過去の警告ログから判定される異常な警告イベントを強調するIDSの可視化手法に関する２，３の考察

木村 知史 †　　稲葉 宏幸 †

† 京都工芸繊維大学大学院工芸科学研究科
606-8585 京都市左京区松ヶ崎橋上町
kimura08@sec.is.kit.ac.jp, inaba@kit.ac.jp

あらまし 近年，ネットワークへの不正アクセスによる被害が増大し，深刻な問題となっている．その対策として，侵入検知システム (IDS) の重要性が高まっている．IDS とは，ネットワーク上への不正なアクセスの兆候を検知し，システム管理者に通報するシステムである．IDS の問題点として，大規模なネットワークのアクセスを監視すると，膨大な警告情報が出力され，解析に多大な時間と労力が必要となることが挙げられる．この問題を解決するために，過去の傾向に基づくログ分析手法や，ログの可視化などの手法が知られている．本論文では，過去の傾向に基づくログ分析手法を用いて，異常な警告イベントを強調する可視化システムを提案する．

Notes on IDS Visualization System that Emphasize the Anomalous Warning Events based on Past Tendency

Satoshi Kimura†　　Hiroyuki Inaba†

†Kyoto Institute of Technology
Hashigami-tyou, Matsugasaki, Sakyo-ku, Kyoto-shi, Kyoto 606-8585, JAPAN
kimura08@sec.is.kit.ac.jp, inaba@kit.ac.jp

Abstract Recently, illegal access to the network is increasing. It has been a serious problem. To overcome this problem, necessity of Intrusion Detection System(IDS) is increasing. IDS is the notifying system of network manager to inspect symptoms of the illegal access. The problem of IDS is tremendous warning logs especially for large scale network. Analyzing these logs apply a large amount of load to a network manager. To overcome this problem, there exist several methods for analyzing logs based on past tendency and some visualization methods for the logs. In this paper, we propose the visualization system that emphasizes the anomalous warning events.

1　はじめに

近年，インターネットの急激な発展に伴い，ネットワークへの不正アクセスによる被害が増大し，深刻な問題となっている．その対策として，侵入検知システム (Intrusion Detection System, 以下 IDS) の重要性が高まっている．IDS とは，ネットワーク上の不正なアクセスの兆候を検知し，システム管理者に通報するシステムである．システム管理者に通報する方法として，警告情報をログとして出力する方法が一般的である．しかし，IDS の問題点として，ある程度

の規模のネットワークを監視すると，膨大な警告情報が出力され，解析に多大な時間と労力が必要となることが挙げられる．この問題を解決するために，過去の傾向に基づくログ分析手法 [1][2][3] や，ログの可視化などの研究が行われている [4][5][6][7].

本論文では，過去の傾向に基づくログ分析手法を用いて，異常な警告イベントを強調する可視化システムを提案する．提案手法は，異常と判定された警告イベントを強調して表示することにより，システム管理者の負担を従来より軽減できると考えられる．なお，本研究における IDS のソフトウェアとして，オープンソースのシグネチャマッチ型 IDS である Snort[8] を用いている．

2　関連研究

本章では，本研究に使用したログ分析手法および視覚化手法の関連研究について述べる．

2.1　変動係数を用いたログ分析手法

検知ログの分析に関する研究として，管理者の主観に頼らない変動係数を用いたログ分析手法 [1] が存在する．

文献 [1] では，対象とする警告イベントごとに，過去の期間 T_1，T_2 の検知数を比較し，T_2 の検知数に対する T_1 の検知数の増減の割合が，あるしきい値を超えた場合に警告を出すという手法を用いている．この T_1，T_2 は集約期間 T と呼ばれる期間で，過去 3ヶ月のログを分析し，動的に決定される．この手法のイメージを図 1 に示す.

