サービス接続時に登場する SNI の分布に関する一考察

浅岡 諒† 中尾 彰宏‡ 小口 正人†† 山口 実靖†

†工学院大学 〒163-8677 東京都新宿区西新宿 1-24-2
‡東京大学 〒113-0033 文京区本郷 7-3-1
††お茶の水女子大学 〒112-0012 文京区大塚 2-1-1

E-mail: †{cm22002@ns, sane@cc}.kogakuin.ac.jp, ‡nakao@nakao-lab.org, ††oguchi@is.ocha.ac.jp

あらまし TLS1.2 を用いる通信のサービスの同定手法として，登場する SNI に基づく手法が提案されている．当該手法は，接続対象サービスにより出現する SNI が異なることを前提とし，出現確率の偏りにより接続先サービスを推定，同定する．一方で，実サービスにおいて出現する SNI の偏りや，各 SNI のサービスごとの出現確率の偏りの詳細な調査などは行われていない．本稿では，実サービス接続時に出現する SNI 実例を示し，登場する SNI 情報の重複などについて考察する．

キーワード セキュリティ・プライバシ，Web 情報システム，情報抽出

1. はじめに

Web 上では，動画配信サービスや地図検索サービス，メールサービスなどの様々な Web サービスが提供されている．ユーザが接続するサービスをネットワークフローに基づき特定することは様々な分野への応用が期待できる．例えば，災害時においてメールサービスなどの安否確認を行えるサービスのフローを優先して流すような優先制御や，従量課金制の制度において特定のサービスへの接続に限り課金対象から除外をするゼロレーティングサービスへの応用が挙げられる．本稿では，Web サービスのサービス同定に着目する．

ネットワークフローの解析に基づくサービス同定手法として，IP アドレスやポート番号に基づく方法[1]が考えられるが，精度は十分ではない[2]．複数のサービスで同一の IP アドレスを用いる場合や，ポート番号が 80(HTTP)か 443(HTTPS)に限定される Web ブラウザを用いる場合には有効ではない．したがって，パケットのペイロードの解析に基づかない方法には精度に限界がある．

より高精度でサービス同定をするためにパケットのペイロードを解析する DPI(Deep Packet Inspection)[3]が注目されている．Web ブラウザを用いた通信の多くは TLS によって暗号化されているが，TLS1.2 においては SNI(Server Name Indication)などの一部のフィールドは暗号化されず，解析に用いることができる．DPI に基づいたサービス同定手法として SNI の出現に基づく手法[4]が提案されている．本手法では，接続対象サービスにより出現する SNI に偏りがあることを前提としており，この偏りにより接続サービスを推定する．しかし，本手法ではサービスへのアクセス時に出現した SNI を全て用いることが精度を低下させる主な要因となっている．そこで，同定に用いる SNI を選定することで精度を向上させる手法がいくつか提案されており，SNI の出現確率に基づく手法[5]，SNI の出現確率の標準偏差に基づく手法[6]，エントロピーに基づく手法[7]がある．しかし，これらの研究においても実際サービス接続時の登場 SNI の偏りや，各 SNI の登場サービスの偏りや類似性の調査や考察はされておらず，精度向上が生じる原因などは明らかなされていない．そこで，本稿では実サービスへの接続時に出現する SNI の実例を示し，登場する SNI 情報の重複などについて考察する．

2. 関連研究

2.1. SNI

Web ブラウザを用いた通信の多くは TLS によって暗号化されている．暗号化されたパケットのペイロードを解析することはできないが，TLS1.2 においては SNI などの一部のフィールドは暗号化されないため解析をすることができる．

TLS による暗号化通信を開始するには TLS ハンドシェイクが必要となる．TLS ハンドシェイクでは，通信内容を暗号化するための共通鍵を作成する．TLS1.2 における TLS ハンドシェイクのメッセージフローを図 1 に示す．TLS ハンドシェイクは，非暗号部および暗号部から構成されており，ClientHello から ServerHelloDone までは非暗号部となる．ClientHello は，TLS のバージョンを表す ClientVersion や TLS の拡張機能を表す Extension などのフィールドから構成されている．SNI は ClientHello の Extension に含まれている．

2.2. SNI の出現に基づくサービス同定手法

SNI を用いたサービス同定手法として SNI の出現に基づく手法[4]が提案されている．当該研究では，Web ブラウザを用いてサービスに 1 回アクセスすることを 1 アクセスと定義しており，本稿でもこの定義を用いる．本手法では，1 アクセスで出現した SNI に基づき

