2003－DSM－31　(4)
2003／11／26

トラフィック傾向可視化による P2P ファイル共有通信検出支援モデルの提案

戸川 聡 *　　金西計英 **　　矢野米雄 ***

* 徳島大学大学院工学研究科
** 徳島大学高度情報化基盤センター
*** 徳島大学工学部

概要：近年普及しつつある Peer-to-Peer（P2P）ファイル共有ソフトウエアを利用した著作権侵害が問題となっている．P2P ファイル共有ソフトウエアの使用そのものは問題ない．しかし，インターネット上に展開されるファイル共有コミュニティで流通する大半のデータは，音楽 CD や DVD から抽出された著作物であり，これらのほとんどは著作権者の許諾を得ないまま不法に公開されている．企業や大学のキャンパスネットワークでは，P2P ファイル共有ソフトウエアの使用自体が禁じられている場合がある．しかし既存のポートフィルタリングなどでは P2P ファイル共有通信を制限することは難しい．結果，管理者はトラフィックを常時監視し，P2P ファイル共有通信を検出しなければならない．本稿では，管理者がおこなう P2P ファイル共有通信検出作業を軽減するため，検出支援モデルを提案する．

Proposal on the Peer-to-Peer Traffic Detection Assistance Model Using the Traffic Activity Visualization

Satoshi Togawa*, Kazuhide Kanenishi** and Yoneo Yano***

*Graduate School of Engineering, University of Tokushima
**Center for Advanced Information Technology, University of Tokushima
***Faculty of Engineering, University of Tokushima

Abstract: In this research, we have proposed the assistance model for peer-to-peer traffic detection. Recently, an illegal file has been exchanged with peer-to-peer file exchange software. These files are extracted from music CD and DVD. Most files do not obtain the copyright person's approval and are open to the public. Neither enterprise nor the Campus Network user of the university must acquire these files from the problem in morality. However, the illegal file is actually acquired via Campus Network. The network administrator should observe the users' peer-to-peer communication.

In this paper, first of all, We explain a problem of peer-to-peer file exchange system. Next, we explain the proposal on the peer-to-peer traffic detection assistance model. Finally, we conclude it.

1　はじめに

Peer-to-Peer（P2P）通信によるファイル共有が問題となっている．これを実現するソフトウエアとして，WinMX[1] や Winny[2]，KaZaA[3] などが存在する．一般的にこれらのソフトウエアは，公開対象としたディレクトリ中に存在するファイル群を，インターネット上の不特定多数の利用者に公開する．公開可能なファイルの種別に技術的制約はない．コンピュータシステム上にファイルとして存在可能であれば，そのデータ内容に関わらず公開が可能となる．

P2P 共有ソフトウエアを用いて，ファイルをインターネット上に公開する行為そのものに違法性はない．しかし，音楽 CD から抽出した楽曲データや DVD から抽出した動画データなど，著作件法で保護される著作物をデータファイル化し，著作件者の許諾を得ず公開，共有することはできない．これらの行為は，著作権法で規定される公衆送信権，送信可能化権の侵害にあたる．

最近普及しつつある Winny は，あるファイルの取得動作を始めた時点で取得データをキャッシュデータ化する．Winny は対象ファイルの取得動作と同時にキャッシュデータを送信可能状態に置く．このた

め Winny 利用者は，著作権法で保護されるデータの取得を開始した時点で，送信可能化権を侵害する可能性を否定できない．

著作権法で保護されるデータをファイル化し，インターネット上で共有する行為は以前から行われてきた．古くは音楽 CD から抽出した楽曲データを MPEG-1/Audio Layer3（MP3）[4] 形式などにエンコードし，Anonymous FTP や無料ホームページに掲載する形が主流であった．これらは従来型のクライアント-サーバ型ファイル共有モデルである．

P2P 型，クライアント-サーバ型を問わず，インターネット上に不法公開されるデータファイルを取得する行為はモラル上好ましいとは言えない．特に企業や大学のキャンパスネットワークを経由してファイル取得を行うとき，多くの場合キャンパスネットワークの利用規定に反する目的外利用に該当する．さらに使用する P2P 型共有ソフトウエアによっては，データファイルを取得後自動的に共有状態に置く．これは当該ソフトウエア利用者が意図する，しないに関わらず送信可能化権侵害の可能性が残る．

クライアント-サーバ型ファイル共有モデルの場合，不法データを公開するサーバは頻繁には移動しない．したがって企業や大学のネットワーク管理者が，不法なクライアント-サーバ型ファイル共有の利用を制限する場合，公開サーバの IP アドレスを特定し，そのサーバへの通信をフィルタリングすれば一定の効果を実現できた．

