ログ解析によるマルウェア侵入検知手法の提案

田中功一[†1] 堀川博史[†2] 峰野博史[†3] 西垣正勝[†3]

企業など組織におけるマルウェア侵入に対し，組織のネットワーク管理者を支援するために，インタネットなど組織外部に対して通信を中継するプロキシ装置のログに着目し，これを効率的に縮退する手法を検討した．考察では 1 億回/月，記録されたログから，グレイリストと呼ぶ 8000 件程度のアクセス先を抽出する方法を検討し，グレイリストを導き出すためのツールを作成した．グレイリストは，過去，マルウェアが侵入していなかったと想定する期間のアクセスログから切り出した，安全な通信先のみを含むと想定したアクセス先一覧と，侵入後の期間のアクセス先一覧を比較することで作成した．その結果，マルウェア侵入によってなされた通信と識別済みのアクセス先 URL を含むグレイリストを作成することができた．グレイリストはネットワーク管理者の目によって，疑わしいアクセス先か否かの判断材料に利用でき，膨大なログをすべて調査することなく効率的に検知を行う支援環境が構築できた．すべてのタイプのマルウェア侵入を検知することは難しいが，少なくとも組織外に対する不審な通信を調査する目的で，短時間で問題のある URL を検出する手段として利用できる．

A Study of Malware Intrusions Detection Method Based on Analysis of Proxy Log

KOICHI TANAKA[†1] HIROSHI HORIKAWA[†2]
HIROSHI MINENO[†3] MASAKATSU NISHIGAKI[†3]

1. はじめに

マルウェアの脅威は衰えず，標的型攻撃の対策が緊急課題となっている．ウイルス対策ソフトの多くは，パターン（シグネチャ）検索を用いることでウイルスか否かを判断しているが，標的型攻撃で利用されるマルウェアは標的に合わせてカスタマイズされているため，特定するに足りうるパターンを持たず，したがって通常のウイルス対策ソフトウェアでは検出できないことが多い[1]．この背景には，標的型攻撃のマルウェアは，シグネチャをすり抜ける為の個別のカスタマイズや Command and Control（C&C）とよばれるサーバからの操作といった，プログラムに人間の高度な判断が加わるため対策を難しくしていると考えられる．

このような，プログラムに人間の高度な判断が加わる攻撃に対して，防御側も従来のプログラムによる対策に加えて，人間の高度な判断を加えようというのが本稿のログ解析によるマルウェア侵入検知手法の提案である．2 章では標的型マルウェア検知の問題について，3 章では筆者らが実践しているマルウェア検知の体制・手順・ツールを説明する．4 章ではツールの効果について評価する．また，5 章は本稿のまとめである．

†1 静岡大学大学院 （博士課程）
Shizuoka University
†2 三菱電機インフォメーションテクノロジー
Mitsubishi Electric Information Technology Corp.
†3 静岡大学大学院
Shizuoka University

2. 標的型マルウェア検知の問題

2.1 標的型攻撃とは

攻撃自身は 2003 年から始まっているにも関わらず，誰が何のために行っているのかが長年判明されないでいたが，2011 年に米政府が「サイバー空間問題」[2]として扱い始めて以降，「誰が何のために」の答えが公開され始めた．標的型攻撃の意図背景は，組織情報の収集，監視と情報システムの破壊にあると想定される[3].国際対話の流れの中での問題テーマと想定脅威は，「社会インフラ破壊，重要情報窃取（軍事，知財等）」となっている．攻撃の核心部は侵入後に外部の攻撃者によって行われる，情報システム（組織内部）への侵入行為による情報の窃取・破壊である．従って多くの企業や組織にとって標的型攻撃は脅威となっている．なお，本稿では，使用者の不利益となる「不正な活動」を行うプログラムの総称，いわゆる「ウイルス」や「トロイの木馬」についても，マルウェアと呼称している．

一般に，企業などの組織におけるネットワーク管理者は，早期に標的型攻撃のマルウェアを検知，発見に尽力し，またマルウェアの通信を阻止，攻撃者の情報の窃取・破壊に至る過程を防止したいと考えている．

標的型攻撃には一般に表 1([3]を参考に作成)に示す①計画立案段階，②攻撃準備段階，③初期潜入段階，④基盤構築段階，⑤内部侵入・調査段階，⑥目的遂行段階，⑦再侵入の，7 つの段階があるといわれている．

従来の対策は，入口対策とよばれ，マルウェアを社内ネットワークに侵入させない対策であり，これは③初期潜入

段階に対応する．なお，入口対策とは，標的型攻撃の入口に対する対策の総称であり，偽装メール対策，マルウェア感染防止，脆弱性対策など，侵入されないことを前提とした対策を言う．

表 1 標的型攻撃の各段階と対策要否
Table 1 Steps of targeted attacks and countermeasures.

