# Practical Cache Server Operation

## ～量と質、様々なQueryに対応するには～

Internet Week 2004
NTTコミュニケーションズ(OCN)
大島 治彦



# Today's Topic

ｲﾝﾀｰﾈｯﾄの普及によりCache Serverは常に忙しい。
「異常/正常」どちらのQueryも増え続ける。
どうする？

- 様々なﾀｲﾌﾟの大量Query
- Authorative Serverとの関係
- Cache Serverのｽｹｰﾗﾋﾞﾘﾃｨ

# 大量Query(1)



# 同一SrcIPからの大量Query

- DoS、不正ｱｸｾｽ、ﾕｰｻﾞの設定ﾐｽ
  (ServFail、NXDomain、PrivateIP逆引き･･･)
- Webｻｰﾊﾞのﾛｸﾞ解析ﾂｰﾙ
  (大量の逆引き)

## 検知と対処

- CPU使用率、Queryレート(qps)、nslookup遅延の閾値
- QueryログやF/WでのSrcIPの特定
- BINDのBlackholeやF/Wでアクセス拒否



## 大量Query(2)

# 大量のSrcIPからの特定のQuery

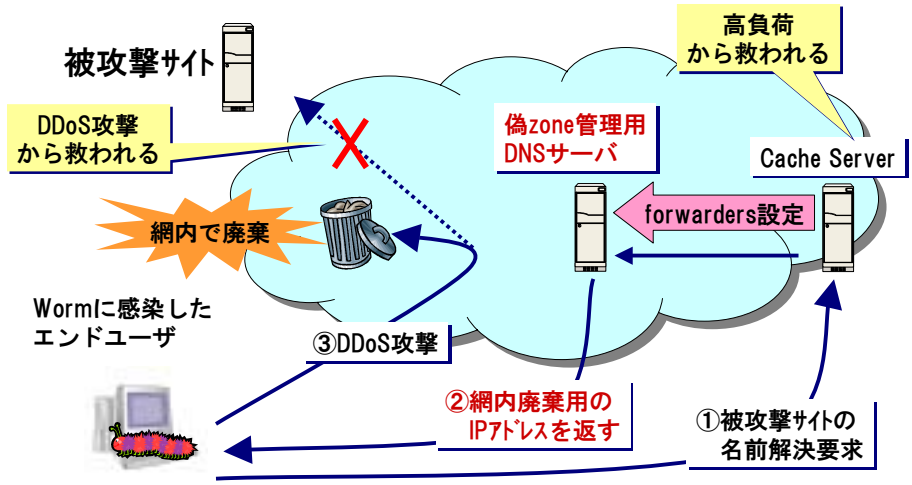
- Wormが特定のｻｲﾄをｱﾀｯｸ
  (MS Blast、Mydoom、Netsky、Antinny…)
- DDoS(FQDNを引くもの)

