マルウェアの転送ログを利用した地域毎のボット活動分析

金井 瑛 †

† インターネットマルチフィード株式会社
〒 100-0004 東京都千代田区大手町 1-5-1
大手町ファーストスクエア EAST タワー 3 階
E-mail: kanai@mfeed.ad.jp

あらまし　インターネット利用者の様々なセキュリティを脅かすボットは，日々多くの亜種が出現し，それらに用いられる高度な遮蔽技術は対策を困難とする．ボットは金銭的利益を得る手段として利用できる背景から，通信や活動全般の形態を日々進化させており，具体的な活動の変化についてはよく知られていない．本稿では，日本国内のハニーポットで観測された 3 年分のマルウェア転送ログを用いて，マルウェアの転送元の国に着目した傾向の変化を分析した．分析の結果，マルウェアの転送は転送元の国の時間帯と連動した転送数がみられることを示した．また，北米及びヨーロッパ地域から 2009 年後半から継続して転送ノードが増加している様子について示した．

Analyzing geo-specific bot based on malware transfer records

Akira Kanai†

†INTERNET MULTIFEED CO.
OTEMACHI 1st SQUARE EAST TOWER,3F,
1-5-1 Otemachi, Chiyoda-ku, Tokyo 100-0004, Japan

Abstract　Bot threatens Internet users' security with malicious behaviors. The new varieties of botware appear everyday. Since those bots conceal their existences and activities, it is difficult to defend from them. In this paper, malware traffic logs, which were captured by honeypots in Japan for three years, were analyzed especially focusing on nations. As a result, it was discovered that the amount of malware transports is coordinated to the local time of where the transport nodes reside in. In addition, it was shown that the number of transport nodes has been continuously increasing in North America and Europe since late 2009.

1　はじめに

インターネット上におけるボットネットの対策は急務であり，近年では多くの研究に加えて国が主導となった予防推進事業が進められている．ボットの傾向を分析し対策を考案するために，静的解析や動的解析を用いたマルウェア自身の解析と，感染したノードが攻撃者から操作される仕組みの解析により，ボットネット全体の動作の調査が進められている．しかし，ボットはボットネット拡大のための感染活動や維持活動を維持するための多様な改良が実施されており，ボットの対策手法を研究や実施するためにはボット活動の傾向分析を継続的に実施する必要がある．

筆者は 2008 年にマルウェアが感染する手順の 1 つである転送に着目し，CCC DATAset 2008 を用いて日本国内のハニーポットで取得された転送イベントのデータセットを分析した [1]．この結果，日本国内で発生する転送イベントは日本のネットワーク形態やトラフィックの流量と深い関係が存在することを示した．

本稿では，ほぼ同様の環境で収集された情報であ

る MWS 2010 Datasets[2](以下，データセット) を用いて，同様の傾向が継続しているかを調査した．

本稿は全 6 節で構成される．第 2 節では，本稿の解析に用いるデータセットの概要と特徴について述べる．第 3 節では，このデータセット期間中のノード数遷移を国ごとに見ていき，その特徴を述べる．第 4 節において，各国の一日における時間ごとのノード数を日本と比較する．第 5 節では，日本における一日における時間ごとのノード数の分布を平日と休日で違いがあるか比較する．最後に第 6 節で本稿の調査で得られた結果をまとめる．

2　調査データ

本稿で用いたデータセット 2010 は「攻撃元データ」と呼ばれるデータセットであり，2007 年 11 月 1 日から 2010 年 04 月 30 日までの 30 月間にわたる攻撃に関わるイベントのログが 6,575,080 件含まれている．データセットに含まれる全イベントのうち，プッシュ型イベントは 222, 294 件 (3.4%) でありプル型イベントは 6, 352, 786 件 (96.6%) である．このイベントは短期間における同一 IP アドレスからの攻撃によるイベントを別に扱う．イベントに含まれる唯一な IP アドレス数 689040 件のうち，プッシュ型イベントのノード数は 95891 件 (13.9%) であり，プル型イベントのノード数は 593, 149 件 (86.1%) である．本稿中ではある期間中に含まれたログの件数をイベント数と呼び，ある期間中のログから重複したアドレスを排除した唯一な IP アドレスの個数をノード数と呼ぶ．

