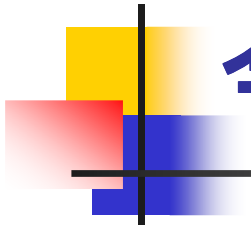
DNS 512bytesの壁
～OCN DNSトラヒック分析より～
+ JPRS Update

吉村 知夏 (NTT Communications)
豊野 剛(NTT)
藤原 和典 (JPRS)



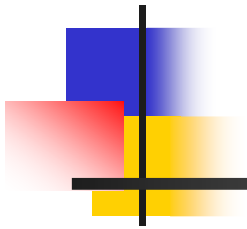
今日の概要

【From OCN】

- DNSパケット長の制限
- 再帰的DNS高負荷事象
- 512bytes超パケットの増加
- 健全な名前解決に向けて

【From JPRS】

- ネームサーバは内部名で

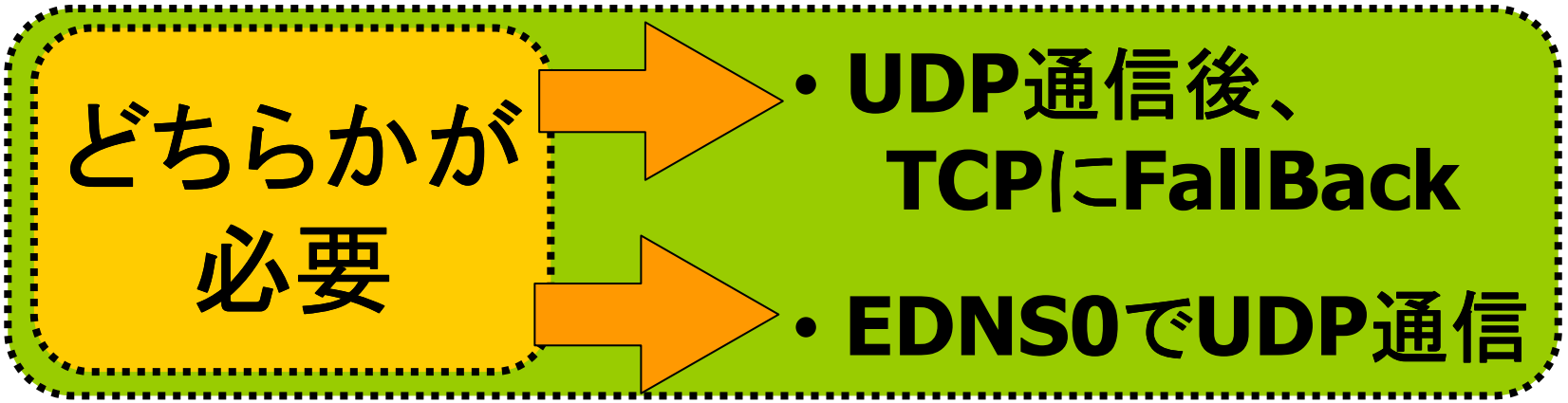


1．パケット長の制限

1-1. パケットサイズの制限

- DNSのパケット長
  - **UDP　512bytes以内**

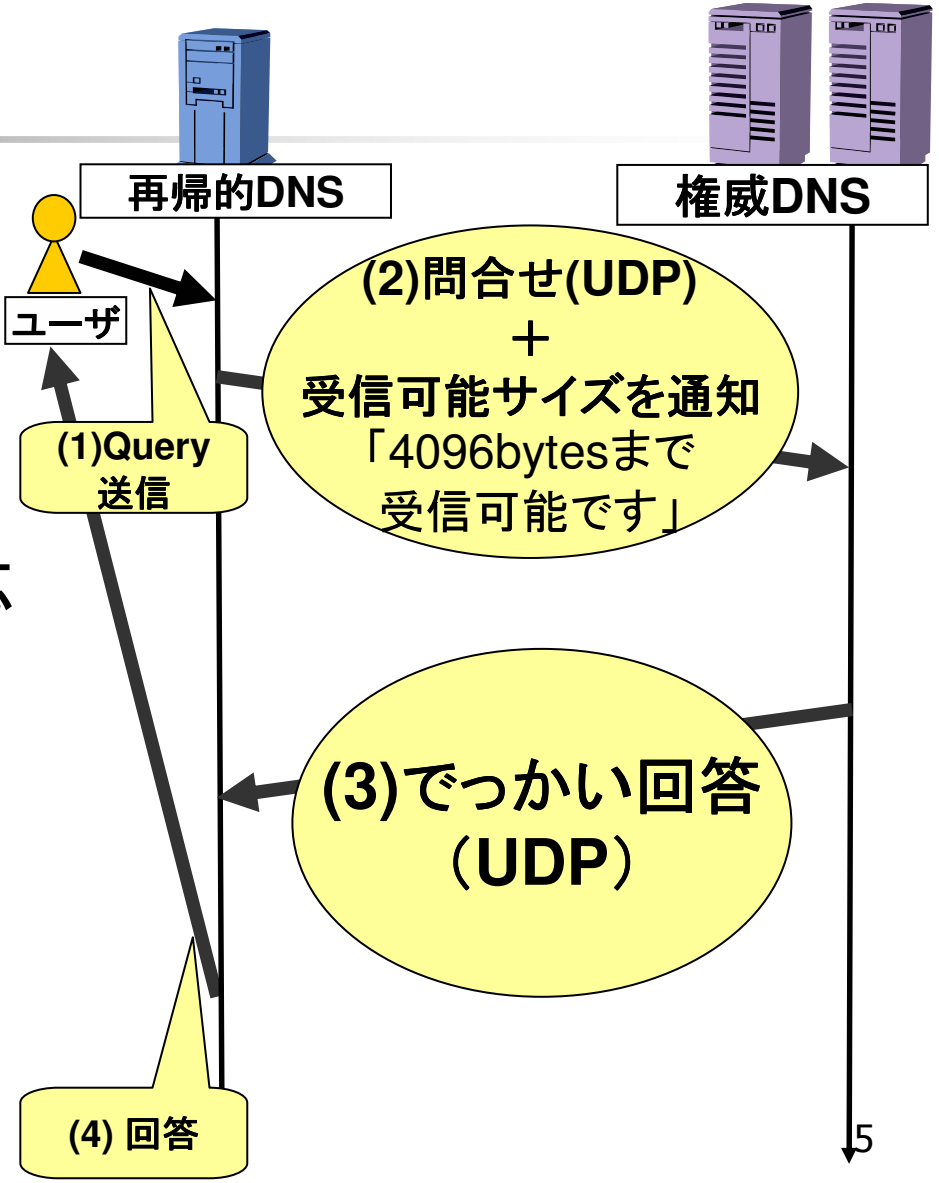
- 512bytes超のパケットを処理するとき

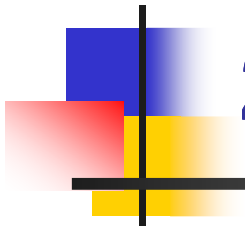


1-2. EDNS0とは

（extension mechanisms for DNS version 0）

- 512bytes以上のパケットをUDPで処理できる仕様
  - RFC 2671 で定義
  - BIND9はデフォルトで対応
  - BIND8.3以降
- 問合せ時に、受信できるパケット長を通知

