フラグメンテーションの今後を考えよう

2014/4/18

JANOG33.5 Interim Meeting

松平直樹
富士通株式会社

自己紹介

- SA46T技術ファミリーの提案者です
  - SA46T
  - SA46T-PR
  - SA46T-PT
  - SA46T-AS
  - SA46T-AT

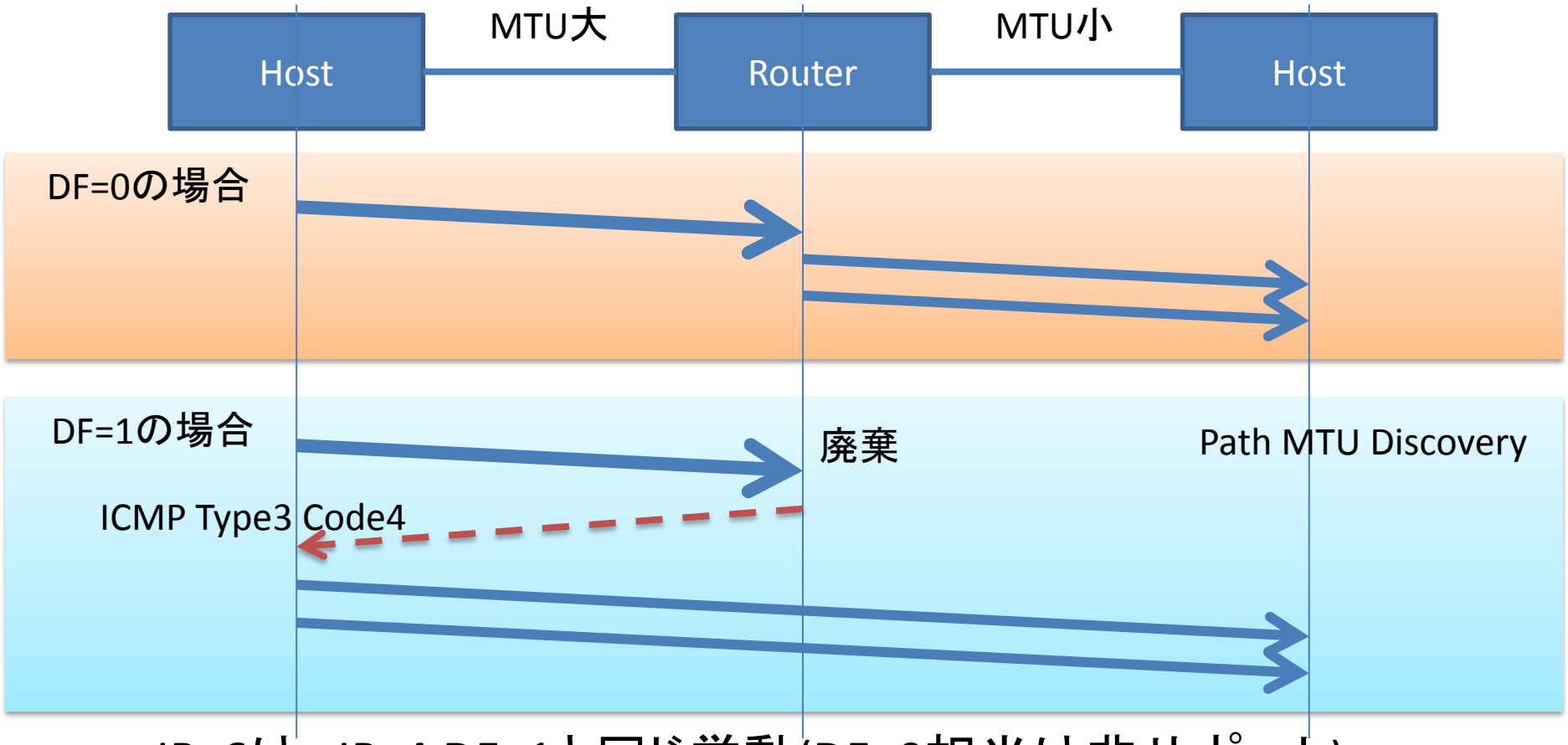
SA46T

Stateless Automatic IPv4 over IPv6 Tunneling

おおまかな流れ

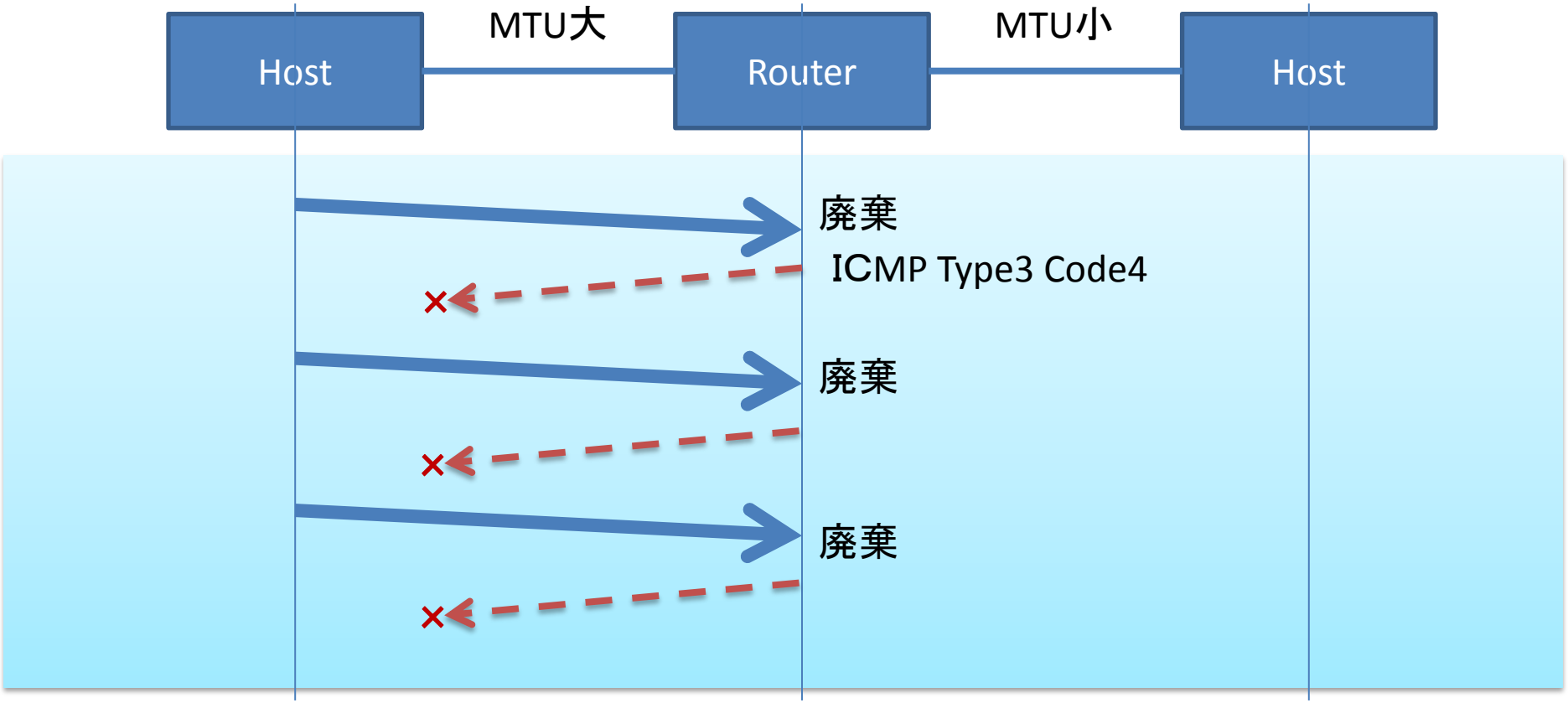
- フラグメンテーション技術のおさらい
- 過去に起きた問題とその対策
- これから起きそうな問題
- 可能性のある対処法
- 最近のIETFの状況
- 測定データのご紹介
- どうしていきましょうか？

フラグメンテーション(IPv4)



- IPv6は、IPv4 DF=1と同じ挙動(DF=0相当は非サポート)
  - IPv6には、そもそもDFビット相当は存在しない

PMTU ブラックホール



- ICMPエラーメッセージがフィルタ(廃棄)される
- Path MTU長が発信ホストに伝わらないので廃棄されるサイズで送信を繰り返す
- 永遠に通信できない
- 何故、フィルタされるのか? あるいはフィルタしていそうなところと、していないところがあるのか?