しかし P2P ファイル共有の利用を制限する場合，データ転送元となるホストが不特定多数かつ可変であることから，IP アドレスベースの通信フィルタリングは効果的ではない．さらに最近主流の Winny は，標準的な待ち受けポート番号を持たず，ランダムに設定された TCP ポートにて接続を待ち受ける．したがってポート番号ベースのフィルタリングも利用制限には効果的ではない．

このため既存のフィルタ技術では P2P ファイル共有の制限は困難である．したがって，ネットワーク管理者が P2P ファイル共有通信の制限を試みる場合，管理するネットワーク内の P2P ノードから受発信される P2P トラフィックを検出し，個別に対応しなければならない．

現在, P2P ファイル共有通信検出のために Snort[5] などの侵入検知システム（Intrusion Detection System:IDS）を使用できる．本来 Snort は不正侵入検知のために構築されたシステムだが，ルールを記述することで P2P トラフィック検出に応用できる．しかし，ルール記述文法が複雑であることから管理者への負担が大きい．また，パターンに適合しないと P2P トラフィックとして管理者に通知しないため，取りこぼしなどの誤検出は避けられない．さらにルールに基づく限り，新たなトラフィックパターンを持つ P2P ファイル共有通信の検出は難しい．

より詳細な調査は，IDS などが出力するログ情報を調べることで実現される．しかし，一般にログ情報は膨大なテキスト情報で構成される．手がかりのない状態でログ情報調査を行っても，効率が悪いうえ多大な労力が必要となる．

そこで本研究では，トラフィック傾向可視化による P2P ファイル共有通信検出支援方式を提案する．監視対象ネットワークから送受信されるトラフィックをもとに特徴量を抽出しモデル化する．生成されたモデルを可視化し特徴マップを生成する．管理者は特徴マップを参照することで，定常のトラフィック傾向の俯瞰が可能となる．定常状態の把握により，低頻度で発生する特異状態に気づきやすくなる．この結果，ログ情報調査時に「あたり」をつけやすくなり，調査負担を軽減できる．

以下本稿では，2 章で P2P ファイル共有通信の現状について述べ，3 章でこれらファイル共有通信の検出支援モデルについて述べる．4 章で本研究で使用するトラフィックモデルの構成と可視化手法を述べ，5 章でまとめる．

2 P2P ファイル共有通信の現状

2.1 P2P ファイル共有の通信形態

現在主流の P2P ファイル共有通信は，以下の 2 つに分類できる．

Hybrid 型： 図 1 に Hybrid 型 P2P 通信の例を示す．リソース探索機能およびノード探索機能は中央サーバに依存し，リソース交換はノード間にて行う方式である．中央サーバに蓄積した索引にて検索を行うため，高速なリソース検索が可能となる．ノードとなるクライアントは，中央サーバから示された相手ノードとファイル交換を行う．しかし中央サーバに障害が発生すると P2P 網そのものが機能しなく