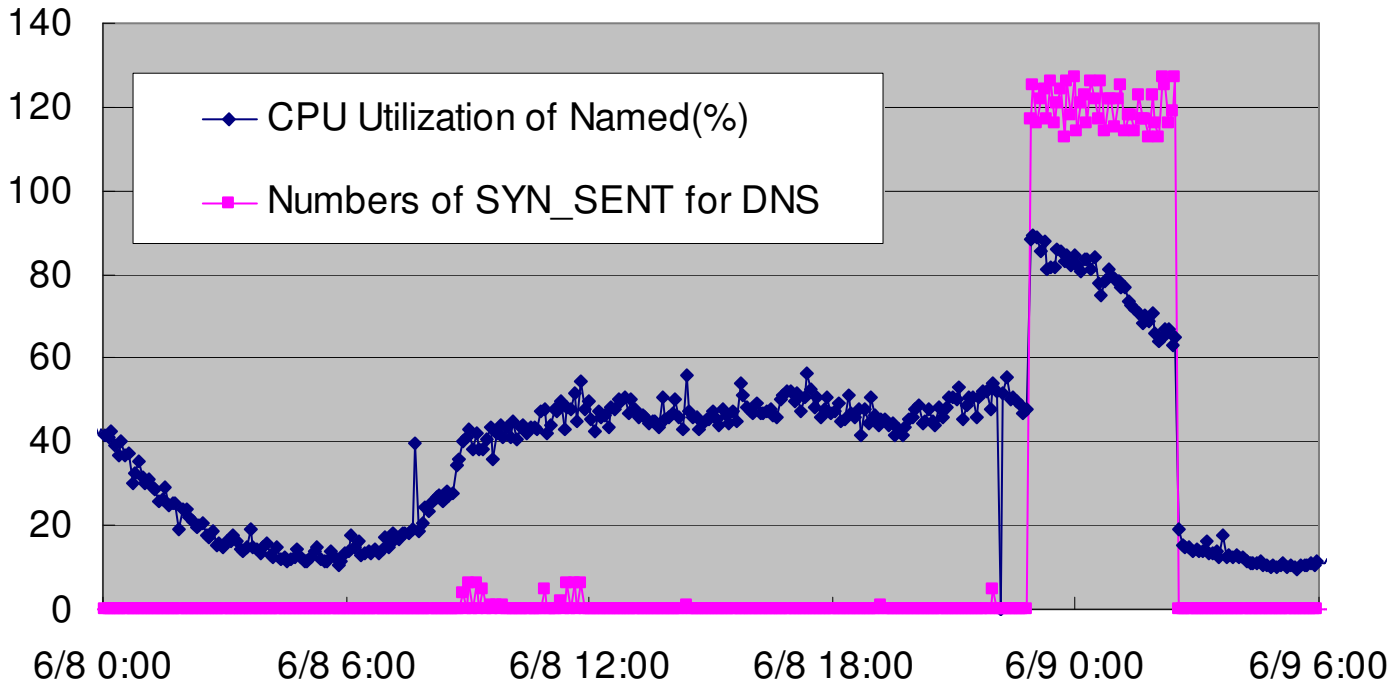




2. 再帰的DNSサーバの高負荷

2-1. OCN の再帰的DNSサーバ

- namedのCPU使用率が急上昇
  - 2004年4月から6月に発生
- 他の権威DNSサーバへTCPセッションを張ろうとしていた



