代表的な特徴量と時間の揺れに基づいた χ 二乗値による異常検知

小 島 俊 輔[†1,†3] 中 嶋 卓 雄[†2] 末 吉 敏 則[†3]

ネットワーク上を流れるパケットの異常検出において，これまでに，発信元 IP アドレスや発信先ポート番号から求めた χ 二乗値により，異常を検知する方法が示されている．これまでの手法は，χ 二乗値の計算に 1 つの特徴量を用いており，同じシンボルが偶然連続して現れた場合に False-Positive が発生する問題があった．そこで本稿では，複数の特徴量から計算した χ 二乗値により，異常を発見する手法を提案する．また，昼夜のトラヒック変化に対応するため，BIN の境界値を動的に変化させる手法を提案する．提案する計算手法を適用して，時間の揺れを考慮した χ 二乗値を計算することにより，False-Positive の削減を実現できた．また，動的な BIN 境界値変更アルゴリズムを適用することにより，昼夜のトラヒック変化に対応させることが可能となった．

Anomary Detection using Chi-Square Value based on Main Feature and Time Deviation

Shunsuke Oshima,[†1,†3] Takuo Nakashima[†2] and Toshinori Sueyoshi[†3]

To detect anomaly packets owing in network, previous researches have shown that the chi-square value for the source IP address or destination port number can detect anomalies. In the previous researches, the chi-square values were calculated from one feature causing the degradation in terms of False-Positive when the same symbol appears sequentially. Therefore, we propose an anomaly detection technique using chi-square value based on multi features. We also propose dynamic BIN division technique to deal with the tra c uctuations such as day and night. Applying our method, the chi-square value based on time deviation could decrease the False-Positive. We also apply the dynamic BIN division technique, our method could adapt the day and night tra c uctuations.

1. はじめに

ネットワークに接続されたコンピュータは，サーバの機能を停止する DoS/DDoS 攻撃の脅威に，常に晒されている．悪意あるユーザは，DoS/DDoS 攻撃の第 1 段階として，ネットワーク接続された脆弱な PC を乗っ取り，BOT と呼ばれるコンピュータを多数作成する．この脆弱な PC を探し出すために利用されるのが，ポートスキャンや IP スキャンと呼ばれる手法である．スキャンに伴い，組織などのネットワークには大量のパケットが流れる．次に，第 2 段階として，悪意あるユーザは，多数の BOT に対して DoS/DDoS 攻撃対象となったサーバへの攻撃指令を出し，サーバとしての機能を麻痺させる．ここでも，サーバと BOT の間で大量のパケットが送受信される．いずれにおいても，大量のパケットがネットワークを飛び交っており，本稿ではこれ以降，サーバや BOT 化された PC で大量に観測される，このような悪意あるパケットを単に異常と呼ぶことにする．DoS/DDoS 攻撃の被害を最小限に抑えるためには，異常を早期に発見するための手法が必要である．このような異常検知手法では，新たに生み出される様々な異常に対応するため，何の知識も必要としない異常の発見手法が望まれる．

我々は，統計的な手法により通常トラヒックと異常トラヒックの違いを見分ける研究を行っている．本稿は，その中でも，期待度数と観測度数の差異を数値化する χ 二乗値に注目し，これにより DoS や DDoS，ポートスキャンなど，トラヒック変化を伴う多くの異常を検出できることを示す．本稿は以下のように構成する．まず，第 2 節では，χ 二乗値に関するこれまでの研究と問題点について述べる．第 3 節では提案する手法について解説する．第 4 節では，実験方法と実験に用いたデータについて説明し，第 5 節にて，今回得られた結果を示す．最後に，第 6 節で結論と今後の方針を述べる．

2. 関連研究と問題点

統計的な異常検知は，送信元 IP アドレスや送信先ポート番号といった特徴量を独立変数と見なして，パケットからシンボルを取り出し，その度数からエントロピーや χ 二乗値を計算することで異常を検知する．エントロピーは，各シンボルの出現度数が均一になると増大し，逆に 1 つのシンボルが集中して出現すると小さくなるという特徴がある．この特徴を用いて，これまでに多くの異常検知に関する論文が発表されている[2)5)7)9)]．エントロピーを

