KDDI R&D LABS

社外秘

ユーザ通信状況分析ツール (2012/08/29)

(株) KDDI研究所
ネットワーク設計グループ
稗圃

KDDI R&D LABS

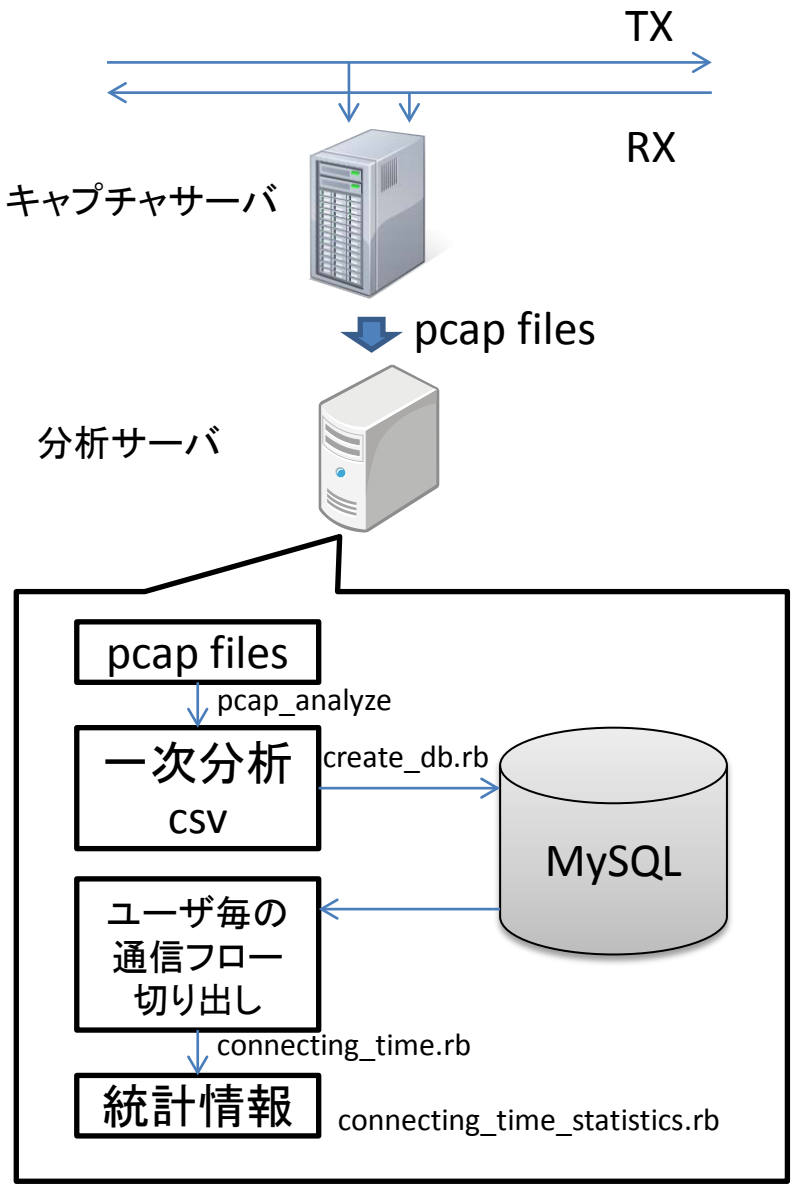
社外秘

概要

ツール群概要

- pcap_analyze
  - 福元研究主査が中心となって開発を進めている，pcap分析ツール.
    - tcp/udp のセッション単位でセッションの確立時刻・伝送量やステートをcsvに出力する.
    - HTTPに限り，接続先情報などHTTPのステータスをcsv出力可能.
    - HTTPボディに位置情報やHTML5関連技術らしい記述が含まれている場合は，それらの情報を抜き出すことができる.
- create_db_tcp.rb
  - pcap_analyze で出力された tcp*.csv ファイルを MySQL 5.1 DBへ登録するためのツール.
  - tcpについて記述された csv のみ取り扱い可能.
- create_db_vol.rb
  - pcap_analyze で出力された volume*.csv ファイルを MySQL 5.1 DB へ登録するためのツール
  - volume*.csv には，ユーザ毎(src_ipaddr毎)に通信が存在している時間(基本的に秒単位)の上下通信ボリュームが記載されている.
- connecting_time.rb
  - MySQL に登録された tcp セッション情報から，ユーザごとに通信の流れを解析，jsonとして出力する.
  - 本資料で説明.