以後の分析で転送元ノードの国に着目し，傾向を考察するには全てノード数を用いる．2008 年の活動分析の際，データセットには約 300 万件のイベント数が含まれていたが，約 3 割のイベントが同一ノードからのイベントであった．このため，イベント数への着目は攻撃イベントの件数を調査できるが，攻撃元となっているノード数の増減を知るには適していない．

2.1　ボットの動作と転送方式の種別

本節では,「攻撃元データ」に含まれる情報を述べるにあたり，ボットの基本的な動作について述べる．ボットが新しい他のノードに感染する際の活動は次の 3 つに分類できる．

1. 攻撃:脆弱なノードの脆弱性を利用し対して小さなプログラム（ダウンローダ）を送り込む．
2. 転送:ダウンローダはインターネット上の配布元ノードからマルウェアを転送する．
3. 接続:マルウェアに感染したノードはボットネットの管理サーバに接続する．以後は管理サーバからの命令に従い活動する．

攻撃はノードの脆弱性を利用するが，多くの攻撃手法は文字列処理などメモリ管理の脆弱性を利用したものであり，攻撃時に多くの情報を攻撃対象のノードに転送できない．そのため，攻撃者は**ダウンローダ**と呼ばれる小さなソフトウェアを送り込みノードに実行させる．この手法には**プッシュ型転送**と**プル型転送**の 2 つの種類がある．この 2 つはそれぞれマルウェアの転送方法が異なる．それぞれの転送方法について図 1 を用いて解説する．

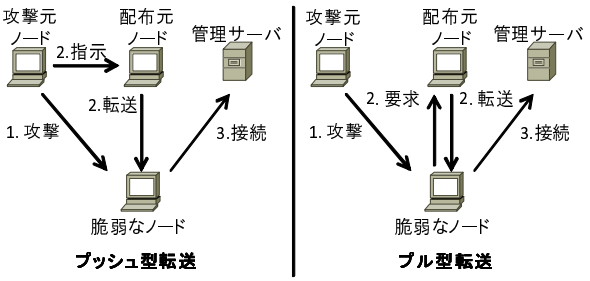

図 1: プッシュ型転送とプル型転送

図左のプッシュ型転送ではダウンローダが実行されると脆弱なノードはあらかじめプログラムされたポート番号でインターネットからの通信を受け付ける．攻撃元ノードは配布元ノードに対して脆弱なノードの IP アドレス情報とポート番号を伝える．攻撃元ノードと配布元ノードは同一ノードの場合もある．配布元ノードは受け取った情報を用いて，脆弱なノードにマルウェアをアップロードする．プッシュ型転送はインターネット上のノードから通信が開始されるため，NAPT など，外部からの通信が制限された環境においてプッシュ型の転送は動作しない．

図右のプル型転送ではダウンローダが実行されると脆弱なノードはあらかじめプログラムされた IP アドレスおよびポート番号で管理サーバに接続し，マルウェアをダウンロードする．プル型転送は脆弱なノードから通信が開始されるため，外部からの通信が制限された環境においても動作する．転送が終了し，マルウェアに感染したノードはボットネット

の管理サーバに接続する．以後は管理サーバからの命令に従い活動する．

2.2 IP アドレスと国の関連付け

本稿では，攻撃元データに含まれる IP アドレスから配布元ノードの所属する国を推定する．推定には IP アドレスブロックから該当アドレスを持つノードが存在する国を推定できるソフトウェアである MaxMind 社の提供する GeoIP[3] を用いた．

3 長期におけるノード数の推移

本節では，データセットに含まれる 30 月間のそれぞれの月のノード数変化を国別に比較することで，マルウェアの活動方式の変化について調査する．2007 年 11 月と 2010 年 04 月のプル型転送とプッシュ型転送のノード数が多かった 5ヶ国について，それぞれ表 1 と表 2 に示す．以降はそれぞれの転送方式の傾向について述べる．

表 1: プル型転送ノード数上位 5 国

2007 年 11 月 計: 42521 ノード		2010 年 04 月 計: 11268 ノード	
国	ノード数	国	ノード数
日本	39496	日本	5061
中国	784	台湾	1480
台湾	582	ロシア	851
フィリピン	490	米国	441
インド	162	中国	379

表 2: プッシュ型転送ノード数上位 5 国

2007 年 11 月 計: 8346 ノード		2010 年 04 月 計: 836 ノード	
国	ノード数	国	ノード数
日本	1533	米国	177
米国	1168	韓国	115
韓国	1044	日本	86
フランス	419	台湾	46
台湾	394	フランス	41

