

FW へのアクセス件数の予測手法を利用した異常検出

1 はじめに

サイバーフォースセンター（CFC）では、全国の警察施設のインターネット接続点においてファイアウォール及び侵入検知装置（Intrusion Detection System : IDS）を通じて攻撃活動等の監視を行っている¹。

本分析レポートは、こうしたファイアウォールに対するアクセスについて、過去のアクセス状況を基に、アクセス件数の推移を予測する手法について検討し、その予測手法の信頼度を評価したものである。

この結果に基づき、当センターでは、異常な状態が発生した際の国民へのタイムリーな情報提供に向けて、本手法による予測値と観測値との比較によってアクセス件数に異常がないかを日常的に監視している。また、RSI（Relative Strength Index）²等の指標によって異常を検出した際に、定常状態であればどのようなアクセス件数となったかを予測し、現在のアクセス件数と比較することによって異常状態の分析に役立てている。

2 予測手法

本分析レポートの予測手法は、ファイアウォールに対するアクセスについて、1週間先までの1時間毎のアクセス件数を予測することを目的としている。

近代の統計学においては、時系列分析の手法として、自己回帰移動平均（autoregressive moving average : ARMA）モデル等の時系列モデルが用いられることが多いが、これらの手法では、「長い期間の予測では、標本平均自体が最適予測となり、予測誤差の分散は確率過程の分散そのものとなる。³」とされており、 $24 \times 7 = 168$ 期先までの予測を必要とする今回の目的にはそぐわない。

そこで、標本データをトレンド成分と周期成分に分解して予測する加法型成分分解手法について検討する。

¹ 「インターネット定点観測」とは

http://www.cyberpolice.go.jp/detect/obs_info.html

² 我が国におけるインターネット治安情勢について

http://www.cyberpolice.go.jp/server/rd_env/pdf/incident_analysis.pdf

³ 「経済の時系列分析」（山本 拓：創文社）

なお、インターネット治安情勢⁴において「国内の時間帯推移」(図1参照)として情報提供しているように、時間帯によってアクセス件数が増減することがわかっており、また、曜日による特色もあると推測されるため、周期成分として1日間周期変動及び1週間周期変動を想定している。

また、トレンド成分としてアクセス件数の長期的な傾向の変動を想定しており、インターネットに接続されているコンピュータの台数やウイルス、ワームに感染しているコンピュータの台数の増減等によって変動するものと考えられる。

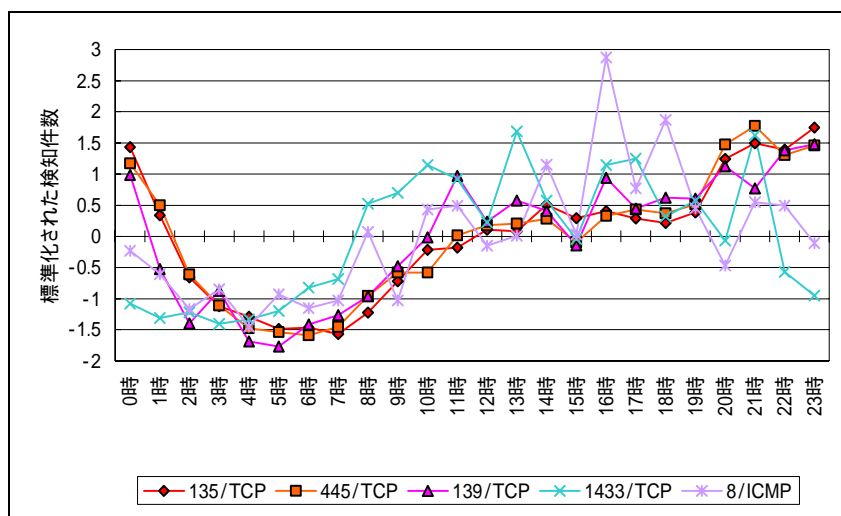


図1 国内の時間帯推移 (インターネット治安情勢から引用)

本分析レポートで提案する予測手法「TFACE: Traffic Forecast by Additive Component Extract」は次のとおりである。

