Analysis of ICMP Quotations

David Malone[1] and Matthew Luckie[2]

[1] Hamilton Institute, NUI Maynooth, David.Malone@nuim.ie
[2] WAND Group, Computer Science Dept., University of Waikato, mjl@wand.net.nz

1 Introduction

RFC 792 requires most ICMP error messages to quote the IP header and the next eight bytes of the packet to which the ICMP error message applies. The quoted packet is used by the receiver to match the ICMP message to an appropriate process. An operating system may examine the quoted source and destination IP addresses, IP protocol, and source and destination port numbers to determine the socket or process corresponding to the ICMP message. In an idealised end-to-end Internet, the portion of the packet quoted should be the same as that which was sent, except for the IP TTL, DSCP, ECN bits, and checksum fields. In the modern Internet, this may not always be the case. This paper presents an analysis of ICMP quotations where the quote does not match the probe.

2 Methodology

2.1 Data Collection

Using `tcptraceroute`, the paths to 84393 web servers used in a previous study [1] were traced serially between the 6th and 12th of May 2005. All TCP SYN packets sent from the measurement source, as well as all ICMP time exceeded, unreachable, source quench, redirect, and parameter problem messages were recorded using `tcpdump`. 1190351 probes were sent, and 858090 ICMP replies were received and matched to a probe from 53768 unique IP addresses. 836456 ICMP responses were of type time exceeded, 21525 were of type unreachable, and 109 were of type source quench. A further 9 ICMP messages were unmatched.

By default, `tcptraceroute` generates a TCP SYN packet to port 80 and assigns the packet a unique IP-ID so that any subsequent response can be matched to its probe. The ECE and CWR TCP flags were set to test the reaction of middleboxes to these flags. The DSCP/ECN IP fields were set to 0x0f to identify systems that might modify these fields as part of the forwarding process. The DF bit was set to identify behaviour related to workarounds for broken path MTU discovery. Finally, each hop was probed once, to a maximum of 25 hops.

2.2 Quote Matching

As we may be matching an ICMP response to a packet that has been modified in flight, we are relatively liberal when matching an ICMP response to a probe.

Based on the responses seen, the following heuristic was devised. A list of the 25 most recently sent probes is kept, as well as an array of the most recently sent probes for each IP-ID value. If an ICMP response can be matched by IP-ID or byte-swapped IP-ID to one of the 25 most recently sent probes, then it is deemed a match. Otherwise, it is not clear if any of these 25 probes match the ICMP response, because it is possible that either the IP-ID in the quoted packet was modified or the response was significantly delayed.

We score each of these 25 probes, as well as the last probes sent with a matching IP-ID or a byte-swapped IP-ID, and select the probe that meets the greatest number of the following criteria: matching destination IP address, matching TCP source port, matching TCP sequence number, no previous matching response, in last 1200 sent. Providing at least one of the IP-ID, destination IP address, TCP source port, or TCP sequence number matches, the probe with the largest number of matching criteria is the matching probe. To validate this technique, more unusual matches were inspected manually, and they appeared to be genuine.

2.3 Modification Classification

A modification may be classified one of three ways. First, if a modification is made to a single field and it appears unrelated to any other modification made, then it is noted as only affecting that field; the modification is further examined to determine if the field was set to zero, byte-swapped, incremented, or altered in some other way. Second, if a modification alters a set of fields in a related way, the modifications are summarised. For example, if an intermediate node inserts a TCP MSS option into a SYN packet as it is forwarded, then to do so correctly it will adjust the IP length field, TCP offset, TCP checksum, as well as include the option itself. Third, if a modification appears to be the result of accidentally overwriting a series of consecutive fields in the quoted packet, such that the integrity of the quoted packet is now compromised, the modification is classed as clobbering the fields.

2.4 Spacial Classification

Where possible, we infer if the modification is made in-flight while forwarding the probe, or is made during the quoting process and therefore localised. Quotes from a pair of adjacent hops are required to spacially classify a modification. A modification is associated with the first IP address of the pair. If a modification is observed at one hop but not the next, it is classed as a *quoter* modification. If the same modification is observed at adjacent hops, it is classed as an *in-flight* modification provided at least one of the corresponding IP or TCP checksums quoted is valid, indicating the change was intentional. This reduces the chance that adjacent quoter modifications are incorrectly classified as an in-flight modification, perhaps due to using the same router model with the same quirk at adjacent hops. Otherwise, if a modification is observed, but there is not a quote from an adjacent hop available for spacial classification, then it is classified as an *edge* modification. Finally, we stop processing a path when a loop is inferred.

Table 1. Modifications made to IPv4 and TCP headers, by quoter IP address.

