Vol.2015-IOT-30 No.3
2015/7/3

EthernetヘッダおよびIPヘッダのみを用いたWebサービスの推定

杉田 大知[1,a)]　伊藤 嘉浩[1,b)]

概要：本論文は，Ethernet ヘッダまたは IP ヘッダのみを用いた Web サービス推定方法を提案し，実験によりその有効性の評価を行うものである．本方式では，Ethernet ヘッダおよび IP ヘッダ内のフレーム長，パケット長の累積相対度数分布を利用するので，暗号化されたセキュアな通信でも推定が可能である．本論文では 5 つのグループに属する 18 のサービスに対して評価を行い．本方式の有効性を確認している．

キーワード：アプリケーション識別，トラヒック特性，トラヒック解析，Ethernet トラヒック，IP トラヒック，Web サービス

Web Service Estimation using only Ethernet frame header or IP packet header

Daichi Sugita[1,a)]　Yoshihiro Ito[1,b)]

Abstract: This paper proposes a Web service estimation method by using only Ethernet frame header or IP packet header and confirms effectiveness of the method by experiment. Within the frame header of packet header, the proposed method uses only information that is not encrypted, such as the length. Therefore, the method can be applied to secure communication. This paper does experiments for 18 Web services, and confirms the effectiveness of the proposed method.

Keywords: Application Identification, Traffic Characteristics, Traffic Analysis, Ethernet Traffic, IP Traffic, Web service

1. はじめに

インターネットの普及により，様々なサービスが Web サービスの形で提供されている．これらは，検索サービス，オンラインショッピングサービス，動画配信サービスなど多岐に渡り，これらは我々の生活にとって欠かせないものとなっている．

Web サービスの重要性が高まるにつれて，そのサービス品質の向上が必要となっている．Web サービスはネットワーク上のサービスであるので，その品質は，ネットワークにおける QoS(Quality of Service) とも密接な関係がある．したがって，Web サービスを向上するためには，ネットワークにおける適切な QoS 制御も必要となる．更に，対象となる Web サービスによって，要求される QoS は異なるので [1]，Web サービスごとに異なる QoS 制御を行う必要がある．そのためには，QoS 制御を行うルータなどのネットワーク機器において現在扱っているトラヒックがどのサービスのものかを判別できなければならない．

あるトラヒックがどのサービスのものかを推定するために，トラヒック内のパケットに含まれるデータをアプリケーション層まで解析することが考えられる．しかし，QoS 制御を行うネットワーク機器がアプリケーション層まで解析を行うと，CPU などの処理量の増加を招くことに

[1] 名古屋工業大学工学研究科創成シミュレーション工学専攻
Nagoya Institute of Technology, Graduate School of Engineering, Department of Scientific and Engineering Simulation
[a)] sugita@en.nitech.ac.jp
[b)] yoshi@nitech.ac.jp

© 2015 Information Processing Society of Japan

なり，必ずしも現実的ではない．他の方法として，IP アドレスや TCP のポート番号，FQDN などのドメイン名を利用するものがある．しかしながら，Google や Yahoo! などの複数のサービスを提供している Web サービスにおいては，ドメイン名や IP アドレス，ポート番号の情報だけではサービスまで判別することは難しい．さらに，IPsec を用いたセキュアな通信では TCP 以上も暗号化されてしまうため，利用できる情報は限定される．

本研究ではデータリンク層である Ethernet ヘッダとネットワーク層の IP ヘッダにおける情報の中で，特に IPsec などの暗号化による影響を受けないもののみを用いて Web サービスを推定する方法を提案する．そして，実験によりその有効性を確認する．

本論文の構成を以下に示す．第 2 章で，推定方式の提案を行う．第 3 章では，提案方式で用いるトラヒックの特徴抽出を行う．第 4 章では，評価実験を述べ，5 章では本論文のまとめと今後の課題について述べる．

2. 推定方式の提案

本研究における，推定方式を以下に述べる．本方式では，Ethernet ヘッダと IP ヘッダから得られるものを考え，Ethernet および IP パケットのフレーム長における累積相対度数分布を用いる．これらは，IPsec などの暗号化により隠蔽されないものである．