- ｱﾀｯｸﾄﾗﾌｨｯｸで被攻撃ｻｲﾄとﾈｯﾄﾜｰｸが輻輳
- ｸﾗｲｱﾝﾄがNXDomainをｷｬｯｼｭせず、連続QueryでCache Serverが高負荷



# 対処

- Forwardersと偽ｿﾞｰﾝｻｰﾊﾞでの代理応答

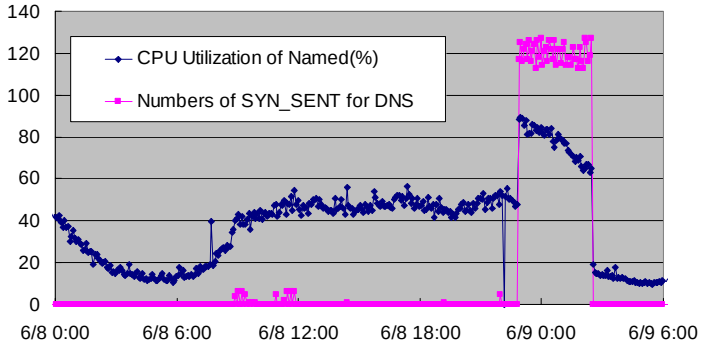


# Authoritative Serverとの関係



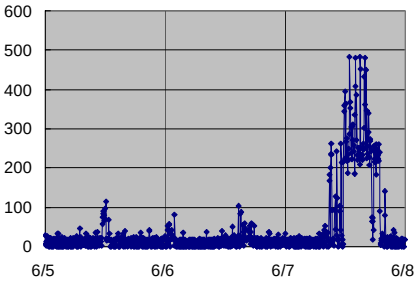
# 原因不明のCache Server高負荷

- Query数は正常、でもCache Serverは高負荷
- 何故かTCPのSYN_SENTの数と連動

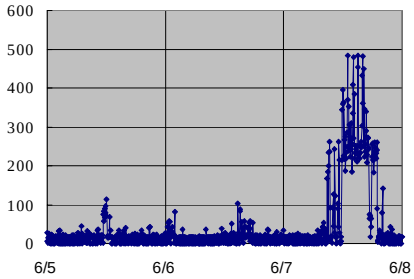


# TCP Connectionが何故増える？

- TC(Truncated)ﾋﾞｯﾄが立ったﾊﾟｹｯﾄ数と連動




Number of TCP packets sent to authoritative servers (per 5 min.)

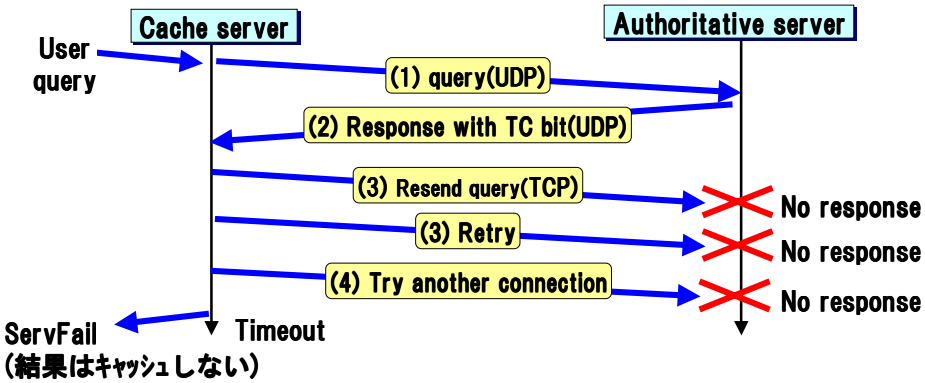



Number of truncated answers from authoritative servers (per 5 min.)



# SYN_SENT発生のメカニズム

- Authorative Serverが大きなRR Setを登録
- しかも、EDNS0に対応せずTCPもﾌｨﾙﾀ




注) 高負荷が発生するのはBIND8。BIND9では発生しない。

# 対処は？

- Authoritative Serverの管理者への依頼
  - RR Setのｻｲｽﾞを小さく
  - TCPでのｱｸｾｽを許可
  - EDNS0への対応
- BINDのBlackhole設定で該当ﾕｰｻﾞからのQueryを拒否
- Forwarders設定と偽ｿﾞｰﾝｻｰﾊﾞでの代理応答



# EDNS0とTCPﾌｨﾙﾀの関係

| Cache Server | Authoritative Server | |
|---|---|---|
| EDNS0 対応 | EDNS0 対応 | TCP 対応 (RFCでは必須) |
| ○ | ○ | × |
| ○ | × | × |
| × | ○ | × |
| × | × | × |

- TCPを開ければ問題なし。
- TCPが開けられないとEDNS0が必須。

TCPをﾌｨﾙﾀすると512ｵｸﾃｯﾄ超でSYN_SENTが発生ないのはここだけ

## 512ｵｸﾃｯﾄ以下が基本だが．．．

- AAAAでTCPへのfallbackが増加へ
  現在のAAAAのQuery比率⇒約2%
- EDNS0も直ぐには対応できない
- ｾｷｭﾘﾃｨ面、性能面でもTCPは開けたくない
- そもそもAuthoritative Server依存の対策は辛い