†1 熊本高等専門学校 ICT 活用学習支援センター
†2 東海大学 産業工学部
†3 熊本大学 自然科学研究科

用いた異常検知では，エントロピーが増加するか減少するかは異常の種類によって異なるため，未知の攻撃については，エントロピーがどちらに傾向するかをあらかじめ知ることは難しい．そこで，本稿では，DoS/DDoS に関係なく，パケット全体の分布の変化の具合を数値化する χ 二乗値に注目した．χ 二乗値はまた，Z 値 (Z-score) など他の多くの統計的手法と異なり，情報源のガウス分布を仮定しなくてよい．χ 二乗値は，以下の式で計算する．

$$\chi^2 = \sum_{i=1}^{B} \frac{(O_i \quad E_i)^2}{E_i} \tag{1}$$

ここで，B は観測されるシンボルの数，E_i はシンボルの期待度数，O_i は観測度数である．以後，シンボルの観測に用いる区間を窓，区切る基準長を窓幅 W と表記する．式 (1) の値は自由度 $B \quad 1$ の χ 二乗分布に従い，期待度数と観測度数の間に差異があるほど大きくなる．この χ 二乗値を用いた異常検知に関する文献をいくつか挙げておく．まず，論文 10)11) では，Solaris BMS の audit ログのイベント発生度数を基に，移動平均を計算し，それを基に χ 二乗値を計算することで異常が検知できることを示している．また，論文 3) や論文 12) では，プリンタの使用頻度やページ数，あるいはホームページの URL アクセス頻度などの情報から，通常と異なるユーザの使用を検知する方法が提案されている．さらに，論文 2) は，到達パケットの送信元 IP アドレスから求めた χ 二乗値が，DDoS 攻撃の検知に有効であることを示している．

χ 二乗値を用いた手法では，第 1 の問題点として，たまたま短期間に同じイベントが集中した場合に，False-Positive(以後 FP と略) と誤判断することが挙げられる．たとえば，パケットの異常検知を例に挙げると，近年の Web ページは，フレームや画像，CSS など多数のパーツから構成されており，さらに RSS の情報も勝手に届くため，わずか数ページを閲覧しただけで，同じ IP アドレスやポート番号を持つ多数のパケットが飛び交う．このように特徴量の揃ったパケットの集中は，まさに DoS/DDoS と同じ状況であり，χ 二乗値が増加して FP を引き起こす．そこで，特徴量の揃った場合にも χ 二乗値が大きくならない手法が望まれる．

第 2 に，χ 二乗値の安定と，早期あるいは小さい異常の発見にはトレードオフの関係がある．統計的手法では，連続したパケット列を一定の区間で区切り，その中のシンボルの度数をカウントする．窓幅が狭いと，同じ特徴を持つパケットが窓の中に偶然集中する確率は高く，FP は増える．そこで，論文 2) では，パケットの分布を均一に保つため，窓幅 $W = 10000$ を推奨している．一方で，逆に窓幅 W が広いと，パケットの収集に時間を要し，早期の異常発見は困難となる．さらに，異常パケットの個数と比較して窓幅が広い場合は，少ない異常パケットが多くの正常パケットに埋もれ，発見は難しい．そのため，窓幅を狭くした FP を抑える手法が望まれる．

第 3 の問題点として，χ 二乗値は，期待度数の 20%以上が 5 未満となる場合，精度が悪くなることが知られている[8]．そこで，通常は期待度数が 5 以上となるように，隣接する E_i 同士を結合するなどの前処理を行う．論文 2) では，送信元 IP アドレスを度数順にソートした後，全体を数個の BIN と呼ばれる 5 個程度の入れ物に分割し，(1) 式を適用する方法を提案している．この手法では，度数の小さいシンボルほど多くのシンボルを 1 つの BIN にまとめることで，度数の大きいシンボルとのバランスを取る．このまとめたシンボル数を以降ではビン幅と表現する．BIN 幅は，異常検知に先立ってあらかじめ決定しなければならず，一旦決定した後は勝手に変更できない．しかし，トラヒックのパケット分布が変化する夜中や週末においても，パケットの期待度数が確実に 5 以上となる BIN 幅を設定することは困難であり，そのため，BIN 幅を自動的に決定し，昼夜のパケット分布に動的に対応するような仕組みが必要となる．

