

R/S Pox レッグライン特性

R/S Pox Leg-Line Characteristics

高橋 秋典[†]
Akinori Takahashi五十嵐 隆治[†]
Ryuji Igarashi上田 浩[‡]
Hiroshi Ueda岩谷 幸雄[§]
Yukio Iwaya木下 哲男[¶]
Tetsuo Kinoshita

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は従来モデルであるポアソン過程には従わず、時系列の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 [2][3] や輻輳・非輻輳の臨界領域の影響 [4]、また、ネットワークアプリケーションによる影響 [5] や、DDoS 攻撃のような非定常的な異常トラフィックによる影響 [6] などの報告があり、トラフィック事象変化に対して自己相似性の様相が変化するということが観測されている。

この自己相似性を表すハーストパラメータ H の導出法の一つである R/S 解析法は、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 [7]、B.B. Mandelbrot により数学的な基礎付けがなされた統計的解析法 [8] で、Leland et al. [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ものである。この解析法は「グラフ的な方法」と呼ばれるもので、 H は観測時系列データにおける任意長区間から算出される R/S 統計量をグラフにした R/S Pox Diagram のプロット点群の傾きより導出される。この R/S 解析法に対しても様々な検討が行われており、R/S Pox Diagram の H 計算範囲の妥当性 [9][10] や計算高速化のためのアルゴリズム改良 [11] などが提案されている。

これらの研究においては、R/S 解析法から導出される H に着目しトラフィック事象に対する検討が行われているが、非定常的に突発的トラフィック量増加を示すレベルシフトが発生する場合や、間欠的に到着し、且つ到着期間内では周期列となるようなトラフィック時系列の場合、R/S Pox Diagram に特徴的なプロット形状が現れることが報告されている [12][13][14]。我々の調査においては、特に周期的時系列に対しては、プロット点群が一つの傾きではなく、途中から折れ曲がり、二つの傾きを呈することが観測されている。

これらのプロット形状は、非定常性を表す特徴であると考えられるが、信頼性に欠ける H 推定を与えてしま

うため、統計解析においては除去すべき対象として検討されてきた。Mandelbrot et al. がコンピュータシミュレーションを用いた研究で、ランダム過程に決定論的過程、具体的には sin 波を重畳したとき、その R/S Pox Diagram 形状に変化が現れることを指摘していた [8]。しかし、非定常性が重畳されたときの R/S Pox Diagram の形状変化から事象変化を積極的に利用しようとした研究は見当たらない。著者らは、その非定常性に対するプロット形状の定量化は事象変化の検知に積極的に応用できると考えている。また、R/S 解析法の計算過程において、R/S 統計量を求める任意長区間のサイズによって未計算部分が存在し、直近の取得データが解析に反映されないことがある。これに対して問題視する研究も見当たらない。

本研究では、まず R/S 解析法の未計算区間により時間特性に対する即応性が低下する問題点を指摘し、その改良法を提案する。さらに、その改良法を用いた R/S Pox Diagram に現れる特徴的なプロット形状の要因について明らかにし、その形状を定量的に扱える R/S Pox レッグライン特性を提案する。この名称は、周期的時系列に対して折れ曲がるようなプロット点群の形状が人間の脚部に似ていることに由来する。この特性を用いて周期的時系列に対する周期推定法を検討し、シミュレーション時系列に対する性能評価を行った。また、その際、実環境から観測された長期的ポートスキャン攻撃トラフィックに対して適用を試み、そのパケット到着間隔を推定することで本手法の有効性を示した。

本研究の構成を以下に示す。2 章では、R/S 解析法を用いた R/S Pox Diagram の導出ならびに未計算区間に対する改良について述べる。3 章では、周期性を有する時系列に対する R/S Pox Diagram の特徴について述べ、4 章ではその特徴を定量的に扱う R/S Pox レッグライン特性を提案する。5 章では、シミュレーション並びに実トラフィック時系列データに対する適用を試み、6 章でまとめる。

2. R/S 解析法

2.1. 導出手順

R/S 解析法 [1] による R/S Pox Diagram の導出手順を図 1 に示す。

まず、対象とする観測データから解析対象とする時系列データ $X_t : t = 1, 2, \dots, N$ を得る。パケットトラフィックにおいては、時間軸上での点過程であるトラフィックを、計測単位時間 Δt 毎の到着パケットを計数して得ることになる。この時系列 X_t に対してデータが

[†] 秋田大学大学院工学資源学研究所, 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

[§] 東北学院大学工学部, Faculty of Engineering, Tohoku Gakuin University

[¶] 東北大学大学院情報科学研究科, Graduate School of Information Sciences, Tohoku University

重複しない任意長区間 $n: 1 \leq k \leq n$ を定め, 式 (1) により当該区間平均 $(\bar{X})_n$ を求める.

