情報処理学会論文誌 Vol.54 No.12 2461–2471 (Dec. 2013)

多変量解析による標的型攻撃の分類

三村 守[1,a)] 田中 英彦[1]

受付日 2013年1月18日, 採録日 2013年9月13日

概要：機密情報や個人情報の搾取を目的とする標的型攻撃は多くの組織にとって脅威である．近年の標的型攻撃では，すでにマルウェアに感染した端末が踏み台にされ，情報の送信先は刻々と変化するため，真の攻撃者を識別することは困難となっている．攻撃者を識別するためには，複数の標的型攻撃のパラメータの共通性を分析し，攻撃者ごとに分類する必要がある．しかしながら，どのパラメータが最も攻撃者の特徴を示しているかは明確ではないため，攻撃者ごとに分類するのは容易ではない．この論文では，複数の標的型攻撃に関するパラメータを数値化し，概略の傾向を主成分分析で調査する．次に，因子分析により標的型攻撃を説明する要因を明らかにし，攻撃者と相関が高いパラメータを抽出する．さらに，因子負荷量からパラメータの優先度を決定し，クラスタ分析で複数の標的型攻撃を攻撃者ごとに分類する．

キーワード：標的型攻撃，多変量解析，主成分分析，因子分析，クラスタ分析，多次元尺度構成法

Using Multivariate Statistics to Classify Targeted Attacks

Mamoru Mimura[1,a)] Hidehiko Tanaka[1]

Received: January 18, 2013, Accepted: September 13, 2013

Abstract: Targeted attacks that exploit confidential information or personal information are serious threats for many organizations. Recently, attackers use the infected terminals as stepping stones, and often change destination of the stolen information. Thus, it is difficult to identify and reveal the true attacker. To identify the true attacker, we need to analyze commonality between targeted attacks and classify the attacks under each attacker. However, it is not clear which parameters indicate characteristic of attackers most, and not easy to classify the attacks under each attacker. In this paper, we use principal component analysis to investigate a tendency of targeted attacks. Next, we use factor analysis to find the factors that indicate characteristic of targeted attacks, and select high correlation parameters between an attacker. Furthermore, we determine priority of parameters by computing the corresponding factor loadings and use cluster analysis to assign a set of attacks into attackers.

Keywords: targeted attack, multivariate statistics, principal component analysis, factor analysis, clustering, multi dimensional scaling

1. はじめに

近年，組織が保有する機密情報や個人情報の搾取を目的とするサイバー攻撃の脅威が顕在化している．2011 年には国会，政府関係機関，民間企業等において大規模なサイバー攻撃が相次いで発覚し，大きな社会問題となったのは記憶に新しいところである．サイバー攻撃の中でも特に脅威が指摘されているのは，主に機密情報や個人情報の搾取を目的とし，ある組織や個人に標的を絞って実施される標的型攻撃である．経済産業省が実施した調査によると，2007 年には標的型攻撃を受けた経験がある企業は 5.4%にとどまっていたが，2011 年には約 6 倍の 33%に拡大している [1]．標的型攻撃の中でも，ある組織に特化した，時間および手法を問わずに継続的に行われる一連の攻撃は APT（Advanced Persistent Threat）や新しいタイプの攻撃 [2] と呼ばれることもあり，大きな脅威となっている．近年の標的型攻撃は，何らかの目的をもって実施されているもの

[1] 情報セキュリティ大学院大学
Institute of Information Security, IISEC, Yokohama, Kanagawa 221–0835, Japan
[a)] dgs104101@iisec.ac.jp

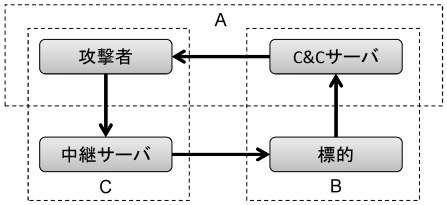

図 1 標的型攻撃の概念モデル
Fig. 1 A conceptual model of a targeted attack.

と考えられる．攻撃者は，個別の組織に特化した様々な手法を駆使して目的の達成を試みる．標的型攻撃に用いられるマルウェアは，攻撃者が目的を達成するための一手段にすぎない．これに対し，情報処理推進機構では標的型攻撃を段階的に区分し，各段階における具体的な対策を検討している [3]．このような対策を検討するためには，各々の標的型攻撃をマルウェアの種類ごとに分類し，その対策を検討するだけでなく，攻撃者の目的を根本から妨げる手段を検討した方が効果的であると考えられる．攻撃者の目的を推定するためには，各々の標的型攻撃を攻撃者ごとに分類する必要がある．我々が想定しているのは，観測された結果である標的型攻撃に関する情報から，攻撃者の目的や背景の推定を試みることである．現状の標的型攻撃への対策は，主にその手法に対して検討されているが，必ずしもその攻撃者に特化した対策が検討されているわけではない．攻撃者を識別し，標的型攻撃を攻撃者ごとに分類することができれば，攻撃者の目的や背景を推定し，その目的を根本的に妨げるようなより効果的な対策を検討することが可能になるものと考えられる．

典型的な標的型攻撃の概要を次に示す．まず攻撃者は，業務を装った件名やファイル名をつけた標的型メールで不正プログラム（マルウェア）を送信する．標的型メールを受信したユーザが，業務を装った件名やファイル名を不審に思わず，添付ファイルを開封した場合，その端末で不正な命令が実行され，マルウェアに感染する．マルウェアに感染した端末は踏み台とされ，攻撃者のコマンド&コントロールサーバを介した遠隔操作により任意の命令が実行され，不正な情報の搾取に利用される．搾取された情報は，外部のサーバに送信され，最終的に攻撃者によって回収される．本研究における標的型攻撃の概念モデルを**図 1** に示す．図中の「攻撃者」から「中継サーバ」を経由した「標的」までの矢印は，攻撃者が送信するマルウェアの流れを示している．図中の「標的」から「C & C サーバ」を経由した「攻撃者」までの矢印は，攻撃者によって回収される情報の流れを示している．本研究ではこの攻撃の流れに着目し，以下の観点で標的型攻撃に関する情報を分類する．

- 情報の送信先 (A)
- マルウェアの挙動 (B)
- マルウェアの送信元 (C)

近年の標的型攻撃では，コマンド&コントロールサーバは頻繁に変更され，情報の送信先は刻々と変化する．また，発信元を秘匿するために，すでにマルウェアに感染させ，遠隔操作が可能な一般ユーザの端末が踏み台にされ，標的型メールが送信される場合も珍しくない．さらに，不正に搾取した情報を用い，新たに業務を装った件名やファイル名を付与する悪質なケースも目立つようになってきている．このような理由から，標的型攻撃の真の攻撃者を識別することは，近年ではよりいっそう困難となってきている．標的型攻撃の真の攻撃者を識別するためには，複数の標的型攻撃の共通性を分析し，攻撃者ごとに分類する必要がある．しかしながら，どのパラメータが最も攻撃者の特徴を示しているかは明確ではないため，攻撃者ごとに分類するのは容易ではない．