3. 提案する異常検出手法

本稿では，前説で述べた 3 つの問題点の解消のため，複数の特徴量から χ 二乗値を求める手法を提案する．さらに時間の揺れを考慮することで，狭い窓においても FP が低減できることを示す．また，昼夜のトラヒック変化に対応するため，動的な BIN 幅変更アルゴリズムを提案する．

3.1 提案する複数の特徴量を用いた χ 二乗値

特徴量，すなわち独立変数を 2 つ使用して χ 二乗値を計算する場合，通常は $n \quad m$ 分割表を作成し，各セルの値を観測値 O_i とみなせばよい．この場合，求めた χ 二乗値は $(n \quad 1)$ $(m \quad 1)$ の自由度を持つ χ 二乗分布に従う．しかし，ネットワークを流れるパケットに対して，この手法をそのまま適用することは難しい．なぜなら，送信元 IP アドレスやパケット長など，パケットが持つ特徴量の各シンボルの出現頻度は冪乗則に従っており[1]，単純に分割表を適用すると，パケットが集中するセルと，ほぼ 0 となるセルが生じるためである．どちらの特徴量から見ても，50% ずつに配分されるような分割例を図 1 に示した．特徴が図 1(a) であれば，分割表が適用できるが，パケットには図 1(b) のような傾向があり，さらに特徴量の数や分割の数が多いほど，この傾向は顕著となる．

そこで，図 2 に示す手順により表を作成し，BIN に格納する手法を提案する．図は特徴量が 2 つの場合を説明しており，特徴量が N の場合は，次元数を N に拡張する．

提案する χ 二乗値の計算手順を以下に示す．

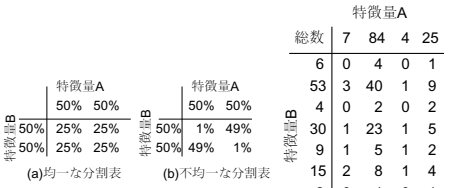

図 1 2 つの特徴量を用いた分割表の例

図 2 2 つの特徴量による χ 二乗値の計算

(a) 特徴量の数を N として，各特徴量ごとのシンボルをインデックスとする N 次元の配列を作成する．次に窓幅 W に入ったパケットについて順にシンボルを取り出し，該当する配列のセルの度数を 1 だけ増加し，同時に各特徴量ごとの総数も増加する．

(b) 窓内のすべてのパケットの集計が完了したら，ソート対象の特徴量以外の場所を保ったまま，特徴量ごとに総数でソートする．異常時におけるパケットは，送信元 IP アドレスなどのシンボルは集中もしくは分散どちらかの傾向があり，ソートにより異常なシンボルの位置が近づく．

(c) 各特徴量ごとの順位から，総合的な評価値 s を決める．求めた s を基に，用意した BIN に分割し BIN 中の総数をカウントし，この値を観測度数とする χ 二乗値を求める．

具体的な BIN への分割手順を示す．まず，パケット j の評価値 s_j を以下のように定める．

$$s_j = \sum_{k=1}^{N} \frac{r_{j,k}}{m_k} \tag{2}$$

ここで，N は使用している特徴量の数，$r_{j,k}$ は特徴量 k におけるパケット j の順位，m_k は特徴量 k のシンボル数である．たとえば，特徴量 A のシンボル数を 4，特徴量 B のシンボル数を 8 とし，特徴量 A におけるパケット j の順位を 2，特徴量 B の順位を 3 とすれば，$s_j = 2/4 + 3/8 = 0.875$ という値を得る．その後，$d_{i\ 1} < s_j \quad d_i$ ならば該当するパケットを BIN_i へ割り振る．ここで，d_i は BIN_i の境界値である．例えば，$d_1 = 0.5, d_2 = 0.75, d_3 = 1$ とすれば，s_j が 0.5 以下であれば BIN_1 へ，0.75 以下なら BIN_2 へ，1 より大きい場合は BIN_4 へ，といった具合に振り分ける．例示した $s_j = 0.875$ は，この場合 BIN_3 となる．この特徴量 N の数が増えると，ソート時の時間計算量と領域計算量の増加が懸念されるが，シンボルの度数の順位を先に調べ，シンボル順位への変換テーブルを作成すれば解決する．

