JANOG21 Meeting:
トラフィック流量の閾値監視に向けて

2008/1/24

樽井 行保 (インターネットマルチフィード株式会社)
谷津 航 (KDDI株式会社)
原田 薫明 (NTTサービスインテグレーション基盤研究所)
廣川 裕 (NTT情報流通プラットフォーム研究所)

背景

- きっかけ
  - 2007年1月 JANOG19 の “Beyond MRTG”
    - 閾値監視関連に Q&A が‥
      - 株サイトの daily 変動にちゃんと対応できてます？
      - 大手ISPの計画作業時等、どうやって運用してるの？等
    - 多くのエンジニアが興味を持っているんだな、と再認識

- 沖縄で思いついたこと
  - やっぱり「使える」閾値監視ツールが欲しい
  - 高精度な閾値監視ツールを運用に組み込みたい！

  - → 今回、この発表を企画した樽井の動機

本日の論点

□ 目指すところ

- 閾値監視ってどうしてますか？
- 「使える」閾値監視ツールってどういうのだろう？
- 高精度な異常検知アルゴリズムを作るために必要な点、考慮すべき点について情報共有・議論したい。

□ 目指さないところ

- 時系列データは既に手元にあることが前提
- トラフィックデータの収集・蓄積方法、ツールのお話

Agenda

topic	presenter
1, 背景説明、要求条件整理 ・運用現場はこんなツールが欲しい。	KDDI **谷津**
2, 解決手法 – 現場編 ・オペレータによる解決策の摸索	MF/JPNAP **樽井**
3, 解決手法 – 研究者編 ・アルゴリズム屋の研究紹介 ・実データでの比較結果	NTT SI研 **原田** NTT PF研 **廣川**
4, 議論 ・皆さん、どんなツールを動かしてる？ ・高精度なアルゴリズムの精度を上げるにはどのような点に考慮が必要か？	

アウトライン

- (1)「流量の閾値監視」の必要性　　…　谷津
  - 閾値監視の特性
  - ネットワーク運用現場からツールへの要求条件
- (2) 解決策の摸索
- (3) 先端研究の紹介

自己紹介 – KDDI/谷津

- TTNet入社。伝送その他いろいろ保守。
  - 電源、空調、トイレ、花壇、etc.
- パワードコム。AS4716。設計→運用。
- KDDI。AS4716。その他。
- バックボーン運用しつつ色々。お客さんのお手伝いとか。

1-1，まず、流量監視とは？

- 監視業務の1つ。
  - ①定期的に更新されるトラフィックグラフの流量をチェック
  - ②流量が通常から逸脱しているかどうか判定
  - ③想定外の変動を検知した場合にはオペレータに通知

- 検出したいもの：
  - A，**短時間での急激な増加、減少**
    - いずれも通信に影響を与える可能性があり、対応を要する場合が想定される。
  - B，**中・長期的な増加、減少**
    - 設備更改等、最適なプロビジョニングを可能にする。
  - → 今回の狙いは短時間での変動検知

1-2，流量監視の必要性

- いろいろ監視しているけど、充分じゃないかも知れない
  - 拒否されなければ、etherアクセスはping監視。
  - ifIn/OutErrorみてたり。
- アラーム無し故障。 (Silent Down)
  - なんだか調子が悪いとか、エラーが出てるとか。
- 流れるトラフィック量からネットワークの状況変化を敏感に感じとれることがあるかもしれない。
  - これらの傾向変化を数百･数千ものグラフに埋もれさせない。
- (DoSなど)意図しない急激なトラフィック変化。
  - たとえ対応不要だとしても、やっぱり知りたい運用者。
  - 心の準備。
- など。

1-3，実際、どのように監視するのか？

- 運用サイドの選択肢：
  - A，オペレータが定期的にグラフを目視確認。
    - 稼働、運用コストが見合うなら、アリ。
    - 正常、異常の判断は属人的。「職人」が頑張る世界。
  - B，流量監視しない。
    - 諦める。エンジニアとしては負け。
  - C，閾値監視ツールを作ってしまえ！
    - Routine 業務を機械に任せてオペレータは楽したい。
    - ツールに監視させれば、品質も一定になる

