Internet Draft
draft-fpeng-ecn-04.txt

Fei Peng
Beijing University of posts and telecomm.

Jian Ma
Nokia Research Center

May 2001

A proposal to apply ECN into Wireless and Mobile Networks

Status of this Memo

This document is an Internet-Draft and is NOT offered in accordance with Section 10 of RFC2026, and the author does not provide the IETF with any rights other than to publish as an Internet-Draft

Internet-Drafts is working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts. Internet-Drafts is draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet- Drafts as reference material or to cite them other than as "work in progress."
The list of current Internet-Drafts can be accessed at
http://www.ietf.org/ietf/1id-abstracts.txt
The list of Internet-Draft Shadow Directories can be accessed at
http://www.ietf.org/shadow.html.

1. Abstract

TCP congestion control has been developed on the assumption that congestion in the network to be the only cause for packet loss. Thus, it drops its transmit window upon detecting a packet loss. In the presence of high error rates and intermittent connective characteristic of wireless link, these results in an unnecessary reduction in link bandwidth utilization for packet losses are not mainly due to congestion.

Current ECN proposal proposed by IETF obtains a part separation of congestion control and packet losses with the purpose of preventing unnecessary packet drops due to buffer overflow. However, network

congestion can not be completely avoided and because of the losses of ECN by corruption and congestion, ECN will have to interact with the existing congestion control mechanism in TCP. This paper provides an effective way to improve TCP performance in wireless and mobile networks and cooperate ECN into such environment. Upon each dropped packets due to buffer overflow or the RED routers, an ISQ (ICMP Source Quench) is generated and sent back to the source of each dropped packet. The approach reduces congestion losses and the reaction time to congestion in the network. With the ECN control, network congestion is alleviated. When CE bit can be marked, no ISQ is required to be generated and ISQ messages created will not cause wasting of bandwidth. To ensure system more secure in case of losses of ECN or ISQ messages, a window threshold is calculated to allow that packet losses should initiate window reduction during times without ISQ and ECN messages coming back. It is important to note that this method is simple to implement for it makes minimal modification of current systems.

2. Introduction

Congestion control is one of the key mechanisms to accommodate the increasingly diverse range of services and types of traffic in the Internet. Initially Internet was intended to support best-effort service, and TCP congestion control method that was actually implemented has been developed on the assumption that the network would be treated as a black box. This means that the end nodes do not exercise control by directly ascertaining the state of routers and transmission line, but rather regulate the traffic by inferring the network load indirectly from packet loss and response time fluctuations. In wired networks, this may not induce serious problems as packet losses are mainly due to congestion. However, in the presence of high error rates and intermittent connective characteristic of wireless links typically like wireless and mobile networks or the satellite environment, this reliance on packet drop as an indication of congestion causes a significant degradation in TCP performance, for TCP reacts to packet loss as it would in the wired environment: it drops its retransmit window size before retransmiting packets, initiates congestion control or avoidance mechanisms and resets its retransmission timer, thereby result in an unnecessary reduction in link bandwidth utilization, which causing poor throughput and very high interactive delays.

Recently, several schemes have been proposed to alleviate the effects of non-congestion-related losses on TCP performance over networks that have wireless or similar high-loss links.
One of the researches concerns on improvement of transport underlying protocol stacks. For example, there have been several proposals for

reliable link-layer protocols as forward error correction (FEC) and retransmission of lost packets in response to automatic repeat Request (ARQ) messages. However, it is the main worry about link-layer protocols that an adverse effect on certain transport-layer protocols such as TCP is very possible. Another solution of network layer protocol called snoop protocol is proposed to cache packets at the base station and perform local retransmissions across the wireless link. Like link-layer solutions, the snoop approach could also suffer from not being able to completely shield the sender from wireless losses. In practice, with the enhancement of TCP underlying protocols, the link error rate will still remain 1E-6 bits/sec. So it is essential to give a solution in TCP protocol stack.

Several schemes modified at transport layer have been proposed to alleviate the effects of non-congestion-related losses on transport performance. The indirect-TCP is one of the first protocols to distinguish different losses by splitting a TCP connection between a fixed and mobile host into two separate connections at the base station, a more optimized wireless link-specific protocol tuned for better performance can be used over a one-hop wireless link. The advantage of the split connection approach is that it achieves a separation of flow and congestion control of the wireless link from that of the fixed network. However, there are some drawbacks of this approach such as loss of semantics, application re-linking and software overhead, etc. And there is no need to sacrifice the semantics of acknowledgments in order to achieve possible good performance.