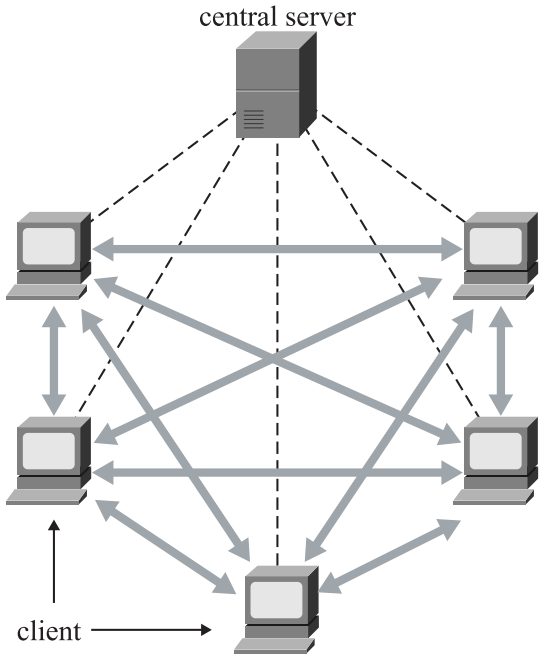

図 1: Hybrid 型 P2P 通信モデル

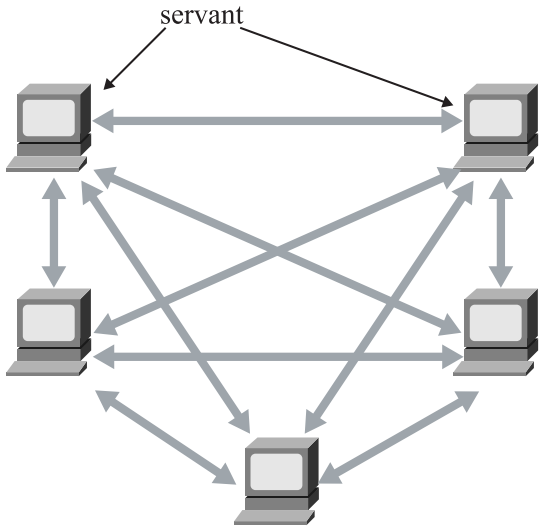

図 2: Pure 型 P2P 通信モデル

なる．このため Hybrid 型は中央サーバが単一障害点になり得ると言える．

Pure 型： 図 2 に Pure 型 P2P 通信の例を示す．Pure 型はリソース検索機能やノード探索機能を管理する中央サーバを持たない．これらの機能は各ノードに実装される．あるノードがリソースを探索する場合，対等に存在する他ノードの連携により探索が実現される．各ノードはサーバ機能，クライアント機能の両方を実装するため，サーバントと呼ばれる．このため理論上単一障害点が存在せず，高可用性を実現できる．しかしリソース探索を P2P 網全体に対し分散的に行うため，レスポンスが遅くなりやすい．

Hybrid 型 P2P ファイル共有を実装している代表的ソフトウエアとして，Napster や WinMX が存在する．また，Pure 型 P2P ファイル共有を実装する代表的ソフトウエアとして，Gnutella[6] や Winny, Freenet[7] が存在する．

2.2 Pure 型 P2P ファイル共有通信の特徴

現在代表的な Pure 型 P2P ファイル共有ソフトウエアと言える Winny を例に，そのトラフィック特徴を述べる．Pure 型 P2P ファイル共有通信の発生時からファイル取得までを 4 フェーズに分け，それぞれの特徴を述べる．

2.2.1 コミュニティ参加フェーズ

インターネット上に存在する P2P コミュニティに自分ノードを参加させるフェーズである．P2P コミュニティに参加するためには，まず初期ノードリストを入手する．初期ノードリストには既にコミュニティに参加しているホストの IP アドレスとポート番号が暗号化され記載される．宛先ポート番号には標準値が存在しない．Winny の待ち受けポートはソフトウエア導入時にランダム選択される．この結果，初期ノードリストに記載される宛先 IP アドレスと宛先ポート番号に一貫性は存在しない．

コミュニティに参加しようとする自ノードは，事前取得した初期ノードリストのうち 20〜30 ノードを対象に，一斉にコネクションを生成する．前述の理由より，宛先 IP アドレスおよび宛先ポート番号には一貫性がない．このため表層的には多数のランダムな宛先に対しコネクションが生成される．初期値設定にも左右されるが，初期コネクションから 2〜4 ノードを選択し上流ノードに設定する．

コミュニティ参加フェーズにおける，送信元，宛先の各 IP アドレス，ポート番号の例を表 1 に示す．

2.2.2 待機フェーズ

上流ノードが選択された後，待機フェーズとなる．待機フェーズではおおむね 2〜4 程度の上流ノード，および同数程度の下流ノードとリンクを確立し，定常状態に入る．この間，自ノードを中心とした宛先ノードの増減は少ない．しかしリンクを確立した上流ノード，下流ノードとの間で継続的にリンク情報

表 1: コミュニティ参加フェーズにおける接続状態（抜粋）

送信元 IP アドレス	送信元 TCP ポート	宛先 IP アドレス	宛先 TCP ポート
aaa.bbb.114.10	1565	ccc.45.178.161	20008
aaa.bbb.114.10	1566	ddd.111.97.41	3768
aaa.bbb.114.10	1567	eee.187.56.93	1826
aaa.bbb.114.10	1568	fff.145.49.71	1511
aaa.bbb.114.10	1569	ggg.164.131.58	7700
aaa.bbb.114.10	1570	hhh.228.180.73	31528
aaa.bbb.114.10	1571	iii.124.29.240	7667
aaa.bbb.114.10	1572	jjj.213.133.144	9119
aaa.bbb.114.10	1573	kkk.127.236.73	7743
aaa.bbb.114.10	1574	lll.111.106.118	12036
aaa.bbb.114.10	1575	mmm.107.244.165	7410
aaa.bbb.114.10	1576	nnn.110.154.21	24646
aaa.bbb.114.10	1577	ooo.199.8.124	4000

