Measurement and Classification of Out-of-Sequence Packets in a Tier-1 IP Backbone

Sharad Jaiswal†, Gianluca Iannaccone§, Christophe Diot§, Jim Kurose†, Don Towsley†

§Sprint ATL
{gianluca,cdiot}@sprintlabs.com

†CS Dept. UMass, Amherst
{sharad,kurose,towsley}@cs.umass.edu

I. Introduction

An important characteristic of any TCP connection is the sequencing of packets within that connection. Generally, if sequence numbers are monotonically increasing, then all is well - data flows through that connection without loss, and the network does not introduce pathological problems such as in-network duplication and reordering. Conversely, out-of-sequence packets indicate that the connection suffers from loss, duplication or reordering. It is thus of interest to study the magnitude of out-of-sequence packets within Internet TCP connection, and to identify their causes, as the magnitude of out-of-sequence packets is a good indicator of the "health" of a TCP connection and the path that it is traversing.

In this paper we present measurements and a classification of out-of-sequence packets in TCP connections within the Sprint IP backbone. Informally, we will say that a packet is out-of-sequence if it has a sequence number that is smaller than that of a previously observed packet in that connection. An out-of-sequence packet can be caused by three different events:

- **Retransmission**. In this case, a sender infers that a packet has been lost and retransmits the packet. The retransmitted packet will have a sequence number that is smaller than previously observed packets at the measurement point and hence will be deemed "out-of-sequence."
- **Network duplication**. In this case, a non-sender-retransmitted copy of a packet is observed. This can occur when the measurement point is within a routing loop (and hence the same packet is observed more than once), or if the network itself creates a duplicate copy of the packet.
- **In network-reordering**. In this case, the network inverts the order of two packets in a connection (for example, because of parallelism within a router [1] or a route change).

We would like to emphasize that in this paper, the terms "out-of-sequence" and "reordered" do not have the same meaning. A reordering event in our classification is just a subset of all possible events which result in an out-of-sequence packet.. Previous studies have often used the terms "out-of-sequence" and "reordered" interchangeably.

Our contributions are twofold. The first contribution is methodological. Because we measure out-of-sequence packets at a *single* point in the backbone (rather than by sending and measuring end-to-end probe traffic at the sender or receiver [1], [4]), a new methodology is required to infer the causes of a connection's out-of-sequence packets based only on measurements taken in the backbone, i.e., in the "middle" of this connection. An advantage of having such a measurement point within the backbone is that we are able to characterize the behavior of flows between a very large number of source-destination pairs, without having to instrument the individual senders and receivers. While the use of a single measurement point has the advantage of sampling traffic from a very large number of connections, it also poses challenges. Because we are taking measurements in the "middle" of a TCP connection, we do not know whether a packet observed at our measurement point is received at the intended destination and/or the action taken at the packet's destination; we can only infer this from the previously- or subsequently-observed packets from that connection and our knowledge of how TCP behaves.

Our second contribution is the characterization of the out-of-sequence behavior itself. We characterize the out-of-sequence behavior of 14 million TCP connections, generated in more than 3,400 unique Autonomous Systems. The links at which measurements were performed varied in terms of their capacity (OC-3/12/48 links), utilization, and location (e.g., intra-POP, inter-POP and peering links). Our data thus represents a diverse mix of end-to-end paths and traffic characteristics.

In our work, we use the information available in the TCP header, the identification field in the IP datagram, and knowledge of the functioning of the TCP protocol, to infer the cause of the out of sequence packet. Figure 1 summarizes the decision process that implements the rules below to classify out-of-sequence packets. The edges leading to leaf nodes (classifications) in the decision tree in Figure 1 are annotated with the decision rules (R1 through R6) corresponding to those classifications. Due to space considerations, we shall not discuss the classification rules in detail, and refer the reader to the technical report [2], which has a comprehensive discussion of the heuristics used. We show that using these simple techniques, it is possible to classify almost all out-of-sequence packets in our traces and that we can characterize the uncertainty in our classification.

Several of our classification techniques require an estimate of the sender's TCP RTO (retransmission timeout interval) and

This work is supported by the National Science Foundation under grants ANI-0077039, ANI-9805185 and ITR-0085848. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Part of this work was carried out when Sharad Jaiswal was an intern at Sprint ATL.

