

インターネット上の高精度な時刻配信サーバの運用

2005.1.20

伊藤大輔, ito@mfeed.ad.jp

インターネットマルチフィード株式会社

Contents

- A. 本発表について
 - 概要
 - 自己紹介
 - ネットワークによる時刻供給の背景
- B. NTPの特色
- C. 共同研究とその成果について
- D. 安定した時刻情報配信について

概要

■ 目的

- インターネット上の高精度な時刻サーバの運用について、共同研究の成果を報告するとともに、高精度かつ安定した時刻情報配信についてその運用・管理手法を積極的に議論していきたい。

■ 内容

- 概要
- NTPの特色
- 共同研究とその成果について
- 安定した時刻情報配信について

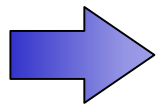
自己紹介

- 所属: インターネットマルチフィード株式会社
 - 1999年よりインターネット上の時刻配信に関する共同研究に携わるとともに、2001年に実験公開したNTPサーバの実際的な運用・管理を担当

- インターネットマルチフィード社について
 - 1997年設立
 - 東京・大手町にてデータセンタを運営
 - 2001年に「JPNAPサービス」を開始
 - 商用インターネット相互接続(IX)サービス
 - 現在MAX 45Gbps相当のトラフィックを交換

ネットワークによる時刻供給の背景

- インフラ基盤としてのインターネットの確立
 - インターネット上での電子商取引の増加
 - 証券取引等における時刻情報の重要性
- 時刻同期の重要度の向上
 - ネットワーク機器の時刻合わせ(logの精度向上)
 - 録画予約などによる時間指定のアプリケーションの存在
 - 家電のネットワーク化



時刻同期に対するニーズの高まり

ネットワークによる時刻供給の背景

- イン트라ネット等、ネットワーク利用の普及
 - ネットワークに接続された計算機の増加
 - 簡易な時刻情報取得の方法としてネットワークを利用

本発表では、ネットワーク上の時刻同期プロトコルとして、インターネット上で非常に多く使われているNTP (Network Time Protocol)を取り上げる

Contents

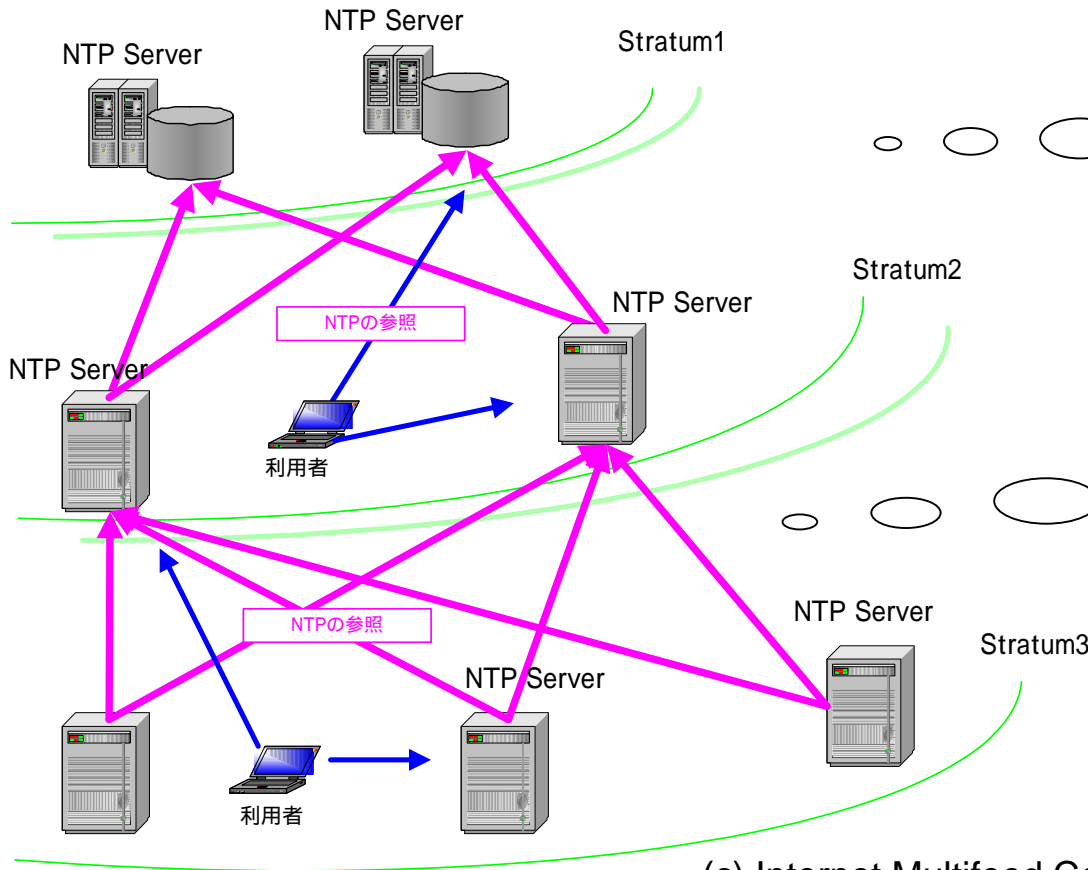
- A. 本発表について
- B. NTPの特色
 - NTPの概要
 - NTPの特色
 - SNTPの概要
- C. 共同研究とその成果について
- D. 安定した時刻情報配信について

NTPの概要

- RFC1305
 - Network Time Protocol (Version 3) Specification, Implementation
 - David L. Mills
- インターネットなどのネットワークに接続された計算機内蔵時計の時刻を合わせるための技術
- 多くはサーバ上にて動作
 - ntpd (<http://www.ntp.org/>)
 - 事実上の標準。多くのディストリビューション、機器に実装
 - Version4を開発中
 - openntpd (<http://www.openntpd.org/>)
 - openbsdプロジェクトが開発

NTPの特色 階層化

- 計算機ネットワークに接続された計算機を階層状に接続
- ネットワークで接続される計算機間の時刻同期のための通信を一つのサーバに集中させない



階層構造

時刻同期のための通信を一つのサーバに集中させない

NTPの特色 Stratum

- NTPにおける計算機の階層名
- Stratum1が最上位の階層(第一階層)
 - GPS,原子時計などの時刻源(Stratum0)に直結のサーバを示す
- Stratum16が最下位
 - Stratum16に同期することは出来ない
- ntptraceで階層状態の確認が可能

NTPの特色 複数サーバからの選択

■ サーバの選択

■ Stratum優先

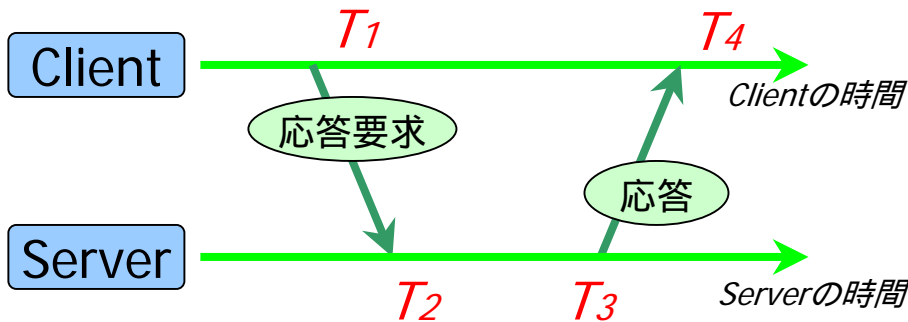
- どんなに不安定で精度が悪いStratum1でも、Stratum2より優先して選択される