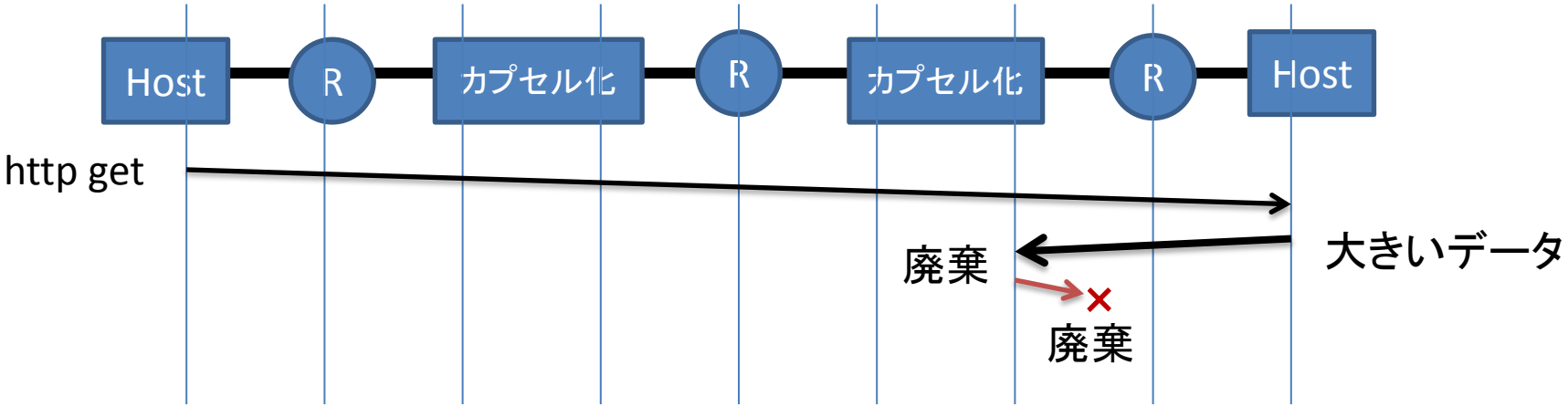
フラグメンテーション関連で過去に何が起きたか

- 何故、PMTUDが追加されたか
  - FDDI
- TCPの挙動とUDPの挙動
  - NFS
- 過去、大きく2度話題になった
  - 何が起き、どう対処したか／しなかったか:TCP MSS
  - 問題にならなかった時代:PPP

フラグメンテーション関連でこれからどういう問題が予測されるか

- DNS EDNS0, DNSSEC
- IPv4-IPv6移行技術(カプセル化、IPv4-IPv6変換)
- NVO3 (VxLAN, NVGRE, STT等)

カプセル化とフラグメンテーション



- アクセスできるサーバとアクセスできないサーバの存在
- アクセスできないサーバは、TCP/MSSの操作で、アクセス可能に
  - ICMP Type3 Code4メッセージのフィルタリングと推測

可能性のある対処法(現実的かどうかはともかく)

- アプリケーション層
  - NFSのような方法(確か：記憶が定かなら)
  - UDPを使用しているアプリをTCPを使うように変更する(DNS等)
- トランスポート層
  - TCP/MSS(解決策というより緊急避難?, IPsec時破綻)
  - UDPやGREの解は無い
  - (ICMPv4/ICMPv6は大丈夫か??)
  - RFC4821: “Packetization Layer Path MTU Discovery”
- ネットワーク層
  - DFを0に書き換える(規約違反：解決策というより緊急避難?)
- 物理層／データリンク層
  - Ethernetのジャンボフレームを使う
  - FDDI, ATMを使う
  - 最低1500Byteのパケットが届くように網設計する

フラグメントが悪であるという議論

- フラグメントの貧弱性
- iw2013: DNSのメッセージサイズについて考える～ランチのおともにDNS～
  - 第一フラグメント便乗攻撃
  - http://jprs.jp/tech/material/iw2013-lunch-L3-01.pdf

IETF会議に於ける議論の状況

- 85th IETF (Atlanta): 2012/11
  - v6ops: Why Operators Filter Fragments
- 86th IETF (Orlando): 2013/3
- 87th IETF (Berlin): 2013/8
  - 6man: IPv6 Fragment Header Deprecated
  - intarea: GRE MTU
- 88th IETF (Vancouver): 2013/11
  - IEPG88: Fragmentation and Extension Header Support in the IPv6 Internet
- 89th IETF (London): 2014/3
  - 6ops: Why Operators Filter Fragments
  - intarea: GRE MTU