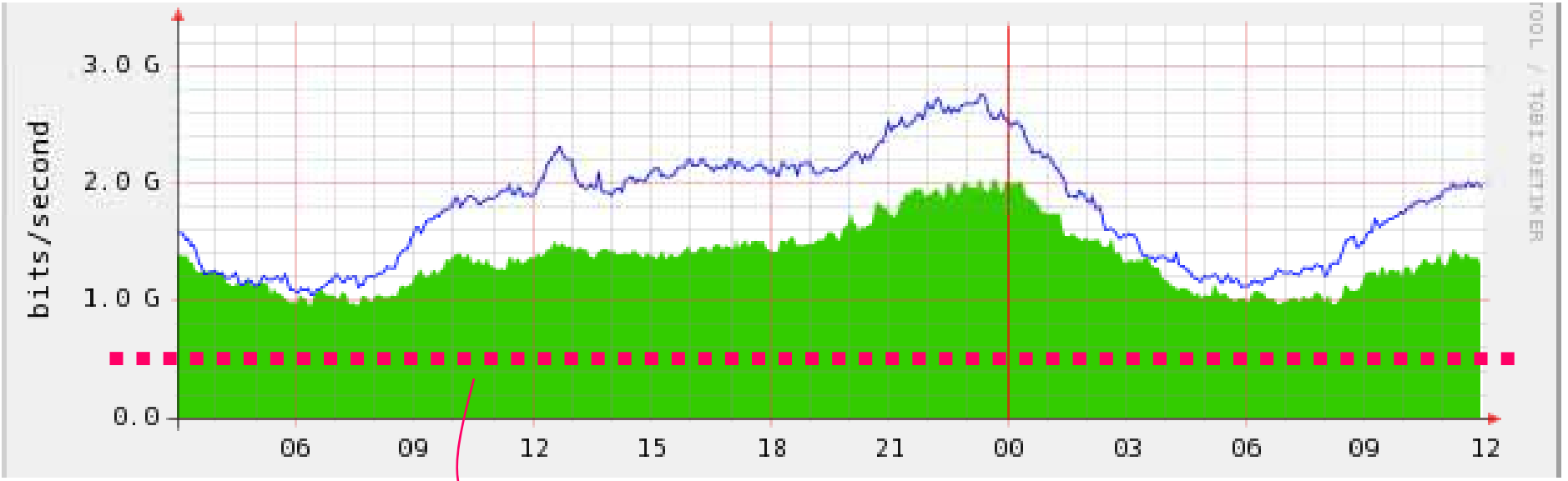
1-4，閾値監視の種類

- 大きく別けて2種類
  - ①固定閾値
    - static
    - 上限・下限の閾値(border)が一定
  - ②動的閾値
    - dynamic
    - チェックの都度、閾値(border)が変動。
  - ご参考:
    - JANOG19「トラフィックの監視、管理って皆さんどうしているのでしょう？ -Beyond MRTG-」

閾値監視の種類

JANOG19の
発表資料より

- 2種類に分けられる。 ... static, dynamic
  - ① 静的に閾値を定義　(固定閾値)
    - 対象毎に上限 or/and 下限値を設定。