■ 同じStratumの場合

- 参照毎に精度(offset)と分散(dispersion)を計算し、もっとも安定していると思われるサーバをReferenceとする

NTPの特色 時刻合わせ

- 双方向通信で行われ、回線による遅延時間をキャンセルして時刻同期の精度を高める様に考慮



$T_2 - T_1 = \text{伝送遅延(delay)} + \text{時刻のずれ(offset)}$
 $T_4 - T_3 = \text{伝送遅延(delay)} - \text{時刻のずれ(offset)}$

$$\text{offset} = \frac{((T_2 - T_1) - (T_4 - T_3))}{2}$$

- 得られたoffsetをClientは時刻のずれとして計算
- 伝送遅延が行きと帰りで同じであることが前提

SNTPの概要

- RFC2030
 - Simple Network Time Protocol(SNTP) Version 4 for IPv4, IPv6 and OSI
 - D.Mills

- RFC1361
 - Simple Network Time Protocol (SNTP)
 - D.Mills

- NTPの簡易版。階層の末端で利用することを推奨

- 多くはクライアント上にて動作
 - 多数のアプリケーションが存在。OSとして標準装備も。

Contents

- A. 本発表について
- B. NTPの特色
- **C. 共同研究とその成果について**
 - インターネットへの日本標準時供給サービスに関する共同研究
 - 研究成果
 - 時刻供給ネットワーク
 - ネットワーク構築時の注意点: ARPの影響
 - ネットワーク構築時の注意点: 参照台数
 - 日本標準時配信実験としてサーバを公開
 - Stratum1とStratum2の精度について
- D. 安定した時刻情報配信について

インターネット上への 日本標準時供給サービスに関する共同研究

■ 目的

- インターネット上に安定した高精度な時刻情報配信・配送網構築を行うこと

■ 参加団体

- 独立行政法人情報通信研究機構(NICT)
 - 旧 通信総合研究所(CRL)
- 日本電信電話株式会社(NTT)
- 株式会社インターネットイニシアティブ(IIJ)
- インターネットマルチフィード株式会社(MFEED)

■ 期間

- 平成11年6月23日 ~ 平成14年3月31日

■ 利用プロトコル

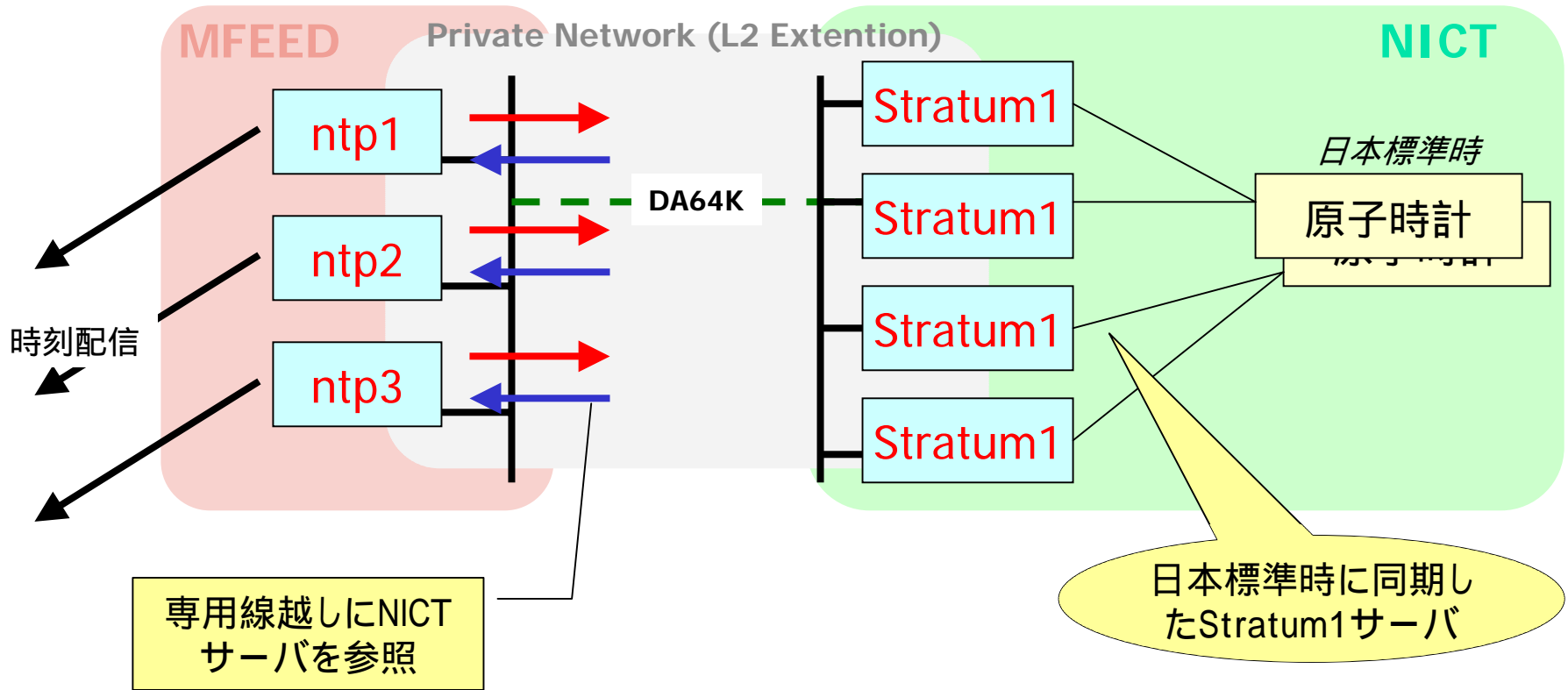
- NTP

研究成果

- NICTの日本標準時をMFEEDから配信可能に
 - NICTの原子時計の時刻を大手町のMFEEDより配信
- 公開NTPサーバを利用してアクセス状況を観測
 - 多アクセス
 - 高精度