3.1 プル型転送の傾向

2 つのグラフでプル型転送の傾向を示す．まず，図 2 に，各国の 2007 年 11 月のノード数を 1 とした際の，以降の月におけるノード数の割合を示した．この図では，日本，中国，フィリピンとインドについて示した．また，図 3 に各国の月におけるノード数を示した．この図は割合ではなく，ノード数を示しており，ロシア，米国と台湾についてこの図で示す．これらの国は 2007 年 11 月を基準に考えると期間内に大きな増加が見られたものである．

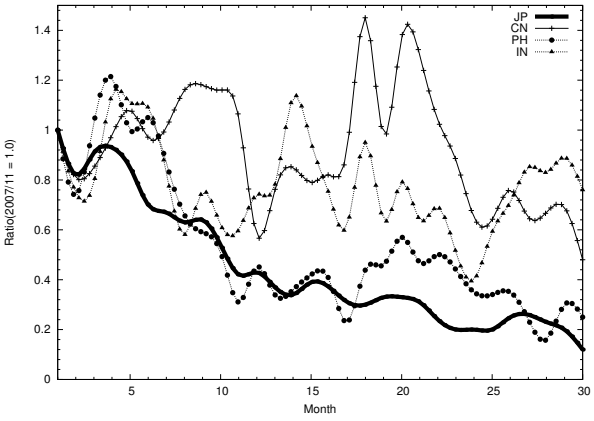

図 2: 2007 年 11 月からのプル型転送ノード数

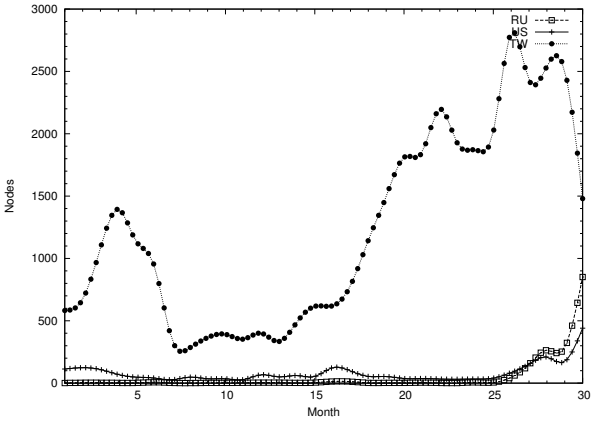

図 3: 急激に増加したプル型転送ノード数

全体的なイベントの減少と同様にプル型転送の総数は 2007 年に比べて減少傾向にある．ただし，国毎に着目すると国ごとに 3 つの傾向が見られる．

1. **減少傾向**: 期間中，減少している傾向．今回の結果では，日本，フィリピンと台湾が当てはまる．これらの国では 2007 年に比べて利用者の意識向上や，国が主導となったボット対策などが施行されていると推定できる．台湾に

関しては一時的に大きな減少と増加が確認できたが，現在は 2008 年前半とほぼ同水準まで減少している．

2. **変化なし**: 期間中，若干の変動はあるが，大きくノード数の変化がない傾向．今回の結果では中国とインドが当てはまる．2007 年の時点で十分にインターネットが発展しており，特に大きな対策などが施行されていないと推定できる．

3. **2009 年末からの急速な増加傾向**: 2007 年から 2009 年前半までは低いノード数であったが，2009 年から極めて大きな増加が見られた傾向．今回の結果ではロシアと米国でこの傾向が見られた．この傾向については第 3.2 節にて詳解に述べる．

3.2 プル型攻撃の 2009 年後半の急速な増加

第 3.1 節の考察において，ロシアと米国からのプル型ノード数は 2009 年半ばから急激に増加している傾向が分かった．この様子は 2010 年 4 月の時点でも継続しているため，2010 年 4 月時点でプル型ノード数が多い上位 15ヶ国について同様の傾向が存在するかを確認した．その結果，米国，カナダ，ドイツ，スペイン，フランス，イギリス，ポーランド，ルーマニア，ロシアにおいて同様の傾向が確認された．これらの国の地域性を考えると，ヨーロッパ及び北米においてこの傾向が見られる．多くの場合，地理的に隣接した国のノードは所属する IP アドレスの管轄組織が同じであるため，アドレスブロックが隣接しており，これらの地域で集中的なボット感染の拡大が発生したと推測できる．なお，台湾もノード数は 2009 年半ばから急激に増加している様子が観測されたが，2010 年に入ってから大きく減少しており，現在も継続している事象ではないと考える．

