IoT 機器に対する効率的な広域ネットワークスキャンを実現するための機器推定用データ作成手法

大崎光洋[†1]，種茂文之[†1]，和氣弘明[†1]，石岡裕[†1]，松下一仁[†1]

概要：
近年著しく数が増加している IoT 機器のセキュリティ設定問題を調査する広域ネットワークスキャンシステムにおいては，ネットワークに悪影響を与えない程度の必要最小限の通信量で，十分な調査が実施できる必要がある．このような効率的なスキャンシステムを実現するためには，スキャン結果から IoT 機器を機器推定することにより，対象ポートを限定してスキャンを行える必要があるが，これによりスキャンによる通信量を最小化するためには，IoT 機器推定の精度を可能な限り高める必要がある．
本研究は，IoT 機器推定によってスキャンの対象ポートを限定する方法や，スキャン時の機器推定の方法について整理した上で，大量に取得したバナー情報から機器推定するために必要なデータを効率的に作成する方法を確立することを目的としている．本稿では，実際のスキャンデータを用いて機器推定のためのデータを作成した結果を紹介し，作成したデータを用いてシミュレーションにより機器推定を実施する．

キーワード：IoT，広域ネットワークスキャン，機器推定

Data Creation Method for Device Estimation to realize efficient Wide-Area Network Scan for IoT devices

Mitsuhiro Osaki[†1], Fumiyuki Tanemo[†1], Hiroaki Waki[†1], Yutaka Ishioka[†1], Kazuhito Matsushita[†1]

Abstract: In recent years, the number of IoT devices has increased significantly. The wide-area network scan system for investigating the security setting problem of these IoT devices needs to enable sufficient investigation with the minimum amount of communication without adversely affecting the network.
Such an efficient scan system makes it possible to narrow down the target ports by estimating the IoT device from the scan results. In order to minimize the amount of communication traffic, it is necessary to improve the accuracy of IoT device estimation.
In this research, after organizing the method of limiting the target port of scan by IoT device estimation and the method of device estimation at the time of scanning, we aim to establish a method to efficiently create data required for device estimation from a large amount of banner information acquired. Also, we present the result of creating data for device estimation using actual scan data, and perform device estimation by simulation using the created data.

Keywords: IoT devices, Wide-area Network Scan, Device Estimation

1. はじめに

近年, IoT 機器の数は著しく増加しており，それに伴い，不十分なセキュリティ設定等の脆弱性を突いたマルウェア感染等により，DDoS 攻撃のようなサイバー攻撃に悪用される IoT 機器も増加している[1]．このようなサイバー攻撃を防ぐためにも，国内の IoT 機器に対して網羅的にポートスキャンを実施してセキュリティ設定を調査する広域ネットワークスキャンの実施が不可欠な状況となってきている．

IoT 機器を対象とした広域ネットワークスキャンについては数多くの研究がなされている一方で[2]，数多くの IoT 機器を調査対象とするスキャンについては，それに係る通信量も膨大になる恐れがあり，特にワイヤレスの通信エリアでは通常の通信サービスに悪影響を及ぼすことも懸念される．そのため，IoT 機器のセキュリティ設定についての十分な調査を可能にしつつ，必要最小限の通信量でスキャンを行えるような広域ネットワークスキャンシステムの実現が急務である．

このような広域ネットワークスキャンシステムを実現するためには，特にスキャンによって見つかった IoT 機器の機種やファームウェアのバージョンなどを可能な限り推定して，その機種・ファームウェアのバージョンに応じて対象ポートを絞ったスキャンを実施することが考えられる．この対象ポートの絞り込みにより，通信量を最小化しながら当該 IoT 機器のセキュリティ設定調査に必要なスキャンを実施することが可能となる．

このようなスキャンの効果を最大化するためには, スキャン状況から IoT 機器推定を行う精度を可能な限り高めるこ

†1 エヌ・ティ・ティ・アドバンステクノロジ株式会社
NTT Advanced Technology Corporation

© 2019 Information Processing Society of Japan

とが重要である．

機器推定のための従来技術としては，nmap[3] 等の著名なスキャンツールが実装する OS・アプリケーション等の推定機能や，SHODAN[4] や Censys[5] 等のスキャン結果を提供するオンラインサービスが実装している機器推定機能がある[6][7].