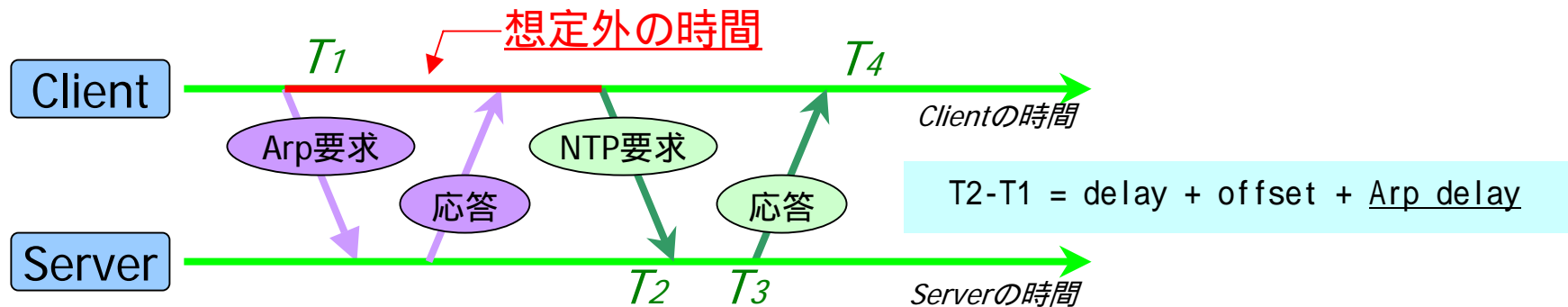
時刻供給ネットワーク

- NICTのサーバを複数台参照したStratum2で時刻を供給



ネットワーク構築時の注意点: ARPの影響

- ARPによる遅延をNTPは計算していない

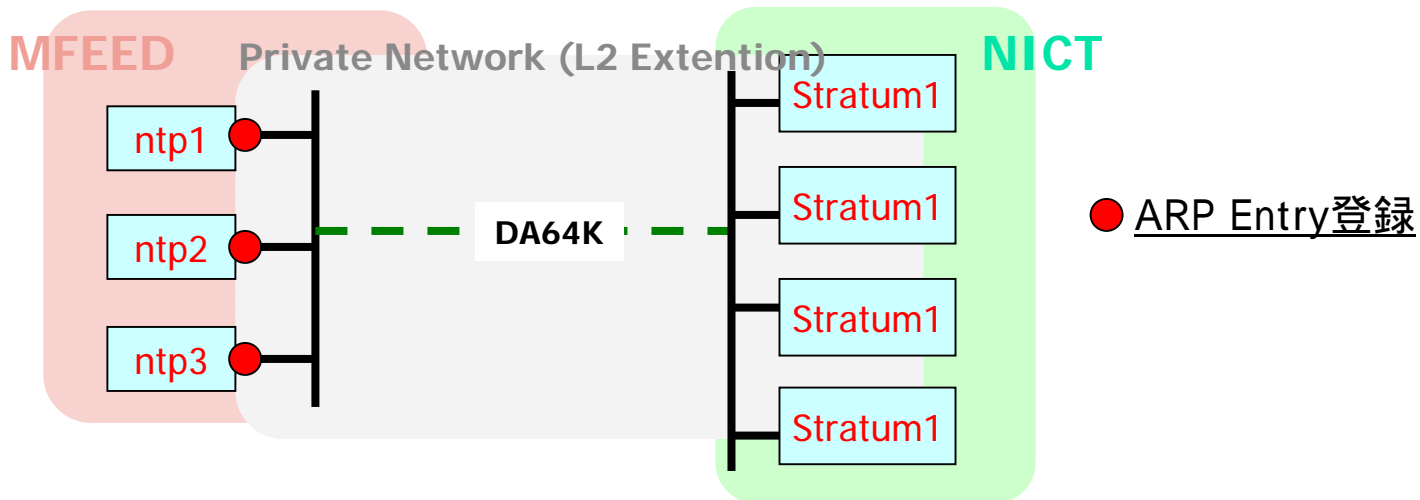


$$\frac{((T_2 - T_1) - (T_4 - T_3))}{2} = \text{offset} + \text{Arp_delay}$$

- Arp の遅延により、得られるoffsetが大きく変化
 - Arp CacheのClearのタイミングで精度が変化
- 実験の精度計測により発覚

ネットワーク構築時の注意点: ARPの影響 (continued)

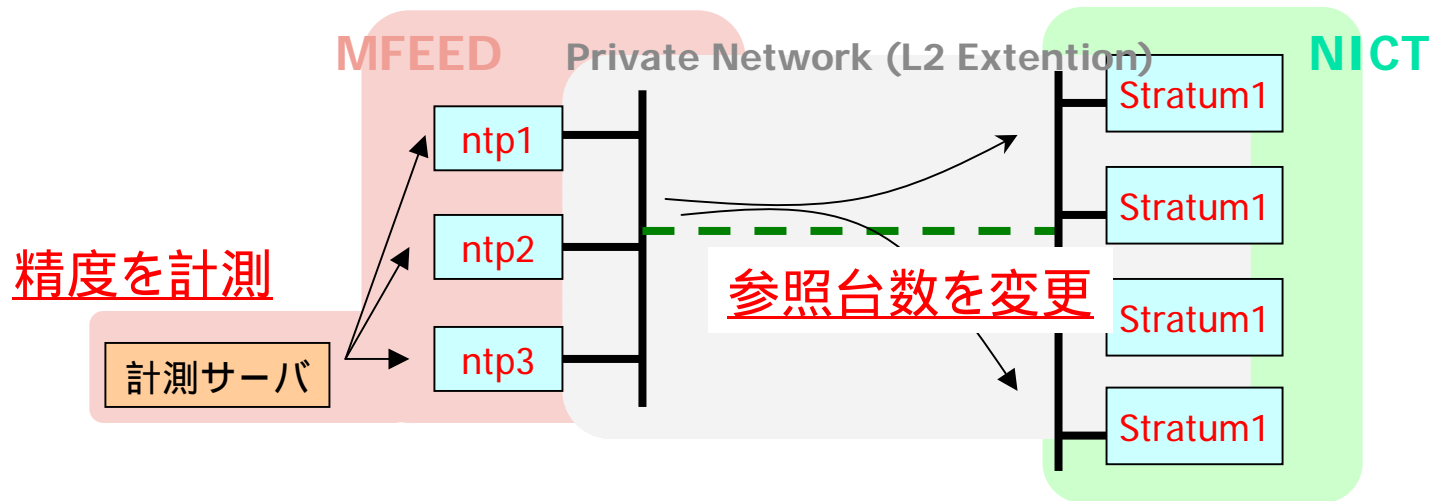
■ StaticなArp Entryを設定して解決



- Gatewayへのdelayが大きいところでは同様の影響が想定される

ネットワーク構築時の注意点: 参照台数

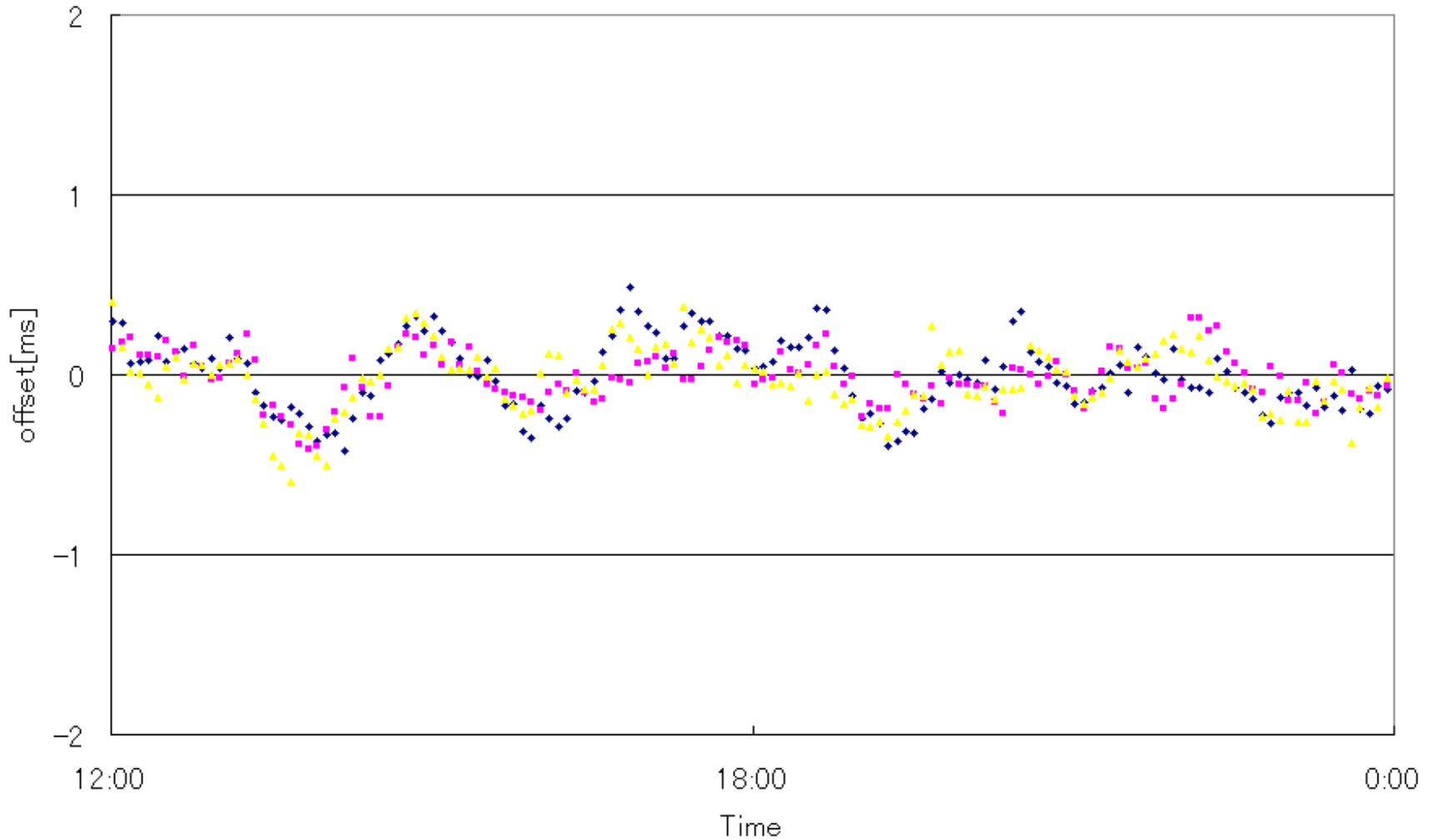
- 何台のStratum1を参照すればStratum2サーバの精度は十分に保てるか？



