5E-03

サイバー攻撃被害範囲特定のための サーバログトレース方法に関する提案

松本 光弘† 白木 宏明† 大松 史生†
三菱電機株式会社 情報技術総合研究所†

1 はじめに

近年、サイバー攻撃は巧妙化しており、その被害は後を絶たない。入口対策だけで攻撃を防ぐことには限界があり、攻撃を受けることを前提とした対策が必要である。攻撃を受けた場合は、その被害範囲を最小限に抑えるために、まず攻撃による被害状況を把握する必要がある。

被害状況を把握するためには、サーバの操作記録(ログ)から攻撃に関する情報を抽出する必要がある。特に、サーバのシステムログやアプリケーションログ(以下、アプリログ)から攻撃に関する情報を特定することは、攻撃者がサーバ内でどのような操作を行っていたのかを特定することにつながるため、被害状況を把握するのに有効である。しかし、複数に渡る膨大なログから攻撃に該当する操作内容(イベント)を特定することは困難であり、セキュリティの専門家であっても多くの時間を要する。

そこで本論文では、セキュリティ監視装置であるIDS/IPS, FW, プロキシ等から通知される攻撃検知情報を基に、攻撃に関連したアプリログを特定する方法を提案する。通信ログとアプリログの時刻情報を基に通信とアプリケーションの関連付けを行うことで、上記セキュリティ監視装置で検知した攻撃に関する通信を発生させたアプリケーションを特定することが可能となる。これにより、セキュリティ監視者は攻撃検知情報から関連するアプリログをいち早く特定することができるため、アプリケーションが不正に利用されている・不正なアプリケーションが利用されている等、攻撃内容を迅速に把握することができる。

2 関連研究

ログのトレースを効率良く行うためのデータモデルを提案している研究がある [1]。順序関係や階層関係、クラス間関係から各事象を結びつけることによって、ログトレースに特化したログ DB を構築しており、メールの配信経路や送受信ログの追跡を実現している。性能評価では、従来の関係データモデルと比べて効率良くメール配信経路の取得が可能であることを確認している。

上記研究 [1]では、順序関係や親子関係がログ内に明記されていることを前提としているが、複数の異なるログでは、そのような情報はないため、本論文ではログ間の関係を特定する。

また、ファイルと Web ページの関連性を抽出する研究 [2]があり、左記論文は本論文との関連抽出方法と考え方が類似している。左記論文 [2]では、ファイルアクセス時間と Web 閲覧時間を用いて、同じ時間に利用されていた場合に、関連があるとしている。

本論文も同じ時刻に記録されたログは関連があるとしているが、多くのログに関連が見られるログについては、関連度を下げることで関連付けの精度を上げている。

3 システム構成

図 1 のネットワークシステムにおいて、インターネットからの不審な通信をセキュリティ機器が検知し、セキュリティ警告ログを出力する。セキュリティ管理者は、セキュリティ警告ログを基に不審通信がサーバにどのように影響を与えたのかを調査する必要があり、そのためには不審通信に関する情報をサーバ内に記録された通信ログとアプリログから特定する必要がある。

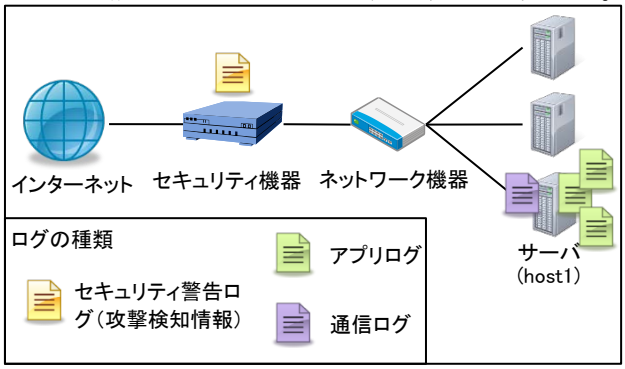


図 1 ネットワーク構成

セキュリティ機器が出力するセキュリティ警告ログとサーバが出力する通信ログは、共に通信に関する情報を出力するため、ログの内容に類似点が多く、関連付けを行うこと(図 2 の①)は容易である。一方、アプリログには、アプリケーション内でのユーザの操作やアプリケーションの動作等に関するログが記録されているため、通信ログとアプリログの関連付けを行うこと(図 2 の②)は困難である。

そこで、本論文では通信ログとアプリログの関連付けを行う方法を提案する。

A method to combine server logs for confirming the detail of the cyber-attack damage
Mitsuhiro Matsumoto†, Hiroaki Shiraki†, Fumio Ohmatsu†
Information Technology R&D Center, Mitsubishi Electric Corporation. †

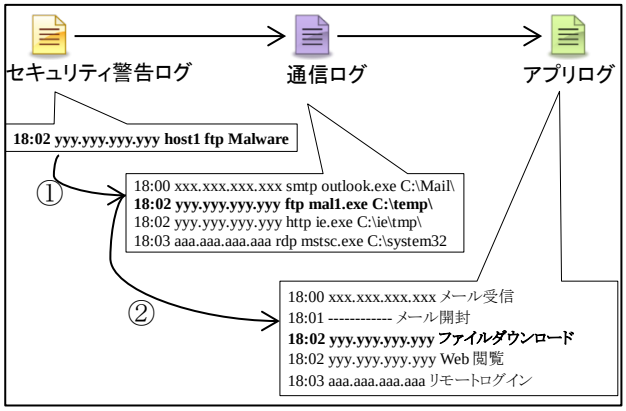


図 2 通信の詳細把握までのログトレース

4 提案方法

通信ログから関連するアプリログを特定するために、TF-IDF を基にした関連度導出方法を提案する。TF-IDF は文章中の重要語を抽出する手法であり、情報検索や文章要約等の分野で利用されている。