サービス同定を行う．1 アクセスで出現した SNI は図 2 に示すような登場 SNI ベクトルで表現される．登場 SNI ベクトルの値は，1 アクセスで SNI が 1 回以上出現した場合は 1, SNI が 1 回も出現しなかった場合は 0 で表現される．図 2 の例では，Web 検索サービスに接続をした結果，1 アクセスで a.com が 1 回，b.com が 2 回，c.com が 1 回出現したため，登場 SNI ベクトルは (1, 1, 1, 0)となる．

本手法は，調査フェーズと同定フェーズから構成される．

調査フェーズでは，接続サービスと登場 SNI ベクトルの関係を表したデータベースを作成する．具体的には以下の(1)～(3)の処理を行う．

(1) Web ブラウザを用いて同定候補のサービスにアクセスし，トラフィックを取得する．
(2) 取得したトラフィックの登場 SNI ベクトルを作成する．
(3) 接続サービスと登場 SNI ベクトルの関係を表すデータベース(登場 SNI ベクトル DB)を作成する．

同定フェーズでは，同定対象トラフィックの登場 SNI ベクトルと登場 SNI ベクトル DB の関係に基づきベイズ推定により同定結果を求める．具体的には以下の(1)～(4)の処理を行う．

(1) Web ブラウザを用いて同定対象のサービスにアクセスし，トラフィックを取得する．
(2) 同定対象トラフィックの登場 SNI ベクトルを作成する．
(3) 図 3 に示すように同定対象トラフィックの登場 SNI ベクトルと登場 SNI ベクトル DB 内の全ての登場 SNI ベクトルを比較する．比較では，登場 SNI ベクトルの全要素の値が一致(以下，完全一致)しているかを判定する．
(4) 比較した結果に基づき，式(1)に示すようなベイズ推定により各サービスの確率を求める．同定候補のうち確率が最も高いサービスを同定結果とする．式(1)において，A はサービス A に接続される事象，X は登場 SNI ベクトル X が作成される事象を表す．

$$P(A|X) = \frac{P(X|A)P(A)}{P(X)} \quad (1)$$

ただし，同定対象トラフィックの登場 SNI ベクトルが登場 SNI ベクトル DB 内のどの登場 SNI ベクトルとも完全一致しない場合，本手法は同定を放棄する．本手法において精度を低下させている主な要因は同定放棄である．ベイズ推定において全ての候補の確率が 0%になるため，同定結果を求めることができず放棄となる．