多くのスキャンツールは，スキャン結果から得られるバナー情報に基づいて IoT 機器等の機器推定を行うが，バナー情報から得られる特徴的な文字列から機器の推定を自動化する研究や[6]，実機やファームウェアから推定情報を生成する研究[8]，特定機器の推定までには至らないが，機器のウェブ管理画面の画像に基づいたクラスタリングによって機器をカテゴライズする研究[9]などが行われている．

本稿では，我々が提案する通信量を最小化する広域ネットワークスキャンシステムと当該システムにおける機器推定機能の概要について説明するとともに，日本国内の IP アドレスへの大規模な広域ネットワークスキャンにより大量に得られたスキャン結果中のバナー情報から，当該システムで利用可能な機器推定用のデータベースを効率的に作成するための方法について説明する．また，実際に行ったスキャン結果から抽出した機器推定用データを紹介し，そのデータを使った場合の機器推定の精度についてシミュレーション結果等から推測する．

本稿の構成は次の通りである．次章にて，我々が構築している広域ネットワークスキャンシステムとその構成について説明する．第３章では，スキャンシステム内の機器推定の仕組みと機器推定データベースの構成を示す．第４章では，スキャン結果から得られた大量のバナー情報から機器推定データベースの情報を効率的に作成する方法について説明する．第５章では，実際のスキャン結果から作成した機器推定用データについて紹介し，シミュレーション結果から得られた機器推定の精度等について説明する．

2. 広域ネットワークスキャンシステム

本章では，本研究における広域ネットワークスキャンシステムの動作概要と，スキャン効率化の考え方，システムの構成とその動作内容について説明する．

2.1 広域ネットワークスキャンの動作概要

本研究では，IoT 機器のセキュリティ設定や脆弱性を確認するための前段として，機器の存在有無と応答確認をポートスキャンによって定期的に実施するための広域ネットワークスキャンシステムを構築する．

本システムは，対象とする広域な IP アドレスの範囲を事前に設定し，設定した IP アドレスへのポートスキャンを定期的に実施して（半日に 1 回等），スキャン結果を蓄積する．そのスキャン結果を分析することにより，各 IoT 機器のオープンポートの状態等を確認することができる．

2.2 スキャン効率化の考え方

本システムは，広域な IP アドレス範囲を定期的にポートスキャンするため，大量のスキャンパケットを送る必要があり，それに係る通信量が膨大になる恐れがある．本システムでは，スキャンで見つかった IoT 機器の機種やファームウェアのバージョン等を可能な限り推定して，その機種・バージョンに応じて対象ポートを絞ったスキャンを実施することで効率化を行い，スキャンに係る通信量を削減することを目指している．

初回のスキャンの実施においては，機器推定情報が利用できないため，標準的な対象ポート集合を利用して一般的なスキャンを行うが，このスキャンの結果から機器の機種・バージョン等が推定できた場合には，次回のスキャンから，その機種・バージョン毎に絞り込んだスキャンを実施する．

機種・バージョンに応じた対象ポートとしては，(1)当該機器が機能提供等の目的で公開している可能性の高いポートや，(2)当該機器を対象としたマルウェア・攻撃ツールなどが作成・待ち受け等するポートなどが考えられるが，いずれにせよ通常のスキャンが対象とする標準的な対象ポート集合から，より少ないポート集合とすることで，スキャンに係る通信量を最小化しながら，当該 IoT 機器のセキュリティ設定調査に必要なスキャンを実施することが可能となる．

実施したスキャン結果が，前回のスキャン結果から大きく乖離している場合には，機器等の条件が変化したものとして，再度，標準的なスキャンから実施していく．

2.3 システムの構成

本研究で実現する広域ネットワークスキャンシステムは，図 1 に示すように４つの構成要素から構成される．

- 機器特性情報データベース

スキャンの対象とする IP アドレスのリストを格納すると共に，各 IP アドレス毎に，直近のスキャン結果や，スキャン対象ポート集合，過去のスキャン結果から判明した機器推定情報などを蓄積する．機器特性情報データベースは，毎回のスキャン結果や機器推定状況等に基づいて随時，内容が変わっていく．

