IoT 機器に関する広域ネットワークにおける最適化したスキャンポートおよび頻度を用いた効率的なスキャン方式について

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

概要：近年著しく数が増加している IoT 機器等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，効率的にスキャンを実施し，脆弱な機器等を検出できることが望ましい．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，その IoT 機器を推定し，推定した機器に対してスキャン対象ポートを限定したり，機器の応答状況に応じてスキャンする頻度を可変とする仕組みが有効であると考える．本稿では，最初に日本国内全域を対象とした広域ネットワークスキャンと TCP のフルポートスキャンを実施した結果を紹介し，その結果から検討した広域を網羅的にスキャンするための標準スキャンポート，機器の変化に追従するための標準スキャン頻度を提示する．さらに，通信量を削減しつつ効率的にスキャンするためのポート最適化方式，頻度最適化方式と，それらを用いた効率的なスキャンシステムを紹介すると共に，日本国内を対象に実施した広域スキャンによる実測データを用いて，通信量削減効果とポート検出率について標準的なスキャン方式と比較・評価した結果を報告する．

キーワード：IoT，ネットワークスキャン，ポート

Efficient scanning method with optimized scanning ports and frequency in wide area network for IoT devices

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

Abstract: In recent years, the number of IoT devices has increased significantly. The wide-area network scan system for investigating the vulnerable problem of these IoT devices needs to enable sufficient investigation with the minimum network traffic. Such an efficient scan system makes it possible to narrow down the target ports by estimating the IoT device from the scan results. We will introduce the results of the wide area scan covering the whole of Japan and TCP full port scan. The standard scan port for comprehensively scanning the wide area are determined from the scan results and the standard scan frequency for following the changes of the equipment are presented. In addition, we introduce port optimization method and frequency optimization method for efficient scanning while reducing the network traffic. We will report the results of comparison and evaluation of the network traffic reduction effect and port detection rate with the standard scan using the actual measurement data obtained by the wide area scan conducted in Japan.

Keywords: IoT devices,　Network scanning,　ports

1. はじめに

近年，IoT 機器の数は著しく増加している．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[2]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[3][4][5]．一方で，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おそれがあり，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，スキャン結果から得られた情報を用いて，IoT 機器の機種やカテゴリーなどを可能な限り推定して，その機種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．この対象機器に応じたポートの絞り込みにより，無駄なスキャンを省き通信量を最小化することが可能となる．機器が推定できないケースについても少ないスキャントラフィックで効果的なスキャンをするための最適なスキャンポートを検討する必要がある．また，応答がない IP アドレスに対して，機器の新たな接続を速やかに検出するスキャンポートやスキャン頻度の検討も必要となる．頻度に関しては，機器の応答特性に応じた最適な頻

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

© 2020 Information Processing Society of Japan

度の検討が必要である．モバイル機器のようなエリア移動によってIPアドレスが頻繁に変化するような機器と，設置してから長期間運用される監視カメラ等では，スキャンの最適な頻度は異なっており，無駄なスキャントラフィックを省き通信量を最小化するために機器に応じた頻度の設定が必要である．

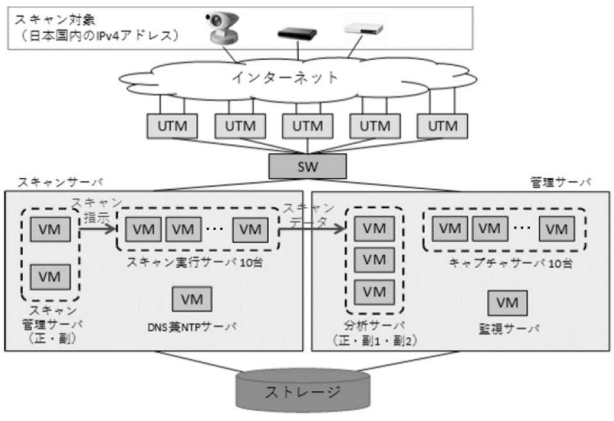
本稿の構成は次の通りである．次章にて，我々が構築した広域ネットワークスキャンシステムの構成について説明する．それを用いて国内全域を対象に広域ネットワークスキャンを実施した結果も紹介する．第3章では，広域ネットワークスキャンの結果を基に設定した網羅的にスキャンするための標準的なスキャンポートと，応答の変化に追従できる標準的なスキャン頻度を採用した標準的なスキャン方式について定義する．第4章では，標準的なスキャン方式を基準として，より効率的にスキャンするためのポート最適化と頻度最適化を用いた効率的なスキャン方式について説明する．第5章では，実ネットワークにおいて実施したスキャン結果を用いて，標準的なスキャン方式と効率的なスキャン方式を比較し，通信量の削減効果とポート検出状況を評価する．