過去 3ヶ月のログを分析する手法として，検知数のばらつきを見るために，集約する期間ごとに検知数の標準偏差を求める．平均検知数は各警告イベントによって大きく異なるため，標準偏差そのものをばらつきの尺度として用いることは適当ではないと考えられる．そこで，平均値の影響を除いたばらつきの尺度として，標準偏差を平均検知数で正規化した値を用いている．これにより，各警告イベントの検知数の相対的なばらつきを見ることができる．この尺度が変動係数とよばれるものである．集約期間 T（hour）における変動係数を $C_v(T)$，標準偏差を $S_d(T)$，平均検知数を $\overline{X}(T)$ とするとき，変動係数は以下の式で定義されている．

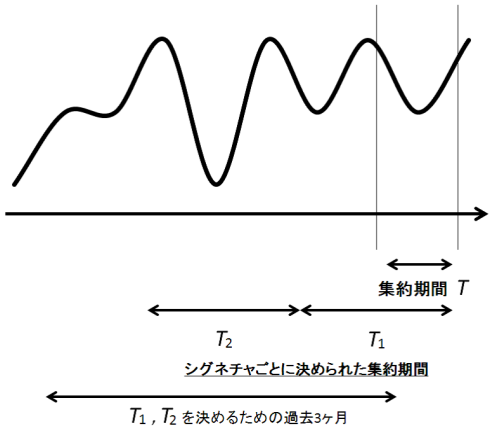

図 1: T_1, T_2 を決定するイメージ図

$$C_v(T) = \frac{S_d(T)}{\overline{X}(T)} \tag{1}$$

一般に変動係数の値の変化量は，集約期間が長くなるほど小さくなっていくと考えられる．そこで式 (1) を用いて，集約期間 T（hour）における増減率 $R_d(T)$ を以下の式で定義している．

$$R_d(T) = \left| \frac{C_v(T+1) - C_v(T)}{C_v(T)} \right| \tag{2}$$

期間 T_1，T_2 の値として，増減率 $R_d(T) < R_{th}$ となる期間 T を用いる．なお R_{th} は任意の値を設定できるが，文献 [1] では 0.01 としている．

期間 T_1，T_2 での検知数をそれぞれ N_1，N_2 とし，増減の割合 F を以下の式で定義する．

$$F = \frac{N_1}{N_2} \tag{3}$$

F の値がしきい値 T_h を超えたとき，異常な検知数であると判断し，警告を出す.

変動係数の値は相対的なばらつきを表すものであるので，変動係数の値が大きい警告イベントほどしきい値を上げることが望ましいと考えられる．そこでしきい値 T_h は，以下の式で定義されている．

$$T_h = \delta \times C_v(T_1) \quad (4)$$

ただし $T_1 = T_2$ とし，また δ は警告の出しやすさを決める定数であり，警告イベントによらず一定としている．

実験結果として，データ取得期間を過去 3ヶ月とし，各警告イベントの変動係数を求めた結果，その値はおおよそ 1.0 程度に収まるということが判明している．式 (4) によるしきい値で用いられている定数 δ を 1.1 から 0.1 刻みで 1.5 まで変化させた結果，過去の傾向から異常な検知数かどうかを判断することに成功している．

文献 [1] では，管理者の主観に頼らない変動係数を用いることにより，しきい値の設定を動的に決定することを最大の利点としている．一方，問題点として，集約期間 T を決定するための過去のデータ取得期間 3ヶ月における検知数が安定しているという前提条件が必要であることが挙げられる．

2.2 その他のログ分析手法に関する研究

その他の IDS に関する既存手法として，文献 [2] では，実運用されている IDS のログを用いて観測値から理論統計分布を求めている．従来までは単位時間中に検知されるイベントの頻度にのみ注目していたが，新たに警告イベント毎の到着間隔および継続回数の計 3 つのパラメータに注目することにしている．その結果，頻度に関してはポアソン分布にモデル化でき，到着間隔ならびに継続回数については指数分布にモデル化できることが示されている．実験結果として，一見してランダムに検出されるイベントが理論的な統計分布に従うことがわかり，この分布モデルから外れるイベントを異常として検出する手法の妥当性が確認されている．

文献 [3] では，IDS のログを分析するために，従来，頻度分析が用いられているが，それに加えて出力されるイベントの変化量に注目した分析手法を提案している．これは，出力数が様々に変動するイベントの中から，異常に変化したイベントを検出するために，過去の長期間の出力特性と，最近の短期間の出力特性についての平均値の比較を行う比率分析と，稀率分析を用いる手法である．実験結果として，従来からの頻度分析結果の中から不要なイベントを特定でき，さらに発見が難しい頻度値が小さいながらも急激に増加するイベントを特定できることが確認されている．