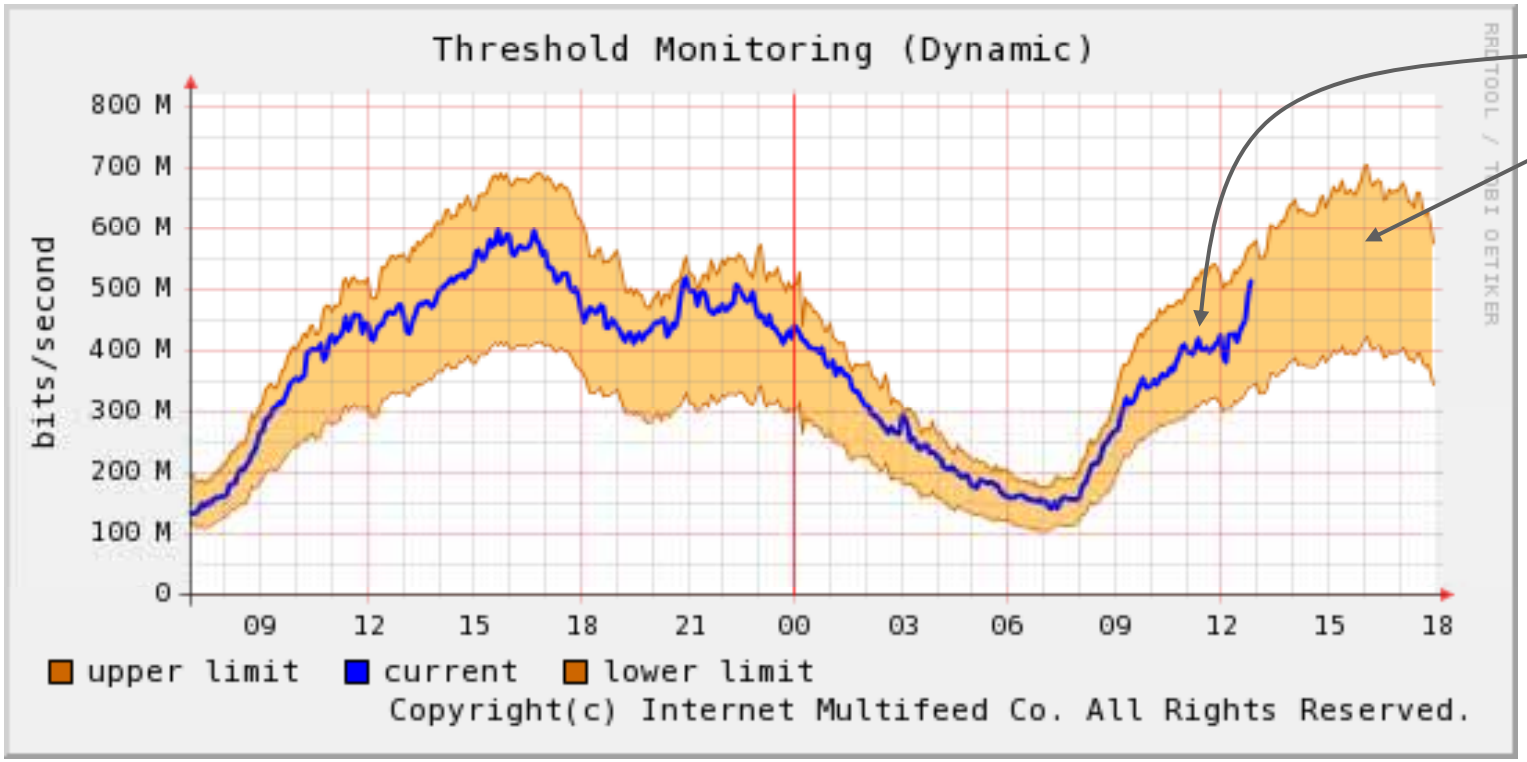


この例では下限の閾値: 500Mbps

continued.

JANOG19の
発表資料より

- ② 動的な閾値監視
  - 過去データのトラフィック傾向を参考に閾値を算出。
  - 「基準線・予測範囲」を導く。

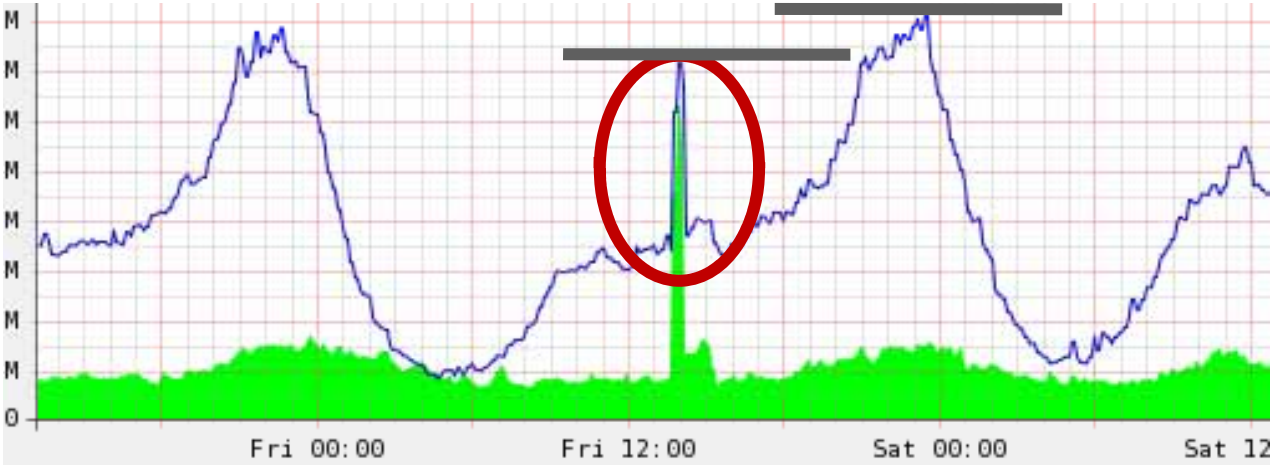


実traffic

予測範囲

1-5，固定(＝静的な)閾値監視じゃダメ

- 通知してほしいポイントを見逃がしがち
  - 固定閾値の場合、１日の Min/Max を考慮に入れると上限･下限の閾値は極端な数値に設定せざるを得ない。



- 設定にそれなりに手間も掛かる。
  - 中長期的なトラフィック量の変動とともに閾値設定も追従させる必要がある。

→ 動的な閾値監視システムが欲しい　★ 本発表の目的

1-6, 現場から要求条件 (まとめ)

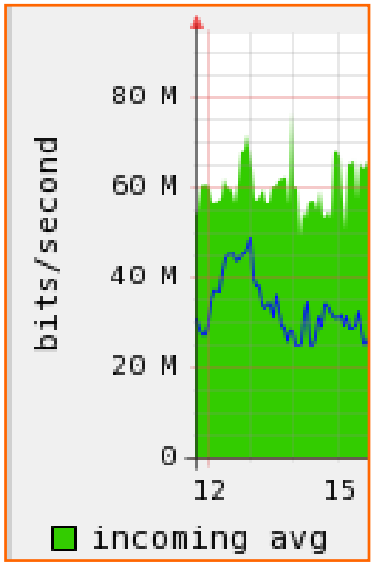
- A, **高精度なアルゴリズム**
  - **何をおいても異常検知精度の高さが命!!**
  - **波形パターン、流量規模を問わず有用なもの**

