

ユーザの活動状況が個々のネットワークトラフィックに与える影響

柿崎歩^{†,a)} 上岡英史^{†,b)}

情報通信技術の発達に伴い、電話や電子メールだけではなく、LINE や Skype などの通信アプリケーション、Twitter や Facebook などの SNS が登場し、人々のコミュニケーション手段は多様化してきている。しかし、このような現状においても、「通話」は最も頻りに利用されるコミュニケーション手段の一つである。それは、相手とリアルタイムにコミュニケーションが取れるためであるが、その特性ゆえにコミュニケーションのすれ違いが起こるという問題がある。本研究では、ネットワークトラフィックの周期性に着目した解析を行うことによって、ユーザの状況を推定する手法を探求する。本稿では、ユーザの利用するアプリケーションをネットワークトラフィックの特徴を用いて判別し、それをもとにユーザが工作中であるのか、休憩中であるのかを推定する方式を提案する。また、予備実験を通して提案手法の有効性を評価する。

Influence of User Activity on Individual Network Traffic

AYUMU KAKIZAKI ^{†,a)} EIJI KAMIOKA ^{†,b)}

As the spread of information and communication technologies, communication means have been diversified. For example, VoIP services such as LINE and SKYPE, and SNSs such as Twitter and Facebook are widely used. Even in such a convenient communication environment, “telephone call” is one of the most popular services because it enables a real-time communication. However, the real-time service, sometimes, cannot reach the correspondent as the characteristic shows, namely, he/she may be in an office meeting, for instance. This study explores an approach to estimate user’s situations focusing on the periodicity of network traffic. In this paper, a framework for inferring if a user is working or taking a rest using the characteristics of network traffic which each network application generates. In addition, the effectiveness of the proposed method will be evaluated through a preliminary experiment.

1. はじめに

スマートフォンやタブレット型 PC といったモバイル端末の普及によって、ブロードバンド回線によるホームネットワークに留まらず、外出先においてもユーザがネットワークに常時接続されている環境が整備されつつある。さらに、情報通信技術の発展は目覚しく、このような環境とともに、人々はサービスプロバイダから様々なサービスを受けられるようになってきた。

しかし、サービスが多様化することで、ユーザが望んでいないサービスを提供される事も多くなっている。

そこで、センサなどから周囲の情報を取得し、それを元にユーザの置かれた状況に適したサービスを提供するコンテキストウェア・サービスの研究が盛んに行われており、実世界の情報を取得する方法とそれを用いた状況判別といった部分において、研究者毎の違いが現れている。

また、これらの通信環境が整ってきた現在においても、「通話」というコミュニケーション手段は頻りに利用されている。これは、メールなどとは違い、相手とリアルタイムにコミュニケーションが取れるためであるが、そのリアルタイム性ゆえに、相手の状況によっては通話が行えない場合も考えられる。例えば、電車の中にいる時などは、電

話が繋がったとしても、マナーの観点から通話を遠慮しなければならない。さらに、この時の着信が留守番電話に接続され、それに気づいたときは相手に電話を掛けなおすと考えられるが、その際、最初に電話してきた相手が必ず電話に出られるとは限らない。この時、相手の状況を通話前に知ることができれば、相手の都合が良い時に通話を行うことが可能になり、コミュニケーションのすれ違いを減らすことが出来ると考えられる。

本研究は、ユーザの置かれた状況を推定することで、このコミュニケーションのすれ違いを解消することを目的とする。実世界の情報を取得する方法としては、ネットワークが利用可能な環境において、ユーザが端末を操作した際に発生するネットワークトラフィックを用いる。著者らは、このネットワークトラフィックは利用するアプリケーションによってその周期性に違いが現れるということ、これまでの研究で明らかにしてきた。そして、ユーザの状況によって、利用しているアプリケーションは異なると考えられるため、利用アプリケーションを判別することが出来れば、ユーザの状況を推定することが可能であるという仮説のもと、その検証を行なう。

本稿では、会社員の一日を想定したスケジュールから、ネットワークを利用した行動をリストアップし、ネットワークトラフィックの周期性の違いを利用したアプリケーション判別によって、ユーザが「工作中」であるか、または、「休憩中」であるかの推定を行うことを目標とした。さらに、その正解率を評価実験によって求め、提案手法の有効

