ウィルスっていったいどの位感染しているのかな？

小堀　智弘

1. はじめに

インターネットの急速な普及に伴い，ウィルスによる被害が急増している．Wittyワームは感染すると 20,000 箇所のランダムな送信先を攻撃する[3]．このような，感染後の送信頻度や潜伏期間は，ウィルスの種類や感染先のネットワーク環境に依存して異なる．

しかし，ウィルスの平均感染期間が分かれば，ウィルスシグネチャー（定義ファイル）の更新頻度やオンラインスキャンを行うタイミングを決定する 1 つの指標を得ることが出来る．

本研究では，ウィルス平均感染期間を同定するために，数学的モデルに基づいて適切な閾値を求める方法を提案する．年間平均で何回ウィルスに感染するか，その感染する期間はどれ位なのかを明らかにする．

2. 活動期間の同定原理

2.1　基本定義

j 台の不正ホストが期間[0, t]にポートスキャン（以後スキャン）するセンサからログデータを抽出する．センサとは，不正ホストからのスキャンを観測する正規ホストとする．

不正ホストの活動は，ウィルスに感染してから始まり，ウィルスが検知されて駆除されることで終了する．このサイクルをラウンドと呼び，1 年間のラウンドの回数を r とする．また，ラウンドの長さを感染期間 d で表す．

2.2　期間同定のアルゴリズム

(1)　サンプリング

不正ホストの集合から，ランダムに 100 個の発信アドレスをサンプリングし，r，d を主観評価によって判別する．

(2)　固定閾値

スキャン間隔 t が閾値 T を超えたときを 1 ラウンドの終結と判断する．　最適な T を変化させた時の r，d を求めるプログラムによって求める．

(3)　適応閾値

ラウンドの間隔はパケットの数によって変わる．そこで、パケットの到着間隔がポアソン分布に従うことを仮定し，期間 t でパケットが届かない確率を同定する．λはホスト毎の平均パケット到着率であり，$\lambda=c/d_0$ と定義する．cはホストの年間の総カウント数，d_0は観測期間における最初と最後のパケットの時間差である．この時，全くパケット到着しない確率が 1%となる閾値は，$T^*=-\ln(0.01)/\lambda$で求められる，T^*を越える時，その前後を違うラウンドと考える．

3. 調査評価

ISDAS[2]のデータ 2004 年 9 月 1 日～2005 年 9 月 30 日における独立した 12 台のセンサについて解析した．

固定閾値（T=30）と適応閾値により算出した平均感染期間の分布を図 1 に示す．各同定手法の解析結果を表 1 に示す．

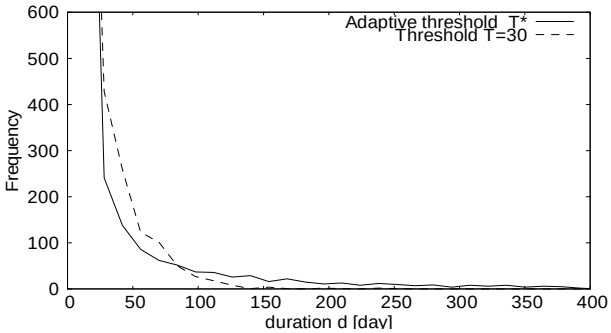

図 1. 固定閾値と適応閾値による感染期間 d の分布

図 1 の 2 つの曲線の比較をすると，80 日あたりを境に固定閾値と適応閾値の頻度の多さが入れ替わる．これにより，適応閾値の方が不正ホストの特徴が見られる．また表 1 より，ラウンド数は適応閾値の方がサンプリング値に近い．これらのことから固定閾値より適応閾値の方が主観評価に近いことが言える．

表 1. 閾値と平均感染期間

同定手法 閾値　T	サンプリング	固定閾値 T=30	適応閾値T^*
算出アドレス数	100	1586	1586
平均ラウンド数 r	1.49	1.671	1.567
平均感染期間 d	24.6	18.152	32.3

4. おわりに

最適な感染期間はそれぞれの不正ホストによって異なり，一概に固定の閾値による分類は当てはまらない．不正ホストは年間平均で 32 日間の寿命である．さらに，年間平均で 1.5 回感染を繰り返していることが分かった．今後の課題として，スキャン先の数の違いによってどれくらい感染期間の推定を行う．

参考文献

[1]小堀, 他, ISDAS 分散観測: ウィルスの平均寿命はいくらか?，情報処理学会，コンピュータセキュリティシンポジウム CSS2006，pp.519-524，2006.

[2]戸田，他，ISDAS: Internet Scan Data Acquisition System，CSS2004，pp. 199-204，2004.

[3]A. Kumar，V. Paxson and N. Weaver, "Exploit-ing Underlying Structure for Detailed Recon-strucion of an Internet-scale Event", USENIX，Internet Measurement Conference，pp. 351-364，2005.