2. 広域ネットワークスキャンシステム

2.1 広域ネットワークスキャンシステムの概要

国内のIPv4アドレスに対して，広域ネットワークスキャンによる調査を実施するためのシステムを設計して構築した．図1にスキャンシステムの構成を示す．スキャンツールは高速にスキャンすることが可能なMasscan[6]を採用した．国内約1.5億のIPv4アドレスをターゲットとし，多数のIPアドレスへのスキャンを高速に実行するため，10台のサーバおよび10回線を使用して，並列にスキャン実行が可能な構成とした．本システムでは，1,000万IPアドレスに対して約7,500ポートのスキャンを72時間で実行することが可能である．

図1　スキャンシステム構成図



2.2 広域ネットワークスキャンの実施

広域ネットワークスキャンを実施する際にフルポートでのスキャンは網羅性の観点では最適であるが，効率性・通信量の観点からは非効率で現実的ではないため，広域ネットワークスキャンに最適なポートに関する先行研究がないか探索したが見つけることができなかった．そのため本研究の前段調査として，まずは国内のインターネット上の約1,200万のIPv4アドレスに対してTCPのフルポートである65,535ポートでスキャンを実施し，応答するポートの特性，応答するポートの状況等を調査することにした．

2018年12月にTCPフルポートによる広域ネットワークスキャンを行った．スキャンは，TCP SYNパケット送信によるポートスキャンに併せて，バナーの取得も実施した．送信先から何らかの応答があったIPアドレス数は328,240件であった．このスキャン試行に対するIPアドレスの応答率は2.7%であった．広域ネットワークスキャンで応答のあったポートについてはランキングを作成して今後のスキャン対象ポートの選定に利用した．表1 に，フルポートスキャンの応答上位10位までのランキングを記載した．このポートランキングは，4章 効率的なスキャン方式のポート選定で利用している．

表1　ポートの応答ランキング

順位	ポート番号
1	80
2	443
3	25
4	21
5	110
6	587
7	7547
8	143
9	340
10	995

2019年4月に日本国内全域を対象とした応答特性を調査するため，国内約1.5億のIPv4アドレスに広域ネットワークスキャンを実施した．約1.5億とは国内に割り当てられている約2億のIPv4アドレスから主に大手パブリッククラウド等で使われているIPアドレス帯を除外したIPアドレスである．またポートについては，フルポートで全てのIPアドレスのスキャンを実施する場合は1年以上の時間を要することから，先述のフルポートスキャンで応答IPアドレスが多い順にポートをスキャン対象ポート集合に追加していく操作を行い，そのスキャン対象ポート集合内のポートに応答したIPアドレス数の95%となるポート集合である約7,500ポートを用いて実施することにした．

この広域ネットワークスキャンでは 3,423,652 件の応答があり，応答率は 2.30%であった．

広域ネットワークスキャンにおいて実施時期，ポート数，IP アドレス数は異なるものの，応答率，応答ポートともに概ね近い数値を示しており，異なるサンプルにおける広域ネットワークスキャンであっても応答のある割合は 2%～3%程度であると推測する[7].

