DarkNet における遅延相関とアドレス間距離との位置関係の解析

大田 昌幸 + 杉本 周 † 菅原 俊治 † 福田 健介 ‡ 廣津 登志夫 *
+ 早稲田大学理工学部 † 早稲田大学基幹理工学研究科情報理工専攻 ‡ 国立情報学研究所 * 豊橋技術科学大学

1 序論

インターネット上のあらゆるサービスは DoS,Virus,Worm などの妨害攻撃に悩まされており、これら攻撃に対する検知・防御が求められている。攻撃検知手法の一つとして Darknet が挙げられる [1]。この手法では、実際にはホストが存在しない「巨大な」観測専用アドレス空間宛のパケットを収集し、解析を行うことで攻撃の情報を得る。しかし、全ての組織が「巨大な」アドレス空間を用意して観測することは不可能である。

そこで、各組織が割り当てられたネットワークのアドレス空間のうち一部を用いて攻撃の監視を行い、複数の組織が協調し全体として広いアドレス空間の監視を実現する「分散協調監視アーキテクチャ」を提案してきた [3]。そのためには、より小さな観測領域で効果的な観測が必要となる。本研究では時間相関とアドレス間の距離の関係を示し、分散協調アーキテクチャを利用するのに適切なアドレス空間の大きさを推定する。

2 解析手法

2.1 相互相関係数

単位時間毎に観測された到着パケット数の時系列を $T_l = \{t_0, t_1, t_2, ..., t_{n-1}\}$ とする。l 番目の時系列と m 番目の時系列の相互相関係数 $C_p(T_l, T_m)$ は次式で定義される。

$$C_p(T_l, T_m) \stackrel{\text{def}}{=} \frac{E[(T_l(t_i) - E[T_l(t)])(T_m(t_{i+p}) - E[T_m(t)])]}{\sqrt{V[T_l(t)]}\sqrt{V[T_m(t)]}}$$

ここで、E[*],V[*] はそれぞれ時系列の平均および分散を表わす。C は $-1 \leq C \leq 1$ の範囲を取り、$C = 0$ では、2 つの時系列は無相関、$0 < C \leq 1$ のときは正の相関、$-1 \leq C < 0$ のときは負の相関があることを表す。$|C|$ が 1 に近いほど 2 つの時系列データに強い相関がある。

2.2 遅延相関系列

相互相関係数を用いて、遅延相関系列を次のように定義する。n 個の観測点 $O_0, O_1, O_2, ..., O_{n-1}$ が等間隔かつ一直線に並んだ観測系 O を考える。各観測点において観測された時系列を $T_0, T_1, T_2, ...T_{n-1}$ とする。2 つの観測点 O_l, O_m の距離を $distance(O_l, O_m) = |l - m|$ と定義する。$distance(O_l, O_m) = d(0 \leq d \leq n-1)$ となるような全ての O_l, O_m のペアについて、ラグ $p(0 \leq p \leq n-1)$ における相互相関係数 $r_p^{b,d} = C_p(T_l, T_m)$ を計算する。ここで $b = min(l, m)$ であり、ラグ p は時間に関するラグである。相互相関係数 $r_p^{b,d}$ を d の昇順に並べて作った系列 $R_p^{b,d} = \{r_p^{b,0}, r_p^{b,1}, r_p^{b,2}, ..., r_p^{b,n-1}\}$ を「観測点 O_b における、ラグ p のときの遅延相関系列」と定義する。このとき O_b をベース観測点と呼ぶ。さらに、$R_p^{b,d}$ において、b についての相加平均をとった $r_p^d = \sum_b r_p^{b,d}/(n-1)$ を並べた系列 $R_p^d = \{r_p^d | 0 \leq d \leq n-1\}$ を「観測系 O 全体における、ラグ p のときの遅延相関系列」と定義する。