2-2. 何が起こっていたか

(1) ユーザがqueryを送信

(2) 権威DNSへUDPで問合せ

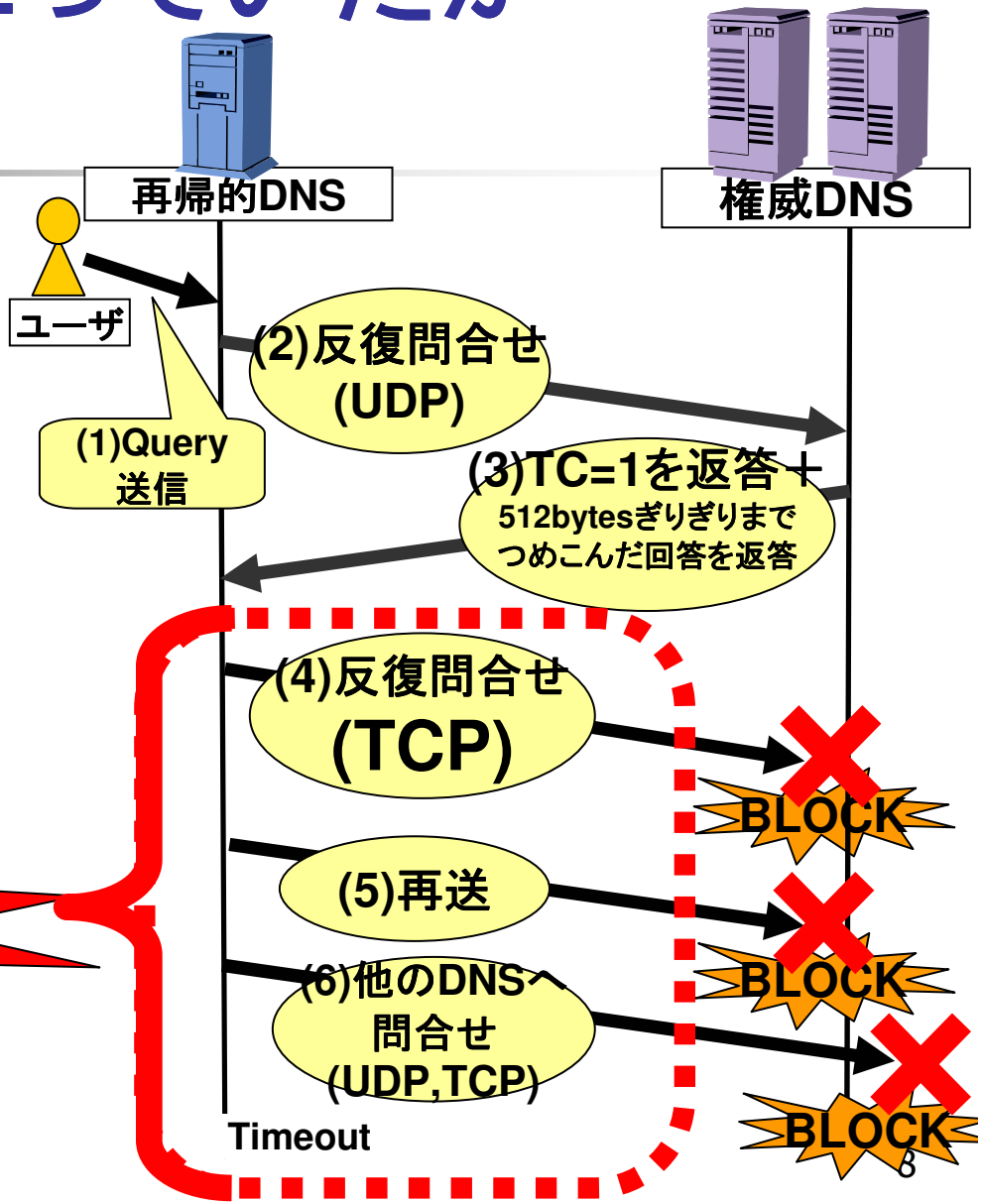
(3) truncated bitが立った返答
【truncated bit】
＝返答パケットサイズが
UDP512bytesに収まりきらない

