Internet Area WG
Internet Draft
Updates: 791,1122,2003
Intended status: Proposed Standard
Expires: September 2010

J. Touch
USC/ISI

March 5, 2010

Updated Specification of the IPv4 ID Field
draft-touch-intarea-ipv4-unique-id-03.txt

Status of this Memo

Abstract

The IPv4 Identification (ID) field enables fragmentation and reassembly, and as currently specified is required to be unique within the maximum lifetime on all IP packets. If enforced, this uniqueness requirement would limit all connections to 6.4 Mbps. Because this is obviously not the case, it is clear that existing systems violate the current specification. This document updates the specification of the IP ID field to more closely reflect current practice and to more closely match IPv6, so that the field is defined only when a packet is actually fragmented and that fragmentation occurs only at originating hosts or their equivalent. When fragmentation occurs, this document recommends that the ID field be unique within the reordering context, rather than an arbitrary, unenforced upper bound on packet 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 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, as has already been noted.

This document updates the specification of 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 and in which fragmentation occurs only at the source. It also updates the recommended uniqueness interval to support the impact of reordering on reassembly, rather than using an arbitrary and unenforceable packet 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 requirement using the key words listed above. This convention aids reviewers in quickly identifying or finding this document's explicit requirements.

3. The IPv4 ID Field

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:

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

- In IPv6, fragments are indicated in an extension header that includes an ID, Fragment 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 ID field is not present for non-fragmented packets, and thus is 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 IPv4 ID Field

The IPv4 ID field was originally intended for fragmentation and reassembly [RFC791]. 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 can also be useful for duplicate avoidance and ICMP validation. The field can be used at routers or receiving hosts to remove duplicate packets. The IP ID field can be used to validate payloads of ICMP responses as matching the originally transmitted packet at a host [RFC4963]. At a tunnel ingress, the ID enables returning ICMP messages to be matched to a cache of recently transmitted packets, to support 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. Background on IPv4 ID Reassembly Issues

The following is a summary of issues with IPv4 fragment reassembly in high speed environments raised previously [RFC4963]. Readers are encouraged to consult RFC 4963 for a more detailed discussion of these issues.

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 or inhibited, and whether a packet is fragmented or not.

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 the field is absent, and so IPv6 speeds are not limited by the ID field uniqueness.

Note also that 120 seconds is only an estimate on the maximum packet 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.

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.

6. Updates to the IPv4 ID Specification

This document updates the specification of the IPv4 ID field in three distinct ways, as discussed in subsequent subsections:

- Use the ID field only for fragmentation
- Avoid ID repetition and its impacts
- Encourage more safe use of the ID field

There are two kinds of packets used in the following discussion:

Atomic packets: packets not yet having been fragmented (MF=0 and fragment offset=0) and for which further fragmentation has been inhibited (DF=1), i.e., as a C-code expression:

```
(DF==1)&&(MF==0)&&(frag_offset==0)
```

- Non-atomic packets: packets which have either already been fragmented, i.e.:

```
(MF=1)||(frag_offset>0)
```

or for which fragmentation remains possible (DF=0), i.e.:

(DF==0)||(MF==1)||(frag_offset>0)

or (equivalently):

~((DF==1)&&(MF==0)&&(frag_offset==0))

6.1. IPv4 ID Used Only for Fragmentation

Although at least one document suggests the ID field has other uses, we assert here that the ID field is defined only for fragmentation and reassembly.

- >> The IPv4 ID field of MUST be ignored except for packet reassembly.

Such devices typically include receiving hosts and tunnel egresses, but may include any intermediate device that reassembles a packet, such as a firewall or NAT. The ID field is thus meaningful only for non-atomic packets that have actually been fragmented, either at the source or elsewhere along the path, and have not been reassembled before being examined. In atomic packets, the ID field has no

meaning, and thus its values are always to be ignored. Atomic packets are detected by their DF, MF, and fragmentation offset fields as defined in Section 6, because such a test is completely backward compatible; this document thus does not reserve any ID values, including 0, as distinguished.