- スキャン実行機能

スキャン実行機能は，機器特性情報データベースに格納されているスキャン対象の IP アドレスに対して，スキャンパケットを送信して，その応答を受信する．受信情報から得られたバナー情報，オープンポート情報を機器特性情報データベースに格納する．

- 機能推定機能

機器推定機能は，機器特性情報データベース内から取得したバナー情報，オープンポート情報を基に，機器推定データベース内の情報を用いて機器推定を実施し，機器推定結果を機器特性情報データベースに格納する．

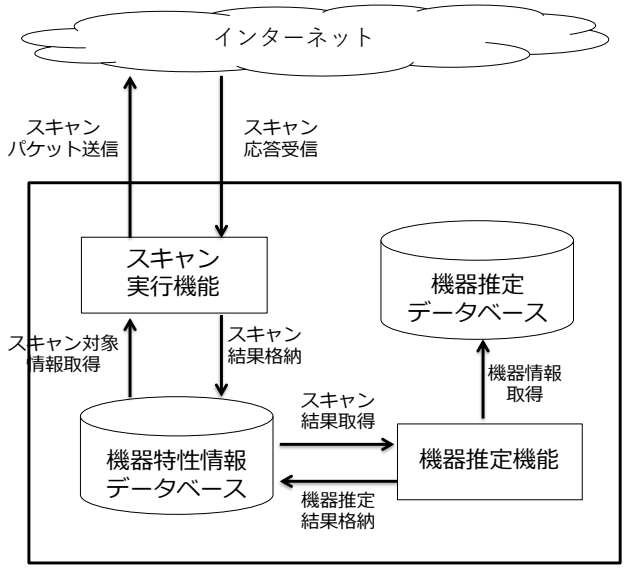

図 1　スキャンシステムの構成

- 機能推定データベース

機器推定データベースは，機器毎に固有のキーワード情報，オープンポート情報等が蓄積されたデータベースであり，機器推定機能がスキャン結果とこのデータベースの情報を照合することで機器の推定を行う．

本システムによるスキャン開始時には，機器推定データベースを事前に準備するともに，機器特性情報データベースにスキャン対象 IP アドレスのリストを登録する．対象 IP アドレスに対して，スキャン実行機能が一連のスキャンを実施し，その結果が機器特性情報データベースに格納される．また，スキャン結果のバナー情報やオープンポート情報から機器推定機能が推定作業を行い，推定できた機器情報が，また機器特性情報データベースに登録される．これらの作業を繰り返すことで，定期的なスキャンが効率的に実施される．

3. 機器推定の仕組み

本章では，本スキャンシステムの機器推定機能が行う機器推定の実施フローと，機器推定データベースの構成について説明する．

3.1 機器推定の実施フロー

本スキャンシステムで得られた各対象 IP アドレスのスキャン結果毎に，図 2 に示す実施フローに基づいて機器推定が実施される．

まず，対象 IP アドレスのスキャン結果中のバナー情報，オープンポート情報に基づき，機器推定データベースの全ての機器情報に対して(1)バナー類似度と(2)ポート類似度を算定し，これらの類似度の加重平均を各機器の類似度とする．バナー類似度とポート類似度の算出方法は後述する．

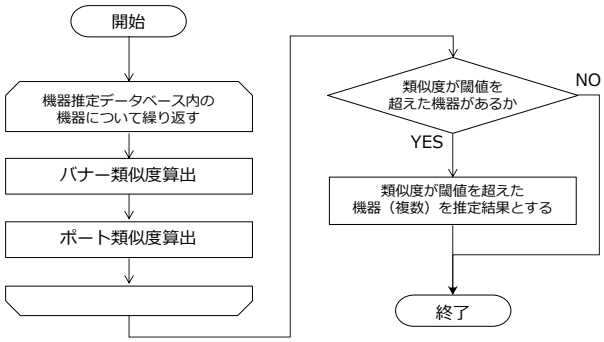

図 2　機器推定の実施フロー

次に，機器推定データベース内の全ての機器情報の類似度のうち，一定の閾値以上の機器情報を当該 IP アドレスに接続された機器に関する情報として推定する．閾値を超えた機器情報が複数ある場合には，いずれも当該 IP アドレスに紐づく機器であると推定する(例えば，NAPT 等によって複数機器が接続されている等)．一方，類似度がいずれも閾値を超えない場合には，機器が推定できなかったものと判断する．

