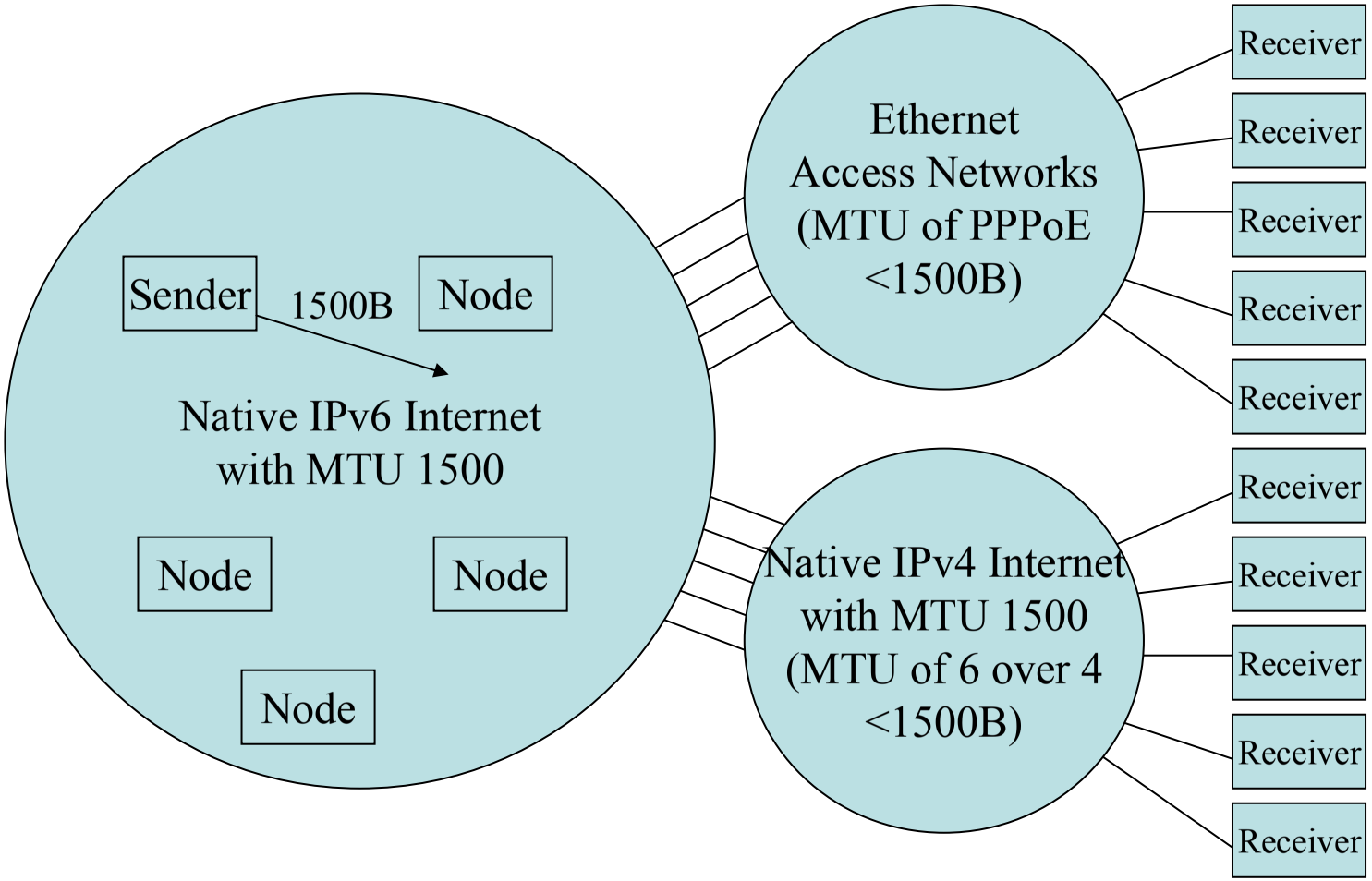
Tunnels at the Last Hop with Typical IPv6 Deployment



