ウェーブレット解析によるIDSログの中長期的な傾向調査に関する試み

木村　知史†　　稲葉　宏幸†

†京都工芸繊維大学
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 インターネットの普及に伴い，不正アクセスの被害が増大している．不正アクセスの被害を防止するための手段の一つとしてIDSが知られている．IDSは危険性のあるパケットを検知して，管理者に通報するシステムである．しかしIDSは一般的に大量の検知アラートが発生するという問題が存在する．著者らは短期的な検知アラート系列が周波数領域でどのような検知傾向を持っているのかを調査した結果を報告している．本報告では，従来の解析結果に加えて，さらに長い期間における検知アラート系列がどのような検知傾向を持っているのかを調査し，さらに国別の時差を考慮した検知傾向の調査結果を報告する．

Examination of Long Term IDS Log Analysis by Wavelet Transform

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 With the spread of the Internet, the illegal access to the network is increasing. To overcome this problem, various technologies are known. IDS is the one of those technologies. IDS detects the packets including danger information and notifies it to a network operator. However, IDS has a critical problem that IDS generally outputs a tremendous number of logs. In our previous work, we reported some periodic trend in the detected alerts by discrete wavelet transform(DWT). In this paper, we report the results that we investigate some long term trend and a time difference by country in the detected alerts.

1　はじめに

近年，インターネットの普及に伴い，サイバー攻撃が激化している．サイバー攻撃によって不正侵入や情報漏えいの被害が増大しており，攻撃を事前に防止する技術，または攻撃後の情報漏えいの早期発見が可能な技術が必要とされている．その技術の一つとして，Intrusion Detection System(IDS)が広く用いられている．しかしIDSは一般に大量のアラートが発生するという問題が存在する．大量のアラートの中から本当に危険な攻撃であると考えられるアラートを検出することは困難であり，ネットワーク管理者の大きな負担にもなる．これらの大量に検知されるアラートは，一見してランダムに検知されているように見えるが，長期間にわたって観測するとアラート種別ごとに特徴的な傾向を

示している場合がある．著者らは，従来より検知アラートの時系列データに対して離散ウェーブレット変換による周波数解析を行うことにより，大量に検知されるアラートの中から特徴的な周期傾向を持つアラートを抽出することを試みている [1]．従来の研究では一見して周期性を持つことが明らかなシグネチャに対して，一週間という期間を定めて周波数解析を行った．結果として，一見して周期性を持つことが明らかなシグネチャについては，各レベルに対応した周波数成分においても周期性を見出すことができ，従来より正確かつ客観的に周期性を持っていることと，その周期を確認することができた．また，一見すると周期性を持たないようなシグネチャについても，各周波数レベルを局所的に観測することにより周期的に変動している箇所を見出すことができ，曜日や時間帯に依存する何らかの周期性等，これまで見逃されていた傾向を発見できる可能性を見出すことができた．

本稿では従来から行ってきた研究をさらに発展させ，一ヶ月程度のさらに長い期間における周期性を離散ウェーブレット変換によって見出すことができるかを試みた．また，従来ではシグネチャ名と検知数，検知時刻に着目して周波数解析を行っていたが，さらに IP アドレス値による国別の情報を付加することにより，新たな周期性を見出すことができるか調査した結果を報告する．

なお，本研究では IDS 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[2] を用いている．

2　ウェーブレット変換による周波数解析

[1] で報告したように，本学キャンパスネットワークで監視している IDS により検知されるアラートの中で，一見して明らかに周期性を持つアラートと，一見して周期性を持たないアラートが存在することが判明した．それぞれの例を図 1，図 2 に示す．

