Vol.2012-CSEC-59 No.1
2012/12/14

因子分析による標的型攻撃の傾向調査

三村 守[1,a)]　田中 英彦[1,b)]

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．標的型攻撃に関する多くの情報から，攻撃者との関連が強いパラメータを抽出するためには，複数の標的型攻撃の相関を分析し，パラメータの傾向を知る必要がある．本稿では，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，主成分分析でその傾向の概略を調査する．さらに，因子分析により標的型攻撃の傾向を示し，標的型攻撃を構成する因子について考察するとともに，攻撃者との関連が強いパラメータを明らかにする．

キーワード：標的型攻撃，多変量解析，主成分分析，因子分析

Investigating bias in targeted attacks by factor analysis

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

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 select high correlation parameters between an attacker from much information about targetted attacks, we need to analyze correlation between targetted attacks and know the bias. In this paper, we express parameters of targetted attacks numerically and use principal component analysis to investigate an outline of the bias. Furthermore, we use factor analysis to indicate the bias, consider what the factors of targetted attacks are, and reveal high correlation parameters between an attacker.

Keywords: Targeted Attack, Multivariate Statistics, Principal Component Analysis, Factor Analysis

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%に留まっていたが，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) と呼ばれることもあり，大きな脅威となっている．

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，攻撃者の遠隔操作により任意

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

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，ファイアウォールやプロキシを介してコマンド＆コントロールサーバに送信される．最後に，コマンド＆コントロールサーバに送信された情報は攻撃者によって回収される．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．標的型攻撃に関する多くの情報から，攻撃者との関連が強いパラメータを抽出するためには，複数の標的型攻撃の相関を分析し，パラメータの傾向を知る必要がある．

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つく情報とは言えない．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つくものと考えられる．

そこで本稿では，近年複雑化している標的型攻撃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を数値化し，主成分分析でその傾向の概略を調査する．さらに，因子分析により標的型攻撃の傾向を示し，標的型攻撃を構成する因子について考察するとともに，攻撃者との関連が強いパラメータを明らかにする．

2. 関連研究

標的型攻撃の相関分析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挙げられる．

文献 [2] 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．本稿においても，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，数値化の手法は異なる．また，マルウェアの動的な挙動のみならず，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本稿とは異なる．

文献 [3] 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．本稿の分析対象はマルウェアだけでなく，標的型攻撃全般の広範囲であるため，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．

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．これに対し，本稿は標的型攻撃の傾向調査を目的としており，分析の対象はマルウェアのみならず，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広範囲となる点が，従来の研究とは最も大きく異なる．

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 パラメータの選定

主成分分析に用いる標的型攻撃のパラメータを表 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た．

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接続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振る舞い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 [4] を考慮すると，必ずしも攻撃者に結びつく情報とは言えない．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メール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．

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サーバの登録レジストラ
	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

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，その要素の値に一つの数値を付与する．

STEP2 次の要素の値が未知であれば新たな任意の数値を付与し，既知であれば同一の数値を付与する．

STEP3 STEP2 をそのパラメータのすべての要素に対して実施する．

STEP4 STEP1〜STEP3 をすべてのパラメータに対して実施する．

標的型攻撃の数を n とすると，この手順により，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 列の多次元ベクトルに数値化される．

3.3 主成分分析

2009 年から 2011 年の 3 年間にある複数の組織で発生した標的型攻撃において，分析に用いる 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．その選定した約 500 件の標的型攻撃に関する 22 のパラメータを，先に示した手法で n 行 22 列の多次元ベクトルに数値化し，主成分分析を実施した．主成分分析の計算には，R[5] 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 2 に示す．もとのパラメータが名義尺度であるため，標準偏差の値には絶対的な意味はない．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，４〜５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易い傾向がある．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 2 主成分分析の結果の概要
Table 2 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

主成分分析の結果，４～５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[*1]を実施した．因子の抽出法は最尤法[*2]とし，因子の回転[*3]はプロマックス法[*4]で実施した．探索的因子分析の計算には，R[5] の pfa[6] 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[*5]を表 3 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．