ある標的型攻撃に用いられたメールの件名や本文，添付ファイルの名称や拡張子，マルウェアの種類，コマンド&コントロールサーバ等のパラメータは，いずれも単体では確実に攻撃者と結び付く情報とはいえない．なぜならば，ほとんどのパラメータは攻撃者が任意に変更することが可能だからである．しかしながら，変更に要する攻撃者のコストはパラメータの種類によって異なる．たとえば，メールの件名や本文，添付ファイルの名称等はコストを気にせずに容易に変更することが可能である．これに対し，マルウェアの種類，コマンド&コントロールサーバ等は，実際にマルウェアを作成したり，コマンド&コントロールサーバを準備するコストが発生したりするため，それほど容易に変更することはできないものと考えられる．ゆえに，複数の標的型攻撃に関するパラメータには，変化にある程度の傾向が生じている可能性がある．したがって，その変化の傾向から，攻撃者の特徴を示すパラメータを抽出することができれば，攻撃者の識別に結び付くものと考えられる．

そこでこの論文では，近年複雑化している標的型攻撃の真の攻撃者を識別するために，複数の標的型攻撃に関するパラメータの相関を分析し，攻撃者ごとに分類することを研究の目的とする．そのためにまず，複数の標的型攻撃に関するパラメータの概略の傾向を主成分分析で調査する．次に，因子分析を用いてパラメータの特徴を解釈し，攻撃者と相関が高いパラメータを抽出する．さらに，抽出したパラメータの因子負荷量の計算によって優先度を決定し，クラスタ分析で複数の標的型攻撃を攻撃者ごとに分類した結果を考察する．

2. 関連研究

標的型攻撃の分類に類似する研究としては，マルウェアの亜種の分類に関する研究があげられる．

文献 [4] では，マルウェアの動的な挙動を多次元ベクトルとして数値化し，ベクトル間のハミング距離からマルウェアの亜種を判定する手法が提案されている．本研究においても，標的型攻撃に関係するパラメータを多次元ベクトル

として数値化するが，数値化の手法や類似度の判定にユークリッド距離を用いる点等の様々な違いがある．手法に関する最も大きな相違点は，あらかじめ因子分析を用いて相関が高いパラメータを抽出し，分類に用いるパラメータの優先度を考慮する点である．また，マルウェアの動的な挙動だけでなく，標的型攻撃全般に関するパラメータが分析の対象である点も異なっている．

文献 [5] では，過去に収集されたマルウェアとの機械語命令列の類似度を算出する手法に加え，マルウェアのアンパッキング手法および逆アセンブル手法を組み合わせた自動分類システムが提案されている．この手法では，類似度を算出するために機械語命令列の最長共通部分列を用いている．本研究の分析対象はマルウェアだけでなく，標的型攻撃全般を対象とし，広範囲であるため，基本的には各パラメータが一致するか否かで攻撃の類似度を判定する．しかしながら，似ている傾向がある一部のパラメータの類似度を求める際には，最長共通部分列ではなく最長共通部分文字列を用いている．

これらの研究は，マルウェアの亜種の分類に関する研究であり，分析の対象はあくまでもマルウェアのみである．本研究では，マルウェアを攻撃者が目的を達成するための一手段と解釈する．そのように解釈すると，これらの研究は，マルウェアという手段への対策を検討するためには有用な内容であるといえる．これに対し，本研究は標的型攻撃を攻撃者ごとに分類することを目的としており，分析の対象はマルウェアという手段だけでなく，メールの件名や本文，添付ファイルの名称や拡張子，マルウェアの種類，コマンド&コントロールサーバ等の標的型攻撃全般に関する情報となっている．文献 [6] では攻撃者に関係する様々な情報を分析の対象としており，複数の関連する標的型攻撃において共通する情報がある事例が示されている．しかしながら，事例の数が少なく，その情報と攻撃者の関連性の強さも定量的に示されていない．本研究では，約 500 件の標的型攻撃の事例を対象として，攻撃者に関係する情報を分析し，その情報と攻撃者の関連性の強さを定量的に分析する．さらに，その情報と攻撃者の関連性の強さを考慮して，標的型攻撃を攻撃者ごとに分類する．標的型攻撃を攻撃者ごとに分類することができれば，攻撃者の目的を根本的に妨げるようなより効果的な対策に結び付くことが期待できる．

3. 主成分分析によるパラメータの分析

標的型攻撃に関するパラメータを多次元ベクトルに数値化し，主成分分析によってパラメータの傾向を調査する．

3.1 標的型攻撃に関するデータの概要

本研究で分析の対象とする標的型攻撃に関するデータの概要について説明する．分析の対象とする標的型攻撃は，2009 年から 2011 年の 3 年間に 26 の組織で発生した約 500 件の標的型攻撃において，後述する 22 のパラメータを取得することができたものを機械的に選定したものである．これらの標的型攻撃に用いられたマルウェアは，いずれもユニークなハッシュ値を持つ検体である．ファイルの種類は，24.1%が実行形式，32.3%が pdf 形式，9.6%が doc 形式，6.1%が xls 形式となっており，残りはその他の形式である．ウイルス対策ソフトの検知名によるマルウェアの種別は，トロイの木馬タイプが 64.4%，バックドアタイプが 19.2%，残りはその他のタイプとなっており，ほぼすべてのマルウェアに遠隔操作をするための機能が実装されている．マルウェアに感染したホストに指令を送り，制御するために用いられるコマンド&コントロールサーバのドメイン名については，.com が 36.1%で最も多く，ついで.org が 26.2%，.net が 9.8%となっている．

表 1 標的型攻撃に関するパラメータ
Table 1 Parameters about targeted attacks.

	No.	Parameter	説明
A	1	host name	C & C サーバのホスト名
	2	IP address	C & C サーバの IP アドレス
	3	domain name	C & C サーバのドメイン名
	4	DNS server	C & C サーバの DNS サーバ
	5	resistrant	C & C サーバのドメイン登録者
	6	resistrar	C & C サーバのドメインレジストラ
	7	address owner	C & C サーバの IP アドレス所有者
B	8	virus name 1	ウイルス対策ソフト 1 の検知名
	9	virus name 2	ウイルス対策ソフト 2 の検知名
	10	virus name 3	ウイルス対策ソフト 3 の検知名
	11	virus name 4	ウイルス対策ソフト 4 の検知名
	12	temp file	マルウェアが作成するファイル名
	13	registry	マルウェアが作成するレジストリ名
	14	mutex	マルウェアが作成するミューテックス名
	15	protocol	マルウェアの通信規約
C	16	subject	メールの件名
	17	transit	経由したメールサーバ
	18	X-Mailer	端末で用いられたソフトウェア名
	19	time zone	端末の時刻帯
	20	from	メールの送信者
	21	attached file	メールの添付ファイル名
	22	specimen	解凍後の添付ファイル名

次に，標的型攻撃に関連するパラメータを**表 1** に示す．これらのパラメータは，マルウェアのコマンド&コントロールサーバ等のブラックリストで共有されている情報や，セキュリティ関係企業のマルウェアの解析サービスで提供される情報を参考として選定し，図 1 で示した概念に基づいて分類したものである．

グループ A のパラメータ 1〜7 は，標的型攻撃に用いられるマルウェアの情報の送信先に関する情報である．1 はマルウェアが接続するコマンド&コントロールサーバのホスト名，2 はその IP アドレス，3 は 1 のドメイン名を示す．

