Vol.2012-CSEC-57 No.1
Vol.2012-IOT17 No.1
2012/5/10

ネットワークトラフィック変化検知のための視覚的表現法に関する検討

小西航[†]　高橋秋典[†]　五十嵐隆治[†]　上田浩[††]
岩谷幸雄[†††]　木下哲男[††††]

インターネットには多様なトラフィックが疎通しており，正常トラフィックに混在する不正アクセスのような異常トラフィックを検知するためには，高度な専門知識が必要となる.そのため，知識を持たないエンドユーザはその脅威を認識できないという問題点がある．そこで，我々は R/S Pox Diagram の特性値を用いたトラフィックの視覚的表現法の検討を行い，トラフィック変化検知に対する性能について検証した.

Study of the Visual Presentation for Network Traffic Anomaly Detection

Wataru Konishi[†]　Akinori Takahashi[†]　Ryuji Igarashi[†]　Hiroshi Ueda[††]
Yukio Iwaya[†††]　Tetsuo Kinoshita[††††]

Various kinds of traffic pass over the Internet and the flows include many anomalies such as misuses or attacks. In order to detect such anomalies sophisticated knowledge is required in case we intend to detect anomalies. The requirement of the specialty knowledge becomes a burden to end users because they are not so trained to discern anomalies from normal traffic flows. This is the present purpose of our proposal to utilize R/S pox diagrams to visualize the change of traffic flows when they are exposed to various anomalies. In the present study we investigate the change detection characteristics of the R/S Pox Diagrams.

1. はじめに

インターネットのパケットトラフィック時系列が自己相似性に起因する長期記憶過程であることが発見されてから[1], その自己相似性の要因を調査する研究[2]が行われてきた. その要因には, 上位層プロトコルの TCP 輻輳制御[3][4]や輻輳・非輻輳の臨界領域の影響[5], また DDoS 攻撃のような非定常的トラフィックによる影響[6]など様々な報告があり，トラフィック事象変化に応じて自己相似性の様相が変化することが確認されている.

我々はこれまでトラフィック特性として自己相似性の度合いを表すハーストパラメータ H に着目し，異常トラフィックに対する H の変化を観測してきた[7][8][9]. このとき H 導出法の一つである R/S 解析法[1]において，導出過程で生成される R/S Pox Diagram に異常トラフィックに応じて特徴的なプロット点群が現れることを確認している[9]. 通常，この特徴的プロット点群はトラフィックの自己相似性の程度を検討する場合，非定常性を持つ事象として除外して検討されるものである．しかし，逆を言えばこの特徴的プロット点群は非定常性の特徴が反映されたものであり，このプロット形状変化をとらえることができれば，異常トラフィック検知につながることが期待できる．また，特徴を定量的にとらえるだけでなく，視覚的に特徴変化を呈示することができれば，専門知識を有しない利用者においても感覚的に異常と認識できる可能性も期待できる.

そこで, 本研究では R/S Pox Diagram の特徴的プロット形状の定量化手法を提案し，トラフィック特性の変化を視覚的に把握できる表現法に関する検討を行った．また，提案手法の有効性の検証として，長期的ポートスキャン攻撃に対する特徴量の効果を示し，検知指標としての可能性を評価した.

本稿の構成は以下のようにする．第 2 章では R/S 解析における累積範囲 R と標準偏差 S の計算および，R/S Pox Diagram による特徴量の導出法，また第 3 章では R/S Pox Diagram, および特徴量の効果的表現法を検討したトラフィック可視化ツール，第 4 章ではシミュレーションによる疑似攻撃トラフィックデータおよび実環境でのポートスキャントラフィックに対する本手法の効果，そして，第 5 章ではまとめを述べる.