本稿では，第 1 の特徴量として，論文 2) などにおいて有効とされた送信元 IP アドレスや送信先ポート番号を，また，第 2 の特徴量として，あるパケットが到達してから次のパケットが到達するまでの到達時間の揺れ Δt を用いる．Δt を考慮した理由は，Web アクセスや ssh, telnet など人の操作によって生じたパケットの Δt が一定とならないためである．発信元 IP アドレスのみを観測すると DoS に見える分布でも，Δt を加えた 2 次元平面上の分布を見ることで，DoS とは区別できる．

3.2 提案する BIN 幅変更アルゴリズム

BIN 幅は，特徴量の数 N，特徴量の中に含まれる IP アドレスやポート番号に含まれるシンボルの数，窓幅 W，BIN 数 B などの多くのパラメータが影響しており，境界を決定するのは難しい．しかし，BIN に求められる条件は，通常時における BIN 中のパケット数が 5 以下とならないように BIN 幅を設定することのみである．そこで，BIN へ分割する際の境界値 $d_i(t)$ を変更する以下の手法を提案する．

$$d_i(t) = d_i(t \quad 1) \left(1 \quad \eta \frac{\overline{p_i}(t \quad 1) \quad E_i}{E_i}\right) \tag{3}$$

ここで，期待度数 E_i は W/B などの定数であり，また，η は窓幅の変更係数である．この式は，境界値 $d_i(t)$ によって区切られたそれぞれの BIN の中の平均パケット度数 $\overline{p_i}(t)$ を常に期待度数に近づける．$\overline{p_i}(t)$ は時刻 t における窓内のパケット数 $p_i(t)$ より以下の指数移動平均の式で求める．

$$\overline{p_i}(t) = \lambda p_i(t) + (1 \quad \lambda)\overline{p_i}(t \quad 1) \tag{4}$$

ここで，λ は平滑化係数と呼ばれ，0 から 1 までの値を取る．

論文 2)10)11) では，1 日や 1 週間といった期間で変化するトラヒック変化に対応するため，過去に発生したイベントの発生を基に指数移動平均を用いて，固定した BIN 幅から動的に期待度数 E_i を計算する．提案する式は，逆に期待度数 E_i を固定し，BIN 境界 $d_i(t)$ を動的に変化させる．これにより，提案する式は，期待度数が 5 以上という条件を満足するだけでなく，トラヒック変化に追従する機能を持った式となる．BIN の境界 d_i は特徴量の数 N やシンボル数，窓幅 W，BIN 数 B などの多くのパラメータによる影響が考慮されており，従来必要だったパラメータの変更に伴う，面倒な BIN 幅の再決定の作業は不要となる．

3.3 BIN 境界変更アルゴリズムのシミュレーション

BIN 境界変更アルゴリズムを用いて，冪関数 $y = a \quad (x + 1)^{\ b}$ の y を E_i ずつに分割する点 $x = d_i$ を 5 箇所見つけるシミュレーションを行った結果を図 3 に示す．ここで，$\eta = 0.05, \lambda = 0.2, a = 10000, E_i = 2000$ とした．図の横軸はシミュレーションの繰り返し数 t，縦軸は求めた冪関数の区切り $x = d_i$ である．昼夜のトラフィック変化をシミュレート