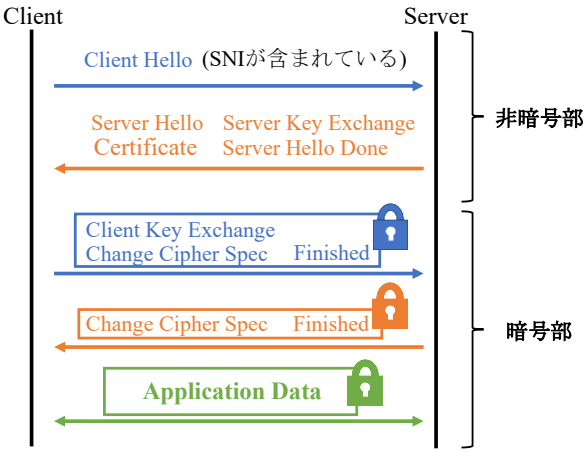

図 1 TLS1.2 における TLS ハンドシェイクのメッセージフロー

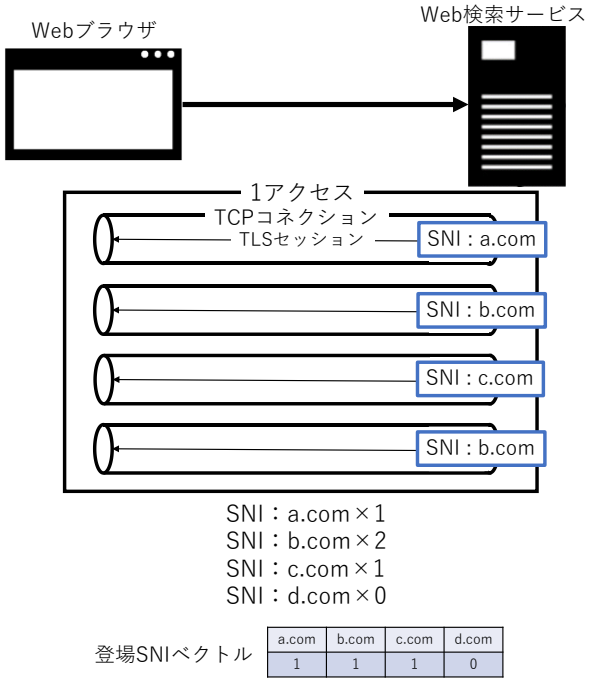

図 2 登場 SNI ベクトル

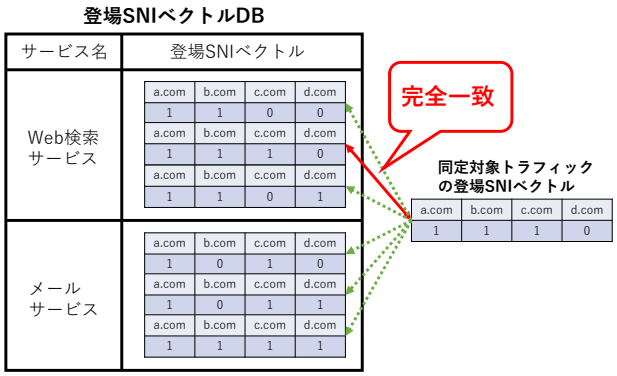

図 3 同定フェーズにおける登場 SNI ベクトルの比較

2.3. SNI 選定によるサービス同定の精度向上手法

SNI の出現に基づく手法[4]には，2.2 節で述べたように同定放棄が主な原因で精度が低下しているという課題がある．そのため，精度を向上させるには同定放棄に対処する必要がある．そこで，同定放棄発生時は同定をする上で重要ではないと予想される SNI を除外することにより同定放棄を解消し，精度を向上させる手法がいくつか提案されている．除外された SNI は同定に用いない．

SNI が 1 アクセスで少なくとも 1 回出現する確率を出現確率と定義する．例えば，100 回アクセスしたうち，ある SNI が 30 アクセスで出現した場合，その SNI の出現確率は 30%となる．SNI の出現確率に基づく手法[5]では，SNI の出現確率が低い順に同定放棄でなくなるまで SNI を除外する．出現確率が低い SNI は，同定対象トラフィックにおいても出現する確率が低く，同定をする上で重要ではないと考えられるため除外を行う．

SNI の出現確率の標準偏差に基づく手法[6]では，サービスごとにおける SNI の出現確率の偏りに着目し，SNI の出現確率の標準偏差が低い順に同定放棄でなくなるまで SNI を除外する．ある 1 つのサービスで必ず出現し他のサービスでは必ず出現しない SNI の例を考えると，本 SNI は接続サービスを絞り込むことができるため，同定をする上で重要であると考えられる．このように，サービスごとにおける SNI の出現確率に偏りがあると SNI の出現確率の標準偏差は高くなる．一方で，どのサービスにおいても出現確率が 80%となる SNI の例を考えると，本 SNI は接続サービスを絞り込むことができないため，同定をする上で重要ではないと考えられる. このように, サービスごとにおける SNI の出現確率に偏りがないと SNI の出現確率の標準偏差は低くなる．本手法では，サービスを絞り込む能力が低い SNI，すなわち SNI の出現確率の標準偏差が低い順に SNI を除外する．

エントロピーに基づく手法[7]では，SNI を同定に用いることで削減されるエントロピーの大きさが小さい順に同定放棄でなくなるまで SNI を除外する．接続サービスを事象系 X，ある SNI が出現するか否かを事象系 Y としたとき，ある SNI を同定に用いることで削減されるエントロピーの大きさは平均相互情報量 I(X;Y) に等しい．平均相互情報量は，片方の事象系に関する情報を得たときに，もう片方の事象系に与える影響の大きさを表している．この影響の大きさは，平均相互情報量の値が大きいほど大きくなる．平均相互情報量が大きい SNI は，接続サービスの同定に大きく影響を与える SNI であり，同定をする上で重要であると考えられる. 本手法では, 平均相互情報量が小さい順に SNI を除外する.