が交換される．しかし下流ノードからのリソース検索要求が不定期に依頼されるため，自ノードへのコネクションが増減する．

2.2.3 検索フェーズ

取得するリソースを探索するため，検索語を入力し検索する．この段階では既に上流，下流の検索リンクが確立しているため，接続コネクションの大きな変動はない．

2.2.4 取得フェーズ

検索結果として示されたリソース一覧から，ファイル取得を行う．Winny は 1 ヶ所のノードからファイル転送を行うのではなく，コミュニティ全体から複数選択された転送元から分散的にファイル転送を行う．したがって取得フェーズにおいては接続コネクションの増加が観測できる．同時に転送元に選択された相手ノードから，自ノードに対し継続的なデータ転送が行われる．

2.3 フィルタリングによる利用制限の検討

ネットワーク利用者への通信制限実現のために，ポートベースや IP アドレスベースのフィルタリング技術によるアクセス制限手法が存在する．本節では既存技術であるフィルタリングを用いた P2P ファイル共有通信の制限を検討する．

Hybrid 型 P2P ファイル共有通信は中央サーバが単一障害点となるため，中央サーバへの経路を遮断すれば理論上容易にリソース検索機能を遮断できる．このため既存のフィルタリング技術は，Hybrid 型 P2P ファイル共有通信の制限に関しては一定の効果が期待できる．しかし現実には複数の中央サーバが運用されているため，これらをすべて網羅することは難しい．

Pure 型 P2P ファイル共有通信の制限は，さらに困難をともなう．Pure 型 P2P ファイル共有利用者は，インターネット上に無数に存在する P2P サーバントのうち，いずれか 1 つに接続できれば P2P ファイル共有コミュニティに参加できる．さらに Freenet や Winny では，待ち受けポート番号がランダムに生成されるため，既存のポートフィルタ技術は適用できない．

3 検出支援モデル

2 章では P2P ファイル共有通信の現状について述べた．P2P ファイル共有通信形態を Hybrid 型, Pure 型にそれぞれ分類し，特に今後主流になりつつある Pure 型 P2P ファイル共有通信につき，トラフィックを表層現象としてとらえた場合の特徴を述べた．本章では，P2P トラフィック検出支援モデルについて述べる．

3.1 検出支援の枠組み

2.3 で述べたように，P2P ファイル共有通信の利用制限に既存技術であるポートフィルタリングを用いることは難しい．特に Pure 型 P2P トラフィック

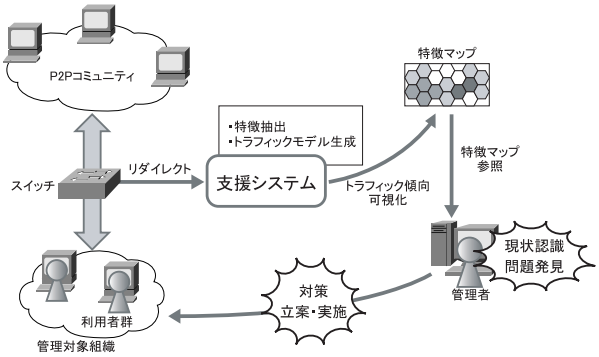

図 3: 検出支援モデル

の特性であるランダムな宛先ポート選択が，フィルタリング技術の適用を困難にしている．

しかし見方を変えれば，P2P ファイル共有トラフィックは広範囲な宛先 IP アドレスと宛先ポートを対象に通信を行っていることがわかる．従来のクライアント- サーバ型通信のように，コネクション確立後に同一宛先ポート間にて通信を行うのではない．頻繁に宛先 IP アドレス，宛先ポートを変更し，自ノードに接続を試みる他ノードも頻繁に発生することがわかる．

IP アドレスを X 平面，宛先ポートを Y 平面として宛先 IP アドレスと宛先ポート番号をプロットすると，三次元空間に一様分布することがわかる．一方，従来のクライアント-サーバ通信は，ほとんどが同一ポートを宛先とするため，三次元空間上に線，または点として現れる可能性が高い．

監視対象トラフィックの送信元，宛先の各 IP アドレス，ポート番号および発生頻度とそれぞれの間の通信量を抽出し可視化すれば，管理者は監視対象のトラフィック傾向を俯瞰可能となる．通常は点ないし線として表出する特徴マップ上に，一定の平面を占める部分が表出すれば，通常とは異なるトラフィックが発生したことを示唆できる．

図 3 に本研究で提案する検出支援モデルを示す．利用者群が行う通常の通信は，キャンパスネットワーク内各所に設置される Layer2 スイッチを経由する．スイッチに実装されるポートフォワード機能により，監視対象トラフィックを本研究で提案する支援システムに転送する．支援システムは，取得したトラフィックから特徴量を抽出しモデル化する．本研究では，これをトラフィックモデルと呼ぶ．トラフィックモデルには, 送信元, 宛先の IP アドレス, ポート番号をインデックスとして，コネクション生成数と単位時間当りのデータ転送量を集積する．