Note that this excludes some current practices that use the ID field and the remainder of the IP header as a unique tag. This tag has been suggested as a way to detect and remove duplicate packets, e.g., at congested routers, although this has been noted and no current deployments are known [RFC1122]. Some hosts use this tag to validate received ICMPs, in which the ICMP payload – an IP packet prefix – is matched against a cache of recently transmitted IP headers. This ensures that the received ICMP reflects a transmitted packet, though it does not prevent spoofing of ICMPs for attackers that can see those packets, and like ID reuse will cause problems at high packet rates. A similar sort of matching can be used in tunnels, to enable ICMP relaying at the tunnel ingress, with similar challenges

[RFC2003].

Deprecating the use of the IPv4 ID field for these non-reassembly uses should have little - if any - impact. IPv4 IDs are already frequently repeated, e.g., over even moderately fast connections. Duplicate suppression was only suggested, and no impacts of ID reuse have been noted. Routers are not required to issue ICMPs on any particular timescale, and so ID repetition should not have been used for validation, and again repetition occurs and probably could have been noticed [RFC1812]. ICMP relaying at tunnel ingresses is specified to use soft state rather than a packet cache, and should have been noted if the latter for similar reasons [RFC2003].

6.2. Avoiding IPv4 ID Repetition and Its Impacts

This document specifies that IPv4 be modified to more closely match IPv6's fragmentation constraints, to permit fragmentation only at devices that control the uniqueness of the IP ID field, e.g., sources, tunnel ingresses (for the outer header), and packets emitted from a NAT to its public side (see Section 8).

- >> Sources SHOULD set DF=1.

- >> IPv4 fragmentation SHOULD be limited to the originating source, even when the DF field allows it.

Keep in mind that a source is any device that uses one of its assigned IP addresses as a source IP address in emitted packets. This

includes hosts, routers when originating packets, packets emitted from NATs (see Section 8), and tunnel ingresses.

It may not be possible for sources to know whether all of the above specifications are satisfied. As a result, we recommend that:

- >> Sources unable to meet the non-repeating IP ID requirement above MUST NOT emit non-atomic packets.

In other words, such sources can emit only non-fragmented packets where DF has been set. Such sources can repeat the ID field for atomic packets, as it is intended to be ignored.

Sources emitting non-atomic IPv4 packets need to set the ID field sufficient to support reassembly, and encourages the use of stronger transport layer validation where possible. Uniqueness over a two minute interval may be excessive to support reassembly in some environments, and is clearly already being ignored.

- >> Sources 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 fragment reordering would affect reassembly.

It is impractical to assert "MUST NOT" here, because there is no strict enforcement on packet lifetime and because sources may not be able to determine the reordering period.

- >> Sources 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 [RFC5320]. Applications SHOULD NOT use UDP without checksums [RFC793], and SHOULD be very careful in their use of UDP-Lite [RFC3828] in such environments.

Additional integrity checks can be employed using tunnels, as in SEAL, IPsec, or SCTP [RFC4301][RFC4960][RFC5320]. Such checks can avoid the reassembly hazards that can occur when using UDP and TCP checksums [RFC4963].

6.3. Encourage Safe ID Use

This document makes further changes to the specification of the IPv4 ID field and its use to encourage its safe use as follows.

RFC 1122 discusses that TCP retransmits a segment it may be possible to reuse the IP ID (see Section 7.2). This can make it difficult for a source to avoid ID repetition for received fragments. RFC 1122

concludes that this behavior "is not useful"; this document formalizes that conclusion as follows:

- >> The IP ID MUST NOT be reused when sending a copy of an earlier non-ATOMIC packet.

RFC 1122 also suggests that fragments can overlap [RFC1122]. Such overlap can occur if successive retransmissions use different

packetizing but the same reassembly Id.

This overlap is noted as the result of reusing IDs when retransmitting packets, which this document deprecates. Overlapping fragments are themselves a hazard [RFC4963]. As a result:

o >> Overlapping packets MUST be silently ignored during reassembly.