2.3 Hashing Alert Matrix

送信元 IP アドレスおよび送信先 IP アドレス情報を一つの画面で把握するために，横軸が送信元 IP アドレス，縦軸が送信先 IP アドレスの 2 次元平面を考える．警告の IP アドレスに対応する位置に点をプロットすることで，警告の有無を表示する．警告の種類は点の色で，警告の量は点の色の濃さで表現する．

しかし，この手法の問題点として，一般的なモニタの解像度では，異なる複数の点が同一の点としてプロットされてしまうため，警告状況を正確に把握することができない．これは，限られた表示解像度に対して 32 ビットの全 IP アドレス空間の表示を試みているからである．そこで，ハッシュ関数を用いることにより，限られた解像度においても，描画領域を有効に活用できるように対策を施したシステムが Hashing Alert Matrix[4] である．

Hashing Alert Matrix は，IP アドレス情報をそのまま用いるのではなく，ハッシュ関数を用いることで，IP アドレス情報の下位 1 ビットでも異なると IP アドレス情報の数字が大きく変わり，プロットされる点の位置が変わるという性質を用いた手法である．IP アドレス値をモニタ上の座標値に変換する際に，IP アドレス値をハッシュ関数に入力して得られたハッシュ値を座標値 (横軸ω(ドット)，縦軸 h (ドット)) とする．すなわち，IP アドレス値 x から縦軸と横軸のハッシュ値を取得し，ハッシュ関数をそれぞれ $H_\omega(x)$ と $H_h(x)$ としたとき，送信元 IP アドレスが x，送信先 IP アドレスが y である警告は，モニタの座標位置 $H_\omega(x)$ と $H_h(y)$ にプロットされる．このため，プロットされる点は分散し，一般的に使用される限られた解像度においても 1500 個程度までの検出数であれば，衝突する点は数個以下であることが報告されて

いる．以上の方法によって，限られた描画領域を有効に活用し，検知ログを直感的に把握することが可能となる．

2.4 その他の可視化に関する研究

その他のIPアドレスに関連する可視化の研究として，平安京ビュー [5] やIDS RainStorm[6]，UnEqual Scaling Alert Matrix[7] が存在する．

平安京ビュー [5] は，IPアドレスの階層構造に着目した検知ログの視覚化手法である．IPアドレスの階層構造を利用することによって，観測されたすべてのIPアドレスを二次元平面上に配置することが可能であり，一つの画面で検知状況の全体的な把握が可能である．

IDS RainStorm[6] は，IPv4におけるクラスBを利用し，32ビットのIPアドレス空間の個数を減らすという点に着目した検知ログの視覚化手法である．送信元IPアドレスと送信先IPアドレスの関係を時間軸に沿って把握することができるという特徴も有している．

UnEqual Scaling Alert Matrix[7] は，送信元IPアドレスおよび送信先IPアドレスのオクテット毎の検知数によりプロット位置を決定するという点に着目した可視化手法である．IPアドレスから画面に点をプロットする際に，IPアドレスの階層構造に着目し，検知ログに含まれるIPアドレスのオクテット毎の割合によって，不均一に点のプロットを行っている．この手法を用いることにより，IPアドレス階層構造を失うことなく，描画領域を最大限有効に活用することが可能になっている．

それぞれの研究手法に共通する項目は，広大なIPアドレス空間を限られた解像度において，如何にして直感的に情報を把握するかという点にある．しかし，全ての情報を均一に扱って描画するのではなく，統計的手法を用いて，特定の警告イベントを特に強調して描画するという手法は我々が知る限り提案されていない．

3 異常警告イベント強調型視覚化手法

本章では，管理者の主観に頼らない変動係数を用いたログ分析手法 [1] と Hashing Alert Matrix[4] を組み合わせ，それぞれに改良を加えた上で，異常な警告イベントを強調する視覚化手法を提案する．