3. 標準的なスキャン方式

前章の国内約 1,200 万の IP アドレスに対して実施した TCP のフルポートスキャンと国内 1.5 億の IP アドレスを対象に約 7,500 ポートで実施したスキャンの応答結果を基に，広域ネットワークを網羅的にスキャンするためのポートや，モバイル機器をはじめとした IoT 機器の動的変化にも追従できるような標準的なスキャン方式を検討することにした．本研究において，標準的なスキャン方式とはスキャンポートとスキャン頻度の基準値となるものであり，この標準的なスキャン方式でスキャンした際に得られるスキャン結果と同等の結果を目指しつつ，スキャンポートや頻度を削減し，効率的にスキャンする方式を検討するかを命題とした．

3.1 標準的なスキャンポート

脆弱な IoT 機器を検出するため，定期的に広域ネットワークスキャンを実施する場合，どのくらいのポート数やポート種類でスキャンを実施すべきなのかを検討する．

ポートの最適化手法を検討するにあたり，比較対象として網羅的にスキャンを行うために応答状況を把握し，標準的なスキャン対象ポートを以下のような手順で定めた．

日本国内の約 1,200 万のグローバル IPv4 アドレスに対して，TCP フルポートスキャンを実施し，IoT 機器を含む全ての機器を対象に実際に使用されているポートの応答状況を調査する．

IoT 機器とは考えづらいハニーポットなどと想定される応答ポート数が 100 以上の IP アドレスは，除外する．

更に，極少数の IP アドレス（全応答 IP アドレス数の 0.001%未満）からしか応答がないポートを特異点として除外する．

上記の条件下において，定めたポート（数とポート種類）を標準的なスキャン方式で対象とするポートとする。

表 2　TCP フルポートスキャン結果による標準ポート数

対象 IP 数	応答数	応答数の 0.001%	標準的なポート数
1,200 万	328,240	3.28	約 11,000

表 2 は，TCP フルポートスキャンにて 1,200 万件の IP アドレスに対して，2018 年 12 月に調査した内容を示したものである．応答数の割合は 2.7%程度であり，応答があった IP アドレス約 32.8 万の 0.001%未満に相当する約 3 未満の IP アドレスからしか応答がなかったポートを除外することで，標準的なポート数（ポート種類）を約 11,000 ポートとした．

3.2 標準的なスキャン頻度

脆弱な IoT 機器を検出するため，定期的に広域ネットワークスキャンを実施する場合，どのくらいの頻度でスキャンを実施すべきかを検討した．標準的なスキャン頻度とは，モバイル機器等が移動した際などに IP アドレスが変化するケースに追従できるようにスキャンするための最小頻度である．2018 年 12 月に実施した約 1,200 万の IP アドレスに対しての TCP フルポートの広域ネットワークスキャンで応答があった約 27 万件の IP アドレスに対して，2019 年 1 月 18 日から 2 月 8 日までの 3 週間において 1 日に 2 回のスキャンを継続して行う定点観測スキャンを実施した．3 週間の定点観測スキャンにおいて，約 70%の IP アドレスは継続して同じ応答があり，約 30%の IP アドレスは応答に変化があった．変化があった 30%の IP アドレスについて，変化の頻度が最も多かったのは，12 時間であった．実際にはモバイル機器等は数時間あるいは数十分程度で変化している可能性も高い．しかしながら，それらの機器の IP アドレスの変化に追従するためにスキャン頻度を設定すると頻繁にスキャンを実施することになり攻撃と見なされかねない．そのため，同一 IP アドレスに対して 1 日に 2 回，12 時間に 1 回のスキャンを上限とした．本研究においては，変化に追従できる最小頻度である 12 時間を標準的なスキャン頻度と設定した．