2. R/S Pox Diagram 特徴量

2.1 R/S 解析法

R/S 解析は，H. E. Hurst がナイル川の流量変動の統計的解析に導入した後[10]，B. B. Mandelbrot により数学的な基礎付けがなされた統計解析法で[11]，Leland[1]が初めて，ネットワークトラフィックの自己相似性の解析に導入した

† 秋田大学大学院工学資源学研究科
Graduate School of Engineering and Resource Science, Akita University
†† 京都大学学術情報メディアセンター
Academic Center for Computing and Media Studies, Kyoto University
††† 東北大学電気通信研究所
Research Institute of Electrical Communication, Tohoku University
†††† 東北大学大学院情報科学研究科
Graduate School of Information Sciences, Tohoku University

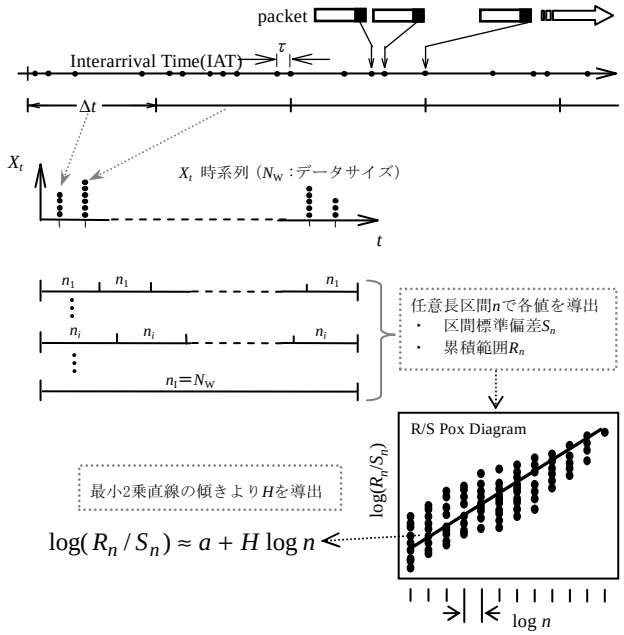


図 1　R/S 解析法を用いた H 導出

ものである．本解析法は“グラフ的な方法”と呼ばれるもので[1], [2], [12]，自己相似性を表すパラメータであるハーストパラメータ H は Pox Diagram を用いて導出される．H の導出過程を図 1 に示す.

まず，時間軸上での点過程であるパケットトラフィックを，単位時間Δt 毎の到着パケットを計数して得られる時系列データ X_t, $t = 1, 2, \cdots, N_W$ とする．時系列 X_t 内の任意長区間 n ，$1 \leqq k \leqq n$ において，当該区間平均 $(\overline{X})_n$ を用い，以下の(1)~(4)式により求めた区間標準偏差 S_n と累積和 W_k を用いて導出した累積範囲 R_n を求める．図 1 に示したように，$\log(R_n/S_n)$と $\log n$ をプロットしたものが Pox Diagram で，最小 2 乗法により求めたプロット点群の傾きを H とするものである.

$$(\overline{X})_n = \sum_{k=1}^{n} X_k / n \tag{1}$$

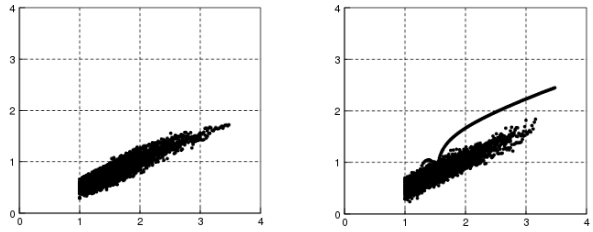
$$S_n = \sqrt{\sum_{k=1}^{n} {X_k}^2 / n - (\overline{X})_n^2} \tag{2}$$

$$W_k = \sum_{j=1}^{k} X_j - k(\overline{X})_n \qquad \{0, W_1, \cdots, W_k, \cdots, W_n\} \rightarrow k = 1 \sim n \tag{3}$$

$$R_n = \max\{0, W_1, \cdots, W_k, \cdots, W_n\} - \min\{0, W_1, \cdots, W_k, \cdots, W_n\} \tag{4}$$