3.1 変動係数を用いたログ分析の改良手法

本節では，変動係数を用いたログ分析の改良手法について述べる．2.1節で述べたように文献 [1] では，式 (3) が式 (4) のしきい値を超えたときに異常な警告だと判定している．しかし，変動係数は式 (1) で定義されているように，標準偏差を平均検知数で正規化したものである．それに定数 δ を乗じたしきい値と，式 (3) で定義される量とを比較することは必ずしも適切ではないと考えられるため，本研究では式 (3) の代わりに以下の式 (5) で定義される量を用いる．

$$F = \frac{(N_1 - N_2)}{N_2} \tag{5}$$

上式において N_1，N_2 は2.1節で定義したものである．式 (5) は，直近の警告数の増減量を期間 T_2 の検知数 N_2 で正規化する形になっており，その値はおおよそ変動係数の値と同程度になることが期待できる．なお，異常な警告であるかどうかの判定は，2.1節と同様に式 (5) の値 F が以下の式 (6) のしきい値を超えた場合に異常な警告であると判定する．

$$T_h = \delta \times C_v(T_1) \tag{6}$$

なお，N_1 の個数が N_2 の個数よりも少なく検出された場合にも異常な警告である可能性があるが，本研究では警告数が減少することは異常な状態ではないとみなしている．

3.2 Hashing Alert Matrixの改良手法

本研究では，視覚化システムとして，Hashing Alert Matrixを参考にし，それに改良を加えた

ものを用いている．改良点として，3.1 節で述べた変動係数を用いたログ分析手法により異常であると判定された警告イベントは，5 × 5 ドットの四角形で大きめにプロットし，異常であると判定されなかった警告イベントは 3 × 3 ドットの円形で小さめにプロットすることにした．また，異常であると判定された警告イベントは赤色で，プライオリティが 1 の警告イベントは黒色で，2 の警告イベントは緑で，3 の警告イベントは青で表現することで，色覚的にも注意を促すようにしている．なお描画領域は 800 × 800 ドットとしている．また，時間的な流れを把握するために，1 時間ごとに画面を連続的に描画できるようなアニメーション機能も追加している．

作成した異常イベント強調型視覚化システムと，その一部の拡大図を図 2 に示す．総検知数は 8982 件で，異常警告イベントが強調されて描画されているのがわかる．プロット点をクリックすることで，その警告イベントの詳細情報を端末上に表示することも可能であり，プロット点が重なっている場合は，その情報を全て端末上に表示するように実装している．

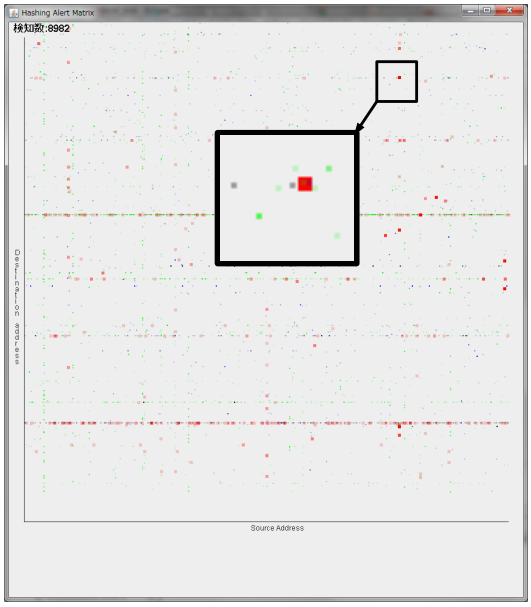

図 2: 異常イベント強調型視覚化システムと，その一部の拡大図

4　実環境における性能評価

本章では，まず提案手法の性能を実環境において評価するための実験条件について述べる．次に，異常警告イベント数に関する評価と，異常警告イベント数の視覚化への影響についての考察を行う．

4.1　実験条件

提案手法の実装のために，プログラミング言語である Java を用いた．また性能評価のために必要な検知データは，本学のキャンパスネットワーク環境を用いて得られたデータを用いた．Snort から出力されたログは MySQL データベースへと保存されるので，そのデータを用いて評価を行う．データ取得期間は，2012 年 5 月 1 日から 5 月 7 日までを使用する．期間 P は 3ヶ月とし，集約期間 T は 1 時間単位としている．