表 1 Yahoo10 サービスのサービス名および URL

サービス名	URL
Yahoo Account	https://login.yahoo.com/
Yahoo Advertising	https://advertising.yahoo.com/
Yahoo Celebrity	https://www.yahoo.com/entertainment/celebrity/
Yahoo Finance	https://finance.yahoo.com/
Yahoo Mail	https://mail.yahoo.com/
Yahoo News	https://www.yahoo.com/news
Yahoo Web Search	https://search.yahoo.com/
Yahoo SmartTV	https://smarttv.yahoo.com/
Yahoo Sports	https://sports.yahoo.com/
Yahoo Lifestyle	https://www.yahoo.com/lifestyle/style

表 2 MSN11 サービスのサービス名および URL

サービス名	URL
MSN Account	https://login.live.com/
Amazon	https://www.amazon.co.jp/
Facebook	https://www.facebook.com/
MSN Mail	https://outlook.live.com/
MSN Map	https://www.bing.com/maps/
MSN News	https://www.msn.com/ja-jp/news/
Rakuten	https://www.rakuten.co.jp/
MSN Web Search	https://www.bing.com/
MSN Skype	https://www.skype.com/
MSN Store	https://www.microsoft.com/
Twitter	https://twitter.com/

本節で述べた手法は，2.2 節で説明した SNI の出現に基づく手法[4]に比べて同定精度を大きく向上させることができたが，実サービスにて出現する SNI を考慮することで精度の向上の余地がある．

2.4. HTTPS トラフィックからのサービス同定

Shbair らは HTTPS のトラフィックからのサービス同定手法に関するサーベイ結果を報告している[8]。本サーベイでは，SNI を用いたプロトコルベースの方法などの, 様々な同定方法を紹介している. Bortolameotti らは，SNI と SSL 証明書情報を使用して悪意のある TLS 接続を検出する手法を提案している[9]。彼らは，著名な 100 の Web サイトについて，SNI のサーバ名文字列の構造や文字列の形式を調査し，報告している．Shbair らは, SNI に基づく HTTPS トラフィックのフィルタリングに注目し, HTTPS トラフィックの識別とフィルタリングにおける SNI の信頼性を評価している[10]. また, ファイアウォールを回避する方法を示して

表 3 Google15 サービスのサービス名および URL

サービス名	URL
Google Web Search	https://www.google.com/
Gmail	https://mail.google.com/
YouTube	https://www.youtube.com/
Google Map	https://www.google.com/maps/
Google News	https://news.google.com/
Google Drive	https://drive.google.com/
Google Photo	https://photos.google.com/
Google Calendar	https://calendar.google.com/
Google Sheets	https://docs.google.com/spreadsheets/
Google Document	https://docs.google.com/document/
Google Translate	https://translate.google.com/
Google Account	https://myaccount.google.com/
Google Play	https://play.google.com/
Google Plus	https://plus.google.com/
Google Scholar	https://scholar.google.com/

いる．ただし，これらの研究は単一のコネクションのみを対象としており，本稿の研究の様に複数のコネクションを用いた考察には到っていない．Kim らは，サービスによりネットワークトラフィックを分類する方法を提案している[11]．彼らは暗号化されていない TLS Handshake の証明書フィールドを使用し分類を実現している．当該研究は，本研究と同様に TLS の非暗号化部分を使用しており，セッション ID なども使用している．しかし，本研究とは異なり，SNI の登場ベクトルなどの複数のコネクションを考慮した情報は用いていない．また，本研究の評価のようなより複雑な状況でも高い精度が得られることは示していない．

3. 実サービスにおける SNI の調査結果

本章では，実サービスに接続した際に出現した SNI の実例を示し，SNI の出現の偏りと重複について考察する．調査対象のサービスは Yahoo の 10 サービス(表 1)，MSN の 11 サービス(表 2)，Google の 15 サービス(表 3)とした．Mozilla Firefox(ver.85.0.1)を用い，各サービスに 100 回アクセスし，トラフィックをキャプチャした．Yahoo および Google には 2021 年 1 月，MSN には 2021 年 2 月にアクセスした．TLS1.2 によって TLS セッションが確立されたため，SNI は暗号化されず解析可能であった．

3.1. 実サービスにおいて出現する SNI の偏り

SNI の出現に基づく手法[4]では，サービスごとにおける SNI の出現の偏りに基づき接続サービスを推定する．本節では，実サービスへの接続時に出現した SNI のサービスごとにおける偏りについて考察する．