本方式では，各 Web サービスから抽出した累積相対度数分布と，推定を行うトラヒックの累積相対度数分布との相関係数を求め，この値から各 Web サービスの推定を行う．

3. 各 Web サービスの特徴抽出

提案方式で用いる特徴量を求めるため，本章では以下の実験を行う．この実験環境を**図 1** に示す．被験者は，キャンパスネットワークを介して，インターネット上の実 Web サービスを利用し，指定されたタスクを行う．そして，タスクの実行中に観測されたトラヒックから，Web サービスの特徴を抽出する．被験者は，Web クライアントとして Firefox[2] を使用する．なお，被験者は同時に一つの Web サービスしか利用しないものとする．実験者は，プロトコルアナライザである Wireshark[3] を用いてトラヒックの観測を行う．被験者は 20 代の男女 21 名である．

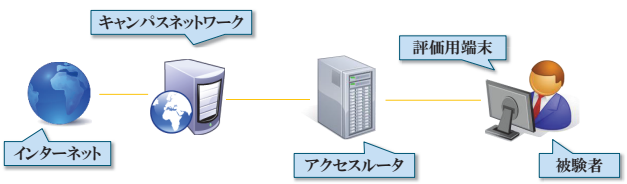

図 1 実験環境
Fig. 1 Experimental environment

本実験では，[4] より，日本国内でアクセス数の多い 15 個の Web サービスを対象とする．**表 1** に対象となる 15 個の Web サービスを示す．そして，これらの Web サービスを，その目的別にオンラインショッピングサービス (以下，shop)，動画配信サービス (video)，地図検索サービス (map)，インターネット百科事典サービス (wiki)，Web 検索サービス (search) の 5 つのグループに分類する．shop グループは，商品の検索や購入を行うことができるものである．video グループは，投稿されている動画の検索や視聴を行うための Web サービスであり，map グループは，地図の閲覧や各施設の検索を行うためのものである．wiki グループは，記載されている記事の閲覧や検索を提供する Web サービスである．また，search グループは，インターネット上で公開されている情報をキーワードなどを使用して検索するものである．

実験結果より得られた累積相対度数分布の例 (shop グループ) を**図 2**，**図 3**，**図 4** に示す．図 2 は，shop1 の Ethernet のフレーム長に対する累積相対度数分布を示している．図 3 と図 4 は，それぞれ，shop2 と shop3 の Ethernet のフレーム長に対する累積相対度数分布である．図中の横軸は，フレーム長であり，縦軸は，相対度数と累積相対度数である．

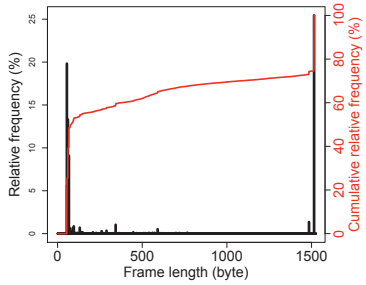

図 2 shop1 における累積相対度数分布
Fig. 2 Relative frequency of shop1

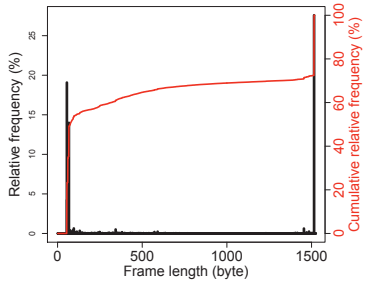

図 3 shop2 における累積相対度数分布
Fig. 3 Relative frequency of shop2

図 2〜図 4 を見ると，shop グループ内のサービスにおいては，フレーム長の分布が非常に類似していることがわかる．これらの 3 つのサービス間で相関係数を計算したところ，平均 0.98 となり高い相関を示した．また，shop グループとしての特徴量を求めるため，shop1，shop2，shop3 をまとめたトラヒックに対して累積相対度数分布を求めたも

