数理解析研究所講究録
第1548巻 2007年 154-161

統計的不正アクセス検知モデルの改良

林坂 弘一郎†, 酒井 悠人†, 土肥 正†
Koichiro Rinsaka†, Yuto Sakai† and Tadashi Dohi†
†広島大学大学院工学研究科情報工学専攻

1. はじめに

コンピュータやネットワークへの不正アクセス，コンピュータウイルス感染といった情報セキュリティに関する問題は現代の情報化社会において重要であるという認識は既に社会に浸透している．不正アクセスやコンピュータウィルスによって，システムの停止，内部情報の流出，破壊などの危険があることは周知の通りである．したがって，このような情報セキュリティ確保のための対策は組織にとって重要な課題である．

コンピュータやネットワークに対する情報セキュリティ確保のためのシステムとして，不正アクセスの監視を行う侵入検知システム (IDS: Intrusion Detection System) とフィルタリング機能を持つファイアウォールが設置される．ファイアウォールは不正アクセスに利用されるパケットを遮断するが，不正アクセスの中には通常の通信に用いられるパケットを利用するものも多く，ファイアウォールではそのような不正アクセスに対しては効力を発揮しない [1]．これに対し，IDSは不正アクセスの監視にも有効である．

IDSが不正アクセス（攻撃）を検知するために採用されている技術は次の2種類に大別される．すなわち，不正検出と異常検出である．不正検出は，既知の攻撃からリポジトリと呼ばれるデータベースを作成し，アクセス状況をこのリポジトリと照合することにより攻撃を検知するものである．この手法はアルゴリズムがシンプルであり，誤報が少ないなどの長所を有するが，リポジトリに登録されていない未知の不正アクセスについては検出できない問題点がある [2]．一方，異常検出は各種の統計的手法を用いて通常状態の活動の統計情報を記録した通常プロファイルを作成し，現在のアクセス状況と通常プロファイルとの間に大きな差があるときに，その活動を攻撃の兆候と見なす手法である．すなわち，異常検知は未知の不正アクセスについても検出が可能である．

異常検知に対して，これまでに様々な統計手法が攻撃検知のために採用されている [1–5]．浅香ら [3] は判別分析による侵入検知手法を提案した．Yeら [4] は指数重み付け移動平均法によって監査データから不正アクセスを検知する手法を考えた．宮本ら [2] は観測されたネットワークトラヒックデータを数値化し，サポートベクタマシンを用いたクラスタリング手法を適用することにより異常検知を行っている．

更に，Yeら [5] はホストマシン上の活動を記録した監査データに対して離散時間マルコフ連鎖（以下，DTMC）を適用することで攻撃を検知する枠組みを提案している．Yeら [5] によって提案されたDTMCに基づいた攻撃検知はアルゴリズムが単純であるために，ホストマシン上における監査活動の負荷を軽減することが可能である．観測された監査事象の時系列データをパスとして取り出し通常プロファイルと比較することで攻撃を検知する．しかしながら，DTMCモデルは通常状態の活動と攻撃活動が混在するようなより現実的なパスに対して攻撃検知精度が低下するという問題を持っている．

本稿では，監査データに対して多変量解析手法である数量化理論4類とクラスター分析を用いることで類似したパスを1つのクラスターに集約し，攻撃検知する手法を提案する．これにより，攻撃と通常状態の情報が混在する場合における攻撃検知精度の向上を図る．

本稿の構成は以下の通りである．2. ではYeら [5] によって提案されたDTMCに基づいた攻撃検知手法について概観し，その問題点について言及する．3. では多変量解析に基づいた攻撃検知手法を提案する．ここで提案する攻撃検知手法は数量化理論4類とクラスター分析を組み合わせることにより攻撃の兆候を検知するものである．4. のシミュレーション実験を通して，DTMCモデルと多変量解析モデルの攻撃検知精度を比較し，提案手法の有効性を示す．

2. 離散時間マルコフ連鎖に基づいた攻撃検知

ここでは，Yeら [5] によって提案された離散時間マルコフ連鎖 (DTMC) に基づいた攻撃検知手法を概観する．

2.1. DTMC による攻撃検知

ホストマシン上で観測された時系列監査データをそれぞれ事象の到着と見なし，各到着を離散時間の単位時間 $t\ (=0,1,2,\cdots)$ とみなす．また，観測された事象の系列が DTMC に従うものと仮定する．DTMC において，時刻 $t+1$ でのシステムの状態は時刻 t での状態のみに依存し，それ以前の状態には依存しない．すなわち，時刻 t において観測された状態を X_t とすると，

$$\Pr\{X_{t+1}=i_{t+1}|X_t=i_t, X_{t-1}=i_{t-1}, \cdots, X_0=i_0\} = \Pr\{X_{t+1}=i_{t+1}|X_t=i_t\} \tag{1}$$