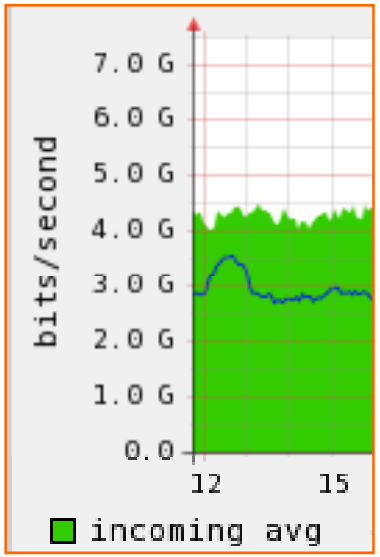
- B, **なるべく** simple **なものを**
  - **設定の簡略化 → 運用負荷の軽減**
  - **計算負荷も軽く**

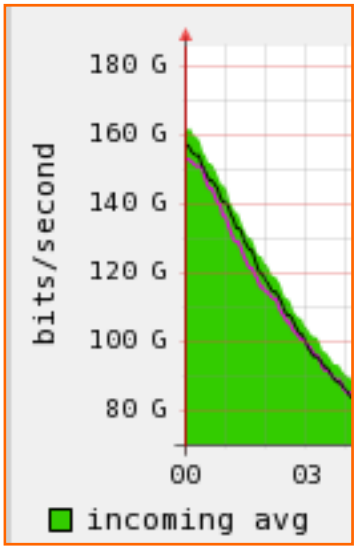
- C, **監視システムとしての付加機能**
  - **変動が発生したタイミングのみ通知してほしい**
  - **注意域、警戒域のレベル分け**

1-6-A, 高精度なアルゴリズム

- 事業者別、および計測ポイントにおいて波形は様々
  - e.g. 波形の特徴、ピーク時間帯、流量の規模



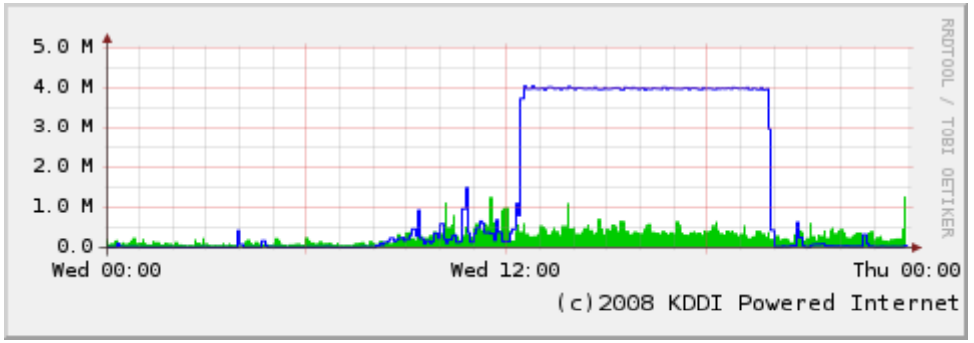




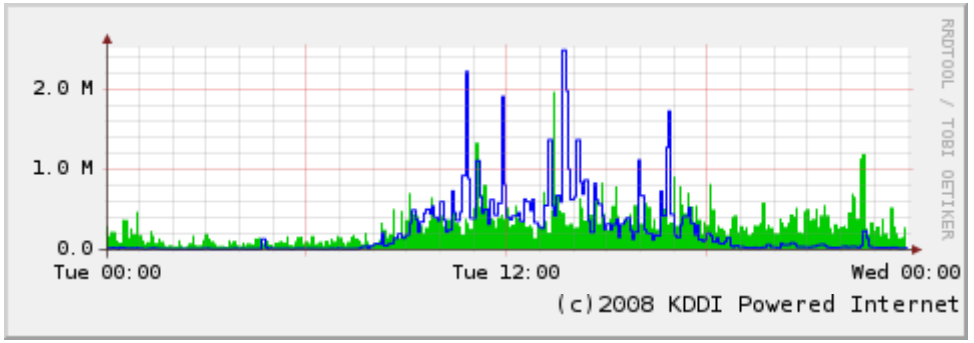
  - → 条件を問わず、精度が一定な判定アルゴリズムが必要

1-6-A, 高精度なアルゴリズム (たとえば)