*1 因子の数，変数等を変え，試行錯誤を繰り返しながら因子を探索する手法
*2 標本データから母集団を推定する手法の一つであり，最も尤もらしい値を母集団の推定値とする手法
*3 仮に決めた因子空間の座標系を変換し，新しい座標系を決定する操作であり，データを解釈し易くするために実施する．
*4 回転後の因子軸が直交しない斜交回転の一手法であり，因子間に相関がある場合に用いられる．
*5 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 3 探索的因子分析の因子負荷量
Table 3 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

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[*6]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[5] の cfa 関数 [7] を用いた．その結果，RMSEA(Root Mean Square Error of Approximation)[*7]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考えられる．検証的因子分析の各パラメータの因子負荷量を表 4 に示す．また，その因子間相関行列を表 5 に示す．

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．

*6 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
*7 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 以下であれば当てはまりがよく，0.1 以上であれば当てはまりが悪いとされる．

表 4　検証的因子分析の因子負荷量
Table 4　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 5　因子間相関行列
Table 5　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	0.506	0.438	0.645	1.000

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振る舞い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 1 に示す．図中の太い実線は 0.8 以上の強い相関を示し，細い実践は 08〜0.4 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，他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因子との相関がもっとも低い．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．

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 5 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．

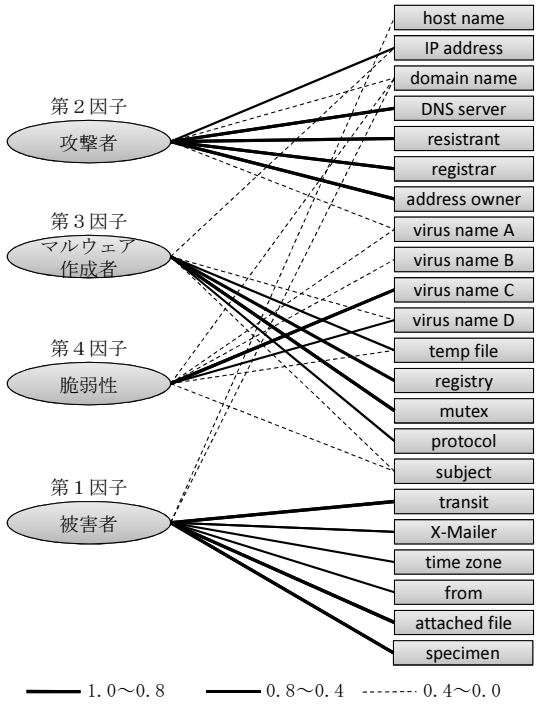

図 1 因子分析のパス図
Fig. 1 A path diagram of factor analysis

5. 考察

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．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て差し支えないであろう．

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，むしろ攻撃者との関係が低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指標にはなり得ず，あくまで参考程度に解釈するのが妥当であろう．

6. おわりに

本稿では，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，攻撃者との関連が強いパラメータを明らかにすることを目的とした．そのためにまず，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，主成分分析でその傾向の概略を調査した．さらに，因子分析により標的型攻撃の傾向を示し，標的型攻撃を構成する４つの因子を発見するとともに，攻撃者と相関が高いパラメータを抽出した．

今後の課題としては，攻撃者と相関が高いパラメータを使用し，実際に攻撃者毎に標的型攻撃を分類することが挙げられる．しかしながら，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．したがって，各パラメータの因子負荷量から，分類のための優先度を決定する必要があるものと考えられる．

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，経済産業省（オンライン），入手先 ⟨http://www.meti.go.jp/press/2011/05/20110527004/20110527004.html⟩（参照 2012-08-12）(2011).
[2] 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).
[3] 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).
[4] 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，情報処理推進機構（オンライン），入手先 ⟨http://www.ipa.go.jp/security/vuln/report/newthreat200907.html⟩（参照 2012-08-12）(2009).
[5] The R Project for Statistical Computing（オンライン），入手先 ⟨http://www.r-project.org/⟩（参照 2012-09-14）.
[6] 青木繁伸：因子分析（オンライン），入手先 ⟨http://aoki2.si.gunma-u.ac.jp/R/pfa.html⟩（参照 2012-09-14）.
[7] 青木繁伸：検証的因子分析（オンライン），入手先 ⟨http://aoki2.si.gunma-u.ac.jp/R/cfa.html⟩（参照 2012-09-14）.