(1) トレンド成分と周期成分の抽出

予測する時間を N (1週間先までを予測する場合は $N = 168$)、標本データを $\{y_1, y_2, \dots, y_T\}$ ($T = \{kN \mid k \text{は自然数}\}$) とする。この標本データの N 時間移動平均を求め

$$\bar{y}_t = \frac{1}{N} \left(\frac{1}{2} y_{t-\frac{N}{2}} + \sum_{i=-\frac{N}{2}+1}^{\frac{N}{2}-1} y_{t+i} + \frac{1}{2} y_{t+\frac{N}{2}} \right) \quad \left(1 + \frac{N}{2} \leq t \leq T - \frac{N}{2} \right) \quad (2.1)$$

$$\Delta y_t = y_t - \bar{y}_t \quad \left(1 + \frac{N}{2} \leq t \leq T - \frac{N}{2} \right) \quad (2.2)$$

⁴ 我が国におけるインターネット治安情勢について (平成 18 年 1 月期)
<http://www.cyberpolice.go.jp/detect/pdf/20060214.pdf>

とおく。ここで、 \bar{y}_t 及び Δy_t は移動平均を用いて定義しているため、標本データ ($1 \leq t \leq T$) に対して、先頭と末尾の半周期分 ($1 \leq t \leq \frac{N}{2}, T - \frac{N}{2} + 1 \leq t \leq T$) については定義されない。

この \bar{y}_t が「トレンド成分」、 Δy_t が「周期成分 + ノイズ」を表している。

(2)トレンド成分の予測

トレンド成分 \bar{y}_t の N 時間先との差を $\Delta_N \bar{y}_t$ とする。

$$\Delta_N \bar{y}_t = \bar{y}_{t+N} - \bar{y}_t \quad \left(1 + \frac{N}{2} \leq t \leq T - \frac{3N}{2}\right) \quad (2.3)$$

この差が正規分布するものと仮定して、トレンド成分の変化率の予測値とその信頼区間を求める。

$\Delta_N \bar{y}_t$ の平均値を $\overline{\Delta_N \bar{y}}$ 、 $\Delta_N \bar{y}_t$ の標準偏差を $\sigma_{\Delta_N \bar{y}}$ とすると

$$\overline{\Delta_N \bar{y}} = \frac{1}{T - 2N} \sum_{i=1+\frac{N}{2}}^{T-\frac{3N}{2}} \Delta_N \bar{y}_i \quad (2.4)$$

$$\sigma_{\Delta_N \bar{y}} = \sqrt{\frac{(T - 2N) \sum_{i=1+\frac{N}{2}}^{T-\frac{3N}{2}} (\Delta_N \bar{y}_i)^2 - \left(\sum_{i=1+\frac{N}{2}}^{T-\frac{3N}{2}} \Delta_N \bar{y}_i \right)^2}{(T - 2N)^2}} \quad (2.5)$$

と表される。ここで $\Delta_N \bar{y}_t$ は N 時間の変化を示しているため、

$$a = \frac{1}{N} \overline{\Delta_N \bar{y}} \quad (2.6)$$

$$a_{CI} = \frac{1}{N} \left(\overline{\Delta_N \bar{y}} \pm 2\sigma_{\Delta_N \bar{y}} \right) \quad (2.7)$$

が、それぞれ 1 時間あたりのトレンド成分の変化の予測値とその信頼限界となる。

トレンド成分の最新の値 $\bar{y}_{T-\frac{N}{2}}$ 並びに予測した変化分 a 及び a_{CI} を用いて、「トレンド

成分の予測値」 $\hat{Y}_{trend}(t')$ と「トレンド成分の信頼限界」 $\hat{Y}_{trend-CI}(t')$ を

$$\hat{Y}_{trend}(t') = \overline{y_{T-\frac{N}{2}}} + a \cdot \left(\frac{N}{2} + t'\right) \quad (1 \leq t' \leq N) \quad (2.8)$$

$$\hat{Y}_{trend-Cl}(t') = \overline{y_{T-\frac{N}{2}}} + a_{Cl} \cdot \left(\frac{N}{2} + t'\right) \quad (1 \leq t' \leq N) \quad (2.9)$$

と定義する。

(3) 周期成分の予測

周期成分の振幅はトレンド成分の大きさに比例すると仮定し、トレンド成分 $\overline{y_t}$ で正規化した Δy_t の大きさ C_t を

$$C_t = \frac{\Delta y_t}{y_t} \quad \left(1 + \frac{N}{2} \leq t \leq T - \frac{N}{2}\right) \quad (2.10)$$

