Internet Area WG J. Touch
Internet Draft USC/ISI
Intended status: Updates 791,1122,2003 M. Mathis
Expires: September 2009 PSC
March 9, 2009

IPv4 ID Uniqueness Requirements
draft-touch-intarea-ipv4-unique-id-01.txt

Status of this Memo

Copyright Notice

Abstract

The IPv4 Identification (ID) field enables fragmentation and reassembly, but is required and must be unique within the maximum segment lifetime on all packets. If implemented as required, this uniqueness would limit all connections to 6.4 Mbps; since this is ubiquitously not the case, it is clear that existing systems violate the current requirement. This document updates the requirements for the IP ID field to more closely reflect current practice, and to more closely match IPv6, in which the field is defined only when a packet is actually fragmented. Even when fragmented, this document recommends that the ID field uniqueness consider the reordering context, rather than an arbitrary, unenforced upper bound on segment lifetime.

Table of Contents

1. Introduction

In IPv4, the IP Identification (ID) field is a 16-bit value that is unique for every packet for a given source address, destination address, and protocol, such that it does not repeat within the Maximum Segment Lifetime (MSL) [RFC791][RFC1122]. All packets between a source and destination of a given protocol must have unique ID values over a period of an MSL, which is typically interpreted as two

minutes (120 seconds). This uniqueness is currently specified as required for all packets, regardless of fragmentation settings.

The uniqueness of the IP ID is a known problem for high speed devices, because it limits the speed of a single protocol between two endpoints to 6.4 Mbps for typical MTUs of 1500 bytes [RFC4963]. This strongly indicates that the uniqueness of the IPv4 ID is moot.

This document updates the requirements for the IP ID field to more closely reflect current practice, and to more closely match IPv6, in which the field is defined only when a packet is actually fragmented. It also updates the recommended uniqueness interval to support the impact of reordering on reassembly, rather than using an arbitrary and unenforceable segment lifetime.

2. Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC-2119 [RFC2119].

In this document, the characters ">>" proceeding an indented line(s) indicates a compliance requirement statement using the key words listed above. This convention aids reviewers in quickly identifying or finding this RFC's explicit compliance requirements.

3. Current Requirements

IP supports packet fragmentation, where large packets are split into smaller components to traverse links with limited maximum transmission units (MTUs). Fragments are indicated in different ways in IPv4 and IPv6:

- o In IPv4, the header contains three fields: Identification (ID), Offset, a "Don't Fragment" flag (DF), and a "More Fragments" flag (MF) [RFC791]

- o In IPv6, fragments are indicated in an extension header that includes an ID, Offset, and MF flag similar to their counterparts in IPv4 [RFC2460]

IPv4 and IPv6 fragmentation differs in a few important ways. IPv6 fragmentation occurs only at the source, so a DF bit is not needed to prevent downstream devices from initiating fragmentation. The IPv6 fragment header is present only when a packet has been fragmented, so the fields - notably the ID field, as will be shown later - are not present for non-fragmented packets, and thus are meaningful only for

fragments. Finally, the ID field is 32 bits, and unique per source/destination address pair for IPv6, whereas for IPv4 it is only 16 bits and unique per source/destination/protocol triple.

This document focuses on the IPv4 ID field issues, because in IPv6 the field is larger and present only in fragments.

4. Uses of the ID Field in IPv4

The IPv4 ID field was originally intended for fragmentation and reassembly. Within a given source address, destination address, and protocol, fragments of an original packet are matched based on their IP ID. This requires that IDs are unique within the address/protocol triple when fragmentation is possible (e.g., DF=0).

The ID field has been discussed as useful in other ways. It can be used to detect and discard duplicate packets, e.g., at congested routers (see Sec. 3.2.1.5 of [RFC1122]).

The ID field may also be useful in tunnels. ICMP along tunnels may return only a portion of the information needed by a tunnel ingress to relay information back to the packet source. Encapsulators may retain copies of recently sent packets, to enable ICMP relaying [RFC2003].

These latter uses require that the IP ID be unique across all packets, not only when fragmentation is enabled. This document deprecates all such non-fragmentation uses.