2.2 特徴的プロット形状および提案特徴量

R/S Pox Diagram は，理論的な 2 次の自己相似過程である FGN（Fractional Gaussian Noise）のような時系列の場合，図 2(a)に示すように log(*n*)に対して単調的な増加傾向を呈する．しかし，この定常状態に非定常的な系列が混入した場合，図 2(b)に示すように，log(*n*)に対してその性質に応じた特異かつ興味あるプロット形状を示すことが観測されている[9]．このような状況は，ネットワークトラフィックにおいて正常通信が行われているとき，不正アクセスなどによる異常トラフィックの発生として想定できる．この特徴的プロット形状は，突発的なパケットトラフィックの増減のようなレベルシフト的時系列に対しては定常状態のプロット点群の上部に新たな点群が発生する．また，周期列と見なせるような間欠的なパケット到着事象が重畳されたトラフィック時系列においては，一方向の傾きから途中で折れ曲がるように 2 つの傾きを示すようになる．



(a) 定常時　　(b) LS 発生時

図 2　R/S Pox Diagram

2.3 提案特徴量

2.2 で示した特徴的プロット形状を定量的に扱うために，これまで各任意長区間長 n におけるプロット上限点の一群の傾きを H_{Sup}，下限点群の傾きを H_{Inf}，全プロット点の傾きを H として導出してきた[9]．このとき，H_{Sup} においてはレベルシフト的系列に効果的であったが，周期的トラフィックに対しては H_{Sup}，H_{Inf} ともに傾きが小さくなる傾向は示したものの，周期性を安定的に特徴づけることができなかった．これは，周期的トラフィックによる特徴的プロット形状が図 3 に表すように途中から傾きが変化する場合，1 方向の傾きによる評価が困難であったためと考えられる．

そこで本研究ではこれらの特徴量を改良し，R/S Pox Diagram の n における前半部と後半部の傾きを示す特徴量を新たに提案する．その特徴量を図 3 に示す．具体的には，