とおく。この C_t の周期毎の値 $\{C_t, C_{t+N}, C_{t+2N}, \dots\}$ が正規分布するものと仮定し、 C_t の予測値と信頼区間を求める。なお、ホワイトノイズは長期間の平均を取ると 0 になるため、この重ね合わせによりノイズが軽減されることとなる。

周期の数 M は $M = (T/N) - 1$ であり、 $\{C_t, C_{t+N}, C_{t+2N}, \dots\}$ の平均値を $\overline{C}(t')$ 、標準偏差を $\sigma_c(t')$ とすると

$$\overline{C}(t') = \begin{cases} \frac{1}{M} \sum_{i=1}^M C_{t'+Ni} & (1 \leq t' \leq \frac{N}{2}) \\ \frac{1}{M} \sum_{i=1}^M C_{t'+Ni-N} & (\frac{N}{2} < t' \leq N) \end{cases} \quad (2.11)$$

$$\sigma_c(t') = \begin{cases} \sqrt{\frac{M \sum_{i=1}^M (C_{t'+Ni})^2 - \left(\sum_{i=1}^M C_{t'+Ni}\right)^2}{M^2}} & (1 \leq t' \leq \frac{N}{2}) \\ \sqrt{\frac{M \sum_{i=1}^M (C_{t'+Ni-N})^2 - \left(\sum_{i=1}^M C_{t'+Ni-N}\right)^2}{M^2}} & (\frac{N}{2} < t' \leq N) \end{cases} \quad (2.12)$$

と表され、 C_t の予測値を $\hat{C}(t')$ 、 C_t の信頼限界を $\hat{C}_{Cl}(t')$ とすると、それぞれ

$$\hat{C}(t') = \overline{C}(t') \quad (1 \leq t' \leq N) \quad (2.13)$$

$$\hat{C}_{Cl}(t') = \overline{C}(t') \pm 2\sigma_c(t') \quad (1 \leq t' \leq N) \quad (2.14)$$

となる。

ここで、トレンド成分の予測値 $\hat{Y}_{trend}(t')$ 、信頼限界 $\hat{Y}_{trend-Cl}(t')$ と上記 $\hat{C}(t')$ 、 $\hat{C}_{Cl}(t')$ を用い、「周期成分の予測値」 $\hat{Y}_{cycle}(t')$ と「周期成分の信頼限界」 $\hat{Y}_{cycle-Cl}(t')$ を

$$\hat{Y}_{cycle}(t') = \hat{C}(t') \cdot \hat{Y}_{trend}(t') \quad (1 \leq t' \leq N) \quad (2.15)$$

$$\hat{Y}_{cycle-Cl}(t') = \hat{C}_{Cl}(t') \cdot \hat{Y}_{trend-Cl}(t') \quad (1 \leq t' \leq N) \quad (2.16)$$

と定義する。

(4) アクセス件数の予測

トレンド成分と周期成分の予測を合成し、本手法では「アクセス件数の予測値」 $\hat{Y}(t')$ と「アクセス件数の信頼限界」 $\hat{Y}_{Cl}(t')$ を

