4V-7

トラフィックパターン解析に基づく P2P ファイル共有ソフトウェアの利用検出

元木 伸宏† 泉 裕†† 塚田 晃司†††

†和歌山大学大学院システム工学研究科 ††和歌山大学システム情報学センター

†††和歌山大学システム工学部

1 はじめに

情報インフラの遍在性，ネットワーク技術の安定性が高まるにつれて，高品質な通信環境が整いつつあり，新たな通信デバイス・ソフトウェアが次々と作り出されている．インターネット上にオーバレイ環境を構築する P2P（Peer to Peer）ファイル共有ソフトウェアも上記の一つである．ファイル共有ソフトウェアは，不特定多数のユーザ間で特定ファイルを容易に共有できるため，多くのユーザに利用されている．しかし，ユーザによる P2P ファイル共有が企業や教育機関でも見受けられる一方で，著作権の侵害，情報の漏洩，あるいは帯域の圧迫などの深刻な問題も表面化している．各問題を回避するため，セキュリティポリシにより利用を制限する事例が多いが，潜在的な利用者を排除することは困難である．一般にはパケットのデータ・ペイロードを検閲する手法や製品があるが，特定のファイル共有ソフトウェアに特化したものであり，検閲に要するオーバヘッドも大きい．

本研究では，ネットワーク上での TCP セッションや UDP パケットの流れから，P2P ファイル共有の挙動を検出する汎用的な手法を提案する．さらに，管理者によるファイル共有ソフトウェア検出が可能な，本手法に基づいた支援システムを構築し，和歌山大学内ネットワークを対象にした実験結果について述べる．

2 技術概要

検出対象であるファイル共有ソフトウェアについて説明し，本研究におけるトラフィックパターンの概念について述べる．

2.1 ファイル共有ソフトウェア

ファイル共有ソフトウェアは，不特定多数の PC 間で専用のプロトコルを用いて通信し，ファイルを共有する．さらに，同ソフトウェアはファイルを共有する際に独自のオーバレイ環境を構築し，規模の拡大が容易である P2P 方式を採用している．以下は，今回の実験で検出対象とした 9 つのファイル共有ソフトウェアである．

- Winnyp…Winny の後継
- WinMX…Napster 互換クライアント
- LimeWire…Gnutella クライアント
- Cabos…LimeWire の後継
- Share10_ex2…Share の TCP 版
- Share10_nt5…Share の UDP 版
- eMule…eDonkey の後継
- Azureus…BitTorrent クライアント
- BitComet…BitTorrent クライアント

2.2 トラフィックパターン

本研究で用いるトラフィックパターンとは，「一定時間でどれだけのホストとどのように通信したか」というトラフィック全体の様子を指す．さらに，データ・ペイロードの情報を除く，パケットに含まれる IP ヘッダの情報（送受信の IP アドレスなど）のみを用いてトラフィックパターンとする．よって，個々のパケットを精査するのではなく，ファイル共有ソフトウェアが動作する端末の振る舞いを監視する．上記に類似するトラフィックパターンは，様々な研究で用いられている[1][2].

3 提案手法

3.1 ファイル共有ソフトウェアの挙動

表 1 に，2.1 節で述べた 9 つのファイル共有ソフトウェアを動作させ，3 分間ほどの短時間のデータを採取した結果を示す．同結果から，ファイル共有ソフトウェアには，瞬間的に通信先ホストの数が増加する傾向がある．同結果を得る過程において，eMule を除くソフトウェアは単にアプリケーションを起動させ，eMule は起動後，サーバと接続させた．いずれの場合においても，特定ファイルをダウンロードせずに，アプリケーションの起動のみで多数のセッションが生成されることが確認できる．

P2P file sharing detection based on analysis of traffic pattern
†Nobuhiro Motoki, Graduate School of Systems Engineering, Wakayama University
††Yutaka Izumi, Center for Information Science, Wakayama University
†††Koji Tsukada, Faculty of Systems Engineering, Wakayama University

表 1: 採取結果

名称	使用プロトコル	通信先ホストの数
Winnyp	TCP	19
WinMX	TCP	4
LimeWire	TCP, UDP	13, 243
Cabos	TCP, UDP	15, 215
Share10_ex2	TCP	58
Share10_nt5	UDP	89
eMule	TCP, UDP	7, 21
Azureus	TCP, UDP	66, 1290
BitComet	UDP	9161