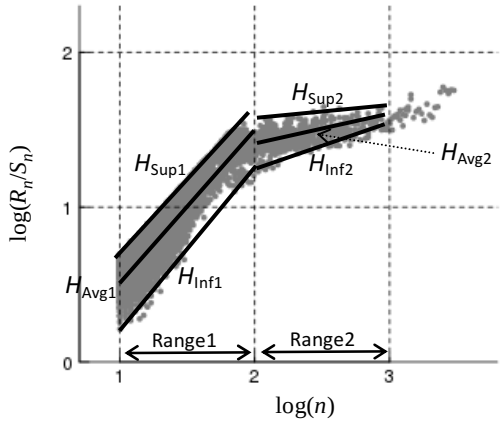


図 3　提案特徴量

Vol.2012-CSEC-57 No.1
Vol.2012-IOT17 No.1
2012/5/10

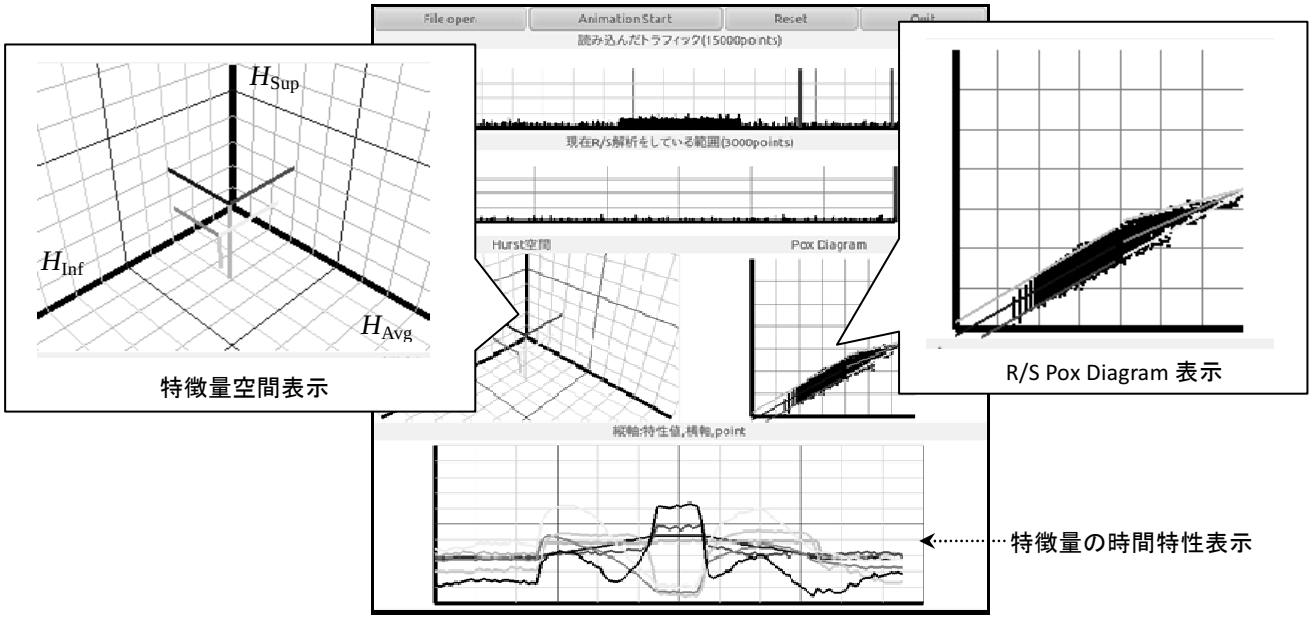


図 4　トラフィック可視化ツール

$1<\log(n)<2$ の範囲を前半部[Range1]，$2<\log(n)<3$ を後半部[Range2]と設定して，各範囲におけるプロット点群より傾きを導出する．それぞれの範囲における上限点群の傾きを H_{Sup1}，H_{Sup2}，下限点群の傾きを H_{Inf1}，H_{Inf2}，また，各 n のプロット点の平均値 $\overline{R_n/S_n}$ の傾きを H_{Avg1}，H_{Avg2} とする．このとき，H_{Avg1} は従来法である R/S 解析法で導出される H に相当する．これらの特徴量は概ね 0 から 1 までの値を示すが，最小 2 乗法による回帰直線の傾きであるゆえ，1 以上，または 0 以下の値となることもある．

この提案特徴量は性質として，レベルシフト的トラフィックに対してはプロット点が上部に現れるため H_{Sup1}, H_{Sup2} が高くなり，周期的トラフィックに対してはプロット点が水平方向に傾くため，Range2 の特徴量が小さくなる．

3. トラフィック可視化

本研究におけるトラフィック特性の可視化は，ネットワーク管理者に対する監視支援機能，およびエンドユーザに対する注意喚起を目的としたトラフィック事象確認機能といった観点で開発を行うものである．そこで，トラフィック特性変化をより効果的に認識できる表現法を検討するため，提案特徴量を用いたトラフィック可視化ツールを開発した．その可視化ツールの画面表示を図 4 に示す．

今回は提案特徴量の表現法として，R/S Pox Diagram 表示，3 次元特徴量空間表示，時間特性表示を検討し，様々なトラフィック時系列に対する動的変化を観測した．

3.1 R/S Pox Diagram 表示

R/S Pox Diagram 表示は，特徴的プロット形状を呈する R/S Pox Diagram をそのまま動的に表示するものである．その際，各特徴量の傾きを表す回帰直線も表示することで，より特徴量の変化を表すことができる．定量化する以前のプロット表示を確認することにより，数値に現れない変化も認識できると推測される．

3.2 特徴量空間表示

特徴量の新たな表現法として，H_{Sup}，H_{Inf}，そして H_{Avg} をそれぞれ 3 次元直交座標系に対応させた 3 次元特徴量空間を提案する．これにより，各範囲 Range1 および Range2 から得られるそれぞれ 3 つの特徴量の状態を，3 次元空間内の 1 点として表現することができる．空間内と 1 点として表現することで，それぞれの状態から式(5)で示す距離 D を得ることができる．

$$D=\sqrt{(H_{Sup2}-H_{Sup1})^2+(H_{Inf2}-H_{Inf1})^2+(H_{Avg2}-H_{Avg1})^2} \quad (5)$$

トラフィックが定常状態ならば各範囲の傾きがほぼ近い値となり，D の値は小さくなる．非定常的トラフィックにより特徴量に変化が生じると D も変動し，特に折れ曲がるプロット形状の場合，Range1 と Range2 それぞれの傾きが異なるので，D の値は大きくなる．