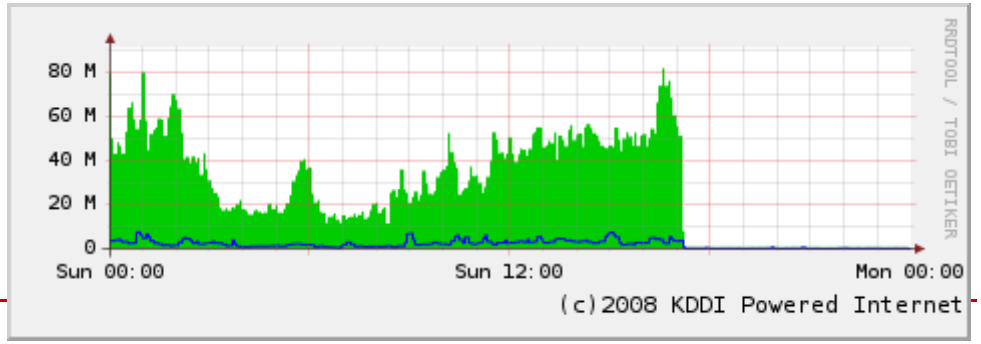
- おおい日
  - アラートが欲しい



- ふつうの日
  - こちらは全部正常
  - 1個もアラートしたくない



- サイレントダウン



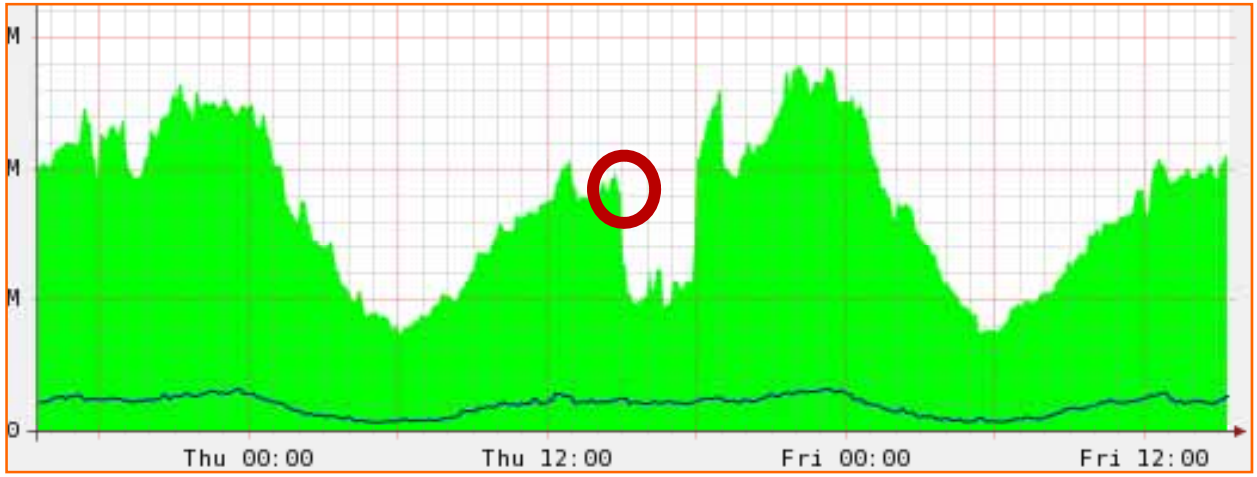
1-6-B, なるべく simple なものを。

- 設定項目、パラメータが少ないもの
  - 設定の簡略化
    - 検知精度が良くても、職人用の道具になっては「負け」
    - 細かいチューニングはできるだけ避けたい。
  - メンテナンスフリーに
    - 一旦、設定投入したら変更不要なものが望ましい。

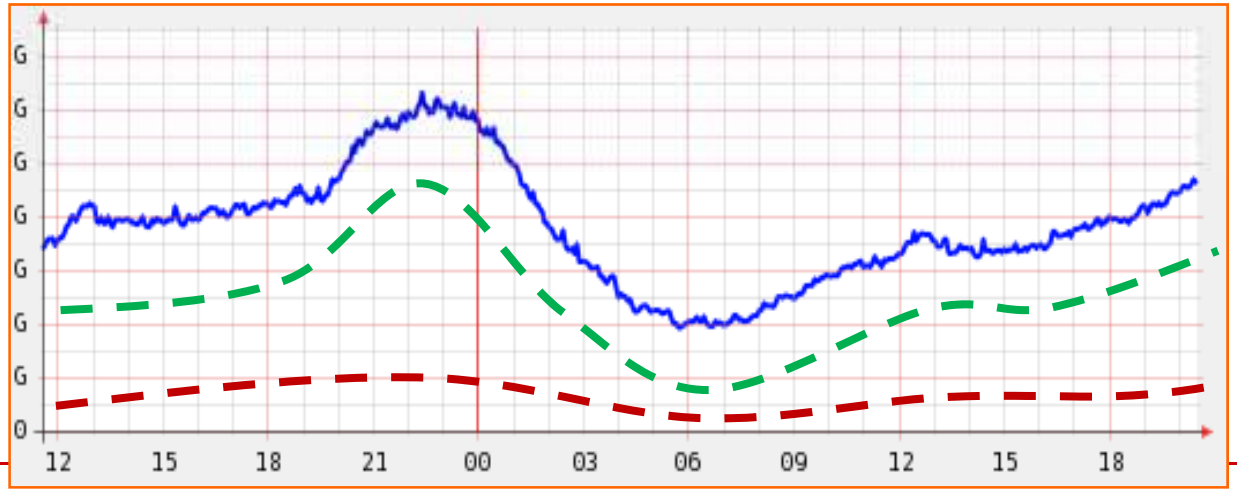
- 計算負荷を軽く
  - このツールのために新たな箱は買えない。
    - e.g. 既存のMRTGサーバに add on して動かす。
    - 数百、数千の Target でも5分以内に完了できるもの。