[†] 芝浦工業大学
Graduate school of Engineering and Science, Shibaura Institute of Technology

a) E-mail: ma11048@shibaura-it.ac.jp
b) E-mail: kamioka@shibaura-it.ac.jp

性について議論する。

以下、第2節ではネットワークトラフィックを利用した利用アプリケーションの判別手法の関連研究について説明し、第3節では今回提案する自己相関係数およびDPマッチングを用いたアプリケーション判別システムの構成について述べる。第4節では、実際に測定したネットワークトラフィックのデータに対して提案手法を適用し、正解率による評価および考察を行う。最後に、第5節にて本稿のまとめと今後の課題を述べる。

2. 関連研究

従来の研究において、データを送受信しているアプリケーションの判別には、ポート番号を利用した手法が一般的であった。これは、通信で使用されるポート番号がアプリケーションによって固定されているため、そのポート番号を識別することで、利用アプリケーションを判別するというものである。しかし、最近ではWEBブラウザ上で様々なクラウドアプリケーションを実行するSaaS(Software as a Service)などが普及し、ポート番号だけで利用アプリケーションを推定することは難しくなっている。

これに対し、ポート番号を用いないアプリケーション判別手法として、トラフィックの変化に着目した研究がある。

八木氏ら[1]は、フローの初期段階における挙動がアプリケーション毎に固有であることに着目し、フロー開始直後数パケットのペイロード長の遷移パターンに着目した判別手法を提案している。なお、ここでのフローの定義は、送受信IPアドレス、送受信ポート番号、プロトコルの5-tupleの組合せが一致するパケットの集合と定義されている。

山下氏ら[2]は、ネットワークトラフィックから得られる特徴量の時系列データに着目し、パターン認識を用いたサービス分類によって、サービス毎に適したネットワーク品質を提供する方法について検討している。

3. 提案手法

本稿で提案するアプリケーション判別システムを図1に示す。

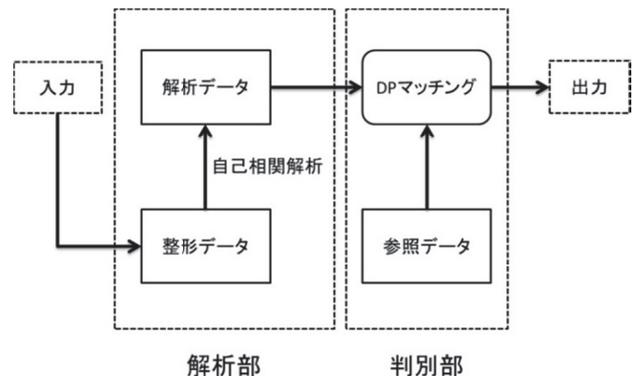


図1 アプリケーション判別システムの概要

3.1 解析部

解析部では、まず、入力されたネットワークのトラフィックデータに対して、ブロードキャストされたパケットなどを除去するため、IPアドレスによるフィルタリングを行なって、利用しやすい形へと整形する。その後、整形データに対して、自己相関係数を用いた解析を行うことで、データの周期的な特徴を検出する。

ここで用いる自己相関とは、あるデータがそれ自身を時間シフトしたデータとどれだけ相関があるかを求めた尺度であり、この平均を取ったものが自己相関係数と呼ばれ、(1)式で与えられる。

$$R(\tau) = \frac{1}{N} \sum_{t=1}^N \{v(t) \cdot v(t + \tau)\} \quad (1)$$

ここで、 t は時間[s]、 N は総時間[s]、 τ は時間差[s]、関数 v は入力データである。関数 R の値が1に近ければ正の相関が、-1に近ければ負の相関が強いことを示し、0に近ければ相関がないということが言える。

3.2 判別部

判別部では、解析後のデータと参照データをパターンマッチングによって比較し、その類似度により利用アプリケーションを判別する。

パターンマッチングのアルゴリズムとして、今回はDP(Dynamic Programming)マッチング[3][4]を利用した。DPマッチングとは、パターンの要素間に定義された類似度にもとづいて、パターンの伸縮まで考慮に入れたマッチング方式であり、文字、音声、モーションなど様々な分野で適用されている。入力データと参照データは、同じアプリケーションを利用した場合においても、データが一致するとは考えにくいいため、伸縮性を考慮したマッチング方式を採用した。

4. 評価実験

ここでは、各行動（工作中、休憩中）におけるトラフィックデータを取得・解析し、その結果を判別システムに入力することで利用アプリケーションの判別を行い、それをもとに推定する行動の正解率を用いて提案手法の評価を行う。