それぞれの時系列データに対して離散ウェーブレット変換による周波数解析を行った結果，

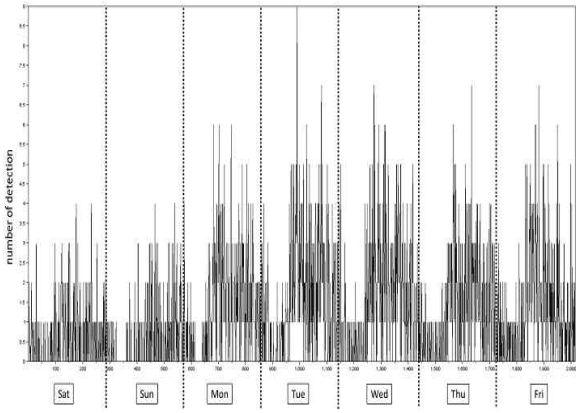

図 1: 一見して明らかに周期性を持つアラート

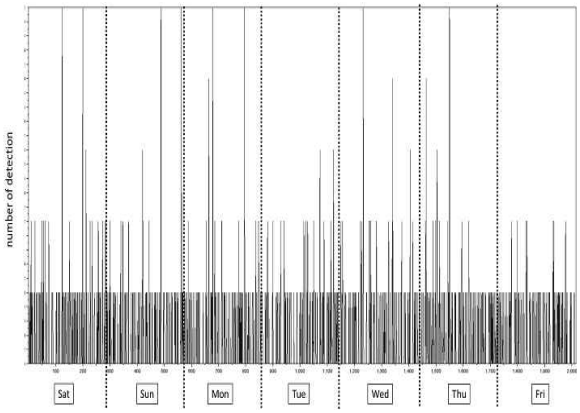

図 2: 一見して周期性を持たないアラート

周波数領域においても何らかの周期性を見出すことができた．

2.1　ウェーブレット変換

周波数解析法として広く用いられているフーリエ変換では基底関数として周期関数である三角関数を用いているため，変換後は時間領域の情報は失われてしまう．一方，ウェーブレット変換では，基底関数として局在化している関数を用いるために，変換後も時間領域の係数をある程度残したまま周波数解析を行うことが可能である．これは，ある時系列データに対してウェーブレット変換を施した場合に，どの位置にどの周波数成分が存在しているかを調べることが可能であることを意味している．ウェーブレット変換では基底関数を拡大縮小，または移動させ

ることによって時間周波数領域における効率的な解析を行うことが可能である．

以下式 (1)，(2) にウェーブレット変換と，逆ウェーブレット変換の式をそれぞれ示す．

$$(W_\psi f)(b,a) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{a}} \overline{\psi\left(\frac{x-b}{a}\right)} f(x)dx \quad (1)$$

$$f(x) = \frac{1}{C_\psi} \int\int_{\Re^2} (W_\psi f)(b,a) \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right) \frac{dadb}{a^2} \quad (2)$$

ここで，ψ はマザー・ウェーブレットと呼ばれる基底関数である．変数 a は ψ を拡大縮小させる変数であり，変数 b は ψ を平行移動させる変数である．式 (1)，(2) は連続なウェーブレット変換，逆変換の定義である．これを離散化するために，整数 j,k によって座標 $\left(b, \frac{1}{a}\right) = (2^{-j}k, 2^j)$ とおいて離散化を行うと，式 (1) は

$$d_k^{(j)} = 2^j \int_{-\infty}^{\infty} \overline{\psi(2^j x - k)} f(x)dx \quad (3)$$

となり，逆変換の式 (2) は

$$f(x) \sim \sum_j \sum_k d_k^{(j)} \psi(2^j x - k) \quad (4)$$

となる．

また，式 (4) の右辺に現れる 2 乗和の一方を

$$g_j(x) = \sum_k d_k^{(j)} \psi(2^j x - k) \quad (5)$$

と書き，また式 (4) を式 (5) における $g_j(x)$ を用いて

$$f_j(x) = g_{(j-1)}(x) + g_{(j-2)}(x) + \cdots \quad (6)$$

と書くことができる．ここで $d_k^{(j)}$ は詳細化成分と呼ばれ，整数 j に対応する周波数の大きさを示している．整数 j はレベルと呼ばれ，このレベルが高いほど高周波数成分を抽出することができる．

著者らは [1] において，離散ウェーブレット変換を様々な時系列データに用いることによって，各レベルに対応した周波数成分がどれ程含まれているかを観測した．その結果，周波数領域において何らかの周期性を見出すことができた．しかし，一週間という短い期間を対象として行った調査であり，さらに長い期間での調査や国別の時差を考慮した調査が求められていた．

2.2 周波数成分の定義

著者らは [1] で，周波数領域で周期性を見出すために，式 (4) における詳細化係数 $d_k^{(j)}$ を用いて調査を行った．しかし，中長期傾向の周期性を見出すためには，期間が増えることによってレベル数も増大することから，個々のレベル毎で周期性を見出すことは困難になることが予想される．そこで，式 (5) におけるいくつかの $g_j(x)$ を積算することにより，*high-band*，*mid-band*，*low-band* という 3 つの周波数成分を抽出することとする．それぞれの成分の説明を下記に示す．

