攻撃元特定のためのトラヒックマトリクス推定手法

Traffic matrix estimation for identification of attack sources

大下　裕一 [1]　　阿多　信吾 [2]　　村田　正幸 [1]
Yuichi Ohsita　　Shingo Ata　　Masayuki Murata

大阪大学 大学院情報科学研究科 [1]
Graduate School of Information Science and Technology, Osaka University
大阪市立大学　大学院工学研究科 [2]
Graduate School of Engineering, Osaka City University

1　はじめに

近年インターネットにおいて，悪意を持った第三者がサービスを提供する計算機に攻撃を行い，一般ユーザの利用を妨げる分散サービス拒否 (Distributed Denial of Service; DDoS) 攻撃が深刻な問題となっている．

DDoS 攻撃への対策として，攻撃元を特定し，攻撃元でフィルタリング等を行うことが効果的である．しかし，[1, 2] などの既存の攻撃元特定手法は，ルータの大規模な置き換えが必要となる，攻撃元と正常な通信相手の区別が難しい等の問題があり，広く利用されていないのが現状である．

DDoS 攻撃では，攻撃者は被害者に対し，大量のパケットを送ることでその通信を阻害する．もし，各送信元・宛先間のトラヒックを観測することができれば，トラヒック急増の原因となるトラヒックの送信元を特定することができることから，攻撃元の特定および対策が可能となる．また，トラヒックの増加量に基づく攻撃元の特定は，トラヒックを急増させない正常な通信相手を攻撃元と誤検出することはないと考えられる．しかし，送信元・宛先間のトラヒックを正確に観測するためには，ネットワークのエッジルータにおいて送信元・宛先ごとのフロー統計情報を観測する必要があり，実現が容易ではない．

フロー統計情報の観測を必要としない手法として，送信元宛先間のトラヒック量 (トラヒックマトリクス) をリンクのトラヒック量から推定する手法 [3] が提案されている．リンクのトラヒック量は SNMP をもちいて既存のルータから収集可能であるため，この手法は既存のネットワークにも容易に適用可能である．しかし，この手法では通常トラヒックのように，宛先が分散しているトラヒックを仮定しており，攻撃トラヒックのような、ある一つの宛先に偏ったトラヒックを精度よく推定することができない．

そこで本稿では，トラヒックの増加分を正確に把握するために，トラヒック増加量に注目したトラヒックマトリクス推定手法を提案する．トラヒックの総量をもとに推定する既存のトラヒックマトリクス推定手法ではトラヒックの変化を推定することは難しいが，提案手法では，トラヒック増加量をもとに推定を行うため、攻撃検出に有効な情報となるトラヒック増加量を正確に推定することができる．

以降，2 で提案手法の概要について述べる．そして，3 で提案手法の評価を行い，被害者宛トラヒックの送信元ごとの増加量を推定可能であることを確認する．最後に，4 でまとめと今後の課題について述べる．

2　トラヒック増加量推定方法

X をリンクのトラヒック量，A をネットワークのルーティング情報をあらわす行列，T を各送信元・宛先間のトラヒック量とすると，以下の関係が成り立つ．

$$X = AT \tag{1}$$

式 (1) のうち，X，A は SNMP や OSPF 等のルーティング情報を集めることにより，取得可能である．そこで，式 (1) が成り立つ T を求めることができれば，トラヒックマトリクスを推定することができる．しかし，この関係が成り立つ T は多数存在するため，トラヒックの特性を仮定する必要がある．[3] では，gravity model を用いてトラヒックマトリクスの推定を行うことを提案している．gravity model では，送信元・宛先間のトラヒックの流量は，その入り口となるリンクのトラヒック量，出口となるリンクのトラヒック量に比例すると仮定している．しかし，gravity model を用いると，攻撃による出口のトラヒックの増加は，送信元のトラヒック量にしたがって分配されてしまい，トラヒック増加の原因となる送信元を把握することが難しい．そこで，トラヒックの増加を正確に把握するために，各リンクごとのトラヒックの増分をもとに，送信元・宛先間のトラヒック増加量の推定を行う．