対策の段階	攻撃段階	内容
-	① 計画立案	攻撃目標の設定，事前準備
-	② 攻撃準備	標的型メール，C&C サーバの準備
従来からの対策	③ 初期潜入	標的型メールの送付
内部対策が必要	④ 基盤構築	ネットワークや端末の情報入手，バックドア構築
	⑤ 内部侵入・調査	他の端末への侵入，管理情報窃取
	⑥ 目的遂行	情報窃取，破壊
-	⑦ 再侵入	バックドアからの再侵入

ウイルス検知技術としては，三種に大別される技術が使われてきた．

(1) パターン検知型
ウイルスの特定のバイナリパターンをデータベース（DB）化し，それとの逐次照合による検知方法

(2) 振る舞い検知型
問題となる機能ブロックへのアクセスを察知し，不審な動きを検知し，通信を行っているモジュールが正規のものであるかどうか認証する手段とあわせて，不正な動きを察知する

(3) 不審サイトへのアクセス
不審もしくは危険な通信先を，あらかじめ DB 化し，通信先を照らし合わせることで不正な通信を知る

(1) パターン検知型と (3) 不審サイトへのアクセスは，ウイルス対策ソフト会社が事前に調査し，DB 化したものでないと，侵入されても察知することができないという問題を持つ．また，(2) 振る舞い検知型は，ウイルスが，一般の機能モジュールに巧妙に紛れ込む，また信頼されたソフトウェア，たとえばブラウザやメーラの一部など，あたかも OS 機能の一部として動作するように作成される時がある．この場合，侵入されてもどれが問題の発生源か気づきにくいという課題を持つ．さらに近年の「標的型マルウェア」はカスタマイズされたソフトウェアであり，(1)～(3) の技術を利用した，従来ソフトウェアでの検知が極めて困難になってきた．

マルウェアのカスタマイズに関しては，攻撃者はマルウェアがウイルス対策ソフトで検出される場合，侵入後の検出を避けるため，攻撃前にツールを使ってプログラムをカスタマイズする．カスタマイズツールは多数存在し，例えば，プログラムの動きは変えずにプログラムコードを圧縮するツールや，コードを改変して難読化するツール，コードを暗号化するツールなどがある．

標的型攻撃で使われるトロイの木馬型マルウェアは，侵入後に通信ポートを開設し，特定コンピュータと通信を行う．この特定コンピュータを C&C サーバとよび，C&C サーバとの通信をコネクトバック通信と呼ぶ．C&C サーバとは，Command and Control サーバの略で，攻撃者がマルウェアに対して指令となるコマンドを送信し，マルウェアが仕掛けられたコンピュータの動作を制御するために用いられる．コネクトバック通信は，外部の攻撃者が端末へ侵入する時の通信方式として，端末側が接続元となって通信を発し，それに応答する形で攻撃者が端末に接続し，侵入する際の通信のことをさす．これは主に攻撃者がファイアウォールをすり抜けるために用いられる．

たとえば，あるトロイの木馬型のマルウェアの場合，侵入し，C&C を受け付けるための通信ポート（バックドア）の開設を行い，次に，侵入したことを悪意のあるマルウェア作者あるいはオペレータに伝えるための通信を開始するという動作となる（図 1）．

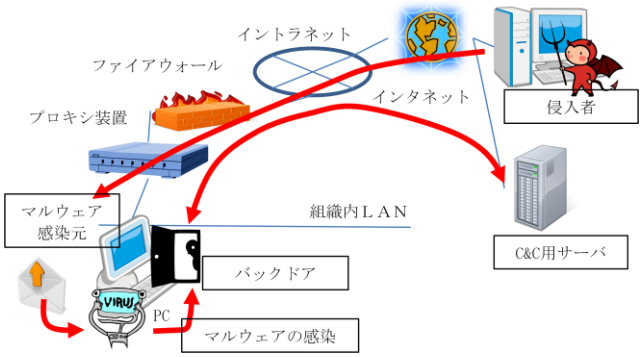

図 1 マルウェアの侵入とバックドア，C&C
Fig. 1 An intrusion of malware, backdoor and C&C.

一般的にマルウェアは，インタネット上の特定のサイトにあるプログラムを通してオペレータとの通信路を確保し，オペレータは各種コマンドを操作することで，侵入した装置の制御を遠隔から実施する．結果として装置上にある企業機密などの情報を盗み出し，破壊活動，また侵入口を経由した環境の探査などを行い，自らの目的を達成しようとする．

近年，入口対策に加えて出口対策が注目されている．出

口対策とは，マルウェアが重要情報を外部に送信することを防ぐことである．企業ネットワークとインタネットの境界で，外向きの通信をチェックし，マルウェアによる情報送信を検出した場合には遮断する．これにより，PC（パソコン）がマルウェアに感染した場合でも，重要情報が流出しないようにする．

④基盤構築段階は，最初の出口対策段階であり，この段階での検知，発見が望まれる．本段階以降は，システム内部に開通したコネクトバック通信を用いて行う「外部攻撃者によるシステム内部への侵入行為」となる．攻撃者は，各種ツールを用い，試行錯誤しながら手動で攻撃を行う．この段階での攻撃目的は，侵入済みの PC（もしくは通信装置）を起点にして，ネットワーク及びサーバの位置情報等を収集することにある．④基盤構築段階での攻撃プロセスを次に示す．

- 攻撃者は目標組織への侵入成否を管理し，継続的に侵入可能な目標の管理（戦果管理）を行う
- 侵入端末から ID とパスワードハッシュを窃取し，システム内侵入拡大の準備を行う