4. 効率的なスキャン方式

本来であれば IoT 機器の状態を把握するためには標準的なスキャン方式で実施することが望ましい．しかしながら，今後増加する IoT 機器に対し，前章の標準的なスキャン方式を用いてスキャンをすると，すべての IP アドレスに一律に同じポート集合，同じ頻度でスキャンをすることになりスキャン通信量が膨大になってしまう．オープンしていないポートへのスキャンは無駄であり，また常に同じ応答を返す機器に対してはスキャンの省略が可能かもしれない．それらを理由に通信量を削減したスキャン方式を検討した．スキャンするポート[8]と頻度[9]を機器の応答状況に応じて最適化することで，標準的なスキャン方式で検出できるポートの検出数をできるだけ維持しつつ，より少ない通信量でスキャンする方式を検討した．効率的なスキャン方式については、スキャンにて応答がある IP アドレ

スに対してのスキャンと、応答がない IP アドレスに対してのスキャンでそれぞれ最適なスキャンポートとスキャン頻度を検討した．

応答がある IP アドレスに対してのスキャンにおけるポートは，機器が返すバナー情報やポート情報から機器を推定することで，その機器に応じたポートに最適化できると考えた．最初のスキャンで取得したバナー情報やポート情報を，予め作成しておいた機器推定用データベース（機器が利用するポートやバナーに含まれるキーワードから構成）と照合することによって機器を推定する[10]．機器が推定できたものについては次回以降のスキャンで，推定した機器が利用するポートでスキャンすることにした．頻度に関してはスキャンの応答が前回と同じであれば同じ機器が継続して応答していると判断して，徐々にスキャンする間隔を延伸させる方式を検討した．

一方，スキャンで応答がない IP アドレスは，2.2 節で示した応答率 2～3%の結果から，全体の約 97%と推定されるため，できるだけ少ないスキャンポートとスキャン頻度で，効果的に機器を検出する設定を検討する必要がある．応答がない IP アドレスに対してのスキャンで利用するポートは，フルポートスキャンの応答ランキングをもとに効率的に機器を検出するポート集合を作成した．頻度に関しては，定点観測スキャンにて定常的ではなく，散発的に応答を返した頻度をもとに設定した．

応答がない IP アドレスに対してのスキャン実施時に応答があった場合は，次回は応答がある IP アドレスに対してのスキャンに移行する．逆に応答がなくなった場合は応答がない IP アドレスに対するスキャンに移行する．すべての IP アドレスは，いずれかのスキャン方式でスキャンすることにした．

4.1 応答がある IP アドレスに対するスキャンのポートと頻度

効率的なスキャン方式では，バナー情報などを基に機器推定を行うことで対象機器を特定し，その機器に応じたポートと頻度でスキャンを行うことにより通信量を削減することを目標に検討した．応答がある IP アドレスに対するスキャンにおけるスキャンポートと頻度は以下のように設定した．

最適なポート選定方式の検討に先立ち，以下の式（1）に示すポート検出率（標準的なスキャン結果と効率的なスキャン結果の 2 つのスキャン結果のポート集合に含まれる共通要素の割合）により標準的なスキャンと比べてポート検出の能力があまり劣化しない条件となる 95%を目標とした．

$$\text{ポート検出率} = \frac{|\ Ps \cap Pe\ |}{|\ Ps \cup Pe\ |} \quad (1)$$

Ps … 標準的なスキャンの応答ポートの集合
Pe … 効率的なスキャンの応答ポートの集合

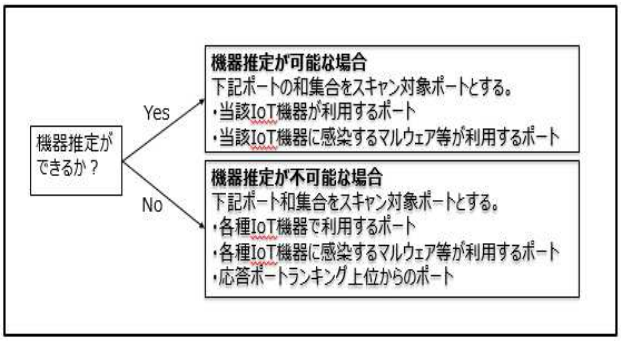


図 2　最適なポート選定方式

検討した最適なポート選定方式を図 2 に示す．機器推定が可能な場合は「当該 IoT 機器が使用するポート」と「当該 IoT 機器に感染するマルウェア等が使用するポート」の和集合とし，スキャン対象ポート数は，推定機器として機器推定データベースに取得格納されている機種が利用するポート（最小利用ポート数は 1，最大利用ポート数は 28，平均利用ポート数としては約 2.2）とした．

