様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２８ 年　 ６ 月 １６ 日現在

機関番号：３２６８９
研究種目：研究活動スタート支援
研究期間：2013～2014
課題番号：２５８８００２０

研究課題名（和文）超高速ネットワーク詳細モニタリング技術の研究

研究課題名（英文）Fine-grained traffic monitoring for ultra high-speed networks

研究代表者
森　達哉（Mori, Tatsuya）

早稲田大学・理工学術院・准教授

研究者番号：６０７０８５５１

交付決定額（研究期間全体）：（直接経費）　　2,100,000円

研究成果の概要（和文）：本研究課題の遂行にあたり、特にアプリケーションとして有望であるテーマから研究を進めた。具体的には暗号化通信の通信先ホスト名を推定する問題に取り組みんだ。この問題を解決するために、ドメインネームグラフと呼ぶデータ構造とアルゴリズムを提案し、DNSの観測情報から暗号化された通信の宛先ホスト名を高精度に推定できることを実証した。結果を国際会議 TMA 2015 (採録率 29.6%)で発表、スケーラビリティに関する課題を克服した結果を 2015年度に Computer Comminications 誌にて発表。国内特許出願1件と同出願のPCT出願1件を実施。国内招待講演を1件実施。

研究成果の概要（英文）：We began with a research topic that focuses on the most motivating application of the proposed scheme. After several trials, it turned out that the requirements of the applications are fairly complex. Therefore, we decided to focus our attention on the methodologies to establish that application, rather than trying to generalize the application. The application was to infer the hostnames of encrypted traffic. To this end, we developed a new scheme called DNG (domain name graph), which captures complex dynamics of DNS name resolutions. Our main research outcomes are as follows: We first published a research paper at IFIP TMA 2015 (acceptance rate = 29.6%). We then published the extended version of the previous paper (we extended the methodology to drastically improve scalability) at Elsevier Computer Communications. We also filed two patents for domestic, and another two patents as PCT. We also presented an invited talk at a domestic conference.

研究分野：情報ネットワーク

キーワード：トラヒック　推定　DNS　暗号化　グラフ　スケーラビリティ

１．研究開始当初の背景

インターネットでは日々様々な障がいやセキュリティインシデントが発生するため、ネットワークのモニタリングが必要不可欠である。一方でバックボーンやデータセンターで広く普及しているイーサネットの高速化が顕著であり、転送レートが毎秒テラビットを越える技術の開発が進められている。そのような超高速ネットワークを詳細にモニタリングするためには非常に膨大な量のデータを高速に処理する技術が求められる。

２．研究の目的

上述の背景に鑑み、リアルタイムで詳細なトラヒック情報を獲得可能なインターネットトラヒックモニタリング技術を開発する。

３．研究の方法

高速なモニタリングに必要なデータ構造とアルゴリズム（確率的連想配列）の開発、および具体的なアプリケーションとして、暗号化通信を対象としたホスト名推定技術に取り組む。特に後者に関しては実データを用いた性能評価を行う。

４．研究成果

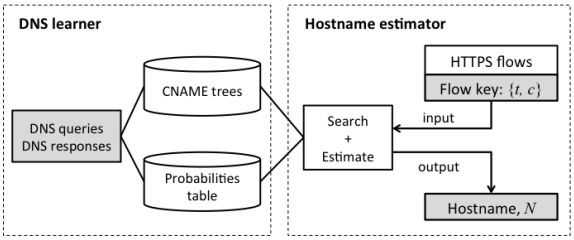
【研究の経緯】

データ構造とアルゴリズムを考案する上で応用からの要求条件を明確にするため、まずは応用から取り組んだ。この結果、応用の要求条件が当初想定した以上に複雑であるため、まずはそのような複雑な応用に対処可能なデータ構造とアルゴリズムの開発を中心に取り組む方針とした。この結果、当初想定していたような汎用的なデータ構造とアルゴリズムを開発するという路線を変更することになったが、元々のモチベーションとなっていた応用に関しては良好な成果を得ることができ、結果として超高速モニタリング。

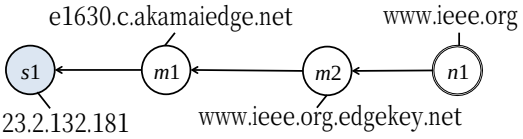
【具体的な成果】

超高速ネットワークモニタリングの具体的なアプリケーションとして、SSL/TLS によって暗号化されたウェブ通信におけるホスト名推定する技術に取り組む。通信の暗号化が行われると通信の宛先 サーバのホスト名が不明となるため、通信事業者は自網の回線がユーザにどのように利用されているかを把握することが困難である。一方、ウェブ通信の多くが暗号化されるようになってきており、通信事業者がネットワークモニタリングを行う上での障壁となっている。