As lost packets can be simply divided into congestion-related losses and non-congestion-related losses, an ELN protocols can be used to differentiate the packet loss by adding an explicit loss notification option to TCP acknowledgments when a packet is dropped on the wireless link. Future cumulative acknowledgments corresponding to each lost packet must be always marked to identify that a non-congestion-related loss has occurred, then the sender may perform retransmissions without invoking the associated congestion-control procedures. In practice, this algorithm brings burden to the implementation nodes for judgment is required for each dropped packets and each lost packet due to transmission errors needs marking otherwise it will invoke congestion control. Additionally, it might be difficult to identify which packets are lost due to errors on lossy link, for example, it may be hard to determine the connection that a corrupted packet belongs to since the header could itself be corrupted.

This paper provides a mechanism used in wireless and mobile networks. Like ELN, it is a mechanism by which

the reason for the loss of a packet can be communicated to the TCP sender. In particular, it provides a way by which senders can be informed that a loss happened because of reasons related to congestion), it is very easy to identify which packets are lost due to buffer overflow or RED mechanism in routers. The window threshold set to avoid unnecessary reduction of window size by
Packet losses ensure that most of the times the congestion control mechanism is invoked by ISQ or ECN.

3. Current ECN proposal in IETF

Bits 10 and 11 in the IPV6 header are proposed respectively for the ECT(ECN Capable Transport indicator) and CE (Congestion Experienced indicator). Bits 6 and 7 of the IPV4 header TOS field are also proposed as the ECT and CE placeholders respectively. TCP header is modified to add an additional flag, the ECN Echo flag, to notify the sender (from the receiver) that it is contributing to congestion. The flag's bit-space is borrowed from the reserved field in the TCP header. This bit is also interchangeably referred to as the ECT bit in this text.

The ECT bit is set by the sender end system if both the end systems are ECN capable. This is confirmed in the pre-negotiation during the connection setup phase in TCP. Packets encountering congestion are marked (CE bit) by a router on their way to receiver end systems (from sender end systems), with a probability proportional to their bandwidth usage following the procedure used in RED [RFC2309] routers. When the receiver end system receives packet with CE and ECT bits set, it informs the sender end system that it is contributing to congestion by the setting of ECT bit in the ACK packet.
The sender end system reacts by halving the congestion window upon receiving the ACK packet. And it reacts only once to ECT messages per in-flight window of messages.

4. Limitations of the Current ECN Proposal in wireless/mobile networks [RFC 2481]

It is assumed that the participating routers are capable of RED or some other active queue management mechanism. In such a router, a packet has a probability of being dropped where this probability is dependent on average queue size.
Because of the complex condition of the networks, packet drops due to buffer overflow can not be completely prevented with ECN mechanism, in addition, ECN itself will be lost by congestion or the corruption.

According to above assumption, it is most required that ECN proposal shall coupled with congestion control mechanism in TCP. In networks over imperfect links where packet losses are not mainly due to congestion, the unnecessary reduction of throughput will occur for the congestion window has been shielded half before ECN could come back.

It should mention that Current ECN mechanism does not have to change any way with our proposal at the transport layer.

5. ICMP Source Quench

All gateways must contain code for sending ICMP Source Quench messages when they are forced to drop IP datagram due to congestion. Although the Source Quench mechanism is known to be an imperfect means for Internet congestion control, and research towards more effective means is in progress, Source Quench is considered to be too valuable to omit from production gateways. [RFC 1009]

There is some argument that the Source Quench should be sent before the gateway is forced to drop datagram. For example, a parameter X could be established and set to have Source Quench sent when only X buffers remains. Or, a parameter Y could be established and set to have Source Quench sent when only Y per cent of the buffers remain.

Two problems for a gateway sending Source Quench are (1) the consumption of bandwidth on the reverse path, and (2) the use of gateway CPU time. To ameliorate these problems, a gateway should be prepared to limit the frequency with which it sends Source Quench messages. This may be on the basis of a counter (e.g., only send a Source Quench for every N dropped datagrams overall or per given source host), or on the basis of a timer (e.g., send a Source Quench to a given source host or Overall at most once per T milliseconds). The parameters (e.g., N or T) must be settable as part of the configuration of the gateway; How to give a suitable value of N or T in practice is also a problem.