- NICT側Stratum1サーバの参照台数を変更しつつ、Stratum2サーバの精度を計測サーバから観測
 - 計測サーバは日本標準時と1msec以内に同期
 - 計測サーバは被計測サーバに十分に近い

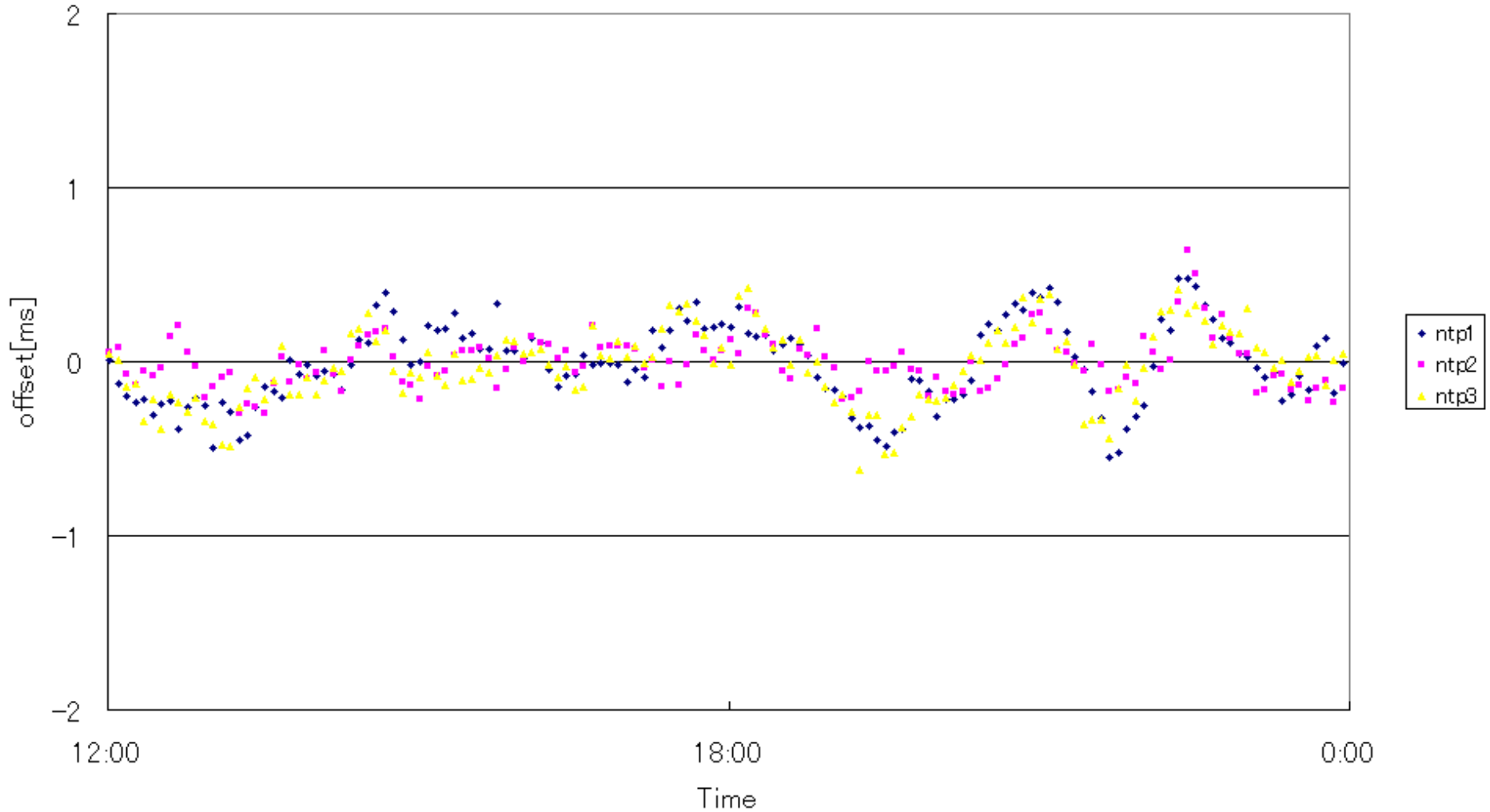
4台参照時の精度 (graph)

4台参照時の精度



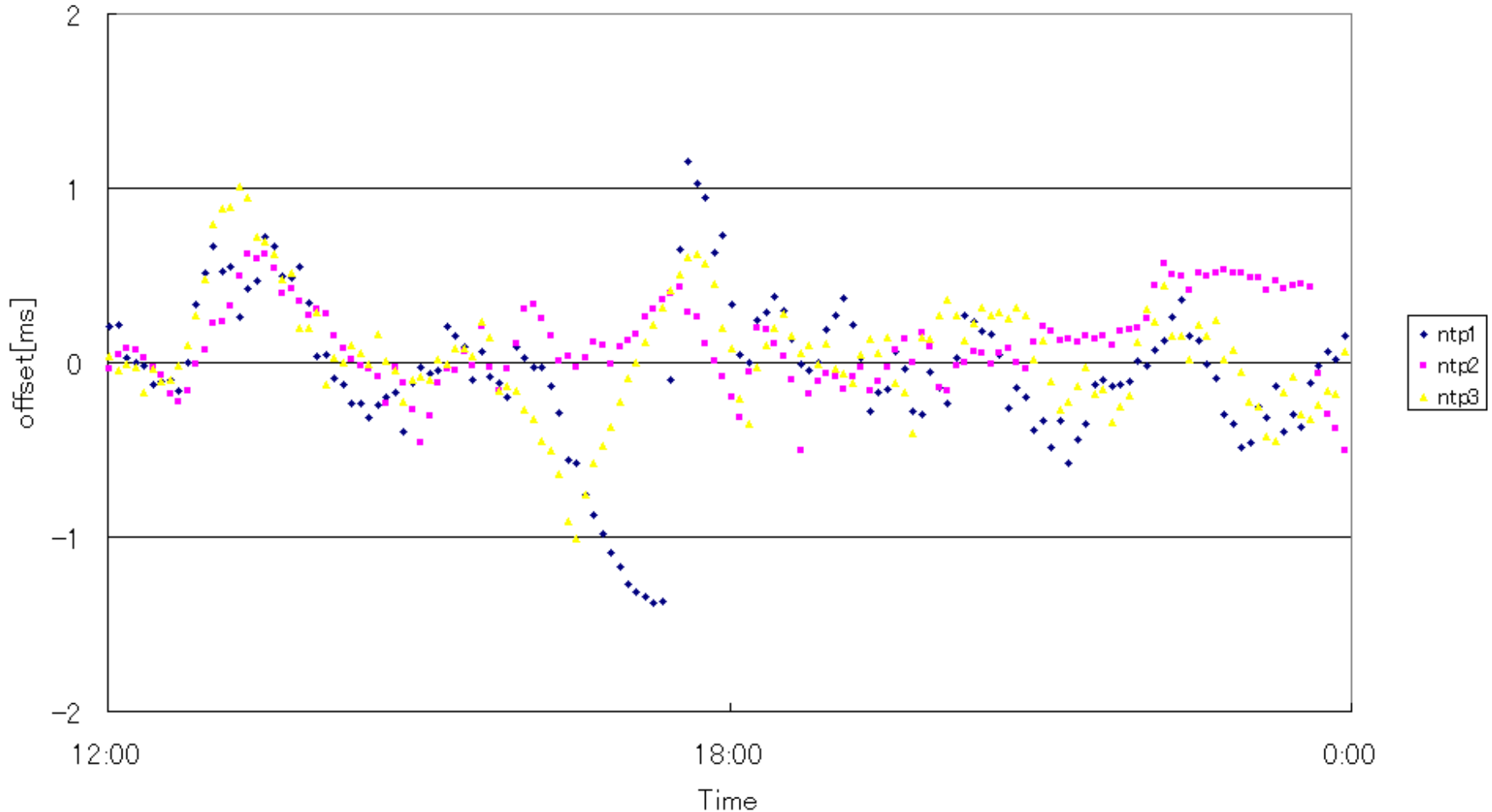
3台参照時の精度 (graph)