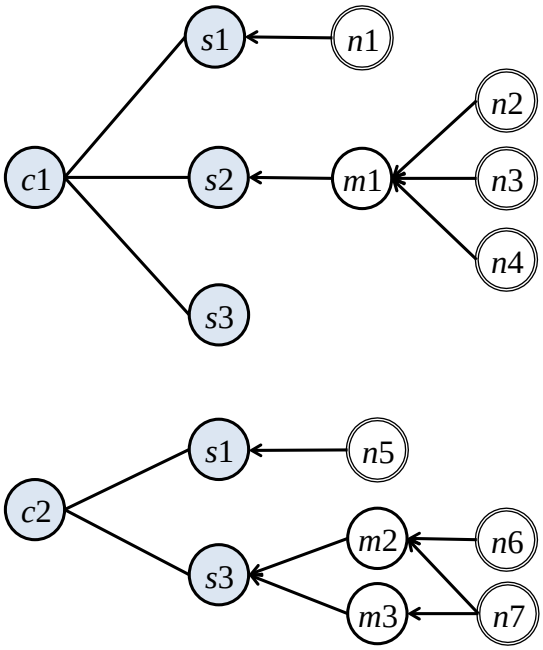
この問題を解決するために DNS クエリ・応答情報を用いて HTTPS 通信の ホスト名を推定するフレームワークである SFMap (Service-Flow map) を開発した。SFMap の主要なアイディアは CNAME を経由した IP アドレスと複数ドメイン名の関係をグラフとして保持し、そのグラフを元に所与のサーバ IP アドレス、クライアント IP アドレス情報から尤もらしいサーバのホスト名を推定することにある。下図に SFMap システムの全体図を示す。システムは DNS モニタリング情報をもとに DNG の学習を行う部分(DNS learner)と、観測した HTTPS フローをもとにウェブサーバのホスト名を推定する部分(Hostname estimator)から成る。



下図は DNS 名前解決の一例である。



ここで注意すべき点は名前解決には中間的なノードである CNAME が含まれる点である。すなわち、単純にサーバ名とホスト名の関係を保持するだけでは不十分である。CNAME も含めた名前解決情報を保持することにより、DNS レゾルバにおけるキャッシュが存在したとしても適切に名前解決をトレースすることが可能となる。DNG は CNAME を含んだ名前解決情報をグラフとして表現したものである。下図は DNG のイメージ図である。



この図において C はクライアント、S はサーバ、m は CNAME、n はサーバのホストネーム(FQDN)である。DNG の構築にあたっては(s,c,n)のタプルが出現した回数、および(s,n)のタプルが出現した回数もあわせて管

理する。
ホスト名の推定は HTTPS 通信を観測してクライアントとサーバの IP アドレスの組(c, s)が与えられた時、DNG とタプル出現回数の情報から尤もらしいホスト名 n を最尤推定する問題に帰着する。
本研究課題では 2 の観測ポイントで収集したデータを用い、この方式の精度を比較した。この結果が下表である。LAB, PROD は観測ポイントの名前である。

	LE	LE-NTE	UE	UE-NTE	DN-Hunter
LAB	54.98%	68.08%	71.59%	92.25%	67.90%
PROD	79.90%	88.29%	90.88%	90.88%	85.40%

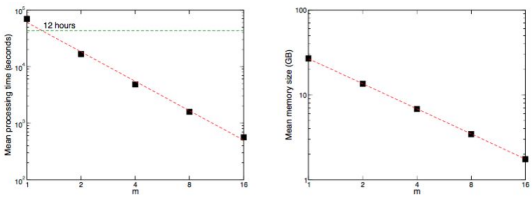
LE、LE-NTE, UE, UE-NTE は提案手法のバリエーションであるが、ここでは UE-NTE と名付けた DNG をクライアントの区別をせず一括に管理し、かつ TTL による expiration を無視した方式がもっとも精度が良かった。また、既存手法である DN-Hunter と比較しても推定精度が高いことが示された。
下表は exact match ではなく、public suffix までの比較にしたケースである。