- *high-band*
  レベル 1 から 3 の $g_j(x)$ について積算を行う．これにより，高周波数成分を表現するレベル 1 から 3 までの周波数成分をまとめて表現することが可能となる．

- *mid-band*
  レベル 4 から 6 の $g_j(x)$ について積算を行う．これにより，中周波数成分を表現するレベル 4 から 6 までの周波数成分をまとめて表現することが可能となる．

- *low-band*
  レベル 7 以上の $g_j(x)$ について積算を行う．これにより，低周波数成分を表現するレベル 7 以上の周波数成分をまとめて表現することが可能となる．

なお，上述の定義は，実験的にレベル 1 から 3 までを高周波数成分，レベル 4 から 6 までを中周波数成分，レベル 7 以上を低周波数成分としている．

3 周期傾向の調査

文献 [1] で行った実験条件は，データ取得期間を一週間とし，データ取得間隔は，データ取得期間中に検知された各シグネチャのアラート数を 5 分間隔で積算したデータを用いていた．ここで用いていたデータの特徴量はシグネチャ名と検知数，検知時刻である．

本稿では，データ取得期間を一週間及び一ヶ月とし，さらにデータの特徴量としてシグネチャ名と検知数，検知時刻に加え，新たに IP アドレス値による国別の特徴量を追加した．なお，調査を行ったシグネチャは一見して周期性を持つことが明らかなシグネチャを 2 種類と，一見して周期性を持たないが，断続的に検知されているシグネチャ2 種類を取り上げて調査を行った．それぞれのシグネチャの名称を表 1，2 に示す．なお，シグネチャの名称に付けられた番号は便宜的に付けたシグネチャ番号である．

表 1: 周期性を持つシグネチャの名称

シグネチャの名称
11 : DNS SPOOF query response with TTL of 1 min. and no authority
21 : WEB-MISC SSLv2 openssl get shared ciphers overflow attempt

表 2: 断続的に検知されているシグネチャ

シグネチャの名称
7 : ATTACK-RESPONSES 403 Forbidden
8 : WEB-MISC robots.txt access

3.1 シグネチャ毎の国別の統計結果

データ取得期間を 2013 年 6 月 1 日から 6 月 7 日までの一週間とし，シグネチャ毎の上位 3ヶ国の国別統計結果を表 3 に示す．なお，シグネチャ番号 11，21，8 については学外から学内へのパケットを検知するシグネチャであったため，IP アドレス値は送信元 IP アドレス値から国名を割り出した．シグネチャ番号 7 については学内から学外へのパケットを検知するシグネチャであったため，IP アドレス値は送信先 IP アドレスから国名を割り出している．

シグネチャ番号 11，21 は一見して日本時間の深夜から昼にかけて次第に検知数が少なくなり，その後夜にかけて徐々に検知数が増加するという傾向を示しており，日本時間において多くの人がパソコンを使用する時間帯に検知数が増えるシグネチャであった．シグネチャ番号 11 は学内から学外に対してクエリを出し，それに対してのレスポンスを検知するシグネチャであったために，一見して明らかに周期的な傾向が現れていると考えられる．シグネチャ番号 21 は学外から学内に対してのアラートであるが，表 3 の検知元である国名 1 位が日本 (JP) であり，2 位のフランス (FR)，3 位のアメリカ (US) に比べ検知数が圧倒的に多いために，全体としても日本時間での周期的な傾向が現れていると考えられる．シグネチャ番号 7，8 は一見して周期傾向を持たないアラートであった．シグネチャ番号 7 は，表 3 の検知元である国名 1 位のアメリカ (US) と 2 位の日本 (JP) の間で検知数に大きな差はなく，国別の時差の影響によって周期傾向が現れていない可能性がある．また，シグネチャ番号 8 の検知元 IP アドレス値の大半は検索エンジンのクローラーによるものであった．

3.2 様々なシグネチャに対する周波数解析の評価

データ取得期間を 2013 年 6 月 1 日から 6 月 30 日までの 1ヶ月とし，一見して周期性を持つことが明らかなシグネチャを 2 種類取り上げて解析を行った．対象としたシグネチャは表 1 に示したシグネチャである．

