サンプリングによるNetFlowの生成とデータの観測について

榊原裕之† 北澤繁樹† 河内清人† 藤井誠司†
†三菱電機株式会社 情報技術総合研究所

1. はじめに

筆者らは未知のワームなどの不正アクセスを早期に検知するため，ネットワークの正常な状態からの変動を主成分分析により発見する Anomaly 型の不正アクセス分析システム「セキュリティー攻撃予兆分析システム」の開発に取り組んでいる [1].

近年，企業において情報漏えいが問題となっているが，原因の 1 つとしてマルウエアが考えられるため，開発システムをマルウエアによる情報漏えいの検知へ適用することを検討している．実例に基づき，イントラネットの PC に感染したマルウエアが PC 内部のデータをインターネット上の攻撃者の Web サーバへアップロードする攻撃データを想定し，この攻撃データに起因するネットワークデータの変動を検知する方針とした．

一方，想定する監視対象のネットワークの流量と開発システムの処理性能を考慮しサンプリングベースの NetFlow v5[2]のデータを分析することとした．しかし，サンプリングを行うことにより想定する攻撃データが NetFlow データとして観測されないことが懸念された．

そこで，開発システムを情報漏えいの検知に適用する準備として，実験環境で，サンプリングベースの NetFlow により攻撃データが観測されるか確認の実験を行ったので，結果と考察を述べる．

2. マルウエアによる情報漏えいの攻撃データ

開発システムをマルウエアによる情報漏えい検知に適用するにあたり，検知する攻撃データを想定する必要がある．本章では，マルウエアによる情報漏えいの実例に基づき想定した攻撃データについて説明する．

2.1. 攻撃データの想定に参照したマルウエア

情報漏えいを起こしたマルウエアの実例として Infostealer.Monsters がある．当マルウエアは PC に感染すると雇用サービス企業の Web サイトに登録された個人データを不正にダウンロードし，その後インターネット上の攻撃者の Web サーバにアップロードする[3]．多くの組織で許可される http を漏洩の手段として用いており今後も脅威となりうると考えたため，当アップロード行為を情報漏えいのモデルとして参照した．

Observation Of Attack Packets On Sampling-based NetFlow
Hiroyuki Sakakibara†, Shigeki Kitazawa†,Kiyoto Kawauchi†, Seiji Fujii†
† Information Technology R&D Center, Mitsubishi Electric Corporation

2.2. 想定する攻撃データ

セキュリティ関連の Web サイトで報告されている Infostealer.Monsters の情報に基づき，以下の様に，検知対象とする攻撃データを想定した．

まず，雇用サービス企業の Web サイトにおける個人データ登録ページの登録項目から，個人データ 1 件あたりを約 500 バイトと見積もった．次に，100 万件以上のデータを漏洩させたと推定されているため，100 万件×500 バイトのデータを PC が起動しているビジネスアワーおよそ 8 時間に漏洩させたと仮定した．また，データのアップロードは 1 件ずつ行うと非効率であるため，まとまった件数をファイル化して行ったと仮定した．

これらから，5M バイトのファイル(1 万件相当)を 5 分に 1 回の頻度でビジネスアワーの期間に http POST でアップロードするデータを検知対象の攻撃データとして想定した．開発システムでは，当攻撃データに起因するネットワークデータの正常状態からの変動を分析する．

3. サンプリングベースの NetFlow

開発システムでは，ネットワークデータとして NetFlow データを分析対象としており，監視対象とするネットワークの流量と開発システムのデータ処理性能を鑑み，サンプリングベースの NetFlow データを分析する．

3.1. NetFlow v5

NetFlow v5 は宛先 IP，発信元 IP，宛先ポート，発信元ポート，プロトコルタイプが同じパケットを集約するトラフィックデータの観測方式である．1 つの NetFlow データには観測されたパケット数，転送サイズが含まれるため，どの端末同士が何の通信をどれくらいの量行ったか監視することが可能である．

3.2. サンプリングベースの NetFlow

ネットワークの流量によっては NetFlow データの数が大きくなり分析処理性能に合わないことがある．解決策としてパケットをサンプリングし，サンプリングされたパケットから NetFlow データを生成する方法がある．図 1は，パケット 5 個につき 1 つをサンプリングし，得られたパケットから NetFlow データを生成する例である．

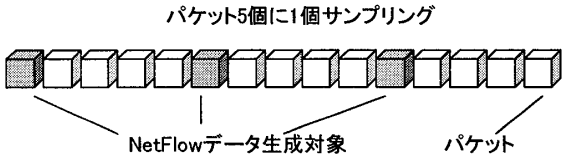

図 1 サンプリングベースの NetFlow

4. サンプリングベースの NetFlow による攻撃データの観測実験の必要性

サンプリングベースの NetFlow データの生成において，攻撃データのパケットがサンプリングされない場合は NetFlow データとして観測されないため分析の対象とならず検知漏れとなる．従って，開発システムの検知の適用前に，想定する攻撃データが，監視対象のネットワークにおいてサンプリングベースの NetFlow データで観測され，分析対象となるか確認する必要があった．