実サービス接続時に出現する SNI のサービスごとにおける偏りが大きい，すなわち，サービスごとに出現確率に差がある SNI は多数存在する．調査対象のサービスのうち 1 つのサービスで出現確率が 100%となり，その他のサービスで出現確率が 0%となる SNI を one-hot SNI と定義する．例えば，”advertising.yahoo.com”という SNI は one-hot SNI であり，Yahoo Advertising にのみ出現し，かつ同サービスにおいて出現確率が 100%となっている．Yahoo では，Yahoo Account，Yahoo Celebrity, Yahoo News 以外の全てのサービスで one-hot SNI が存在している．MSN では，MSN Account 以外の全てのサービスで one-hot SNI が存在している．Google では，Google Document および Google Sheets 以外の全てのサービスで one-hot SNI が存在している．このように，one-hot SNI は多くのサービスで存在している．one-hot SNI は接続サービスを絞り込むことができる．そこで，one-hot SNI を考慮した手法[12]が提案されており，精度の向上が確認されている．one-hot SNI 以外にも出現確率に偏りが大きい SNI として，複数のサービスで出現確率が 100%となり，その他のサービスで出現確率が 0%となる SNI は Yahoo で 3 個，MSN で 5 個，Google で 13 個存在している．例えば，”www.yahoo.com”という SNI は，Yahoo Celebrity, Yahoo News，Yahoo Style でのみ出現し，かつ出現確率が 100%となっている．本 SNI も接続サービスを絞り込むことができるため，同定をする上で重要な SNI であると考えられる．

一方で，実サービス接続時に出現する SNI のサービスごとにおける偏りが小さい，すなわちどのサービスにおいても同程度の出現確率となる SNI が存在する．”s.yimg.com”は Yahoo の全サービスにおいて出現確率が 100%となる SNI である．全サービスにおいて出現確率が 100%となる SNI は本 SNI のみである．Yahoo のサービスで用いられる画像は全て”s.yimg.com”に保存されており，Yahoo のサービスにアクセスした際は必ず本 SNI が出現する．”safebrowsing.googleapis.com”は Yahoo，MSN，Google の全サービスにおいて出現する SNI であり，どのサービスにおいても出現確率が 80～90%程度となっている．Google セーフブラウジングは，ユーザが危険な Web サイトにアクセスしようとした際に警告を発

表 4 出現確率が一致する SNI(one-hot)の数

	one-hot SNI 数	one-hot サービス
Yahoo	25	advertising
	3	finance
	2	search
	2	sports
MSN	13	amazon
	2	map
	50	rakuten
	2	search
	4	skype
	13	store
	7	twitter
Google	3	account
	2	map
	3	photo
	2	play
	2	youtube

する，Web ブラウザの機能である．本機能が働く際に本 SNI が出現すると予想される．したがって，接続するサービスに依存せずに本 SNI が出現する．以上のように，どのサービスにおいても同程度の出現確率となる SNI は接続サービスを絞り込むことができず，同定をする上で重要ではないと予想されるため，除外をすべきであると考えられる．

3.2. 実サービスにおいて出現する SNI の重複

つぎに，各 SNI の登場の有無により得られる情報の冗長性について調査する．前述の様に，ある SNI の登場の有無がサービスにより異なる場合は，その SNI の登場の有無に関する情報は，サービスを特定する上で有益となる(サービスのエントロピーを削減する)．ただし，ある SNI のサービスごとの登場の有無と，別の SNI のサービスごとの登場の有無が完全に一致している場合は，片方の SNI の登場の有無の情報がもう片方の SNI の情報と重複しており，片方の SNI の情報が得られればもう片方の情報は有益ないことになる．本節では，ある SNI と別の SNI が与える情報の冗長性(重複性)について調査する．

まず，サービスごとの出現確率が完全に一致する SNI を調査した結果を示す．one-hot SNI が複数存在しているサービスにおける one-hot SNI の数を表 4 に示す．表より，出現確率が完全に一致する one-hot SNI が複数存在していることが分かる．これら SNI は冗長であり，1 つの SNI を除いて除外することが可能であると期待できる．具体的な SNI としては，Yahoo Advertising における ”advertising.yahoo.com” や”t.co”, ”www.oath.com”などが完全に出現確率が一致する SNI であった．

出現確率が完全に一致する n-hot SNI($1 < n <$ 全サービス数)は，存在していたが，表 4 と比較し数が非常に少なく，MSN で 2 個存在するのみであった．よって，重複の多くは one-hot SNI であり，あるサービスへのアクセス時にのみ接続される機能の SNI であると予想される．