4.1 測定環境

ネットワークトラフィックの測定環境を表 1 に示す。

表 1 測定環境

端末	Windows ノートパソコン
OS	Windows 7
無線 LAN	802.11g
パケットキャプチャソフト	Wireshark[5]
WEB ブラウザ	Google Chrome
解析ソフト	MATLAB[6]

この他にも、Perl プログラムによる測定の自動化などを行っている。

また、利用した大学のネットワーク環境は、大学内イントラネットからインターネットにアクセスする場合、必ずプロキシサーバを経由する仕様となっている。

4.2 設定した行動

今回は、会社員の一日を想定し、その生活の中でネットワークを利用するものを測定する行動として設定した。「メール」や「ドキュメント編集」に関しては、ローカル環境で利用することが多いと考えられるが、今回は WEB ブラウザ上でアプリケーションを使用するクラウドコンピューティングサービスを用いた。

そして、この設定した行動について、表 2 および表 3 のように、工作中・休憩中という 2 つの状態に分類し、各行動時に利用しているアプリケーションと関連付けた。

表 2 「仕事」の行動とアプリケーション

「仕事」	アプリケーション
電子メール	Gmail
音声通話	Skype
ビデオ通話	Skype
ドキュメント編集	Google Drive

表 3 「休憩」の行動とアプリケーション

「休憩」	アプリケーション
SNS サイト利用	Twitter
音楽鑑賞	Grooveshark
動画視聴	ニコニコ動画
ネットラジオ聴取	radiko.jp

4.3 各行動の測定条件

測定は、表 4 の条件で行った。行動は 60 秒間行い、各 6 回ずつ測定した。

表 4 行動の測定条件

行動	概要
電子メール	電子メールの閲覧・送受信
音声通話	Skype で音声通話
ビデオ通話	Skype でビデオ通話
ドキュメント編集	ファイルの編集・保存
SNS サイト利用	Twitter の閲覧・投稿
音楽鑑賞	Grooveshark で音楽を聴く
動画視聴	ニコニコ動画で動画視聴
ネットラジオ聴取	radiko.jp でラジオ聴取