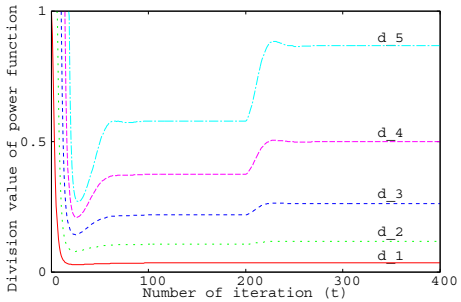

図 3 冪関数を均一に区切る値を求めるシミュレーション

するため，冪関数の指数部は，繰り返し数 $t < 200$ では $b = 1.2$，$t \geq 200$ では $b = 1.5$ に変更した．図から，d_i の値はシミュレーションの進行に従って収束しており，シミュレーション開始時で約 70 回，$t \geq 200$ では，約 45 回で目標とする値の 5%以下の誤差に収束した．これにより，提案する式はトラヒック変化に追従することを確認できた．ただし，異常を検知した場合は，異常に追従しないように，d_i の更新を停止するなどの工夫が必要となる．

4. 実験方法と実験データ

実験に用いたデータは DARPA1999[6] である．このデータは組織内と外部を接続するルータを行き交う約 5 週間分のパケットキャプチャを含んでおり，大小あわせて 100 万パケットからなる異常なパケットが含まれている．実験では，異常のない第 2 週を BIN 幅の決定用として，また，すべての異常パケットがラベルづけされている第 4 週と第 5 週のデータを異常検知のデータとして用いた．時間の揺れ Δt は，同じ送信元 IP アドレス，もしくは送信先ポート番号を持つパケット間の時刻差とした．Δt をそのまま使用すると度数の計測が困難となるため，123.456789[sec] であれば 100, 0.004321[sec] であれば 0.004 といった具合に，最上位の有効数字桁で丸める．また，到着したパケットと同じ IP アドレスの履歴が無い場合は，Δt を無限大と定義する．これにより偽装 IP アドレスを用いる DDoS 攻撃では，すべての Δt が無限大となり，また単調な IP アドレスを持つ DoS 攻撃は，Δt が単調となるため，いずれも Δt に偏りが生じる．

5. 実 験 結 果

5.1 BIN 幅変更アルゴリズムのトラフィック変化への追従性

まず，BIN の境界変更アルゴリズムの昼夜トラフィック変化への追従性を検証する．DARPA

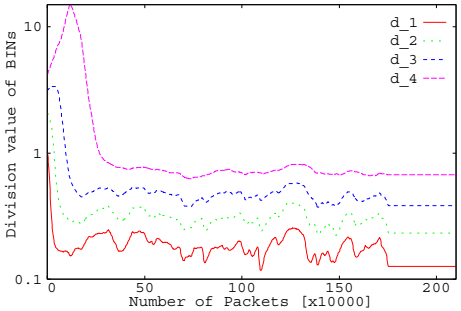

図 4 DARPA 第 2 週目のデータによる BIN の境界値の変化

第 2 週のデータを用いて，BIN の境界を求めた結果を図 4 に示す．パラメータは窓幅 $W = 10000$，BIN 数 $B = 5$, $\eta = 0.05$, $\lambda = 0.2$ とした．BIN の境界値 d_i が時間とともに変化し，約 70 回程度の反復計算後に窓幅は安定する．この回数は，図 3 の結果とほぼ同数であった．また，BIN の境界値，特に d_1 に周期的な動きを観測した．この周期性は，第 2 週目のデータに含まれる 5 日分のトラヒックによるものである．パケットのタイムスタンプは昼間に集中しており，周期性はこの昼夜のトラヒック変化により生じている．この結果から，提案する式は，DARPA1999 における実際のトラヒック変化に対する追従性を有することが確認できた．

5.2 異常検知の実験

前節の BIN 境界 d_i の決定後，第 4 週，第 5 週に含まれる異常を検知する実験を行った．実験では，データに多量に含まれる異常パケットによる BIN への悪影響を抑えるため，第 4 週以降で d_i の更新を停止した．ここで，従来の方法と同様に発信元 IP アドレスのみから求めた χ 二乗値を図 5 に，また，今回提案する発信元 IP アドレスと Δt の 2 つの特徴量から求めた χ 二乗値を図 6 に示す．なお，χ 二乗値は BIN 数 N や窓幅 W などによって最大値が大きく変わるため，最大値を 1 とする正規化を行っている．