一方で，機器推定が不可能な場合は「各種 IoT 機器が使用するポート」と「各種 IoT 機器に感染する各種マルウェア等が使用するポート」，および「応答ポートランキング上位からのポート」の和集合をスキャン対象ポートとした．このスキャン対象ポートは，各種 IoT 機器で利用するポート（約 1,800 機種で合計として約 500 ポート）および各種 IoT 機器に感染する各種マルウェア等が使用するポート（173 種類のマルウェアで合計として約 50 ポート）の和集合として約 540 であるが，これだけではポート検出率は 75%程度になり目標の 95%に到達できないため，標準的なスキャン方式で検出できるポートとできるだけ同等になるように，応答ポートランキングにおける上位ポートを加えた和集合とした．

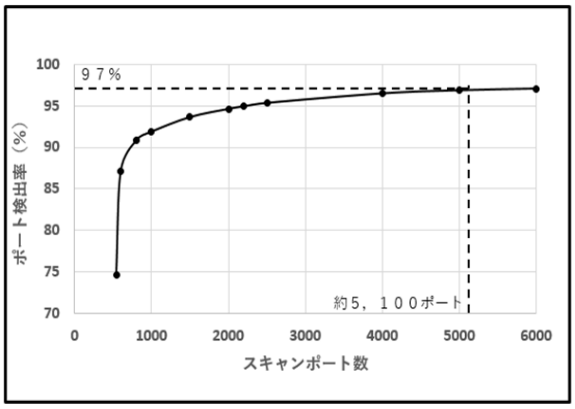


図 3　スキャンポート数とポート検出率

図 3 にスキャンポート数とポート検出率の関係をシミュレーションしたものを示す．これは，横軸がスキャンポート数，縦軸がポート検出率を表していて，スキャン対象ポート数は，実ネットワーク上での実測としてポート検出率の目標である 95%を下回らないように余裕をみて，シミュレーションにより 97%となるスキャンポート数として 5,100 とし，これを機器推定不可能な場合に使用するスキャン対象ポート集合とした．

スキャン頻度については，標準的なスキャン頻度を基に，機器の特性に応じた最適な頻度でスキャンを行うことで通信量を削減するスキャン方式を検討した．削減する方法として，スキャンで取得したバナー情報，ポート情報が前回のスキャン結果と変わらない場合は，同じ機器が接続されていると推定してスキャン頻度を延伸させる方式を策定した．

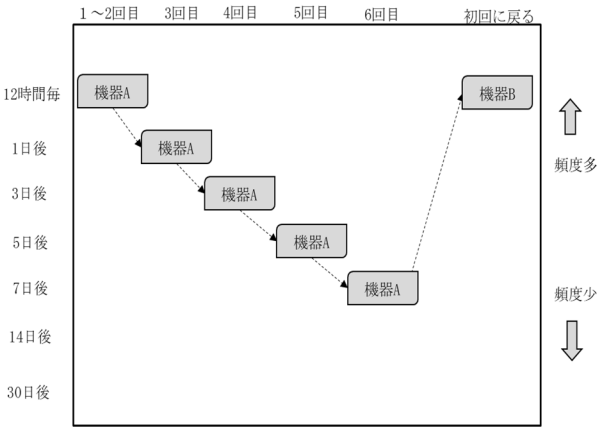


図 4 スキャン頻度の延伸イメージ図

図 4 にスキャン頻度の延伸イメージを示す．縦軸が次回スキャンまでの日数，横軸がスキャン回数を表している．初回のスキャン実施後の 12 時間後に 2 回目のスキャンを実施し，前回と同じ応答を得た場合は，1 日，2 日，3 日，5 日，7 日，14 日と徐々に延伸させていき，最長で 30 日まで延伸する． 図 4 の機器 A は 6 回目のスキャンにおいて，前回から応答が変わったため，機器が変わったと判断して初回に戻り，頻度も初期値である 12 時間になったことを示している．