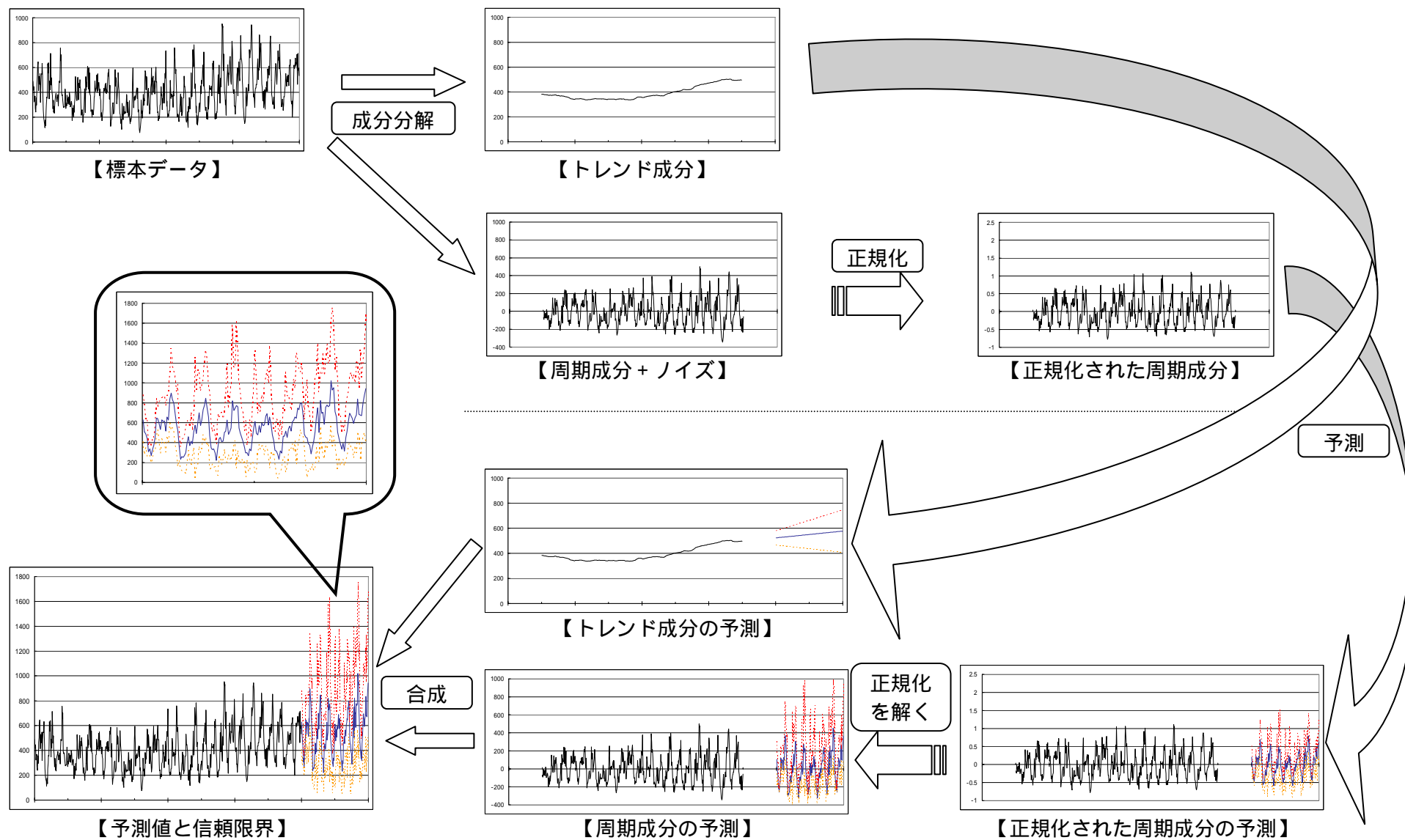
$$\hat{Y}(t') = \hat{Y}_{trend}(t') + \hat{Y}_{cycle}(t') \quad (1 \leq t' \leq N) \quad (2.17)$$

$$\hat{Y}_{Cl}(t') = \hat{Y}_{trend-Cl}(t') + \hat{Y}_{cycle-Cl}(t') \quad (1 \leq t' \leq N) \quad (2.18)$$

と定義する。

以上で述べた本手法の概要を図 2 に示す。

図2 本レポートにおける予測手法（概要）



3 予測の例

(1) 定常状態の例

定常状態における予測を例示するため、次の条件で予測を行った。

項目	条件
標本として使用する観測拠点	57 拠点
予測するポート	445/TCP
標本データ期間	平成 17 年 7 月 9 日～ 9 月 30 日(12 週間)
予測期間	平成 17 年 10 月 1 日～ 10 月 7 日(1 週間)