分析のフロー



1. 専用キャプチャサーバ(Swiftwing SIRIUS)でキャプチャ→ pcap ファイル取得
2. pcapファイルを分析サーバに渡す
3. pcap_analyze ツール群でtcp/http分析
4. create_db.rb でMySQL DB登録
5. connecting_time.rb によりユーザ毎の通信の流れを分析→JSON化
6. XMLからユーザ全体の統計情報を切り出し

KDDI R&D LABS

社外秘

connecting_time.rb 入出力概要

KDDI R&D LABS

社外秘

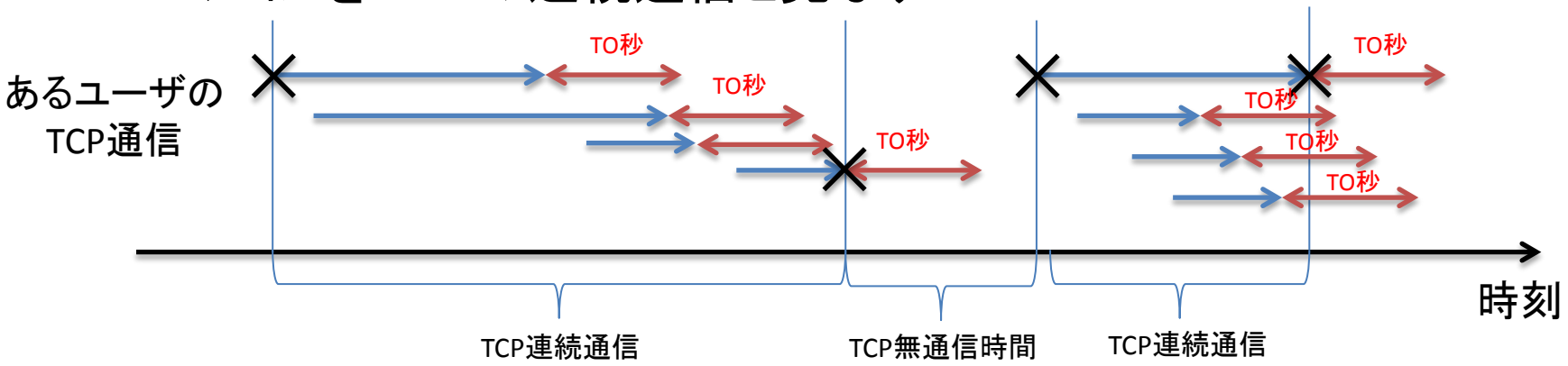
- 入力
  - MySQL に格納された pcap_analyze のセッション毎の分析結果
- 出力
  - JSONファイルを出力
    - ユーザ毎の通信フローは二種類
      - tcp*.csv から算出した TCPセッションの流れを時系列で整理
      - volume*.csv から算出したユーザのオンライン状況を時系列で整理
- 注: 連続通信の概念
  - ユーザ毎のTCPフローやオンライン状況のフローはいずれも連続通信を識別し，タグで切り分けで出力する.
  - 連続通信の定義は次ページ

TCP連続通信の識別

KDDI R&D LABS

社外秘

- TCP連続通信
  - あるユーザがTCPセッションを確立し，該時刻から該セッションの終了時刻+TIMEOUT秒以内に発生した該ユーザのTCPセッションを一つの連続通信と見なす．



  - ×～× までを連続通信と見なす
  - あるTCP連続通信から次のTCP連続通信までの間をTCP無通信時間とする
    - 前TCP連続通信の最終TCP確立終了時刻～次TCP連続通信の最初のTCP確立時刻＝TCP無通信時間
    - 注意: TCP連続通信内であっても実際の通信が発生していないタイミングもある(Keep-Aliveなど)