可視化ツールにおける特徴量空間表示は，各範囲の特徴量から空間座標を決定し，各面からの垂線で特徴量の大きさを表現する．これより，数値的距離 D の認識だけでなく，空間座標による情報も認識向上へ寄与すると考えられる．

3.3 時間特性表示

3.1 および 3.2 での表示では，観測時刻での状態の把握はできるが，特性変化に対する各値の変動傾向は確認できない．そこで，それぞれの値の経時変化を確認するため，本手法における特徴量および三次元特徴空間での距離 D を時間特性グラフとして表示する．

4. シミュレーション

本手法の有効性を検討するため，ポートスキャン攻撃を想定したシミュレーショントラフィック時系列データに対する時間依存特徴量 $H_{Sup1}(t), H_{Sup2}(t), H_{Inf1}(t), H_{Inf2}(t), H_{Avg1}(t), H_{Avg2}(t)$，$D(t)$を計測した．本稿では，ある時刻で瞬間的に発生する短期的ポートスキャン，時間間隔をおいて周期的に発生する長期的ポートスキャン，実環境のキャンパスネットワークで観測されたポートスキャンという 3 種類のトラフィック時系列データを用いた．各シミュレーション時系列サイズ N は 15000 点，特徴量を導出するための観測時系列 X_t のサイズ N_W は 3000 点とし，時刻 t を 1 ずつ進めるごとに特徴量を計算した．最初の特徴量導出のためには 3000 点が必要となるため，求められる時間依存特徴量のサイズは 12000 点となる．

シミュレーションデータにおいてポートスキャントラフィックを付加する前の定常時系列データは，各点においてパケットが発生する確率 p=0.01 を設定したベルヌーイ試行を全点に対して行い，50 回試行した系列を重畳したものを使用した．

4.1 短期的ポートスキャン

短期的ポートスキャンに対する適用例を図 5 に示す．図 5(a)に示すように，ポートスキャントラフィックは時刻 t=2000 に発生してΔt=5 間に 100 パケット到着すると想定したものである．

図 5(b)より，ポートスキャン開始時に H_{Sup1} が感度良く反応し，その後，約 3000 点の間，値を持続した．これは，発生時のトラフィック増加に対して導出時の累積範囲 R_n の値が大きくなることが要因と考えられる．その後，値が持続するのは，時系列 X_t にポートスキャントラフィックが観測されている間，R_n が大きくなる任意区間 n が存在するためであると推測される．図 5(c)において，時間が経過しポートスキャントラフィックが時系列 X_t から確認できなくなる直前，H_{Sup2} が 0 に近くなることが観測できた．それに合わせて，図 5(d)に示すように距離 D は値が大きくなる．

短期的ポートスキャンのような単発パルス的時系列に対する特徴量のふるまいから，以下のような知見が得られた．

- 突発的トラフィック量の増加に対する検知指標として H_{Sup1} が有効である．
- 観測時系列 X_t に対するポートスキャンが存在しなくなる判別には，H_{Sup2} が有効である．
- 単発のポートスキャントラフィックでは，H_{Inf}，H_{Avg} に際立った変動は見られない．

4.2 長期的ポートスキャン

長期的ポートスキャンとは，送信する調査パケットを時間間隔をおいて送信することで，パケット数を抑制させて行うスキャン攻撃である．これはトラフィック量のしきい値による異常検知法では検知しにくいものである．特に，