4.4 「仕事」行動の解析結果

ここでは、「仕事」と分類した行動について、その自己相関解析結果を図 2-図 5 に示す。なお、見やすさを考慮し、各行動 3 回分をプロットしている。

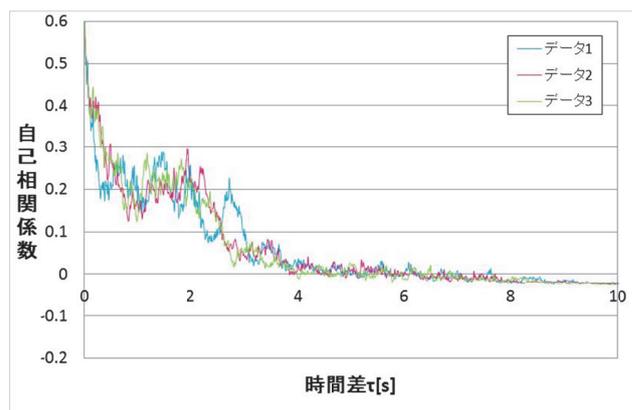


図 2 電子メールの自己相関解析結果

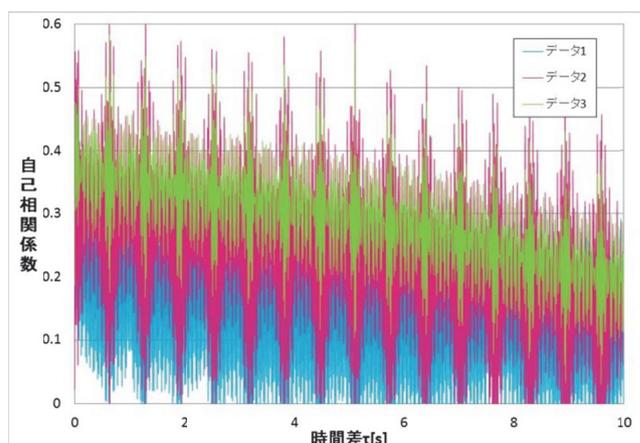


図 3 音声通話の自己相関解析結果

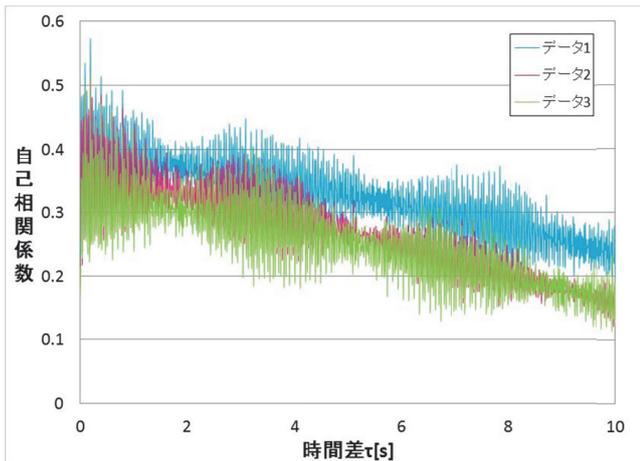


図 4 ビデオ通話の自己相関解析結果

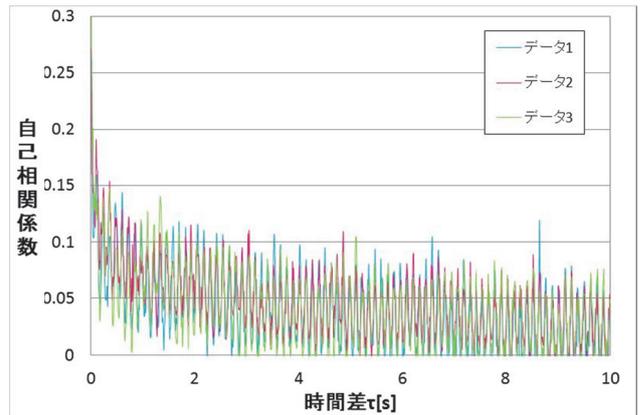


図 6 SNS サイト利用の自己相関解析結果

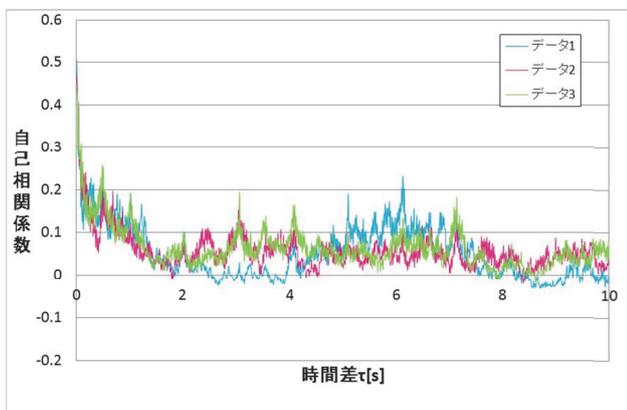


図 5 ドキュメント編集の自己相関解析結果

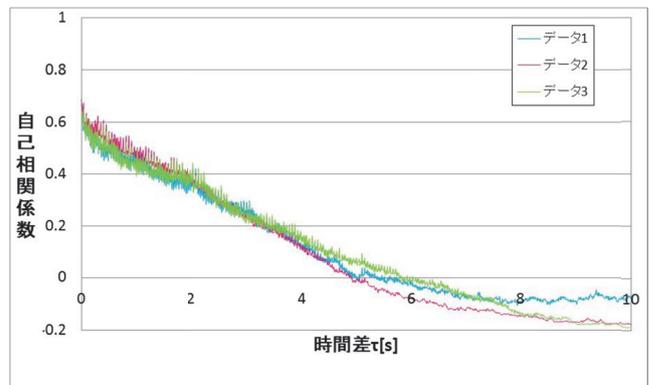


図 7 音楽鑑賞の自己相関解析結果

「仕事」行動で共通の特徴というものを見つけることが出来ない。しかし、同じアプリケーションを実行した場合、6 回分の解析データはすべて類似した傾向が得られている (図では 6 回中 3 回分の測定結果のみ表示している)。このことから、ネットワークトラフィックの周期性を用いて、利用アプリケーションの種類を区別できる可能性があることが確認できる。

