2ZC-08

トラフィックデータを用いた機械学習による ネットワーク状態推定手法に関する研究

吉田蓮† 松下葵† 和泉諭† 髙橋晶子†
仙台高等専門学校†

1. はじめに

サイバー攻撃などのネットワーク犯罪の増加に伴い，ネットワーク管理者によるネットワークに接続されている機器の構成管理や状態監視が重要となっている．しかし，スマートフォンなどのモバイル機器の普及や，IoT 技術の登場によりネットワークに接続される機器の数は急速に増加している[1]． また，ネットワーク利用者の通信のプライバシへの関心が高くなっており，暗号化された通信が急激に増加している．このことからポートベース方式やペイロードベース方式など従来手法によるトラフィック解析は困難になっている．

本研究ではネットワークに接続されている機器やその状態の把握を自動化することでネットワーク管理者の負担の軽減を目指す．本稿では実際に収集したトラフィックデータから統計的特徴を抽出し、それを特徴量とした機械学習（ニューラルネットワーク）を用いたモデルを作成し，ネットワークで利用されているサービス及び、機器の推定を行う．

2. 関連研究と提案

ネットワークに接続している端末や利用者の推定・把握に関する関連研究や既存技術として，主にユーザ毎の認証 ID や機器毎の MAC アドレスなどで管理が行われている．IoT 環境の普及によりネットワークに接続している端末数やその利用者が増加していることから，これら個々の ID や MAC アドレスを管理することは管理者の負担が増大する．また仮想化技術や MAC アドレスのランダム化などの発展により， MAC アドレスによる端末認証も限界がある．

そこで，ネットワーク上のトラフィックを解析することで，ネットワーク状態を推定する手法がある．従来のトラフィック解析では主に，ポートベース方式が用いられており，サービス毎に使用されるポート番号によってサービスの種類を推定している．また，他にもペイロードベースによる解析も行われている．この手法ではパケットに含まれるペイロード部の中の文字列から，特定の文字列を抽出・照合しトラフィックデータの解析を行う．

ネットワーク状態の推定に関する課題として，従来手法を用いたネットワークトラフィックの解析や推定が困難なことが挙げられる．その要因として，複雑で動的なプロトコルの登場や，動的ポートやトンネリングなどの技術の発展などが挙げられる．これにより，これまで主流であった IP アドレスや MAC アドレス，ポート番号を解析して端末や利用者，サービスの推定は困難である．

また，利用者のプライバシ保護の観点から，暗号化されたセキュアな通信の利用が増加している．これにより通信の内容を参照することが不可能になるため，特定の文字列を照合するペイロードベースの解析を行うことも困難である．合わせて，ネットワークに接続されている端末が増加していることから，ネットワークトラフィック自体も増加傾向にあることから，ペイロードベースの解析は大量のディスク容量や高い処理能力が必要となる．

そこで，本研究では図 1 に示す概要に従い，トラフィック情報の統計的データを特徴量として，ニューラルネットワークを用いたネットワーク状態推定手法を提案する．推定には偽装や秘匿が困難である統計的特徴であるパケット長などの統計的データを用いて，ネットワークに接続されている端末，利用者，サービスなどを推定する．これにより，ネットワークの障害の

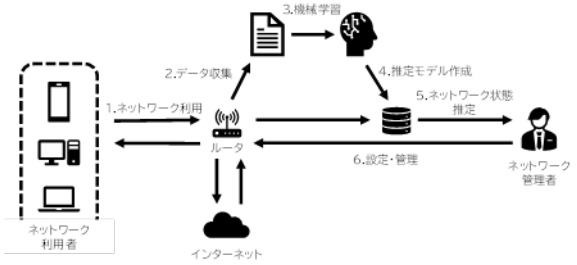

図 1 ネットワーク状態推定手法の概要

Network State Estimation Method by Machine Learning using Traffic Data
†Ren Yoshida, Aoi Matsushita, Satoru Izumi, Akiko Takahashi・National Institute of Technology, Sendai College