「2 予測手法」で述べた手法で予測した「予測値」及び「信頼限界」並びに実際の「観測値」を図3に示す。

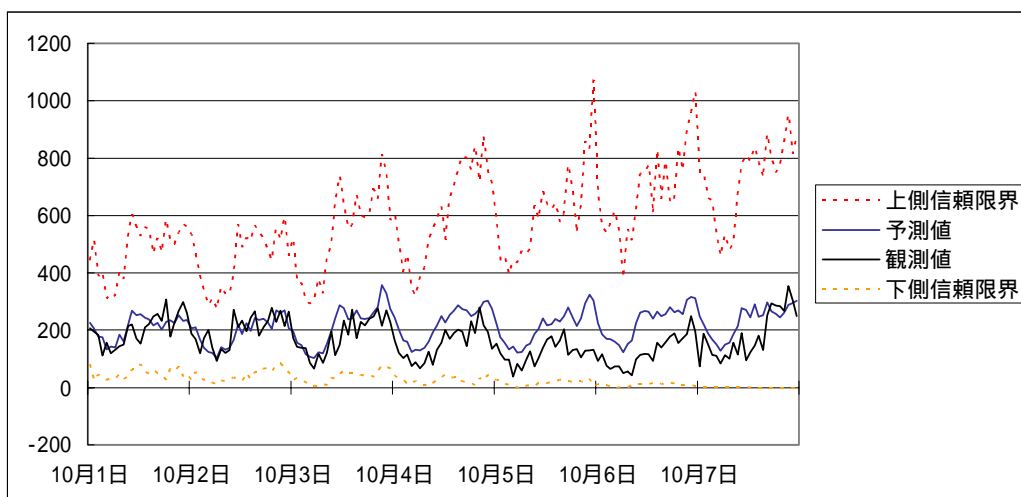


図3 定常状態の例

このグラフからわかるように、予測値と観測値が共に「アクセス件数が昼間に増加し、夜中に減少する」という似た推移を描いており、すべての点において観測値は信頼区間に収まっている。また、予測期間の後になればなるほど信頼区間が広がっているが、これは時間の経過毎にトレンド成分の信頼区間が広がっているためである。

(2) 異常状態の例

異常状態における予測を例示するため、次の条件で予測を行った。

項目	条件
標本として使用する観測拠点	57 拠点
予測するポート	445/TCP
標本データ期間	平成 17 年 5 月 22 日～ 8 月 13 日(12 週間)
予測期間	平成 17 年 8 月 14 日～ 8 月 20 日(1 週間)

異常状態として 8 月中旬の 445/TCP を選択した理由は、Zotob ワーム等の活動に起因すると推測されるアクセス急増があったためである。

「2 予測手法」で述べた手法で予測した「予測値」及び「信頼限界」並びに実際の「観測値」を図 4 に示す。

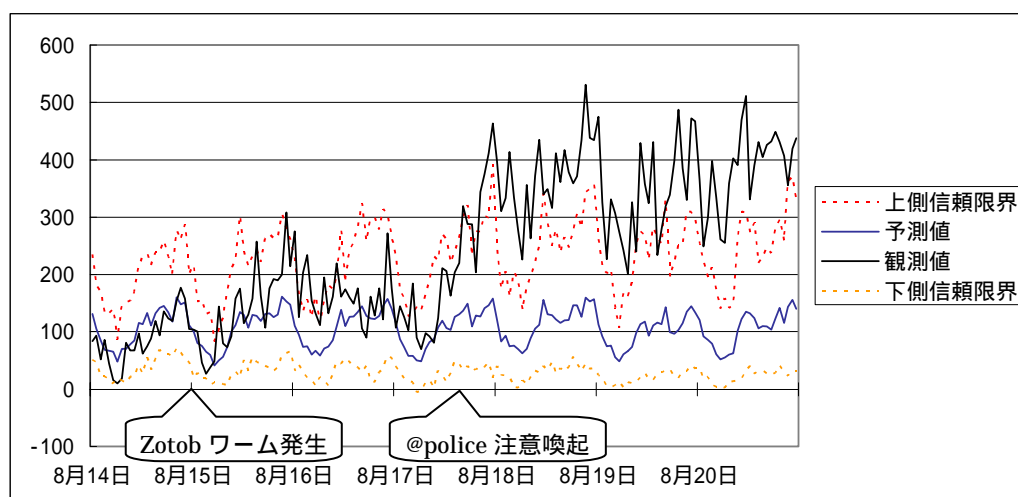


図 4 異常状態の例

このグラフからわかるように、Zotob ワームが発生した日本時間 8 月 15 日頃から 445/TCP に対するアクセス件数が増加し始め、15 日では 6,15,22 時台に、16 日では 0,2,3,4,7,10 時台に信頼区間から外れている。17 日 20 時以降については、ほぼすべての時点において観測値が上側信頼限界を越えており、異常なアクセス急増があったことがわかる。