4.5 「休憩」行動の解析結果

「休憩」と分類した行動について、その自己相関解析結果を図 6-図 9 に示す。なお、こちらも見やすさを考慮し、各行動 3 回分をプロットしている。

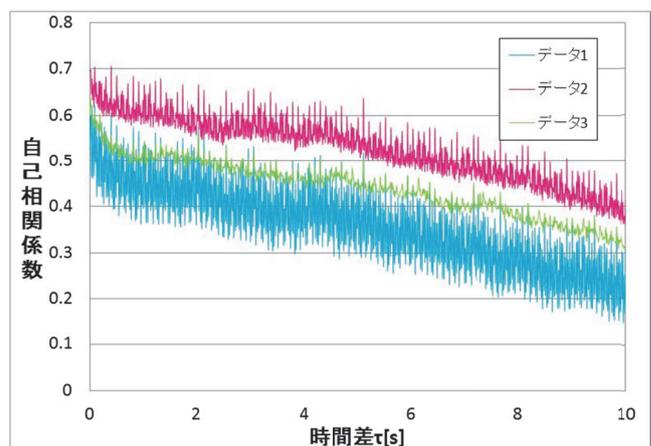


図 8 動画視聴の自己相関解析結果

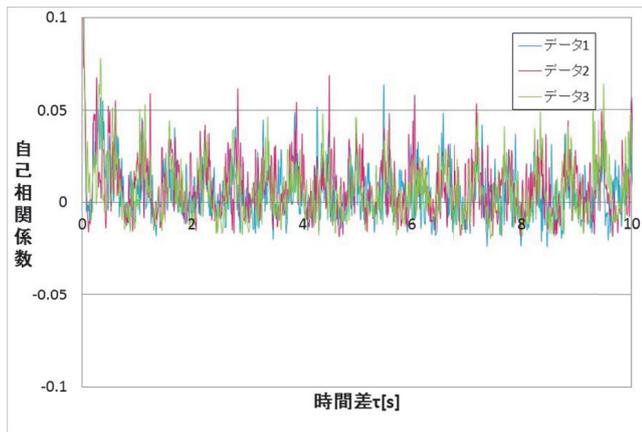


図 9 ネットラジオ聴取の自己相関解析結果

「仕事」行動と同様に、「休憩」行動においても共通の特徴は得られなかったが、アプリケーションごとには類いの傾向が見られるのがわかる。

4.6 評価手法

DP マッチングなど、参照データと比較することで結果を得る判別手法では、各行動の正しい参照データを用意することが重要である。しかしながら、必ずしもアプリケーション固有のトラヒックパターンが観測されるとは限らず、正しい参照データを定義することは難しい。

今回の評価では、DP マッチングによる判別において、参照データ数を増やした時に、正解率がどのように変化するかを見ることで、提案の有効性を確認する。

4.7 評価結果

今回、各行動の解析データを DP マッチングによって判別し、その正解率によって評価を行う。複数の参照データを用いる場合は、DP マッチングのスコアを平均し、平均後の値が最小スコアとなる行動を判別結果とする。

表 5 「仕事」の行動の正解率 [%]

参照データ数	電子メール	音声通話	ビデオ通話	ドキュメント編集
1	76.7	60.0	76.7	86.7
2	81.7	63.3	83.3	90.0
3	85.0	61.7	83.3	93.3
4	96.7	63.3	83.3	96.7
5	100.0	66.7	83.3	100.0
平均	88.0	63.0	82.0	93.3

表 6 「休憩」の行動の正解率 [%]

参照データ数	SNS サイト利用	音楽鑑賞	動画視聴	ネットラジオ聴取
1	86.7	100.0	96.7	66.7
2	85.0	100.0	100.0	78.3
3	83.3	100.0	100.0	75.0
4	83.3	100.0	100.0	83.3
5	83.3	100.0	100.0	83.3
平均	84.3	100.0	99.3	77.3

表 7 「仕事」と「休憩」の正解率 [%]

参照データ数	仕事	休憩
1	75.0	87.5
2	79.6	90.8
3	80.8	89.6
4	85.0	91.7
5	87.5	91.7
平均	81.6	90.3