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

当該区間の線形な傾向 $k(\bar{X})_n$ からのずれを表す累積和 $W_k: 1 \leq k \leq n$ を式 (2) より求める.

$$W_k = \sum_{j=1}^k X_j - k(\bar{X})_n \quad (2)$$

さらに, 式 (3) より, この累積和 W_k の最大値と最小値の差を表す累積範囲 R_n を求める. この R_n と式 (4) から求められる当該区間標準偏差 S_n の比を用いて, R/S 統計量を導出する.

$$R_n = \max\{0, W_1, \dots, W_k, \dots, W_n\} - \min\{0, W_1, \dots, W_k, \dots, W_n\} \quad (3)$$

$$S_n = \sqrt{\sum_{k=1}^n X_k^2/n - (\bar{X})_n^2} \quad (4)$$

このとき, 任意長区間 n で導出される R/S 統計量の数は, $mn \leq N$ を満たす最大値 m となる. さらに, n を Δn ずつ増加させながらそれぞれの R/S 統計量を導出し, 式 (5) に基づいてハーストパラメータ H を推定

する.

$$\log(R_n/S_n) = H \log(n) + \log c \quad (5)$$

具体的には $\log(R_n/S_n)$ を被説明変数, $\log n$ を説明変数, $\log c$ を定数とした回帰モデルと想定し, 最小 2 乗法により H を推定する. これをグラフ的に表すと, 横軸 n , 縦軸 R_n/S_n のグラフにプロットされた点群の傾きを求めることになる. このグラフが R/S Pox Diagram である.

2.2. R/S 解析法の未計算区間

R/S 解析法は, サイズ N の時系列 X_t に対して任意長区間 n を定め R/S 統計量を導出する. しかし, 図 1 に示すように, N が n で割り切れる場合, X_t 全区間データに対して計算が行われるが, 割り切れない場合は X_t の時間軸上後半の部分に計算されないデータが存在する. これは n が大きくなるに従い, 未計算区間も大きくなることになる.

通常, 時系列解析において観測時系列の時間軸は時刻 t が小さい方向が”過去”を表し, 大きい方向が”直近”を表している. つまり, 未計算区間にはより直近のデータが存在することになり, 解析に反映されにくくなる. 観測対象によっては即応性が重視されることもあり, 直近に発生したデータが解析に反映されないという事実は, 特徴量の経時変化特性の検討上, 問題となる場合がある.

2.3. 反転時系列による改良

この問題点を解決するために, より”直近”に近いデータを X_t の前半に, ”過去”のデータを後半に反転させた時系列 X'_t を生成する. これは, 式 (7) に示すように, 時系列データを単純に反転させたものである.

$$X_t = \{x_1, x_2, \dots, x_N\} \quad (6)$$

$$X'_t = \{x_N, \dots, x_2, x_1\} \quad (7)$$

この改良法では, 直近の取得データが時系列 X_t の前半 $N/2$ までに存在しているならば, どの任意長区間 n においても解析に反映されることになる.