関連Internet Draft

- Why Operators Filter Fragments and What It Implies
  - draft-taylor-v6ops-fragdrop-02
- IPv6 Fragment Header Deprecated
  - draft-bonica-6man-frag-deprecate-02
- A Fragmentation Strategy for Generic Routing Encapsulation (GRE)
  - draft-bonica-intarea-gre-mtu-04

IPv6フラグメントヘッダ廃止の議論

- 87th IETF Berlinの6man WGで提案
- 背景
  - ICMPv6 Packet Too Bigのフィルタリング
  - 拡張ヘッダのついたIPv6パケットのフィルタリング
- 上位レイヤでの対応を期待
  - PLPMTDU(RFC4821)
- 建設的な提案なのか？
  - 悲鳴なのでは？
- 現在、I-DはExpire

PLPMTUD

- Packetization Layer Path MTU Discovery
- 56th IETF San Francisco (2003/3)でBOF開催
- 2007/3にRFC4821として発行
- Packetization Layerで、Path MTUをprobeする
  - Loss reporting mechanisms
  - congestion control algorithms, rate limited
  - diagnostic tools
- 小さいMTUから少しずつ大きくしていく
- PMTUDが動く場合は容易に学習可能PMTUに合わせてパケット化
  - PMTUDが動くにこしたことはない

GRE MTU

- GRE(RFC2784)に於いてフラグメントに関する記載が不足(ベンダ依存)
- 現状の調査と整理
- RFC2784の改版を目的としない？
- 取り得る処理
  - ペイロードを廃棄(PMTUD/PLMTUD)
  - ペイロードをフラグメントしてカプセル化
  - カプセル化後にフラグメント

Why Operators Filter Fragments

- 仮説
  - 実装の問題と運用の問題の双方が原因
- 具体的な記載
  - Stateful inspection
    - 組み立てなおすことによる性能劣化
  - Stateless ACLs
    - 2個目以後のフラグメント
  - Performance
    - Forwarding planeでなくControl planeで処理
  - Other
    - バグなど
- 現在、活動が見えなくなっている模様

測定データのご紹介

- 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

ICMP PTBのフィルタ(MTU=1500)

IPv4

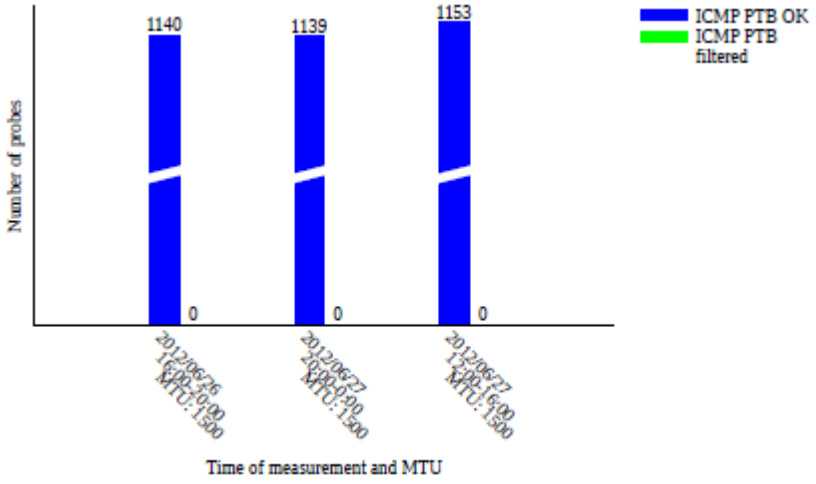


Figure 10: ICMPv4 PTB filtering - MTU: 1500

IPv6

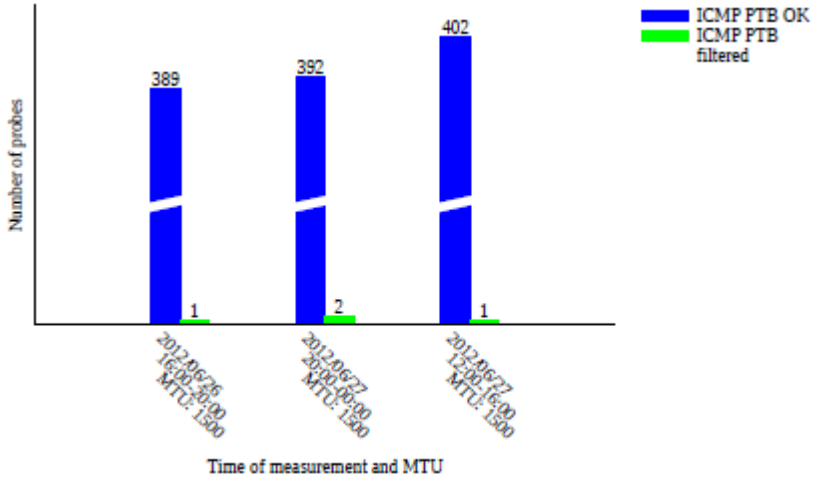


Figure 11: ICMPv6 PTB filtering - MTU: 1500

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

ICMP PTBのフィルタ(MTU=1280)