次項で説明するが、この条件での信頼区間の信頼度は 97%以上と考えられるため、正常状態において 2 回連続で信頼区間から外れる確率は 0.09%以下である。そこで、閾値を「2 回連続で信頼区間外となった場合」として異常検出を試みた場合、16 日 4 時前の時点で異常を検知することができたこととなる。

この信頼度について次項で検討する。

4 信頼区間の信頼度

本手法による信頼区間の信頼度を検証するため、次の条件で予測を行い、そのそれぞれの結果について、実際の観測値が信頼区間内に収まっている割合（以下、ヒット率という。）を計算した。

項目	条件
標本として使用する観測拠点	57 拠点
予測するポート	上位 10 ポート 1
標本データ期間	予測期間の直前の 4 週間～25 週間
予測期間	1 週間 2

- 1 平成 17 年 4 月 1 日～9 月 30 日のアクセス件数の上位 10 ポート（135/TCP、445/TCP、139/TCP、1433/TCP、ICMP、4899/TCP、1434/UDP、1026/UDP、22/TCP、137/UDP）
- 2 平成 17 年 10 月 1 日～11 月 30 日のそれぞれを初日とする 1 週間

上記の条件により、予測するポート 10 種類、標本データ期間 22 種類、予測期間 61 種類のそれぞれの組合せである 13,420 通りについて予測を行った。条件としているポート・期間においては、特段の大きなインシデントは無かったため、これらのヒット率を信頼区間の信頼度とみなす。

その結果について、標本データ期間及び 1 時間あたりの平均アクセス件数によって集計した表を次頁に示す。なお、表中のヒット率は、同枠に入るヒット率を平均したものである。また、表中の空白欄は、該当する条件が無かったものであり、そのヒット率は不明である。