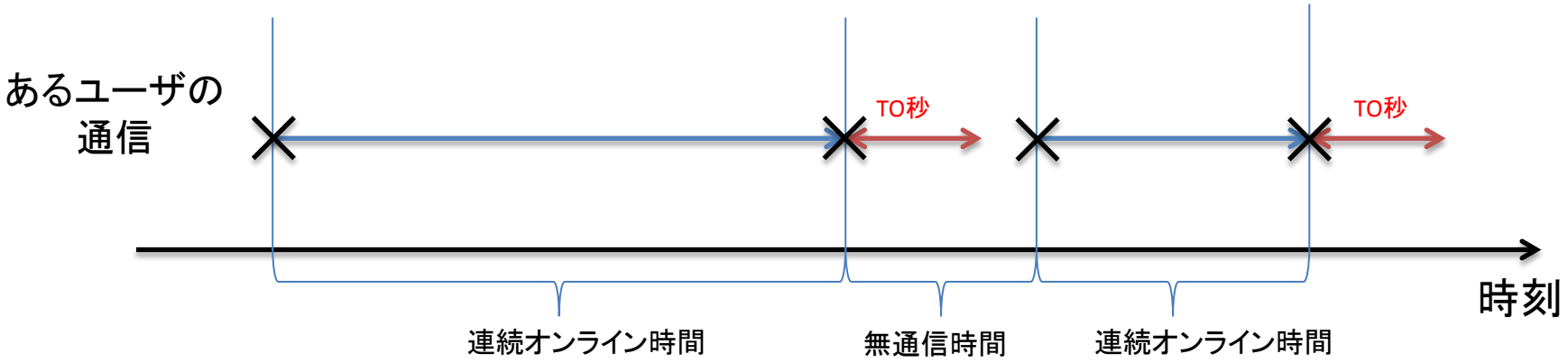
連続オンラインの識別

KDDI R&D LABS

社外秘

■ 連続オンライン

– あるユーザが何らかの(L2以上の)通信を開始したとき，当該通信がTIMEOUT秒以上途切れたときまでを一つの連続的なオンライン状態と見なす．



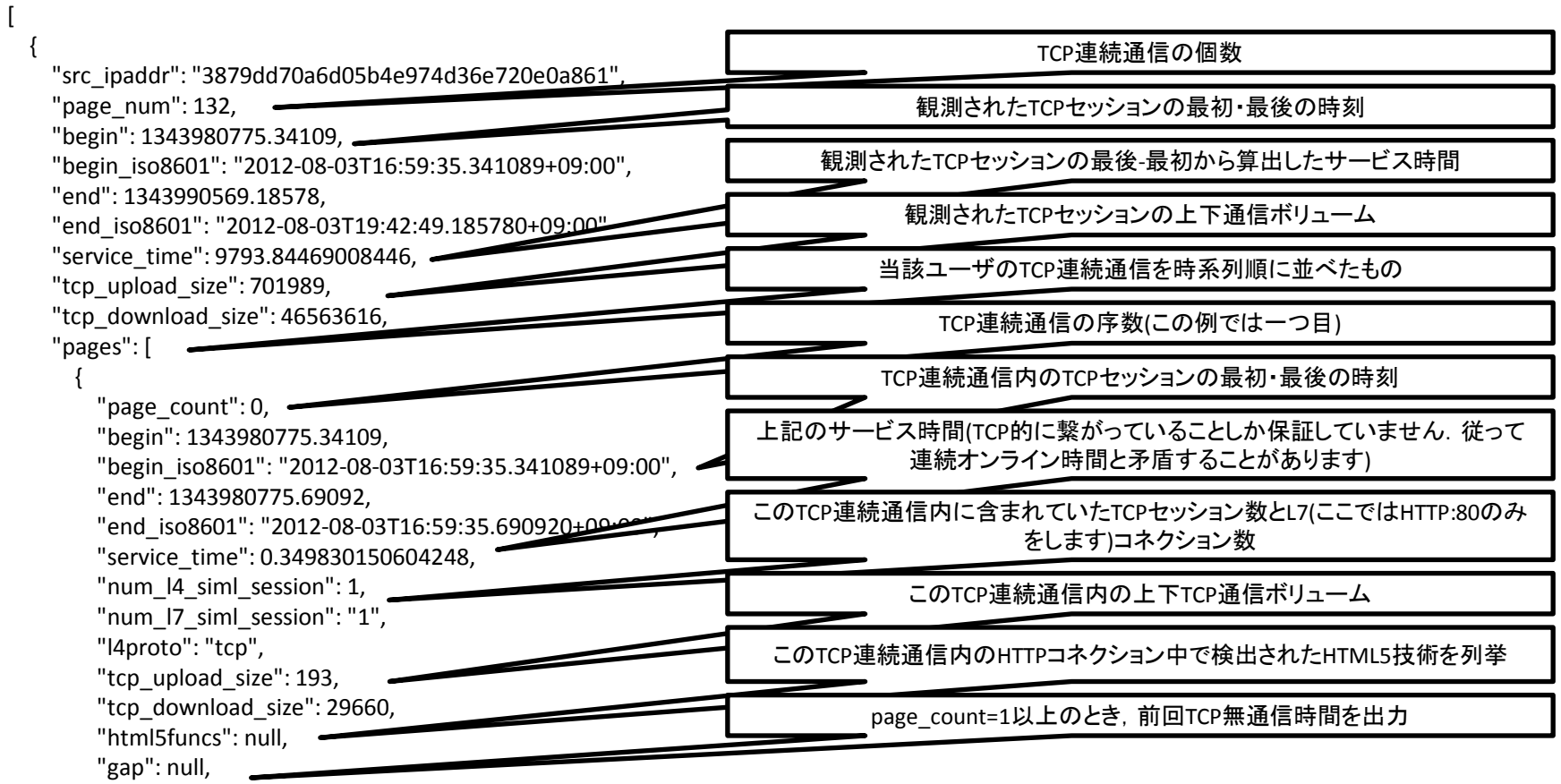
– × ～ × までを連続オンライン時間と見なす

– ある連続オンライン時間から次の連続オンライン時間までの間を無通信時間とする

JSON出力(1)

KDDI R&D LABS

社外秘



JSON出力(2)

KDDI R&D LABS

社外秘


```
"l3l4": [ {
        "tcp_hash": "5626d615728b0339d4600f3722532137",
        "dst_ipaddr": "219.103.34.226",
        "src_port": 32767,
        "dst_port": 8937,
        "begin": 1343980775.34109,
        "begin_iso8601": "2012-08-03T16:59:35.341089+09:00",
        "end": 1343980775.69092,
        "end_iso8601": "2012-08-03T16:59:35.690920+09:00",
        "service_time": 0.349830150604248,
        "tcp_upload_size": 193,
        "tcp_download_size": 29660,
        "l7proto": "other"
    }
  ]
},
```


このTCP連続通信内のTCPセッションを設立時刻順に表示
ただし，本項目はTCPセッションごとに出力されるとは限りません．
同一TCPセッション上にHTTPコネクションが複数ある場合，HTTPコネクションごとに出力されます(pcap_analyzeの仕様通り)

上記に関連して，TCPセッションを識別するためのハッシュ値
本値が一致する項目は，同一TCPセッションと見なしてください

当該TCPセッションの設立開始時刻と終了時刻

当該TCPセッションのサービス時間

当該TCPセッションの上下通信ボリューム

L7コネクションの種類．http/https/other のいずれか
dst_port で判断しており，それぞれ 80, 443, それ以外です．
l7proto が http かつ分析可能であれば，HTTPコネクションの詳細が項目”l7”に記述されます(次ページ)

JSON出力(3) ～L7proto='http'のとき～

KDDI R&D LABS

社外秘


```
{
        "tcp_hash": "410e14cf3179d06166a310e12ad56bf6",
