4E-03

Firewall ログを用いたマルウェア感染端末の検知手法

神谷　和憲[†]　青木　一史[†]　中田　健介[†]　佐藤　徹[†]　倉上　弘[†]　谷川　真樹[†]
NTT セキュアプラットフォーム研究所[†]

１．はじめに

近年，DDoS 攻撃，スパムメール，情報漏洩など，マルウェアを起因としたインシデントが多く発生している．一方で，マルウェアの種類は多種多様であり，市販のウイルス検知ソフトでは約半数のマルウェアを検知することができない［1］．マルウェアのシグネチャをベースとした事前対策では防御は十分ではなく，マルウェアの感染を前提とし，感染後の被害拡大を抑える事後対策が重要となっている．

事後対策としては，ネットワークログの観測により，マルウェアの感染活動を検知する SIEM(Security Information and Event Management)が使われ始めている．SIEM の技術として，特に HTTP Proxy サーバ（本稿では以下，Proxy と呼ぶ）のログをもとに不正な URL へのアクセスを検知する手法が知られている［2］．ただしこの手法では HTTP 以外で通信するマルウェアの挙動を捉えられないため，他の手法と組み合わせた多層検知の手法が求められている．

本稿では，Proxy と並んで一般的に使用される Firewall からログを収集し，マルウェア感染端末を検知する手法を提案する．具体的にはマルウェアを動的解析した結果から Firewall 用の悪性リストを作成しマッチングを行なう．

本手法により，Proxy ログの分析だけでは検知できない種類のマルウェア感染端末を検知できることを示した．

２．企業網における Firewall ログの特徴

図 1 に，企業網で一般的である Proxy と Firewall を用いた網構成を示す．

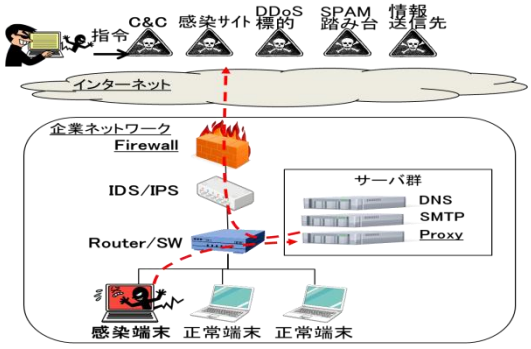

図 1 Firewall と Proxy の設置例

ここで Firewall の OUTBOUND 通信に対するポリシーは次のように設定することが多い．

Firewall OUTBOUND ポリシーの設定例
Proxy サーバを発信元とする通信のみ ACCEPT
それ以外の OUTBOUND 通信は DENY

ログの収集は Proxy，Firewall でそれぞれ可能である．ログは各デバイスから Syslog で送信して収集することが多い．Firewall，Proxy におけるログの収集可否を表 1 に示す．

TCP/IP 通信に関する情報は Firewall ですべて収集可能であり，Proxy では HTTP で通信したものについて TCP/IP レベルのログを収集可能である（表中では△で表記した）．HTTP レイヤに関する情報は Proxy のみで収集可能である．

表 1 Firewall と Proxy におけるログの収集可否

	Firewall	HTTP Proxy
時刻	○	○
アクション	○	×
ip_proto	○	×
src_ip	○	△
dst_ip	○	△
dst_port	○	△
http_url	×	○

以上より，図 1 の企業網においては，Proxy ログに加えて，Firewall の OUTBOUND DENY ログを収集することで，HTTP 以外の通信を把握することができる．これにより，あるマルウェアが HTTP 以外のプロトコルで通信する場合であっても，その挙動を捉えるのに必要なログを収集することができる．

３．感染端末検知手法

本稿では Firewall ログで収集可能なフィールドについて，マルウェア特有の悪性リストを抽出し，トラフィックログとのマッチングにより，感染端末を検知する手法を提案する．

３．１．悪性リスト抽出手法

マルウェア通信に特有の値リスト（悪性リスト）の抽出手順を図 2 に示す．

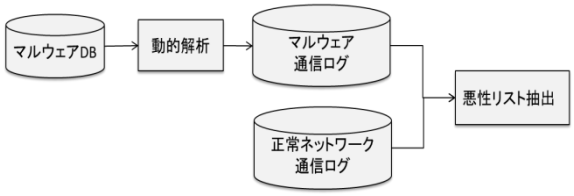

図 2 悪性リスト抽出手法

まず，多種多様なマルウェアを動的解析し，マルウェア通信ログを収集する．動的解析では，実ネットワークに被害を与えないよう，擬似応答をするなどしてログを収集する［3］．