## Cache Serverの自律的な回避

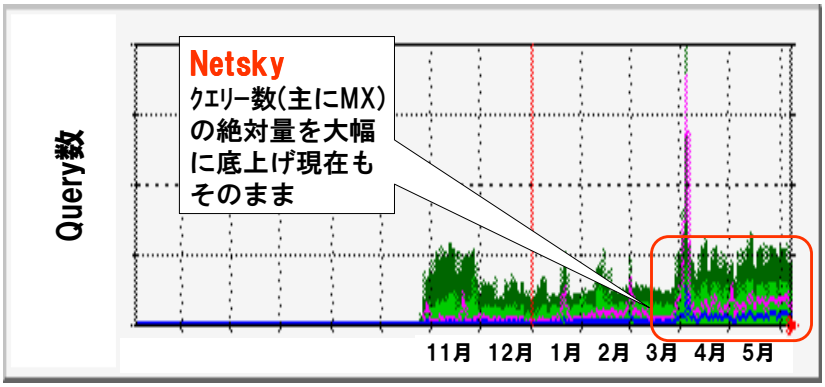
- Cache Serverが自分で身を守る(実装変更)
  - TCﾋﾞｯﾄが立ったﾊﾟｹｯﾄが帰ってきてもTCPにfallbackしない(RFC違反)
  - 一度TCPにfallbackしても応答の無いRRはTTLの間Cacheし、再帰問合せをせずにServFailを返す(Internet Draft)

# Cache Serverのｽｹｰﾗﾋﾞﾘﾃｨ

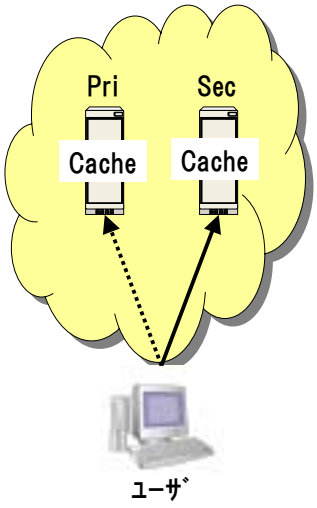


# 規制できないQuery数の増加

- ﾌﾞﾛｰﾄﾞﾊﾞﾝﾄﾞ化と利用動向の変化
- 慢性的に存在するｳｨﾙｽ、ﾜｰﾑの影響



# Cache Serverの負荷分散



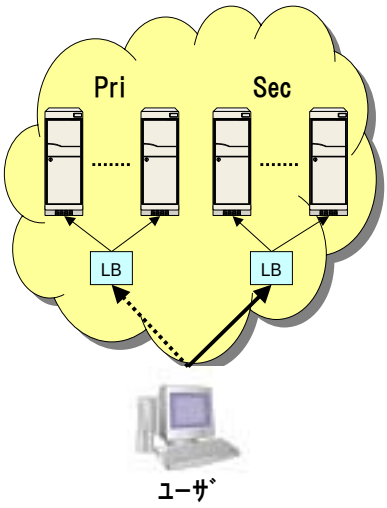

- 地域毎、ｻｰﾋﾞｽ毎に利用するCacheを分散し、PrimaryとSecondaryで運用
- 分散数を増やし、Cache1台当たりのqpsは低めで運用
- qpsが上がり負荷が厳しくなれば、再分散 or 設備更改



# Pri/Secでの構成の限界

- Cache1台あたりは数千qps程度。
  ⇒ｽｹｰﾙしない。
- ﾕｰｻﾞが設定するのはPriとSecの2つ
  ⇒PrimaryとSecondaryの切り替えが遅い
- 高信頼性&大容量のCacheｻｰﾊﾞが有効

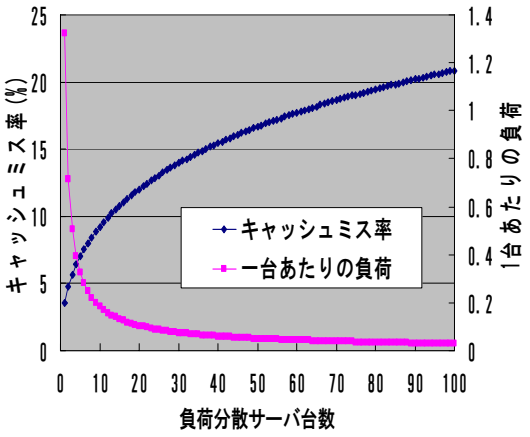
# 負荷分散装置＆複数台Cache




- 負荷分散装置をフロント、裏側に複数のCache配置でQueryを分散
- ユーザは負荷分散装置のVIPを利用
- Cache故障時は負荷分散装置が自動切り離し



# Cacheの台数の最適値？




- 闇雲に台数を増やせばキャッシュが有効に機能しない
- 実際のトラフィックを分析すると10台程度までが最適台数

(注)OCNの実トラフィックの場合

# まとめ

- Authoritative Serverが注目されがちだが、Cache Serverも非常に大切
- Cache Serverは自分で自分の身を守る
- Queryの中身にも敏感に