表 3: 4 つのシグネチャに対する国別の統計結果の順位

シグネチャ番号	1 位/検知数	2 位/検知数	3 位/検知数
11	US/2580	TW/1247	JP/3
21	JP/1130	FR/46	US/31
7	US/209	JP/159	CN/31
8	US/3277	CN/266	JP/183

シグネチャ番号 11 の時系列データのグラフと，*high-band*，*mid-band*，*low-band* の周波数成分のグラフを図 3 に示す．また，シグネチャ番号 21 の時系列データのグラフと，*high-band*，*mid-band*，*low-band* の周波数成分のグラフを図 4 に示す．

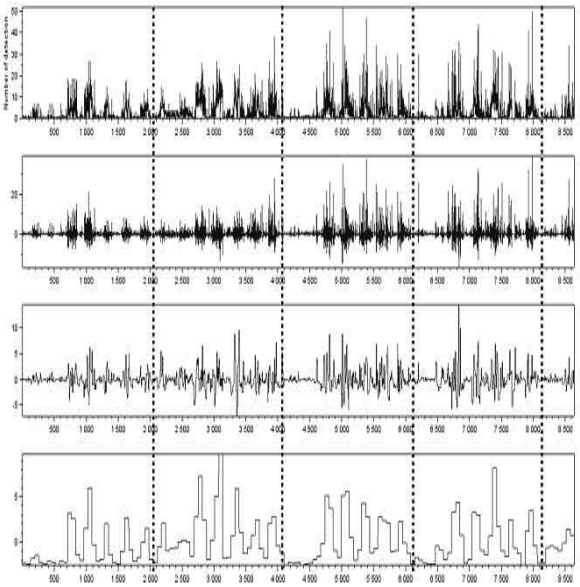

図 3: シグネチャ番号 11 の時系列データと *high-band*，*mid-band*，*low-band* の周波数成分

図 3 と図 4 において，最上段のグラフはデータ取得期間中の各シグネチャの検知数を表す時系列データであり，2 段目のグラフは *high-band* の周波数成分，3 段目は *mid-band* の周波数成分，最下段は *low-band* の周波数成分を示している．なお，図中の 4 つの破線は一週間毎の区切りを示している．

時系列データにおいて，データ取得期間の初日である 6 月 1 日は土曜日，6 月 2 日は日曜日であり，平日に比べて検知数が少なくなっている傾向が見られる．図 3，図 4 の両方において，最下段の *low-band* で平日に相当する位置の周波数成分が大きくなっており，土曜日及び日曜日は周波数成分が小さくなるという傾向が出ており，特に *low-band* の周波数成分において明確に周期性を確認することができる．

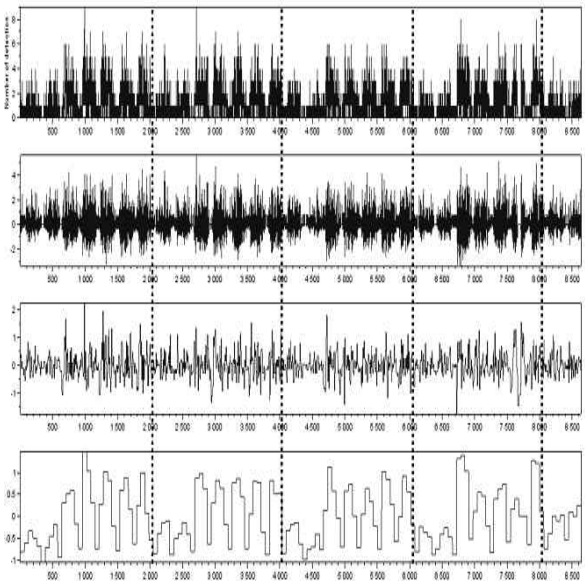

図 4: シグネチャ番号 21 の時系列データと *high-band*，*mid-band*，*low-band* の周波数成分

次に，一見して周期性をもたないが，断続的に検知されているシグネチャ2 種類の解析を行った．対象としたシグネチャは表 2 に示したシグネチャである．

シグネチャ番号 7 の時系列データのグラフと，*high-band*，*mid-band*，*low-band* の周波数成分のグラフを図 5 に示す．また，シグネチャ番号 8 の時系列データのグラフと，*high-band*，*mid-band*，*low-band* の周波数成分のグラフを図 6 に示す．