4.2　異常警告イベント数に関する評価

式 (6) で定義されるしきい値を決める定数 δ の値を 1.1 から 1.5 まで 0.1 刻みで変更し，異常であると判定された警告イベント（以下，異常警告イベント）の数の減少の推移を示したグラフを図 3 に示す．図 3 より，定数 δ の値が大きくなるほど異常警告イベント数が減少しているのがわかる．なお，過去 3ヶ月間における検知ログから集約期間 T を求める際に，検知数が非常に少ないシグネチャの集約期間は統計的に意味がないと考えられるので，1 時間に 1 個以上検出されたシグネチャ（以下，対象シグネチャ）のみについて集約期間を決定している．本実験では対象シグネチャ数は計 25 個であった．

25 個の対象シグネチャの内，比較対象として適切であると考えられる 5 つの対象シグネチャについての詳細を表 1 に示す．なお，比較項目として重要視したものは，以下の 4 つの項目である．

- 変動係数 $C_v(T)$

表 1: 5 つの対象シグネチャについての詳細

番号	平均検知数	集約期間 T	$C_v(T)$	$\delta = 1.1$	$\delta = 1.2$	$\delta = 1.3$	$\delta = 1.4$	$\delta = 1.5$
7	3.8	10	0.182	68	56	42	29	23
8	32.4	11	0.161	577	542	385	358	307
11	5.7	7	1.558	639	631	611	601	601
22	2.2	15	1.279	449	447	424	409	408
89	10.2	25	1.880	0	0	0	0	0

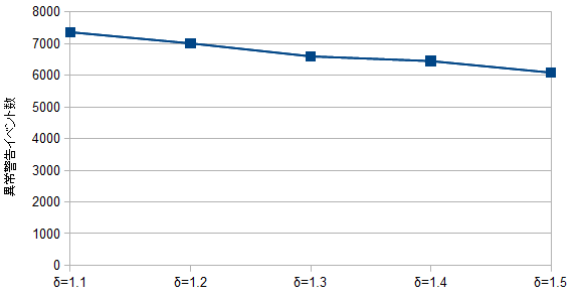

図 3: δ の値による異常警告イベント数

- パラメータ δ の値による異常警告イベント数の減少割合
- 平均検知数
- 集約期間 T

表 1 における平均検知数とは，対象シグネチャの 1 時間における平均検知数である．T は式 (2) により求めた集約期間であり，$C_v(T)$ はその集約期間における変動係数の値を示している．

なお，表 1 における番号は，シグネチャの種類を区別するためのものである．対応するシグネチャの名称を表 2 に示す．

表 1 において，変動係数の値が小さなシグネチャは δ の値が大きくなると異常警告イベント数が大幅に減少するが，逆に変動係数の値が大きなシグネチャは δ の値が大きくなっても異常警告イベント数があまり減少しないことがわかる．これは変動係数の値が小さいものほど過去 3ヶ月の間で安定して検出されているシグネチャであり，変動係数の値が大きいものほど過去 3ヶ月の間で，短時間で大量に検出されるような不安定なシグネチャであるからだと考えられる．

表 2: 各シグネチャの名称

シグネチャの名称
7 : ATTACK-RESPONSES 403 Forbidden
8 : WEB-MISC robots.txt access
11 : DNS SPOOF query response with TTL of 1 min. and no authority
22 : SNMP request tcp
89 : ICMP superscan echo

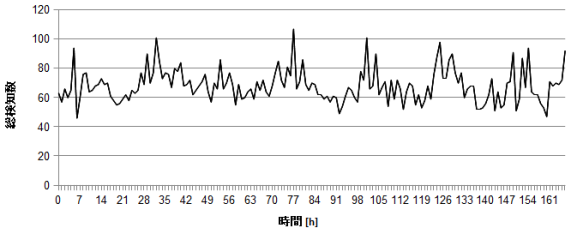

図 4: シグネチャ番号 8 の検知数の推移

データ取得期間における 1 時間単位でのシグネチャ番号 8 の検知数の推移を求めたグラフを図 4 に，検知数の割合 F の推移を図 5 に示す．

さらに，シグネチャ番号 11 の検知数の推移を求めたグラフを図 6 に，検知数の割合 F の推移を図 7 に示す．なお，図 5，図 7 における破線は各シグネチャに対して決定されたしきい値 T_h を示している．