7. Updates to Existing Standards

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

7.1. Updates to RFC 791

RFC 791 states that:

> The originating protocol module of an internet datagram sets the identification field to a value that must be unique for that source-destination pair and protocol for the time the datagram will be active in the internet system.

And later that:

> Thus, the sender must choose the Identifier to be unique for this source, destination pair and protocol for the time the datagram (or any fragment of it) could be alive in the internet.
>
> It seems then that a sending protocol module needs to keep a table of Identifiers, one entry for each destination it has communicated with in the last maximum packet lifetime for the internet.
>
> However, since the Identifier field allows 65,536 different values, some host may be able to simply use unique identifiers independent of destination.
>
> It is appropriate for some higher level protocols to choose the identifier. For example, TCP protocol modules may retransmit an identical TCP segment, and the probability for correct reception

> would be enhanced if the retransmission carried the same identifier as the original transmission since fragments of either datagram could be used to construct a correct TCP segment.

This document changes RFC 791 as follows:

- >> The IP ID is not defined if the packet (datagram) is atomic. IP packet sources MAY use any value as ID; all such values MUST BE ignored on examination at intermediate nodes and destinations.

- >> The IP ID of non-atomic packets MUST BE unique for the time where fragments are expected to overlap.

- >> Hosts SHOULD emit only atomic packets (i.e., not fragmented at the source, and with DF=1).

We do not expect that it will be useful to involve higher-level protocols in determining ID values.

7.2. Updates to RFC 1122

RFC 1122 states that:

> 3.2.1.5 Identification: RFC-791 Section 3.2
>
> When sending an identical copy of an earlier datagram, a host MAY optionally retain the same Identification field in the copy.
>
> DISCUSSION:
>
> Some Internet protocol experts have maintained that when a host sends an identical copy of an earlier datagram, the new copy should contain the same Identification value as the original. There are two suggested advantages: (1) if the datagrams are fragmented and some of the fragments are lost, the receiver may be able to reconstruct a complete datagram from fragments of the original and the copies; (2) a congested gateway might use the IP Identification field (and Fragment Offset) to discard duplicate datagrams from the queue.

This document changes RFC 1122 as follows:

- >> The IP ID field MUST NOT be used for duplicate detection or removal.

- o >> IP ID values MUST NOT be repeated when packets are retransmitted.

- o >> IP packet fragments MUST NOT overlap.

7.3. Updates to RFC 1812

There are no updates to RFC1812.

7.4. Updates to RFC 2003

RFC 2003 states that:

> Identification, Flags, Fragment Offset
>
> These three fields are set as specified in [RFC791]. However, if the "Don't Fragment" bit is set in the inner IP header, it MUST be set in the outer IP header; if the "Don't Fragment" bit is not set in the inner IP header, it MAY be set in the outer IP header, as described in Section 5.1.

This document changes RFC 2003 as follows:

- o >> IP-in-IP tunnels SHOULD emit only atomic packets.

Note that this recommendation applies to all tunnels, but the focus of this document is IPv4 requirements, so its explicit requirements focus on IPv4 cases.

8. Impacts on NATs and Tunnel Ingresses

Network address translators (NATs) and address/port translators (NAPTs) rewrite IP fields, and tunnel ingresses (using IP encapsulation) copy and modify some IP fields, so all are considered sources, as do any devices that rewrite any portion of the IP source, IP destination, IP protocol, and IP ID tuple for non-atomic packets [RFC3022]. As a result, they are subject to all the requirements of any source, as has been noted.

NATs present a particularly challenging situation for fragmentation. Because NATs overwrite portions of the reassembly tuple in both directions, they can destroy tuple uniqueness and result in a reassembly hazard. Not only do NATs need to behave as a source for the purposes of this document, but also:

- o >> NATs MUST either silently drop fragments or reassemble them before translating and emitting them.

Problems with transmitting fragments through NATs are already known; translation is based on the transport port number, which is present in only the first fragment anyway [RFC3022]. This document underscores the point that not only is reassembly (and possibly subsequent fragmentation) required for translation, it is required for IP ID uniqueness.