4〜6 は 3 のドメイン名に関する情報であり，7 は 2 の IP アドレスの所有者を示す．これらのパラメータの特徴は，変更や維持にある程度の手間とコストがかかるため，容易に変更することができないということである．標的型攻撃の目的が情報の搾取であった場合，これらは搾取した情報の送信先に関係する情報であることを考慮すると，真の攻撃者を追求するための最も信頼性が高い情報であると考えられる．

グループ B のパラメータ 8〜15 は，標的型攻撃に用いられるマルウェアの挙動に関する情報である．8〜11 は大手ベンダのウイルス対策ソフトにおけるマルウェアの検知名を示す．12〜14 はマルウェアのコンピュータ内部での挙動に関する情報であり，各々作成する一時ファイルの名称，レジストリの名称およびミューテックスの名称を示す．15 はマルウェアがコマンド&コントロールサーバとの通信に利用する通信規約を示す．これらのパラメータも，変更や維持にある程度の手間とコストがかかるため，容易に変更することは困難であると考えられる．しかしながら，近年指摘されているウイルス作成ツールの存在 [7] を考慮すると，必ずしも攻撃者に結び付く情報とはいえない．異なる攻撃者が，同じツールを使用してマルウェアを作成した可能性も考えられるためである．

グループ C のパラメータ 16〜22 は，標的型攻撃に用いられるマルウェアの送信元に関する情報である．16 はメールの件名，17 は当該メールが経由したメールサーバの IP アドレスを示す．18 はメールを送信した端末で用いられたソフトウェアの名称，19 はその時刻帯を示す．20 はメールの送信者，21 は添付ファイルの名称，22 は添付ファイルが圧縮されている場合の解凍後のファイルの名称を示す．添付ファイルが圧縮されていない場合，21 と 22 は同じ値となる．これらのパラメータの特徴は，変更や維持にほとんど手間やコストがかからないため，容易に変更することが可能であるということである．しかしながら，メールの件名や添付ファイルの名称には，攻撃者の特徴が現れる可能性は否定できない．

3.2 パラメータの数値化

表 1 に示した標的型攻撃に関するパラメータは名義尺度（カテゴリデータ）であり，そのままでは主成分分析を実施することはできない．そこで，標的型攻撃に関するパラメータを以下の手順で数値化する．

STEP1 あるパラメータを選択し，その要素の値で昇順にソートする．

STEP2 ユニークな要素の値に対し，1 から順に数値を付与する．

STEP3 そのパラメータの要素の値を，対応する付与した数値に置き換える．

STEP4 STEP1〜STEP3 をすべてのパラメータに対して実施する．

基本的に，名義尺度であるパラメータの比較結果は一致するか否かである．したがって，数値化した値の差，比率および順序を比較することに意味はない．そこで本研究では，名義尺度であるパラメータを昇順にソートし，1 から順に数値を付与することで，類似の名義尺度が相対的に近い数値となるように数値化する．この工夫により，数値の差や比率には意味はないが，順序には意味が生じる．よって，主成分分析や因子分析を実施することが可能となる．

本研究ではさらに，部分一致した場合にも相互に関係があると考えられるパラメータがある点に着目する．たとえば，コマンド&コントロールサーバの IP アドレスのネットワークアドレス部分が一致し，ホストアドレス部分が一致しない場合，その IP アドレスは部分一致することになる．この例の場合，コマンド&コントロールサーバは同一のネットワークに存在している．ゆえに，そのネットワークで動的に IP アドレスを割り当てている場合には，コマンド&コントロールサーバの IP アドレスは容易に変更することが可能となる．よってこのような場合には，比較結果は一致すると考えることも可能である．この部分一致の判定は，表 1 において IP アドレスを示すパラメータである 2 および 17 に対して適用する．また，マルウェアが複数のコマンド&コントロールサーバや DNS サーバを用いる場合，一部のサーバのみ一致する攻撃も相互に関係すると見なすべきである．この部分一致の判定は，表 1 において複数の値から構成される可能性があるパラメータである 1〜8 に対して適用する．ほかの例としては，同一の固有名称に，括弧や記号が付加される場合等がある．このような場合には，最長共通部分文字列を検索して部分一致を検出することとする．この部分一致の判定は，表 1 における 1 および 3〜7 に対して適用する．部分一致の判定を適用するパラメータに関しては，STEP2 において部分一致する要素をユニークとは見なさずに数字を付与し，STEP3 において同一の数値に置き換える．

標的型攻撃の数を n とすると，この手順により，標的型攻撃に関するパラメータである n 行 22 列の名義尺度が，n 行 22 列の順序尺度に数値化される．本研究で分析の対象とする約 500 件の標的型攻撃に関するデータを，これまでに示した手法で数値化した．数値化の際に，部分一致して同一の値に数値化されたパラメータとその件数は**表 2** に示すとおりである．表中の件数は，数値化した後のユニークな値の件数であり，部分一致件数は同一の値に数値化された件数を示す．1〜3 のコマンド&コントロールサーバに関するパラメータについては，全体の半数弱のユニークな値に数値化された．部分一致の内訳については，そのほとんどが複数のコマンド&コントロールサーバを用いており，一部のサーバが一致したものであった．4〜7 の DNS サーバおよび所有者に関するパラメータについては，コマ

表 2 部分一致したパラメータとその件数
Table 2 The parameters that partially matched and the number.

No.	Parameter	件数	部分一致件数
1	host name	195	11
2	IP address	206	30
3	domain name	149	16
4	DNS server	69	19
5	resistrant	93	22
6	resistrar	68	21
7	address owner	122	8
8	virus name 1	68	-
9	virus name 2	311	-
10	virus name 3	129	-
11	virus name 4	12	-
12	temp file	103	-
13	registry	35	-
14	mutex	26	-
15	protocol	32	-
16	subject	327	-
17	transit	237	94
18	X-Mailer	99	-
19	time zone	11	-
20	from	310	-
21	attached file	344	-
22	specimen	348	-

ンド&コントロールサーバに関するパラメータよりもさらに多くの重複が確認できた．これらのパラメータの部分一致も，ほとんどが複数のコマンド&コントロールサーバを用いていたために一致したものであった．この結果から，同一の DNS サーバおよび所有者により，複数のコマンド&コントロールサーバが用いられているものと推定できる．8〜11 のウイルス対策ソフトの検知名については，各ベンダで検知名の命名規則が異なるため，ユニークな件数は様々であった．12〜15 のウイルスの挙動に関するパラメータについては，作成するファイル名については件数がやや多い一方で，通信規約のように件数が少ないパラメータもあることから，類似の挙動を示す亜種が含まれているものと考えられる．16〜22 のマルウェアの送信元に関するパラメータについては，容易に変更することが可能であることから，ユニークな件数に顕著な傾向は見られなかった．経由したメールサーバの部分一致については 5 件に 1 件の割合と最も多く，同じサービスの負荷分散されたメールサーバの IP アドレスが部分一致したものがほとんどであった．

3.3 主成分分析

数値化した約 500 件の標的型攻撃に関するデータについて，主成分分析を実施した．主成分分析の計算には，R [8] の prcomp 関数を用いた．第 5 主成分までの主成分分析の結果の概要を**表 3** に示す．寄与率は，第 1 主成分で 64.2%程度の低い値であった．このことから，すべての標的型攻撃に関するパラメータが，同じ傾向を示すわけではないことが予想できる．累積寄与率は，第 4 主成分から第 5 主成分あたりで 90%以上の値となった．換言すると，4 つ〜5 つの主成分で，標的型攻撃に関するパラメータの 9 割が説明できることになる．しかしながら，主成分分析の目的は，できるだけ少ない主成分に変数を集約することにあるため，相関が低い変数も含まれやすい傾向がある．よって，各主成分と各パラメータの相関から各主成分が何であるかを解釈することは困難である．各主成分の解釈を容易にするためには，因子分析を実施するのが適当である．

表 3 主成分分析の結果の概要
Table 3 A summary of principal component analysis.