が任意の時刻 t，及び状態 $i \in S$ について成立する．ただし，S は状態空間である．今，時刻 t から $t+1$ での状態推移が時間と統計的に独立であれば DTMC は定常推移確率

$$p_{i,j} = \Pr\{\text{the system is in a state } j \text{ at time } t+1 \mid \text{the system is in state } i \text{ at time } t\} \tag{2}$$

を持つ．このとき，式 (2) の DTMC モデルの推移確率行列は次式によって与えられる．

$$P = \begin{bmatrix} p_{1,1} & \cdots & p_{1,s} \\ \vdots & \ddots & \vdots \\ p_{s,1} & \cdots & p_{s,s} \end{bmatrix}. \tag{3}$$

ただし，$\sum_{j=1}^{s} p_{i,j} = 1$ である．また，初期確率分布を

$$\mathbf{Q} = [q_1, q_2, \ldots, q_s] \tag{4}$$

とする．ここで，

$$q_i = \Pr\{\text{the system is in state } i \text{ at time } 0\} \tag{5}$$

である．

DTMC においてシステムの状態系列が $X_1, ..., X_T$ となる確率は次式によって与えられる．

$$\Pr\{X_1, ..., X_T\} = q_{x_1} \prod_{t=2}^{T} p_{x_{t-1}, x_t} \tag{6}$$

式 (6) の対数をとることで，次式が得られる．

$$\log(\Pr\{X_1, ..., X_T\}) = \log\left(q_{x_1} \prod_{t=2}^{T} p_{x_{t-1}, x_t}\right). \tag{7}$$

DTMC モデルにおいて，式 (3) で与えられる推移確率行列，及び式 (4) の初期確率分布は，システムの状態の観測結果 $X_0, X_1, X_2, ...X_{N-1}$ から以下のように推定可能である．

$$p_{i,j} = \frac{N_{i,j}}{N_{i.}}, \tag{8}$$

$$q_i = \frac{N_i}{N}. \tag{9}$$

- $N_{i,j}$: X_t が状態 i で X_{t+1} が状態 j である観測された X_t と X_{t+1} のペアの数.
- $N_{i.}$: X_t が状態 i で X_{t+1} が状態 $(1, \cdots, s)$ のいずれかの状態である観測された X_t と X_{t+1} のペアの数.
- N_i : X_t が状態 i である数.
- N : 全体の観測数.

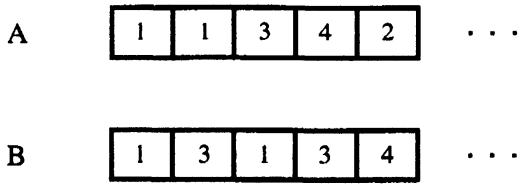

図 1: パスの類似性の問題.

したがって，観測された監査データが通常状態におけるホストマシンの活動から得られたものと仮定すれば，それらから $p_{i,j}$ 及び q_i を推定することで，通常状態の活動を表す統計情報である通常プロファイルを得ることができる．その後，時刻 t において監査したい活動 E_t について最近到着した T $(=1,2,\cdots)$ の事象 $E_{t-T+1},\cdots,E_t$ を取り出す．ここで T はウィンドウサイズと呼ばれる．式 (7) によりウィンドウサイズ T 中の活動の生起確率を計算する．もしもその確率が大きければそれが通常状態での活動と見なし，小さければ通常状態の活動ではあり得ない，すなわち，攻撃状態にあると見なすことができる．つまり，事象系列の発生確率に対して任意のしきい値を設定し，式 (7) で求めた値がしきい値よりも大きければ通常状態の活動，小さければ攻撃状態であると判定する．続いて，時刻 $t+1$ では $E_{t-T+2},\cdots,E_{t+1}$ までの T 個の事象について監査する．これを事象到着の都度行うことで，攻撃をリアルタイムに検知することができる．