時間をスロット化して考え，n 番目の時間に観測された各リンクのトラヒック量を要素とする行列を L_n とし，各リンクの正常時のトラヒック量を要素とする行列 $\bar{L}_n$ とする．以下のように各リンクのトラヒックの正常時と比べた増加量を要素とする行列 G_n を定義する．

$$G_n = L_n - \bar{L}_n \tag{2}$$

そして，観測された G_n をもとに，gravity model をもとにした式 (3) を用いて，観測時刻 n における i から j へのトラヒックの増加量 $f'_{i,j,n}$ を推定する．ここで，$g^{\mathrm{in}}_{i,n}$ は，入り口リンク i におけるトラヒックの増加量，$g^{\mathrm{out}}_{j,n}$ は出口 j におけるトラヒックの増加量とする．

$$f'_{i,j,n} = \begin{cases} g^{\mathrm{in}}_{i,n} \dfrac{g^{\mathrm{out}}_{j,n}(j)}{\sum_{\{k:(g^{\mathrm{out}}_{k,n}>0)\}} g^{\mathrm{out}}_{k,n}} & (g^{\mathrm{in}}_{i,n} > 0, g^{\mathrm{out}}_{j,n} > 0) \\ -\left| g^{\mathrm{in}}_{i,n} \dfrac{g^{\mathrm{out}}_{j,n}(j)}{\sum_{\{k:(g^{\mathrm{out}}_{k,n}<0)\}} g^{\mathrm{out}}_{k,n}} \right| & (g^{\mathrm{in}}_{i,n} < 0, g^{\mathrm{out}}_{j,n} < 0) \\ 0 & (\text{others}) \end{cases} \tag{3}$$

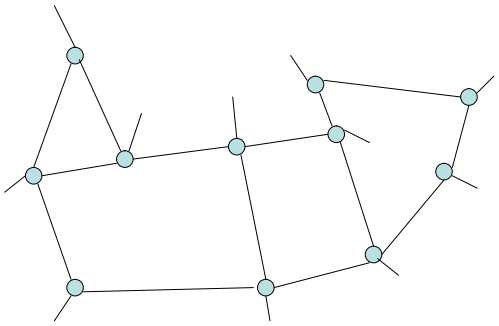


図 1　Abilene のバックボーントポロジ

しかし，$f'_{i,j,n}$ は，出入り口以外のリンクの観測結果は反映されていない．そこで，$f'_{i,j,n}$ に対して，式 (4) のようにして，途中のリンクの観測結果を反映させ，最終的なトラヒック増加量の推定結果を得る．ここで，A^{-1} は A の擬似逆行列とし，F'_n は $f'_{i,j,n}$ を要素とする行列とする．

$$F_n = F'_n + A^{-1}(G_n - AF'_n) \tag{4}$$

最後に，この推定に必要な正常時のトラヒック量の定め方について述べる．正常時のトラヒック量 $\bar{L}_n$ としては，$\bar{L}_{n+1} = \alpha L_n + (1-\alpha)\bar{L}_n \quad (0 < \alpha < 1)$ のように過去の観測されたトラヒック量の重みつき平均をすることが考えられる．しかし，一時的なトラヒックの急増が $\bar{L}$ に反映されてしまうと，$\bar{L}$ が高い値となってしまい，その後のトラヒック急増の検出に悪影響を与える．そこで，以下のように，トラヒックマトリクス推定の結果をもとに，トラヒック急増の影響を除外して，$\bar{L}_n$ を定める．

まず，式 (5) のように，$\hat{f}_{i,j,n}$ を定める．ここで，F_n の i から j 宛てに対応する要素を $f_{i,j,k}$ とあらわし，$f_{i,j,n}$ の最近 N 個の平均，標準偏差をそれぞれ $\mu_{i,j}$，$\sigma_{i,j}$ とあらわす．β は，トラヒックの急増を判断する閾値を決めるパラメータである．

$$\hat{f}_{n,k} = \begin{cases} f_{i,j,n} & (f_{i,j,n} < \mu_{i,j} + \beta\sigma_{i,j}) \\ 0 & (\text{others}) \end{cases} \tag{5}$$