4.2 応答がない IP アドレスに対するスキャンのポートと頻度

国内の広域ネットワークスキャンの特性として，2.2 節広域ネットワークスキャンの実施において判明した応答率は 2%～3%程度であるため，97%超は応答がない．これら大多数の応答のない IP アドレスに対して新たに機器が接続された，あるいは起動されたことを検出するために定期的にスキャンをする必要がある．本来であれば応答のない IP アドレスに対してのスキャンも標準的な頻度である 12 時間で実施することが望ましいが，スキャン対象の範囲が広く，トラフィック量が膨大になるため，いかに少ないポートと頻度でスキャンを実施するかが重要となる．

応答がない IP アドレスに対するスキャンポートの検討に先立ち，以下の式（2）に示す機器検出率（応答がない機器に対するスキャンは，1 ポートでも検出したら検出可能とし，標準的なスキャン方式に対する機器検出能力を表す）として 95%を目標とした．

$$\text{機器検出率} = \frac{De}{Ds} \qquad (2)$$

De … 効率的なスキャンで 1 ポートでも検出した IP アドレス数
Ds … 標準的なスキャンで 1 ポートでも検出した IP アドレス数

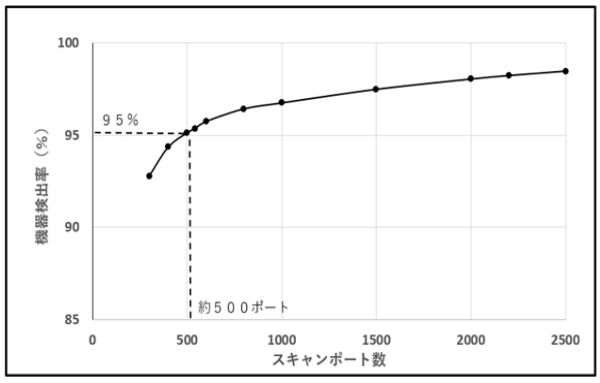


図 5　スキャンポート数と機器検出率

応答がない IP アドレスに対するスキャンポートの検討として，図 5 にスキャンポート数と機器検出率を示す．これは，シミュレーションにより機器検出率を算出したもの

であり，横軸にスキャンポート数（応答ランキング上位のポートから選定），縦軸に機器検出率を表した．

機器検出率が 95%となる部分が，スキャンポート数として約 500 となり，これを応答がない IP アドレスに対するスキャンで使用するスキャンポートとした．

スキャン頻度については，1 日 2 回の定点観測の結果を用い応答頻度を測定した．表 3 の定点観測における散発的な応答は縦の項目が定点観測スキャンにおける IP アドレス，横の項目がスキャンを実施した日時であり，〇が応答状況を示したものである．表 3 の IP アドレス（a.a.a.a）のような 8 割以上の応答を返した IP アドレスは常時機器が接続されていると想定して対象から外し，それ以外の散発的な応答を返した IP アドレスにおける 2 週間のスキャンでの応答数は平均 3.45 回であった．これは 4 日に 1 回の応答であり，すなわち 4 日未満でスキャンを行えば遅延なく応答を検知できる想定である．そのため，応答がない IP アドレスに対してのスキャン頻度は 3.5 日に 1 回とした．

表 3　定点観測における散発的な応答

	9月18日		9月19日		9月20日		9月21日		9月22日	
IPアドレス	0時	12時	0時	12時	0時	12時	0時	12時	0時	12時
a.a.a.a	〇	〇	〇	〇	〇	〇	〇	〇		〇
b.b.b.b				〇					〇	〇
c.c.c.c		〇					〇			
d.d.d.d				〇	〇			〇		
e.e.e.e	〇						〇			