2.2. DTMC モデルにおけるパスの類似性の問題

多くのネットワークにおいて，ホストマシンには複数のユーザーから同時にアクセスが行われている．また，不正アクセスが行われている最中にも正規ユーザーから通常のアクセスがある場合も少なくない．このような場合，連続した攻撃状態の監査事象の中に通常状態の監査事象が紛れ込むことになる．これは一種のノイズであり，このノイズが観測ウィンドウ内に存在することで，式 (7) で計算される事象の生起確率は上昇することになる．したがって，攻撃状態に正常状態のノイズが混在することで，不正アクセスを検知できない可能性がある．また，正常状態に攻撃状態のノイズが混在すれば，正常なアクセスが攻撃であると認識されてしまうこととなる．

更に，ノイズの混在するタイミングがわずかに異なる場合を考える．図 1 は 2 種類の観測系列（パス）を示している．パス A と B は似たパスであるが，DTMC モデルにおいては全く別のパスとして認識される．仮に A, B がともに正常状態であったとしても，パスの生起確率は別個に計算されるため，生起確率が低く推定され，結果として攻撃状態である判定されるという問題点を内包している．

3. 多変量解析を用いた攻撃検知

DTMC による攻撃検知では **2.2** で述べたようなパス類似性に起因する検知精度の低下に問題がある．本研究では多変量解析の手法である数量化理論 4 類とクラスター分析を用いて類似したパスをいくつかのクラスターに分類することにより攻撃を検知する手法を提案する．

3.1. 数量化理論 4 類によるパスの類似度の数量化

パス同士の類似度を数量化するために数量化理論 4 類を適用する．まず，観測されたすべてのパスについての親近性行列を作成する．この親近性行列とは，各対象間の類似度を表す行列である．本研究では親近性行列の非対角要素を以下の方策によって決定する．すなわち，選択された 2 種類のパスに出現した各事象の回数をカウントし，一致した個数を親近性行列の要素とする．この一対比較をすべてのパスについて行い非対角要素を決定する．また，対角要素については 0 とする．図 2 には 4 種類のパスの例を示す．例えば，パス A と B を比較すると，一致する事象は事象 1 が 2 個，事象 2 は 0 個，事象 3 は 1 個，事象 4 も 1 個となり，その合計 4 が親近性行列の要素となる．具体的には親近性行列の要素を $V_{i,j}$ とし，i $(=1,2,\cdots)$ 番目のパスにおいて観測された事象 k $(=1,\cdots,s)$ の個数を Z_{i_k} とすると

$$V_{i,j}=\begin{cases}0 & (i=j)\\ \sum_{k=1}^{s}\min(Z_{i_k},Z_{j_k}) & (i\neq j)\end{cases} \tag{10}$$

となる．すなわち，2 種類のパス i, j の関連が高いものには，$V_{i,j}$ の値が大きくなる．

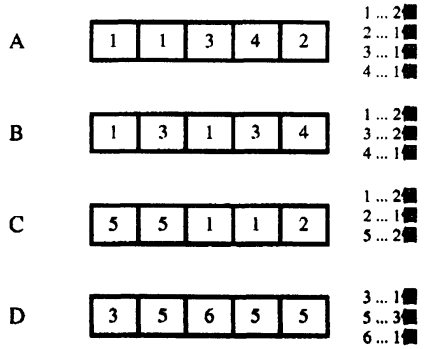

図 2: パスの例.

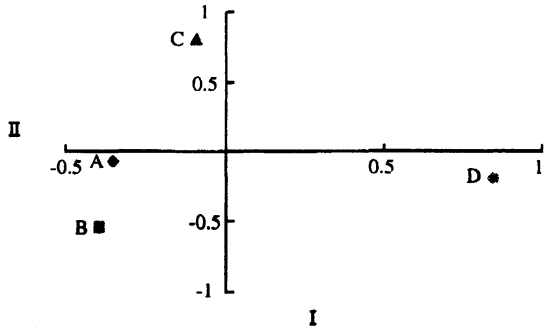

図 3: パス数量の散布図 (n=2).

表 1: 親近性行列と H 行列.

	A	B	C	D
A	0	4	3	1
B	4	0	2	1
C	3	2	0	2
D	1	1	2	0

親近性行列

	A	B	C	D
A	-16	8	6	2
B	8	-14	4	2
C	6	4	-14	4
D	2	2	4	-8

H行列

表 2: 2 次元の数量.