(4) 権威DNSへTCPで再度問合せ

⇒ 権威DNS側でTCP/53を開けてない

TCPの再送＆
タイムアウト待ちが発生

【サーバ負荷上昇】



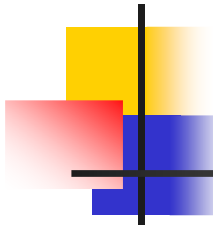
2-3. ***.***.***.*** について

- 名前解決できなかった
- 363個のPTRが設定されている

..***.***.in-addr.arpa. IN PTR **********.com.
..***.***.in-addr.arpa. IN PTR **********.com.
..***.***.in-addr.arpa. IN PTR *******.com.
..***.***.in-addr.arpa. IN PTR *********.co.jp. など

- “dig -x ***.***.***.***” の返答パケットサイズは10000bytesを超える

512bytes超の対策が必要



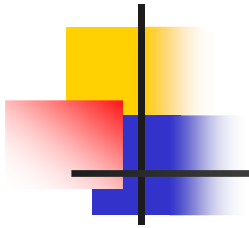
2-4. OCNの対応

- 権威サーバの管理者（某社）とコンタクトを取った
  - TCP/53の開放をしてほしい
  - レコードを512bytes以内へ縮小してほしい
- ユーザからのアクセスを一時的に遮断した
  - BINDのBlackhole設定を使用
  - “***.***.***.*** PTR”以外のqueryもブロックしてしまう
  - ISPとして、決して望ましい対応ではない
- 対応パッチを検討・作成した（参考）
  - TC=1 が返ってきても、TCPにfallbackしない
  - ユーザへは”servfail”として返答する

2-5. 本来あるべき姿

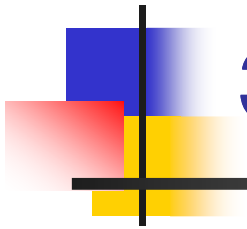
- **512bytes超のレコードを設定するならば…**
- 権威サーバ側でTCP/53を開けるべき
  - TCPの開放はRFCで推奨されている
    - RFC 1123
      DNS servers must be able to service UDP and **should** be able to service TCP queries … it **should not refuse to service a TCP query** just because it would have succeeded with UDP.
- 権威サーバ側でEDNS0に対応するべき

2004年10月頃、TCP開放・EDNS0ともに対応した模様
＝名前解決ができるようになった

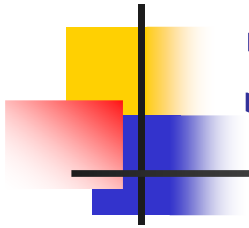


2-6. 沸きあがる疑問

- OCNだけの問題なのか？
- 512bytes超対策をしていないサーバがあればどこでも起きうる事象
- 512bytes超のパケットはどのくらいあるのだろう？



3. 512bytes超パケットの増加



3-1. パケットの肥満要因

- パケットが太る要因が盛りだくさん
  - ユーザの利用形態の変化
    - ラウンドロビン（ワームが使うドメイン、メールサーバ）
    - LDAPなどのリソース探索(SRVクエリ)
    - DNSSEC
    - ENUM　などなど…
  - IPv6対応ノードの増加(AAAAクエリ)
    - Windows Longhorn(もう来年！)は標準でIPv6対応

    ⇒ユーザがAAAAqueryを平気で出す時代がやってくる

3-2. 512bytesを超えるレコード(ワーム)