この段階以降の攻撃プロセスは，マルウェア感染問題ではなく，攻撃者の手動操作によるハッキングプロセスとなり，⑤内部侵入・調査段階に移行する．

コネクトバック通信は，システムのネットワーク設計ルールに従った通信プロトコルを用いる正常通信（プロキシ装置を経由した HTTP，HTTPS，CONNECT）として行われるため，他の一般的な Web アクセスによる通信に混ざりこんだ場合，その検知が困難な場合が多い．

2.2 既存の出口対策

既存の出口対策は，ネットワーク設計時にアクセス区画を整理することで，重要な情報をユーザ端末からアクセスできないようにしたり，マルウェアが攻撃元に通信する特殊な通信をファイアウォールで遮断したりする対策が主である[3][4]．

一般に，会社などの組織で運用されるファイアウォールでは,外部へのアクセス手順として HTTP プロトコル以外の通信を止めていることがあるため，マルウェアの出口は HTTP プロトコルが用いられることが多い．攻撃が成功した場合の攻撃者とマルウェアの通信は，業務で使う通信経路を使うため，ウイルス対策ソフトは，通常の通信と悪意のある通信が区別できない．

ネットワーク設計ルールに従った通信プロトコルを用いる正常通信を検知する方法は検討段階にあり,「ブラウザ通信を模倣する HTTP 通信検知」に対する手法が求められている．

ブラウザ通信を模倣する HTTP 通信検知（プロキシ装置を介して行う HTTP プロトコルに関する通信)に対しては次の 4 個がアイデアレベルとして提示されているが，継続検討となっている．

- プロキシ装置の認証機能（ID，パスワードによる）を用いる[3]
- ユーザが端末を使用しない時間帯に外向けの通信をファイアウォールで強制的に遮断し，その後の通信を観察する[3]
- プロキシ装置で Java や Script や META タグを利用しリダイレクトさせる[4]
- 通信ヘッダーのシンプルさからマルウェアを検出する[4]

従って，現時点ではブラウザ通信を模倣する HTTP 通信を検知する方法は検討段階にあると言える．

3. 提案する対策体制・手順とツール

従来のウイルスが予めプログラムされている動作を実施するのに対して，標的型攻撃のマルウェアは，シグネチャをすり抜ける為の個別のカスタマイズや C&C といった人の操作や判断が加わるため，対策を難しくしている．マルウェアの侵入がなされた後で，可能な限り早期にその検出と対策をすべきなのは言うまでもなく，このような，プログラムに加えて,人間の高度な判断が加わる攻撃に対して,防御側も従来のプログラムによる対策に加えて，人間の高度な判断を加えようというのが本稿のログ解析によるマルウェア侵入検知手法の提案である．

3.1 マルウェア検知の体制と手順

企業などの組織におけるネットワーク管理者は，早期に標的型攻撃のマルウェアを検知，発見を試み，そして，マルウェアの通信を阻止し，攻撃者の情報の窃取・破壊に至る過程を防止したいと考えている．そのため，組織の情報システム部門は，大なり小なり，侵入検知の仕組みや機器を導入しているのが当たり前であり，そのために多くのリソースを割いている組織は少なくない．

ネットワーク管理者は，各種ウイルス対策ソフトウェア,ファイアウォールなどを駆使し，通信を阻止したり，検知したりしようとする．

ところが，多くの標的型マルウェアの場合，各種対策ソフトは，たとえ情報漏えいにつながる C&C 経路を確立されたとしても，検知ができず通常の通信として扱ってしまうため，発見が遅れがちである．一方で，検知の仕組みは，ツール以外にもたとえばファイアウォールのログを目視確認などの方法でその兆候を暴き出す，いわば「ノウハウ」を使うものもある．不審なホストに対する通信を早く見つけ出す方法があれば，検出すべき URL，すなわち，ブロックすべき URL を早期に発見できる可能性が高くなる．

3.2 組織におけるネットワークの構成

組織におけるネットワークは，侵入検知システム（IDS: Intrusion Detection System）などのファイアウォール搭載機能，専用セキュリティ装置，そしてウイルス対策ソフトなどに搭載される各種施策によって監視され，保護されている．また，多くの組織では，インタネットの出口にあたる上位組織での管理とその下部，会社組織でいう支店などの単位で階層的管理がなされているのが現状である．

外部に対する通信トラヒックがさほどない時代，また組織内での通信がさほどなく，またインタネット対一組織の通信が主であった時代に比べ，現在は，組織内での Web 利用などインタネット技術を使ったアプリケーションや通信手段が一般的に利用されている．

ここで，脅威が外部組織から持ち込まれることを想定すると，従来は入口管理をしておけばそれで足りたが，多くの組織は同時並行的にインタネットや外部組織とのやり取りを複数の方法,つまり,組織内ネットワークのみならず，媒体や専用の通信手段を使って実施することを考えれば，それぞれの組織単位で多重の防御手段を設置し，人材を配置し，監視を強化せねばならない．

一方，マルウェアの通信経路を考えると，組織外と通信を行わねばならないため，一般的にプロキシ装置と呼ばれる組織内ネットワークおよびイントラネットとインタネットの接点に配置されるネットワーク機器を通過することになる（図 2）．