発見した異常は合計 7 つであり，図中に DARPA で使用されている異常の名称を記した．図中の括弧書きは異常ではないのに χ 二乗値が大きくなったものであり，(t), (f), (s), (h) はそれぞれ，telnet, ftp, ssh, http を誤判定したことを意味する．図より，発信元 IP アドレスのみを用いた従来の方法ではこのような FP が多く含まれている．また，図の (m) は，組織外のクライアントから組織内の複数のサーバに対して，pop, http, ssh, telnet の多数のセッションが確立さた区間であり，1 台のクライアントから多くのパケットが短時間に集

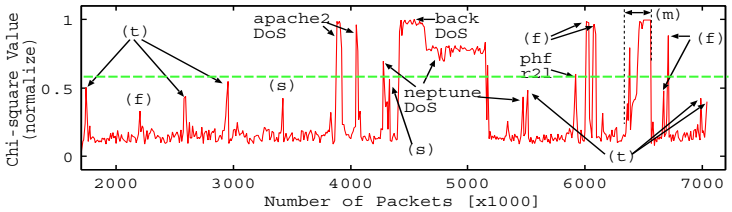

図 5 発信元 IP アドレスより求めた χ 二乗値 (窓幅 W=10000)

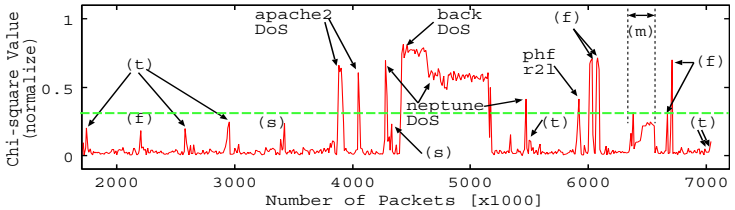

図 6 発信元 IP アドレスと Δt より求めた χ 二乗値 (窓幅 W=10000)

中している．そのため図 5 の χ 二乗値は異常値を示した．一方，Δt を考慮した図 6 では，telnet や ssh のほか，(m) の χ 二乗値は減少している．この結果から，提案する手法は，発信元 IP アドレスが集中した場合でも，人的な操作で発生したパケットの χ 二乗値を減少させる効果がある．図 5 では，ftp, ssh, telnet などを攻撃と誤検知しないしきい値を設定するといくつかの異常を見逃してしまうが，図 6 では，ftp や ssh, telnet の多くを攻撃と誤検知しないしきい値 (0.25-0.3 付近) を設定することができる．

一方，提案する手法では，ftp の χ 二乗値は大きなままであり改善されない．これは，ftp が機械的なパケットの送受信であり，Δt が単調になるため，Δt を考慮しただけでは異常との区別がつかきにくいことが原因である．さらに，ssh 接続についても，scp と思われるものは ftp 同様に χ 二乗値が大きくなった．このことから，ftp や ssh については，Δt 以外のパラメータを検討したり，知識ベースの侵入検知方法と併用するなどの工夫が必要となる．

次に，発信先ポート番号について，同様の実験を行った．従来の発信先ポート番号による結果を図 7 に，提案する Δt を考慮したものを図 8 に示す．図 8 から Δt を考慮した場合は，FP は ftp の 3 件のみと大幅に低減できた．一方で，図 7 で発見された 5 つの異常を見逃した．これは，Δt による平滑化が進みすぎたためである．もともとポート番号のシンボル数は IP と比較して非常に少ない．窓幅が広い場合は，1 つのポート番号に対して様々な Δt が観測されていまい，異常の持つ特徴が失われてしまう．

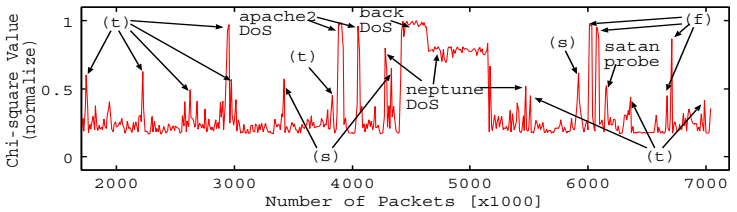

図 7 発信先ポート番号より求めた χ 二乗値 (窓幅 W=10000)