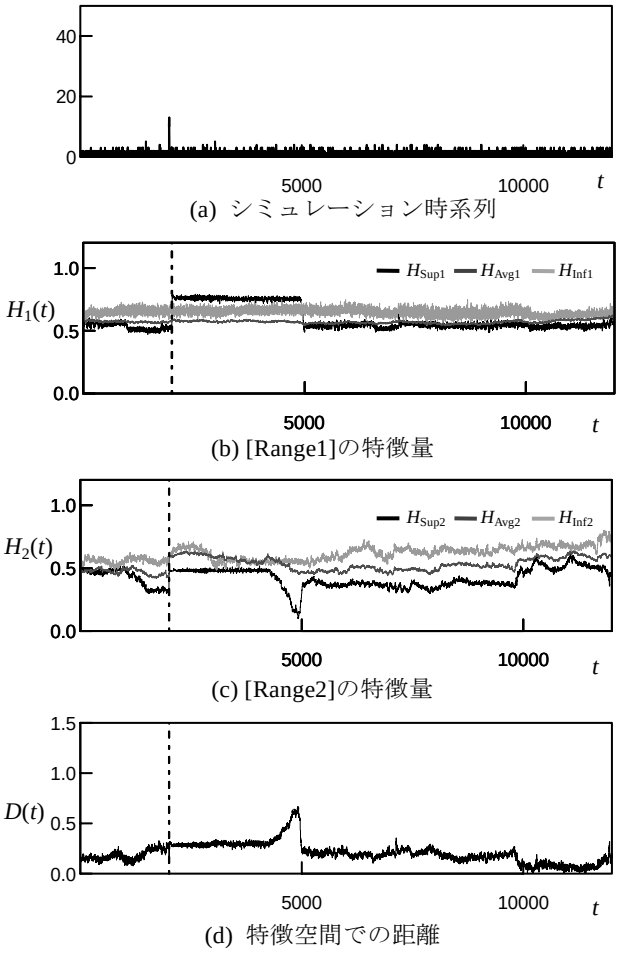


図 5　短期的ポートスキャンでの時間特性

間欠的に発生するバースト性を有しているので，しきい値の設定によっては，同一の攻撃者が行っている一連のポートスキャン攻撃を複数の攻撃として検知してしまう問題点もある．また，攻撃のトラフィックレートが低いので，正常通信パケットでも誤検知となる場合も多くなるので，注意が必要となる．

長期的ポートスキャンに対する適用例を図 6 に示す．図 6(a)に示すように，ポートスキャントラフィックは時刻 t=2000 からΔt=5 間に 100 パケット到着し，時間間隔 T=100 を置いて周期的に到着すると想定したものである．終了時刻は t=7000 とした．

図 6(b)より，4.1 と同様に，攻撃開始時は H_{Sup1} が感度良く反応し，その後，時系列 X_t 内に攻撃トラフィックが観測されなくなってから 3000 点後まで値を持続した．H_{Avg1} は攻撃発生後，徐々に高くなり，時系列 X_t 全体に攻撃トラフィックが反映した時点で最も高い値となった．攻撃終了後は徐々に低くなる傾向を示した．H_{Inf1} は時系列 X_t 全体に攻撃トラフィックが反映している間，定常状態に比べ高い値を示した．図 6(c)より Range2 においては各特徴量とも攻撃開始から徐々に低くなり，全体に反映した時点で全ての値とも 0 に近い値となった．これは Range2 におけるプロッ

