Fast Congestion Notification Packet (CNP) in RoCEv2 Networks

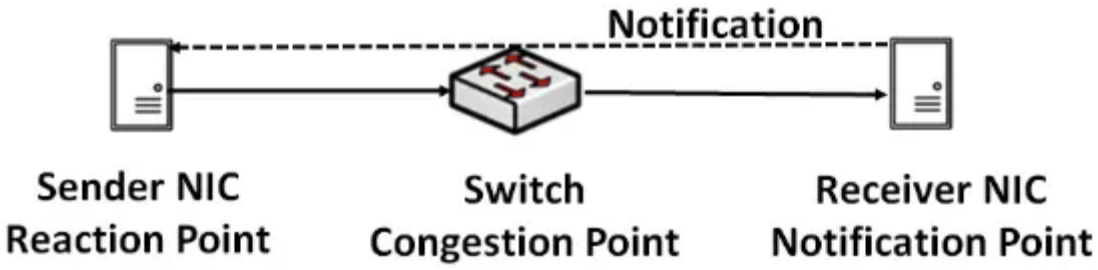
draft-xiao-rtgwg-rocev2-fast-cnp-01

Xiao Min ZTE

Hesong Li ZTE

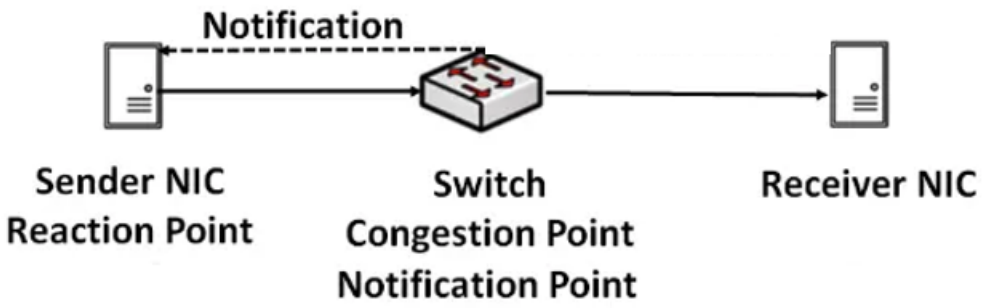
Luigi Iannone Huawei

What's CNP in RoCEv2?



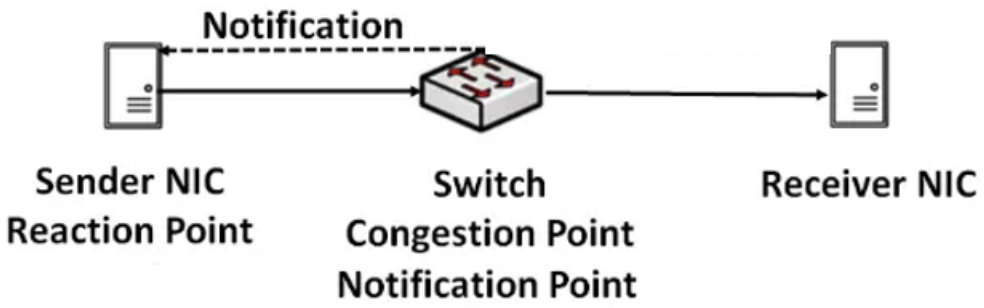
- Step 1, Congestion detected by Switch.
- Step 2, ECN bits marked by Switch.
- Step 3, Marked ECN bits detected by Receiver.
- Step 4, CNP sent to Sender by Receiver.
- Step 5, Transmission rate reduced by Sender.

What's Fast CNP in RoCEv2?



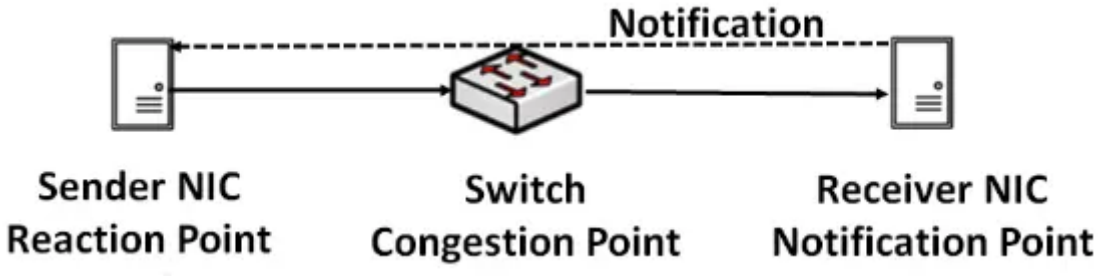
- Step 1, Congestion detected by Switch.
- Step 2, ECN bits marked by Switch. (Optional)
- Step 3, Fast CNP sent to Sender by Switch.
- Step 4, Transmission rate reduced by Sender.