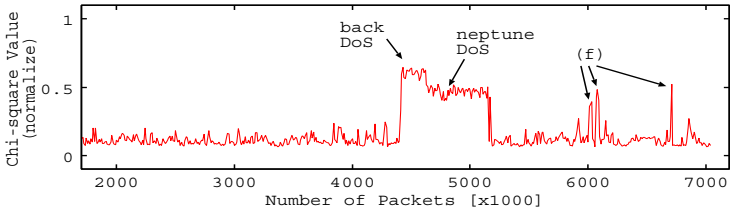

図 8 発信先ポート番号と Δt より求めた χ 二乗値 (窓幅 W=10000)

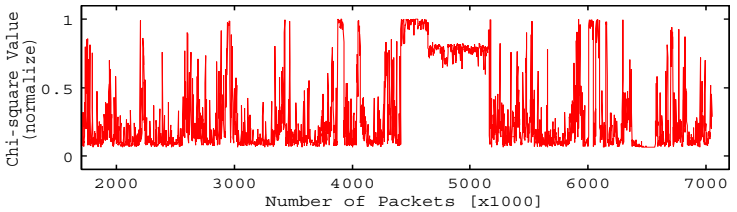

図 9 発信先ポート番号より求めた χ 二乗値 (窓幅 W=2000)

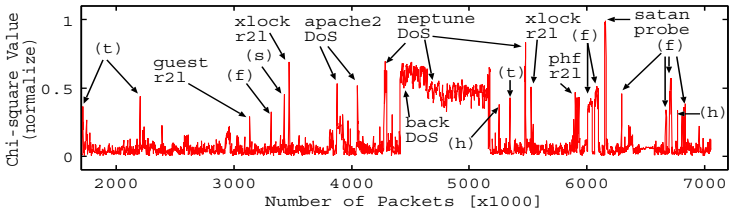

図 10 発信先ポート番号と Δt より求めた χ 二乗値 (窓幅 W=2000)

そこで，他のパラメータや条件は同じとし，窓幅のみを $W = 2000$ として，従来の方法による χ 二乗値を図 9 に，また，提案する Δt を考慮したものを図 10 に示した．図 9 の中

表 1　従来手法と提案する手法の比較

	従来手法			提案手法		
	W=10000		W=2000	W=10000		W=2000
	srcip	dstport	dstport	srcip+Δt	dstport+Δt	dstport+Δt
FP の数	△	△	×	○	○	△
小さな異常の発見	×	×	△	×	×	○

央の異常についてはかろうじて判別できるが，全体的に χ 二乗値が大きく揺れており，他の異常については区別がつかない．一方，図 10 は，図 9 に比べて χ 二乗値が大幅に減少している．理由は，Δt により，人為的に作成されたパケットの影響を抑えたためである．さらに，3134(1000) パケット目付近の guest r2l のように，図 8 で発見されなかったいくつかの異常が新たに発見された．この r2l は異常パケット数が合計で 417 パケットしかなく，窓幅 $W = 10000$ ではわずか 4.2%であるが，窓幅 $W = 2000$ では 21%となる．つまり，図 10 の結果から，提案する手法は窓幅が狭い場合にも有効であり，窓幅を狭くしたことで，DARPA1999 における小さな異常の発見が可能となった．最後に，従来の手法と提案する手法について，異常検出の適用範囲についてまとめたものを表 1 に示しておく．

6. おわりに

本稿では，BIN 幅を自動的に決定する手法の提案を行った．提案した手法を用いることで，シミュレーション実験では反復計算によりトラヒック変化へ追従した．また，実際のパケットデータを用いた実験では，昼夜のトラヒックに対応することを確認した．次に，複数の独立変数から χ 二乗値を求める手法を提案し，独立変数として Δt を考慮した．これによって，窓幅 $W = 10000$ における χ 二乗値では，ssh や telnet などの人為的に作成されたパケットの χ 二乗値が減少した．さらに，提案する手法は，窓幅が $W = 2000$ の場合にも有効であり，窓幅を狭くすることで DARPA1999 における小さな異常を発見できた．提案する手法は窓幅が狭い場合も良好な性能を示しており，早期に異常を発見することができる．