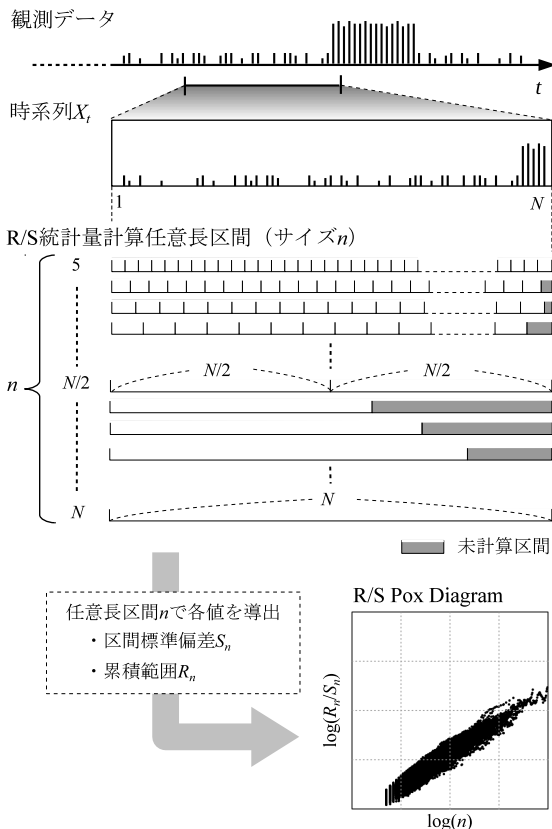


図 1 R/S 解析法

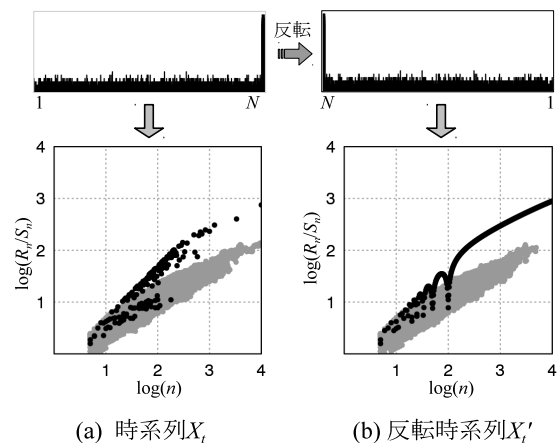
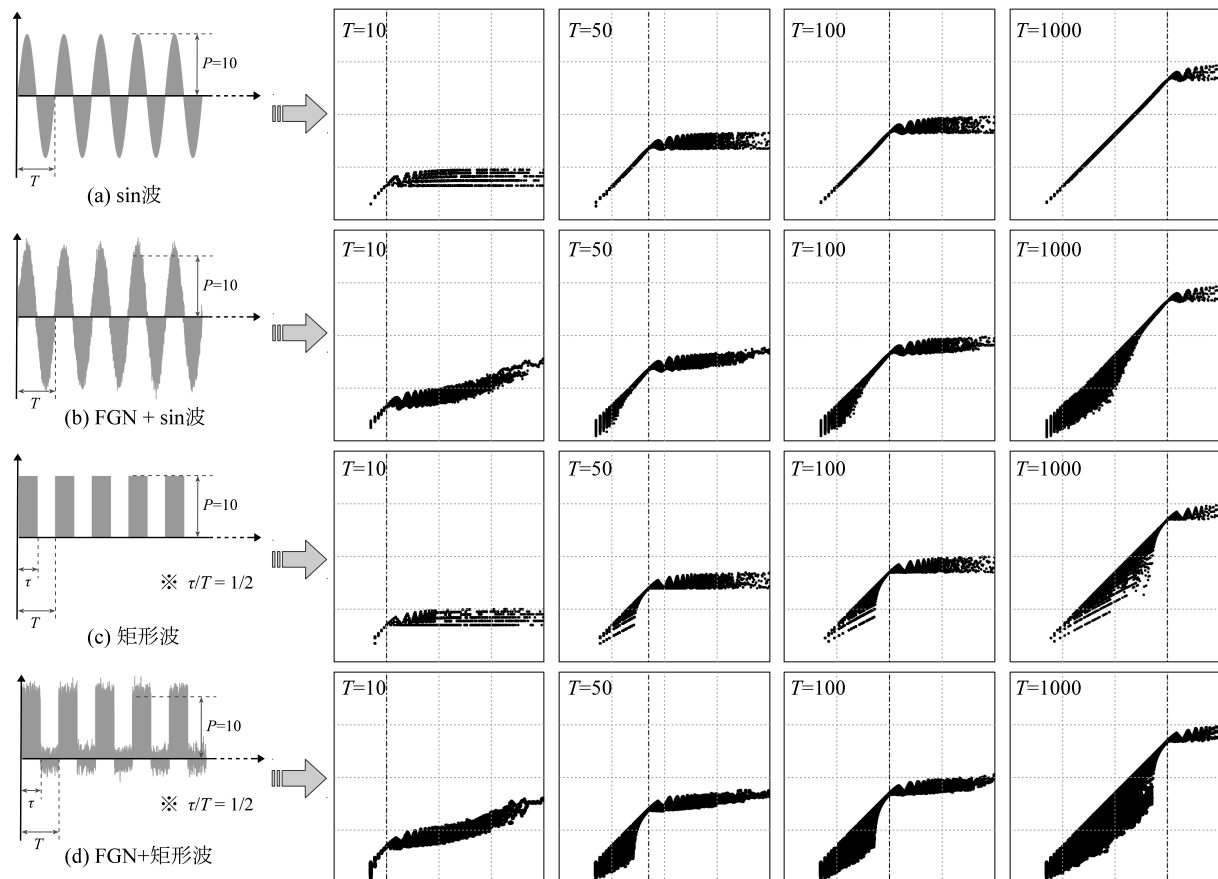


図 2 反転時系列の効果

図 3 周期 T の影響

2.4. 改良法の効果

改良法の効果を示すため、突発的にトラフィックデータが増加した（レベルシフト）シミュレーション時系列を用いて、従来型の時系列 X_t と改良型の反転時系列 X'_t の 2 つから得られる R/S Pox Diagram を比較した。用いたシミュレーション時系列および導出された Pox Diagram をそれぞれ図 2(a), 図 2(b) に示す。シミュレーション時系列 X_t は、分散 1.0, 平均 3.0, $H = 0.5$ と設定したサイズ $N = 10000$ の FGN(Fractional Gaussian Noise) に対し、区間 [9901-10000] に振幅 $P = 10$ のレベルシフト系列を重畳させたものである。これは、観測対象である時系列に対して、非定常的トラフィック系列が早い段階で観測されたことを想定したものである。R/S Pox Diagram の灰色のプロット点は全区間におけるプロット点を表し、黒のプロット点はレベルシフトの変化点 $t = 9901$ が含まれる任意長区間から導出されたものを表す。

図 2(a) から、 X_t に対する R/S Pox Diagram では、レベルシフトの変化点が含まれるプロット点がまばらに観測され、特に $\log(n)$ の後半部分では未計算区間が多く存在している。しかし、図 2(b) より、反転時系列 X'_t では全ての任意長区間 n に対してレベルシフトの変化点

が計算に反映されていることがわかる。これより、反転時系列を用いることで早い段階で観測される非定常的系列に対して即応性が高まる効果が期待できる。

3. 周期性に対する R/S Pox Diagram

本研究では、間欠的に到着するパケットトラフィックのように周期性を有する時系列に着目し、その特徴を明らかにした。本解析法に関する検証は、解析に用いるパラメータが多すぎて数学的検証が困難と考えられる。そこで、本研究ではシミュレーション時系列を用いた実験的検証を行った。検証に用いたシミュレーション時系列のデータサイズ N は 10000 とし、R/S 解析の任意長区間 n の増分 Δn は 0.01 とした。周期的時系列の波形として、周期 T の影響に対する検証には sin 波および矩形波を用いたが、その他の検証はどちらもほぼ同様な結果となったため、本稿では矩形波のみの結果を示した。

3.1. 周期 T の影響

周期 T に対する R/S Pox Diagram のプロット形状変化の様子を図 3 に示す。図 3(a) は振幅 $P = 10$, 周期 T の sin 波、図 3(b) は分散 1.0, 平均 0.0, $H = 0.6$ とした FGN を定常状態時系列とし (a) の sin 波を重畳させたもの、図 3(c) は振幅 $P = 10$, パルス幅 τ , 周期 T とする矩形波、図 3(d) は (b) と同様の FGN に対し