5. IPv4 ID Exhaustion

With the maximum IPv4 packet size of 64KB, a 16-bit ID field that does not repeat within 120 seconds means that the sum of all TCP connections of a given protocol between two endpoints is limited to roughly 286 Mbps; at a more typical MTU of 1500 bytes, this speed drops to 6.4 Mbps [RFC4963]. This limit currently applies for all IPv4 packets, regardless of whether fragmentation is enabled, used, or inhibited.

Note that IPv6, even at typical MTUs, is capable of 18.7 Tbps when fragments are present, due to the larger 32-bit ID field. When fragmentation is not used, IPv6 speeds are not limited by the ID field uniqueness.

Note also that 120 seconds is only an estimate on the maximum segment lifetime. It is loosely based on half maximum value of the IP TTL field, which is represents 0-255 seconds, although it must be

decremented by 1 second for each router on a path even when held for less than a second [RFC791]. Network delays are incurred in other ways, e.g., satellite links, which can add seconds of delay even though the TTL is not affected. There is no enforcement mechanism to ensure that packets older than 120 seconds are discarded.

6. Current Practice

Wireless Internet devices are frequently connected at speeds over 54 Mbps, and wired links of 1 Gbps have been the default for several years. Although many end-to-end transport paths are congestion limited, these devices easily achieve 100+ Mbps application-layer throughput over LANs (e.g., disk-to-disk file transfer rates), and numerous throughput demonstrations have been performed with COTS systems at these speeds for over a decade. This strongly suggests that IPv4 ID uniqueness has been moot for a long time.

7. Recommended Practice

There are two kinds of packets, defined herein, for which recommended practice is described:

o Atomic packets: packets not yet having been fragmented (MF=0 and offset=0) and for which further fragmentation has been inhibited (DF=1), i.e.: ((DF==1)&&(MF==0)&&(offset==0))

o Non-atomic packets: packets which have either already been fragmented (MF=1 or offset>0 or both), or for which fragmentation remains possible (DF=0), i.e.: ((DF==0)||(MF==1)||(offset>0)), or (equivalently), ~((DF==1)&&(MF==0)&&(offset==0)).

Although at least one document suggests the ID field has other uses, it useful to confirm here that the ID field is defined only for fragmentation:

o >> Gateways (i.e., routers) and receiving hosts MUST ignore the contents of the IPv4 ID field for atomic packets. The egresses of IP encapsulation tunnels act as receiving hosts, and thus MUST follow this requirement.

o >> The IPv4 ID field MUST NOT be utilized for purposes other than fragmentation and reassembly.

Fragments that repeat the IP ID risk being reassembled incorrectly, especially when fragments are reordered or lost [RFC4459]. Although such errors may be detected at the transport layer, this results in

excessive overall packet loss, as well as wasting network bandwidth. As a result, this document notes that:

o >> Hosts emitting non-atomic IPv4 packets MUST set the ID field uniqely per source IP, destination IP, and protocol tuple, sufficient to support reassembly.

Note that "sufficient to support reassembly" need not require unique IDs over a two minute interval. It should be sufficient that:

o >> Hosts emitting non-atomic IPv4 packets SHOULD NOT repeat ID field values within a given source IP, destination IP, and protocol tuple over the period that the receiver is expected to experience fragment reordering.

Note that it is impossible to ensure a "MUST NOT" in this requirement, because there is no strict enforcement on segment lifetime; as a result the requirement is listed as a "SHOULD NOT" only.

This suggests that the host employ rate limiting on each source/estination/protocol triple. The recommendations above are most appropriate at the host (or tunnel ingress), and can be difficult to enforce at routers. As a result, we recommend that for IPv4, as for IPv6:

o >> IPv4 fragmentation SHOULD be limited to the originating source, e.g., the host or tunnel ingress. IPv4 fragmentation SHOULD NOT be performed where the IPv4 ID field is not under direct control, e.g., at routers.

Note, however, that it may not be possible for applications to know whether any of the above three requirements are satisfied at a host or on tunnels along a path (esp. those employing outer fragmentation). As a result, we recommend that:

o >> Hosts unable to meet the non-repeating IP ID requirement above MUST NOT fragment outgoing IP packets, and MUST also set the DF flag to prevent subsequent fragmentation.

o >> Applications that cannot ensure safe IPv4 ID generation and that allow DF=0 SHOULD employ integrity checks that would detect mis-reassembled fragments, e.g, as in SEAL [Te??]0. E.g., applications SHOULD NOT use UDP without checksums [RFC793], and SHOULD be very careful in their use of UDP-Lite [RFC3828] in such environments, even existing UDP and TCP checksums may not be sufficient [RFC4963].