	I	II
A	-0.35	-0.07
B	-0.39	-0.54
C	-0.09	0.81
D	-0.84	-0.1

次に，得られた親近性行列について，非対角要素を 2 倍し，対角要素は各行の非対角要素の和について正負を入れ替えた値を代入することで，表 1 に示す H 行列が得られる．このようにして得られた H 行列の固有値問題を解き，得られた第 $n(\geq 1)$ 固有値までを取り上げ，これらに対応する固有ベクトルを n 次元の数量とすればよい．表 2 には H 行列から得られた 2 次元の数量を示す．この数量より，図 3 に示すようなパス間の類似度に関する散布図が得られる．

3.2. クラスター分析によるパスの分類

類似したパスを同じクラスターに集約するために，クラスター分析を適用する．まず，**3.1** で求めた n 次元の数量を各パスのクラスター分析における変量とし，各パス間の非類似度（ユークリッド距離）を求める．非類似度が小さいほどパス同士の類似性が高いと言える．次に，ユークリッド距離を元にデンドログラムを作成する．本稿では個々の点をまとめたクラスター間の距離を決定する際において最短距離法を採用している．

図 4 には表 2 の数量からクラスター分析によって得られたデンドログラムを示す．図のデンドログラムにおいて非類似度 0.5 の高さで切断したとき，パスは (A, B), (C), (D) のクラスターに分類される．一方，非類似度 1.0 の高さで切断したときには，パスは (A, B, C) と (D) のクラスターに集約される．デンドログラムを 0.5 で切断すると 4 種類のパスが 3 つのクラスターに，1.0 で切断すれば 2 つのクラスターに分類されるので，以降ではそれぞれ，75%, 50% のように表記する．

今，任意の 2 種類のパス $X_1, ..., X_T, Y_1, ..., Y_T$ が 1 つのクラスターに集約されたとすると，このクラスターの生起確率は次式によって求めることができる．

$$\log(\Pr\{X_1, ..., X_T, Y_1, ..., Y_T\}) = \log\left\{(q_{x_1}\prod_{t=2}^{T} p_{x_{t-1},x_t}) + (q_{y_1}\prod_{t=2}^{T} p_{y_{t-1},y_t})\right\}. \quad (11)$$

3.3. 攻撃検知手順

本研究では DTMC と同様にあらかじめ観測された監査事象から通常プロファイルを作成し，後に観測された事象と通常プロファイルを比較することで攻撃の検知を行う．以下に通常プロファイルの作成手順，及び攻撃

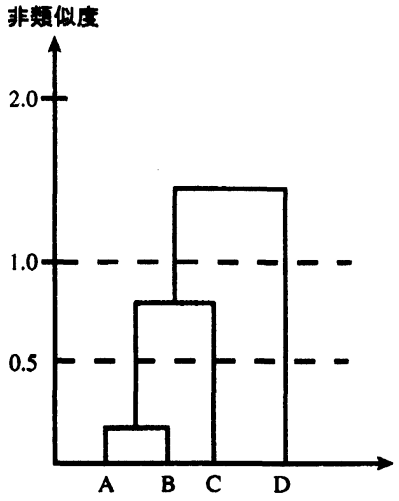

図 4: パスのデンドログラム.

検知手順を示す.

通常プロファイル作成手順

Step1 通常状態の監査データから観測された各パスに対する親近性行列を求める.

Step2 親近性行列の固有値と対応する固有ベクトルを求め, 大きい固有値 n 個を取り上げ, これに対応する固有ベクトルをパスの数量とする.

Step3 Step2 で求めた数量から個々のパス間のユークリッド距離を求め, いくつかのクラスターに分類する.

Step4 同じクラスターに属するパスの生起確率からクラスターの生起確率を求め, 通常プロファイルを作成する.

攻撃検知手順

Step1 観測されたパスが通常プロファイルに含まれる場合は Step2 へ, 通常プロファイルに含まれないパスが出現した場合は Step3 へ.

Step2 しきい値を基準に, 到着事象が通常活動によるものか攻撃活動によるものかを判別し, 終了.

Step3 通常プロファイルに到着事象を加えた各パスのユークリッド距離を求める. 観測されたパスが, いずれかのクラスターに属した場合, クラスター内のユークリッド距離の最も近いパスを選択し, しきい値を基準に到着事象が通常活動によるものか攻撃活動によるものかを判別する. どのクラスターにも属さなければ攻撃活動と判別し, 終了.