	第 1 主成分	第 2 主成分	第 3 主成分	第 4 主成分	第 5 主成分
標準偏差	235.5	116.5	71.5	54.4	44.7
寄与率	0.642	0.157	0.059	0.034	0.023
累積寄与率	0.642	0.799	0.858	0.892	0.915

4. 因子分析によるパラメータの抽出

主成分分析の結果，4 つ〜5 つの主成分で，標的型攻撃に関するパラメータを説明できる可能性が示された．そこで，標的型攻撃のパラメータの傾向を説明できる因子を探索するために，因子分析を実施する．さらに，因子分析の結果を用いて攻撃者と相関が高いパラメータを抽出する．

4.1 探索的因子分析

主成分分析を実施した約 500 件の標的型攻撃に関する n 行 22 列の多次元ベクトルに対し，探索的因子分析 [9] を実施した．因子の抽出法は最尤法 [10], [11] とし，因子の回転は斜交回転 [12] の一手法であるプロマックス法で実施した．探索的因子分析の計算には，R [8] の pfa [13] 関数を用いた．その結果，固有値 1.0 以上を基準とすると，4 因子構造が妥当であるという結論に達した．4 因子を仮定した探索的因子分析の因子負荷量[*1]を**表 4** に示す．

第 1 因子は，当該メールが経由したメールサーバ，メールを送信したソフトウェアの名称，添付ファイルの名称と強い相関があり，時刻帯および送信者とやや強い相関がある．また，マルウェアが接続するコマンド&コントロールサーバのホスト名およびドメイン名との弱い相関も認められる．

第 2 因子は，マルウェアが接続するコマンド&コントロールサーバの DNS サーバ，ドメイン登録者，レジストラおよび IP アドレスの所有者と強い相関があり，コマンド&コントロールサーバの IP アドレスとやや強い相関がある．また，ウイルス対策ソフト 1 の検知名との弱い相関も認めら

[*1] 因子と分析に使用した変数との相関係数に相当する値

表 4 探索的因子分析の因子負荷量
Table 4 Factor loadings of exploratory factor analysis.

Parameter	第 1 因子	第 2 因子	第 3 因子	第 4 因子
host name	0.234	0.023	−0.228	0.197
IP address	0.087	0.613	0.213	0.116
domain name	0.358	0.220	0.195	0.315
DNS server	0.041	0.902	−0.033	0.023
resistrant	−0.098	0.952	0.057	−0.116
resistrar	−0.057	0.974	0.058	−0.112
address owner	−0.105	0.944	0.015	−0.069
virus name 1	0.139	0.327	−0.295	0.211
virus name 2	0.116	0.125	−0.157	0.252
virus name 3	0.016	−0.126	0.170	0.984
virus name 4	−0.032	−0.157	0.269	0.685
temp file	−0.316	−0.035	0.648	0.268
registry	0.221	0.153	0.665	0.032
mutex	0.136	0.009	0.854	−0.101
protocol	0.093	0.021	0.689	−0.135
subject	0.003	0.121	0.343	0.370
transit	0.969	−0.040	0.066	−0.019
X-Mailer	0.831	−0.022	−0.009	−0.063
time zone	0.570	−0.126	0.008	0.133
from	0.557	0.060	−0.020	−0.068
attached file	0.957	−0.030	0.079	−0.079
specimen	0.931	−0.074	0.023	0.043

れる．

第 3 因子は，マルウェアが作成するミューテックス名と強い相関があり，一時ファイル名，レジストリ名および通信規約とやや強い相関がある．また，コマンド&コントロールサーバの IP アドレス，ウイルス対策ソフト 4 の検知名およびメールの件名との弱い相関も認められる．

第 4 因子は，ウイルス対策ソフト 3 の検知名と強い相関があり，ウイルス対策ソフト 4 の検知名とやや強い相関がある．また，コマンド&コントロールサーバのドメイン名，ウイルス対策ソフト 1 および 2 の検知名，マルウェアが作成する一時ファイル名およびメールの件名との弱い相関も認められる．

4.2 検証的因子分析

次に，探索的因子分析で発見した 4 つの因子と各パラメータの関係を，検証的因子分析[*2]で確認する．探索的因子分析の結果，相関が認められた各因子とパラメータの間に相関があることを仮定し，検証的因子分析を実施した．検証的因子分析の計算には，R [8] の cfa 関数 [14] を用いた．その結果，RMSEA（Root Mean Square Error of Approximation）[*3]の値は 0.075 となった．したがって，検証的因子分析の結果は必ずしも最適というわけではないが，許容範囲内であると判断できる．必ずしも最適とはならなかった要因の 1 つには，第 3 因子と第 4 因子にやや強い相関があり，両者を同一の因子と解釈する見方もありうるということが考えられる．両者を同一の因子と解釈する見方に関しては，後に実施する因子の命名時に再度考察する．検証的因子分析の各パラメータの因子負荷量を**表 5** に示す．また，その因子間相関行列を**表 6** に示す．

[*2] ある程度の仮説が設定されており，変数に基づいて仮説とした因子構造が妥当かどうかを検証する手法
[*3] モデルの分布と真の分布との乖離を示す指標であり，一般 0.05 以下であればあてはまりがよく，0.1 以上であればあてはまりが悪いとされる．

表 5 検証的因子分析の因子負荷量
Table 5 Factor loadings of confirmatory factor analysis.

Parameter	第 1 因子	第 2 因子	第 3 因子	第 4 因子
host name	0.263	0.000	0.000	0.000
IP address	0.000	0.568	0.378	0.000
domain name	0.381	0.300	0.000	0.390
DNS server	0.000	0.897	0.000	0.000
resistrant	0.000	0.898	0.000	0.000
resistrar	0.000	0.933	0.000	0.000
address owner	0.000	0.884	0.000	0.000
virus name 1	0.000	0.240	0.000	0.170
virus name 2	0.000	0.000	0.000	0.160
virus name 3	0.000	0.000	0.000	0.999
virus name 4	0.000	0.000	0.103	0.667
temp file	0.000	0.000	0.474	0.146
registry	0.000	0.000	0.903	0.000
mutex	0.000	0.000	0.801	0.000
protocol	0.000	0.000	0.601	0.000
subject	0.000	0.000	0.416	0.353
transit	0.970	0.000	0.000	0.000
X-Mailer	0.787	0.000	0.000	0.000
time zone	0.603	0.000	0.000	0.000
from	0.533	0.000	0.000	0.000
attached file	0.933	0.000	0.000	0.000
specimen	0.935	0.000	0.000	0.000

表 6 因子間相関行列
Table 6 Factor correlation matrix.