シグネチャ番号 8 の変動係数は比較的小さく，図 4 に示されているように，イベントが継続して検出されていることがわかる．また，シグネチャ番号 11 の変動係数の値は比較的大きく，図 6 に示されているように，イベントが断続的に検出されていることがわかる．なお，その他の

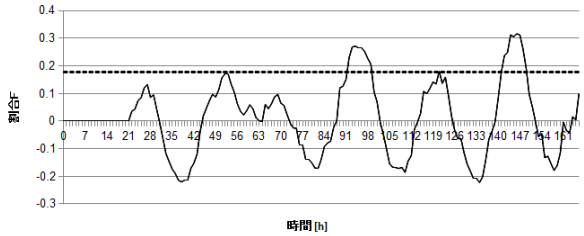

図 5: シグネチャ番号 8 の検知数の割合 F の推移

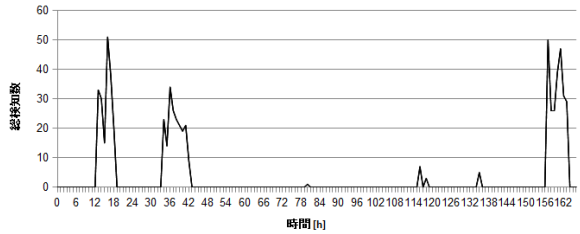

図 6: シグネチャ番号 11 の検知数の推移

対象シグネチャにおいても同じような傾向が見られた．

一方，シグネチャ番号 89 は変動係数の値が比較的大きく，過去 3ヶ月の間での平均検知数が 10.2 となっているのに対し，異常警告イベント数は δ が最小の 1.1 の場合においても 0 であった．なお，データ取得期間におけるシグネチャ番号 89 の総検知数は 1046 である．データ取得期間における 1 時間単位でのシグネチャ番号 89 の検知数の推移を求めたグラフを図 8 に示す．また，シグネチャ番号 89 の検知数の割合 F の推移を図 9 に示す．なお，しきい値を定めるパラメータ δ は図 5，図 7，図 9 のいずれにおいても最小の 1.1 としている．

図 8 において，シグネチャ番号 89 における時間変化の検知数の特徴は，7(h) から 20(h) までの間に多くの検知がされており，その後 40(h) 前後で 100 個程度の検知がされている点である．

表 1 において，シグネチャ番号 89 の集約期間 T は 25(h) と長く，データ取得期間における初期の段階で多くの検知がされている．そのためパラメータ δ が小さな値であったとしても，異常警告イベントだとは判定されなかったのだと考えられる．実際，図 9 を見ると，シグネチャ

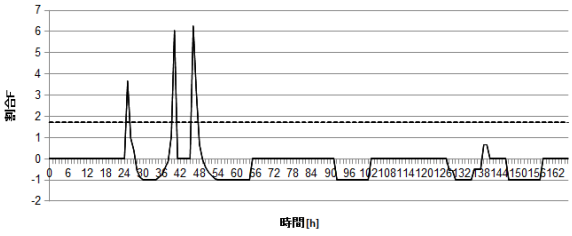

図 7: シグネチャ番号 11 の検知数の割合 F の推移

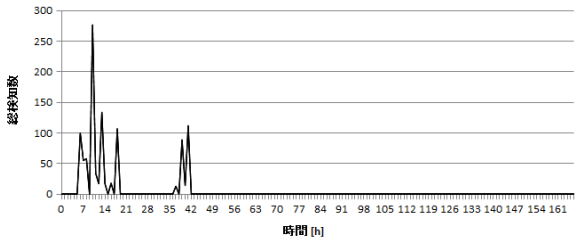

図 8: シグネチャ番号 89 の検知数の推移

番号 89 の検知数の割合 F が，それに対応するしきい値を上回ることがなかったことが示されている．

4.3 異常警告イベント数の視覚化への影響

パラメータ δ の値が 1.1 における視覚化を図 10 に示す．

図 10 より，データ取得期間における異常警告イベントが強調されてプロットされていることがわかる．本実験ではデータ取得期間が 7 日間と比較的長期間で実験を行っているが，異常イベント強調型視覚化システムをさらに有効に活用するには，データ取得期間を短くするか，1 時間ごとに画面を連続的に描画できるアニメーション機能を活用することが有効であると考えられる．