1-6-C，監視システムとしての付加機能性

■ 変動が発生したタイミング「のみ」通知してほしい



■ 注意域、警戒域のレベル分け



1-6, 現場から要求条件 (まとめ)

- A, **高精度なアルゴリズム**
  - **何をおいても異常検知精度の高さが命!!**
  - **波形パターン、流量規模を問わず有用なもの**

- B, **なるべく** simple **なものを**
  - **設定の簡略化 → 運用負荷の軽減**
  - **計算負荷も軽く**

今回はA&Bに主眼

- C, **監視システムとしての付加機能**
  - **変動が発生したタイミングのみ通知してほしい**
  - **注意域、警戒域のレベル分け**

アウトライン

- (1)「流量の閾値監視」の必要性
- (2) **解決策の摸索　　… 樽井**
  - **現場のオペレータも知恵を絞ってみた**
  - **浮かび上がってくる課題点や限界等**
- (3) 先端研究の紹介

2-1，ここ数年のMultiFEEDでの運用実績

- 周期性を頼りに急激な減少を検知するツール
  - 第1条件：　直近比較　(現在値と15分前を比較)
  - 第2条件：　前週比較　(1,2,3週前の同時刻の平均値と比較)
    - 両条件で閾値を下回ったら異常判定 → ALERT

- 課題点
  - 平日昼間のみ動くツール
  - 祝日は考慮できず
    - 祝日当日の監視は全く役立たず
    - 祝日の翌週、翌々週も監視精度が低下
  - 異常データが発生した翌週、翌々週も同様に精度が低下

2-2，今までの知見と今回の閃きを大切に

- A，祝日対応用のカレンダー
  - 年1回程度、holiday.txt を更新するくらいの手間は許容範囲だろう
- B，やはり、トラフィックの周期性
  - 経験則からしても直観的にもこのアプローチは正しそう
  - 1週前より前日の方がより近い流量になる
- C，(故障等の) 異常データの取り扱い
  - 中央値が解になりそう
  - → 1日前、2日前、3日前のデータを持ってきて最大値と最小値を取り除いた上で比較してみよう

2-3，改良版アルゴリズムの骨子

- 条件は2つのまま
  - 第1条件
    - 直近比較　(現在値と15分前との比較)
  - 第2条件
    - 前日比較　(1,2,3日前の同時刻の中央値と比較)