	第 1 因子	第 2 因子	第 3 因子	第 4 因子
第 1 因子	1.000	0.228	0.504	0.507
第 2 因子	0.228	1.000	0.574	0.438
第 3 因子	0.504	0.574	1.000	0.645
第 4 因子	0.507	0.438	0.645	1.000

各パラメータの因子負荷量には，探索的因子分析と比較して特に大きな変化は認められなかった．よって，探索的因子分析で発見した 4 つの因子と各パラメータの関係は妥当であると考えられる．各因子間の相関については，全般的にやや強い相関が認められた．しかしながら，第 1 因子と第 2 因子の相関のみ弱いという結果となった．

4.3 因子の命名

探索的因子分析および検証的因子分析の結果，4 つの因子を抽出し，各パラメータとの関係の妥当性を検討した．

次に，4つの因子と各パラメータとの関係を考察し，各因子に名称を付与する．

第1因子は，主として当該メールが経由したメールサーバ，メールを送信したソフトウェアの名称，時刻帯，送信者，添付ファイルの名称で構成されている．これらはいずれも，標的型攻撃に用いられるメールに関する情報であり，容易に変更することが可能である．しかも，第1因子は最も攻撃者に関係すると考えられるマルウェアの接続先に関する情報との相関が低い．よって第1因子は，すでにマルウェアに感染し，攻撃者に操られた被害者の端末である可能性が考えられる．したがって，第1因子を被害者因子と命名する．

第2因子は，主としてマルウェアが接続するコマンド&コントロールサーバのIPアドレス，DNSサーバ，ドメイン登録者，レジストラおよびIPアドレスの所有者で構成されている．これらは容易に変更することができない情報であり，最も攻撃者に関係する因子であると考えられる．したがって，第2因子を攻撃者因子と命名する．

第3因子は，主としてマルウェアが作成する一時ファイル名，レジストリ名，ミューテックス名および通信規約で構成されている．これらは標的型攻撃に用いられるマルウェアの挙動に関する情報であり，容易に変更することは困難であると考えられる．これらは第2因子である攻撃者因子との相関も高く，これは攻撃者とマルウェアの作成者が同一である可能性を示しているものと考えられる．したがって，第3因子をマルウェア作成者因子と命名する．

第4因子は，主としてウイルス対策ソフト3および4の検知名で構成されている．ウイルス対策ソフト1および2の検知名は，ベンダ独自の命名規則に基づいて決定されている．これに対し，ウイルス対策ソフト3および4の検知名は，脆弱性の名称を基に決定される場合が多いという特徴がある．ゆえに，第4因子を脆弱性因子と命名する．脆弱性因子は，第3因子であるマルウェア作成者因子との相関も高いため，マルウェア作成者因子に抱合して考えても差し支えないであろう．

4.4 パラメータの抽出

検証的因子分析のパス図を**図 2** に示す．図中の太い実線は0.8以上の強い相関を示し，細い実線は0.8～0.4のやや強い相関を示している．また，破線は0.4未満の弱い相関を示している．

この結果から，第1因子である被害者因子は，ほかの因子との相関が低く，独立していると解釈することができる．被害者因子とコマンド&コントロールサーバのホスト名およびドメイン名との弱い相関は，真の攻撃者が表面的に被害者を偽っている可能性を示しているものと考えられる．しかも，第1因子は攻撃者を示す第2因子との相関が最も低い．よって攻撃者を識別するためには，第1因子を除外するのが妥当であると考えられる．

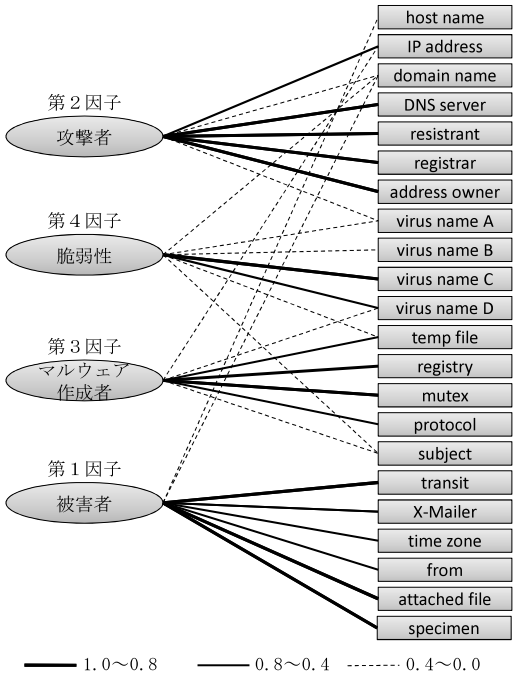

図 2 因子分析のパス図
Fig. 2 A path diagram of factor analysis.

第2から第4因子である攻撃者，マルウェア作成者および脆弱性因子は，表6によると因子間の相関がやや強いことからも，相互に関係しているものと解釈することができる．標的型攻撃の中には，一部の第2因子とのみ相関が強い特定のマルウェアや，脆弱性も確認されている．よってこれらの因子間の相関は，攻撃者が自ら脆弱性を収集して活用し，マルウェアを作成している可能性を示しているものと考えられる．したがって，真の攻撃者を識別するために有効なパラメータは，第2から第4因子に関係するパラメータであると考えられる．

5. クラスタ分析による分類

主成分分析および因子分析により，標的型攻撃において攻撃者に関係すると考えられるパラメータを抽出した．次に，各パラメータの相関の強さに応じた優先度を考慮し，クラスタ分析により標的型攻撃を攻撃者ごとに分類する．

5.1 優先度の決定

各パラメータの優先度を計算するために，第3および第4因子を，第2因子との相関比率で組み合わせる．攻撃者を示す第2因子の因子負荷量はそのまま用いる．第3因子および第4因子の因子負荷量は，各因子と第2因子との相関比率と乗算する．各因子の因子負荷量と相関比率を乗算した値の合計を，最終的な各パラメータの優先度とする．各パラメータの優先度の計算結果を**表 7** に示す．表中の数値は，各因子の因子負荷量と相関比率を乗算した値である．

因子負荷量は，因子分析における相関係数に相当する値

表 7 各パラメータの優先度
Table 7 Priority of parameters.