Modification	In-flight	Quoter	Edge	Total Unique
IPTOS_MOD	1533 (2.9%)	146 (0.3%)	1674 (3.1%)	3030 (5.6%)
IPLEN_SWAP	0 (0.0%)	0 (0.0%)	1 (0.0%)	1 (0.0%)
IPLEN_MOD	0 (0.0%)	174 (0.3%)	322 (0.6%)	480 (0.9%)
IPID_SWAP	0 (0.0%)	29 (0.1%)	469 (0.9%)	494 (0.9%)
IPID_MOD	0 (0.0%)	1 (0.0%)	19 (0.0%)	20 (0.0%)
IPDF_MOD	4 (0.0%)	1 (0.0%)	30 (0.1%)	35 (0.1%)
IPOFF_SWAP	0 (0.0%)	32 (0.1%)	49 (0.1%)	80 (0.2%)
IPDST_MOD	29 (0.1%)	36 (0.1%)	1189 (2.2%)	1248 (2.3%)
TCPSRC_MOD	0 (0.0%)	3 (0.0%)	43 (0.1%)	46 (0.1%)
TCPDST_MOD	1 (0.0%)	2 (0.0%)	129 (0.2%)	132 (0.3%)
TCPSEQ_MOD	1 (0.0%)	0 (0.0%)	12 (0.0%)	13 (0.0%)
TCPACK_MOD	0 (0.0%)	0 (0.0%)	19 (0.0%)	19 (0.0%)
TCPMSS_ADD	4 (0.0%)	0 (0.0%)	19 (0.0%)	23 (0.0%)

3 Results

3.1 Observed Quote Lengths

Most quoters observed (87.60%) quote the first 28 bytes of the probe, which is the minimum amount permitted. 8.60% quote 40 bytes, corresponding to the size of the probe sent, and 2.14% quote 140 bytes, corresponding to ICMP quotations with MPLS extensions included. Therefore, at least 10.7% of quoters allow the complete IP and TCP headers of a probe to be compared with its quote.

3.2 Modifications to IPv4 Headers

The most frequent in-flight modification observed is to the DSCP/ECN byte: 2.9% of quoters were observed to modify this byte. 31 were inferred to use inconsistent values when overwriting the byte. This could be a measurement artifact due to attributing the change to the first IP address where it was observed, rather than an IP address of the previous hop where the change may have been made. 1073 quoters were observed to clear the DSCP, but leave the ECN bits intact, while 429 quoters were observed to clear the complete byte. 71 were observed to assign a DSCP of '001000', indicating some networks may prioritise HTTP traffic using the IP precedence bits. The second-most frequent in-flight modification observed was a modification of the destination IP address; of the 29 quoters that made modifications, 16 used RFC 1918 private addresses. Finally, 4 quoters were observed to clear the DF bit of an in-flight packet.

The quoter modifications observed on the IPv4 header indicate artifacts of processing the packet. The most frequent quoter modification observed is to the IP length field; of the 174 modifications, 160 were to change the field from having a value of 40 (0x28) to a value of 60 (0x3c), while the remaining 14 changed it to a value of 0x2814. Both modifications suggest the length of the IP header was added during processing; in the second case, the field was byte-swapped first.

3.3 Modifications to TCP Headers

Table 1 shows that most modifications to the TCP header were not identified as quoter or in-flight modifications; therefore, we examine all unique modifications. 46 quoters modified the TCP source port; 10 of these quoters used port 1, while the rest chose values that were only seen once. 132 quoters modified the TCP destination port. Some values were observed from multiple quoters and show signs of port redirection; for example, port 81 was seen from 11 quoters, and port 8080 from 13 quoters. Other port values chosen were seen once or twice.

23 quoters revealed that an MSS option was added in-line to the TCP header, probably to work around paths with broken path MTU discovery. 15 quoters revealed an MSS of 536 bytes had been set, 5 an MSS of 1460, and a series of quoters revealed cases of 1360, 1414, and 1436.

3.4 Quote Clobbering

Some probe modifications (not listed in Table 1) are due to inadvertent clobbering or modification of the quote in the quoting process. We group these modifications into four categories. 71 quoters (0.1%) quoted the first 28 bytes of a probe correctly, but then clobbered more than half the remaining bytes in the quote. 203 quoters (0.4%) quoted 60 bytes in the response; as the probe was only 40 bytes in size, the remaining 20 bytes were from a previous user of the memory, as described in US CERT note VU#471084. An additional 4 quoters over-quoted by 10 bytes, although at least the first six extra bytes were zero. Finally, 14 quoters over-quoted by 8 bytes, and set the last 16 bytes to zero.

3.5 Observed RTTs

Some probe packets required a long wait for an ICMP response. 56 required arrived over 10 seconds after the probe was sent, 34 more than 100 seconds, and 30 more than 300 seconds. The reason for these long round trips is not obvious; perhaps the probe is triggering link establishment and the probe is forwarded when the link is complete.

4 Conclusion

This paper presents a methodology for analysis of ICMP quotations, and uses it to analyse a dataset collected with `tcptraceroute` to a large number of web servers. Many in-flight changes are able to be attributed to known packet rewriting techniques. In the data collected for this paper, relatively few quoters are inferred to modify packets in-flight, or indeed to modify them during processing.

References

1. Medina, A., Allman, M., Floyd, S.: Measuring interactions between transport protocols and middleboxes. In: Internet Measurement Conference. (October 2004)