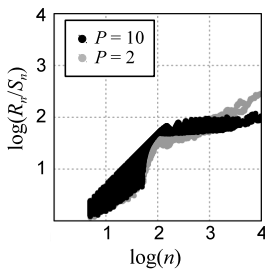


図 4 P の影響

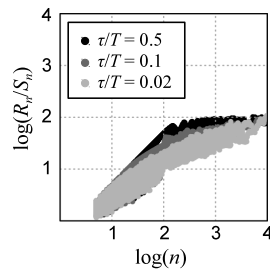


図 5 D の影響

て矩形波を重畳させた時系列である．それぞれの時系列に対して周期 $T = 10, 50, 100, 1000$ と変化させたときの R/S Pox Diagram を示した．グラフの一点破線は $\log(n) = \log(T)$ を示している．矩形波のデューティー比 τ/T は全て 0.5 とした．

結果より，全てのシミュレーション時系列において $n = T$ となるポイントを境にプロット点群の傾きが水平方向に折れ曲がる傾向が顕著に現れた．また， n が前述のポイントより小さい範囲では，各 n における R_n/S_n の最大値の点群による傾きが，全てのシミュレーションにおいてほぼ 1 となった． n が前述のポイントより大きい範囲では， $T = 10$ のとき，FGN が重畳されていると傾きが若干曲線的になったが，その他はほぼ傾きが直線的に，かつ値が 0 に近くなった．この折れ曲がるポイントが明確に周期 T を示したことを用いて，2 つの傾きを表す回帰直線の交点から周期を推定する手法を検討した．

3.2. 振幅 P の影響

3.1 節で用いた FGN に周期 $T = 100$ の矩形波を重畳させた時系列に対して，振幅 $P = 2$ および $P = 10$ としたときの R/S Pox Diagram を図 4 に示す．

$P = 10$ に比べ $P = 2$ の場合，プロット点群の折れ曲がる傾向が弱くなっている．これは，定常状態に比べある程度の振幅変化がないとき，周期的特徴が顕現しにくくなると予測される．

3.3. デューティー比 D の影響

矩形波を特徴づける指標であるデューティー比に対する影響を検討する．パルス幅 τ ，周期 T のとき，デューティー比 D は式 (8) で求める．

$$D = \tau/T \quad (8)$$

3.2 節と同様に FGN に周期 $T = 100$ の矩形波を重畳させた時系列において，パルス幅 τ を変化させてデューティー比 $D = 0.02, 0.1, 0.5$ としたときの R/S Pox Diagram を図 5 に示す．

振幅の影響と同様に， D の値に応じてプロット点群の折れ曲がる特徴に影響が現れたことから， D の値によっても周期的特徴が顕現しにくくなると予測される．

4. R/S Pox レッグライン特性

第 3 章で示した周期的時系列に対するプロット形状は，人体脚部の太腿，膝，脛と形状が非常によく似てい

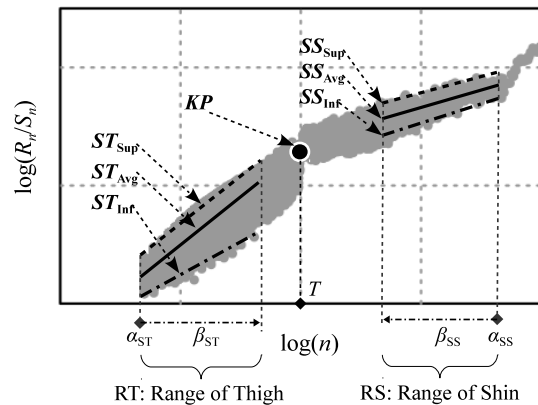


図 6 R/S Pox レッグライン特性

ることから，本研究では R/S Pox レッグライン (Leg-Line) 特性と呼ぶことにする．この特性から得られる特徴量を図 6 に示す．折れ曲がるプロット形状の特徴は，脚部の太腿 (Thigh) と脛 (Shin) の部分に見立て， H 導出法と同様にそれぞれの部分におけるプロット点群の傾きで表す． $n < T$ における導出範囲を RT (Range of Thigh)，傾きを ST (Slope of Thigh)， $n > T$ における導出範囲を RS (Range of Shin)，傾きを SS (Slope of Shin) と表記する．また， $n = T$ となるポイントは，脚部の膝 (Knee) と見立て， KP (Knee Point) と表記する．本章では，この特徴量の定量化法について述べる．

4.1. 各範囲における傾き (Slope Value)

R/S Pox Diagram の各範囲におけるプロット点群は，図 3(a) のように一直線上に並ぶものや，図 3(c), (d) のように縦軸方向に広がるものなど様々な傾斜の形状を呈する．そこで，各範囲において一方向の傾きではなく，プロット点の各 n における上限点群 $\max(R_n/S_n)$ ，平均点群 $\text{avg}(R_n/S_n)$ ，下限点群 $\min(R_n/S_n)$ から求まる三方向の傾きで表す．導出範囲 RT におけるそれぞれの傾きを $\{ST_{\text{Sup}}, ST_{\text{Avg}}, ST_{\text{Inf}}\}$ ，導出範囲 RS におけるそれぞれの傾きを $\{SS_{\text{Sup}}, SS_{\text{Avg}}, SS_{\text{Inf}}\}$ と表し， H を求めた式 (5) と同様に最小 2 乗法を用いて推定する．