- >> Applications SHOULD set DF=1 for all packets exiting a source host, regardless of whether those packets are fragmented at the source or not.

[should this document also deprecate overlapping fragments?]

8. Updates to Existing Standards

The following sections address the specific changes to existing protocols indicated by the requirements in this document.

8.1. Updates to RFC 791

[to be completed]

8.2. Updates to RFC 1122

[to be completed]

8.3. Updates to RFC 2003

[to be completed]

9. Impacts on NATs and Tunnel Ingresses

Network Address (and port) Translators (NATs) rewrite IP fields, and tunnel ingresses (using IP encapsulation) copy and modify some IP fields, so both need to follow host requirements. As a result:

>> NATs MUST NOT ignore the DF bit.

>> NATs SHOULD NOT fragment, even when allowed by the DF bit.

>> Tunnel ingresses MUST NOT ignore the DF bit of the interior packet.

>> Tunnels that fragment MUST do so by fragmenting the outer IP header; they MUST NOT fragment the inner header even when allowed by the DF bit.

[further discussion would be useful, esp. of carrier-grade NATs]

10. Transitioning to These New Requirements

During the transition period, there may continue to be tunnel ingresses and NATs that fragment even when the DF bit is set. It may be useful to use a small ID space to help detect such behavior without causing full disruption, as might occur by using a single value when the DF flag is set (e.g., 0).

As a result, during the transition period, this document recommends that:

>> During the transition period, a small ID space SHOULD be used to assist with debugging and detection; such a space SHOULD use the lower bits (i.e., lower 4 bits) of the ID field and clear (i.e., zero) the remaining high order bits.

11. Security Considerations

This document attempts to address the security considerations associated with fragmentation in IPv4 [RFC4459].

When the IPv4 ID is ignored on receipt (e.g., for atomic packets), its value becomes unconstrained; that field then more easily be used as a covert channel.

[talk about the impact on steganography - if the ID field is ignored, should it be set to zero at any given hop arbitrarily? Should a security gateway set it to zero to prevent a covert channel?]

12. IANA Considerations

There are no IANA considerations in this document.

The RFC Editor should remove this section prior to publication

13. References

13.1. Normative References

[RFC791] Postel, J., "Internet Protocol," RFC 791 / STD 5, September 1981.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

13.2. Informative References

[RFC793] Postel, J., "User Datagram Protocol," RFC 793 / STD 6, August 1980.

[RFC1122] Braden, R., Ed., "Requirements for Internet Hosts - Communication Layers," RFC 1122 / STD 3, October 1989.

[RFC2003] Perkins, C., "IP Encapsulation within IP," RFC 2003, October 1996.

[RFC2460] Deering, S., R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification," RFC 2460, December 1998.

[RFC3828] Larzon, L-A., M. Degermark, S. Pink, L-E. Jonsson, Ed., G. Fairhurst, Ed.L., "The Lightweight User Datagram Protocol (UDP-Lite)," RFC 3828, July 2004.

[RFC4459] Savola, P., "MTU and Fragmentation Issues with In-the-Network Tunneling," RFC 4459, April 2006.

[RFC4963] Heffner, J., M. Mathis, B. Chandler, "IPv4 Reassembly Errors at High Data Rates," RFC 4963, July 2007.

[Te??] Templin, F., Ed., "The Subnetwork Encapsulation and Adaptation Layer (SEAL)," (work in progress), draft-templin-seal-22, June 2008.

14. Acknowledgments

This document was inspired by of numerous discussions among the authors, Jari Arkko, Lars Eggert, Dino Farinacci, and Fred Templin, as well as members participating in the Internet Area Working Group.

This document was prepared using 2-Word-v2.0.template.dot.

Authors' Addresses

Joe Touch
USC/ISI
4676 Admiralty Way
Marina del Rey, CA 90292-6695
U.S.A.

Phone: +1 (310) 448-9151
Email: touch@isi.edu

Matt Mathis
PSC
300 South Craig st.
Pittsburgh PA, 15213
U.S.A.

Phone: +1 (412) 268-3319
Email: mathis@psc.edu