- **explodark.osirc.net**
  - WORM_SDBOT.BRに感染した人が出すquery
    (1) WORM_SDBOT.BRに感染する
    (2) 感染したClientはIRCボットになる ⇒ explodark.osirc.netに接続する
    (3) 悪意ある人がexplodark.osirc.net越しにClientを操作する
- 32個のAレコードラウンドロビン

```
explodark.osirc.net.     IN CNAME  splotto.demo.httpdnet.com.
splotto.demo.httpdnet.com. IN A  211.55.111.138
splotto.demo.httpdnet.com. IN A  211.74.154.213
splotto.demo.httpdnet.com.  IN A  211.219.153.177
splotto.demo.httpdnet.com.  IN A  211.220.20.226  などなど
```

- “dig explodark.osirc.net”の結果は650bytesくらい
- OCNでは、5分間で5000query来ることもある

3-3. 512bytesを超えるレコード(メールサーバ)

- **mail{1,2,3,4}.saveinternet.net**
  - saveinternet.netのMXレコード
  - 今は名前解決できない(2005.01現在)
- Aレコードラウンドロビンを組んでいる

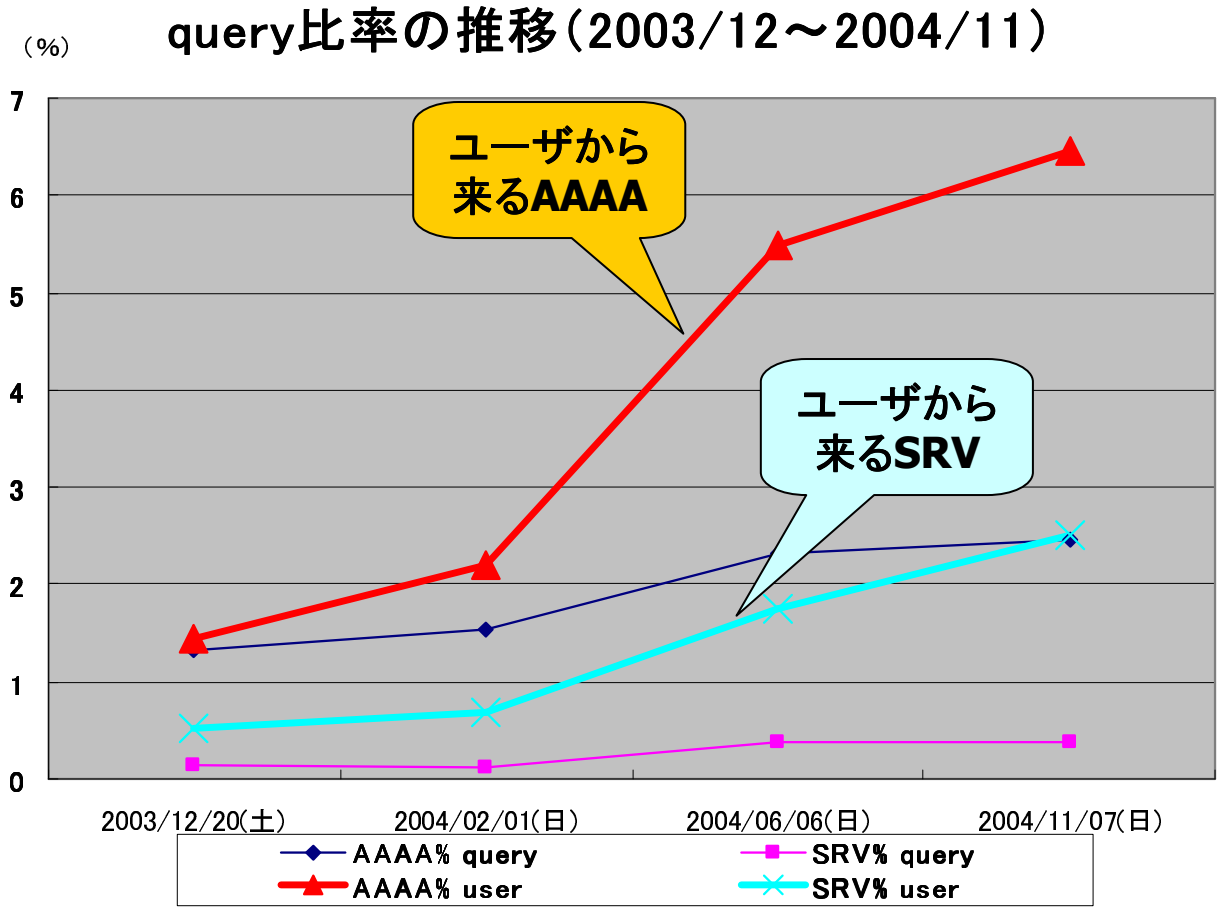
```
saveinternet.net.          IN MX mail1.saveinternet.net.
                           IN MX mail2.saveinternet.net.
                           IN MX mail3.saveinternet.net.
                           IN MX mail4.saveinternet.net.
mail4.saveinternet.net.    IN A  69.42.106.9
mail4.saveinternet.net.    IN A  69.42.107.9
mail4.saveinternet.net.    IN A  69.42.108.9
mail4.saveinternet.net.    IN A  69.42.109.9
mail4.saveinternet.net.    IN A  69.42.110.9
mail4.saveinternet.net.    IN A  69.42.111.9
mail4.saveinternet.net.    IN A  69.42.112.9
mail4.saveinternet.net.    IN A  69.42.113.9   などなど
```