3台参照時の精度



2台参照時の精度 (graph)

2台参照時の精度



ネットワーク構築時の注意点: 参照台数 (Continued)

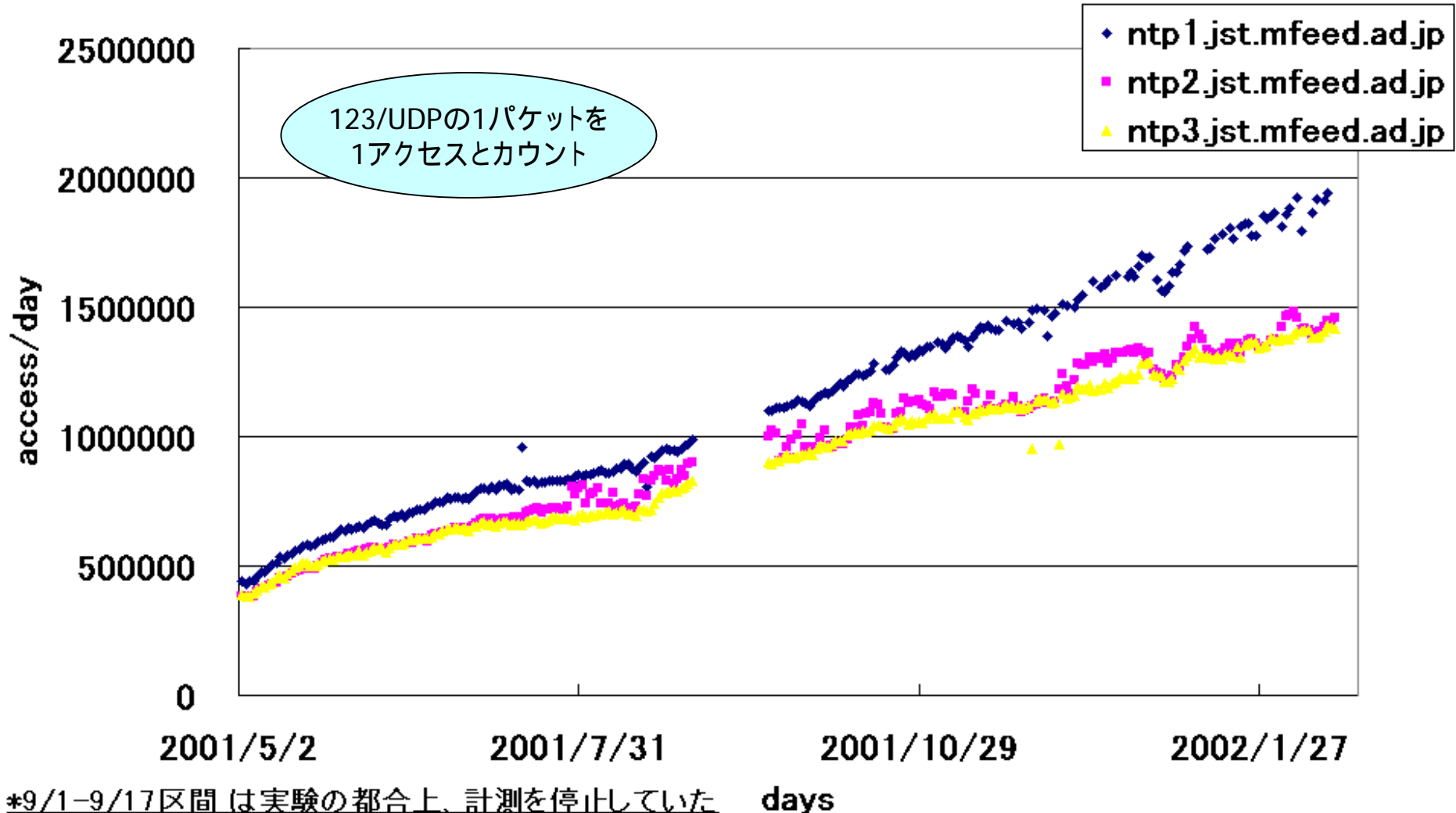
- 3台以上で1msec以内の精度が維持可能
 - ただし実験ネットワークは参照用ネットワークが独立している
 - ほとんど輻輳がない
 - 遅延が一定 (約40msec)
- 参照サーバが2台でも一定の精度は保持
 - 1msecは保てないこともあった
- 実験ネットワークでは4台を参照
 - 冗長性を考慮

日本標準時配信実験としてNTPサーバを公開

- ntp1.jst.mfeed.ad.jp
- ntp2.jst.mfeed.ad.jp
- ntp3.jst.mfeed.ad.jp
 - Celeron 500MHz / Memory 128MB
 - xntpd-3.5.93 (私家版パッチ済)
- 2001年4月10日より公開開始
 - 耐久度の目安としてアクセス数/精度を次頁より紹介

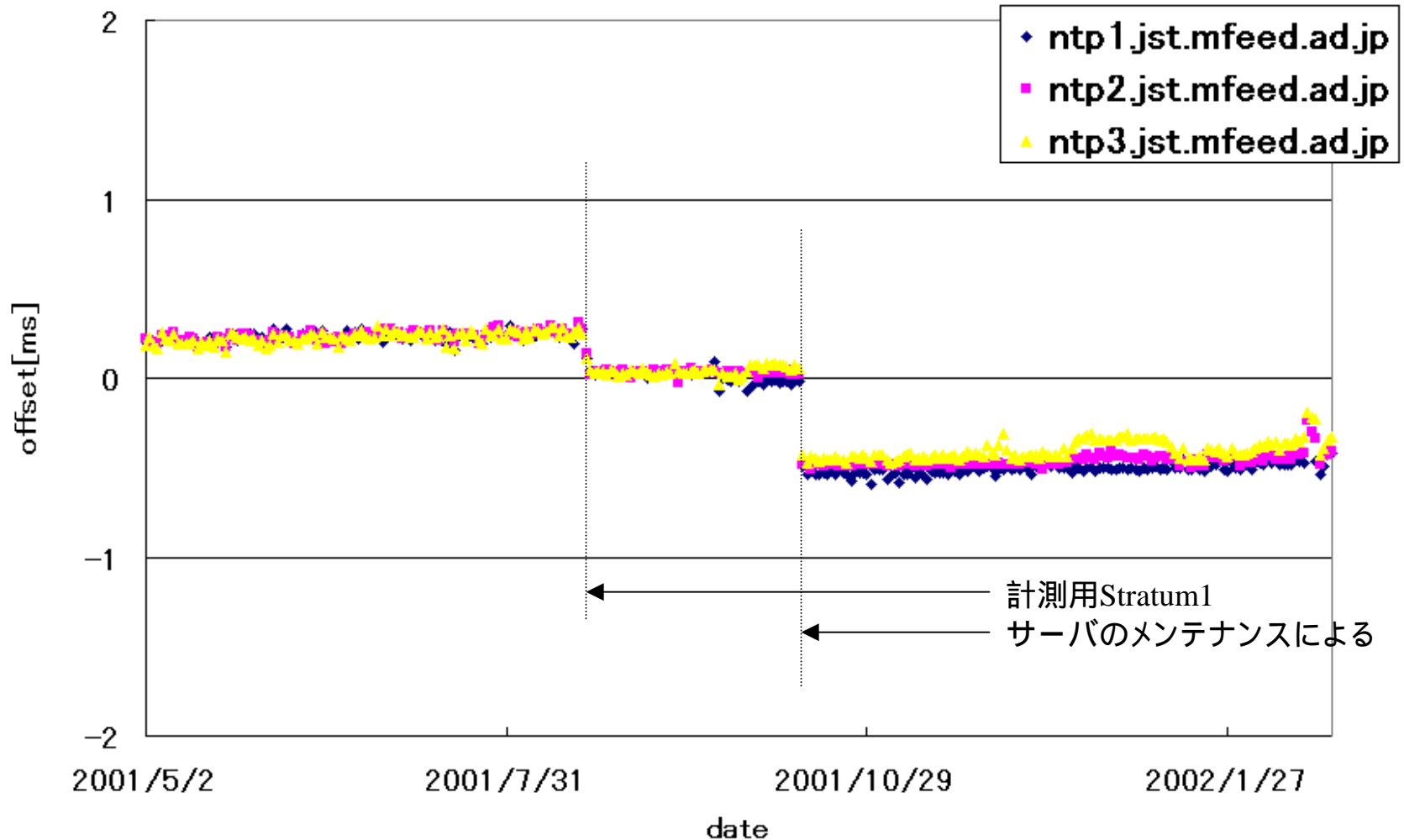
公開時のアクセス状況 (graph)