続いて，サービスごとの出現確率が，完全に一致ではないが非常に近い SNI の調査結果を示す．Yahoo において，サービスごとの出現確率のベクトルのマンハッタン距離が 0%でなく 1%以下であり，celebrity, finance, news, sports, style における出現確率が約 100%である SNI が 12 個存在している．これらは，完全に冗長ではないが，情報量は非常に低いといえ，一つを残して他の SNI は除外することが好ましいと予想できる．同様に MSN では amazon と news にのみ約 100%で登場する SNI や，news と rakuten にのみ 100%登場する SNI が複数存在していた. Google に関しては，その様な SNI の実例は確認されなかった．

3.3. 実サービスにおいて出現する出現確率が低い SNI

つぎに，1 アクセスにのみ出現する SNI について調査する．1 アクセスにのみ出現する SNI が学習データでのみ出現し同定対象データにおいて出現しなかった場合，本 SNI が出現した登場 SNI ベクトルは同定対象トラフィックと完全一致しないため，同定放棄の原因となる．また，1 アクセスでのみ出現する SNI が学習データにおいて出現せず同定対象データでのみ出現した場合，同定対象トラフィックの登場 SNI ベクトルは登場 SNI ベクトル DB 内のどの登場 SNI ベクトルとも完全一致しないため，確実に同定放棄となる．本節では，同定放棄の要因となる 1 アクセスでのみ出現する SNI について調査する．

1 アクセスでのみ出現する SNI は，Yahoo では 724 個存在しており，”changelogs.ubuntu.com” や”bcn.yahoo.com”などの SNI が観測された. また, MSN では 20 個存在しており，”r.clicktale.net” や”geo.moatads.com”などが観測された．また，Google では 25 個存在しており，”maps.google.com”や” p5-75ovu3x2lu4vm-24epstplkz3fta6d-452153-i1-v6exp3.v4.metric.gstatic.com”などが観測された．

Yahoo Finance や Google Web Search などの一部のサービスでは自動生成されたと考えられる SNI が出現することが知られている[13]．自動生成された SNI は 1 アクセスでのみ出現し，別のアクセスで出現することはない．本調査においても実サービス接続時に自動生成された SNI が多く確認された．自動生成された SNI の例として，Yahoo においては” p4-drtkfln25jxri-hmhjjlgl2cirgxfo-548198-i1-v6exp3.ds.metric.gstatic.com” や ” p4-drtkfln25jxri-hmhjjlgl2cirgxfo-548198-i2-v6exp3.v4.metric.gstatic.com”が本調査にて観測されており，これらの SNI は 2 文字を除いて共通している．

Yahoo にて観測された自動生成 SNI は 4 パターンあり，合計で 703 個確認された．”p4-*.metric.gstatic.com”のパターンの SNI は 185 個存在する．”v-*.wc.yahoodns.net”のパターンの SNI は 223 個存在しており，具体例として”v-cz5mcpdlsm.wc.yahoodns.net”がある．”r-*.wc.yahoodns.net”のパターンの SNI は 220 個存在しており，具体例として”r-cz5mcpdlsmreport.wc.yahoodns.net”がある．”dns-.*.sombrero.yahoo.net”のパターンの SNI は 75 個存在しており，具体例として”dns-r767owhz6.sombrero.yahoo.net”がある．

MSN にて観測された自動生成 SNI は 1 パターンのみ存在している．”a*.cloudfront.net”のパターンの SNI は 7 個存在しており，具体例として”ab952ffa85625367a9b1d9504b23b3696.profile.syd1-c2.cloudfront.net”がある．

Google にて観測された自動生成 SNI は 1 パターンのみ存在している．”p5-*.metric.gstatic.com”のパターンの SNI は 24 個存在しており，具体例として”p5-75ovu3x2lu4vm-24epstplkz3fta6d-452153-i1-v6exp3.v4.metric.gstatic.com”がある．

以上のように，1 アクセスでのみ出現する SNI は多く存在している．前述の様に，本 SNI は同定をする上で役に立たないため，全て除外をすべきである．

4. シミュレーション結果

3 章にて実サービス接続時に出現した SNI の調査結果を示した．本章では，調査結果を踏まえ，Web サービスに依存しない，Web ブラウザの機能の SNI を除外することの有効性を示す．具体的には，Yahoo，MSN，Google の全サービスにて出現する SNI である”safebrowsing.googleapis.com”の除外を行った上で SNI の出現に基づく手法によりサービス同定を行う．