通信ログとアプリログにおいても、TF-IDF の考え方が利用できると考える。ユーザがアプリケーションを操作することによって通信が発生した場合、アプリログにはユーザ操作に関するログが、通信ログには通信に関するログが同時刻に記録される。そのため、同時刻にアプリログと通信ログが記録されている場合は、関連するログである可能性が高く TF 値が高くなる。一方で、アプリログには通信に関係しないログも含まれる。そのため、通信に無関係のログが、偶然他のアプリの通信ログと同時刻に記録された場合、通信ログが無関係なアプリと関連付けられる可能性がある。特に、通信に無関係なログを多く出力するアプリは、他の多くの通信と同時刻にログが記録されてしまう可能性が高いため、このようなアプリログは無関係な多くの通信ログと関連付けられてしまう。他の通信ログと同時刻に記録されているアプリログの IDF 値は小さくなるため、TF-IDF の値は小さくなり、無関係な通信ログとアプリログが関連付けられる可能性を低くすることができる。

図 3 より TF-IDF の計算式を示す。各通信ログに対して最も大きな TF-IDF 値を示すアプリログが関連するアプリログとなる。

$$tf_{i,j} = \frac{n_{i,j}}{n_j}$$

$n_{i,j}$：アプリイベントt_iの通信イベントd_jが同時刻に発生する回数

n_j：通信イベントd_jの発生回数

$$idf_i = \log\left(\frac{N}{df_i}\right) + 1$$

N：通信ログの種類数

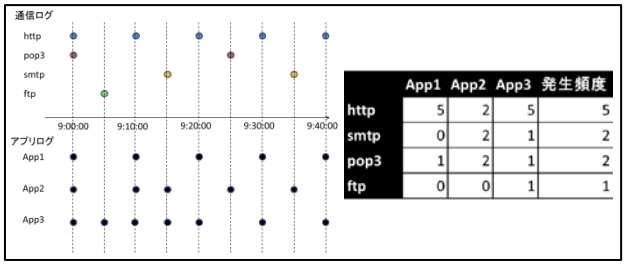
df_i：アプリイベントt_iと同時刻に発生する通信イベントの種類数

$$tfidf = tf \times idf$$

図 3 TF-IDF 値の算出式

5 処理事例

例えば、図 4 のようなログが記録されたとする。図 4(左)は、通信ログとアプリログの時間情報を基に時系列にデータをプロットした図であり、通信ログには 9:00:00 に http と pop3 のログが記録され、アプリログには同時刻に App1,App2,App3 のログが記録されたことを示す。図 4(右)は、図 4(左)から同時刻に記録されたログの回数をカウントした表で、http と App1 のログが同時刻に記録された回数は 5 回である。また、発生頻度の項目から http は通信ログに 5 回記録されていることが分かる。



	App1	App2	App3	発生頻度
http	5	2	5	5
smtp	0	2	1	2
pop3	1	2	1	2
ftp	0	0	1	1

図 4 ログ記録時刻の時系列マップ（左）とログの同時刻発生頻度（右）

図 4(右)から、各通信ログとアプリログの TF-IDF 値を導出し（図 5(左)）、図 5(左)から通信とアプリの関連付けを行う（図 5(右)）。http と App1 の TF 値は 5/5=1 であり、IDF 値は log(4/2)=2(底は 2)となるため、TF-IDF 値は 2 となる。http において、もっとも TF-IDF 値が高いのが App1 となるため、http と App1 が関連付けられる。すべての通信に対して関連するアプリを表したのが図 5(右)である。

	App1	App2	App3
http	2	0.57	1
smtp	0	1.4	0.5
pop3	1	1.4	0.5
ftp	0	0	1

通信サービス	関連アプリ
http	App1
smtp	App2
pop3	App2
ftp	App3

図 5 TF-IDF 値の表（左）と関連ログの対応表（右）

6 まとめ

本論文では、セキュリティ監視装置から攻撃検知情報が通知された際に、迅速に関連するサーバのアプリケーションを特定するために、通信ログとアプリログを関連付ける手法を提案した。

本手法は、時刻情報を基に通信ログとアプリログを自動で関連付けることができるため、専門知識を要することなく、攻撃に関連したアプリログを迅速に特定することができる。

今後は、実データを用いて本手法の関連付けの精度や処理速度等を検証し、有用性を示していく。

参考文献

1. **宋強, 渡辺陽介 , 横田治夫.** ファイルと Web ページの共起頻度に着目した関連性抽出手法の評価. DEIM Forum, 2011.

2. **平井規郎, 森山令子 , 郡光則.** 履歴追跡型データモデルの評価. 日本データベース学会論文誌, Vol.7, No.3, 2008.