The [draft-salim] proposal concerns: ISQ message is generated by the intermediate RED router when it capture a packet with ECT bit set by the active queue management. Before the origination of the ISQ, the packet, which was chosen by RED probability, should be marked if it has not already marked. If the ECT bit is not set, the packet will be dropped whether RED chooses it or the average queue size goes above the maximum threshold. The purpose of this approach is to reduce the reaction time to congestion in the network and provide multilevel congestion feedback. In the following section, we

propose our ISQ mechanism to fit wireless and mobile environment.

6. Source Quench in Wireless/mobile networks

While there are some applications/environments where it might be highly advantageous for the sender to receive some indication of congestion without having to wait a roundtrip time, this is not the common case. Source Quench packets add traffic in the reverse direction on what might be a congested path. Even with multilevel function of ISQs, the congestion window and the slow start threshold value are only halved at TCP source. Without the corresponding reaction of the source behavior, the multilevel ISQ lose its significant. Moreover, if threshold is set appropriately lower, ECN is also to be considered effective to alleviate network congestion in time. So, this section propose the ISQ messages are not required to be generated if only ECN can be set by the RED or other queue mandation without the packet being dropped in such case. Only upon the dropping of packets due to buffer overflow or queue management without ECN, ISQ messages are sent back to the corresponding sources. The algorithm in the router can be described as follows:

If the incoming message causes average queue size go above maximum threshold or causes buffer overflow, the packet is dropped and an ISQ then sent back to the source of the incoming packet.

If the incoming message causes the average queue to go between the minimum and maximum thresholds then:

```
    If the RED probability picks this packet then:
  If the ECT bit is set and the CE bit is not already marked then:
       Mark the packet (CE bit)
  Else if RED chooses this packet and ECT bit is not set then:
       Send an ISQ back and drop the packet.
```

As long as ECT bit is set, CE bit is marked in most of times when condition permits. And also with ECN mechanism in networks, the frequency of generation of ISQ messages is reduced which result in saved bandwidth and avoid implementation complexity though CPU time is no longer a constrained resource today. When ECN is not supported in some cases, Reasonable performance of the protocols that use IP (e.g., TCP) requires an IP datagram loss rate of less than 5%[RFC 1009]. Moreover, it has been quantitatively shown in simulations [kcho-97] that less packet drops happen (only about 1-5%) of the packets are marked or dropped in a RED gateway under incipient congestion. This implies the amount of processing needed at the router is reduced and little waste of bandwidth by generation of ISQ.

So, ICMP Source Quench message (ISQ) are not generated by the intermediate congested RED router if only that router decides to mark the CE bit. ISQ are usually not generated for a packet that has already been marked previously by another router regardless of whether that packet is contributing to some congestion; however, when the router queue level mandates the dropped packet then an ISQ is sent back to the source regardless of whether the packet was marked previously or not. This function of ISQ has been supported in the routers [rfc1009] and since each router is required to provide a disable parameter, only configuration operation is taken to enable its function.

7. Support at TCP layer

The requirements for the end host's reaction to ISQ are at the moment [RFC1122]. The source reacts at the transport protocol level by lowering its data throughput into the network. In TCP, upon identifying the flow causing the congestion, the sender reacts by halving both the congestion window and the slow start threshold value for that flow. The sender does not react to an ISQ message more than once per window. For multiple ECN and ISQ come back from the networks, The source only reacts the first one in a window. Upon receipt of the first ISQ or ECN at time t, it notes the packets that are outstanding at that time (sent but not yet acked) and waits until a time u when they have all been acknowledged before reacting to a new ISQ or ECN message.