3.1.1 バナー類似度

機器推定データベースに保持している機器毎の推定用キーワードおよび当該キーワードの重み係数を基にバナーの類似度を算出する．キーワードは正規表現パターンで保持しており，機器推定対象のバナーがこの正規表現パターンにマッチするか否かにより判定する．

バナー類似度 S_b を求める式を(1)に示す．

$$S_b = \sum_{i=1}^{n} k_i w_i \quad (1)$$

ただし，式(2)を満たすこととする．

$$\sum_{i=1}^{n} w_i \leq 1 \quad (2)$$

ここで，n は当該機器推定用のキーワードの個数，k_i は i 番目のキーワードにマッチした場合は 1，マッチしなかった場合は 0，w_i は i 番目のキーワードの重み係数を示す．

3.1.2 ポート類似度

機器推定データベースに保持している各機器が使用するポート集合と機器推定対象の IP アドレスがオープンしていたポート集合の Jaccard 係数をポート類似度として採用する．ポート類似度 S_p を求める式を(3)に示す．

$$S_p = \frac{|P_s \cap P_d|}{|P_s \cup P_d|} \quad (3)$$

ここで，P_s は対象 IP アドレスへのスキャンにより応答が得られたポートの集合，P_d は機器推定データベース内の当

該機器が使用するポートの集合を示す.

3.1.3 機器類似度

3.1.1 で求めたバナー類似度 S_b および 3.1.2 で求めたポート類似度 S_p から，機器類似度 S を求める式を(4)に示す.

$$S = \alpha S_b + (1-\alpha) S_p \quad (4)$$

ここで，α はバナー類似度の重み係数を示す.

3.2 機器推定データベースの構成

機器推定データベースにおいて，各機器に関するものとして保持する情報を以下に示す.

- キーワード，重み係数
  当該機器とのバナー類似度を算出するために使用するキーワード. 正規表現パターンの形式で当該キーワードの重み係数と共に保持する．1 つの機器に対し，キーワードは複数保持可能とする.
- メーカ名
  当該機器を製造したメーカ名を保持する.
- 機種名
  当該機器の機種名を保持する.
- 型番
  当該機器の型番を保持する.
- バージョン
  当該機器のファームウェアのバージョンを保持する.
- 使用ポート集合
  当該機器が使用する可能性が高いポートの集合を保持する.

4. 機器推定データベースの作成

本章では，機器推定に用いる機器推定データベースの作成にあたっての考え方と，実際の作成方法を説明する.

4.1 広域スキャン結果からの機器推定データベースの作成フロー

2018 年 12 月～2019 年 2 月にかけて実際に日本国内の一部の IP アドレスに対して広域スキャンを実施して得られたバナー情報から，機器推定データベースに登録する機器毎のデータを作成した.

実施した広域スキャン結果から取得されたバナー情報から機器情報データを作成するにあたっては，大量のバナー情報を処理・確認・調査する必要があり，機器推定データベースを作成するためには，作業の効率化を図る必要があった. 特に，調査すべきバナー情報を絞り込み・正規化・集約化することで，確認・調査すべきバナーを減らすことでデータベース作成の効率化を図った.

4.1.1 フィルタリング処理

広域スキャンにより収集したバナーに対し，まずは明らかに機器推定が不可能なバナーや IoT 機器ではないと判定できるバナーを除外するフィルタリング処理を実施することで，調査対象のバナーの絞り込みを行った.

本フィルタリング処理にて，以下のようなバナーを除外対象とした.