一方で，mailbomb など発見されていない異常がいくつかある．今後，パケットを詳細に調査することで，現在検知できない異常を検知する特徴量を追求していく予定である．mailbomb や portscan, DoS, DDoS など，異常の持つ特徴ごとに FP の少ない発見手法を構築することができれば，それぞれの結果の AND や OR を取ることで，数多くの種類の異常を正確に検知することができるようになると考えている．また，提案手法の広範囲における有効性を確認するため，LBNL[4] や MWS[13] のデータセットによる実験を検討している．

参 考 文 献

1) Falutsos, M., Faloutsos, P. and Faloutsos, C.: On Power-Law Relationships of the Internet Topology, *Proc. of ACM SIGCOMM*, pp.251–262 (1999).
2) Feinstein, L., Schnackenberg, D., Balupari, R. and Kindred, D.: Statistical Approaches to DDoS Attack Detection and Response, *Proceedings of DARPA Information Survivability Conference and Exposition*, Vol.1, pp.303–314 (2003).
3) Goonatilake, R., Herath, A., Herath, S., Herath, S. and Herath, J.: Intrusion Detection using the Chi-Square Goodness-of-Fit Test for Information Assurance, Network, Forensics and Software Security, *Papers of the Fourteenth Annual CCSC Midwestern Conference and Papers of the Sixteenth Annual CCSC Rocky Mountain Conference*, Vol.23, No.1, pp. 255–263 (2007).
4) LBNL/ICSI Dataset, http://www.icir..org/enterprise-tracing/download.html.
5) Lee, K., Kim, J., Kwon, K.H., Han, Y. and Kim, S.: DDoS Attack detection method using cluster analysis, *Expert Systems with Applications*, Vol.34, pp.1659–1665 (2008).
6) Lippmann, R., Haines, J. W., Fried, D. J., Korba, J. and Das, K.: The 1999 DARPA O -Line Intrusion Detection Evaluation, *Computer Networks*, Vol.34, No.4, pp.579–595 (2000).
7) Nychis, G., Sekar, V., Andersen, D.G., Kim, H. and Zhang, H.: An empirical Evaluation of Entropy-based Tra c Anomaly Detection, *Proceedings of the 8th ACM SIGCOMM Conference on Internet measurement*, Vouliagmeni, Greece, pp.151–156 (2008).
8) Siegel, S. and Jr., N. J.C.: *Nonparametric Statistics for the Behavioral Science 2nd edition*, McGraw-Hill (1988).
9) Wagner, A. and Plattner, B.: Entropy Based Worm and Anomaly Detection in Fast IP Networks, *Proceedings of the 14th IEEE International workshops on Enabling Technologies, Infrastructure for Collaborative Enterprise*, Linköping, Sweden, pp.172–177 (2005).
10) Ye, N. and Chen, Q.: An Anomaly Detection Technique Based on a Chi-Square Statistic for Detecting Intrusions into Information Systems, *Quality and Reliability Engineering International 17*, pp.105–112 (2001).
11) Ye, N., Emran, S.M., Chen, Q. and Vilbert, S.: Multivariate Statistical Analysis of Audit Trails for Host-Based Intrusion Detection, *IEEE Transactions on Computers*, Vol.51, No.7, pp.810–820 (2002).
12) Zhou, B., Shi, Q. and Merabti, M.: Intrusion Detection in Pervasive Networks Based on a Chi-Square Statistic Test, *Proceedings of the 30th Annual International Computer Software and Applications Conference(COMPSAC06)*, Chicago, Illinois, pp.203–208 (2006).
13) 畑田充弘他: マルウェア対策のための研究用データセット – MWS 2010 Datasets –, MWS2010 (2010).