To prevent unnecessary reduction of window size, as notes previously, unnecessary window shield back causes low utilization of bandwidth, the measurement of the maximum window (called Wmax) experienced on a given connection. We expect this can change over time, and TCP should track these changes and modify its timeout accordingly. First TCP must measure the Wmax whenever it begins to shun down window by ISQ or ECN, we will use Mw to denote the measured Wmax. Then TCP updates a smoothed Wmax estimator using the low-pass filter

$$Wmax \leftarrow a\ Wmax + (1-a)\ Mw$$

Where a is a smoothing factor with a recommended value of 0.9. This smoothed Wmax is updated every time a new measurement is made. Ninety percent of each new estimate is from the previous estimate and ten percent is from the new measurement by the first receipt of ISQ or ECN in a widow cycle.
Given this smoothed estimator, which changes as the Wmax changes, we recommended the threshold (Tl) to determine when packet losses should call the congestion control is set to

Tl = B Wmax,

where B is a delay variance factor with a recommended value of 1.5. Since in congestion avoidance phase, the window increasing rate is linearly, only one packet is increased after a RTT, so 1.5 value of threshold is enough to avoid unnecessary window reduction by packet losses. The ISQ or ECN gives the initialized value of Wmax after the measure of the first reduction of window by either of them.

If the initialization of window size is caused by packet loss controled by the threshold. The Wmax and Tl are calculated as:

Wmax = a Wmax + (1-a) Tl
Tl = B Wmax.

Since the sender does not reduce window more than once per window, the following ISQ or ECN message do not affect any change of transmission rate until the outstanding data before the sender initiate congestion control by a packet loss upon the reach of threshold.

8. Security

With the addition of ISQ messages, It becomes more security for ECN used in wireless and mobile networks. For example when some node does not support ECN mechanism, ISQ could send because the node will otherwise forced to drop IP datagrams due to congestion. Since all gateways must contain code for sending ICMP Source Quench messages in such case, the source transmission rate will be slow down even in non-ECN support environment. Also ISQ message and ECN will be lost, so it is not reliable, but since the source only shun down window once in a transmission cycle, multiple ISQs an ECNs generated at the network will not cause performance problem even with some losses of them. Moreove, the threshold set for window reduction by a packet loss though might be somehow later, it could guarantee to recover the severe congestion met by the losses of ISQ and ECN.

9. Conclusion

All gateways must contain code for sending ICMP Source Quench messages when they are forced to drop IP datagrams due to congestion [RFC 1009], so only configure operation at each intermediate router in the networks is required to be taken. Without modification of ECN mechanism, we only set a threshold to delay the packt loss to

initiate congestion control and also guarantees the security of the whole system. It is very simple to implement and provides really an effective way to improve TCP performance in wireless and mobile networks for the unnecessary window reduction of losses due to transmission errors is effectively avoided.
To further analyze the benefits of the whole systems, continuous simulations must execute in the future work, which contain multiple congested gateways and two way traffic with either support ECN or non-support ECN systems. Without modification of ECN mechanism, the addition of ISQ in no-support ECN systems would invariably improve TCP performance in wireless and mobile networks through preventing packet losses to initiate window reduction. It is also our future work to investigate to what extent it would contribute to improvement of TCP performance through more optimal active queue management, and a better window adaptive algorithm to suit large wide network configurations is also the requirement of study in the future.

10. References

[kcho-97] Cho, K.J. ALTQ/RED Performance,http://www.csl.csl.sony.co.jp/person/kjc/red/perf.html

[FeiNC] FEI, P., JIAN, M. "Overload control method for a packet-switched network", Patent Application NC 18254, June 1999.

[ICCT98] FEI, P., Jian, M., etc. "TCP Performance Enhancements in Wireless and Mobile Networks", International Conference on Communication Technology (ICCT'98), Oct,1998, pp.S46-07-1:S46-07-5.

[draft-kksjf] K. K. Ramakrishnan, Sally Floyd, "A proposal to add Explicit Congestion Notification (ECN) to IP", Internet Draft-kksjf-ecn-03.txt, Oct, 1998.

[SF94] Sally Floyd, "TCP and Explicit Congestion Notification", Computer Communication Review, V.24 N.5, October 1994, p.10-23.

[RFC1009] R. Braden, J. Postel, Requirements for Internet Gateways

[draft-salim] Hadi Salim,J.,etal, A proposal for Backward ECN for the Internet Protocol (Ipv4/Ipv6), June 1998.

11. Acknowledgments

We would like to appreciate Nokia cooperation for supporting this idea. We also thank Beijing University of posts and telecommunicaitons with great advocacy of our research. And we will particularly mention professor (Mrs. Cheng Shiduan),her encouragement and good advice for our work.