Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: April 19, 2013

J. Jaeggli
Zynga
L. Colitti
W. Kumari
Google
E. Vyncke
Cisco
M. Kaeo
Double Shot Security
T. Taylor, Ed.
Huawei Technologies

October 16, 2012

# Why Operators Filter Fragments and What It Implies
## draft-taylor-v6ops-fragdrop-00

## Abstract

This memo is written to make application developers and network operators aware of the significant probability that IPv6 packets containing fragmentation extension headers will fail to reach their destination. Some assumptions about the ability to use TCP or UDP datagrams larger than a single packet may accordingly need adjustment. This memo provides observational evidence for the dropping of IPv6 fragments along a significant number of paths, explores the operational impact of fragmentation and the reasons why dropping occurs, and considers the effect of fragment dropping on applications particularly including DNS.

## Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are 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."

This Internet-Draft will expire on April 19, 2013.

## Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the document authors.  All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document.  Please review these documents carefully, as they describe your rights and restrictions with respect to this document.  Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Table of Contents

## 1. Introduction

Measurements of whether internet service providers and edge networks deliver IPv6 fragments to their destination reveal that for IPv6 in particular, fragments are being dropped along a substantial number of paths. IPv6 datagrams with fragmentation headers are a non-issue in the core of the internet, where fragments are routed just like any other IPv6 datagram. However, fragmentation creates operational issues at the edge of the network that may lead to administratively imposed filtering or inadvertent failure to deliver the fragment to the application.

Section 2 begins with some observations on how often IPv6 fragment loss occurs in practice. We go on to look at the operational reasons for filtering fragments, a key aspect of which is the threats they pose to security policy and appliances. Section 2.2 then looks at the impact on key applications, particularly DNS.

In the longer run, as network operators gain a better understanding of the risks and non-risks of fragmentation and as middlebox, customer premise equipment (CPE), and host implementations improve, we believe that some incidences of fragment dropping will diminish. However, some of the justifications for filtering will persist in the longer term, and application developers must remain aware of the implications.

This document deliberately refrains from discussing possible responses to the problem posed by the dropping of IPv6 fragments. Such a discussion will quickly turn up a number of possibilities, application-specific or more general; but the amount of time needed to specify and deploy a given resolution will be a major constraint in choosing amongst them. In any event, that discussion is likely to proceed in multiple directions and is therefore considered beyond the scope of this memo.

## 2. Observations and Rationale

[Blackhole] is a good public reference for the incidence of IPv6 fragment filtering. It describes experiments run to determine the incidence and location of ICMP Packet Too Big and fragment filtering. The authors used fragmented DNS packets to determine the latter and found for IPv6 that filtering appeared to be occurring on some 10% of the tested paths. The filtering appeared to be located at the edge (enterprise and customer networks) rather than in the core.

[Co-authors, more to contribute?]

### 2.1. Possible Causes

Why does such filtering happen? One cause is non-conforming implementations in CPE and low-end routers. Along with that, some network managers filter fragments on principle, without taking account of the specific risks involved, just as they may filter ICMP Packet Too Big. Both implementations and management should improve over time, reducing the problem somewhat.

Some filtering and dropping of fragments is done for hardware, performance, or topological considerations. Stateful inspection devices or destination hosts can experience resource exhaustion if they are flooded with fragments not followed by the remaining fragments of the unfragmented packet. Stateless ACLs may be difficult to apply to fragments other than the one in which the upper layer header is present. As [Attacks] demonstrates, inconsistencies in reassembly logic between middleboxes or CPEs and hosts can cause fragments to be wrongfully discarded, or can allow exploits to pass undetected through middleboxes. Stateless Load balancing schemes may hash fragmented datagrams from the same flow to different paths because the 5-tuple may available on only the initial fragment.

Leaving aside these incentives towards fragment dropping, other considerations may weigh on the operator's mind. One example cited on the NANOG list was that of a router where fragment processing was done by the control plane processor rather than in the forwarding plane hardware, with a consequent hit on performance. Another incentive toward dropping of fragments is the disproportionate number of software errors still being encountered in fragment processing. Since this code is exercised less frequently than the rest of the stack, bugs remain longer in the code before they are detected. Some of these software errors can introduce vulnerabilities subject to exploitation. It is common practice [RFC6192] to recommend that control-plane ACLs protecting routers and network devices be configured to drop all fragments.