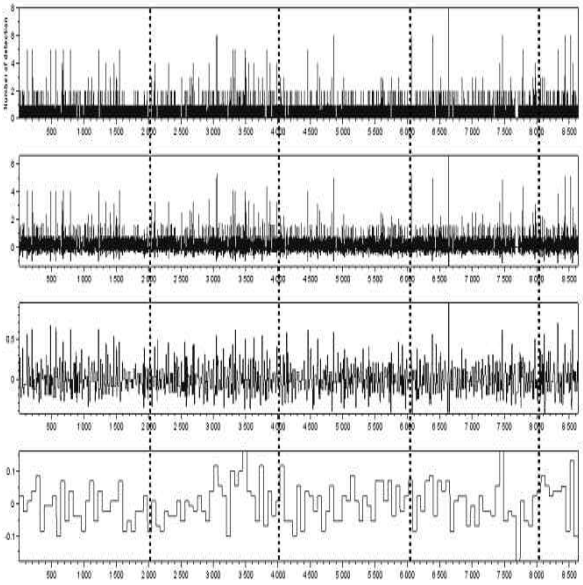

図 5: シグネチャ番号 7 の時系列データと *high-band*, *mid-band*, *low-band* の周波数成分

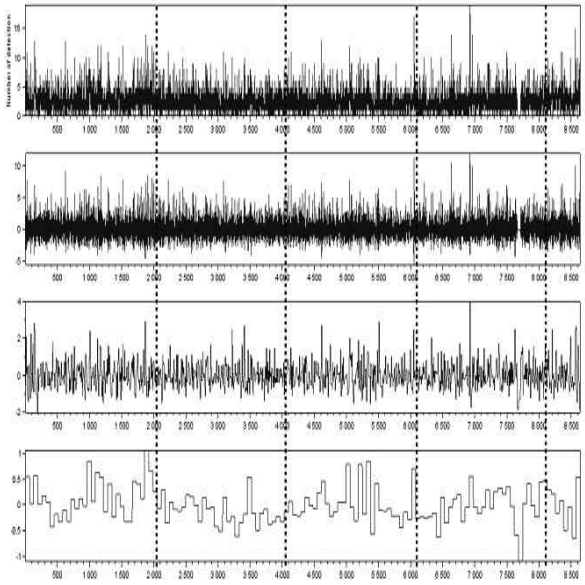

図 6: シグネチャ番号 8 の時系列データと *high-band*, *mid-band*, *low-band* の周波数成分

図 5 及び図 6 において，どの帯域の周波数成分においても図 3，図 4 のような明らかな周期性は見られない．これは 3.1 節において調査を行った結果より，シグネチャ番号 7 及び 8 では，国別の検知数に大きな差がなく，国別では周期性があったとしても全体では周期性が見られなくなっている可能性が考えられる．しかし，*high-band* の周波数成分では突発的に検知数が増減した箇所を抽出できていることが確認できる．

3.3 国別の周波数解析の評価

3.2 節において，シグネチャ番号 7 及び 8 では国別の検知数に大きな差がなく，全体として周期性を見出すことができない可能性があった．そこで，3.1 節で行ったシグネチャ番号 7 おける上位 3ヶ国の国別の時系列データに対して周波数解析を行い，周期性を見出すことができるか調査した．シグネチャ番号 8 で出力されるアラートの送信元は主に検索エンジンのクローラーでるので，人が行う操作によって周期性が生じるとは考えにくく，本稿では調査を行っていない．なお，データ取得期間は 2013 年 6 月 1 日から 6 月 7 日の一週間とした．

まず，シグネチャ番号 7 における全ての国を合わせて検知した時系列データのグラフと，*high-band*, *mid-band*, *low-band* の周波数成分のグラフを図 7 に示す．なお，図中の 6 つの破線は 1 日毎の区切りを示している．

図 7 において，どの周波数成分を見ても周期性は見られない．これは 3.2 節で考察したとおり，シグネチャ番号 7 では，国別の検知数に大差がなく，全体として周期性が見られないのではないかと考えられる．

次に，シグネチャ番号 7 における国別の検知数で 2 位であった日本 (JP) の時系列データのグラフと，*high-band*, *mid-band*, *low-band* の周波数成分のグラフを図 8 に示す．