- プロトコルのメッセージのみで機器に関する情報が含まれないバナー
  【例】 220 FTP Server ready.
- メールサーバのように明らかにサーバとして動作していると判断できるバナー
  【例】.*ESMTP Postfix
- ホスティングサービスであることがわかる証明書など明らかに IoT 機器ではないと判断できるバナー

4.1.2 正規化・集約化処理

フィルタリング処理後，さらなる調査対象のバナーの削減のため，同一の機器が応答したと推測されるバナーの集約化を実施した. ただし, 応答した時刻が格納される HTTP レスポンスヘッダの Date フィールドのように, 同一の機器であっても，バナー取得のタイミングや機器の設定などによりバナーの文字列が異なるフィールドが存在するため，取得したバナーをそのまま集約しても集約化の効果が小さい．そのため，同一の機器であっても，値が異なる可能性が高いフィールドについては正規化処理を実施し，正規化後のバナーに対して，集約化を実施した.

正規化の対象とする文字列の一例を以下に示す.

- HTTP レスポンスの Date フィールド，Expires フィールドなどの値の日時情報
- HTTP レスポンスの ETag フィールド
- HTTP レスポンスの Content-Length フィールド
- TLS レスポンス値

4.1.3 目視による確認・調査

正規化・集約化により集約したバナーについて，目視による確認やインターネットによる機器の調査を実施することで，機器推定データベースへ登録するデータの作成を実施した.

機器推定に使用するキーワードについては，バナー中に現れるメーカ名，機種名，型番など，当該機器を一意に識別できるような特徴的な文字列を設定する．キーワードに付与する重み係数は，キーワードに機器の型番が含まれるなど特定の機器を推測できる文字列が含まれるような場合には，1.0 を使用し，特定の機器までは推定できないが，カメラなどの機器種別が推定できる場合や，組み込み機器用の Web サーバなどが推定できる場合などは, 0.8 など 1.0 より小さい値を使用した.

各機器が使用するポート集合については，設定したキーワードを基に当該機器が接続されていると推定される IP

アドレスを検索し，広域スキャンの応答結果から，当該 IP アドレスがオープンしていたポートの和集合を設定する．

登録するデータについてはメーカ名，機種名および型番のいずれかが不明であっても，ルータやカメラなど何らかの IoT 機器であることが推定される文字列がバナーに含まれていて，複数の IP アドレスから同様のバナーの応答があり，使用しているポート集合が判明しているものであれば，データベースへの登録を実施する．また，バナーから組込み機器用として提供されている Web サーバ用のミドルウェアであることが判定できるものについても，データベースへの登録を実施した．このような組込み機器用の Web サーバのミドルウェア脆弱性を突いた攻撃は多く見られるため，広域スキャンにより使用状況を把握することは重要だと考える．

5. 広域スキャン結果によるデータ作成

提案した機器推定手法の有効性を確認するため，まずは，国内の IP アドレスの一部（1000 万 IP アドレス）に対して，広域スキャンを実施して得られた結果を基に機器推定データベースを作成した．さらに，国内で使用されている IP アドレス（約 1.5 億 IP アドレス）に対して広域スキャンを実施して得られた結果に対し，作成した機器推定データベースによる機器推定を実施し機器推定状況の評価を行った．スキャンには，masscan を使用し，取得するバナーは，masscan で取得できるバナーを対象としている．

5.1 機器推定データベースの作成結果

国内の 1000 万 IP アドレスに対して広域スキャンを実施することで 22,322 の IP アドレスから，のべ 111,674 件のバナーを取得した．これら取得したバナーに対し，フィルタリング処理を実施した結果，調査対象のバナーは 24,413 件に削減され，さらに，正規化・集約化を実施した結果，調査対象のバナーは 8,147 件に削減された．この 8,147 件のバナーに対して，目視による確認・調査を実施した結果，301 件の機器情報を持つ機器推定データベースを作成した．

作成した機器推定データベースにおいて，機器数が多かった機器種別について，機器数の内訳を表 1 に示す．ただし，機器によっては複数のポートから異なるバナーを返すが，いずれのバナーからも機器推定できる機器も存在するため，作成した機器推定データベースには，同一の機器であるが，異なるデータとして登録されているものもあることに留意されたい．

ルータが 71 機器と最も多く，その内 25 機器が家庭向けとして販売されているルータであった．また，ネットワークカメラおよび UTM については，メーカによっては，バナーから型番まで判明する機器が存在し，同機種で型番が異なる機器をデータベースに登録することができたため，機器数が他の機器種別と比較して多くなっている．