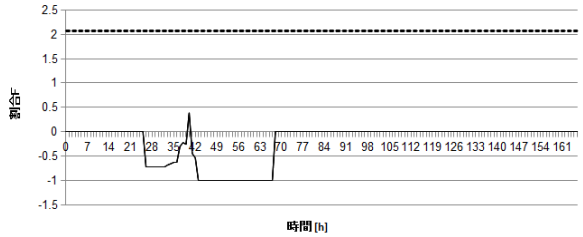

図 9: シグネチャ番号 89 の検知数の割合 F の推移

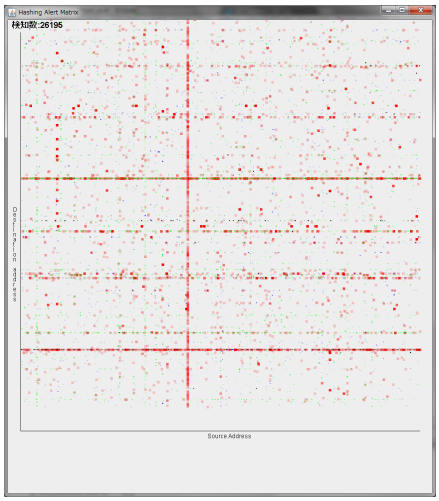

図 10: $\delta = 1.1$ における視覚化

5 おわりに

本論文では，侵入検知システムの視覚化手法において，過去の傾向から判断して特に異常な動きがあった警告イベントの強調を行うことにより，システム管理者の負担を減らすことを目的としている．

提案手法では，異常警告イベントの検出に必要なログ分析手法においてしきい値と比較をする評価値の算出法を改良することにより，適切に異常警告イベントを判定できるようにした．また，視覚化システムについても従来知られている Hashing Alert Matrix の特徴を生かしつつ，異常警告イベントの色や形，大きさを変えることにより，危険性が高いと思われる警告イベントに対してシステム管理者がより注意を向けられるようにしている．

今後の課題として，リアルタイムでの処理に対応するようにシステムに改良を加えると同時に，視覚化システムのユーザーインターフェースの改良を図る必要がある．

参考文献

[1] 戸田剛司，稲葉宏幸:“ 警告イベントの傾向に基づく IDS のログ分析手法に関する考察 ”，信学技報，SITE2010-7，pp.7-12 (2010).

[2] 竹森敬祐，三宅優，田中俊昭，笹瀬巌:“ 攻撃イベント数に関する調査および理論統計分布へのモデル化 ”，電子情報通信学会技術研究報告，vol.103，no.691，pp.20-27，Mar. 2004.

[3] 竹森敬祐，三宅優，中尾康二，菅谷史昭，笹瀬巌:“ Security Operation Center のための IDS ログ分析支援システム ”，電子情報通信学会論文誌，vol.J87-A，no.6，pp.816-825，Jun. 2004.

[4] 李　利，稲葉宏幸，若杉耕一郎，“ 個々の警告イベントを識別可能な侵入検知システムの二次元視覚化手法に関する考察 ”，画像電子学会誌，Vol.40，No.2，pp.369-376 (2011)

[5] T. Itoh, H. Takakura, and K. Koyamada, “ Hierarchical visualization of network intrusion detection data, ” IEEE Computer Graphics Applications, vol.26, no.2, pp.40-47, March/April 2006.

[6] I.R.V.I. Alarms, “ IDS RainStorm: Visualizing IDS Alarms ”, `http://www.cc.gatech.edu/john.stasko/papers/vizsec05.pdf`, 2012/08/18 参照

[7] 溝口峻一，稲葉宏幸，“ ＩＰアドレスの順序関係を考慮した侵入検知システムの 3 次元視覚化手法の提案，” 信学技報，vol.111，no.125，pp.19-24，July 2011.

[8] “ Snort ”, `http://www.iwsec.org/css/2012/`, 2012/08/18 参照