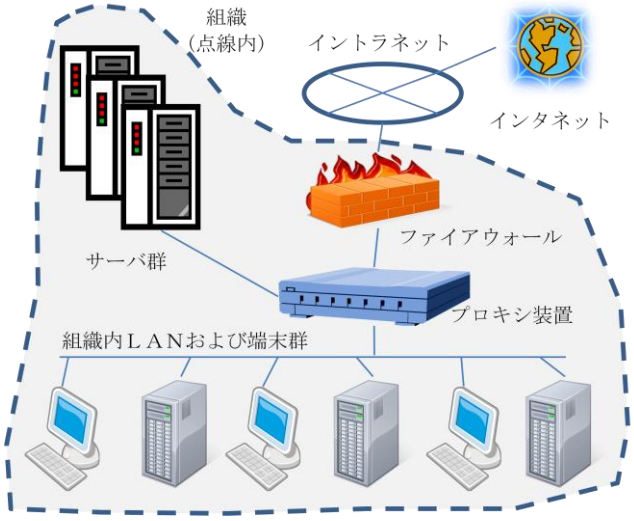

図 2 インタネットと組織の関連，構成図

Fig. 2 Internet, intranet and the system configuration.

3.3 組織内でのマルウェアによる通信

組織からインタネットへの通信は，内部組織を直接インタネットにさらすより，外部へのアクセスを局所化することで，リスクを軽減し，トラヒック制御を実施するためにこのような構成となる．

すなわち，ネットワーク管理者は，プロキシ装置のログを監視すること（図 3）で，マルウェアによる通信を含め，日々のネットワークの状態を把握することができる．

ここでは，ネットワーク管理者が「怪しい」と感じ，マルウェア感染の疑いを把握する経緯と方法についてまとめ，手順化することで検出にかかる時間の短縮を試みる．マルウェアは前述のとおり，種々様々な形態を持つため，ここではまず，検出を試みる条件を明確にしておく．

- すでにマルウェアが侵入し，何らかの活動として組織外部と通信を開始している
- 組織外部の特定のアクセス先に対して複数回の通信を行っている
- 通信はプロキシ装置を経由し，HTTP プロトコルのうち，GET/PUT/CONNECT のいずれかのメソッドを用いて通信を行う

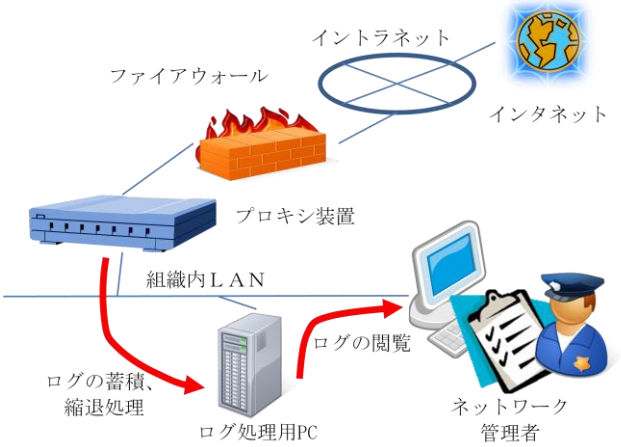

図 3 組織におけるネットワークと管理者

Fig. 3 The network and an administrator.

3.4 プロキシ装置のログ

プロキシ装置のアクセスログを監視すると，時系列で以下のような情報がわかる．

- アクセス者の組織内 LAN の IP アドレス
- アクセス日時
- アクセスメソッド（GET/PUT/CONNECT 等）
- アクセス先 URL（場合によっては IP アドレス）
- アクセス結果（処理ステータス，データ長等）

3.5 不審と判断する理由

プロキシ装置のアクセスログを監視することは，インタネットに対する通信を一挙に把握することができることから，監視を実施するポイントとしては最適である反面，膨大な量のアクセスログが生成されるのは言うまでもない．

1500 人相当の組織の例では 1 日 100 万件，1 か月約 1 億件の通信ログがあり，熟練したネットワーク管理者であっても，人間が処理できる件数を超えている．すなわち，人が不審な通信を検知することが難しい（図 4）．

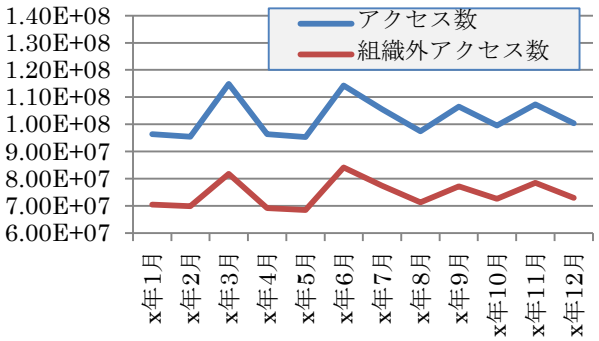

図 4 月別アクセス数，その内組織外アクセス数
Fig. 4 Number of access for Inter/Intranet.