なお, 通常プロファイルの作成においては, 比較的大規模な行列に対して固有値計算を行う必要があることに注意を要する. また, 攻撃検知の Step3 においても固有値計算を再度実行する必要がある.

4. シミュレーション実験

Linux OS 上では, 様々なサービスが稼働しており, 多数の異なったタイプの監査データがログとして記録されている. 本実験では広島大学大学院工学研究科システム信頼性工学研究室のサーバマシンで観測された 2006 年 10 月 1 日から 10 月 10 日までの 12506 個の監査事象を使用し, 連続監査事象を取り出すことにより以下のシミュレーション実験を行った. ここで, 観測された監査事象の事象タイプは 18 種類であり, 各事象の種類によって状態を定義する. また, 事象の到着を 1 単位時間として考える.

多くの場合, 不正アクセスはパスワードの管理や設定の甘さが原因となる. 以下では, パスワード・クラッキングを想定した攻撃をシミュレートし, この攻撃に対して DTMC モデル, 多変量解析モデルの検知精度について検証する.

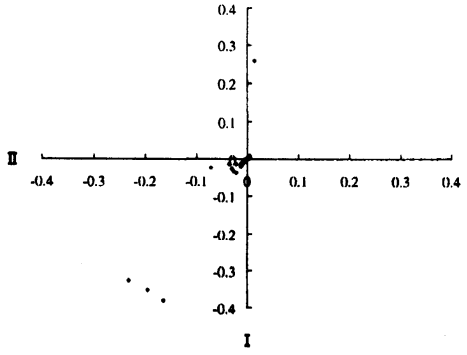

図 5: 多変量解析モデルの二次元プロット 1.

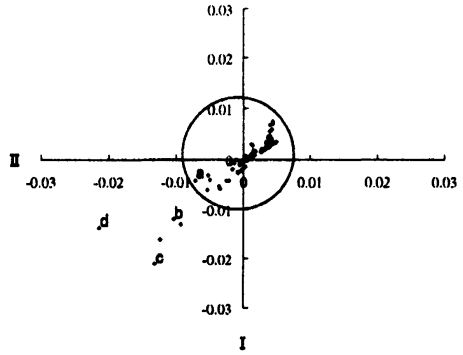

図 6: 多変量解析モデルの二次元プロット（拡大図）.

a	pop3	pop3	dhcpd	dhcpd	sshd
b	named	named	argus	sshd	argus
c	sshd	sshd	named	named	named
d	argus	argus	arugus	argus	named

図 7: パスに含まれる事象の種類.

DTMC モデル及び多変量解析モデル両者の通常プロファイルは 10 月 1 日から 10 月 7 日にかけて観測された 9022 個の監査事象を通常状態と見なして作成した．10 月 8 日から 10 月 10 日にかけて観測された 3484 個の監査事象も通常状態と見なす．攻撃状態の 167 個の監査事象はパスワード・クラッキングを想定しパスワード推測ツール Brutus [6] を使用して作成した．なお，Brutus は辞書に保存された単語を順次組み合わせてパスワードを推測するディクショナリ・アタック，及び辞書を使いあらゆる文字を組み合わせ総当たりで推測を行うブルート・フォース・アタックを行うことができるツールである．これら 3484 個の通常状態の監査事象と 167 個の攻撃状態の監査事象を混合し以下の 4 種類のテストデータを作成した．

data1 通常状態の後に攻撃状態を結合.

data2 通常状態の 3484 個と攻撃状態の 167 個をそれぞれ 2 つに分割し，交互に混合.

data3 通常状態の 3484 個と攻撃状態の 167 個をそれぞれ 4 つに分割し，交互に混合.

data4 通常状態の 3484 個と攻撃状態の 167 個をランダムに混合.