導出範囲 RT, RS の範囲長は，サンプル数とのトレードオフになる．すなわち，図 6 に示した各量の高精度導出を目指す場合，RT ないしは RS の範囲長を大とする必要があるが，この範囲長を大きくし過ぎると，R/S Pox Diagram の縦軸方向でのプロット点数の減少領域をカバーしてしまい，ST や SS の変動 (ゆらぎ) の増加となってしまう恐れがある．また，図 3 に示すように，周期 T の変動によって KP も変動するため，極端な短周期および長周期に対して範囲長を事前に設定するのは困難である．そこで，本研究では観測対象を推定可能とする周期 T の範囲を限定し，その範囲において周期推定を行うことを目標とし，RT, RS の範囲長を検討した．この手法では，周期推定範囲は限られるが，観測データ生成時の時間スケールを調整すれば，実時間上で異なる周期について推定が可能となる．つまり，並列処理によ

り広範囲または限定した周期的時系列の観測システムへの適用も検討できる。本研究では、実時間解析における計算コストを考慮したときの時系列 X_t のデータサイズ N を 10000 以下として検討することから、有意な推定範囲として $10 \leq T \leq 1000$ を目標とした。

導出範囲 RT, RS は以下とした。すなわち図 6 において、 $\log(n)$ ではなく n の値で α_{ST} , および α_{SS} を設定し、これらを各々、RT および RS の起点とする。また、 $\log(n)$ 上の値で RT および RS の範囲長 β_{ST}, β_{SS} を設定する。従って RT は α_{ST} を起点とし、範囲長 β_{ST} , RS は α_{SS} を起点とし、図示の向きの範囲長 β_{SS} となる。RT の起点 α_{ST} は 5 とし、RS の起点 α_{SS} は、任意長区間 n が大きくなると求められる R/S 統計量が少なくなることから、最低三点が求められる $n = N/3$ と設定した。また、範囲長 β_{ST}, β_{SS} は求められる傾きの精度に影響するため調整が必要となるが、本研究では経験的にそれぞれ 0.8 と設定した。

4.2. Knee Point による周期推定

周期 T の特徴となる KP は、導出範囲 RT, RS から求められた各傾きを表す回帰直線の交点の x 座標値から求める。ただし、周期的時系列が重畳されていなくても RT, RS の傾きに少しでも差異が生じた場合、 KP が導出されてしまうことがある。これを回避するために、周期的時系列に対して SS の値がしきい値 γ より小さくなったとき導出を行う制約条件を設ける。また、 KP の値が推定目標とした $10 \leq T \leq 1000$ の範囲外となったとき、導出できなかったものとする。制約条件を考慮した周期推定手順を以下に示す。

- (i). ST_{mode1} および SS_{mode2} を一つずつ選択する。
- (ii). $SS_{mode2} \geq \gamma$ ならば、(i) に戻る。
- (iii). KP 計算
- (iv). $KP < 10$, または $KP > 1000$ ならば、(i) に戻る。
- (v). $T = KP$

mode1, mode2 はそれぞれ、{Sup, Avg, Inf} を表しており、これらを用いたときに導出される KP の数は 9 個となる。それぞれの傾きより計算される KP を表 1 にまとめる。例として、 ST_{Sup} および SS_{Avg} から求められる交点の x 座標を KP_{SA} と表し、以下に計算例を示す。式 (9) は、 ST_{Sup} , 式 (10) は SS_{Avg} の回帰直線を表し、 KP_{SA} は式 (11) より求められる。ただし、 C_{STS}, C_{SSA} は定数である。

$$\log\{\max(R_n/S_n)\} = ST_{Sup} \log(n) + C_{STS} \quad (9)$$

$$\log\{\text{avg}(R_n/S_n)\} = SS_{Avg} \log(n) + C_{SSA} \quad (10)$$

表 1 導出される KP

	SS_{Sup}	SS_{Avg}	SS_{Inf}
ST_{Sup}	KP_{SS}	KP_{SA}	KP_{SI}
ST_{Avg}	KP_{AS}	KP_{AA}	KP_{AI}
ST_{Inf}	KP_{IS}	KP_{IA}	KP_{II}

$$KP_{SA} = \frac{C_{STS} - C_{SSA}}{ST_{Sup} - SS_{Avg}} \quad (11)$$

より精度の高い周期推定が可能となる KP の選定は、第 5 章で行う。

5. シミュレーションによる性能評価

提案する R/S PoX レッグライン特性の性能評価として、第 3 章で示した周期、振幅、デューティー比に対する傾き ST, SS を導出し検討を行った。さらに、シミュレーション時系列および実環境から取得された長期的ポートスキャン攻撃のトラフィック時系列に対する周期推定も行った。シミュレーション時系列は図 3(d) に基づいた FGN に矩形波による周期的時系列を重畳させたものを使用した。

5.1. 周期的特徴に対する傾き

5.1.1. 周期特性

周期 T に対する傾き ST, SS を図 7 に示す。

結果より、 ST において ST_{Sup} は $20 < T \leq 1000$ のとき、安定して約 1.0 の値を保っている。 ST_{Avg} および ST_{Inf} は $T = 30$ のとき最大値をとり、その後 T が増加するにつれて、値が低下している。特に ST_{Avg} は、 $T = 100$ 以降、定常状態とほぼ同じ値に戻っている。ここから、 ST_{Sup} はどんな周期 T に対しても安定した周期的時系列の検知ができると推測できる。また、 ST_{Avg} は、 $T > 100$ となる周期性には反応せず定常状態の特性を表していると推測できる。

SS では、どの傾きも $20 \leq T \leq 800$ において 0.3 以下の値となり、水平方向に傾く特徴を捉えることができた。ただし、 T が 1000 近辺になると傾きが元に戻るようになる。これは、RS の導出範囲設定に起因するものであり、 β_{SS} の調整により修正が可能と考えられる。