変更点

  - 両条件で閾値を上回った or 下回った場合、異常判定

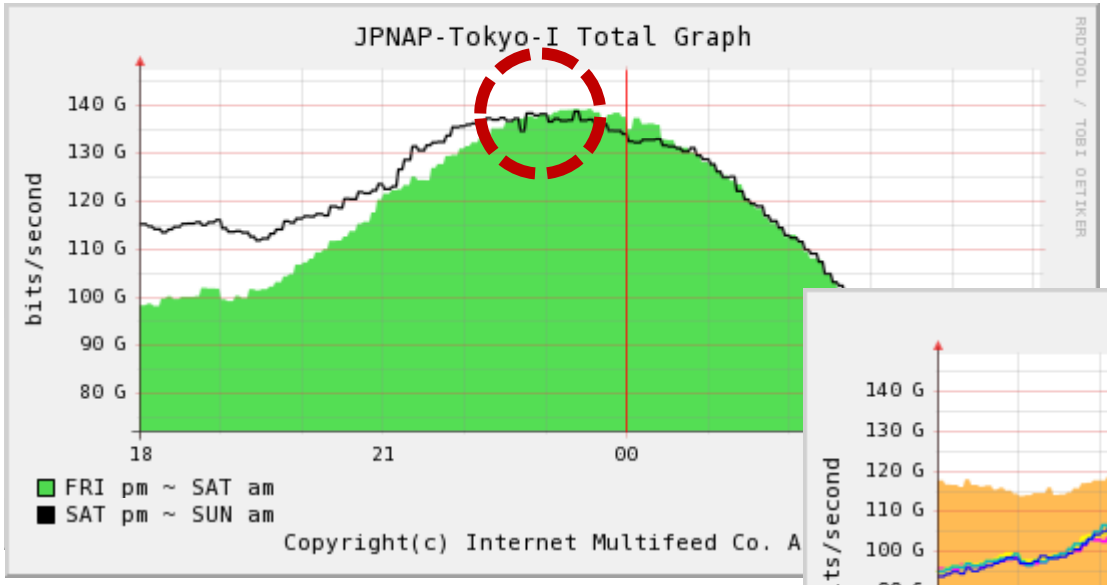
- 至ってシンプル。理想的！

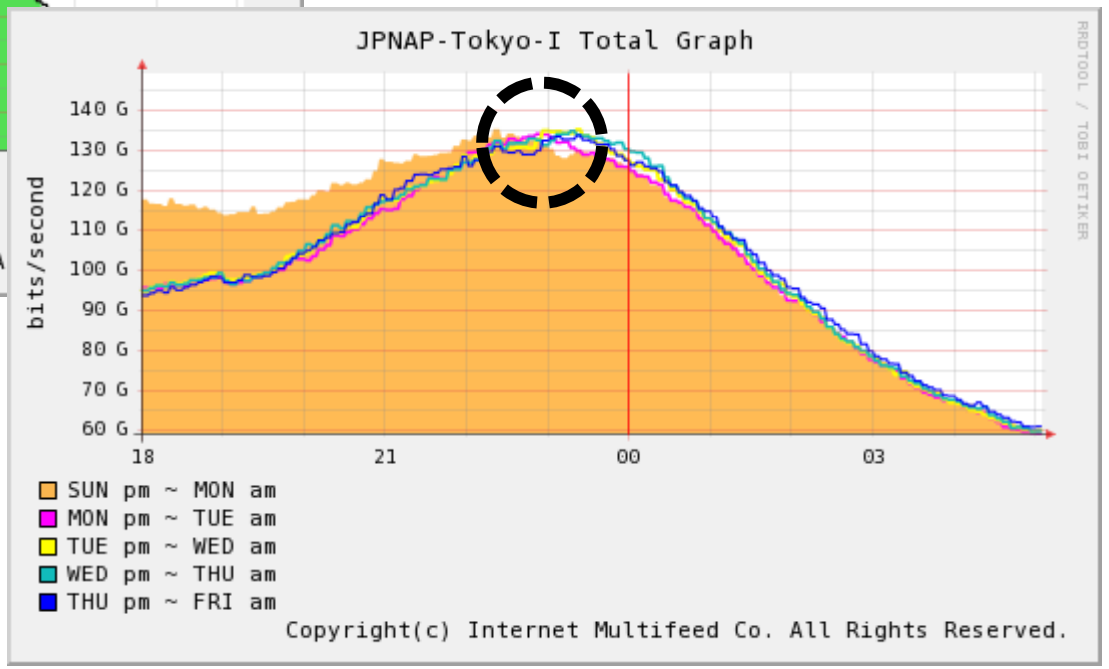
2-4,「前日比較」とした場合のトリック

- トラフィック変動 ＝ 人の動き ＝ 平日･休日に二分
- ここで、トラフィックの平日、休日パタンって？
  - 休日 … 土･日･祝･年末年始等
  - 平日 … 月～金 (つまり、月曜0時～金曜24時が平日？)
- → 厳密には違うのでは？
  - 日曜夜は休日？　金曜夜は平日ですか？
    - 人の動きが変われば、トラフィックにも反映されるはず
- という訳で、実データで可視化してみました
  - 国内最大IXであるJPNAPの総トラフィック量は日本国内の人の動きを推察するのにうってつけ