Operators weigh the risks associated with each of the considerations just enumerated, and come up with the most suitable policy for their circumstances. It is likely that at least some operators will find it desirable to drop fragments in at least some cases.

The IETF can help this effort by identifying specific classes of fragments that do not represent legitimate use cases and hence should always be dropped. Examples of this work are given by [draft-6man-atomic] and [I-D.ietf-6man-oversized-header-chain]. The problem of inconsistent implementations may also be mitigated by providing further advice on the more difficult points. However, some cases will remain where legitimate fragments are discarded for

legitimate reasons. The potential problems these cases pose for applications is our next topic.

### 2.2. Impact on Applications

Some applications can live without fragmentation, some cannot. DNS is one application that may be vulnerable when fragment dropping occurs. EDNS0 extensions [RFC2761] allow for responses in UDP PDU greater than 512 bytes. Particularly with DNSSEC, responses may be larger than the MTU and fragmentation at the sending host in order to respond using UDP is desirable and legal. The current choices open to the operators of DNS servers in this situation are to defer deployment of DNSSEC, fragment responses, or use TCP if there are cases where the rrset would be expected to exceed the MTU. The use of fallback to TCP will impose a major resource and performance hit and increases vulnerability to denial of service attacks.

Other applications, such as Network File System, NFS, are also known to fragment their large UDP packets but this is most often kept within a single organization network and should not be impacted by fragment dropping at the Internet core or edges.

## 3. Acknowledgements

TBD.

## 4. IANA Considerations

This memo includes no request to IANA.

## 5. Security Considerations

[Obviously a few things to say here, or we can find a few good references.]

## 6. Informative References

[Attacks] Atlasis, A., "Attacking IPv6 Implementation Using Fragmentation", March 2012.

http://media.blackhat.com/bh-eu-12/Atlasis/bh-eu-12-Atlasis-Attacking_IPv6-WP.pdf

[Blackhole]

de Boer, M. and J. Bosma, "Discovering Path MTU black holes on the Internet using RIPE Atlas", July 2012.

http://www.nlnetlabs.nl/downloads/publications/pmtu-black-holes-msc-thesis.pdf

[I-D.ietf-6man-ipv6-atomic-fragments]
Gont, F., "Processing of IPv6 "atomic" fragments", draft-ietf-6man-ipv6-atomic-fragments-01 (work in progress), August 2012.

[I-D.ietf-6man-oversized-header-chain]
Gont, F. and V. Manral, "Security and Interoperability Implications of Oversized IPv6 Header Chains", draft-ietf-6man-oversized-header-chain-01 (work in progress), July 2012.

[RFC2761] Dunn, J. and C. Martin, "Terminology for ATM Benchmarking", RFC 2761, February 2000.

[RFC6192] Dugal, D., Pignataro, C., and R. Dunn, "Protecting the Router Control Plane", RFC 6192, March 2011.

[draft-6man-atomic]
Gont, F., "Processing of IPv6 "atomic" fragments (Work in progress)", August 2012.

## Authors' Addresses

Joel Jaeggli
Zynga
924 Mouton Circle
East Palo Alto, CA  94303
USA

Email: jjaeggli@zynga.com

Lorenzo Colitti
Google

Phone:
Email: lorenzo@google.com

Warren Kumari
Google
1600 Amphitheatre Parkway
Mountain View, CA  94043
USA

Phone:
Email: warren@kumari.net

Eric Vyncke
Cisco
De Kleetlaan 6A
Diegem,   1831
Belgium

Phone:
Email: evyncke@cisco.com

Merike Kaeo
Double Shot Security

Phone:
Email: merike@doubleshotsecurity.com

Tom Taylor (editor)
Huawei Technologies
Ottawa, Ontario
Canada

Phone:
Email: tom.taylor.stds@gmail.com