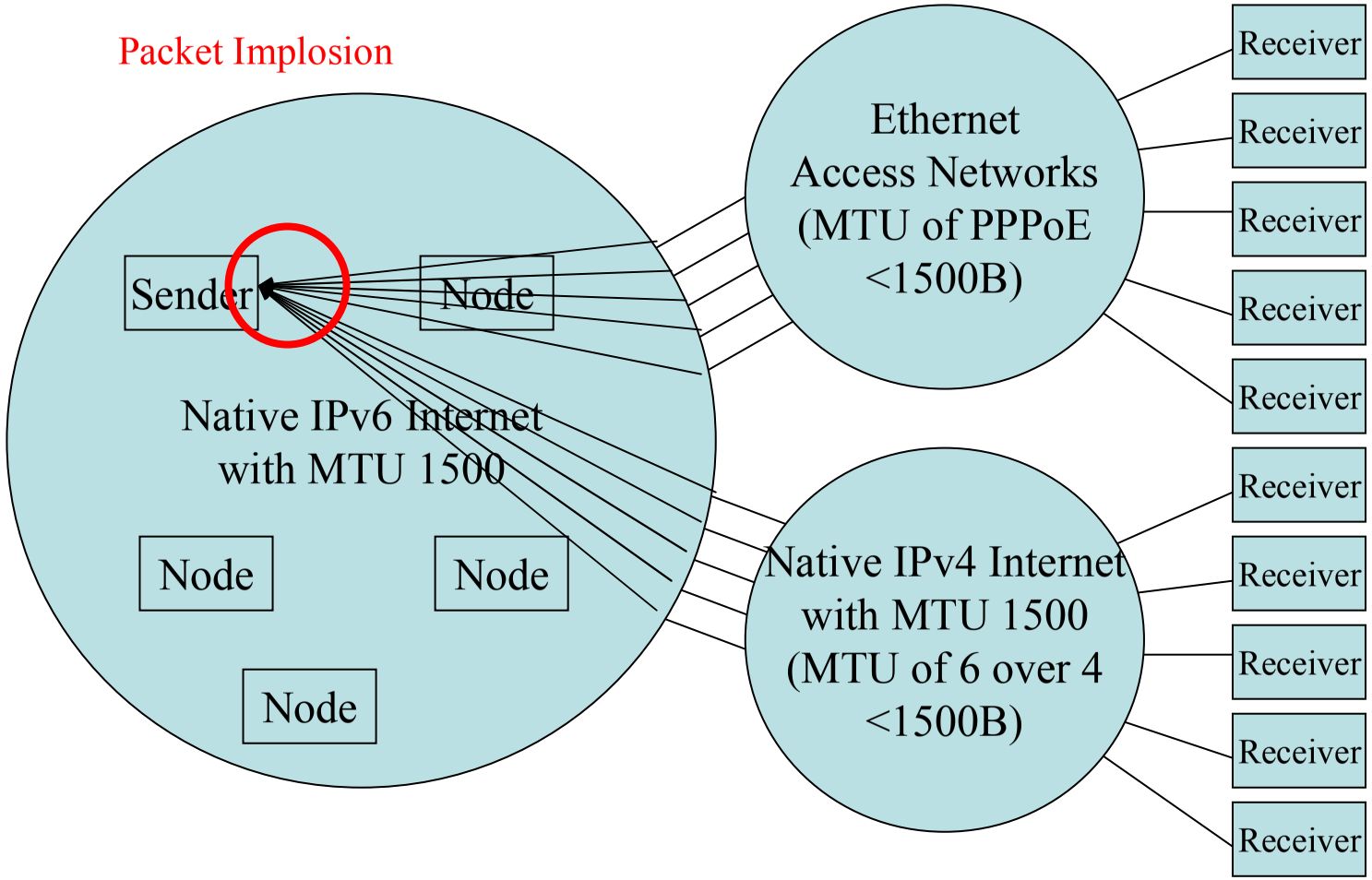
Multicast Path MTU Discovery and ICMP Implosion



Sender Periodically Send Packets with MTU 1500



ICMP Packet Too Big Messages are Generated Near Each Receiver



DOS

- Some multicast routing protocol allows for source address spoofing
  - ICMP may be used for DOS amplifier
  - even if non-link-local multicast is not enabled around a victim

Not a Problem?

- Because almost all ISPs do not enable multicast routing protocol
- ISPs do not allow ordinary users send multicast packets
  - still a problem, because rational ISPs want to avoid to rely on rational operations of other ISPs
  - instead, the multicast PMTUD problem is yet another reason for ISPs to disable multicast
    - multicast PMTUD, to promote multicast and MTUD, ironically killed multicast and MTUD thoroughly

RFC2463 (ICMPv6) Requires

- A Packet Too Big MUST be sent by a router in response to a packet that it cannot forward because the packet is larger than the MTU of the outgoing link. The information in this message is used as part of the Path MTU Discovery process [PMTU].
- Sending a Packet Too Big Message makes an exception to one of the rules of when to send an ICMPv6 error message, in that unlike other messages, it is sent in response to a packet received with an IPv6 multicast destination address, or a link-layer multicast or link-layer broadcast address.
  - Parameter Problem Messages also make an exception

To Prevent ICMP Implosions

- Violate RFC2463 to
  - stop generating ICMP packet too big and parameter problem for multicast packet
  - filter ICMP packet too big and parameter problem for multicast packet
- Or, as it is already a violation, simply
  - stop generating any ICMP
  - filter all the ICMP
  - “it’s against an RFC” is not a valid criticism

Fundamental Solution

- Update RFC2463 to prohibit generation of ICMP against multicast packets
- Write an BCP to Force ISPs not Filter ICMP
- Should take another decade or two
  - unrealistic

Without PMTUD...

- According to RFC2460:
  - It is strongly recommended that IPv6 nodes implement Path MTU Discovery [RFC-1981], in order to discover and take advantage of path MTUs greater than 1280 octets. However, a minimal IPv6 implementation (e.g., in a boot ROM) may simply restrict itself to sending packets no larger than 1280 octets, and omit implementation of Path MTU Discovery.
    - Packet larger than 1280B can not be sent
- IP over IP tunnels (e.g. RFC2473 for MIPv6) needs tunnel MTU 1280B, violating RFC2460
  - or, all the 1280B packets are fragmented, because MTU of 1280B tunnel is smaller than 1280B

Conclusion

- Multicast PMTUD is broken
  - to cause ICMP implosion
- ISPs should filter ICMP Packet Too Big
  - at least against multicast packets but maybe all
- We can’t expect unicast PMTUD work
- We shouldn’t send packets > 1280B
  - except for tunnels