実トラフィックデータに基いた送信先アドレスによる関連付け可否性の評価

三浦 正明† 蓑原 隆‡
拓殖大学大学院工学研究科† 拓殖大学工学部‡

1 はじめに

近年インターネットの利用者の増加に伴いプライバシー保護の観念が強くなってきている．通信のプライバシー保護において暗号化技術を利用することでコンテンツ情報を隠すことが可能だが，アドレス情報により通信の関連付けが行えてしまう問題がある．送信元アドレスをランダムに変更することで関連付けを防ぐ方法[1]が提案されている．しかしアドレス情報を変更していない送信先アドレスから通信の関連付けが行える可能性がある．利用者が限定されるような送信先ノードがあった場合，送信先ノードにアクセスする送信元ノードは同一人物しか存在しないという事から関連付けを行うことが出来る．その為送信先アドレスについてもアドレスを変更する等の対策が必要である．現在送信先ノードのアドレスを変更する手法[2][3]が提案されている．この手法は中継ノードを利用することでアドレス変換を実現しており，送信元アドレスの変更に比べコストが増加する問題がある．全ての通信において送信先アドレスの変更を行うのでは無く，送信先アドレスにより関連付けを行えてしまうようなノードについてはこれらの技術を使用し対応する使い方が好ましい．しかし実際にアドレス変更を必要とする通信がどの程度あるのかは分かっていない．

そこで本研究では実際のトラフィックデータについて送信先アドレスにより関連付けが行えてしまう特徴的な通信を調査した結果から，送信先アドレスによる関連付け対策が必要であることを示す．

2 特徴的な送信先の定義

送信先アドレスによる関連付け対策が必要な場合として，頻繁にアクセスを行うような送信先が考えられる．利用頻度の極端に低い送信先について関連付けを行われても得られる情報は少ないが，利用頻度の高い送信先の場合毎回関連付けを行われるとそれだけ情報を相手に与えてしまうことになる．

関連付けが行える送信先というのは利用者が限定される送信先である為，単一の送信元からのみアクセスを受ける送信先が考えられる．また，送信元に関しては複数の送信先と通信を行っている場合が考えられるが，本研究では送信元についても単一の送信先のみと通信を行っている送信元を対象とする．

以上の事から本研究では送信先アドレスにより関連付けが行える特徴的な送信先を次のように定義する．「利用頻度の高い送信先ノードであり，通信は単一の送受信アドレスからなる．」

3 トラフィックデータの扱い

トラフィックデータを収集する場所に左右されるが，ネットワーク全体のトラフィックを収集した場合データ量は膨大になり，トラフィック処理が困難になる為，本研究ではデータ量を制限して扱う．例えば，数日分のトラフィックデータがあった場合，数分間のトラフィックデータに分割しそれを1つのトラフィックデータとして扱う．

定義を満たす特徴的な送信先を抽出する為の方法を以下で述べる．利用頻度の高い通信を行っている条件は2つの別々の時間のトラフィックデータを比較することで行う．次にそれぞれのトラフィックデータについて一対一の通信を抽出する為に送信先について単一の通信のみ受けている送信先に絞り込む．この状態では送信元アドレスと送信先アドレスについて多対一の関係になっている通信が含まれる為，複数の送信元からアクセスを受けている通信の組み合わせを除外する．

4 実験による分析

特徴的な送信先を抽出した結果に対しての分析を行った．また，抽出された特徴的な送信先に対し，長期に渡って複数の送信元からアクセスを受けない送信先がどの程度存在するか調査を行った．

4.1 実験に用いる実トラフィックデータ

本研究ではより一般的な通信を得る為にトラフィックが集中するバックボーンでのトラフィックデータを対象とする．実験ではWIDEネットワークのバックボーンを対象とし，WIDEバックボーンにおけるトラフィックデータ取得にはMAWI WorkingGroup[4]で提供されて

Evaluation of linkability based on read observed Internet traffic data.
†Graduate School of Engineering, Takushoku University
‡Department of computer Science, Takushoku University

いるトラフィックデータレポジトリを使用する．MAWI WG で提供されているトラフィックデータは 15 分毎に細分化されている．本研究で使用したデータは 2008 年 3 月 18 日，19 日，20 日 (リンク帯域 150Mbps) のトラフィックデータである．

4.2 特徴的な送信先抽出

2008 年 3 月 18 日と 3 月 19 日のデータを元に実験を行い，特徴的な送信先が抽出される割合を出した．トラフィックデータの比較の組み合わせでは，時間の近いデータセットの比較と時間の離れたデータセットの比較についてそれぞれ行った．近いデータセットでは連続的にアクセスしている通信や，たまたまある時間にアクセスした通信が含まれることが考えられる．離れたデータセット比較ではそれらの通信を除外することが出来ると考えられる．抽出結果を 1 に示す．

表 1: 特徴的な送信先抽出結果

	全フロー数	特徴的送信先数	抽出割合
短い比較	923012	6590	0.74 %
離れた比較	862012	3311	0.39 %

それぞれの割合の平均値は，離れたデータセットでは 0.39 %に対し，近いデータセットは 0.74 %という結果が出た．この結果から時間の近いデータセット比較に含まれていた，たまたまアクセスした通信や継続的にアクセスした通信等のノイズが，離れたデータセット比較では除外されたと考えることが出来る．

4.3 単一のアクセスのみ受ける送信先

特徴的な送信先の中から長期のトラフィックにおいて単一の送信元ノードからのみアクセスを受けるような送信先ノードが存在するか調査を行った．

3 月 18 日から 3 月 20 日の 3 日間のトラフィックデータを長期のトラフィックデータと位置づた．使用したデータは，時間の近い比較は 18 日の 12:00 と 12:30 の組み合わせ．時間の離れた比較は 18 日の 12:00 と 19 日の 12:30 の組み合わせである．結果を表 2 に示す．

時間の近いデータセット比較と離れたデータセット比較双方において，長期に渡り単一の送信元からのみアクセスを受ける送信先を抽出することが出来た．通信のサービスの内訳を表 3 に示す．

結果から ssh や pop3s 等の暗号化されたサービスは，長期に渡り単一のアクセスしか受けず，また他人に見られたくない通信を行っていることから非常に特徴的であり，特に対策が必要であると考えられる．

表 2: 単一の送信元からのみアクセスを受ける送信先数

	特徴的な送信先	関連付け可能な送信先
短いデータセット	187	18
離れたデータセット	80	15

表 3: 通信の内訳

ポート	近いデータセット	離れたデータセット
22	1	1
80	9	8
83	1	1
110	1	0
443	3	4
995	1	0
8080	1	0
19513	1	0
50084	0	1

5 まとめ

送信先アドレスによる関連付け対策が必要な送信先を，使用トラフィックデータに制限があった為限定的ではあるが抽出することが出来た．使用したトラフィックデータが限定的な為，他の範囲について実験を行うとより多く対策が必要な送信先が出力される可能性がある．

参考文献

[1] T.Narten and R.Draves: “Privacy extensions for stateless address autoconfiguration in IPv6”, RFC3041(2004).

[2] Sakurai,A., Minohara,T, Sato, R. and Mizutani, K.: One-Time Receiver Address in IPv6 for Protecting Unlinkability, Proc. ASIAN 2007, pp.240-246.

[3] 市川 隆浩，坂野 あゆみ，寺岡 文男: ”匿名性のある IPv6 モビリティ通信プロトコル”，第 8 回インターネットテクノロジーワークショップ，2007，pp36-44.

[4] K.Cho,K.Mitsuya,and A.Kato. Traffic Data Reposity at the WIDE Project. (2000)