異常パケットトレースのアドレス局所性に関する解析*

福田 健介† 国立情報学研究所
廣津 登志夫‡ 豊橋技術科学大学
明石 修§ NTT 未来ねっと研究所
栗原 聡¶ 大阪大学
菅原 俊治‖ 早稲田大学

1 はじめに

世界中の人々が通信インフラとしてインターネットを利用するようになり，ネットワークに対する高い信頼性とセキュリティが必要となっている．インターネットユーザが一部の研究者のみであった時代には，ネットワークは性善説によって構築・運用されてきたが，現在では，日常的にホストやネットワークへの攻撃が行われており，それら攻撃に対する検知・防御が求められている．現在，ワーム，DDoS 等のネットワークを介したホストに対する攻撃の検知には，静的方法と動的方法が知られている．前者 (Darknet) は，ネットワーク中に置いた測定ホストに対するパケットを収集し，攻撃パターンを調査する [1]．後者は (Honeypot)，測定ホストが脆弱性を残した状態で設置され，攻撃パケットに反応することで，攻撃者の攻撃パターンを詳細に特定することが可能となる [2]．

しかしながら，既存の Darknet 攻撃検知手法には 2 つの問題がある．一つは，精度を高くするために大規模な監視アドレスブロックが必要である点である．実際，いくつかの研究では，/8 の未使用アドレスブロックを用いて解析を行っている [1, 3] が，利用可能な IPv4 アドレスは年々少なくなっており，大規模な監視アドレスブロックを今後も使用できる可能性は低い．二つ目の問題点は，監視アドレスブロックを大きく取ることで，収集トラフィック量が増大することである．

これらの問題点に対して，我々は，未使用の小規模のアドレス空間を多数分散配置することで，攻撃パケットの攻撃パターンを早期検知するためのフレームワークを提案している [4]．実現のポイントは，どのように分散配置された小規模のアドレス群から，攻撃パターンを検知するかにある．本研究では，まず，大規模のアドレスブロックへのパケットパターンを解析し，それらを複数の小規模アドレス群として見た場合に，それぞれのパケットパターンがどの程度類似しているかを調査する．

2 収集データ

2006 年 10 月～2007 年 5 月にかけて，国内に設置した “/18” (16384 ホスト相当) の仮想ネットワークに到着したパケットを tcpdump コマンドを用いてキャプチャしたものを基礎データとして使用した．1 日辺りの平均到着パケット数は 130 万，ユニークな IP アドレスの平均は 8 万である．また，データを解析するにあたって，各パケットを以下のトラフィッククラスに分類した．

- tcp-syn: syn フラグのみのついた TCP パケット．これらの大部分はワーム・ウィルスに相当する．
- tcp-synack: syn および ack フラグのついた TCP パケット．これらはソースアドレスが偽造された (D)DoS パケットの跳ね返り (バックスキャッタ) と考えられる
- udp: UDP パケット．多くがワーム・ウィルスである．

これらのデータをさらにポート別に分類する方法もあるが，必要データ処理量が大きくなるため，今回は，よりマクロな挙動に着目することにした．

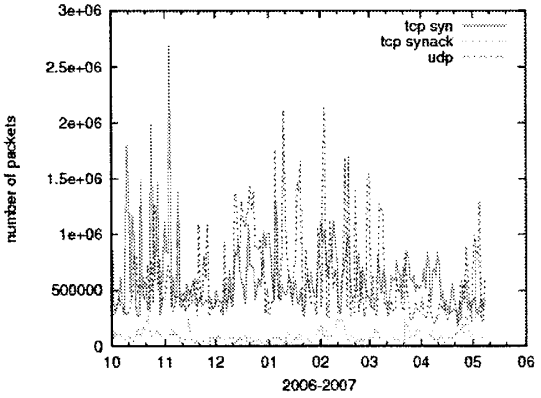

Figure 1: 攻撃トラフィックの変動

図 1 に観測期間中の各トラフィッククラスのトラフィック時系列を示す (ビンサイズは 1 日)．図より，日による変動が大きく 10^6 のオーダのパケット数に達すること，tcp, tcp-syn, udp ではその時間的な変動が大きく異なることがわかる．