© 2015 Information Processing Society of Japan

表 1　対象とする Web サービス
Table 1　Web services

グループ	Web サービス	略名	URL
オンラインショッピングサービス (shop)	Amazon	shop1	http://www.amazon.co.jp
	楽天市場	shop2	http://www.rakuten.co.jp
	Yahoo! ショッピング	shop3	http://shopping.yahoo.co.jp
動画配信サービス (video)	YouTube	video1	https://www.youtube.com
	ニコニコ動画	video2	http://www.nicovideo.jp
地図検索サービス (map)	Google マップ	map1	https://www.google.co.jp/maps/preview
	MapFun	map2	http://www.mapfan.com
	Yahoo! 地図	map3	http://map.yahoo.co.jp
	goo 地図	map4	http://map.goo.ne.jp
インターネット百科事典サービス (wiki)	Wikipedia	wiki1	http://ja.wikipedia.org/wiki/メインページ
	Weblio	wiki2	http://www.weblio.jp
Web 検索サービス (search)	Google	search1	https://www.google.co.jp
	Yahoo!	search2	http://www.yahoo.co.jp
	bing	search3	http://www.bing.com
	goo	search4	http://www.goo.ne.jp

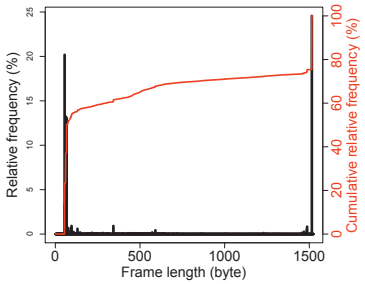

図 4　shop3 における累積相対度数分布
Fig. 4　Relative frequency of shop3

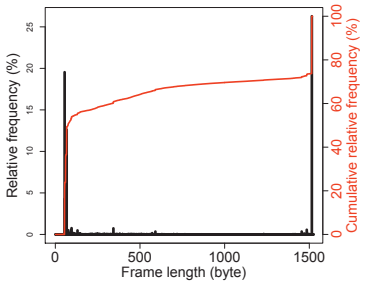

図 5　shop グループにおける累積相対度数分布
Fig. 5　Relative frequency of shop group

のを図 5 に示す．shop グループと同様に，他のグループ内のサービスに対しても相関係数をもとめたところ，同じグループ内の Web サービス同士では高い相関係数を示した．その一方で，違うグループ間 (例えば，shop1 と map1) では低い相関係数しか見られなかった．したがって，Ethernet フレーム長の累積相対度数分布を用いることで，対象とするトラヒックがどのグループに属するものであるかを推定することは可能であると考えられる．図 6，図 7，図 8，図 9 に他の 4 つのグループにおいてグループとしてまとめた累積相対度数分布を示す．図 6 は，video グループの累積相対度数分布を示している．図 7，図 8，図 9 は，それぞれ，map グループ，wiki グループ，search グループの累積相対度数分布である．

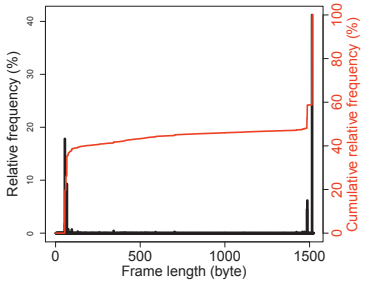

図 6　Video グループにおける累積相対度数分布
Fig. 6　Relative frequency of video group

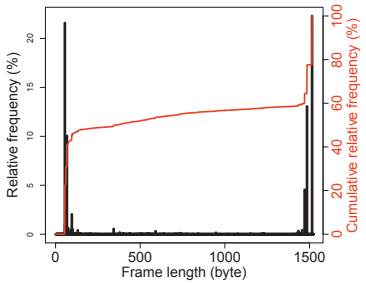

図 7　Map グループにおける累積相対度数分布
Fig. 7　Relative frequency of map group

図 5〜図 9 から，各グループを見ると，shop グループは 50〜55 バイト長のフレームよりも 1500〜1514 バイト長のフレームが占める割合の方が多い．また，video グループ