5.1.2. 振幅特性

周期 T を 50, 100, 500 と設定したときの振幅 P に対する傾き ST, SS を図 8 に示す。 ST においては、周期 T による値の変動はあるものの、振幅 $P > 10$ では全ての傾きにほとんど変化が見られなくなる。つまり、定常状態に対してある程度の大きさの P であれば、観測時系列内で振幅変動があっても、安定した値を呈すると推測できる。また、周期特性と同様に ST_{Sup} は約 1.0 の値を保っていることから、ここでも検知指標としての有効性が推測される。

SS においては、各傾きとも P の増加に伴い値が小さ

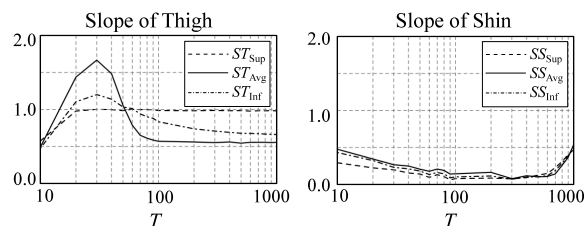
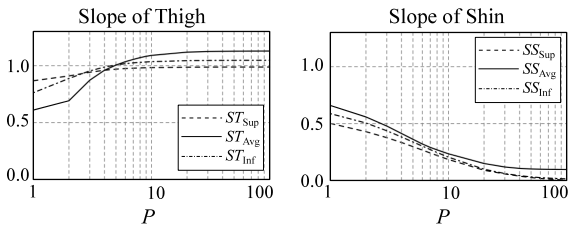
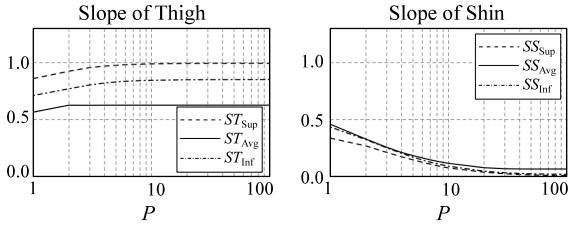
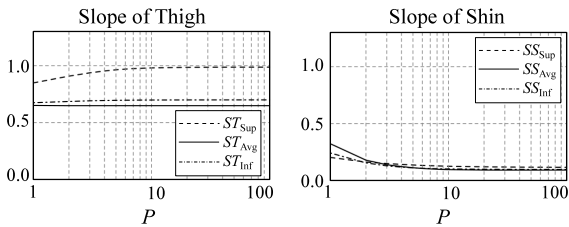


図 7 周期 T に対する傾き

(a) パルス周期 $T=50$ (b) パルス周期 $T=100$ (c) パルス周期 $T=500$ 図 8 振幅 P に対する傾き

くなり、徐々に水平方向に傾く傾向が観測された。ただし、振幅が $1 \leq P \leq 10$ が比較的小さい場合、その傾きは周期 T の影響が優先されることも観測できた。

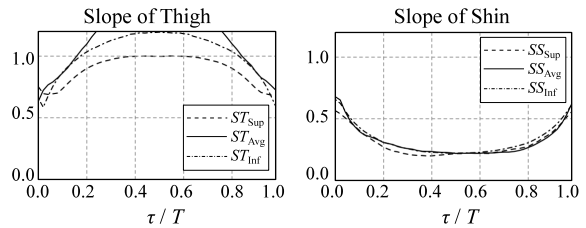
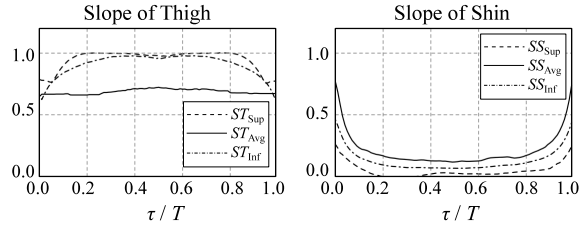
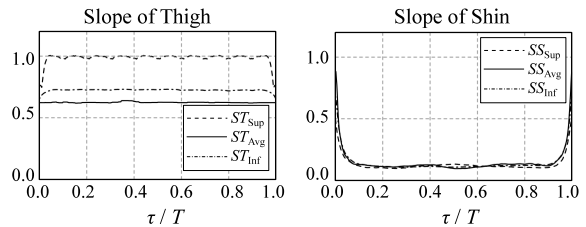
5.1.3. デューティー比特性

周期 T を 50, 100, 500 と設定したときのデューティー比 D に対する傾き ST , SS を図 9 に示す。

ST および SS のどちらの場合においても、 $D = 0.5$ においてそれぞれ最大値および最小値をとることが観測された。このときの値が周期的時系列の特徴を最も捉えた値と推測される。この値が安定して観測できる D の範囲は、周期 T の値に応じて変化することがわかる。 ST_{Sup} に着目してみると、 $T = 50$ ではおおよそ $0.4 < D < 0.6$, $T = 100$ では $0.2 < D < 0.8$, $T = 500$ では $0.04 < D < 0.96$ で安定している。それぞれの T における安定状態開始点の D からパルス幅 τ 計算してみる。

$$\begin{aligned} \tau &= D \times T \\ &= 0.4 \times 50 = 0.2 \times 100 = 0.04 \times 500 \\ &= 20 \end{aligned} \quad (12)$$