標本データ期間と平均アクセス件数によるヒット率の比較（57 拠点）

		標本データ期間																								
		4週間	5週間	6週間	7週間	8週間	9週間	10週間	11週間	12週間	13週間	14週間	15週間	16週間	17週間	18週間	19週間	20週間	21週間	22週間	23週間	24週間	25週間			
平均アクセス件数 / hour	～ 10	68.1%	71.5%	73.7%	75.9%	76.5%	75.4%	75.5%	75.6%	75.7%	76.0%	76.1%	76.2%	76.4%	76.6%	76.7%	76.8%	77.3%	78.2%	79.3%	79.5%	80.3%	80.9%			
	10～ 20	78.0%	86.6%	90.5%	92.4%	93.3%	94.5%	95.3%	95.8%	95.9%	96.1%	96.3%	96.4%	96.5%	96.7%	96.8%	96.7%	95.0%	92.4%	89.9%	90.0%	88.6%	87.6%			
	20～ 30	81.4%	87.2%	89.8%	91.8%	93.6%	93.9%	93.8%	95.7%	96.2%	97.0%	97.2%	97.5%	97.8%	97.9%	98.0%	98.3%	98.4%	98.5%	98.6%	98.6%	98.7%	98.7%			
	30～ 40	84.3%	89.7%	92.1%	93.7%	93.9%	94.7%	95.4%	95.1%	95.6%	96.0%	96.5%	96.8%	96.8%	96.8%	96.8%	96.8%	96.9%	96.8%	96.8%	96.9%	97.2%	97.1%			
	40～ 50	79.2%	86.5%	87.6%	93.7%	98.6%	99.7%	99.9%	99.5%	97.8%	96.7%	96.6%	96.8%	97.3%	97.5%	97.6%	97.6%	97.3%	97.3%	97.3%	97.3%	97.3%	97.1%	97.0%		
	50～ 60	80.8%	87.6%	91.8%	89.9%	91.2%	92.6%	92.2%	96.5%	98.3%	98.7%	99.7%	99.8%									100.0%	99.9%	99.7%		
	60～ 70	83.8%	93.5%	94.3%	93.9%	96.0%	97.9%	96.3%	88.9%	92.4%	97.9%	98.8%	98.1%	97.6%										100.0%		
	70～ 80		98.4%	96.8%	96.1%	96.8%	96.9%	98.9%	99.7%	99.5%	95.1%	95.6%	97.1%	98.2%	99.0%	99.1%	98.8%	98.5%	98.8%	99.0%	99.2%	99.3%	99.3%			
	80～ 90			99.4%	99.4%	99.0%	98.4%	98.3%	98.5%	99.2%	99.5%	99.3%	99.3%	99.3%	98.9%	98.9%	99.1%	99.5%	99.6%	99.7%						
	90～ 100					100.0%	99.9%	99.5%	99.3%	99.4%	99.4%	99.5%	100.0%	100.0%												
	100～ 110																									
	110～ 120																									
	120～ 130																									
	130～ 140																									
	140～ 150		96.3%	98.4%	97.4%																					
	150～ 160		76.0%	90.7%	96.7%	96.6%	99.3%	99.4%	99.3%																	
	160～ 170		76.3%	89.8%	97.1%	98.2%	96.8%	99.4%	100.0%	99.3%	98.2%															
	170～ 180		75.9%	92.4%	98.8%	98.9%	97.8%	96.7%	97.8%	99.9%	99.9%	99.3%	98.8%	99.3%	99.4%	99.2%	98.8%	98.8%	99.9%	99.7%	99.9%	99.8%	99.8%	99.9%		
	180～ 190		85.8%	89.1%	91.6%	92.0%	97.9%	98.3%	97.8%	96.9%	97.5%	98.3%	98.5%	98.5%	99.1%	99.2%	99.5%	99.8%	99.8%	99.9%	99.8%	99.9%	100.0%	99.9%		
	190～ 200		80.6%	83.4%	88.2%	93.7%	95.5%	97.8%	98.4%	98.2%	98.3%	98.5%	99.2%	99.5%	99.6%	99.8%	99.8%	99.5%	99.6%	99.8%	100.0%	99.4%	98.8%			
	200～ 210		59.8%	92.0%	96.1%	89.8%	92.8%	95.2%	91.2%	98.8%	99.7%	99.9%	99.5%	99.7%	99.8%	99.8%	99.6%	99.8%	99.3%	98.5%	99.9%	100.0%	99.4%			
	210～ 220		79.0%	89.3%	96.1%	97.9%	88.9%	88.7%	97.9%	97.0%	98.7%	99.9%	99.8%	99.4%	99.5%	99.8%	99.9%	99.5%	99.9%	99.2%	99.4%	99.8%	100.0%	99.1%		
	220～ 230		92.0%	94.9%	88.1%	94.2%	98.6%	98.6%	96.5%	98.5%	97.5%	98.4%	99.3%		100.0%	99.6%	100.0%	100.0%	99.8%	100.0%	97.8%	100.0%	100.0%	100.0%		
	230～ 240		83.9%	92.9%	94.6%	95.8%	97.9%	99.3%	96.5%	97.5%	99.0%				100.0%	100.0%	99.8%	100.0%	99.4%	99.8%	99.8%	97.8%	100.0%	100.0%		
	240～ 250		83.3%	90.2%	92.9%	97.0%	98.4%	99.1%								100.0%	99.7%	100.0%	100.0%	99.4%	100.0%	99.1%	98.5%	99.9%		
	250～ 260				90.5%	98.4%	98.4%									100.0%	100.0%	99.6%	100.0%	100.0%	99.4%	100.0%	98.8%	98.5%		
	260～ 270				95.4%	98.5%											100.0%	99.6%	99.9%	100.0%	99.9%	99.5%	100.0%	99.2%		
	270～ 280																	100.0%	99.5%	99.9%	100.0%	99.8%	99.6%	99.7%		
	280～ 290																		100.0%	99.6%	99.8%	99.8%	99.9%	99.9%		
	290～ 300																		100.0%	100.0%	100.0%	100.0%	99.9%	99.9%		
300～																					100.0%	100.0%	100.0%			

【凡例】

- 90%未満
- 90%以上 95%未満
- 95%以上 100%未満
- 100%

この表からわかるように、標本データ期間が長ければ長いほど、また平均アクセス件数が多ければ多いほど信頼区間の信頼度が高くなっており、予測に使用する標本データが多くなれば信頼度が高まるという結果であった。

平均アクセス件数が 100～140 件の結果が無いいため断定はできないが、おおむね「平均アクセス件数が 80 件以上であり、標本データ期間が 12 週間以上」での予測では、ヒット率が 97%以上となるものと思われる。

信頼区間の信頼度が 97%である場合、定常状態において 2 回連続で信頼区間から外れる確率は 0.09%であるため、閾値を「2 回連続で信頼区間外となった場合」として異常検出を試みることにより、高い精度で異常を検知することができる。