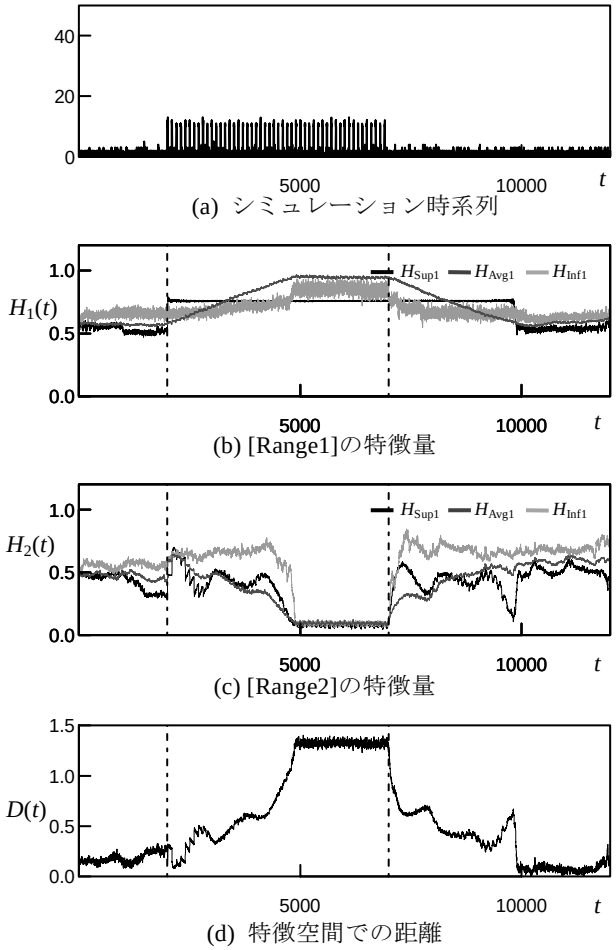


図 6 長期的ポートスキャンでの時間特性

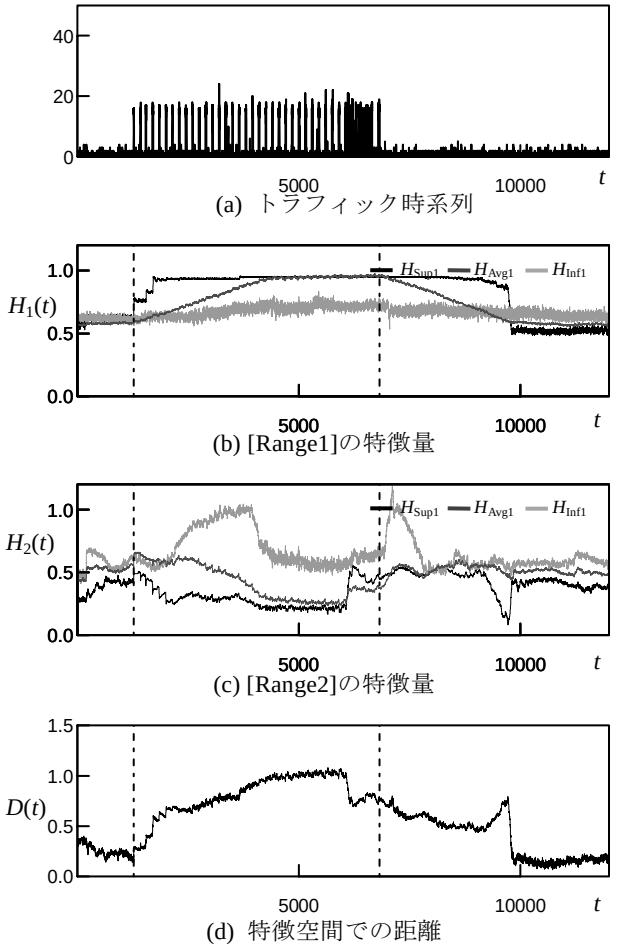


図 7 実環境におけるポートスキャン

ト点群の形状がほぼ水平になったことを示している．このとき，図 6(d)に示すように距離 D は非常に大きくなることがわかる．

以上の観測結果から，以下のような知見が得られた．

- 攻撃発生時は，短期的ポートスキャンと同様に H_{Sup1} が検知指標として有効である．
- 周期的系列の場合，Range2 のプロット形状が水平になることから，各特徴量も 0 に近い値となる．
- 特徴空間での距離 D からも周期的特徴をとらえることが可能である．

4.3 実環境でのポートスキャン

実環境で観測されたポートスキャン攻撃への適用を試みた．本データは，2008 年 8 月 26 日に秋田大学キャンパスネットワークの対外接続ポートで取得されたもので，ポート番号 4899 に対して SYN スキャンを実施している攻撃である．その詳細を表 1 に示す．

表 1．実環境でのポートスキャン詳細

発生日時	2008/08/26 09:02:16
Source IP	one
Destination IP	many
Destination Port	4899
パケット数	64243
攻撃時間	292.7 sec
平均パケットレート	219.4 packets/sec
パルス周期	約 3 sec（T≒150）

このポートスキャン発生時のトラフィック時系列を図 7(a)に示す．これは，パケット計測単位時間Δt=0.02s，TCP SYN フラグのみが立っているパケットをカウントした時系列である．よって，正常通信による SYN パケットもカウントされている．

図 7(b)より，4.2 のシミュレーション時系列とほぼ同様に，攻撃開始時に H_{Sup1} が高くなり，H_{Avg1} は徐々に高くなる傾向を示した．H_{Inf1} は有意な変動は観測できなかった．図 7(c)においても，シミュレーションと同様に H_{Sup1} および H_{Avg1} は低くなる傾向を示した．しかし，H_{Inf1} は徐々に高くなり，攻撃が全体に反映した時点でそれまでより低くなった．