        "dst_ipaddr": "184.26.223.144",
        "src_port": 32767,
        "dst_port": 80,
        "begin": 1343988042.23313,
        "begin_iso8601": "2012-08-03T19:00:42.233129+09:00",
        "end": 1343988053.12571,
        "end_iso8601": "2012-08-03T19:00:53.125710+09:0
        "service_time": 10.8925800323486,
        "tcp_upload_size": 473,
        "tcp_download_size": 24642,
        "l7proto": "http",
        "l7": {
          "begin": 1343988042.3332,
          "begin_iso8601": "2012-08-03T19:00:42.333199+
          "end": 1343988042.61029,
          "end_iso8601": "2012-08-03T19:00:42.610290
          "service_time": 0.277090072631836,
          "response_size": 24642,
          "response_actual_size": 24642,
          "request_host": "platform.twitter.com",
          "request_path": "/widgets/follow_button.1340179658.html",
          "gps_latitude": null,
          "gps_longitude": null,
          "request_referer": "http://sp.lalu.jp/community/show/categoryId/80/topicsId/523112.html",
          "request_user_agent": "Mozilla/5.0 (iPhone; CPU iPhone OS 5_0_1 like Mac OS X) AppleWebKit/534.46 (KHTML, like Gecko) Mobile/9A405
Lalune_iOS/1.0.9",
          "html5": "html5"
        }
    },
```


l7proto='http'かつpcap_analyzeにてHTTPコネクションであると分析されたときに限り，L7項目にHTTPコネクションの情報を出力する．ここだけ順不同なので注意