著者らは，まず，ネットワーク管理者が日常的な業務として実施している，「ログ監視」に着目し，管理者がログを監視している中で「不審」と判断する理由をヒアリングした．ここでは，訓練を積んだネットワーク管理者がどのようにして不審通信と判断するかをまとめた．

【アプローチ 1】既知の URL パターン方式

URL の中には，過去何度か「悪事」に利用されたものがあり，パターンマッチングで判断できるものもある．大量のログから URL を抽出し，それをソート， 同一のアクセス先は 1 つにまとめるユニーク化を行い，これと現在なされている通信によるアクセス先を比較することで，不審なアクセスがないか調査する．発見された場合は，アクセスの元となった組織内利用者にヒアリングを実施，PC の状態を調べて，たとえば不審なプロセスが動いていないか，検査をする．

【アプローチ 2】カウント方式

マルウェアの中には，侵入し C&C を成立し，外部との通信を定期的に特定 URL へアクセスし感染者側の情報を送信し，またその結果として侵入し悪事を行おうとする犯罪者の指令を受け取ることを連続して繰り返すものもある．授受するデータは URL にパラメータとして記載する方法，SSL 通信を用いる方法などいろいろある．このようなマルウェアに対しては，時間軸に対してリニアに増加するアクセス頻度の高い「アクセスログ」を整理し集積すると，数値の大きいものから順に「不審」ランキングができる．同様のマルウェアが侵入した場合は，その特徴通りの通信を実施するため，検出ができる．

このランキングの中には，動画アクセスなどマルウェアと同様に連続し同じ URL にアクセスするものがある．そこで，ランキングに載った URL を１つずつ，どのようなサイトか Whois などを使い，確認してゆく．

動画配信サイト，宣伝サイトなどは除外し， 最終的に，見慣れない URL へと絞り込んでいくことができる．

日々，連続して特定 URL にアクセスする PC はさほどないなどの経験値から，絞り込みを進め，感染している PC の利用者にそのような URL に連続的にアクセスしていることを知っているか等ヒアリングを実施し，「知らない」のであれば，不審として PC の動作を調査する．PC に搭載された OS の，スタートアップに不審なプログラム等の登録有無を調査し，最終的な判断を下す．

【アプローチ３】差分方式

ネットワーク技術者の中には，ログからマルウェアから発生される通信を見抜き，通信先のアドレス（URL）を探し出すことがある．日ごろからログを眺めていると，いつもと違う URL に頻繁にアクセスしているなどの兆候を見つけることができる．1 日前，数日前と比較し，知らない URL のアクセスが連続した場合，その URL を調べるため Whois やサイト評価サービスなどを用いて「怪しさ」のレベルを把握する．その結果，たとえば URL が所属するサーバの管理者情報の記載が無い等である．その他，見慣れない URL を，1 週間，1 か月程度のログの出現日，およびアクセスのあった PC（IP アドレス）について調べ，その特異性に着目する．何台もの PC からアクセスがあれば，新しいサービスなのかもしれないため，ログをさらに調査し，どのような経緯からその URL へアクセスが開始されたか，想定する．

例：事前に動画配信サイトにアクセスしている

例：同じパターンを繰り返す宣伝系サイトの集まり

これらを実施してもなお不審な場合は，アクセスした PC の利用者にヒアリングを実施し，意図せぬアクセスの場合は，PC の状態を調査する．

3.6 方式の選択

3.5 のヒアリングの結果，ネットワーク技術者が，不審と判断した根拠には，突発的かつ通信量が突出した通信，継続な通信（特に夜間）などが挙げられた．突発的とはすなわちウイルスやマルウェアの侵入を示す兆候，継続な通信を攻撃者とマルウェアの通信経路と考えれば，「当たらずと雖も遠からず」の判断が可能と想定できるはずであり，著者らは，この「人間の能力」を前提とし，どのような情報の提供が侵入検知に有効であるかをまとめることとした．なお，3.5 の 3 つのアプローチは，いずれもログだけの結果から不審な URL を見つけ出すため，誰でもある程度訓練を積めば，傾向が把握できるようになると判断している．

3.7 ログの縮退ツール

マルウェア侵入を検知するために，ここでは，ブラウザ通信を模倣する HTTP 通信を検知しやすくする方法として，プロキシ装置のログの通信先情報から，監視を容易化するためにログを縮退するツールについて述べる．

プロキシ装置のログは量が多いことから，人が直接そのログから不審な通信を検知することが難しい．また，世界中のサイトの数は極めて多く，更に，生まれるサイト，消えるサイトが刻々と変わるため，ホワイトリストを作る場

合でも，ブラックリストを作る場合でも，かなりな作業ボリュームが予想される．

そこで，著者らは，特定の企業や組織では，1 日の通信量のパターンや通信先が特定されるという仮定の下に，ホワイトリストとして，同一組織体の過去のログを用いることとした．

【統計的検証】

ある組織のログ 1 年分を，1 か月ごとに差分を取り，新しく現れたサイト，消え去ったサイト，変化がないサイトを調査した（図 5）．

ログに記載された各アクセス記録から，URL を抽出したが，各 URL は GET http://www.abcd1234zzz.com/...”のように記載されており，ホストを示す http://から後の文字列から次の/までと，それ以降は，http の各メソッドに従ったパラメータや，ディレクトリ階層が続く．

その結果，新しく現れたサイトは月あたり 60,000～75,000 サイト，消滅したサイトも同様な数であり，6 月に若干の変化を認めるが，ログ全体の 1 億件/月のオーダと比較し，有意な変化はなかった(6 月以降の増分は組織が関連する所外新サービスの開始による)．

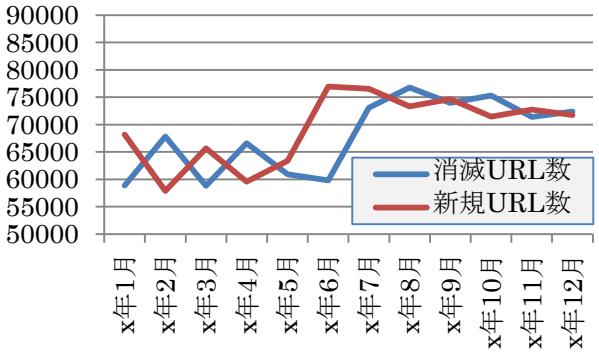

図 5 アクセス先 URL の増減

Fig. 5 Number of URLs.