Stratum2 NTPサーバの一日のアクセス数

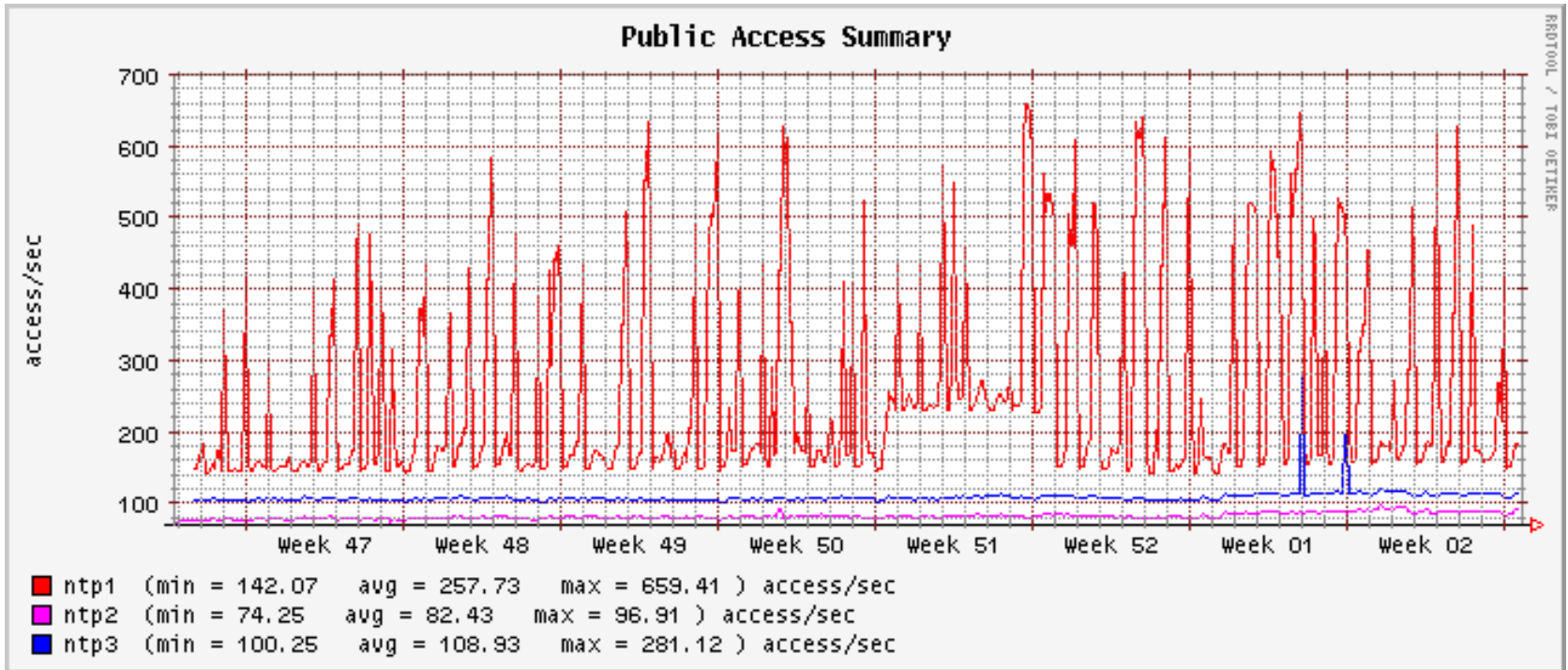


公開時の配信サーバの精度 (graph)

Stratum2 NTPサーバの精度



最近のアクセス状況 (graph)



最近の60日で計測

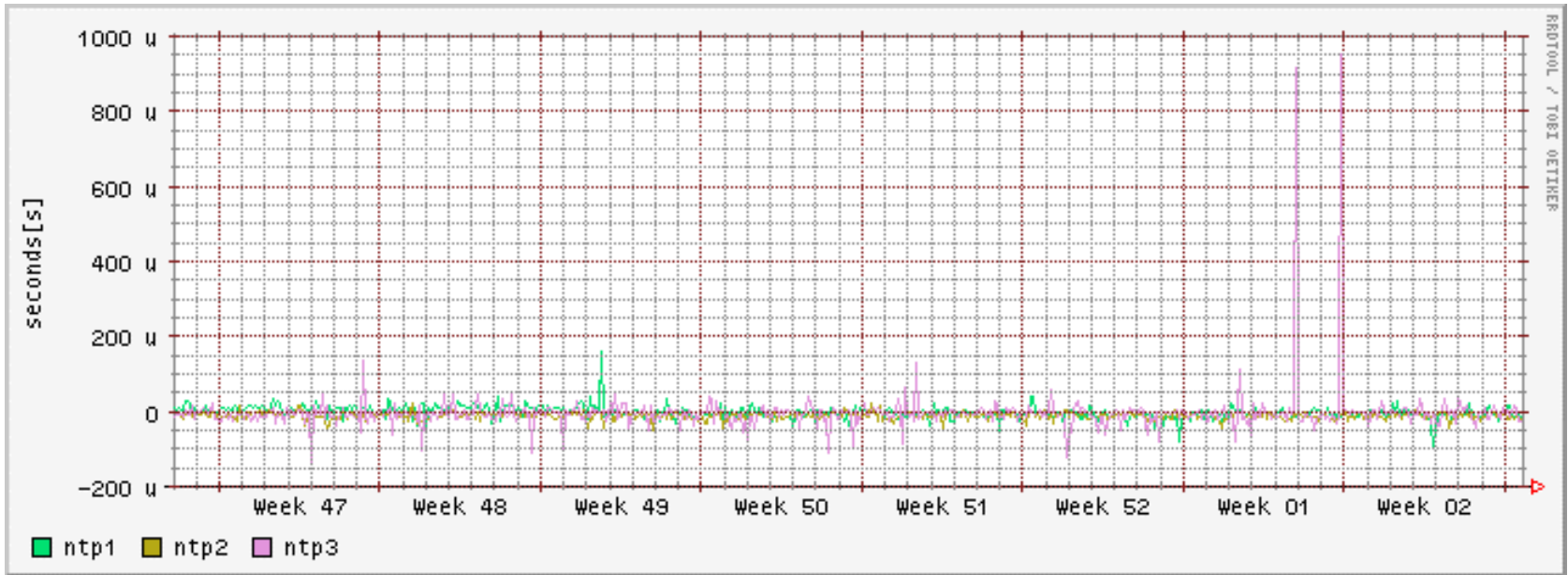
123/UDPの1パケットを
1アクセスとカウント

ntp1: 約2300万アクセス/day

ntp2: 約710万アクセス/day

ntp3: 約940万アクセス/day

最近の配信サーバの精度状況 (graph)