Parameter	第 2 因子	第 3 因子	第 4 因子	優先度
host name	0.000	0.000	0.000	0.000
IP address	0.568	0.217	0.000	0.785
domain name	0.300	0.000	0.171	0.471
DNS server	0.897	0.000	0.000	0.897
resistrant	0.898	0.000	0.000	0.898
resistrar	0.933	0.000	0.000	0.933
address owner	0.884	0.000	0.000	0.884
virus name 1	0.240	0.000	0.074	0.314
virus name 2	0.000	0.000	0.070	0.070
virus name 3	0.000	0.000	0.438	0.438
virus name 4	0.000	0.059	0.292	0.351
temp file	0.000	0.272	0.064	0.336
registry	0.000	0.518	0.000	0.518
mutex	0.000	0.460	0.000	0.460
protocol	0.000	0.345	0.000	0.345
subject	0.000	0.239	0.155	0.393
transit	0.000	0.000	0.000	0.000
X-Mailer	0.000	0.000	0.000	0.000
time zone	0.000	0.000	0.000	0.000
from	0.000	0.000	0.000	0.000
attached file	0.000	0.000	0.000	0.000
specimen	0.000	0.000	0.000	0.000
相関比率	1.0	0.574	0.438	-

である．したがって，攻撃者を示す第 2 因子の因子負荷量は，各パラメータがどの程度攻撃者と関連しているかを示す値である．マルウェア作成者を示す第 3 および脆弱性を示す第 4 因子の因子負荷量も同様に，各パラメータがどの程度各因子と関連しているかを示す値である．したがって，これらの値と各因子と第 2 因子の相関を乗算した値は，第 3 および第 4 因子のパラメータと攻撃者がどの程度関連しているかを示す値となる．よって最終的に，攻撃者に関連すると考えられる各因子の因子負荷量と相関比率を乗算した値は，各パラメータが攻撃者とどの程度関連しているかを示す値と解釈できる．

5.2 優先度を考慮した数値化

因子分析に用いた n 行 22 列の多次元ベクトルでは，パラメータの優先度を考慮することができない．そこで，因子分析で抽出したパラメータの優先度を考慮した別の数値化手法を用いる．以下にその手順を示す．

STEP1 標的型攻撃相互のパラメータを比較し，完全一致または部分一致するパラメータを検出する．

STEP2 完全一致または部分一致するパラメータの優先度の総和を計算する．

STEP3 STEP1〜STEP2 をすべての標的型攻撃相互の組合せに対して実施する．

標的型攻撃の数を n とすると，この手順により，n 行 n 列の多次元ベクトルが得られる．

表 8 コーフェン相関係数
Table 8 Cophnetic correlation coefficient.

分析手法	優先度なし	優先度あり
最短距離法	0.782	0.927
最長距離法	0.496	0.970
群平均法	0.513	0.980
ウォード法	0.776	0.951

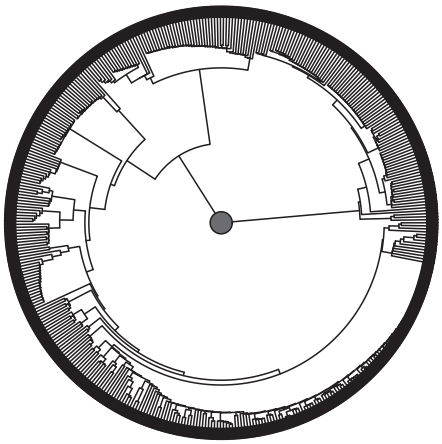

図 3 群平均法によるクラスタ分析の樹形図
Fig. 3 A dendrogram of the cluster analysis by average method.

5.3 クラスタ分析

因子分析に用いた約 500 件の標的型攻撃に関するパラメータを，パラメータの優先度を考慮した手法で n 行 n 列の多次元ベクトルに数値化し，階層的クラスタリングを実施した．各クラスタ間の距離はユークリッド距離で計算し，最短距離法，最長距離法，群平均法およびウォード法を用いた．各手法のコーフェン相関係数[*4]を**表 8** に示す．パラメータの優先度を考慮しない n 行 22 列の多次元ベクトルのクラスタ分析を実施した場合には，コーフェン相関係数は約 0.4〜0.8 のやや低い値となった．これに対し，パラメータの優先度を考慮した手法では，コーフェン相関係数は 0.9 以上のきわめて高い値となった．したがって，いずれの手法においてもパラメータの優先度を考慮した方が，クラスタ分析による歪みが小さくなるということが確認できた．特に，群平均法においては 0.980 の最も高い数値が得られた．群平均法によるクラスタ分析の樹形図は**図 3** のとおりである．図中の円周上には約 500 件の各標的型攻撃の識別番号が並んでおり，大きい半径の弧で結ばれているほど似ており，中心に近い弧で結ばれているほど類似度が低いことを意味している．群平均法では，距離 6 で 20 以上の攻撃が結合した大きなクラスタが 4 つ得られた．その中で最大のクラスタには，132 の攻撃が結合されていた．以

[*4] クラスタ分析の評価指標の 1 つであり，値が大きいほど距離行列と用いた方法のコーフェン行列との歪みが小さいといえる．

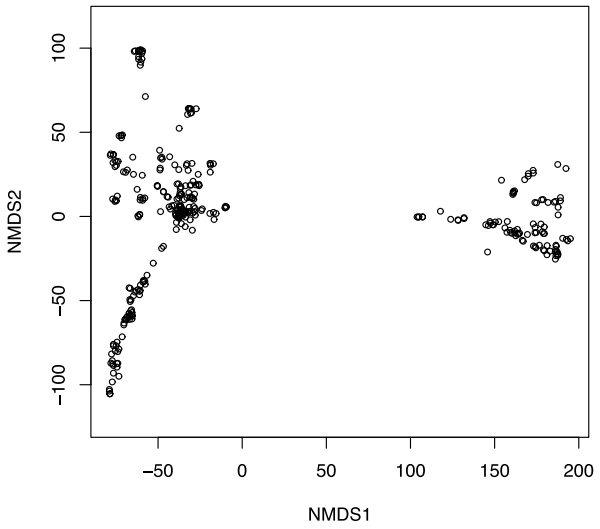

図 4 多次元尺度構成法による攻撃の散布図 1

Fig. 4 Scatterplots of the attacks by multi dimensional scaling 1.

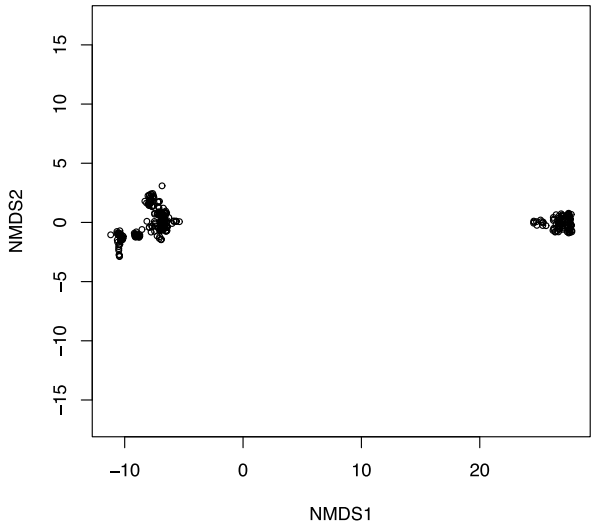

図 5 多次元尺度構成法による攻撃の散布図 2

Fig. 5 Scatterplots of the attacks by multi dimensional scaling 2.