IPv4

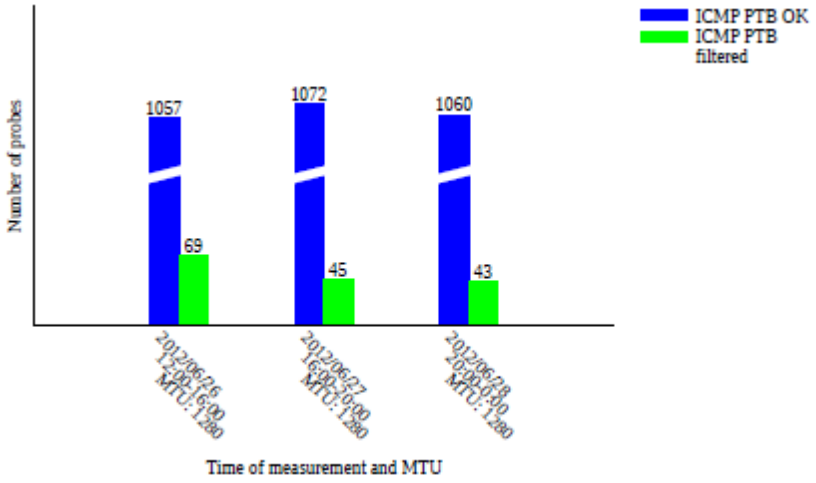


Figure 13: ICMPv4 PTB filtering - MTU: 1280

IPv6

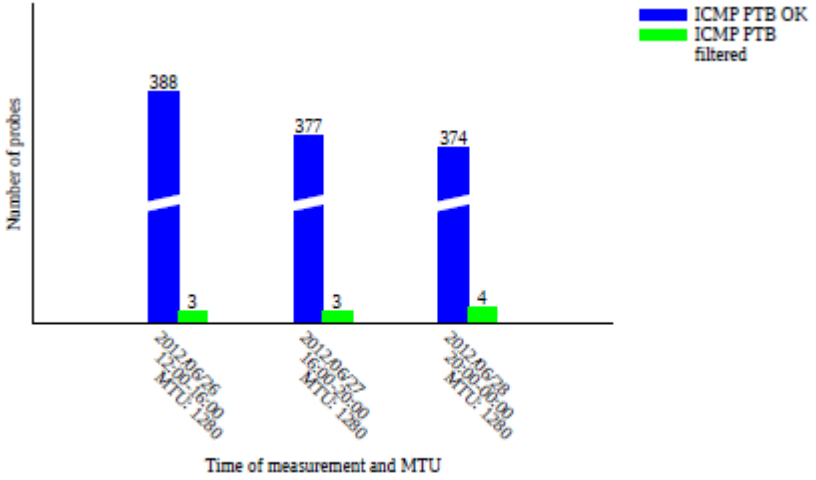


Figure 14: ICMPv6 PTB filtering - MTU: 1280

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

フラグメントのフィルタ(MTU=1500)