	LE	LE-NTE	UE	UE-NTE	DN-Hunter
LAB	57.20%	70.30%	73.80%	94.46%	73.43%
PROD	83.20%	92.12%	94.52%	94.98%	89.98%

提案手法は 95%程度の精度であり、良好な推定結果を得ることができる。さらに推定結果の内、上位 3 位までの推定結果を許容する条件にしたときの評価が下表である。

	Exact matching			Public suffix		
	Hit in 1	Hit in 2	Hit in 3	Hit in 1	Hit in 2	Hit in 3
LAB	92.25	97.23	98.16	94.46	98.16	98.16
PROD	90.88	95.77	96.71	94.98	97.01	97.43

上位 3 までを許容すると推定精度は約 98%近くにまで上昇することがわかる。
さらに超高速トラヒックに対応するためにスケーラビリティを向上させるアルゴリズムの拡張を行った。具体的には DNG のパスが長大になる場合、次数が高いノードのリンクランダムサンプリングし、また DNG が巨大になりすぎないように適当なサイズでグラフを分割統治する処理を導入した。この結果、グラフ分割数を増やすことによって必要な処理時間とメモリ量を大幅に減じることができること（下図）



および分割数を増やしても精度への影響は大きくないことを明らかにした（下表）

	Exact matching			Public suffix		
m	Hit in 1	Hit in 2	Hit in 3	Hit in 1	Hit in 2	Hit in 3
1	0.815	0.886	0.909	0.904	0.936	0.944
2	0.809	0.879	0.902	0.897	0.929	0.937
4	0.806	0.876	0.899	0.894	0.925	0.933
8	0.805	0.875	0.897	0.892	0.923	0.931
16	0.803	0.873	0.895	0.889	0.921	0.929

以上の結果より、提案手法である SFMap は暗号化されたウェブ通信に対し、DNS 情報から DNG を作成し、出現頻度情報から高精度にホスト名を最尤推定できること、およびサンプリングとグラフ分割によって精度を保ちつつもスケーラビリティを大幅に向上させられることが明らかになった。

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計　1　件）

①T. Mori, T. Inoue, A. Shimoda, K. Sato, K. Ishibashi, and S. Goto, “Statistical Estimation of the Names of HTTPS Servers with Domain Name Graphs,” Computer Communications, vol. xx, issue xx, xx 2016, Pages xxxx-xxxx (査読有・出版中)

〔学会発表〕（計　4　件）

①森達哉, 井上武, 下田晃弘, 佐藤一道, 原田薫明, 石橋圭介, 芳賀夢久, 笹生憲, 後藤滋樹,
”名前情報による隠されたトラヒックの顕現化,”
信学技報, vol. 115, no. 370, IN2015-74, pp. 19-24, 2015 年 12 月, 広島県・広島市　(査読なし・招待講演)

②T. Mori, T. Inoue, A. Shimoda, K. Sato, K. Ishibashi, and S. Goto,
“SFMap: Inferring Services over Encrypted Web Flows using Dynamical Domain Name Graphs,”
Proceedings of IFIP Traffic Monitoring and Analysis workshop (TMA 2015), LNCS 9053, pp. 126-139, Apr. 2015.スペイン・バルセロナ (査読有・採択率: 29.6%=16/54)

③森達哉, 井上武, 下田晃弘, 佐藤一道, 石橋圭介, 後藤滋樹,
”DNS クエリ・応答を用いた HTTPS 通信のホスト名推定,”
信学技報, vol. 114, no. 478, IN2014-164, pp. 255-260, 2015 年 3 月　沖縄県・宜野湾市（査読なし）

④ T. Mori and T. Inoue, “Inferring Services over Encrypted Web Flows“, 電子情報通信学会総合大会 2014 年 3 月　新潟県・新潟市（査読なし）

〔図書〕（計　0　件）

〔産業財産権〕
○出願状況（計　2　件）

①②名称：名前特定装置、名前特定方法、及びプログラム
発明者：森達哉、井上武、他
権利者：同上
種類：国内特許出願
番号：特願 2015-028743
出願年月日：2015 年 2 月
国内外の別： 国内、国外(PCT 出願)

○取得状況（計　0　件）

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕
ホームページ等
http://nsl.cs.waseda.ac.jp/projects/sfmap/

６．研究組織
(1)研究代表者
森　達哉　　（MORI, Tatsuya）
早稲田大学・基幹理工学部・准教授
研究者番号：　60708551

(2)研究分担者
（　　）

研究者番号：

(3)連携研究者
（　　）

研究者番号：