Note that NATs/NAPTs already need to exercise special care when emitting packets on their public side, because merging packets from many sources onto a single outgoing source IP address can result in IP ID collisions. This situation precedes this document, and is not affected by it. It is exacerbated in large-scale, so-called "carrier grade" NATs [Ni09].

Tunnel ingresses act as sources for the outermost header, but tunnels act as routers for the inner headers (i.e., the packet as arriving at the tunnel ingress). Ingresses can fragment as originating sources of the outer header, because they control the uniqueness of that IP ID field. They need to avoid fragmenting the packet at the inner header, for the same reasons as any intermediate device, as noted elsewhere in this document.

9. Impact on Header Compression

Header compression algorithms already accommodate various ways in which the IP ID changes between sequential packets. Such algorithms already need to preserve the IP ID. This document relaxes that constraint, making preservation optional for most atomic packets as a result:

>> Header compression MAY preserve the IP ID of atomic packets that are not protected by IPsec AH [RFC4302]. The IP ID of non-atomic packets, and those of packets protected by IPsec AH MUST be preserved.

Note that this can impact the efficiency of header compression in various ways. When compression can assume a nonchanging ID, efficiency can be increased. However, when compression assumes a changing ID as a default, having a non-changing ID can make compression less efficient (see footnote 21 of [RFC1144], which is optimized for non-atomic packets).

10. Transitioning to This Update

?? Do we need this transition?

?? Do we want to say when to stop the transition?

During the transition period, there may continue to be tunnel ingresses and NATs that fragment even when the DF bit is set, or that validate ICMP payloads based on cached packets. It may be useful to use a small ID space to help detect such behaviors without causing full disruption, as might occur by using a single value when the DF flag is set (e.g., ID=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 can more easily be used as a covert channel. For some atomic packets - notably those not protected by IPsec Authentication Header (AH) [RFC4302] - it is possible, and may be desirable, to rewrite the ID field to avoid its use as such a channel.

The IP ID also now adds much less entropy of the header of an IP packet. The ID had previously been unique (for a given IP source/address pair, and protocol field) within 2MSL, although this requirement was not enforced and clearly is typically ignored. IDs of non-atomic packets are now required unique only within the expected reordering of fragments, which could substantially reduce the amount of entropy in that field. The IP ID of atomic packets is not required unique, and so contributes no entropy to the header.

The deprecation of the ID field's uniqueness for atomic packets can defeat the ability to count devices behind a NAT [Be02]. This is not

intended as a security feature, however.

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.

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

[RFC1812] Baker, F. (Ed.), "Requirements for IP Version 4 Routers", RFC 1812 / STD 4, Jun. 1995.

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

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

13.2. Informative References

[Be02] Bellovin, S., "A Technique for Counting NATted Hosts", Internet Measurement Conference, Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurement, November 2002.

[Ni09] Nishitani, T., I. Yamagata, S. Miyakawa, A. Nakagawa, H. Ashida, "Common Functions of Large Scale NAT (LSN) ", (work in progress), draft-nishitani-cgn-03, Nov. 2009.

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

[RFC1144] Jacobson, V., "Compressing TCP/IP Headers", RFC 1144, Feb.

1990.

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

[RFC3022] Srisuresh, P. and K. Egevang, "Traditional IP Network Address Translator (Traditional NAT)", RFC 3022, January 2001.

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

[RFC4301] Kent, S., K. Seo, "Security Architecture for the Internet Protocol", RFC 4301, Dec. 2005.

[RFC4302] Kent, S., "IP Authentication Header", RFC 4302, Dec. 2005.

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

[RFC4960] Stewart, R. (Ed.), "Stream Control Transmission Protocol", RFC 4960, Sep. 2007.

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

[RFC5320] Templin, F., Ed., "The Subnetwork Encapsulation and Adaptation Layer (SEAL)", RFC 5320, Feb. 2010.

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. Detailed feedback was provided by Carlos Pignataro. This document originated as an Independent Stream draft co-authored by Matt Mathis, PSC, and his contributions are greatly appreciated.

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

Author's Address

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