IPv4

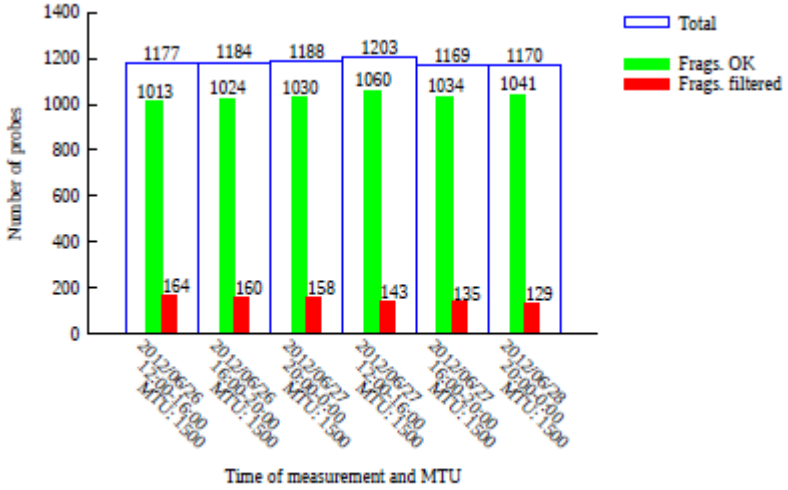


Figure 16: IPv4 fragment filtering - MTU: 1500

IPv6

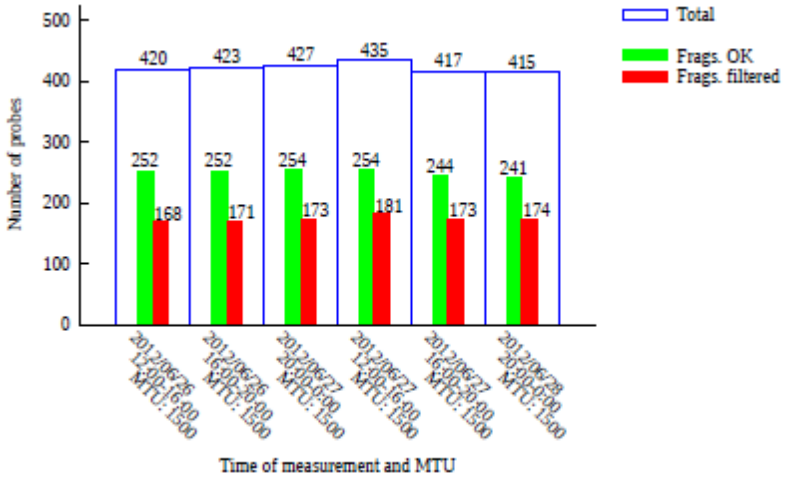


Figure 17: IPv6 fragment filtering - MTU: 1500

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

フラグメントのフィルタ(最小MTU)

IPv4(MTU=576)

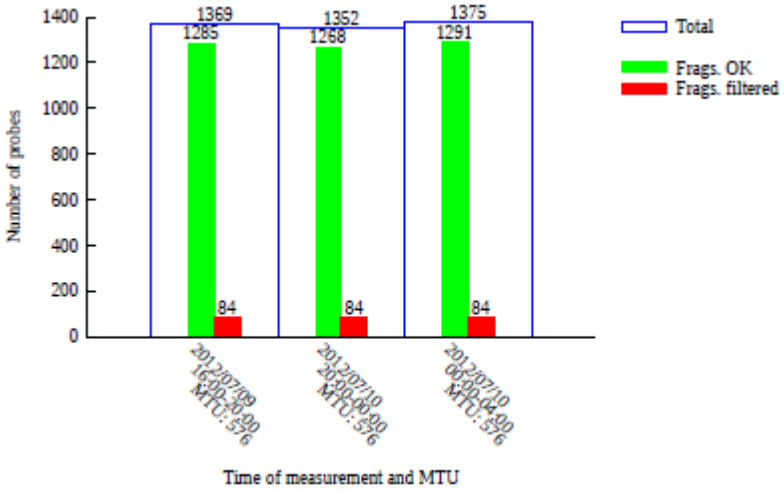


Figure 19: IPv4 fragment filtering - MTU: 576

IPv6(MTU=1280)

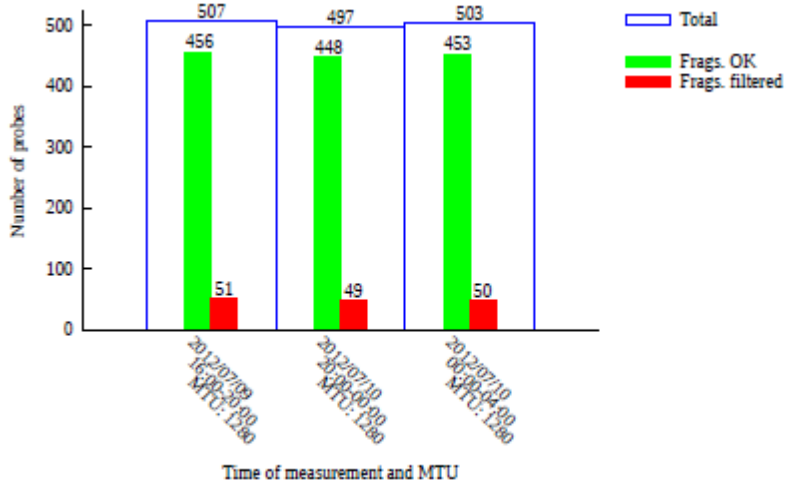


Figure 20: IPv6 fragment filtering - MTU: 1280

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

Path MTU (IPv4)