論より証拠！

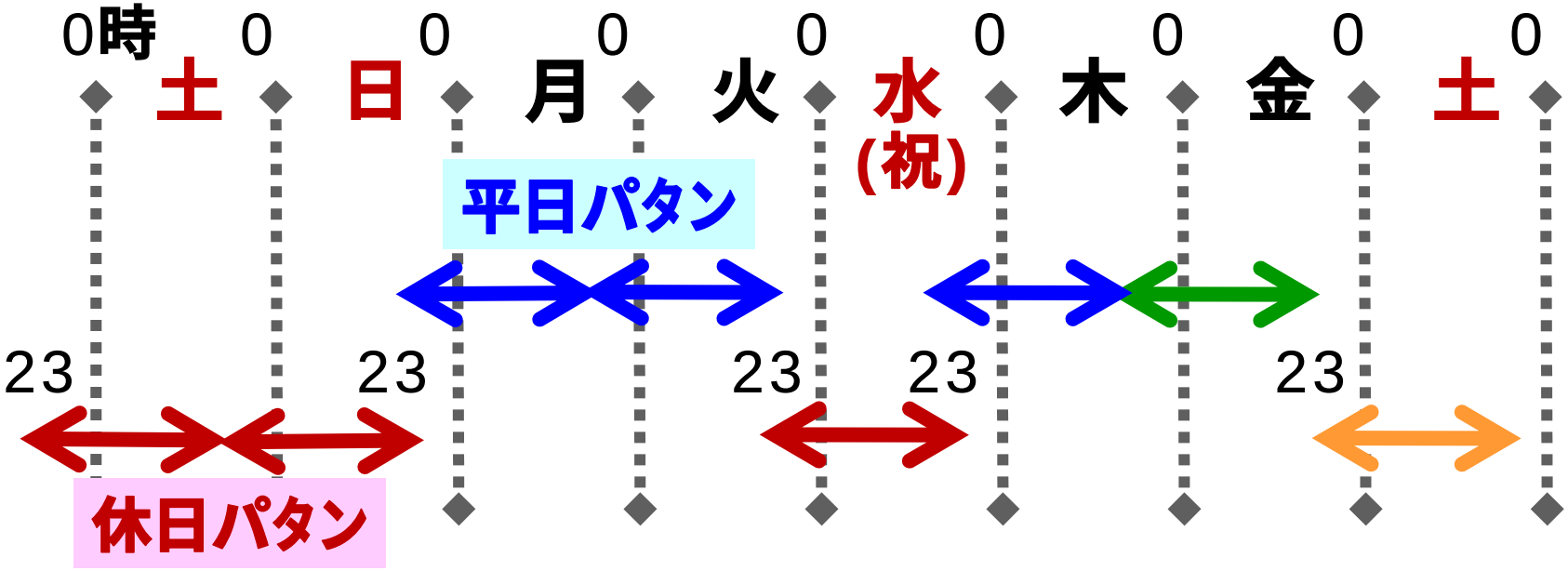
JPNAP TOKYO-I Total
2nd week Dec. 2007





- □ 現在のトラフィック傾向は23時頃が潮目
  - ■ 金曜は23時から休日、 日曜は23時から平日

少しでも精度を向上させるために



- 「前日」のデータと比較する場合：
  - 曜日や祝日だけでなく、時刻も考慮すると良さそう
- → 少し賢いカレンダーを実装し、ツールを改良
  - 時刻を配っていることだし、時間には少しこだわる ^^)

2-5, prototype を作ってみる

- Perl で作成　　(Thx > yuki-k@mfeed)
- RRD から数値を取ってきて、異常有無を判定
- 設定ファイル … XML形式

```
<threshold>
  <upper_percent>125</upper_percent>
  <lower_percent>80</lower_percent>
</threshold>
<target>
  <rrd> /file-to-dir/Target-A.rrd </rrd>
  <rrd> /file-to-dir/Target-B.rrd </rrd>
</target>
<data_source>
  <select>ds0</select>
</data_source>
```

対象(グループ)毎に上限、下限の閾値をn% と設定可。

(100%＝予測値)

2-6, テスト結果 & 課題点、気づき

- 実データでの検知精度もまずまず
  - e.g. MS Update, ニコニコ動画のメンテナンス等の流量変化を検知
- しかし、まだ運用ツールとして満足できる域ではない
- 深夜・早朝帯にポロポロ誤検知あり
  - Trafficが底の時間帯に発生するわずかな動きを異常判定
- 異常発生時の検知はできるが、収束時は検知不能
  - Ping NG は来ても、回復時の Ping OK が出せない感じ

2-7，深まる悩み、運用現場の限界

- 上限・下限の閾値を予測値の何％くらいにすれば良いのか、**パラメータの初期設定やチューニングに悩みそう**
  - エイヤで投入して、当面は精度を見つつ調整？

- 祝日カレンダーを作るくらい簡単、と言ったけれど・・・
  - **ところ変われば、「休日」の概念も異なる**
    - iDCとしては平日だけど、IXPとしては休日パタン、等
  - この類の経験則による設定・判断項目は無くしたい

- → **どうも想定していたほど、楽にならない・・・**

2-8，これって運用現場だけの悩み？

- (今回は例としてトラフィック流量に着目しているが、)
  時系列データを解析し、近未来予測＆異常判定する試みは適用範囲が広い
  - bps, RTT, KVA, **温度**, **メモリ使用量**, CPU, Query, …

- **近年のIT運用分野だけに留まらない (はず)**
  - **歴史ある電話の世界の「トラヒック」**
  - **デンコちゃんの電力需要予測**
  - **為替、株式相場等の経済分野の根幹　・・・**

- **→ 学問や研究課題にもなりそう**

2-9, データマイニングのプロに help !!

- 使えそうな異常検知アルゴリズムはありませんか？
  - rrdtool には未来予測式として Holt-Winters という統計手法が実装されているが、使いモノになるの？
  - その他、世の中にはどんな未来予測のアルゴリズムや先端研究が存在するのだろうか？ (素朴な疑問)

- → というわけで、研究所の皆さん、出番です！

アウトライン

- (1)「流量の閾値監視」の必要性
- (2) 解決策の摸索
- (3) **先端研究の紹介　　…　原田、廣川**
  - **既存の統計学的手法、および先端研究事例の紹介**