生成されたトラフィックモデルを特徴マップとして可視化する．管理者は特徴マップを参照することで，管理対象組織における定常のトラフィック傾向が俯瞰でき，定常状態を把握することで低頻度で発生する異常事象の発見が可能となる．

4　モデル構成と可視化

4.1　トラフィックモデルの構成

本節では，本研究で用いるトラフィックモデルの構成について述べる．トラフィックモデルは次の要素から構成される．

1. 特徴ベクトル
2. 送信元ポート番号
3. 宛先ポート番号
4. 単位時間あたりのデータ転送量
5. コネクション生成数

特徴ベクトルは送信元 IP アドレスごとに生成される．宛先 IP アドレス，送信元ポート番号，宛先ポート番号をインデックスとして，単位時間あたりのデータ転送量，コネクション生成数を特徴量として保持する．

n を送信元 IP アドレス，m を宛先 IP アドレスとすると，トラフィックモデルは次式で表現できる．

$$\boldsymbol{A'} = \begin{pmatrix} \boldsymbol{A_{11}} & \boldsymbol{A_{12}} & \cdots & \boldsymbol{A_{1n}} \\ \boldsymbol{A_{21}} & \boldsymbol{A_{22}} & \cdots & \boldsymbol{A_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{A_{m1}} & \boldsymbol{A_{m2}} & \cdots & \boldsymbol{A_{mn}} \end{pmatrix} \quad (1)$$

トラフィックモデルの各要素には，送信元ポート番号，宛先ポート番号をインデックスとして，単位時間あたりのデータ転送量，コネクション生成数が保持される．

x, y をそれぞれ送信元ポート番号，宛先ポート番号とすると，トラフィックモデルの一要素は次式で表現できる．

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{B_{11}} & \boldsymbol{B_{12}} & \cdots & \boldsymbol{B_{1y}} \\ \boldsymbol{B_{21}} & \boldsymbol{B_{22}} & \cdots & \boldsymbol{B_{2y}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{B_{x1}} & \boldsymbol{B_{x2}} & \cdots & \boldsymbol{B_{xy}} \end{pmatrix} \quad (2)$$

また，格納される特徴量として，a を単位時間あたりのデータ転送量，b をコネクション生成数すると，各要素は次式で表現できる．

$$\boldsymbol{B} = \{a, b\} \tag{3}$$

この結果トラフィックモデルは，送信元 IP アドレスごとに割当てられる特徴ベクトルを集合させた多次元ベクトル集合であると言える．

4.2 自己組織化マップによる可視化

トラフィックモデルは 4.1 で述べたとおり多次元のベクトル集合である．一般に人間が直感的に認識できる次元空間は三次元までである．このため，管理者にトラフィックモデルをそのまま提示しても，人間の空間認識能力をはるかに超えるため，直感的な認知が難しい．

Kohonen により提唱された自己組織化マップ (Self-Organizing Map:SOM) は，2 層のニューラルネットワークで構成される教師なし競合学習モデルである．SOM はデータ間の幾何学的構造を可能な限り保った状態で二次元平面に写像する．同時にクラスタリングも行う．

この結果，管理者は平易な二次元平面にて管理対象組織のトラフィック傾向の俯瞰が可能となる．

5 まとめ

本稿では，企業や大学のキャンパスネットワークで行われる P2P ファイル共有通信の問題について述べ，これらの P2P トラフィックを既存のフィルタリング技術で制限することの困難性について述べた．その上でキャンパスネットワーク内から受発信される P2P トラフィックを検出する手法について検討し，管理者が行う P2P トラフィック検出のための支援モデルを提案した．さらに支援モデルを実現するために必要なトラフィックのモデル化手法について述べ，多次元モデルの認識限界を下げトラフィック傾向の俯瞰を可能にするために行う可視化手法について述べた．

今後は提案手法の有効性を検証するために試作システムを実装し，実証実験を行う予定である．

参考文献

[1] WinMX Web Site, http://www.winmx.com/

[2] Winny Web Site, http://www.geocities.co.jp/SiliconValley/2949/

[3] KaZaA Web Site, http://www.kazaa.com/

[4] Moving Picture Experts Group(MPEG) Web Site, http://www.chiariglione.org/mpeg/index.htm

[5] Snort Web Site, http://www.snort.org/

[6] Gnutella Web Site, http://www.gnutella.com/

[7] Freenet Web Site, http://freenetproject.org/