HTTPコネクションの開始・終了時刻(request_begin ～ response_end)

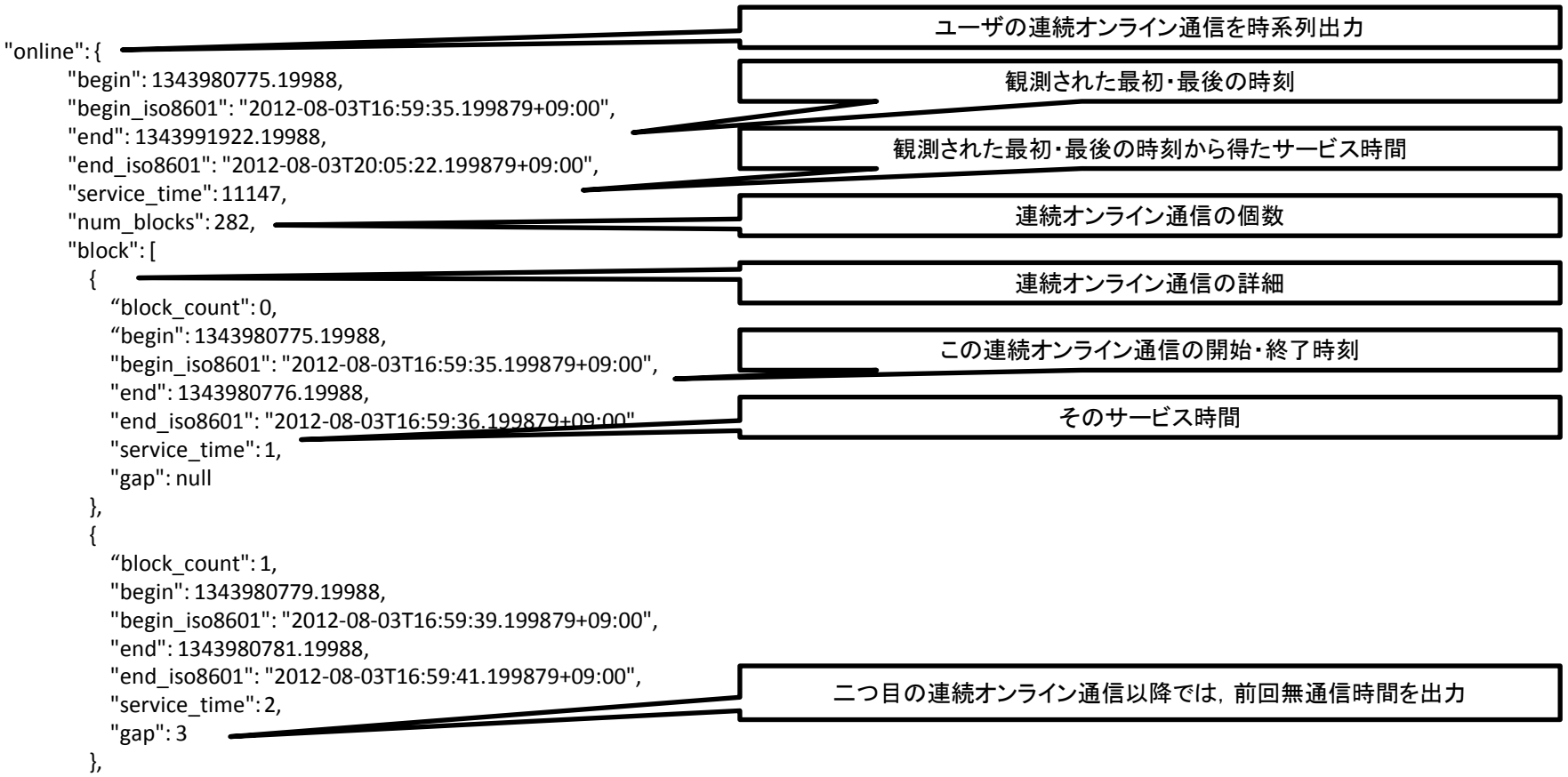
そのサービス時間

pcap_analyzeで得たHTTPコネクションの情報を転記

JSON出力(4) ～ 連続オンライン～

KDDI R&D LABS

社外秘



時間分解能はpcap_analyze の –output-traffic-volume で与えたサンプリング周期で決まります．(デフォルト分解能は1秒)