ただし、1 時間あたりの平均アクセス件数が 80 件以上という条件のため、本手法により異常検出を行うことができるのはアクセス件数が上位のポートに限られ、上の例では、上位 2 つの 135/TCP と 445/TCP のみであった。他のポートについては、標本データが 12 週程度の場合は十分な信頼度が得られないため、57 拠点の標本データによって異常検出を行うことは難しい。

各上位ポートに対する 1 時間あたりの平均アクセス件数を 80 件以上とするために必要な拠点数の目安は次表のとおりである。

	平均アクセス件数 (57 拠点合計)	平均アクセス件数 (1 拠点当たり)	必要拠点数
135/TCP	301.02	5.28	15.15
445/TCP	172.42	3.02	26.45
139/TCP	67.56	1.19	67.50
1433/TCP	51.91	0.91	87.84
ICMP	32.02	0.56	142.43
4899/TCP	26.73	0.47	170.59
1434/UDP	24.97	0.44	182.62
1026/UDP	10.70	0.19	426.15
22/TCP	10.04	0.18	454.32
137/UDP	8.98	0.16	507.80

この表は、平成 17 年 4 月 1 日から 9 月 30 日までの間の上位 10 ポートに対する 1 時間当たりの平均アクセス件数に基づき、1 時間当たりの平均アクセス件数を 80 件とするために必要な拠点数を「必要拠点数」として表したものである。この表からも、57 拠点では上位 2 ポートのみが異常検出可能であることがわかる。また、100 拠点程度では上位 4 ポート、200 拠点程度では上位 7 ポートの異常検出が可能であり、上位 10 ポートの異常検出のためには 500 拠点程度必要となる。

今回の手法では、信頼区間の信頼度を高めるために、式(2.7)及び(2.14)にあるように $\pm 2\sigma$ によって信頼区間を広く取っているが、予測値が0を下回ることもあり、手法の妥当性について今後の検討が必要である。また、今回は標本データ期間として過去半年程度のアクセス件数推移を使用した。更に長い期間のデータを使用した場合には信頼度を高めつつ信頼区間を狭めることができる可能性がある。これらのパラメータについては、状況に応じて適切な値を選定する必要がある。

5 まとめ

本分析レポートでは、ファイアウォールに対するアクセス件数について、過去のアクセス状況を標本データとし、トレンド成分と周期成分に分解してアクセス件数の推移を予測するという加法型の成分分解による手法について検討し、その信頼度の評価を行った。

この結果、1時間あたりの平均アクセス件数が80件以上である場合、12週間以上の推移を標本データとすれば、1週間先までの平常状態の予測を高い信頼度で行うことができることがわかった。

本手法の目的である長期間の予測では、アクセス件数が上位のポートに限られるものの、予測と実際の観測値を比較することにより、ワーム発生等に起因するアクセス増加等の異常状態を検出することが可能であった。

他に異常検出手法として、過去のデータから1時間先のアクセス件数を予測して観測値と比較したり、アクセス件数の変動をRSI等の指標で表したりする方法も考えられるが、本手法では、「3(2) 異常状態の例」に例示したように、予測によって求めた定常状態と異常状態のアクセス件数の推移を比較することができ、異常状態を視覚的に理解しやすいという利点があり、異常状態を分析する際にも有効である。

今回の検証で、本手法によってある程度信頼度の高い定常状態の予測を行うことができることがわかったが、今回使用したパラメータ等は当センターにおけるアクセス状況に基づいたものであり、状況に応じて適切な値を選定する必要がある。

当センターでは、ワームの発生や攻撃活動等を迅速、効率的に調査するため、今後も様々な手法について検討し、国民へのタイムリーな情報提供に向けて取り組んでいく。

参考文献

- (1) 山本 拓 (1988) 『経済の時系列分析』創文社 (現代経済学選書2)
- (2) 北川 源四郎 (2005) 『時系列解析入門』岩波書店
- (3) P.J. ブロックウェル、R.A. デービス (2004) 『入門時系列解析と予測 (改訂第2版)』シーエーピー出版
- (4) 石村 貞夫、ステファニー・リヒャルト (2002) 『Excel でやさしく学ぶ時系列』東京図書
- (5) 損害保険料率算出機構 『過去のデータから将来をどう予測するか』
<http://www.nliro.or.jp/disclosure/risk/risk67-2.pdf>