次に，感染端末がないネットワークのログを正常ネットワーク通信ログ（以後，正常ログとも呼ぶ）として収集する．

ここで，マルウェア通信ログに含まれるあるフィールドαに対する値kについて，正常ログにおける発生率$PL_\alpha(k)$および，マルウェア通信ログにおける発生率$PM_\alpha(k)$を以下の通り計算する．

$$PL_\alpha(k) = \frac{n_k}{N} \quad PM_\alpha(k) = \frac{m_k}{M}$$

上記においてn_kは正常ログのフィールドαにおいて値kが発生した送信元ホスト数，Nは全送信元ホスト数，m_kはマルウェア通信ログのフィールドαにおいて値kが発生したマルウェア検体数，Mは全マルウェア検体数である．

より多くのマルウェア感染端末を検知し，誤検知を少なくするため

$$\Phi_\alpha = \{k \mid PM_\alpha(k) > tm, PL_\alpha(k) < tl\}$$

となる集合をフィールドαに対する悪性リストとして取得する．tm，tlは適宜値を設定する．

３．２．マルウェア感染端末検知手法

あるホストにおいて，悪性リストにマッチするログが一定回数以上発生した場合に検知（回数検知），または一定種類数以上発生した場合に検知（種類数検知）を行なう．

すなわち$k \in \Phi_\alpha$に対しw_kをあるホストにおけるkの発生回数，$s_k = \begin{cases} 0, & w_k = 0 \\ 1, & w_k > 0 \end{cases}$とするときに，悪性リストに単一マッチするログの発生回数w，ログにマッチした悪性リストの種類数sは以下の通り計算する．

$$\mathrm{w} = \max_k w_k \quad \mathrm{s} = \sum_k s_k$$

回数検知は閾値twにより$\mathrm{w} > tw$を満たす条件で発火する．種類数検知は閾値tsにより$s > ts$を満たす条件で発火する．

４．提案手法の評価結果

４．１．データセット

マルウェア通信ログは VirusTotal［4］から取得した数万オーダの検体を動的解析した結果を使用した．検体は，VirusTotal における複数の AntiVirus ソフトによるスキャン結果で 1 つ以上のソフトで「悪性」判定され，かつ動的解析環境によって通信を発生させたものを選択した．使用した検体はファイルの SHA1 ハッシュ値はすべて異なり，特定のマルウェア・ファミリーのみを多く含んだものでないことを確認済である．

正常ログは企業網において収集した Firewall ログ(OUTBOUND DENY)と Proxy ログを使用している．正常ログのホスト数のオーダは数千である．

評価にあたっては，マルウェア通信ログ・正常ログともに 7 分割し，トレーニング用セットと評価用セットを作成し，クロスバリデーションを実施した．

４．２．評価指標

評価指標として，検知率(TPR)・誤検知率(FPR)を使用する．それぞれ，下記の通り計算される．

$$\mathrm{TPR} = \frac{m_{detected}}{M} \quad \mathrm{FPR} = \frac{n_{detected}}{N}$$

ここで$m_{detected}$は検知できたマルウェアの検体数，$n_{detected}$は誤検知した送信元ホスト数である．

４．３．評価結果

提案手法の評価結果を図 3 に示す．図 3 左図は Firewall ログと Proxy ログのそれぞれに対するTPRである．図 3 右図は Firewall ログ，Proxy ログの両方を用いた場合に，Firewall ログのみで検知できた検体のTPRを示している．

いずれの場合も誤検知率は$\mathrm{FPR} < 0.005$となるよう閾値tw，閾値tsを調整している．Firewall ログに対する検知では，検知に使用するフィールドαとして，dst_ip(TCP80 に限定)，dst_port(TCP,UDP に限定)を使用し，回数型検知を行った．Proxy ではフィールドαとして http_url を使用し，種類型検知を行った．

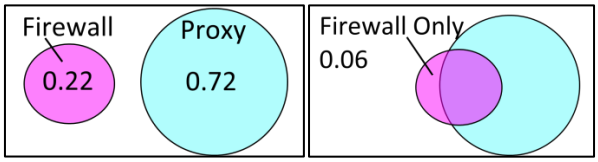

図 3 提案手法評価結果

図 3 左図より，Firewall ログに対する TPRは Proxy ログに対するTPRよりは低い値となっている． 図 3 右図より Firewall ログを使用することにより，Proxy ログからは検知できない検体を6%検知できることが分かる．これらは HTTP を使用しない通信であり，TCP139，TCP445 を使用する通信が多く見られた．中には，TCP80 を使う非 HTTP 通信も見られた．

以上から，Firewall ログを用いた検知は Proxy の悪性リストを補完する用途で有用であり，全体の TPR 向上に貢献することが示された．

参考文献

1. **NTT-COM Security.** グローバル脅威情報レポート. 2014.
2. **NelmsT.** ExecScent: UseNix Security, 2013.
3. **aoki.** Controlling Malware HTTP Communications in Dynamic Analysis System Using Search Engine. : CSS, 2011.
4. VirusTotal. https://www.virustotal.com/.