3 解析方法および結果

トラフィッククラスに分離したパケットトレースをアドレスブロック単位 n ごとに分離し，それぞれの単位時間あたりの到着パケット数の時系列 T_ℓ^n を作る．そして ℓ 番目の時系列と m 番目の時系列の相互相関係数を計算する．

$$C(T_\ell, T_m) = \frac{\sum (T_\ell(t_i) - E[T_\ell(t)])(T_m(t_i) - E[T_m(t)])}{\sqrt{V[T_\ell(t)]}\sqrt{V[T_m(t)]}}, \quad (1)$$

ここで，$E[\cdot], V[\cdot]$ はそれぞれ時系列の平均および分散を表す．また，C は $-1 \le C \le 1$ の範囲を取り，$C = 0$ では 2 つの時系列は無相関，$0 < C \le 1$ では正の相関，$-1 \le C < 0$ は負の相関を表す．C が 1 に近づくにつれ，二つの時系列の揺らぎ方に類似性が大きくなり，-1 に近づくにつれ反対の振る舞いをすることになる．

図 2 に $n = /24$ の分割例を示す．元の観測空間は $n = /18$ であるから，64 個の/24 の空間に分割され，各小規模ブロックは 256 アドレスを監視スペースとして持つことになる．それぞれの時系列間の距離 $D = |\ell - m|$ を変化させた際の相関係数 C_D の減衰の仕方により，どの程度監視ブロックを離れて設置すべきかを判断できる．同様に，アドレスブロック単位 n もまた，分散化する際の監視ブロックサイズの決定に必要なパラメータとなる．なお，今回の解析にあたって，時系列のビンサイズは 1 分，時系列の長さは 2006 年 10 月-2007 年 5 月とした．

*An analysis of spatial locality of Darknet traffic
†Kensuke Fukuda; National Institute of Informatics
‡Toshio Hirotsu; Toyohashi University of Technology
§Osamu Akashi; NTT Network Innovation Labs
¶Satoshi Kurihara; Osaka University
‖Toshiharu Sugawara; Waseda University

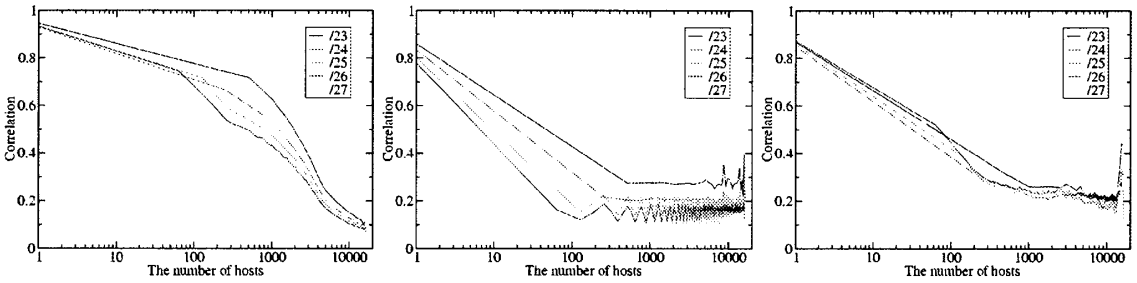

Figure 3: アドレス距離と相関係数: (a) tcp-syn, (b) tcp-synack, (c) udp

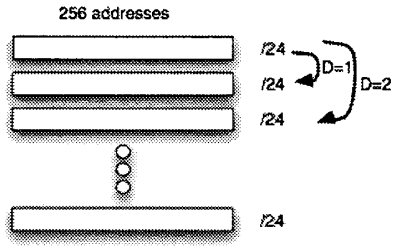

Figure 2: 時系列の分割

図3はnを/23から/27に変化させた際の相関係数の変化を表している(水平軸は距離Dではなく，アドレス数$(D \times n)$で規格化した)．各グラフでは，nが大きいほど相関係数が大きく，距離が大きくなるほど相関係数が小さくなる結果となっているが，トラフィッククラスによりいくつかの違いがある．