3-4. ActiveDirectory

- **LDAPによるSRVレコードの利用**
  - Windows ActiveDirectoryなど
- ドメインコントローラリストなどに利用
  - _kerberos._tcp.dc._msdcs.****.com
  - _ldap._tcp._[MAC-Addr].domains._msdcs.****.edu
  - Answerセクションが大きい。かつ膨大(27個とか)

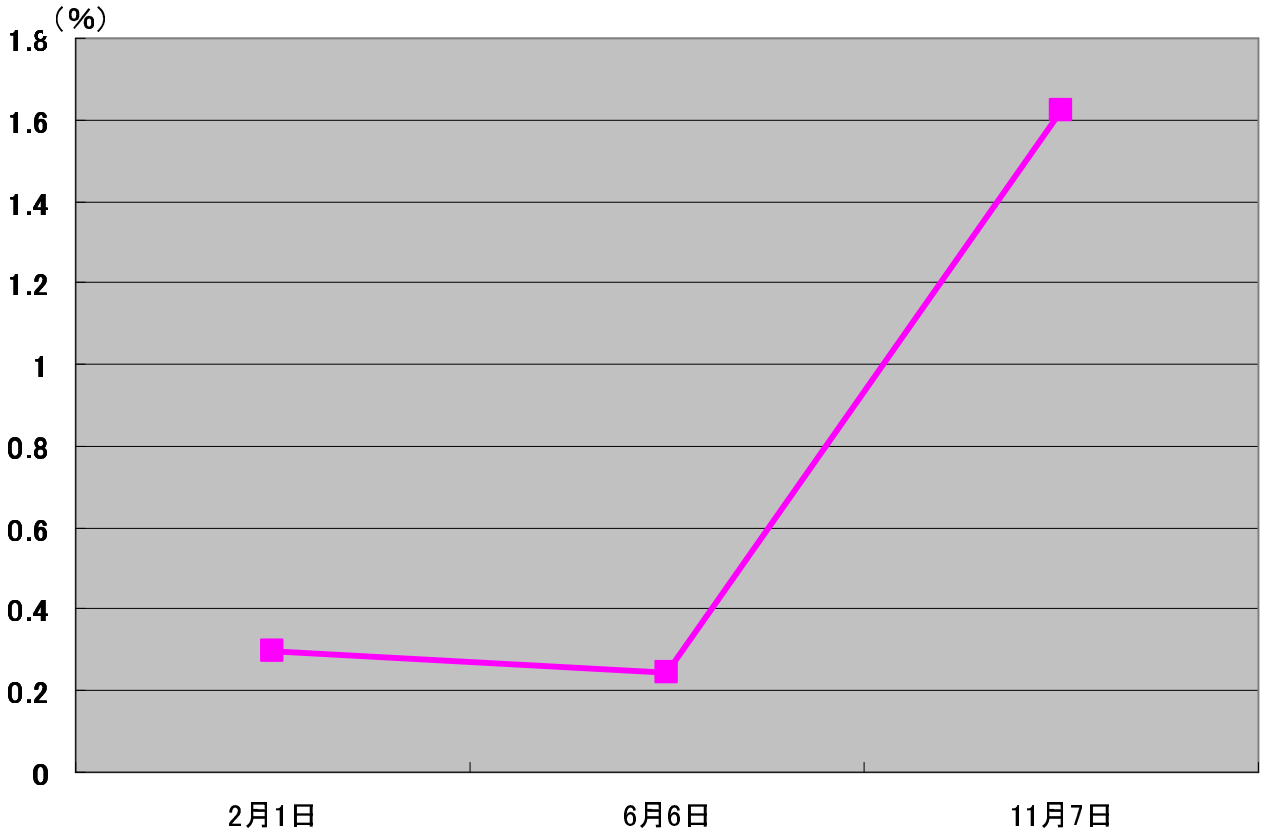
```
_ldap._tcp.dc._msdcs.********.com. SRV ttl 600 ********14.*****.****.com.:389 0 100
_ldap._tcp.dc._msdcs.********.com. SRV ttl 600 ********01.*****.****.com.:389 0 100
_ldap._tcp.dc._msdcs.********.com. SRV ttl 600 ********02.*****.****.com.:389 0 100
_ldap._tcp.dc._msdcs.********.com. SRV ttl 600 ********05.*****.****.com.:389 0 100
_ldap._tcp.dc._msdcs.********.com. SRV ttl 600 ********16.*****.****.com.:389 0 100
_ldap._tcp.dc._msdcs.********.com. SRV ttl 600 ********04.*****.****.com.:389 0 100
_ldap._tcp.dc._msdcs.********.com. SRV ttl 600 ********11.*****.****.com.:389 0 100
_ldap._tcp.dc._msdcs.********.com. SRV ttl 600 ********18.*****.****.com.:389 0 100
```