検出や利用者や端末のネットワーク利用状況などの把握を容易にし，ネットワーク管理者への支援を実現する．

3. 提案手法の設計

本研究ではネットワーク状態を推定するために，以下の手順でトラフィックデータの解析を行う．

Step1. トラフィックデータの収集
ネットワークを構築して様々な状況における通信トラフィックを観測しログデータとして蓄積する．

Step2. 特徴量の抽出
トラフィックデータに含まれる様々な特徴量から，通信の内容を性格づけるような統計的特徴を抽出する．

Step3. 機械学習によるデータ解析
抽出によって得られたデータを基に機械学習によって通信機器の推定を行う．

本実験では学習に用いる特徴量として，パケット長，パケットの到着する間隔を表すパケット送受信間隔，単位時間あたりに転送されるパケット長を表すパケットレート，送信/受信の 4 つを用いた．

4. 実験・評価

手法の性能を評価するために，実際に研究室内で専用のネットワークを構築してトラフィックデータを収集し，実験を行った．今回はトラフィックデータから利用しているサービスと機器を推定する実験を行った．サービス推定実験に用いるデータとして動画のストリーミング再生，E メールの送信，クラウド上での文書編集，ping 送信，ビデオ会議の計 5 種類のサービスにおけるトラフィックデータを収集した．また，機器推定実験に用いるデータとして，iPhone，iPad，laptopPC，desktopPC の計 4 種類の機器におけるトラフィックデータを収集した．

ニューラルネットワークによるアプリケーションの分類を行うため，学習モデルを作成して評価を行った．学習アルゴリズムとしてトラフィックデータを時系列データとして扱うことが可能である RNN[2]，GRU，LSTM 及び，局所的な特徴を抽出することが可能である CNN[3]を用いた．データセットとして実際に研究室で収集されたデータセットを用いて，各データ最大 5 万件のパケットを検証に使用した．RNN，GRU 及び LSTM の時系列シーケンスとして，トラフィックデータを時系列データとみなし，データを 10 個ずつ切り分けて入力とした．CNN でも同様にデータを 10 個ずつ入力とし，畳み込み層，プーリング層，全結合層からなるモデルを作成した．データは 0.8 の割合で学習データ，0.2 の割合でテストデータとした．それぞれの実験で 100 エポックずつ学習を行い，評価指標としてはバランスよくモデルの評価を行うことが可能である F 値を用いた．

表 2　サービス推定に対する F 値

	Streaming	edit	ping	video	mail
RNN	0.9692	0.9953	**1.0000**	0.9696	0.8817
GRU	0.9805	**0.9984**	0.9976	0.9815	0.9255
LSTM	0.9682	0.9963	0.9930	0.9750	0.8900
CNN	**0.9811**	0.9980	0.9876	**0.9854**	**0.9312**

表 2　機器推定に対する F 値

	iPhone	iPad	laptop	desktop
RNN	0.8803	0.9293	0.9172	0.8759
GRU	**0.9062**	0.9362	**0.9255**	0.8619
LSTM	0.8902	**0.9430**	0.9146	**0.8961**
CNN	0.7764	0.8636	0.8472	0.7542

サービス推定実験及び，機器推定実験の結果を表 1 と表 2 に示す．サービス推定実験では，推定するサービスの中でも特に，クラウド上での文書編集と ping の送信が高い精度で推定することができた．一方，E メールの送信は，全体的に精度が低かった．これは E メール送信の特徴量が，動画のストリーミング再生と類似していたためだと考えられる．学習アルゴリズムの中では CNN を用いた推定の評価が高かった．機器推定実験では，サービス推定よりも 1 割ほど精度が低下した．また，推定に用いた機械学習アルゴリズムの中で LSTM や GRU などのアルゴリズムの方が CNN と比べて評価が高かったことから，機器推定においてトラフィックの時系列データはサービス推定の際よりも重要性が高いと考えられる．

5. おわりに

本研究では，ネットワークの状態の把握を容易にすることでネットワーク管理者の負担を軽減することを目的とし，ニューラルネットワークを用いたネットワーク状態の推定手法を提案した．実験結果としてサービス推定のタスクでは CNN が，機器推定のタスクでは RNN を用いた手法が有効であることが確認できた．

【参考文献】

[1] 総務省:IoT デバイスの急速な普及．令和 3 年度版情報通信白書
https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r03/html/nd105220.html（参照 2022-07-09）

[2] Elman J. L.: Finding Structure in Time, Cognitive Science, Vol.14, No.2, pp.179-211 (1990)

[3] Zheng Wu et al.: Online Multimedia Traffic Classification From the QoS Perspective Using Deep Learning, Computer Networks, Vol. 204, No. 26 (2022)