上記の data1～data3 はシミュレーション実験用に意図的に作成されたノイズの少ないテストデータである．一方 data4 は通常状態と攻撃状態の事象が混在したより現実的なテストデータである．

なお，本稿で提案した多変量解析に基づいた攻撃検知手法を上記データに適用する場合においては，700×700 程度の規模の行列について固有値および固有ベクトルを求める必要がある．この固有値計算を高速に行うために，本稿では GNU Scientific Library (GSL) [7] を利用した．

本実験での多変量解析モデルの攻撃検知におけるクラスター分析の際のクラスターと監査事象の関係について以下に述べる．図 5 は上記のデータを使用し，多変量解析モデルの通常プロファイルを作成する際のパスの数量を $n = 2$ としたときの各数量を二次元のグラフにプロットしたグラフである．また，図 6 は図 5 の原点付近を拡大したグラフである．図 6 について記号 a～d はパスの種類を表しており，対応するパスに含まれる事象は図 7 に示している．

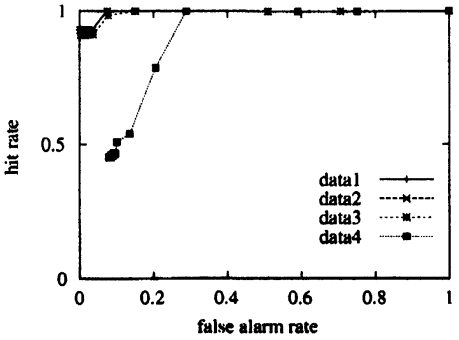

図 8: DTMC モデル.

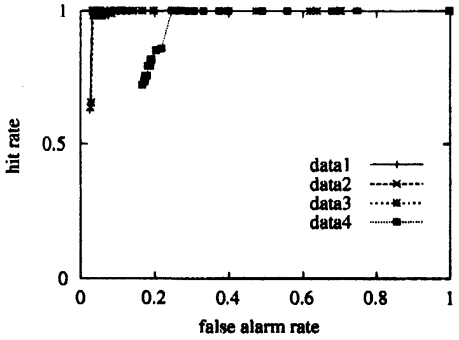

図 9: 多変量解析モデル.

図 6 において，円で囲まれているクラスターが生成されたとする．クラスターのサイズを大きくしていくことで近いパスが同一のクラスターに統合されるためクラスターのサイズを大きくしていくと，b，c，d の順で a が含まれるクラスターに統合される．大きなクラスター内のパスの生起確率は点在する小さなクラスター (含まれるパスが 1 個，2 個など) と比較すると大きくなっているため，大きなクラスターに含まれるパスが観測された場合は通常活動，小さなクラスターのパスが観測された場合は攻撃活動と判定されやすい．

図 6 で a のみが大きなクラスターに属している状態では，事象タイプ named が連続して到着するパスは攻撃活動と判定されやすい傾向がある．しかしデンドログラムの切断位置を変えることでクラスターのサイズを大きくし，b，c のパスも a とおなじ大きなクラスターに統合されると事象タイプ named が連続して到着するパスは通常活動と判定されやすくなる．

4.1. 実験 1

ここでは，4 種類のテストデータについて，最新の到着事象が通常活動によるものか攻撃活動によるものかを判定することにより，DTMC モデルと多変量解析モデルの攻撃検知精度を比較する．なお，両モデルにおいてウィンドウサイズは $T = 5$ と設定している．また，多変量解析モデルにおいて，パスの類似度を数量化する際に利用する数量の次元を $n = 2$ とし，デンドログラムはクラスター数がパス数の 10%になるような非類似度で切断している．両モデルに対して，しきい値を変化させ，不正アクセスの hit rate と false alarm rate をプロットした曲線を図 8，9 に示す．ここで hit rate とは，167 個の攻撃状態の事象のうち，正しく検知できた割合である．一方，false alarm rate とは，3484 個の通常状態の事象のうち，攻撃状態として誤検知された割合である．したがって，hit rate は 1 に近いほど，false alarm rate は 0 に近いほど検知精度が高いと言える．つまり，グラフ上で左上にプロットされるほど検知精度が高いこととなる．

図 8，9 より，作為的に作られた data1, 2, 3 については両モデルで高い検知精度を持っていることが確認できる．hit rate が最初に 1 となる点における false alarm rate は，DTMC モデルでは 0.07 であるのに対し，多変量解析モデルでは 0.03 である．多変量解析モデルの方が false alarm rate が低いため完全に攻撃活動を検知できるしきい値において誤検知が少ないと言える．一方，false alarm rate が 0 となるしきい値が DTMC モデルには存在するため，DTMC モデルには誤検知の全く起こらないしきい値を設定することができる．図 8，9 よりノイズの少ないデータについては両検知手法が有効であると言える．しかし，ノイズの多い data4 では全体的に多変量解析モデルのプロットが DTMC モデルよりも左上にあるため，多変量解析モデルの方が検知精度が高いと言える．したがって，類似したパスを同じクラスターに集約することで攻撃検知精度が向上すると言える．