全ての T において、パルス幅 $\tau = 20$ が得られた。つまり、今回のシミュレーションにおいて、 ST_{Sup} はパルス幅 $20 \leq \tau \leq T - 20$ の周期的時系列に対して安定して約 1.0 の値を示すことがわかった。

(a) パルス周期 $T=50$ (b) パルス周期 $T=100$ (c) パルス周期 $T=500$ 図 9 デューティー比 τ/T に対する傾き

5.2. 周期推定

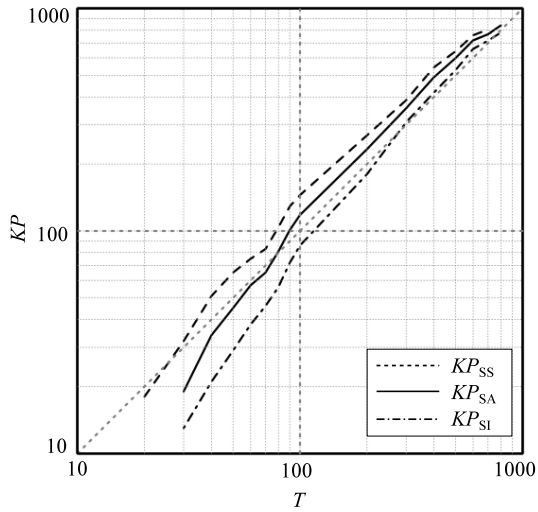
5.2.1. シミュレーション時系列

シミュレーション時系列 (FGN+ 矩形波) を用いて周期を $10 \leq T \leq 1000$ と変化させたときの周期推定法から導出された KP を図 10 に示す。 SS のしきい値 γ は 0.3 とした。図 10(a) は、導出範囲 RT において傾き ST_{Sup} を選択したとき、 RS の傾き $\{SS_{Sup}, SS_{Avg}, SS_{Inf}\}$ から導出された $\{KP_{SS}, KP_{SA}, KP_{SI}\}$, 図 10(b) は、傾き ST_{Avg} を選択したときに導出される $\{KP_{AS}, KP_{AA}, KP_{AI}\}$, 図 10(c) は、傾き ST_{Inf} を選択したときに導出される $\{KP_{IS}, KP_{IA}, KP_{II}\}$ を示している。グラフの対角線にある灰色の破線は、理想直線を表す。プロット点の存在しない T は、提案周期推定手法の制約条件によって導出されなかったものである。

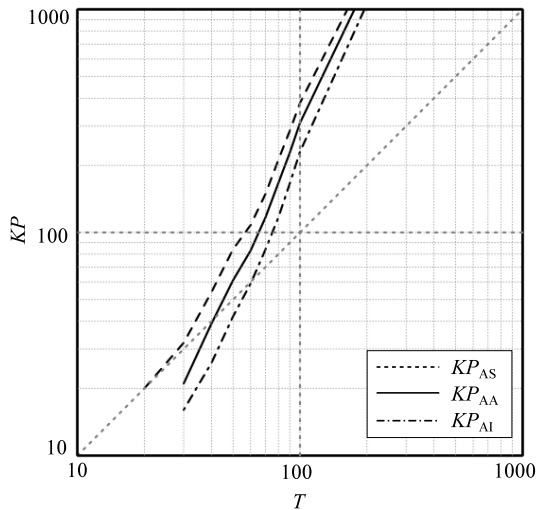
結果より、導出範囲 RT において傾き ST_{Sup} を選択した場合、他に比べ、 RS のどの傾きにおいても理想直線に近似した形となり良好な周期推定が行われている。ただし、 T が小さい場合は KP_{SS} が良好な結果を示し、 $T = 100$ 前後では KP_{SA} , T が大きい場合は KP_{SI} とそれぞれ有効な範囲があることが示された。 $T < 20$ において KP が導出されなかった原因は、5.1.1 項で示したように RT , RS の導出範囲によるものと推測される。

5.2.2. 実環境からのトラフィック時系列

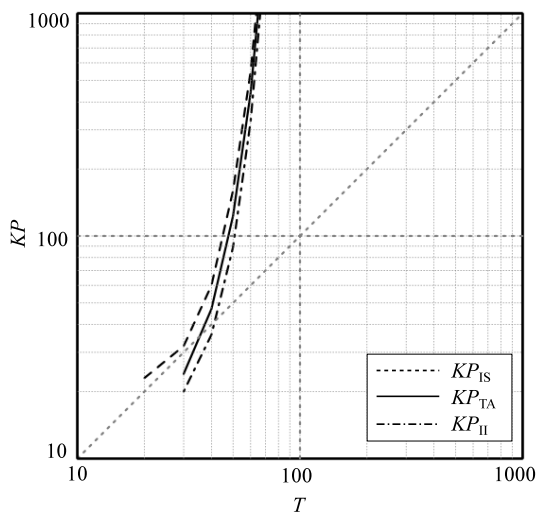
実環境から計測されたトラフィック時系列に対する適用として、長期的ポートスキャン攻撃が含まれたトラフィック時系列を用いた周期推定を試みた。長期的ポートスキャン攻撃とは、ポートスキャンに使用する調査パ