下，同様の距離で大きな順に，33，31，23 の攻撃が 1 つのクラスタに結合されていた．ブートストラップ法 [15] による 1,000 回のリサンプリングの結果，ブートストラップ標本が仮説を支持する相対頻度（ブートストラップ確率）[16] が 95%以上の強固なクラスタ数は 2 であった．

さらに，分類の効果を多次元尺度構成法（MDS：Multi Dimensional Scaling）[*5]で視覚的に確認する．各相関ベクトル間の距離はユークリッド距離，次元数を 2 とし，非計量多次元尺度法である metaMDS [17] を用いた．パラメータの優先度を考慮しない場合の配置図を**図 4**，考慮した場合の配置図を**図 5** に示す．パラメータの優先度を考慮しない場合には，NMDS1 および NMDS2 のいずれの次元においても値の拡散が認められる．Kruscal が提案する多次元尺度構成法の評価基準であるストレスの値は 0.087 であり，分類のあてはまりはまずまずであると評価できる [18]．これに対し，優先度を考慮した場合には，NMDS2 の次元における値の収束が認められる．また，ストレスの値も 0.037 に改善され，分類のあてはまりが良い結果となった．

[*5] 分類対象物の関係を低次元空間における点の布置で表現する手法

6. 考察

6.1 因子分析の妥当性

過去 3 年間の標的型攻撃に関するパラメータの因子分析を実施した結果，被害者因子，攻撃者因子，マルウェア作成者因子および脆弱性因子の 4 つの因子を発見した．各因子間の相関は全般的にやや高いが，被害者因子と攻撃者因子の間でのみ低いという結果となった．被害者因子とコマンド&コントロールサーバのホスト名およびドメイン名との弱い相関に関しては，真の攻撃者が表面的に被害者を偽っている可能性を考慮すれば説明ができる．近年ではダイナミック DNS 等のサービスを利用すれば，コマンド&コントロールサーバのホスト名およびドメイン名を容易に変更することが可能である．したがって，真の攻撃者が表面的に被害者を偽ってコマンド&コントロールサーバのホスト名およびドメイン名を設定することは容易である．残るウイルス対策ソフトの検知名，メールの件名等の独立因子に関しては，いずれも各因子との関係を決定づける合理的な理由はない．ゆえに，攻撃者を識別するためには，攻撃者因子，マルウェア作成者因子および脆弱性因子を構成するパラメータを利用し，被害者因子を構成するパラメータを除外するのが妥当であると考えられる．

興味深いのは，攻撃者因子とコマンド&コントロールサーバのホスト名およびドメイン名との相関が低いという点である．また，攻撃者因子とコマンド&コントロールサーバの IP アドレスとの相関もそれほど高いわけではない．これらの事実は，コマンド&コントロールサーバのホスト名，ドメイン名および IP アドレスは，攻撃者と必ずしも強い関係にあるわけではないという可能性を示している．したがって，コマンド&コントロールサーバのホスト名，ドメイン名および IP アドレスはどこから攻撃されているのかを判断するための信頼できる指標にはなりえず，あくまでも参考として解釈するのが妥当であると考えられる．

6.2 クラスタ分析の妥当性

パラメータの絞り込みを実施せず，優先度を考慮しないでクラスタ分析を実施した場合には，距離行列と用いた各手法のコーフェン行列との歪みが大きい結果となった．また，多次元尺度構成法においても，優先度を考慮した方が分類のあてはまりが良い結果となった．これは，各パラメータの特徴を評価せず，パラメータの優先度を考慮していないためであると推定できる．各パラメータの攻撃者との関係の強さ，情報の信頼性の高さは明らかに同一ではな

い．攻撃者との関係が強いパラメータと弱いパラメータ，または信頼性が高いパラメータと低いパラメータを同じ優先度で評価した場合，分類結果には多量のノイズが含まれることになる．ゆえに，攻撃者との関係が弱いパラメータは除外し，攻撃者との関係の強さおよび情報の信頼性に応じた優先度を考慮するのが妥当である．

因子分析の結果を考慮してパラメータを絞り込み，優先度を考慮してクラスタ分析を実施した場合には，各手法におけるコーフェン相関係数の値はいずれも 0.9 以上となり，距離行列と用いた各手法のコーフェン行列との歪みはきわめて小さいという結果に改善された．これは，各パラメータの特徴を評価し，パラメータの優先度を考慮したためであると考えられる．攻撃者との関係が弱いパラメータを除外し，攻撃者との関係の強さおよび情報の信頼性に応じた優先度を設定することで，分類結果からノイズを除去することができたものと推定できる．ただし，歪みが小さくなったことにより，必ずしも分類の結果が正確になったわけではない．

次に，群平均法により得られた 4 つの大きなクラスタと実際のデータの内容を確認する．これら 4 つのクラスタに結合された攻撃の実際のデータを確認すると，C & C サーバの DNS サーバが共通または重複しているものがほとんどであった．次に多かったのは，C & C サーバのドメイン名，ドメイン登録者，登録レジストラが共通または重複しているものであった．また，C & C サーバのホスト名や IP アドレスが共通している攻撃も多く含まれていた．これらの同一のクラスタに結合された攻撃で共通するパラメータは，いずれも高い優先度になっており，攻撃者との関係が強いと考えられるパラメータであった．したがって，これら 4 つのクラスタは，同一の攻撃者による一連の攻撃に合致しているものと考えられる．一方で，4 つのクラスタ間では，実際のデータから，攻撃者との関係が強いと考えられるパラメータに重複がないことを確認した．さらに，一時ファイルの名称，レジストリの名称，ミューテックスの名称等のマルウェアの挙動に関するパラメータについても重複がないことを確認した．これらの結果から，同一の攻撃者によるものと考えられる 4 つの一連の攻撃は，少なくとも相互に直接的な関係がない 4 名の別の攻撃者によるものであると推定される．同様に得られたクラスタと実際のデータの内容の確認を，異なる距離およびほかのクラスタについても実施したところ，やはり同様の結論が得られる分析結果となった．したがって，クラスタ分析による分類結果は，正しく攻撃者ごとに分類できているものと判断できる．

7. おわりに

本研究では，近年複雑化している標的型攻撃の真の攻撃者を識別するために，複数の標的型攻撃に関するパラメータの相関を分析し，攻撃者ごとに分類することを目的とした．そのために，複数の標的型攻撃に関するパラメータを多次元ベクトルに数値化し，主成分分析により傾向を明らかにした．次に，因子分析を用いて標的型攻撃を説明する 4 つの因子を発見し，攻撃者と相関が高いパラメータを抽出した．さらに，因子分析で抽出したパラメータを，優先度を考慮して多次元ベクトルに数値化し，クラスタ分析によって複数の標的型攻撃を攻撃者ごとに分類した．最後に，因子分析およびクラスタ分析の結果の妥当性について考察した．

この論文では，約 500 件の大量の標的型攻撃に関するデータを，多変量解析によって攻撃者ごとに分類した．標的型攻撃に関するデータを攻撃者ごとに分類するためには，相互の関連性を分析して攻撃者を識別する必要がある．標的型攻撃の相互の関連性を分析するためには，あるパラメータに注目し，共通点を探すというアプローチが一般的である．このような地道なアプローチは，個々の標的型攻撃の関連性を調査するためには有効であるが，大量の標的型攻撃に関するデータを分類する用途には適しているとはいえない．共通点を探すというアプローチでは，どのパラメータに注目するかが重要である．また，複数のパラメータに共通点があった場合，どのパラメータを優先するのかという判断も重要である．どのパラメータに注目し，どのパラメータを優先するかという判断は，従来は分析者の裁量に委ねられる場合がほとんどであった．大量のデータの複数のパラメータの関連性を調査する場合には，分析者が各パラメータの攻撃者との関係の強さを考慮し，使用するパラメータを合理的に決定することは困難である．この論文で示した多変量解析による標的型攻撃の分類手法は，分析者が大量の標的型攻撃に関するデータを短時間で分類したい場合に有用である．その有用性は，統計データに基づいて優先すべきパラメータを抽出し，その優先度を合理的かつ定量的に決定し，さらにその結果を用いて自動的に分類できる点にある．標的型攻撃に関する情報は一般的に公開されることは少ないが，情報共有の重要性が指摘されており，実際に情報共有の取り組みも本格化している [19]．情報共有により得られた大量のデータを分析するためには，この論文で実施した多変量解析による標的型攻撃の分類手法は，きわめて有用であると考える．