4.2. 実験 2

上記の data4 を使用し，多変量解析モデルにおけるデンドログラム切断位置に関する感度分析を行った結果を図 10 に示す．ここで，図中の 90%は通常プロファイルの作成に用いた 9022 個の監査事象には 642 種類のパスが存在し，90%の 578 個のクラスターに集約される位置でデンドログラムを切断したことを意味している．同様に 50%では 321 個のクラスターに，10%では 64 個のクラスターに集約されている．図 10 より，デンドログラムの切断位置を高く設定する，すなわち類似したパスをより集約させ，クラスター数を減少させるほど攻撃検知精度が向上することが読み取れる．

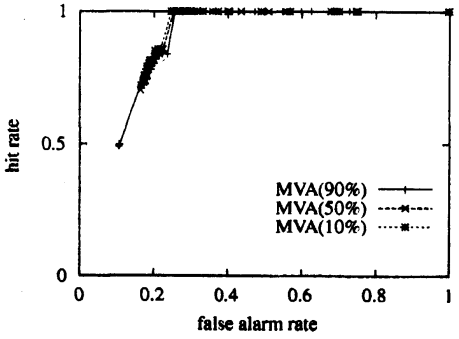

図 10: デンドログラム切断位置の感度分析結果.

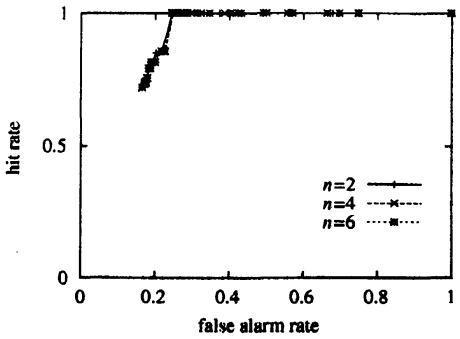

図 11: 次元の変化による感度分析結果.

4.3. 実験 3

ここでも data4 を利用し，パスの類似度を数量化する際に利用する次元 n に関する感度分析を行う．図 11 には $n = 2, 4, 6$ と変化させた場合の多変量解析モデルの hit rate と false alarm rate をプロットしている．図 11 より，次元の量の変化による検知精度の大きな違いは見られない．この次元 n が多変量解析モデルの検知精度に与える影響は軽微であると認められる．

5. むすび

本稿では多変量解析手法である数量化理論 4 類とクラスター分析を採用することで，監査事象系列の類似性を考慮した攻撃検知手法を提案した．数量化理論 4 類では親近性行列から得られた H 行列の固有値問題を解き，得られた固有ベクトルを n 次元の数量とした．この n 次元の数量に対してクラスター分析を行うことで多種の監査事象系列をいくつかのクラスターに集約し攻撃の検知を行った．シミュレーション実験を通して，提案手法はパスワード・クラッキングを想定した監査事象について従来の DTMC モデルよりも高精度に攻撃を検知できることが示された．

参考文献

[1] 油川良太，太田耕平，加藤寧，根本義章 (2003)，分散型ネットワークモニタリングによる不正アクセス早期検出システム，電子情報通信学会論文誌 B，**J86-B**(3)，pp. 410–418.

[2] 宮本貴朗，小島篤博，泉正夫，福永邦雄 (2004)，SVM を用いたネットワークトラヒックからの異常検出，電子情報通信学会論文誌 B，**J87-B**(4), pp. 593–598.

[3] 浅香緑，女部田武史，井上直，岡澤俊士，後藤滋樹 (2002)，不正侵入の痕跡と判別分析によるリモートアタックの検出法，電子情報通信学会論文誌 B，**J85-B**(1), pp. 60–74.

[4] N. Ye, S. Vilbert and Q. Chen (2003), Computer intrusion detection through EWMA for autocorrelated and uncorrelated data, *IEEE Transactions on Reliability*, **52**(1), pp. 75–82.

[5] N. Ye, Y. Zhang and C.M. Borror (2004), Robustness of the Markov-Chain Model for Cyber-Attack Detection, *IEEE Transactions on Reliability*, **53**(1), pp. 116–123.

[6] `http://www.hoobie.net/brutus/`

[7] `http://www.gnu.org/software/gsl/`