3-5. AAAA/SRV queryの増加率



AAAA, SRVなどのサイズの大きなDNSクエリは順調に（？）増加傾向

3-6. 512bytes超パケットの増加率
(OCN再帰的～権威サーバ間)



512bytesを超えるDNS応答の割合は増加

0.3%
(2004/02)
↓
1.6%(2004/11)

3-7. クエリタイプ別 パケットサイズ平均値
(OCN再帰的-権威サーバ間) (2004/11/07　1日計)

クエリタイプ	UDP応答平均 (bytes)	TCP応答平均 (bytes)
全DNSクエリ平均	201	744
A	188	579
PTR	238	1200
NS	415	1520
AAAA	254	無し
MX	246	696
SOA	355	無し
SRV	470	1358
ANY	243	887
TXT	255	無し
その他	176	1034

平均しても700bytes超のパケットをやりとり

NSはパケットサイズが大きい

SRVもけっこう大きい

(注) 各QTYPEに対する応答のうち
・Answerセクションがあるもの
・Answerセクションが
CNAMEだけではないものの
平均値

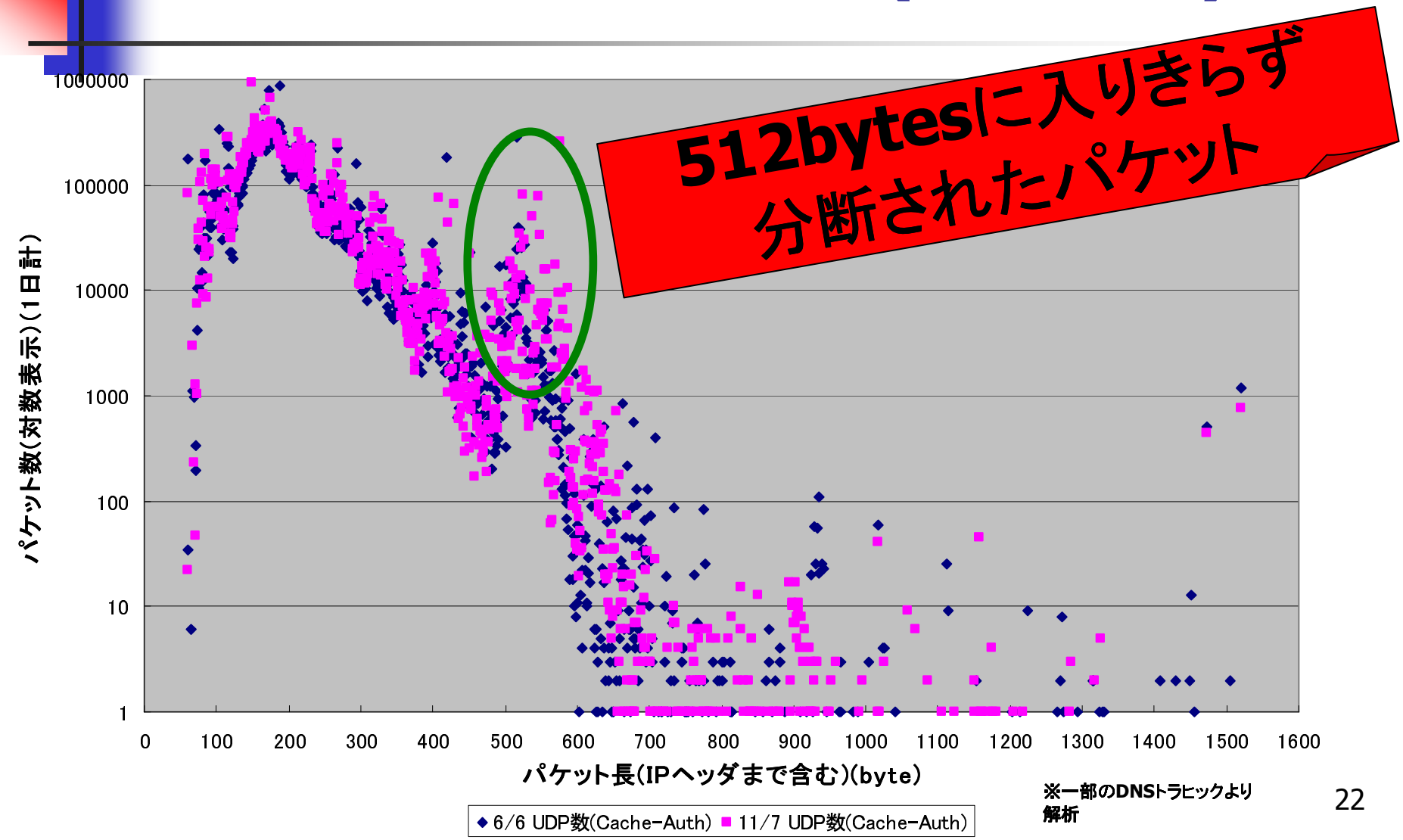
3-8. パケットサイズ動向 まとめ

- 一部のクエリタイプでは512bytes超は当たり前
- 512bytes超パケットは増加傾向にある

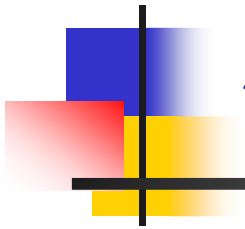
512bytes超をさばく対策が必要
・TCP/53を開放
・EDNS0の対応

- 対策は2通りある
- TCPを使うと、TCPにFallbackするまでに無駄な動きが発生する

OCN再帰的サーバ・他組織権威サーバ間 UDPパケットサイズ分布図(6月/11月)



※一部のDNSトラヒックより解析



健全な名前解決に向けて
～512bytesの壁を乗り越える～

DNSの設定を見直そう (1)

- 512bytes超でも名前解決ができるよう、適切な設定が求められる
  - EDNS0に対応する
  - TCP/53を開放する
- 設定を怠ると、他者へ迷惑をかける可能性がある