また，機器をどの段階まで推定できたかの指標とする推定レベルの定義および作成した機器推定データベースにおける各レベルの機器数の内訳を表 2 に示す．ここでは，機器種別等は不明であるが，組込み機器用の Web サーバであることがバナーから判別可能であり，IoT 機器である可能性が高いものをレベル 0 と定義している．

表 1 機器推定データベース内の機器種別の内訳

機器種別	機器数
ルータ	71
ネットワークカメラ	53
UTM	19
DVR	15
NAS	12
NVR	10

表 2 機器推定データベース内の機器推定レベルの内訳

推定レベル	概要	機器数
レベル 0	IoT 機器である可能性が高いと推定できる	26
レベル 1	機器種別（ルータなど）が推定できる	27
レベル 2	メーカが推定できる	23
レベル 3	メーカおよび機器種別が推定できる	79
レベル 4	機器の型番まで推定できる	146

5.2 実スキャン結果における機器推定状況

国内で使用されている約 1.5 億の IP アドレスに対して広域スキャンを実施して得られたバナー情報およびオープンポート情報に対して，5.1 で作成した機器推定データベースを用いて 3.1 の機器推定フローに基づき，機器推定を実施した．

国内で使用されている約 1.5 億の IP アドレスに対して広域スキャンを実施したところ，1 つ以上のポートから応答があったアドレス数は約 340 万アドレスであった．その中で，何らかのバナーを取得できたアドレス数は 323,629 であり，のべ 1,645,475 件のバナーを取得した．取得したバナーに対して，5.1 で作成した機器推定データベースを用いて，3.1 の機器推定フローに基づき，機器推定を実施した．

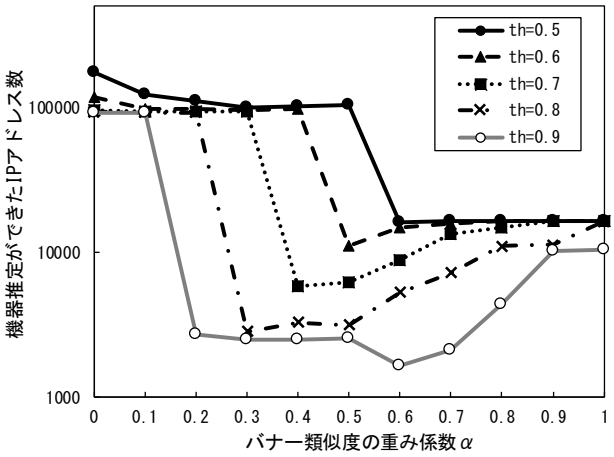

図 3 機器推定データベースによる機器推定結果

機器推定時のパラメータとして，機器推定できたと判定する閾値 th を 0.5 から 0.9 とした場合，それぞれについて，バナー類似度の重み係数 α を 0.0 から 1.0 まで変化させた時の，機器推定フローに基づき，機器を推定することができた IP アドレス数を図 3 に示す.

図 3 より，閾値 th がポート類似度の重みとなる 1-α 以下の場合には，機器推定ができたと判定される IP アドレスが，それ以外の場合と比較して非常に多くなることがわかる．これは，機器推定するための情報として，ポート集合の類似度が支配的な要素となっているためであり，オープンポートの集合が，機器が使用するポートの集合と類似していたために誤推定しているものと考える.

上記閾値 th の値により，機器推定がどの程度正しく行われているかを調査するため，誤推定の尺度として，False Positive および False Negative を以下のように定義する.

(1) False Positive による誤推定

目視によりバナーから機器を推定した結果，いずれの機器にも推定できなかったが，機器推定フローに基づく機器推定では，特定の機器と推定された.

(2) False Negative による誤推定

目視によりバナーから機器を推定した結果，特定の機器を推定することができたが，機器推定フローに基づく機器推定では，機器を推定できなかった．もしくは，別の機器として推定された.

本調査では，正解データにおいて，オープンポートの情報から機器推定できる可能性を考慮しないため，ポート類似度による機器推定の寄与率(1-α)の評価はできないが，それらによる影響も考慮するため，重み係数 α を変更して測定を行った．ポート類似度の比重（1-α）を高くすると今回の調査による機器推定の結果は悪くなるものの，実際の機器推定においてはオープンポートの情報から機器推定できる場合も考えられるため，その点については本調査と異なる可能性がある.