(a) ST_{Sup} 利用時



(b) ST_{Avg} 利用時

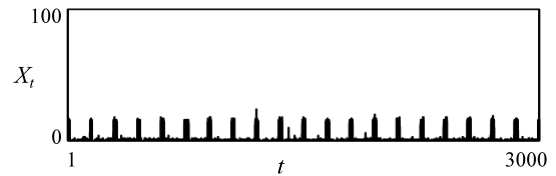


(c) ST_{Inr} 利用時

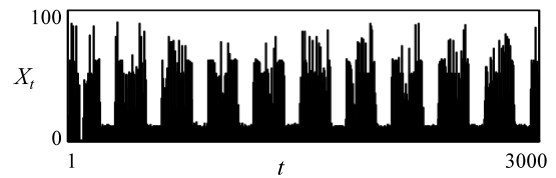
図 10 周期 T に対する KP

表 2 ポートスキャン攻撃の詳細

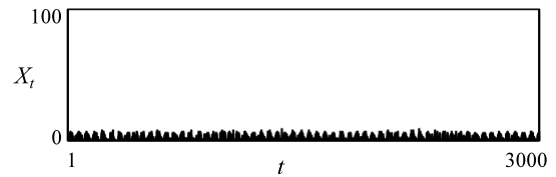
	Scan 1	Scan 2	Scan 3
計測日	2008/08/26	2008/08/27	2008/08/30
宛先ポート	4899	80	23
周期 T	約 150	約 300	約 50
パルス幅 τ	約 10~20	約 200	約 20
振幅 P	約 15	約 30~90	約 5



(a) Scan 1 (2008/8/26/09:02:16)



(b) Scan 2 (2008/8/27 03:12:15)



(c) Scan 3 (2008/8/30 11:00:00)

図 11 ポートスキャン攻撃トラフィック時系列

ケットを短時間で大量に送信せずに、時間間隔を置いて間欠的に少量の調査パケットを送信することで検知しにくくする攻撃である。間欠的に送信されるため周期的時系列の特徴を有しており、本手法によって攻撃検知、ならびに周期推定が可能な観測対象と考えられる。

周期推定に用いたポートスキャン攻撃の詳細を表 2 に、そのときの時系列 X_t のグラフを図 11 に示す。トラフィックデータは秋田大学キャンパスネットワークの対外接続ポートから tcpdump を用いて取得されたパケットデータで、学外から到着した TCP SYN パケットのみをカウントしたものである。この時系列 X_t には、攻撃トラフィック以外にも正常な通信で発生した TCP SYN パケットも含まれる。計測単位時間 Δt は 0.02s、時系列 X_t のサイズ N は 3000 点、つまり解析に用いたデータは 60s 間に計測されたものである。

提案手法による周期推定結果を表 3 に示す。導出範囲 RT の傾きは、前項で示したように安定した値をとる ST_{Sup} を利用して KP を 3 点導出した。結果より、全てのスキャン攻撃に対して、他の値に比べ KP_{SS} の値が比較的精度よく周期を推定することができた。また、Scan1 のようにパルス幅 τ の変動、Scan2 のように振幅 P の変動がランダムに発生している場合でも、良好な周

表 3 周期推定結果

	Scan 1	Scan 2	Scan 3
KP_{SS}	129	299	64
KP_{SA}	108	194	46
KP_{SI}	86	114	32

期推定が行われた。ここから、提案する周期推定法はパルス幅、振幅の変動に対しロバストな手法であると推測できる。

6. まとめ

本研究では、まず、ハーストパラメータ H の推定に用いられていた R/S 解析法の未計算部分の存在を指摘し、反転時系列を適用することで、直近で発生した事象が解析に反映させることを示した。これより、より高い即応性、実時間解析への適用が期待できる。

さらに、R/S Pox Diagram に現れる周期的時系列の特徴に着目し、これを定量化する R/S Pox レッグライン特性を提案した。ここで示した導出範囲 RT , RS の傾きは、周期的時系列の重畳に対して顕著な変化を示すことから、事象変化の検知指標として期待できる。また、プロット形状の折れ曲がるポイントが明確に周期を示すことを利用した周期推定法も提案し、パケットトラフィックを想定した時系列データに対して検証を行った。この結果より、Thigh 部分の上限点群スロープ ST_{Sup} と Shin 部分の上限点群スロープ SS_{Sup} から求められる Knee Point KP_{SS} を用いた場合、良好な周期推定が行えることを示した。

今後は、導出範囲の調整による周期推定の精度向上の実現、ならびに事象の経時変化に対する特徴量の時間特性の明確化により、本特性を利用した長期的スキャン攻撃の検知手法について研究を進める。

謝辞 本研究の一部は科研費 (23500076,23500077) 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] 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フィ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.
- [3] 土井博生, 松田崇弘, 山本幹, TCP ふくそう制御がトラフィ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.
- [4] K. Fukuda, M. Takayasu, and H. Takayasu, A cause of self-similarity in TCP traffic, International Journal of Communication Systems, Vol.18, No.6, pp.603-617,2005.
- [5] 上田浩, 奈須野裕, 岩谷幸雄, 木下哲男, 確率過程による LAN トラフィ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.
- [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] H. E. Hurst: A suggested statistical model of some time series which occur in nature, Nature, 180, 494, 1957.
- [8] 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, May 1969.
- [9] 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.
- [10] 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.
- [11] R. Igarashi, S. Ono, A. Takahashi, Y. Iwaya and M. Sakata, Some Features of Network Traffic Depending on Protocols, Proceedings of ICMR 2005, pp. 426-431, 2005.
- [12] T. Dang and S. Molnar, On the Effects of Non-Stationarity in Long-Range Dependence Tests, Periodica Polytechnica, Ser. El., Vol.43, No.4, pp.227-250, 1999.
- [13] 松葉育雄, 非線形時系列解析, 朝倉書店, 2000.
- [14] 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, Sep. 2010.