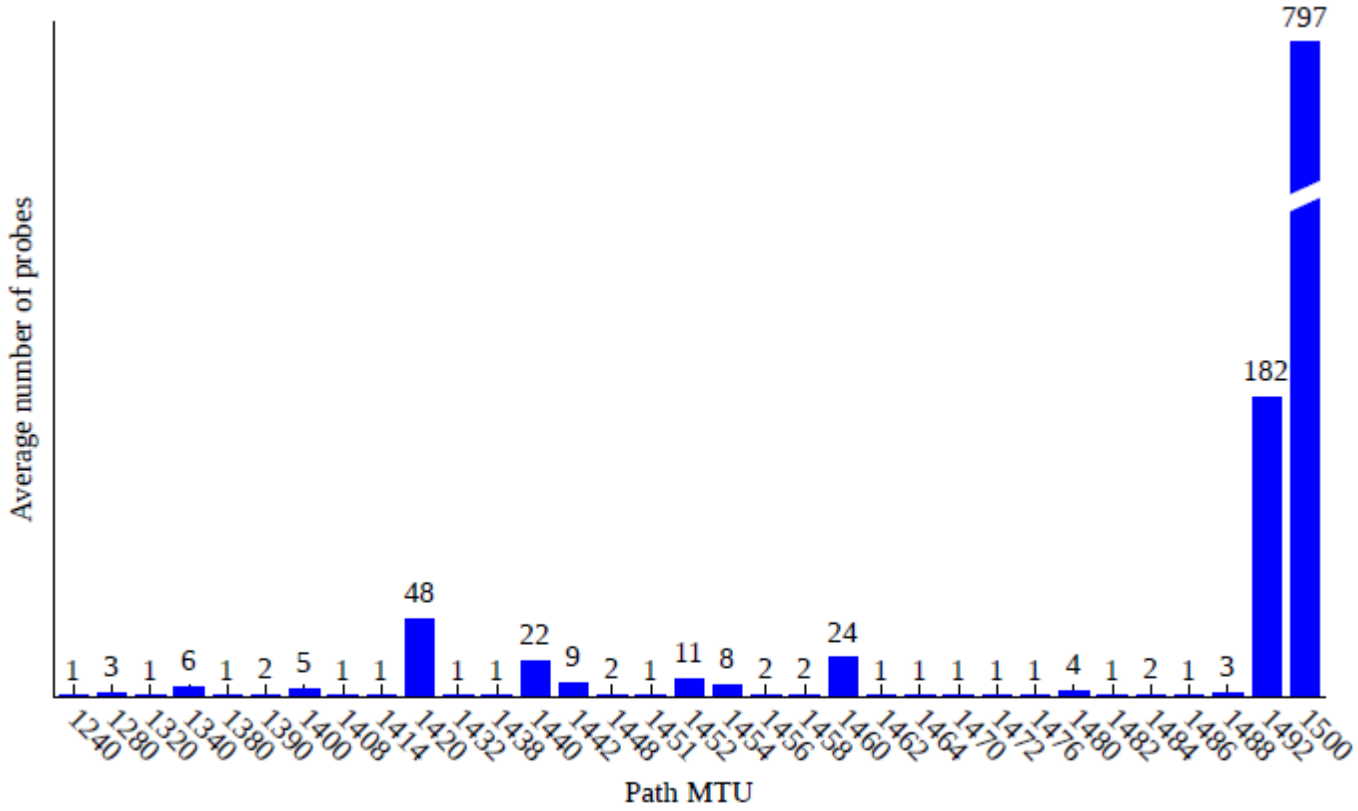


Figure 23: IPv4 common Path MTUs

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

Path MTU (IPv6)