【ホワイトおよびグレイリストの作成】

次に，過去のログ情報からホワイトリストを作成し，その差分を取ることで，チェックするログ（グレイリスト）の件数を縮退させ，人が不審な通信を検知することができるようにする．

ホワイトリストを大きくすれば，グレイリストは小さくはずである．ただし，ホワイトリストが大きい場合，処理時間が多くなる問題を含む．そこで，実験 1 としてホワイトリストとしてのログを，調査期間と同一の 1 ヶ月としてみた．また，次の仮定を置いた．

【仮定】

ある期間 t にアクセスした URL のログ Ut は，安全な通信先とする．

マルウェアの通信が発生している期間 v にアクセスした URL のログ Uv には，当然，マルウェアに関係する URL が含まれている． 期間 t と期間 v は，時間軸上重なることがないこととする．また，期間 t と期間 v の間には 1 か月の緩衝期間を設けることとする（図 6）．

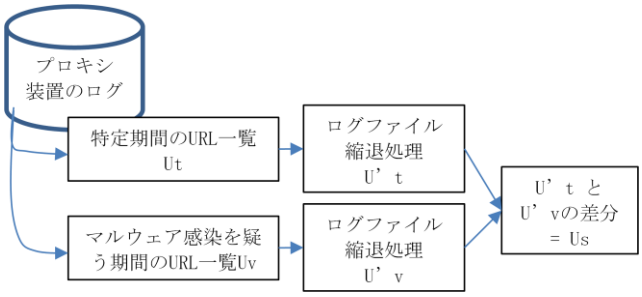

図 6 ログの差分の取得

Fig. 6 System diagram of the log processing.

【実験 1】

期間 t および期間 v を 1 か月とし Ut および Uv を作成，その差分 Us を得てみた．その結果，Ut は約 1 億件，Uv も約 1 億件（マルウェア通信を埋め込んだログ）となり，差分 Us の結果は 100 万件となった．

【実験 1 の課題】

100 万件を有限時間で調査するには時間がかかり過ぎる．

本手法では，マルウェアが発生しているかもしれない日を含む 1 か月間と，過去 1 億件の通信ログとの差分を取ることから開始したが，それでも約 100 万件までしか縮退できないことが分かった．

【実験 1 を踏まえた対策】

Ut および Uv それぞれをプロキシ装置から得られたログそのものではなく，事前に加工し，検索やマッチングの時間を短縮させるため，ログに記録された通信先情報の部分文字列を作ることによって，データの一部破棄を行った．

URL すべてを比較の対象とするのは可能であるが，ログを処理するために膨大な計算量が必要なことは容易に想像できる．事前準備として，1 年分のログから任意に選んだ 1 か月分のログから，URL 中に含まれる/の数を数えてみた（図 7）．

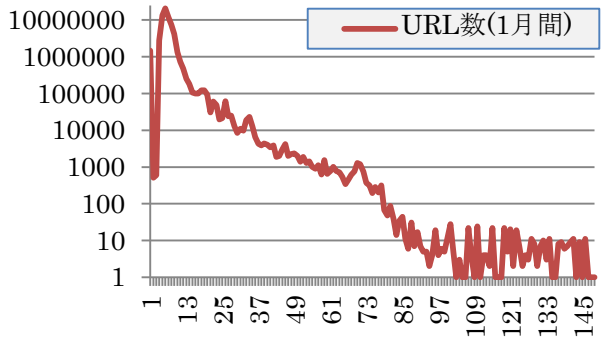

図 7 ある月のログにある URL 中の/の数

Fig. 7 Number of “/” in URL.

その結果，主に CONNECT メソッドで使われるサイト部のみの記録が一番多く，/の数がおおよそ 6 つまでのケースが多数であることが判った．実験 2 として用いるアクセスログの URL は，部分文字列としてサイト＋1 つ目の/までを切り出し使うこととした．この判断根拠は，不審な URL を検出するために与えられた時間はさほど多くはなく，通常の市販 PC においておおよそ 1 時間程度で結果が得られることを条件したためである．データを一部破棄する根拠は，URL そのもの全体を比較しても，URL にはサイトに送信するデータ部分などが含まれており，検索や差分に計算時間を要する割には差分を取る意味がないと判断した事による．

【実験 2】

Ut を上記実験 1 の対策で示す手順で加工し U't とした．次に，各 URL をユニークとなるように URL の長さを http://www.abcd1234zzz.com/xyz/のように限定する．これに加え，HTTP のメソッド（GET，PUT，CONNECT）を限定する．さらに U't から，元のアクセス記録中，1 度しか出現しない URL を削除した．これはマルウェアが C&C サーバに対し連続通信を行うことを想定とした絞り込みである．