特徴量の時間特性はほぼシミュレーションと同様な傾向を示したが，H_{Inf} については異なる傾向も見られた．これは，シミュレーションに比べてプロット形状の折れ曲がる位置が変動していることが要因と考えられる．この特徴変化に対しては，パルス周期 T の変化が影響していると推測される．

5. おわりに

本稿では，トラフィック特性変化に対する R/S Pox Diagram のプロット形状に着目し，形状変化を特徴づける特徴量を提案し，この特徴量を用いたトラフィックの視覚的表現法の検討ならびにポートスキャン攻撃に対する特徴量の効果を検証した．

可視化ツールにおいては，特徴量の時間特性のグラフ表示だけでなく，R/S Pox Diagram および 3 次元特徴空間を合わせて表示することで，実時間での状況および特性変化の傾向がより認識できたと考えられる．

ポートスキャン攻撃開始，終了の検知に対しては，H_{Sup1} ならびに H_{Sup2} の変化傾向に特徴が現れていることを確認した．また，周期的トラフィックに対して Range2 における特徴量が低くなる傾向を確認した．このことは，長期的ポートスキャン攻撃の継続状態を判別する場合に有効であると考えられる．

今後の課題として，パルス周期 T およびピークレートによる特徴量への影響，プロット形状からのパルス周期の推測法の検討などが挙げられる．

謝辞 本研究の一部は科研費（23500077）および東北大学電気通信研究所における共同プロジェクト研究 H22/A14 の助成を受けたものである．

参考文献

1) W. E. Leland, M. S. Taqqu, W. Willinger, D. V. Wilson: On the Self-Similar Nature of Ethernet Traffic, Computer Communications Review, Vol.23, No.4, pp.183-193 (1993).
2) J, Beran, R. Sherman, M. S. Taqqu, and W. Willinger: Long-Range Dependence in Variable-Bit Rate Video Traffic, IEEE Transactions on Communications, Vol.43, No.2/3/4, pp.1566-1579 (1995).
3) 住田義明，大崎博之，村田正幸，宮原秀夫: 上位層プロトコルがネットワークトラヒックの自己相似性に与える影響，信学論 B, Vol. J82-B, No. 6, pp.1126-1137 (1999).
4) 土井博生，松田崇弘，山本幹: TCP ふくそう制御がトラヒックのマルチフラクタル性に与える影響，信学論 B, Vol. J88-B, No. 6, pp.1029-1037 (2005).
5) K. Fukuda, M. Takayasu, H. Takayasu: A cause of self-similarity in TCP traffic, International Journal of Communication Systems, Vol. 18, No. 6, pp.603-617 (2005)
6) M, Li: Change trend of averaged Hurst parameter of traffic under DDOS flood attacks, Computers & security, Vol.25, No.3, pp.213-220 (2006)
7) 上田浩，奈須野裕，岩谷幸雄，木下哲男: 確率過程による LAN トラヒックのモデル化における一考察，情報処理学会論文誌, Vol. 48, No. SIG2, pp. 167-174 (2007).
8) 高橋秋典，五十嵐隆治，上田浩，奈須野裕，岩谷幸雄，木下哲男: オンラインネットワーク監視によるトラフィック異常検知，信学技法, NS2007-64, pp. 57-62 (2007).
9) A. Takahashi, R. Igarashi, H. Ueda, Y. Iwaya and T. Kinoshita: Network Anomaly Detection Based on R/S Pox Diagram, International Journal of the Society of Materials Engineering for Resources, Vol. 17, No. 2, pp. 186-192 (2010).
10) H. E. Hurst: A suggested statistical model of some time series which occur in nature, Nature, 180, 494 (1957).
11) Benoit B. Mandelbrot and James R. Wallis: Robustness of the Rescaled Range R/S in the Measurement of Noncyclic Long-Run Statistical Dependence, Water Res., Vol.5, No.5, pp.967-988 (1969).
12) M. S. Taqqu, Vadim Teverovsky and Walter Willinger: Estimators for long-range dependence an empirical study, Fractals, Vol.3, No.4, pp.785-798 (1995).