そこで，ローカルな実験環境を構築し，監視対象のネットワークのトラフィックレートを想定したバックグランドトラフィックを発生させ，さらに2.2節で定めた攻撃データを流しサンプリングベースの NetFlow データを観測する実験を行うこととした．

5. 攻撃データの観測実験

5.1. 実験環境

①攻撃データ

2章に基づき，攻撃データとして，マルウエアがイントラネットで 1 台の PC に感染し，ファイルを http POST で漏洩することを想定する． 5M バイトのファイルを 5 分に 1 回 POST する．

②バックグランドトラフィック

想定する監視環境のバックグランドトラフィックとして約 240Mbps の http 通信を発生させた．

③NetFlow 生成装置

NetFlow プローブソフトウエア nProbe v4[2]を使用し，NetFlow データを観測した．スイッチのミラーポートからの通信データを，nProbe とコレクタをインストールした PC に取り込み，NetFlow データをファイルに出力した．

④サンプリングレート

開発システムのデータ処理性能とネットワークの流量から，パケット 1000 個に 1 個をサンプリングした．

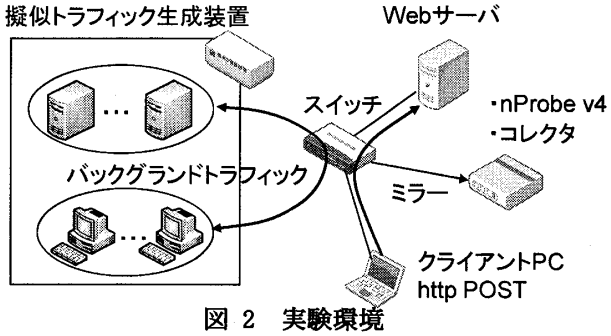

図 2 実験環境

5.2. 実験方法

図 2 の実験系において，クライアント PC から，Web サーバに http POST でファイルを送出させた．これを 60 回繰り返し行い NetFlow データを採取した．

5.3. 実験結果

結果を表 1に示す．5M バイトのファイルについて 60 回の POST 全てを NetFlow データとして観測できた．TCP のセグメントのデータ部分は最大 1460 バイトであり，5M バイトの場合は，送信すると約 3.42K パケットに該当する．サンプリングレート 1000 の場合は攻撃データのパケットが連続して流れたと仮定した場合 3.42K/1000=3.42 パケット観測されると見積もれる．実験ではこの見積もりよりやや大きい値になった．この結果，60 回の POST 全てが NetFlow データとして観測された．

比較として， 2M バイトのファイルについても観測した．同様の計算で 1.37 パケット観測される見積もりとなるが，実測はやや小さく NetFlow データの観測数も減少している．なお，参考として NetFlow データあたりに観測されたバイトサイズも示した．

表 1 バックグランド 240Mbps 測定結果

ファイルサイズ	NetFlow データの観測数	パケット/NetFlow	バイト/NetFlow
5M バイト	60	3.78	5367.00
2M バイト	36	1.12	1561.40

バックグランドトラフィックを 20Mbps にしたところ，5M では NetFlow データの観測数に変わりは無いが，2M では 1.5 倍に増加した(表 2)．

表 2 バックグランド 20Mbps 測定結果

ファイルサイズ	NetFlow データの観測数	パケット/NetFlow	バイト/NetFlow
5M バイト	60	3.91	5423.73
2M バイト	54	1.65	2323.53

表 3は全パケット数に占める送信パケットの割合を POST 時間 5 秒として換算したものである．

表 3 全パケット数に占める送信パケット数の割合

パケット数	240Mbps (198K パケット/秒)	20Mbps (18K パケット/秒)
5M(3.42K)	0.34%	3.66%
2M(1.37K)	0.14%	1.50%

以上より，ファイルサイズのパケット換算数が全通信パケット数に占める割合が大きい程サンプリングされ NetFlow データとして観測されるということが言える．

6. おわりに

実験によれば，想定する監視環境にあわせたサンプリングベースの NetFlow データの生成において，マルウエアによる情報漏えいを想定した攻撃データは，NetFlow データとして観測されることが分かった．今後は同方式で観測したイントラネットの通信データに攻撃データを混入し検知実験を行い，検知性能を確認したうえでイントラネットの監視に適用する予定である．

参考文献

[1]北澤，祢宜，河内，榊原，藤井，“標的型攻撃検知システムの評価”，MWS2009

[2]ntop, http://www.ntop.org/

[3]http://features.techworld.com/security/3626/a-monster-sized-security-mess/