© 2015 Information Processing Society of Japan

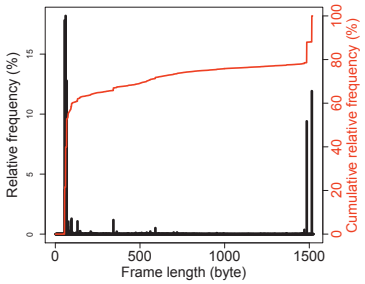

図 8　Wiki グループにおける累積相対度数分布
Fig. 8　Relative frequency of wiki group

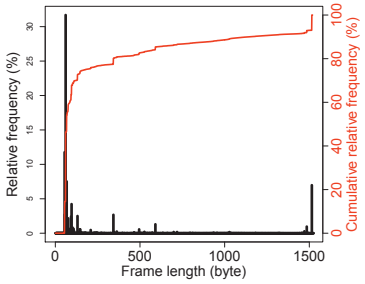

図 9　Search グループにおける累積相対度数分布
Fig. 9　Relative frequency of search group

は 1514 バイト長のフレームが占める割合が 50〜55 バイト長のものの 2 倍程度になっている．一方，map グループの場合は 50〜55 バイト長のフレームが占める割合と 1514 バイト長のものはほぼ同じである．wiki グループは 1514 バイト長のフレームが占める割合よりも 50〜55 バイト長のものの方が多く，search グループは 50〜55 バイト長のフレームが占める割合が 1514 バイト長のものの 3 倍以上となっている．これらのことから，5 つのグループの特徴は異なっていることがわかる．

4. 被験者実験

第 3 章で得られた特徴を用いて，本推定方式の有効性を実験により評価する．本実験では，第 3 章と同様に，実際の Web サービスに対して評価を行う．また，第 3 章で用いなかった Web サービスも評価対象として追加する．追加した Web サービスは，shop グループのもの (shop4[5])，map グループのもの (map5[6])，search グループのもの (search5[7]) の 3 つである．被験者は 20 代の男女 11 名である．なお，本研究では研究の第一段階として，各 Web サービスが単独で動作している状況を想定し，複数の Web サービスが同時に動作している場合の推定は議論しない．推定の方法は，5 つのグループと対象となるサービスの累積相対度数分布の相関係数を求め，最も相関が得られたものをそのサービスの推定グループとする．**表 2** に，本推定方式における推定結果を示す．

表 2 を見ると，IP ヘッダのみ用いた推定方法よりも

表 2　推定率
Table 2　Estimation results

グループ	shop	video	map	wiki	search
Ethernet ヘッダのみ	77.3%	90.9%	20.0%	54.6%	72.7%
IP ヘッダのみ	65.9%	90.9%	14.6%	45.5%	54.6%

Ethernet ヘッダのみを用いた場合の方が推定率が全体的に高くなっていることが分かる．また，Ethernet ヘッダのみを用いた推定方法では video1，map1，search1，search3 の 4 つの Web サービスの推定率は 100%であった．また，未知の Web サービスに関しても，shop4 の推定率が 90.9%であった．一方，map グループの推定率が低くなっているが，これは，map1 が HTTPS による通信を行っているためトラヒックにおけるフレーム長の分布に違いが生じて，map1 の推定率が下がったためである．したがって，HTTPS を用いる場合のトラヒックについては別途考慮する必要がある．

5. まとめ

本論文では，Ethernet ヘッダまたは IP ヘッダのみを用いた Web サービスの推定方法を提案し，実験により，その有効性の評価を行った．実験の結果から，Ethernet ヘッダのフレーム長における累積相対度数分布を使用することで，いくつかの Web サービスにおいては，これを推定することが可能であった．

今後の課題としては，本研究では 18 個の Web サービス，5 つのグループを対象としたが，他の Web サービスおよびグループに対しても調査する必要がある．また，本実験で観測したトラヒックは単一の Web サービスを実行した場合のものであるため，複数の Web サービスが混在するトラヒックにおける推定方式も調査していく．

参考文献

[1] D. Yamauchi and Y. Ito, “A study of effect of MPTCP on Web usability”, Proc.of IEEE GCCE, pp.12-15, Oct. 2014.
[2] “Firefox”, https://www.mozilla.org/ja/Firefox/new/.
[3] “Wireshark”, https://www.wireshark.org/download.html.
[4] “Top Sites in Japan - Alexa”, http://www.alexa.com/topsites/countries/JP.
[5] “Joshin”, http://joshinweb.jp/top.html.
[6] “Mapion”, http://www.mapion.co.jp.
[7] “Infoseek”, http://www.infoseek.co.jp.

© 2015 Information Processing Society of Japan