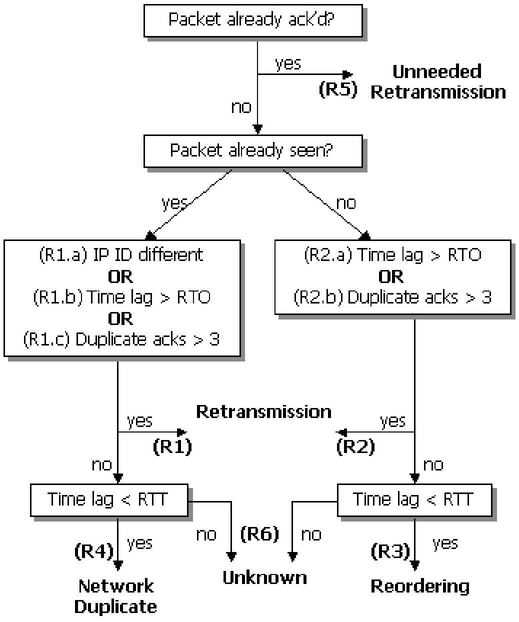

Fig. 1. Decision process for the classification of out-of-sequence packets

RTT (the current round trip delay between sender and receiver). In the absence of knowledge of exact per-sender TCP state at our measurement point, we can again only infer these values from our measurements. In [2], we also describe techniques for estimating per-connection end-to-end RTT from a single measurement point within the network.

II. Measurement and Classification results

In the following, we show the results of our classification over a packet trace collected on a long-haul OC48 (2.5 Gbps) link connecting a POP on the East Coast to one on the West Coast. for our study. Our measurements were obtained using infrastructure developed as part of the Sprint IP Monitoring (IPMON) project [3]. The IPMON measurement system provides packet-level traces from OC-3, OC-12 and OC-48 links in several Points-of-Presence (POPs) in the Sprint backbone. Table I shows the classification results for the observed out-of-sequence packets in an 1-hour stretch of the OC-48 trace. We have observed similar trends to those reported below in several other traces collected on different dates and over different links, [2] has a detailed discussion of figures from these traces.

The first two rows in Table I indicate the total number of observed TCP connections, and the data packets in these connections. The subsequent row enumerates the absolute number and percentage of these packets that were out-of-sequence. Generally, in all our examined traces, the number of out-of-sequence packets is limited to about 5% of the total data packets exchanged by the TCP connections. The last five rows of the table break down the absolute number of out-of-sequence packets and the relative percentage of out-of-sequence packets according to the inferred cause of out-of-sequence behavior. We note that our heuristics classify the vast majority of these data packets, with only between 1% and 5% of the out-of-sequence packets being classified into the *unknown* category.

	OC48	
TCP connections	1.4M	
Data packets	29,925,964	
Out-of-sequence	940,837	(3.14%)
Retransmissions	665,122	(70.69%)
Unneeded Retransmissions	190,326	(20.23%)
Network Duplicates	6,483	(0.69%)
Reorderings	48.035	(5.11%)
Unknown	31,499	(3.35%)

TABLE I
OUT-OF-SEQUENCE PACKET CLASSIFICATION

Table I indicates that the bulk of out-of-sequence packets are due to the retransmission of lost packets. Note that there is not a one-to-one relationship between out-of-order (retransmitted) packets and earlier packet loss, since a source may retransmit more packets than needed to repair a loss, depending on the TCP flavor. However, the number of retransmitted packets does provide a rough upper bound on the total number of packet drops experience by a connection, since every lost packet will result in a retransmission.

We also observe that unneeded retransmissions make up a significant percentage of all out-of-sequence packets. We should clarify that these retransmissions are not really "unneeded", especially from the sender's perspective. If an ACK is lost or delayed, the sender has no choice but to retransmit.

The traces we examine do not show a significant number of network-replicated packets, an event that appears to be rare in all the examined traces.

We also observe that the magnitude of reordering we report is much smaller than the results presented in previous works [1], [4]. Our measurements indicate that reordering affects at most 0.5% of all the data packets and that only 1.57% of the connections experience reordering, substantially less than those in [1], [4]. We would like to point out, however, that there are important methodological differences between the studies, namely the active or passive nature of measurements, spacing between the probes and the scope of the studies, which are discussed in more detail in [2]. We would like to note that in [1] the authors did their experiments in 1998 by sending probes through the MAE-East Internet exchange point. They isolated the cause of reordering as due to parallelism within the main network switching device in the exchange point, i.e. the DEC/Gigaswitch. They further conjectured that reordering in general would be a significant factor in the future Internet as a result of increased parallelism in network devices. Our results, four years later, instead suggest the contrary: network reordering affects a very small percentage of all data packets.

References

[1] J. Bennett, C. Partridge, N. Shectman. Packet reordering is not pathological network behavior. *IEEE/ACM Trans. on Networking*, 7(6), Dec. 1999.

[2] S. Jaiswal et al. Measurement and classification of out-of-sequence packets in a tier-1 ip backbone. Tech. Report 02-17, CS Dept., UMass, May 2002.

[3] C. Fraleigh, S. Moon, C. Diot, B. Lyles, F. Tobagi. Packet-level traffic measurements from a tier-1 IP backbone. Tech. Report TR01-ATL-110101, Sprint ATL, Nov. 2001.

[4] V. Paxson. End-to-end internet packet dynamics. *IEEE/ACM Trans. on Networking*, 7(3):277–292, June 1999.