IIJ
Internet Initiative Japan

Network Initiative

Internet Initiative Japan Inc.

Practical Cache Server Operation

Internet Week 2004
DNS Day

株式会社インターネットイニシアティブ
山本　功司
koji@iij.ad.jp

概要

- IIJ での運用経験を中心に
  - 時間が短いので網羅的な話はしない
- 詳細はJANOG14での発表を参照
  - http://www.janog.gr.jp/meeting/janog14/src/200407janog14-dns-demystified.pdf

Cache Server の運用で重要なこと

Internet Initiative Japan Inc.

- 安定性
  - エンドユーザーが「インターネットを使える」かどうかに直結
  - authoritative server 以上に安定性が求められる
- パフォーマンス
  - スループット
  - レイテンシ
- 正確性/整合性
  - 「正しい」応答を返すこと
  - 特定のドメインが引けなくなったりしない

安定性

Internet Initiative Japan Inc.

- OS レベル、アプリケーションレベルの安定性
- 冗長性
  - ロードバランス
  - anycast
  - resolver レベルの冗長性はあてにしない
    - 遅延大
    - resolver の実装に依存
- 攻撃に対する耐性

攻撃に対する耐性

- 攻撃に対する耐性
  - Cache server に対する DoS は想像以上に容易
  - ブロードバンド接続された単一の PC から、10000qps 以上のクエリを生成可能
  - 悪意がなくても結果として DoS になるケースも
  - mass mailing virus/worm
  - ブロードバンドルーターの異常動作
- 通常、named では数千qps程度が限界
  - 異常クエリ(SERVFAIL)では限界値はもっと低い

パフォーマンス

- 通常スループットが重要
- 自分の運用している cache server の平常値、限界値を知る
  - 監視、モニタリング
  - 限界スループット値計測

限界スループット値計測

- queryperf
  - BIND9 の contrib
- nominum white paper
  - http://www.nominum.com/content/documents/CNS_WP.pdf

モニタリング

- query log
  - 行数を数える
  - 事後計測
- ndc stat / rndc stat
  - cron 等で定期的に出力させ、スクリプトで処理
  - near realtime な計測も可能
- UDP packet 数
  - ほぼリアルタイムに計測できる
  - ただし近似値

UDP packet 数で近似

- カーネルの UDP packet 数の統計を見る
  - 定常状態では、UDP パケット受信数とqpsはほぼ比例
  - netstat -s -p udp
  - SNMP経由で取得
- drop した UDP packet 数が計測できる
  - 異常に増えている時は、プロセスがパケットを処理できていない可能性がある
  - カーネルのチューニングの指針
    - net.inet.udp.recvspace
    - kern.ipc.maxsockbuf
    - 上記はFreeBSDの場合

BINDのknown issue

- BIND8 で lame な設定のドメインが引けなくなることがある
  - だんまりになる
  - BIND8 のバグ
    - BIND8.3.5 - 8.3.7 / 8.4.0 - 8.4.3 で問題
- BIND9 で限界スループットよりはるかに少ないクエリ数でクエリをドロップすることがある
  - 数百qps程度で発生
  - cache cleaning 時にクエリを落とす
    - cleaning interval
  - 本質的な対策はマルチスレッド対応OSを使う
    - とりあえずは余裕を持って使う
    - パッチをあてる

BINDのknown issue(2)

- BIND9でサクサク感がない
  - レスポンスが秒単位で遅延する
  - IPv6 enable でコンパイルされているが、IPv6 connectivity がないと発生
  - 明示的に --disable-ipv6 してコンパイル
  - v6fix 方面でも対応している
    - http://v6fix.net/