2.2 節で紹介した SNI の出現に基づく手法[4]および”safebrowsing.googleapis.com”の除外を行う手法でサービス同定を行い，同定精度の比較をする．同定精度は以下に示すスコアの平均で求められる．サービスを 1 つに絞り込み，正しい同定結果を得られたときのスコアを 1，誤って同定した場合のスコアを 0 とする．サービスを n 個に絞り込み，その中に正しい同定結果が含まれている場合のスコアを $1/n$，含まれていなければ誤答として 0 とする．同定放棄の場合のスコアはサービス数の逆数とする．サービス同定にて用いたアクセスデータは 3 章にて用いたものと同一である．100 アクセス分のデータのうち 90 アクセスを学習データ，残りの 10 アクセスをテストデータとして用いた．

Yahoo の 10 サービスにおける同定精度を図 4 に示す．図中の”既存手法(除外無し)”は SNI の出現に基づく手法[4]，”safebrowsing.googleapis.com を除外”は”safebrowsing.googleapis.com”の除外を行う手法の同定精度を表す．図より，Yahoo Account，Yahoo News，Yahoo Sports の 3 サービスで同定精度が向上していることがわかる．

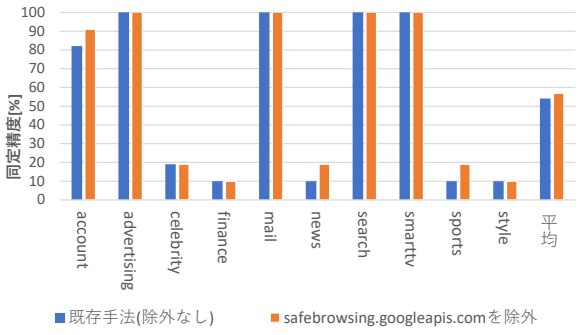

図 4 Yahoo10 サービスにおける同定精度

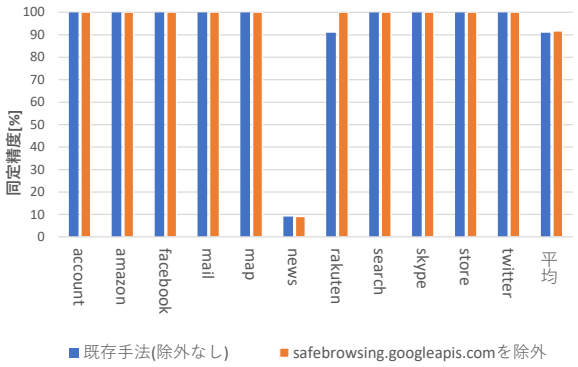

図 5 MSN11 サービスにおける同定精度

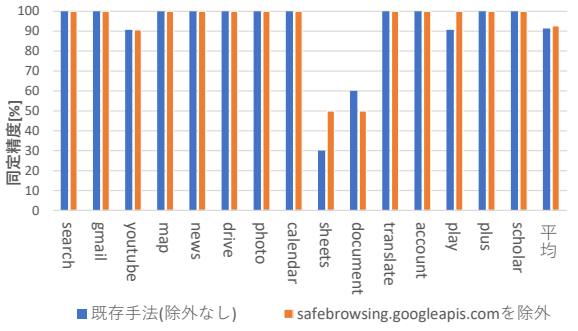

図 6 Google15 サービスにおける同定精度

次に，MSN の 11 サービスにおける同定精度を図 5 に示す．図より，Rakuten にて同定精度が向上していることがわかる．

最後に，Google の 15 サービスにおける同定精度を図 6 に示す．図より，”safebrowsing.googleapis.com”を除外することで平均の同定精度が 91.7%から 92.2%まで向上したことがわかる．また，Google Play および

Google Sheets にて同定精度が向上しており，Google Document で同定精度が低下していることがわかる．Google Sheets および Google Document ではそれぞれ出現する SNI が完全に一致しており，同定精度に限界がある．SNI の出現に基づく手法では，これらのサービスの同定を行うと「接続サービスを 1 つに絞り込んだ上で正答」あるいは「接続サービスを 1 つに絞り込んだ上で誤答」という同定結果が得られる．このような同定結果では同定精度が一定の値にならず不安定になるという問題が発生する．しかし，”safebrowsing.googleapis.com”の除外をし，サービス同定を行うと「接続サービスを Google Sheets と Google Document の 2 つに絞り込んだ上で正答」という同定結果が得られる．この場合，同定精度は 50%で一定となるため安定する．