次に，Uv を U't と同様に加工し U'v とした．U'v で用いる URL の長さも，上述 U't と同様に限定した．

その結果，U't と U'v の差分 U's は約 20 万件となり，差分にはマルウェアから発せられた通信先 URL が含まれた．

ログ全体からみれば，20 万件は十分小さな数値であるが，熟練したネットワーク管理者でも有限時間内において検査を終わらせることは難しい．そこで，ホワイトリストをさらに絞り込みを行うため，引き続き，ホワイトリストの採択期間を変えて結果を見た．

4. 定量的評価

ホワイトリストを作成するための期間 t を v の期間の開始前 10, 5, 3, 1 か月として，複数の新たな Ut 群を作成した．Uv は，実験 2 で用いたものと同一のデータを用いた．その差分結果であるグレイリスト Us の中の URL 数を図 8 にまとめる．

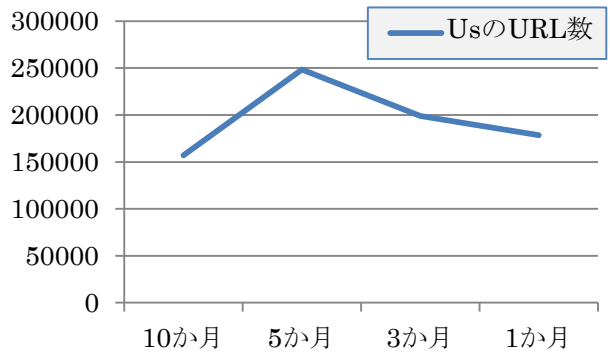

図 8 期間 t の変化と Us の URL 数

Fig.8 Relation between t and number of URLs in Us.

実験の結果，Ut の採取期間からは，有意な変化は計測できず，Us すなわちグレイリストは平均約 18 万件前後となった．

この段階で，Us を目視確認したところ，多くが大手検索サイト，イントラネットで接続された関連企業に関する URL であり，URL の末尾部分だけが変化した状態のものが散見された．

【実験 3】

何人かのネットワーク管理者とグレイリストを眺め，議論を行った結果，Uv に対してさらなるフィルタリングができる可能性が確認できた．以下に該当する項目は，経験的かつ過去にウイルスやマルウェアとの関連が薄いサイトであり，これらを削除することとし，実験 4 を行った．ここでの判断基準は，過去にセキュリティ事故につながる関係にない，組織内部の通信と同様に安全見なせることである．

【実験 4】

実験 3 での Us から，

- 大手検索サイト
- ウイルス対策会社
- 動画配信サイト

などのサイトの URL をフィルタリングし，新たな U't を作成，マルウェアの通信の記録を含む U'v との差分 U's を取った（表 2）．また，図 9 には本稿の範囲で実験した処理の流れ全体を示した．

その結果，8000 件程度が抽出され，その中にはいずれの期間の t においてもマルウェアが発生させた通信による URL が含まれていた．この程度であれば，ネットワーク管理者がグレイリストを眺めて，不審な URL を見つけ易い．

表 2 URL のフィルタリング

Table 2 Filtering result of URL.

t の期間	フィルタリング前	フィルタリング結果	マルウェア URL 有無
10 か月	157000 件	約 7000 件	検出
5 か月	248000 件	約 9000 件	検出
3 か月	199000 件	約 8000 件	検出
1 か月	178000 件	約 8000 件	検出

これらの手法を連続的に行い，また Ut を定常的作業の一環として更新を続けていれば：

- 過去の通信ログだけから，マルウェアが発生する特定 URL への通信を含む集合を抽出し，ネットワーク管理者の調査を効率化することができる
- 信頼する URL の標本も，マルウェアの通信による特定 URL への通信も，1 つのログから導き出しており，コストをかけずに調査の効率化を実現することができる

また，

・ URL の長さや，特徴を抽出するための方法は，今回の実験以外にも採用する「部分」により，いくつかの方法が想定され，ノウハウとして活用できそうである
・ プロキシ装置のログだけの結果から不審な URL を見つけ出す方法のため，誰でもある程度訓練を積めば，傾向が把握できる