  - 権威サーバ側が未対策だと、世の中の再帰的サーバが不幸になる
- 設定を見直しましょう

DNSの設定を見直そう (2)

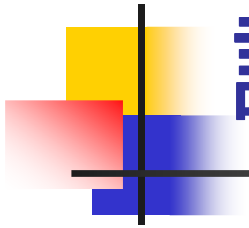
- 具体的な推奨例（EDNS0対応ソフトウェア）
  - BIND9以降
  - BIND8でもEDNS0に対応しているバージョンがあります
    - 8.3.2-RC1 以降
  - CNS、ANS
  - NSD
  - Windows 2003 Server
- リソースレコードの大きさもほどほどに

(おまけ)EDNS0対応状況推測

- 権威サーバの 70%～80%は対応か
  - OCN Cacheサーバから計測
  - OPT RR付クエリ(BIND)に正しく応答したサーバ数

  ⇒　7割の対応数は果たして妥当か？
- EDNS0対応の参考値:
  - 41～55%@K root(NLnet Labs)(2004/03)
  - 約60%@gTLDs(Don Moore)(2004/05)

【参考】http://mydns.bboy.net/survey/
【参考】http://www.nlnetlabs.nl/downloads/edns0.pdf



議論したいこと

- 512bytes超レコードの名前解決の対策は？
  - TCPを開放する？EDNS0に対応する？どちらが望ましい？
- 権威サーバ側がEDNS0に対応していない、TCP/53を開けていない場合どうすればよいか
  - ほかのISPはどう対処されているのですか？
- ユーザ端末（OS）はEDNS0に対応しているか
  - 再帰的サーバにTCPを張られるのは**イヤ！**
- 512bytes超レコードの増加状況について
  - Windows LonghornはIPv6標準対応の模様
  - Rootサーバ、TLDサーバのIPv6化
  - AAAAqueryがボコボコ増えたとき、再帰的サーバは耐えられるのか？

Special Thanks (敬称略･順不同)

- **NTT 情報流通プラットフォーム研究所**

  外山　勝保

  石橋　圭介

  松岡　弘高

- **NTTコミュニケーションズ株式会社（OCN）**

  水越　一郎

  大島　治彦

  石野　雅博