図 8 においては，*low-band* において火曜日を除く平日で，午後以降に徐々に周波数成分が大きくなる傾向が確認できる．シグネチャ番号 7 は外部ネットワークから内部ネットワークの WEB ページへアクセスしたとき，WEB ページが存在するものの，アクセス権限が無いためにアクセスが拒否された際に発生するアラートである．このことから人が多く活動する午後以降に検出数が増加することが予想され，このことが周波数成分を見ることにより確認できたと考えられる．

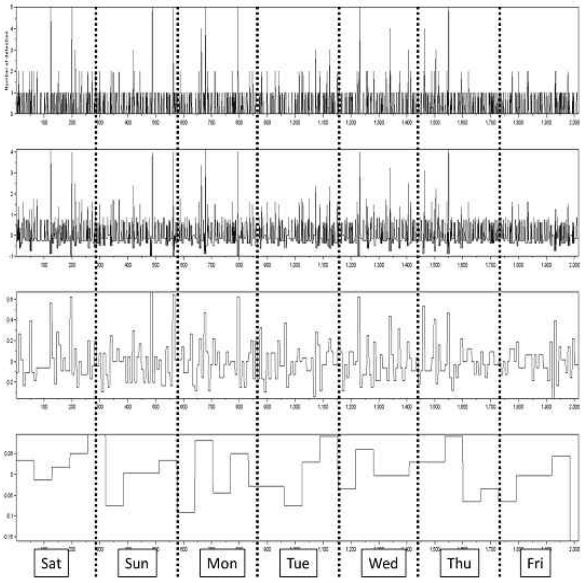

図 7: シグネチャ番号 7 の全ての国を合わせて検知した時系列データと *high-band*, *mid-band*, *low-band* の周波数成分

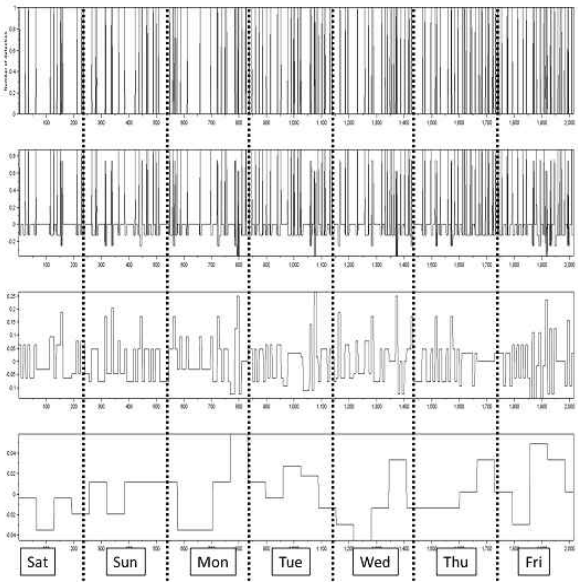

図 8: シグネチャ番号 7 の JP における時系列データと *high-band*, *mid-band*, *low-band* の周波数成分

4 おわりに

本論文では，侵入検知システムのログ分析手法として，離散ウェーブレット変換を用いた中長期的な傾向調査に関する周波数解析を試みた．

一見して周期的に検知されているシグネチャについては，一ヶ月という比較的長い期間においても客観的に周期性を見出すことができ，その周期を確認することができた．また，一見すると周期性を持たないようなシグネチャにおいても，国別による時差を考慮して調査を行った結果，曜日や時間帯による周期性を見出すことが可能であった．さらにそのようなシグネチャについても，*high-band* の周波数成分を見ることにより突発的に大きく検知数が増減する箇所を特定することが可能である．

本実験で行った調査では，深夜では検知数が少なくなり，夕方から夜にかけて検知数が多くなるという周期性が見られた．この傾向が現れた原因を探るために各々のシグネチャについて調査を行ったところ，人が多くパソコンを使用する時間帯に多くの検知数が増大するという結論が得られた．しかし他のシグネチャでは，異なる傾向が見られる可能性がある．今後，様々なシグネチャに対して周期傾向等の調査を進めることで，不正アクセス等の傾向を明らかにすることが必要であろう．

参考文献

[1] 木村知史，稲葉宏幸:“離散ウェーブレット変換による IDS のログ分析手法に関する検討”，信学技報, SITE2013-9, vol.113, no.136, pp.27-32, July. 2013.

[2] “ Snort ”, `http://www.snort.org/`, `2013/08/09` 参照

[3] “ Scilab ”, `https://www.scilab.org/`, `2013/08/09` 参照