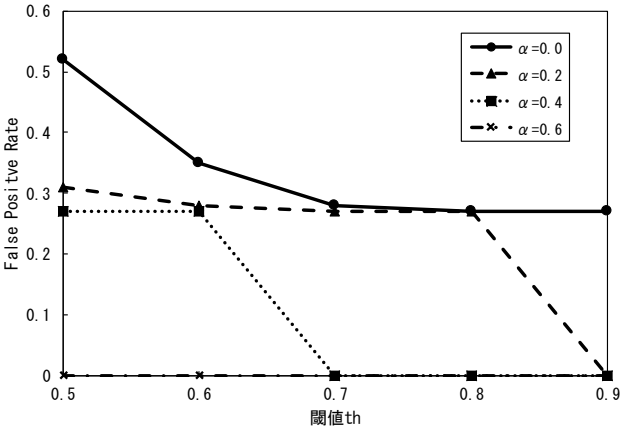

図 4 False Positive による誤推定割合

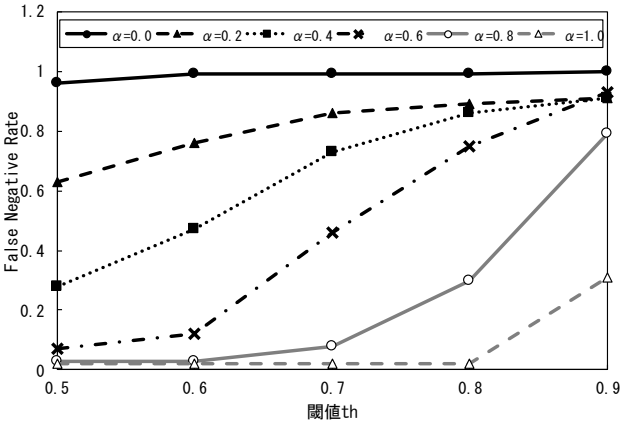

図 5 False Negative による誤推定割合

2 つのパラメータ（th と α）における False Positive および False Negative の誤推定の割合を調べるため，取得したバナーから目視により，機器が推定できると判定できる IP アドレス 100 件および機器が推定できないと判定できる IP アドレス 100 件をサンプリング抽出し，それぞれのパラメータにおける，抽出した IP アドレスの機器推定状況から，False Positive および False Negative の割合を調べた結果を図 4 および図 5 に示す.

図 4 は False Positive による誤推定の割合である．図 4 ではバナー類似度の重み係数 α が 0.6 の場合までしか示していないが，α が 0.6 以上では，いずれの場合でも False Positive の割合は 0 であった．閾値 th が低くなるほど False Positive による誤推定の割合が増える傾向はあるが，バナー類似度の重み係数 α が 0.6 以上の場合には，閾値 th によらず False Positive による誤推定の可能性が低いことがわかる.

図 5 は False Negative による誤推定の割合である．False Positive とは逆に，閾値 th が高くなるほど False Negative による誤推定の割合が増える傾向があるが，特に重み係数 α が大きい場合には，閾値 th が 0.9 より高くなると誤推定の割合が急激に増加する.

これらの調査結果から，バナー類似度の重み係数 α をあ

る程度大きくする場合（0.6 以上）における，閾値 th の最適な値は 0.8 程度と考えられる．

上記の結果を踏まえ，誤推定の割合が最小となるパラメータとして，機器推定できたと判定する閾値を 0.8，バナー類似度の重み係数を 1.0（ポート類似度を考慮しない）として機器推定を行ったところ，16,362 の IP アドレスについて機器を推定できた．これは，バナーを取得できたアドレスに対して，約 5.1%の割合となった．このパラメータにおいて，機器を推定できた機器種別の内訳を表 3 に，機器推定レベルの内訳を表 4 に示す．推定できた機器の機器種別については，機器推定データベース内のデータと同様にルータが最も多く，その内 1,512 機器が家庭向けとして販売されているルータであった．推定できた機器の機器推定レベルについては，機器推定データベース内のデータ構成とは異なり，レベル 3 まで推定できた機器が最も多かった．これは，レベル 4 のデータは同じメーカ，同じ機種の機器が型番毎に作成されるため多くなるが，実際にはレベル 4 の型番まで機器推定できる機器の割合は少なかったためと考えられる．