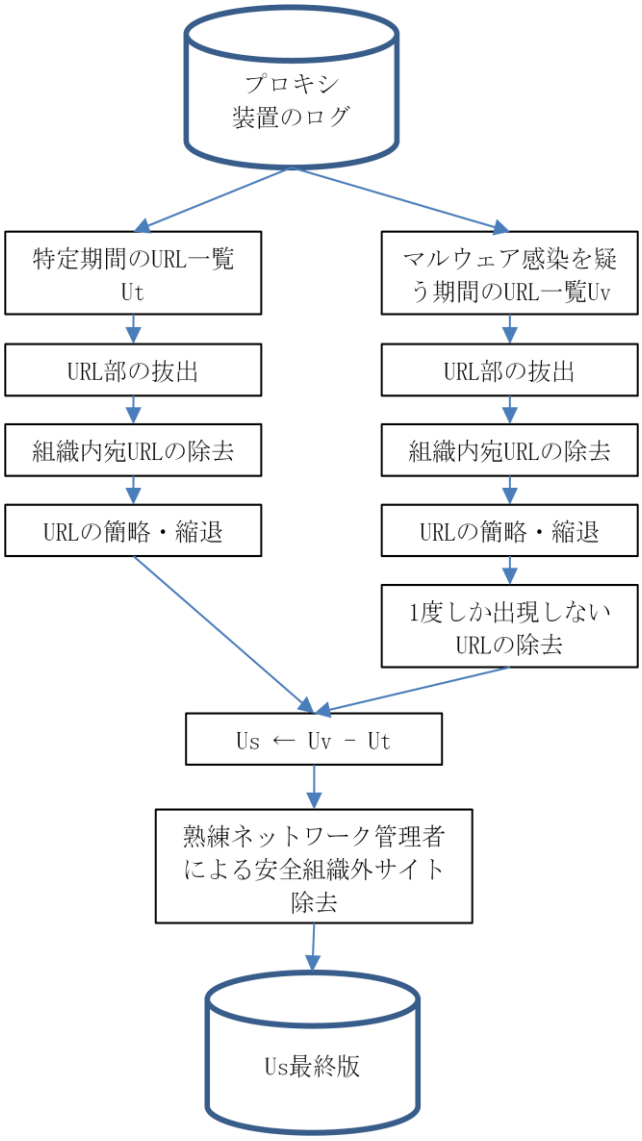

図 9 ログ縮退処理全容

Fig. 9 A log processing concept.

5. おわりに

本手法開発における実験では，マルウェアが発生しているかもしれない日を含む 1 か月間のログと，過去 1 億件の通信ログとの差分を取ることから開始したが，約 100 万件までしか縮退できないことが判ったため，ログに記録された通信先情報の部分文字列を作ることによって，データの一部破棄を行い，最終的には，マルウェアによる通信先を含むと予想される約 8000 件程度のグレイリストを得ることができた．100 行/ページとして 80 ページのログであれば，ネットワーク管理者が有限時間において，その「不審さ」を判別することが可能であり，膨大なログファイルを検査するのと比較し，効率的に作業を行える環境として使える．

今回は，過去，検出されたマルウェアが発生させた通信が判っているログを用いての検証となったが，今後，研究用データセット MWS Dataset 2013 の中より，多くのマルウェア通信データを実ログの中に埋め込み,「マルウェアの通信が縮退されずに残ること」を評価する予定である．

謝辞 本稿の執筆および方向性に関し議論に参加およびアドバイスを頂いた皆様に，謹んで感謝の意を表する．

参考文献

1) IPA 独立行政法人情報処理推進機構セキュリティーセンター, 標的型サーバー攻撃の脅威と対策, http://www.ipa.go.jp/security/event/2013/isec-semi/documents/2013videosemi_targeted_cyber_attacks_v1.pdf, P18(2013).
2) 米大使館政策情報 http://japanese.japan.usembassy.gov/j/p/tpj-20110517a.html サイバースペースのための国際戦略(2011.5.17).
3) IPA 独立行政法人情報処理推進機構セキュリティーセンター,「標的型メール攻撃」対策に向けたシステム設計ガイド, http://www.ipa.go.jp/security/vuln/newattack.html, p63(2013.8.29).
4) IPA 独立行政法人情報処理推進機構セキュリティーセンター,「新しいタイプの攻撃」の対策に向けた設計・運用ガイド 改訂第 2 版, http://www.ipa.go.jp/security/vuln/newattack.html, p58(2011.11).
5) マルウェア対策のための研究用データセット, 神薗 雅紀, 畑田 充弘, 寺田 真敏, 秋山 満昭, 笠間 貴弘, 村上 純一, (社) 情報処理学会 コンピュータセキュリティ研究会 MWS 組織委員会(2013.5.18).
6) 検索エンジンによるマルウェア接続先評価手法の提案 (コンピュータセキュリティ (CSEC) Vol.2010-CSEC-50), 青木 一史, 秋山 満昭, 岩村 誠 他, 情報処理学会研究報告 2010 年度 (2) 2010-08p.6p(2010.6.24).
7) URL ブラックリストの効率的な利用方法の一検討 (インターネットアーキテクチャ), 松木 隆宏, 新井 悠, 電子情報通信学会技術研究報告 : 信学技報 109 (85) 2009-06p.19p～23p(2009.6.11).
8) マルウェアの耐解析機能を逆用した活動抑止手法の提案 (特集 社会を活性化するコンピュータセキュリティ技術), 松木 隆宏, 新井 悠, 寺田 真敏 他, 情報処理学会論文誌 論文誌ジャーナル 50 (9) 2009-09p.2118-2126(2009.9.15).
9) 通信の共通性を利用した悪性プログラム検知手法の実装と評価(特集 社会を活性化するコンピュータセキュリティ技術), 水谷 正慶, 金井 瑛, 武田 圭史 他, 情報処理学会論文誌 論文誌ジャーナル 50 (9) 2009-09p.2137-2146(2009.9.15).
10) マッシュアップによる Web マルウェアの実態調査 (特集 人と共存するコンピュータセキュリティ技術), 松木 隆宏, 新井 悠, 寺田 真敏 他, 情報処理学会論文誌 論文誌ジャーナル 52 (9) 2011-09p.2748-2760(2011.9.15).
11) 平成 25 年上半期のサイバー攻撃情勢について, 警察庁, www.npa.go.jp/keibi/biki3/250822kouhou.pdf(2013.8.2).