最近の60日で計測

計測用Stratum1サーバ
からの時刻差分

1msec以内で非常に安定

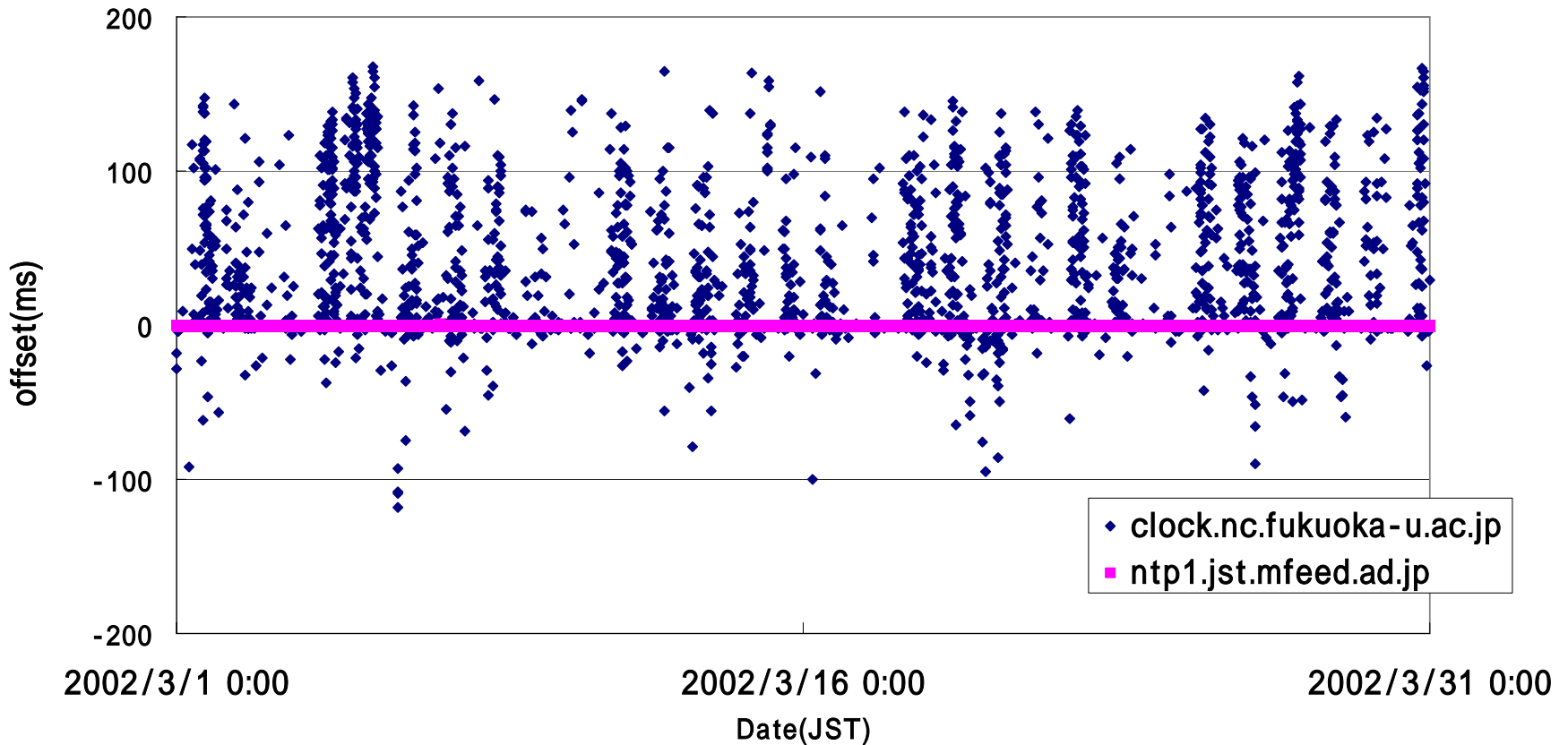
アクセス数と合わせて見ると、NTP
サーバの耐久度は非常に高い

Stratum1とStratum2の精度について

- 近いStratum2と遠いStratum1ではどちらを参照した方が高い精度が得られるのか？
- MFEED内設置の計測用Stratum1サーバより計測
 - Stratum1 サーバは独立して日本標準時に1ms以内の誤差で稼働
 - 計測用Stratum1 サーバと被計測サーバとの時刻の差をoffsetとして記録
- 対象は以下の2台
 - 近いStratum2サーバ: ntp1.jst.mfeed.ad.jp
 - 公開配信サーバのうちの一つ
 - 計測用Stratum1サーバと2hop以内
 - 遠いStratum1サーバ: clock.nc.fukuoka-u.ac.jp
 - 公開されているNTP Stratum1 サーバの一つ
 - 計測用Stratum1サーバとおよそ15hopほど

Stratum1とStratum2の精度について (graph)

Stratum1とStratum2の精度比較



近いStratum2の方が得られる精度がよい

Stratum信仰の誤り

Contents

- A. 本発表について
- B. NTPの特色
- C. 共同研究とその成果について
- D. 安定した時刻情報配信について
 - 遅延の考察
 - サーバの耐久度の考察
 - サーバの選択の考察
 - 監視の考察
 - その他

遅延の考察

- NTPは行きと帰りの遅延が同じであることを仮定
- 高精度な時刻配信ではケアが必要
 - L2 Network
 - GPSが屋上など地理的条件に縛られる場合、可能性あり
 - Gatewayもしくは配信サーバへの距離が遠いネットワーク
 - Arp Requestによる遅延を配慮
 - Ethernet CSMA/CD
 - Switched Networkへの転換
 - L3 Network
 - ClientとServerの通信の経路は行きと帰りで同じか？
 - BGP Multihome環境
 - Hot Potato
 - DynamicRoutingProtocolのcost付け
 - 冗長化プロトコル
 - etc
 - ACL

サーバの耐久度の考察

- NTPサーバ自体の負荷は高くない
 - 3000万アクセス/dayの環境にて、2msec未満の精度を保持
 - トラフィックも1Mbps未満

- 分散は必要か？
 - アクセスの分散目的としては優先度低い
 - それよりも冗長目的で

サーバの選択の考察

- 遠いStratum1より近いNTPサーバを
 - サーバへの距離が遠いほど、行きと帰りの遅延差が大きくなる確率が高い
- ntpdはStratum優先
 - 冗長目的のために、遠いStratum1サーバを参照すると、それがReferenceになる
 - 精度への悪影響の懸念
 - 参照サーバのStratumをよく検討

監視の考察

- 間違った時刻を配信するインパクト
 - 生存確認だけでは監視にならない
- MFEEEDの施策
 - Stratum1からの監視
 - 二種類のしきい値
- 正しい時刻はどちらか？
 - 監視サーバ自体の時刻が狂う可能性
 - 多数決のアルゴリズム

NTPによる高精度な時刻配信のために(まとめ)