表 3　推定機器の機器種別の内訳．

機器種別	推定機器数
ルータ	5,222
ネットワークカメラ	1,540
UTM	926
DVR	1,988
NAS	1,551
NVR	504

表 4　推定機器の機器推定レベルの内訳

推定レベル	推定機器数
レベル 0	3,656
レベル 1	1,749
レベル 2	1,208
レベル 3	8,749
レベル 4	3,209

・

5.3 研究倫理に関する考察

本研究では，日本国内の IoT 機器のセキュリティ状況について網羅的な調査を行うための広域ネットワークスキャンを実施しているが，実施の目的および調査に使用する IP アドレス等については弊社のニュースリリースにて公開の上，実施している(https://www.ntt-at.co.jp)．本ニュースリリースや whois 情報に連絡先情報を記して，スキャン対象先等からの問い合わせには適切に対応するとともに，対象除外の申請があった場合は確実に除外設定を行っている．また，スキャン対象とする IP アドレスをランダマイズ化・分散化することにより，対象先ネットワークにおいて本スキャンによる負荷を低減するよう調整している．

本スキャンにおいては IoT 機器から得られるバナー等の情報を蓄積・活用するが，それらの情報については必要なセキュリティ管理を行い，漏洩等が起こらないようにしている．弊社が実施するスキャンでは，IoT 機器等からの返答パケットの確認まで行うが，その後のログインその他のアクセスは実施しない．

6. おわりに

本稿では，通信量を最小化する広域ネットワークスキャンシステムと当該システムにおける機器推定機能の概要について説明するとともに，大量に得られたスキャン結果中のバナー情報から機器推定用のデータベースを効率的に作成するための方法について説明した．また，実際に行ったスキャン結果から抽出した機器推定用データおよび，それらのデータを活用した場合の機器推定の精度についてシミュレーション結果等から推測した．

6.1 今後の課題

今後は，作成した機器推定データベースを用いて，日本国内のより多くの IoT 機器に対してスキャンを行うことで，機器推定の精度および，機器推定によるスキャンの効率化の程度について実測する．

また，時間経過に伴い増加することが想定される IoT 機器の種別を迅速に識別するための，機器推定 DB 作成の更なる効率化に向けて，バナーから IoT 機器であるかのフィルタリング処理に機械学習を使用する等，自動化に向けて検討していく予定である．

謝辞　本研究は総務省の委託研究「周波数有効利用のための IoT ワイヤレス高効率広域ネットワークスキャンの研究開発」により実施したものである．

参考文献

[1] Yin Min Pa Pa, Shogo Suzuki, Katsunari Yoshioka, Tsutomu Matsumoto, Takahiro Kasama, and Christian Rossow (2016) IoTPOT: A Novel Honeypot for Revealing Current IoT Threats, 情報処理学会論文誌, Vol.57, No.4

[2] 森博志，鉄穎，小山大良，藤田彬，吉岡克成，松本勉，能動的観測と受動的観測による IoT 機器のセキュリティ状況の把握

[3] nmap(https://nmap.org)

[4] SHODAN(https://www.shodan.io/)

[5] Censys(https://https://censys.io/)

[6] X. Feng, Q. Li, H. Wang, and L. Sun, “Acquisitional Rule-based Engine for Discovering Internet-of-Thing Devices,” 27th USENIX Security Symposium, 2018.

[7] A. Costin, A. Zarras, and A. Francillon, “Towards Automated Classification of Firmware Images and Identification of Embedded Devices,” IFIP SEC, 2017
[8] Q. Li, X. Feng, H. Wang, Z. Li, and L. Sun, “Towards Fine-grained Fingerprinting of Firmware in Online Embedded Devices,” IEEE INFOCOM – IEEE Conference on Computer Communications, 2018.
[9] 内田 佳介, 森 博志, 藤田 彬, 吉岡 克成, 松本 勉, ‘‘管理 WebUI の画像的特徴に基づく IoT 機器判別手法の提案,’’ 情報処理学会 Computer Security Symposium, 2017 年.