3 観測データへの適用

I 個の観測アドレスをもつ Darknet において観測された J 時間分のデータについて、この観測アドレス範囲を適当な集約アドレス数 *binIP* ずつ分割してできるサブブロックを観測点とし、各ブロックに到着した集約時間 *binTime* ごとのパケット数の合計を時系列データとして、遅延相関解析を行った。ここで、アドレス方向の距離 = *binIP* $\times (m - l)$、時間方向のラグ $p(0 \leq p \leq J/binTime)$ と定義する。ただし、$l \leq m$ とする。

3.1 遅延相関解析

次の 3 つの条件 (遅延事象条件) を満たす事象を考える。

1. 事象はある一定の伝搬速度 v を持つ
2. 事象は常に観測系 O に対して順方向に進む
3. 事象は途中で一時停止しない

このような事象が観測系 O で発生すると、事象の伝搬速度 v は次式で計算される。

$$v = \frac{l \times arg_d \max_{0 \leq d \leq n-1} r_p^d}{binTime \times p} \quad (address/sec)$$

なお、攻撃は一定の速度 v で到着するため、攻撃を観測する際は $binIP{:}binTime = (l \times arg_d \max r_p^d) : (binTime \times p)$ となるような *binIP* と *binTime* とを設定することで、より正確に解析を行うことができる。

3.2 時間相関とアドレス間距離との関係の解析

binIP と *binTime* との比を一定にしたままで、様々な $binIP(32, 64, 128, 256)$ を用いて遅延相関系列を観測する。その際各 $binIP(32, 64, 128, 256)$ ごとに、高い相関が現れるアドレス方向の最長距離を確認し比較を行う。本研究では、相互相関係数が 0.40 以上を「高い相関」と定義した。

3.3 解析対象データ

Darknet に到着した TCP-syn パケットを対象とした。TCP-syn で「ワーム、ウイルスによる自己増殖試行」を観測できるため、本研究では TCP-syn パケットのみを観測対象とする。

4 パケット到着パターン

解析に先立ち、今回使用したデータについてパケット到着パターンを確認した。パケット到着パターンとは、Darknet に到着した各パケットについて、縦軸に宛先 IP アドレス、横軸に到着時刻をとってプロットした散布図である [3]。パケット到着パターンを見ると、攻撃の特徴を視覚的に把握できる。

図 1 に、某日の 1 日分の TCP-syn のパケット到着パターン

Analysis of relationship between delay correlation and distance of IP addresses in Darknet

Masayuki OHTA†,Syu SUGIMOTO†, Toshiharu SUGAWARA†, Kensuke FUKUDA‡, Toshio Hirotsu*
†Waseda University ‡National Institute of Informatics *Toyohashi University of Technology

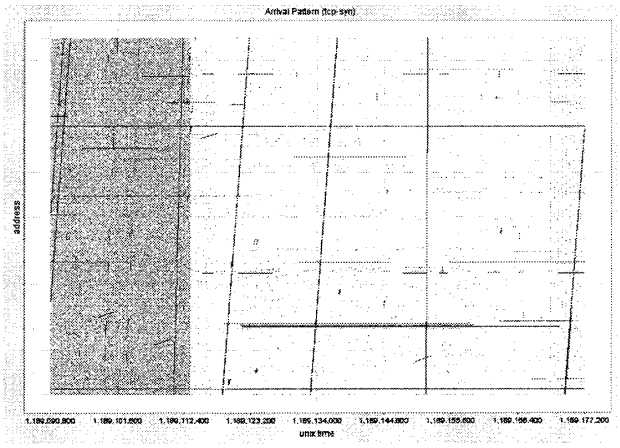

図 1　syn パケット到着の様子

を示す。縦方向、横方向それぞれにはっきりと筋が見える。縦方向の筋は監視アドレス空間全域を走査する攻撃が表れたもの、また横方向の筋は特定のアドレスに対して継続的に攻撃が行われた様子を表わしている。