Why is Fast CNP needed?

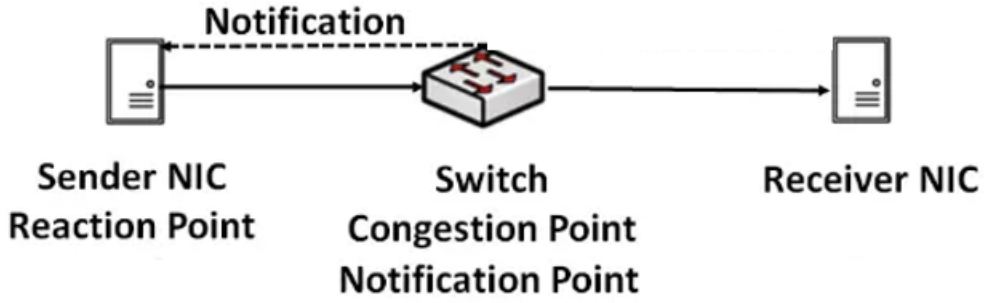


- Fast CNP is faster than CNP
  - Notification Point is closer to Sender.
  - There is significant gain when the distance between Switch and Receiver is non-negligible.
    - Geographically distributed DC and cross DC scenarios

What's the problem for Fast CNP?

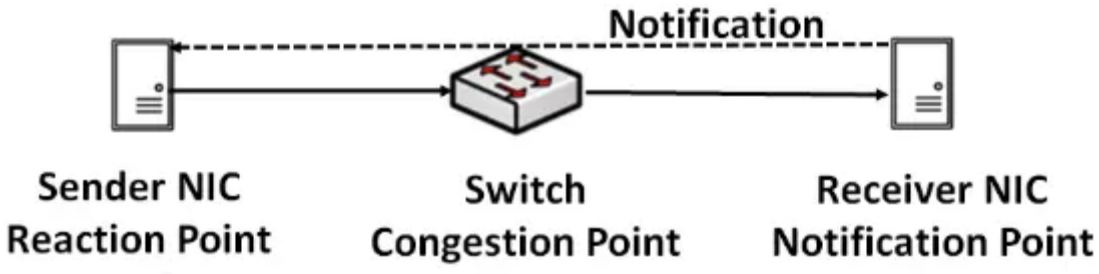


Receiver needs to populate Source QP into CNP. Receiver is able to acquire Source QP from the ECN marked data packet.

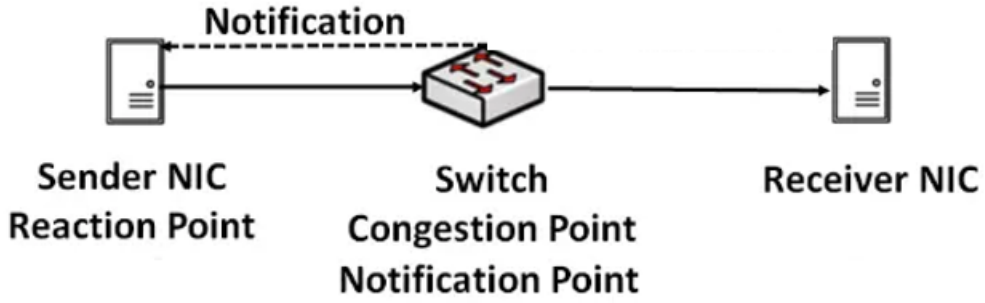


Switch needs to populate Source QP into Fast CNP, but Switch is unable to acquire Source QP from the congested data packet.

What's the proposed solution?



Receiver needs to populate Source QP into CNP. Receiver is able to acquire Source QP from the ECN marked data packet.



Switch prepends CNP an IPv6 extension header, which contains the DA of congested data packet, i.e., the Receiver's address. Sender is able to acquire Source QP from the Fast CNP.

Next steps

- Ask for more reviews and comments
- Revise this draft to improve it

Thank you!