3.2 トラフィックパターンによる検出

前述の通り，ファイル共有ソフトウェアは瞬間的な通信先ホスト数の増加が大きな特徴である．さらに，TCP トラフィックには，ある値から連続した送信元ポート番号が用いられている．上記をふまえ，TCP，UDP それぞれについて以下のような傾向に注目する．

a. TCP トラフィック
- 広範囲のホストと通信している
- 送信元ポート番号がほぼ連続している
- well-known ポート以外を使用している

b. UDP トラフィック
- TCP よりも広範囲のホストと通信している
- well-known ポート以外を使用している

ここで，ファイル共有ソフトウェアは，TCP，UDP 共に，well-known ポート以外で通信するため，well-known ポートを含むトラフィックを除外する．提案手法では，前述のような挙動が現れるトラフィックを検出し，TCP トラフィックの送信元ポートの連続性については，現在は管理者の目視により判断する．

3.3 システム構成

本研究で作成したシステムで使用するソフトウェア，およびシステム構成について述べる．本システムは，学内に流れるトラフィックから IP アドレスやポート番号などの情報をキャプチャし，DB へ登録する mcapture，取得データを格納する DB である MySQL，提案手法により示したトラフィックを管理者が検出するための mshow の 3 つのソフトウェアから図 1 のように構成される．本研究では，mcapture，mshow を作成し，管理者は mshow により指定した時刻の被疑トラフィックを検索し，取得する．

4 実装・実験

4.1 実装環境

本システムは，3.3 節で述べた 3 つのソフトウェアを 1 台の PC 上で動作する FreeBSD に実装し，実験は和歌山大学の対外接続部分で実施した．同部分はトラフィックが収束されており，L2 スイッチの監視ポートを使用している．

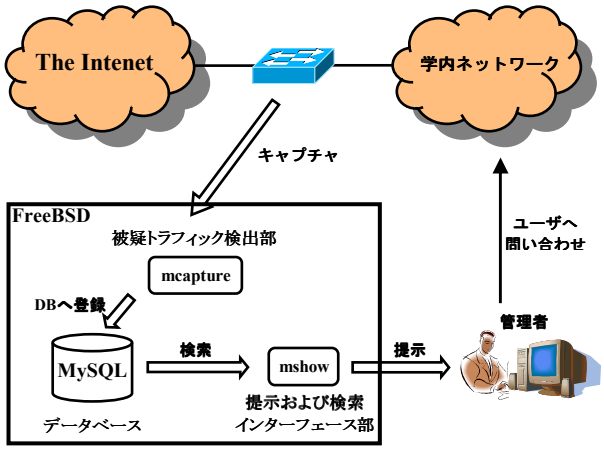

図 1:システム構成図

4.2 実験

2.1 節で述べた 9 つのファイル共有ソフトウェアを，学内で 1 つずつ動作させ実験を行った．なお，通信先ホストの数に関して，閾値を設定する必要があるが，今回は 3.1 節の採取結果により，適切であると考えられる値を閾値として用い，WinMX を除く 8 つを検出した．

5 評価・考察

WinMX を検出できなかった理由には，ハイブリット型であり，通信先ホストの数が少ないことが考えられる．実験の際，対象とは関係のないホストも検出されたが，ユーザへの問い合わせにより，ファイル共有ソフトウェアであることが確認されている．さらに，実験で使用したソフトウェアだけでなく，ネットワーク上の挙動が類似していれば，他のファイル共有ソフトウェアなども検出が可能であり，過去に存在した被疑トラフィックも検出可能である．

6 おわりに

本稿では，トラフィックパターンを用いて P2P ファイル共有の挙動を検出する汎用的な手法を提案し，多種のソフトウェアを検出することで，その有効性を示した．しかし，TCP トラフィックにおける送信元ポート番号の連続性を自動化することや，通信先ホストの数に関する統計学的な閾値設定などが今後の課題として考えられる．

参考文献

[1] 戸川聡，金西計英，矢野米雄，“トラフィックマイニングと可視化による Peer-to-Peer ファイル共有検出支援システムの構築”，情報処理学会，2007-DSM-45，pp.99-104，2007.

[2] 藤井聖，中村豊，藤川和利，砂原秀樹，“通信先ホスト数の変化に注目した異常トラヒック検出手法の提案と評価”，信学論，Vol.J88-B，No.10，pp.1922-1933，2005.