3.3 プッシュ型転送の傾向

図 4 に，各国の 2007 年 11 月のノード数を 1 とした際の，以降の月におけるノード数の割合を示した．この図では，日本，米国，韓国，フランスと台湾について示した．プッシュ型転送は期間全体を通して減衰傾向にあり，ますます減少している転送方法であると推測できる．この結果は，2007 年 11 月と 2010 年 4 月のプッシュ型転送のノード数が多い 5 国において継続して観測されている傾向であることから，新しい攻撃で使われ始めている転送手法でないと考えられる．プッシュ型転送は NAPT 機能を導入しているネットワークでは十分に動作せず，IPv4 アドレス枯渇が騒がれている現在において，攻撃者が本手法を積極的に採用する利点はますます失われている．Web 閲覧やメール機能を主に利用する一般利用者の環境下では，外部から利用者のノードに対して開始される通信は少ないため，活動を発見される可能性が向上する背景から今後も減少する転送方法であるといえる．

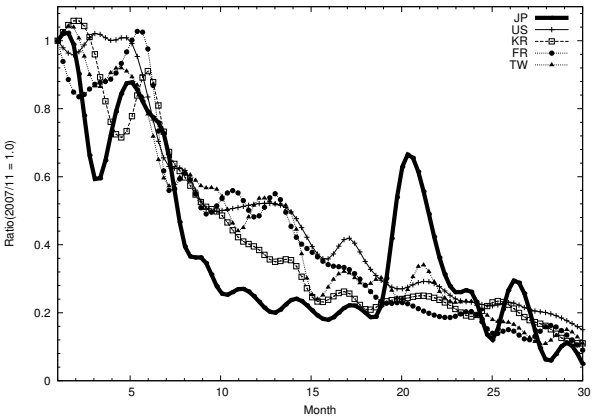

図 4: 2007 年 11 月からのプッシュ型転送ノード数

4 時差を考慮したマルウェア活動の分析

筆者は，2008 年の調査で 1 日における時間ごとのノード数を攻撃元の国ごとに分類すると国の活動と関係した分布となると仮説を立て，日本においてその傾向が見られることを示した．第 3.2 節の考察より，プッシュ型転送は大きく縮退していると判断し，プル型のみに着目する．

本分析には 2 年前の分析と同様の手法を利用する．研究用データセットの期間中おける h 時から $h+1$ 時の平均的なマルウェアの配布元ノードの数を $E_{(h)}$ とする．ここで平均的な 1 日における h 時から $h+1$ 時の一意なマルウェアの配布元ノード数の割合を $TP_{(h)}$ を