■ ネットワーク

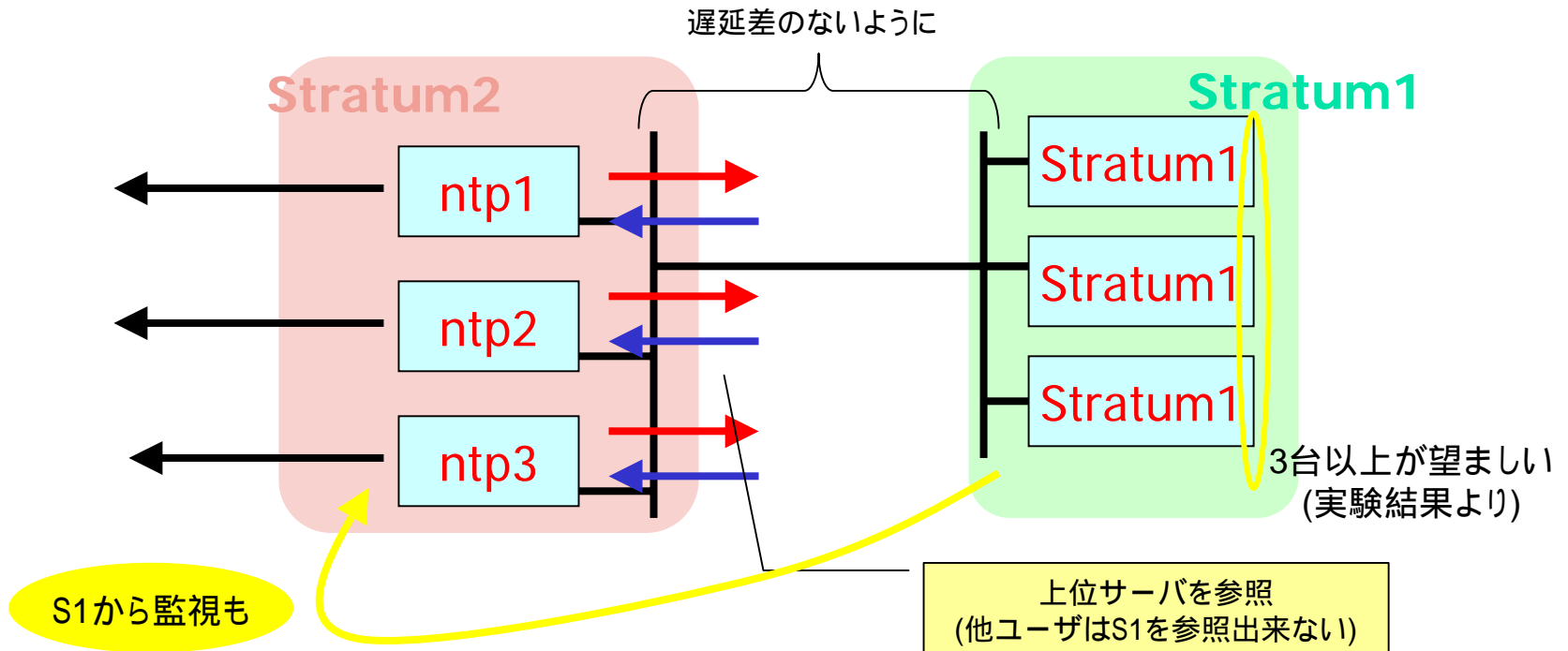
- 通信の行きと帰りの遅延差を極力なくすこと
 - サーバ-サーバ
 - サーバ-クライアント

■ サーバ運用

- NTPの負荷は高くないので、高級なサーバで1台より、安価なサーバで複数台が冗長的に望ましい
- 上位サーバを参照する際は、Stratum1にこだわるより、出来るだけ近くて信頼できるNTPサーバの方が望ましい
- 間違った時刻の配信を避けるための仕組みを
 - Stratum1からの監視
 - 多数決による監視

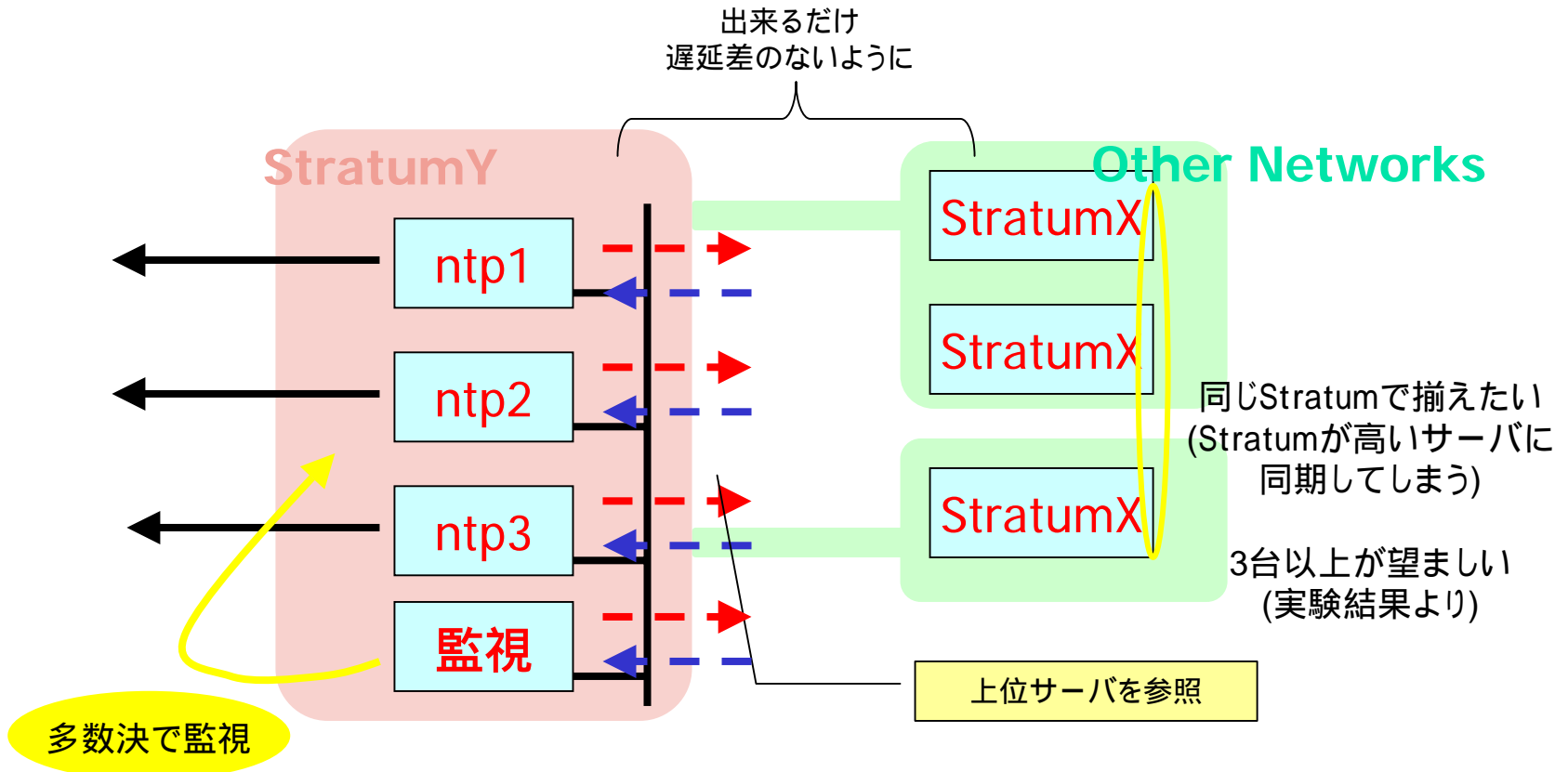
Stratum1がある場合の時刻配信例

- 時刻源となるStratum1と配信用のStratum2を分けると、冗長性が確保でき、かつStratum1の精度を守ることが出来る



Stratum1がない場合の時刻配信例

- ネットワーク的に近く、かつ信頼できるNTPサーバを参照
- 多数決による監視



(参考) NTP version4

- RFC化はまだ行われてない
 - ただしSNTPは先行してRFC化 (RFC2030)
- IPv6 Enable

(参考) 実験公開サーバの今後について

- 同じIPアドレスにて正式サービス化を予定
 - 詳細検討中

Any question or comments?

ご静聴ありがとうございました