以上のように，接続する Web サービスに依存せず，Web ブラウザの機能の SNI を除外することで同定精度が向上することがわかる．

5. おわりに

本稿では，TLS で暗号化されたフローから SNI に基づきサービスを同定する手法に着目し，現実のサービスの同定における各 SNI の登場の有無が与える情報量についての調査結果を示し SNI の価値についての考察を行った．その結果，サービスごとに出現確率に大きな差があり確率が 0%または 100%である SNI が複数あること，あるサービスに 100%登場し他のサービスに全く登場しない one-hot SNI が Yahoo，MSN，Google の多くのサービスに 1 個以上存在すること，サービスごとの出現確率が一致しており冗長な情報しか提供しない SNI が複数存在していることなどが分かった．

今後は，これらの現実のサービスへのアクセス時の SNI の登場の偏りを考慮した同定手法の改善ついて考察していく予定である．

謝　辞

本研究は JSPS 科研費 21K11854, 21K11874 の助成を受けたものである．

参　考　文　献

[1] Server Name and Transport Protocol Port Number Registry, http://www.iana.org/assignments/service-names-port-numbers/service-names-port-numbers.xhtml, Feb 8, 2022.

[2] 岩井貴充，中尾彰宏，“アプリケーション毎のトラフィック制御を目的とする N-gram を用いた網内機械学習によるモバイルアプリケーション同定手法”，信学技報, vol. 115, no. 209, NS2015-78, pp. 41-46, 2015 年 9 月．

[3] M. Hara, S. Nirasawa, A. Nakao, M. Oguchi, S. Yamamoto, and S. Yamaguchi, “Fast Application Identification Based on DPI N-gram,” 2016 IEEE 17th Int. Conf. on High Performance Switching and Routing Workshop Prog., June 2016.

[4] Ryo Asaoka, Yuto Soma, Hiroaki Yamauchi, Akihiro Nakao, Masato Oguchi, Saneyasu Yamaguchi, Aki Kobayashi, "Service Identification of TLS Flows Based on Handshake Analysis," IPSJ Journal of Information Processing (JIP), 2023, Volume 31, Jan. 2023.

[5] Ryo Asaoka, Akihiro Nakao, Masato Oguchi, Saneyasu Yamaguchi, "Accuracy Improvement by Occurrence Probability of Service Identification based on SNI," 2022 Ten International Symposium on Computing and Networking Workshops (CANDARW), 2022.

[6] 浅岡諒，中尾彰宏，小口正人，山口実靖，“SNI の出現確率の分散に基づく SNI ベースサービス同定の精度向上”，信学技報, vol. 122, no. 274, NS2022-114, pp. 79-84, 2022 年 11 月．

[7] 浅岡諒，中尾彰宏，小口正人，山口実靖，“エントロピーを考慮した SNI 選定に基づくサービス同定の精度向上手法”，信学技報，vol. 122, no. 310, NS2022-130, pp. 11-16, 2022 年 12 月．

[8] Wazen M. Shbair, Thibault Cholez, Jerome Francois, Isabelle Chrisment, “A Survey of HTTPS Traffic and Services Identification Approaches,” arXiv preprint arXiv:2008.08339, 2020. doi: 10.48550/arXiv.2008.08339

[9] R. Bortolameotti, A. Peter, M. H. Everts, and D. Bolzoni, “Indicators of malicious SSL connections,” in Network and System Security. Springer, 2015, pp. 162?175.

[10] W. M. Shbair, T. Cholez, A. Goichot, and I. Chrisment, “Efficiently bypassing SNI-based HTTPS filtering,” in Integrated Network Management (IM), 2015 IFIP/IEEE International Symposium on. IEEE, 2015, pp. 990?995.

[11] S.-M. Kim, Y.-H. Goo, M.-S. Kim, S.-G. Choi and M.-J. Choi, "A method for service identification of SSL/TLS encrypted traffic with the relation of session ID and Server IP," 2015 17th Asia-Pacific Network Operations and Management Symposium (APNOMS), 2015, pp. 487-490, doi: 10.1109/APNOMS.2015.7275373.

[12] 浅岡諒，中尾彰宏，小口正人，山口実靖，“単一のサービスにのみ出現する SNI に基づく SNI の選定によるサービス同定の精度向上手法”，信学技報, vol. 122, no. 362, NS2022-157, pp. 43-48, 2023 年 1 月．

[13] Y. Soma, A. Nakao, M. Oguchi, S. Yamamoto, S. Yamaguchi and A. Kobayashi, "Ocurring SNIs for Servicve Indetification," 2020 IEEE 9th Global Conference on Consumer Electronics (GCCE), Kobe, 2020, pp. 586-587, doi: 10.1109/GCCE50665.2020.9292038.