$$TP_{(h)} = \frac{E_{(h)}}{\sum_{n=0}^{23} E_{(n)}}$$

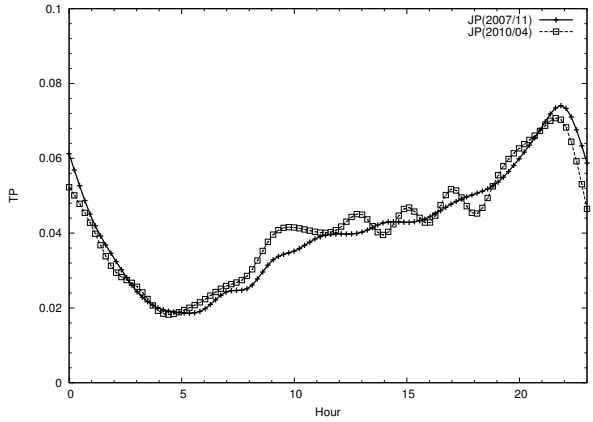

図 5: 2007 年 11 月と 2010 年 4 月の日本における時間毎のノード数の割合

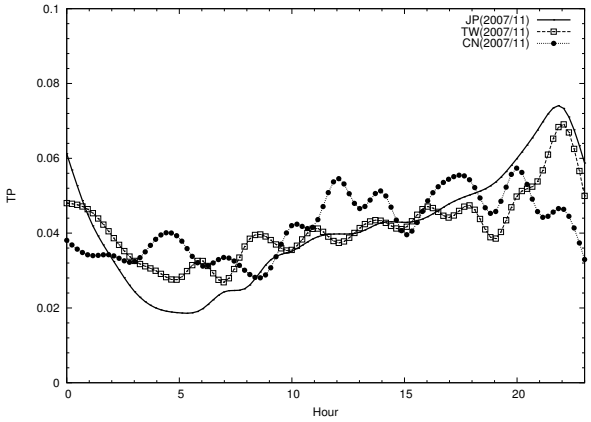

図 6: 2007 年 11 月の各国における時間毎のノード数の割合

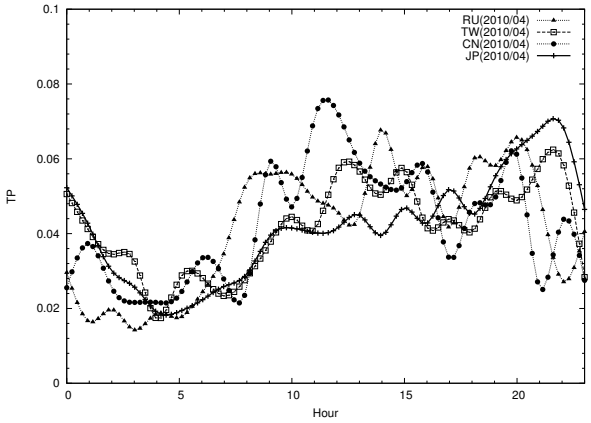

図 7: 2010 年 4 月の各国における時間毎のノード数割合 (RU は時差分について調整済み)

とする．本節では 2007 年 11 月時点と 2010 年 4 月時点でノード数の多かったプル型転送とプッシュ型転送の上位 3ヶ国である，日本，中国，台湾とロシアの 4ヶ国について考察する．

まず，日本における 2007 年 11 月と 2010 年 4 月の時間毎のノード数割合について，図 5 に示す．日本については以前の調査と同様に日本のインターネットエクスチェンジにおけるトラフィック [4] と関係した傾向が存在し，現在もインターネット上の活性なノード数と密接な関係があると考えられる．

さらに，2007 年 11 月と 2010 年 4 月時点の各国における時間毎のノード数割合を，それぞれ図 6 と図 7 に示した．ただし，ロシアについては 2007 年 11 月時点でのプル型件数が極めて少量であり，描画が困難であったため，2007 年 11 月のグラフにロシアは描画していない．また，2010 年 4 月のグラフ描画に際して，ロシアについては描画の基となる情報を日本とモスクワの時差である 6 時間ずらして描画した．これは国の活動とその国に所属する転送ノード数の数に関係があると仮定しているためである．

2007 年 11 月の時点で台湾と中国については，深夜時間帯より日勤帯のほうが若干ノードの割合は多かったが，日本のトラフィックと比較すると，密接な関係があるとは考えられなかった．

2010 年 4 月の時点においては各国において日勤帯と深夜時間帯の差は大きくなり，トラフィックと関係があるようにみえる．特に台湾については日本に近い分布となっている．これらについて，日本と異なる特徴は 2 つある．まず，各国において，昼付近の割合が日本に比べて高い．また，中国とロシアに関しては日本のトラフィック及び本グラフでの割合がピークを迎える夜間に割合が急激に落ちている．特に後者の結果から，中国とロシアでは，日勤帯に用いられるノードが多くマルウェアに感染している可能性が考えられる．あるいは，日本において夜間のみ活性であるノードが非常に多いと考えられる．

5 平日と休日のマルウェア活動の分析

第 4 節において，ある国からの一日のノード数の分布は，その国のタイムゾーンと関係があることを示した．この関係にはインターネットに接続されているノードの数が深く関わっている．インターネットが社会の重要なインフラとなった背景から，これらのノードは家庭を始めとして，学校や企業でも用いられている．社会の一週間における就業状況などを考えると，学校や企業のノード多くは平日に比べて休日は不活性であると考えられ，休日は昼間から