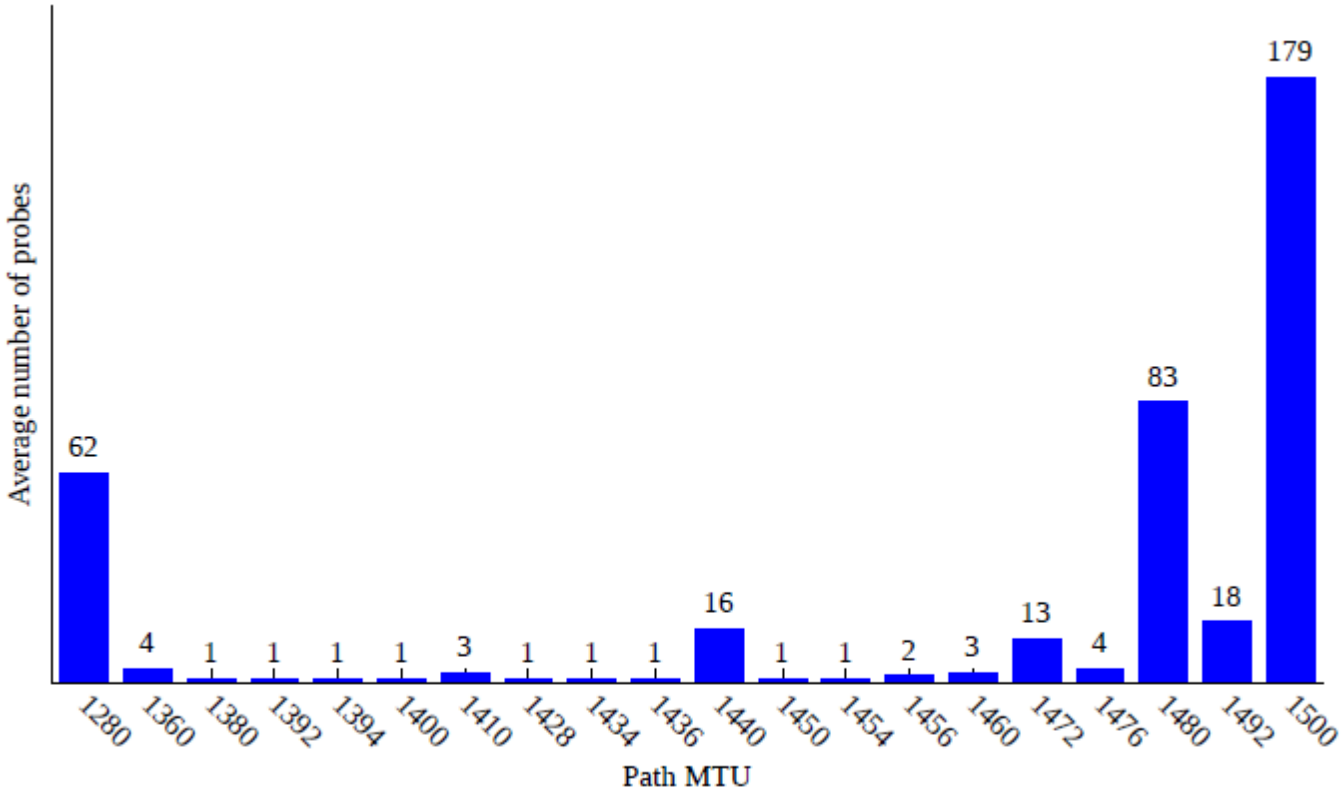


Figure 24: IPv6 common Path MTUs

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

IEPG88: Fragmentation and Extension Header Support in the IPv6 Internet

Failure rate	Duplicate IPv6 addr not removed	Duplicate IPv6 addr were removed
Fragmentation	47.68%	41.57%
Extension Header (8Byte)	52.53%	44.85%
Extension Header (1KByte)	92.17%	89.93%
Oversized Header Chain	71.85%	66.03%

- http://www.iepg.org/2013-11-ietf88/fgont-iepg-ietf88-ipv6-frag-and-eh.pdf

解決に向けて

- 誰が捨てているのか、というか、本当に捨てているの？？
  - ネットワーク、データセンター、キャッシュ、サーバそのもの
  - どのくらいの影響があるのか?
  - TCP/MSSで救われているのはどれくらいか?
- なぜ捨てているのか
  - やむにやまれない理由があるのか
  - 実は、たいした理由は無いとか
- 将来どのような問題が予見されるか(未然に防げれば良い)
  - DNS AAAA(EDNS0): PMTUD動作はIPv6対応に含む
  - DNSSEC：PMTUD動作はセキュアな条件
  - 移行技術(カプセル化、IPv4-IPv6変換)
  - NVO3
- インターネット全体の問題であり、業界を挙げた問題解決が必要なのではないか？
  - PMTUDがきちんと動くことが目指すべき姿だと思います
  - フラグメントの貧弱性を利用した攻撃も可能

ということでみなさん、どうしていきましょうか？