今後の課題としては，分類結果の正確性の評価があげられる．しかしながら，そのためには攻撃者に関する正しい情報が必要であり，現実的には実現は困難である．攻撃者に関する完全な正しい情報が得られる可能性は，きわめて低いものと考えられる．しかしながら近年では，サイバー攻撃の方法，使用したマルウェア，コマンド&コントロールサーバのドメイン名や IP アドレス等の具体的に示し，攻撃者を特定しようとする試みも報告されるようになってきている [20]．このように，攻撃者に関する断片的な情報が

得られた場合には，その情報と攻撃者との関係の強さや情報の信頼性を考慮し，優先度を再設定して分類を実施すればよい．本研究で用いた多変量解析による分類手法は，今後様々な追加情報が得られた場合にも有効に活用することができる汎用的な手法である．このように，新たに得られた情報を評価し，優先度を考慮することで，分類結果はより正確で妥当なもの近づくものと考えられる．たとえ攻撃者に関する完全で正しい情報が得られない場合でも，本研究で用いた分類手法を用いれば，攻撃者との関連性を定量的に評価した分類結果を得ることができる．この関連性を定量的に評価した分類結果から，攻撃者の特徴を整理し，識別に役立てることが可能になるものと考える．

参考文献

[1] 経済産業省：最近の動向を踏まえた情報セキュリティ対策の提示と徹底 (online)，入手先 ⟨http://www.meti.go.jp/press/2011/05/20110527004/20110527004.html⟩（参照 2013-06-03）．

[2] 情報処理推進機構：『新しいタイプの攻撃』に関するレポート—Stuxnet（スタックスネット）等の新しいサイバー攻撃手法の出現 (online)，入手先 ⟨http://www.ipa.go.jp/about/technicalwatch/20101217.html⟩（参照 2013-06-03）．

[3] 情報処理推進機構：『新しいタイプの攻撃』の対策に向けた設計・運用ガイド 改定第 2 版 (online)，入手先 ⟨http://www.ipa.go.jp/security/vuln/newattack.html⟩（参照 2013-06-03）．

[4] 堀合啓一，今泉隆文，田中英彦：マルウェア亜種の動的挙動を利用した自動分類手法の提案と実装，情報処理学会論文誌，Vol.50, No.4, pp.1321–1333 (2009).

[5] 岩村 誠，伊藤光恭，村岡洋一：機械語命令列の類似性に基づく自動マルウェア分類システム，情報処理学会論文誌，Vol.51, No.9, pp.1622–1632 (2010).

[6] Hutchins, E.M., Cloppert, M.J. and Amin, R.M.: Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains, *Proc. 6th Annual International Conference on Information Warfare and Security* (2011).

[7] 情報処理推進機構：脆弱性を利用した新たなる脅威の監視・分析による調査 (online)，入手先 ⟨http://www.ipa.go.jp/security/vuln/report/newthreat200907.html⟩（参照 2013-06-03）．

[8] The R Project for Statistical Computing (online), available from ⟨http://www.r-project.org/⟩ (accessed 2013-06-03).

[9] Holzinger, K.J. and Harman, H.H.: *Factor Analysis: A Synthesis of Factorial Methods*, *1st edition*, University of Chicago Press (1941).

[10] Joreskog, K.: Some contributions to maximum likelihood factor analysis, *Psychometrika*, Vol.32, pp.443–482 (1967).

[11] Lawley, D.N. and Maxwell, A.E.: Factor Analysis as a Statistical Method, *Journal of the Royal Statistical Society. Series D* (*The Statistician*), Vol.12, No.3, pp.209–229 (1962).

[12] Thurstone, L.L. and Thurstone, T.G.: Factorial Studies of Intelligence, *Psychometric Monograph*, No.2 (1941).

[13] 青木繁伸：因子分析 (online)，入手先 ⟨http://aoki2.si.gunma-u.ac.jp/R/pfa.html⟩（参照 2013-06-03）．

[14] 青木繁伸：検証的因子分析 (online)，入手先 ⟨http://aoki2.si.gunma-u.ac.jp/R/cfa.html⟩（参照 2013-06-03）．

[15] Bradley, E.: Bootstrap Methods: Another Look at the Jackknife, *The Annals of Statistics*, Vol.7, No.1, pp.1–26 (1979).

[16] 下平英寿：ブートストラップ法によるクラスター分析のバラツキ評価，統計数理，Vol.50, No.1, pp.33–44 (2002).

[17] Edwards, J. and Oman, P.: Dimensional Reduction for Data Mapping-A practical guide using R, *R News*, Vol.3, No.3, pp.2–7, available from ⟨http://cran.r-project.org/doc/Rnews/Rnews_2003-3.pdf⟩ (accessed 2013-06-03).

[18] Kruskal, J.B.: Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis, *Psychometrika*, Vol.29, pp.1–27 (1964).

[19] 情報処理推進機構：サイバー情報共有イニシアティブ（J-CSIP）2012 年度活動レポート (online), 入手先 ⟨http://www.ipa.go.jp/security/J-CSIP/⟩（参照 2013-06-03）．

[20] Mandiant: APT1: Exposing One of China's Cyber Espionage Units, Mandiant (online), available from ⟨http://intelreport.mandiant.com/Mandiant_APT1_Report.pdf⟩ (accessed 2013-06-03).

三村 守 （正会員）

2001 年防衛大学校情報工学科卒業．同年海上自衛隊入隊．2008 年防衛大学校理工学研究科前期課程修了．同年海上自衛隊保全監査隊勤務．2011 年情報セキュリティ大学院大学博士後期課程修了．博士（情報学）．同年内閣官房情報セキュリティセンター出向．情報セキュリティ大学院大学客員研究員．マルウェア解析，標的型攻撃の相関分析に関する研究に従事．

田中 英彦 （名誉会員，フェロー）

1970 年東京大学大学院工学系研究科電気工学専門課程修了．工学博士．東京大学にて計算機アーキテクチャ，並列処理，人工知能，自然言語処理，分散処理，メディア処理等の教育・研究に従事．東京大学大学院情報理工学系研究科長を経て，2004 年情報セキュリティ大学院大学情報セキュリティ研究科長・教授に就任．2012 年より同学長・研究科長・教授．情報処理学会名誉員，人工知能学会論文賞，ACM SIGGRAPH'99 Impact Paper Award，人工知能学会功績賞，東京都科学技術功労者表彰，経済産業大臣表彰等受賞．情報・システム研究機構教育研究評議会評議員，日本学術会議会員，日本ネットワークセキュリティ協会（JNSA）会長，IEEE Life Fellow，東京大学名誉教授．