5　解析結果

5.1　遅延相関系列

図 2,3 に某日の遅延相関系列を示した。図 2 はパラメータは $I = 4096, J = 6$ 時間,$binIP = 256, binTime = 1200$(sec) であり、図 3 はパラメータは $I = 4096, J = 6$ 時間,$binIP = 64, binTime = 300$(sec) とした、のである。なお、本データには主に 1 つの攻撃 ($v \cong 0.2(addresses/sec)$) のパケット到着パターンを確認している。各 p ごとに相互相関係数のピークがはっきりと山型になって表れる。しかし、この山のピーク値は p の値の増加とともに徐々に減少していく。これは、ラグ p が大きくなって距離が離れるほど元の時系列データとの相関の度合いが下がることを表わす。

5.2　時間相関とアドレス間距離との関係

binIP の値に 32,64,128,256 を使いそれぞれの遅延相関系列を観測し、高い相関が現れるアドレス方向の最長距離を比較し、距離と $binTime$ の高い相関があるかを確認した。$v = 0.2$(addresses/sec) の時の、表 2 には前者と比較して移動の速い $v = 4.0$(addresses/sec) の時の値をまとめた。表 1 より $binIP = 64$ のとき、アドレス方向の最長距離の値が最大になることが分かる。表 2 では、*binIP=32* のとき最大になるが、$v = 0.2$ のときも勘案すると、$binIP = 64$ のときにアドレス方向の最長距離の値は総じて高い値になっている。このことから、/28 のような小さなアドレス空間でも、大きな観測アドレス空間と同等以上の効果を得ると予測される。また、/29 のアドレス空間で観測を行うと、他のアドレス空間に比べて不安定になることもわかった。

6　結論

本研究では、時間相関とアドレス間の距離の関係を示し、分散協調アーキテクチャを利用するのに適切なアドレス空間の大きさを予測した。分散協調アーキテクチャを利用する際は、/28 のアドレス範囲を使うことが効果的であることが分かった。今後はより多くのデータを利用して本結果の検証を行う予定である。

表 1　$v = 0.2$(addresses/sec) の時の binIP と距離との関係

binIP	256	128	64	32
アドレス方向の最長距離	1536	1792	2176	1184

表 2　$v = 4.0$(addresses/sec) の時の binIP と距離との関係

binIP	256	128	64	32
アドレス方向の最長距離	2048	2048	2304	2400

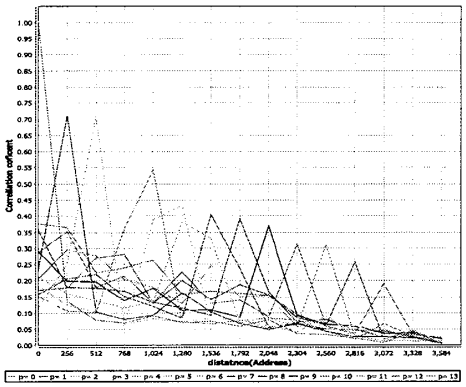

図 2　*binIP* = 256 の場合の遅延相関系列

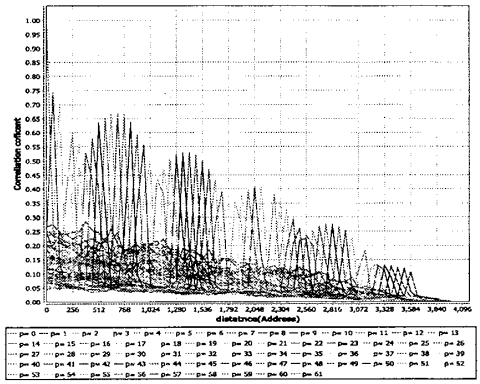

図 3　*binIP=64* の場合の遅延相関系列

参考文献

[1] Pang,R.,Yegneswaran, V.,Barford, P.,Paxson, V. and Peterson, L, "Characteristics of Internet background radiation," IMC'04, pp.27-40 (2004).

[2] 福田 健介, 廣津 登志夫, 明石 修, 栗原 聡, 菅原 俊治, "異常パケットトレースのアドレス局所性に関する解析," 情報処理学会全国大会, 2008.

[3] 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 (83), pp.39-45, 2007.