- tcp-syn: $n = /23$で隣接ブロックとの相関は0.75でありかなり大きな値である．それに対して，同一距離(512アドレス分)だけ離れていた場合に，$n = /27$のブロックサイズでは0.4程度に減衰するが，それでも相関構造が残る結果となった．また，$n = /25, /26, /27$の差は比較的小さく，監視ブロックサイズは比較的小さくできる可能性があることを示唆している．

- tcp-synack: どのブロックサイズにおいても相関係数の値は0.1-0.3と小さい．これは，バックスキャッタで利用される偽造されたアドレスに関連性がないことを表している．また，/25より小さいアドレス空間において，相関値が鋸上の振動をしていることがわかる．これは，/24で見た場合のアドレス空間の4オクテット目の前半(0-127)と後半(128-255)とでトラフィックの振る舞いが異なることを意味する．バックスキャッタで用いられるアドレススペース上の局所性に関しては[4]でも報告されている．

- udp: $n = /23$の相関係数の最大値は0.3程度(距離: 512アドレス)であり，あまり大きな値とはならない．しかし，短いアドレスブロック(例えば$n = /26$)では，アドレス距離が短い場合(64アドレス)に，より大きな相関の値が得られている．

udpとtcp-synはどちらも，ウィルスやワームのパケットがメインであることを考えると，この両者の相関値が大きく異なるのは非常に興味深い結果である．また，tcp-synackおよびudpでは，アドレス距離が1000以上($\approx /22$)離れた場合には，相互の時系列にはほとんど相関がなく，分散監視領域の設定には，少なくともこのサイズ未満にする必要がある．それに対して，tcp-synでは，対象領域を1000アドレスおきに設定し，かつ，ブロックサイズを$n = /27$程度にしても，十分，他のネットワークの状況を推定できるとことを示唆している．

上記の結果は，tcp-synでは小規模分散配置が有効であるが，udpやtcp-synでは有効とは言えないことを表している．しかし，実際に観測されているパケット数では，tcp-synによるものが多くを占めることから，tcp-synでのパケットを分散配置した小規模アドレスブロック群で監視できることは，我々が提案している手法の有効性を示していると言える．

4 結論

本研究では，小規模分散配置された監視アドレスブロックへの攻撃パケットを検知するために，大規模監視アドレスブロックを分割した小規模アドレスブロック間の攻撃パケットの類似性について解析を行った．その結果，TCPによるワーム等の攻撃パケットでは，小規模分割したアドレスブロックへの攻撃に正の相関があり，監視アドレスブロックサイズは/27程度に，監視ブロックの距離は/22程度でも十分有効であることがわかった．それに対して，UDPによるワーム等の攻撃パケットや，DDoSのバックスキャッターでは，アドレスブロック間の相関はほとんどなく，連続的に大きな空間を監視する必要があるとの知見が得られた．今後は，時間差を考慮した相関解析，ポート番号別の攻撃トラフィックの振る舞いを解析する予定である．

謝辞

この研究は科学研究費補助金特定領域研究「IT爆発情報基盤」の支援を受けている．

References

[1] Pang, R., Yegneswaran, V., Barford, P., Paxson, V. and Peterson, L.: Characteristics of Internet background radiation, *4th ACM SIGCOMM conference on Internet measurement (IMC'04)*, pp. 27–40 (2004).

[2] Honeypot project. http://www.honeynet.org/.

[3] Bailey, M., Cooke, E., Jahanian, F., Provos, N., Rosaen, K., and Watson, D.: Data Reduction for the Scalable Automated Analysis of Distributed Darknet Traffic, *5th ACM SIGCOMM/USENIX International Conference on Internet Measurement (IMC'05)* pp. 239–252 (2005).

[4] 廣津登志夫，福田健介，栗原聡，明石修，菅原俊治，"断片アドレスを用いた分散協調インターネット監視に関する一考察"，情報処理学会 OS研究会研究報告 (83), pp.39–45, 2007.

[5] Cooke, E., Myrick, A., Rusek, D., and Jahanian, F.: Resource-Aware Multi-Format Network Security Data Storage, *ACM SIGCOMM LSAD'06*, (2006)