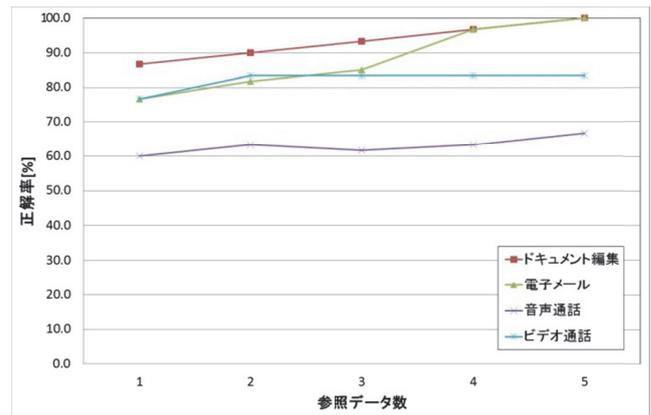


図 10 「仕事」の行動における参照データ数と正解率

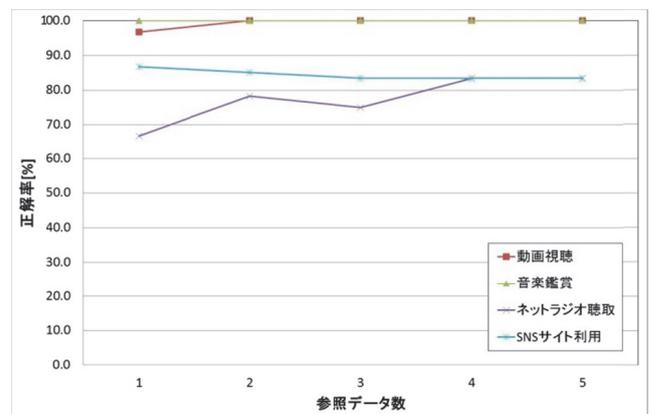


図 11 「休憩」の行動における参照データ数と正解率

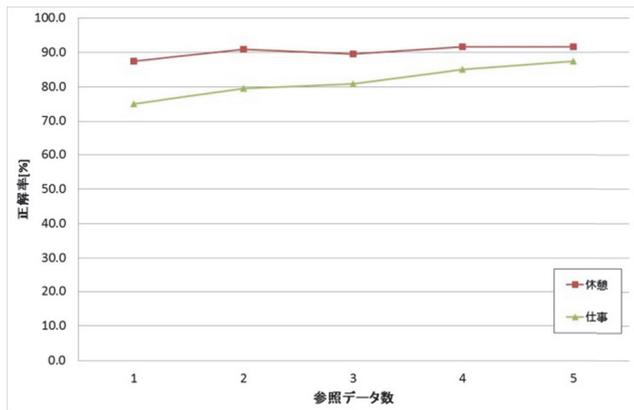


図 12 「仕事」と「休憩」の正解率

4.8 考察

今回の評価によって、「仕事」「休憩」のどちらにおいても、表 5-7 および図 10-12 のように、DP マッチングに用いる参照データを多くするほど正解率が上昇するという結果を得ることが出来た。

目的としていた、「仕事」「休憩」の判別では、それぞれ、81.6%、90.3%という高い判別正解率を得ることが出来たが、今回の評価では、利用したアプリケーションを正確に判別しなければ正解と認めていない。例えば、Skype による音声通話を入力としたとき、ビデオ通話などの他の「仕事」の行動に誤って判別された例が数多く見られた。このような場合には、この条件を緩和すれば正解率はさらに上昇すると考えられる。

5. おわりに

本稿では、コミュニケーションのすれ違いという問題を解決するためにユーザ状況の推定が必要であることを述べた。そして、ユーザがネットワーク端末を操作する際に発生するトラヒックの周期性に着目した、使用アプリケーション判別方法を提案した。その際、自己相関係数を用いた解析を行うことで、過去の自身のデータとの相関を取り、周期性を数値的に表した。マッチングには DP マッチングを用い、その有効性の評価を行った。

評価においては、利用する参照データの数を増やすことで、判別の正解率が増加するという結果を得た。この正解率は、アプリケーションを正しく判別した割合であり、「仕事」「休憩」というより大きな枠での分類をしたときに、より高い正解率を得ることが出来ると考えられる。

今後の課題としては、単一の行動だけではなく、複数行動を同時に行った場合にも適用できるのかを検証していく必要があると考えている。

参考文献

- 1) 八木清之介, 和泉勇治, 角田裕, 根本義章: ネットワークアプリケーション弁別のためのペイロード長の遷移パタンの評価方式に関する一検討, 電子情報通信学会技術研究報告, Vol.107, No.313, pp.1-6 (2007).
- 2) 山下剛史, 星健太郎, 市野将嗣, 小松尚久, 竹下恵, 辻野雅之: トラヒックパターンに着目したサービス分類に関する一検討, 電子情報通信学会技術研究報告, Vol.109, No.36, pp.73-78 (2009).
- 3) 内田誠一: DP マッチング概説 : 基本と様々な拡張, 電子情報通信学会技術研究報告. Vol.106, No.428, pp.31-36 (2006).
- 4) 山下洋一: パターン認識, オーム社, pp.98-115(2000).
- 5) WIRESHARK
<http://www.wireshark.org/>
- 6) MathWorks 社
<http://www.mathworks.co.jp/>