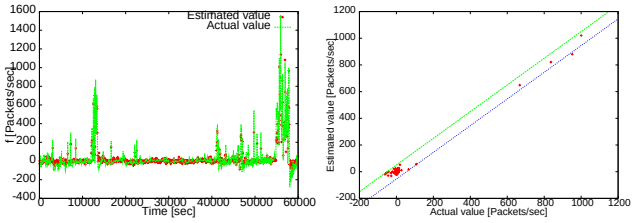
そして，以下のように，$\bar{L}_{n+1}$ を定める．ここで，$\bar{F}_n$ は $\hat{f}_{i,j,n}$ を要素とする行列とする．

$$\bar{L}_{n+1} = \alpha(\bar{L}_n + A\hat{F}_n) + (1-\alpha)\bar{L}_n \tag{6}$$

これにより，トラヒックが急増した箇所の影響をうけず，正常時のトラヒック量を定めることができる．

3　提案手法の評価

提案手法により，トラヒックの増加を正確に推定できることをシミュレーションにより確認する．本稿では，シミュレーショントポロジとして，図 1 を仮定した．送信元・宛先間のトラヒック量は，次のように，大阪大学のゲートウェイでの観測結果をもとにした値を定めた．まず，大阪大学のゲートウェイでパケットの送信元アドレスの上位 16 ビットごとの到着レートを 60 秒間隔で取得した．図 1 では送信元・宛先の組は 110 通りである．そこで，一定個数の送信元アドレスをグループとしてまとめ，そのグループごとにトラヒック量の合計を求めていくことにより，110 通りのトラヒック量の値を定めた．また，正常なトラヒック量を定める際のパラメータとして $\beta = 3$ とした．



(a) 時間変化　　(b) 推定精度

図 2　増加量推定の結果

図 2(a) は，ある送信元・宛先間のトラヒックの増加量とその推定結果の時間変化を表したものである．これにより，トラヒックが急激に上昇した時間帯には，増加量の推定により，その上昇を検出することができることが分かる．

図 2(b) は，4 地点から同一地点へ攻撃として 1000, 830, 660, 500 Packes/sec のトラヒックを加え，その時間帯の送信元・宛先間のトラヒック増加量の推定結果を示したものである．横軸は実際の値，縦軸は推定結果であり，線が示す範囲は $x \pm 50$ である．これより，大規模な攻撃を行った場合であっても，トラヒック増加量の推定精度は ±50 Packets/sec に収まっており，被害者宛トラヒック急増の原因となるトラヒックの送信元を正確に把握できることが分かる．

4　まとめと今後の課題

本稿では，攻撃元を特定するために有効な情報となる，送信元ごとのトラヒック増加量を，リンクごとのトラヒック量の情報から推定する手法を提案した．シミュレーションにより，提案手法がトラヒックの急増を正確に推定することができることを示した．

今後の課題としては，得られた送信元ごとのトラヒック増加量をもとに攻撃元を特定する手法があげられる．

参考文献

[1] S. Wu, L. Zhang, D. Massey, and A. Mankin., "On design and evaluation of intention-driven ICMP traceback," in *Proceedings of IEEE International Conference on Computer Communications and Networks*, Apr. 2001.

[2] C. Partridge, L. A. Sanchez, C. E. Jones, F. Tchakountio, B. Schwartz, S. T. Kent, and W. T. Strayer, "Single-packet IP traceback," in *Proceedings of the ACM SIGCOMM 2001*, Aug. 2001.

[3] Y. Zhang, M. Roughan, N. Duffield, and A. Greenberg, "Fast accurate computation of large-scale IP traffic matrices from link loads," in *Proceedings of ACM SIGMETRICS*, June 2003.