活性な家庭のノードが多いといえる．

本節では一日における時間毎のノード数の分布傾向が平日と休日で同一のものであるかを分析する．本節においても，プッシュ型転送は大きく縮退していると判断し，プル型のみに着目する．これまでに日本における 2010 年 4 月の各曜日毎の各時間に何台の唯一なノード数が見られたかを図 8 に示す．ここでは，平日と休日の差を示すために，平日の月曜日と休日である土曜日と日曜日のノード数を示した．第 4 節では各時間のノード数の割合を示したが，本図では，ノード数を示す．

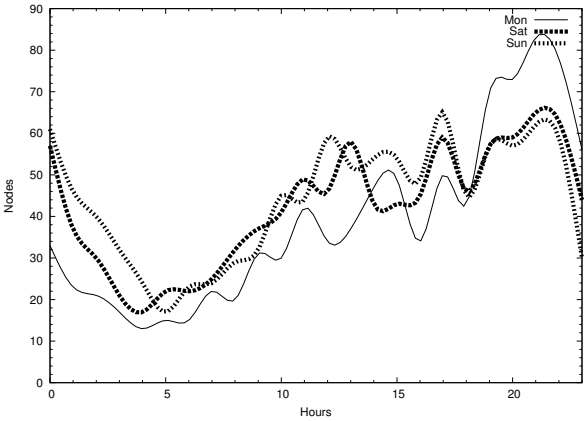

図 8: 日本における平日と休日の時間毎のプル型転送ノード数の割合

平日と休日のトラフィック量を比べると，第 4 節で述べた一日のノード数と同様の特徴を持っている．まず，早朝及び日勤帯において，休日よりも平日の方がノード数が多く観測されている．逆に夕方から深夜にかけては平日が休日に対してノード数が多い．この背景として，ブロードバンド環境の整備が進んだ背景などから，休日は昼間から夜遅くまで家庭でノードを利用していることが考えられる．この 2 つの事象は，一般的なインターネット利用者のトラフィック量に類似しており，日本国内のマルウェア転送のノード数は家庭の活性なノードと密接な関係があると考えられる．

6 まとめ

本稿は攻撃元ホストの帰属する国に注目した長期にわたるハニーポットログの分析を行った．以前の解析では日本の攻撃元ホストは利用者の生活トラフィックに類似したイベント分布であることを示したが，日本以外の国においてはこの傾向は見られなかった．今回の解析では，この傾向は日本内で継続しており，日本以外でも確認されることを示した．

また，分析を曜日毎に試行したところ，平日と休日の転送ノード数の傾向には違いがあることが分かった．この違いは一般的なインターネット利用者のトラフィック量の変化と似ていることから，家庭の活性なノード数と日本国内のマルウェア転送のノード数には関係があると考えられる．本解析では月単位で曜日毎のノード数を集約して分析したが，今後はこの情報と特定の日に絞った分析を実施することが考えられる．これは，一般的なユーザのトラフィック量はオリンピックなどの国民的なイベントの際には大きく変化ことが知られており，これらのトラフィック変化と転送ノード数の変化を比較することで，家庭のノードがマルウェア転送のノード数に占める割合を特定する手掛かりと考えられる．

本稿では，マルウェアの転送イベントのログと，活性なインターネット上のノード数に関係があることを示した．また，約 3 年間にわたるログを分析することで，これらの事象が継続的な関係であると分かった．しかし，第 3 節で示したとおり，本データセットに含まれる転送元ノード数は当初から大きく減少している．この原因には，攻撃者の隠蔽性がより高くなったか，本データセットの情報源となるハニーポットの生息が攻撃者らに広く知れ渡ったなどの理由が考えられる．今後の調査継続に際しては，データセットにて取得したイベントの他ハニーポットとの比較や検体の分析を進め，このデータセットを継続して分析していくことが広く有用であることを示す必要があると考えられる．

参考文献

[1] 金井瑛, 水谷正慶, 武田圭史, 村井純. マルウェアの転送ログを利用したボットの活動分析. **マルウェア対策研究人材育成ワークショップ** *2008* **論文集**, 2008.

[2] 畑田 充弘, 中津留 勇, 秋山 満昭, 三輪 信介. マルウェア対策のための研究用データセット MWS 2010 Datasets. *MWS2010*, October 2010.

[3] MaxMind. GeoIP, 2008. `http://www.maxmind.com/app/ip-location`.

[4] INTERNET MULTIFEED CO. JPNAP 東京 I サービス - トラフィック, Sep 2010. `http://www.mfeed.ad.jp/jpnap-tokyo-i/traffic.html`.