5. 実ネットワークにおける評価

2019 年 12 月から 2020 年 5 月に渡り，前章の効率的なスキャン方式を採用して実ネットワークによる広域ネットワークスキャンを実施した．ランダム抽出した毎回異なる 10 万件の IP アドレスに対して 1 回目と 2 回目は 2 週間，3 回目は 8 週間の計 3 回のスキャンを実施した．また，効率的なスキャン方式で採用したポートと頻度の最適化手法の評価を行うため，同じ IP アドレスに対して，並行して標準的なスキャンを実施した．計 3 回実施した効率的なスキャン方式での通信量について標準的なスキャンと比較した際の削減率と，ポート検出率について示す．また，検討した効率的なスキャン方式で採用するスキャンポートとスキャン頻度についても，それぞれ削減効果を評価する．

5.1 効率的なスキャン方式による通信量の削減効果とポート検出率

標準的なスキャン方式と効率的なスキャン方式の通信量について，ポートと頻度による通信量の削減効果を確認するために，サーバが発出する通信量での比較を行った．表 4 は，標準的なスキャン方式の送信量と効率的なスキャン方式の送信量を 3 回実施したスキャンでそれぞれ比較したものである．効率的なスキャン方式の通信量には応答がある IP アドレスに対するスキャンと応答がない IP アドレスに対するスキャンの通信量を合算している．2 週間のスキャンにおける効率的なスキャン方式の通信量は標準的なスキャン方式の 8.9%程度であり，8 週間のスキャンにおいては 3.95%となった．いずれのスキャンにおいても標準的なスキャンと比較すると大幅な通信量の削減となった．

効率的なスキャン方式は，ポート検出率を標準的なスキャンの 95%以上を保ちつつ通信量を削減することが目標である．通信量の削減が可能であってもスキャン品質としてのポートの検出率が劣化していては脆弱な IoT 機器を検出するためのスキャンシステムとして運用できない．

表 4 には，それぞれのスキャンでのポート検出率を記載した．計 3 回のスキャンにおけるポート検出率の平均値としては，約 95.2%となり目標値を上回った。しかしながら，1 回目は 94.9%，3 回目は 94.3% となり，目標の 95%を若干下回った．スキャン結果を分析したところ，標準的なスキャンと効率的なスキャンは 12 時間内でそれぞれ非同期にスキャンを実施しており，全く同じタイミングでスキャンしていないことが原因で取得したポートに差異があることがわかった．また，ネットワークの輻輳等によるスキャンパケットの損失もポート検出率低下の原因となっていると考えられる．

表 4　標準的なスキャン方式と効率的なスキャン方式における通信量の比較とポート検出率

回数	期間	標準的なスキャン通信量	効率的なスキャン通信量	通信量の削減	ポート検出率
1	2 週間	1,676GB	149GB	8.94%	94.9%
2	2 週間	1,676GB	150GB	8.96%	96.5%
3	8 週間	7,901GB	312GB	3.95%	94.3%
平均	—	—	—	—	95.2%

5.2 最適化したスキャンポートと頻度の削減効果

実ネットワークにおける 3 回のスキャンにおけるポートの削減効果を以下に示す．標準的なスキャン方式におけるスキャンポートは一律 11,000 ポートとした．効率的なスキャン方式におけるスキャンポートは，機器の推定可否によって決まり，応答がない IP アドレスに対しては，無応答スキャンのポート集合でスキャンすることにした．表 5 は 10 万件の IP アドレスに 2 週間，標準的なスキャン方式と効率的なスキャン方式でスキャンした際のポート数を試算したものである．標準的なスキャン方式の対象 IP アドレスである 10 万件に対し，効率的なスキャン方式で

は，応答がない IP は全体の 97%程度であるため約 97,000 件，応答がある IP アドレスのうち，広域ネットワークスキャンでバナー情報が取得でき，さらに機器推定が可能な割合は約 3%程度である[10]ため，試算としては応答した IP アドレス全てでバナーが取得できたと仮定して 90 件とし，残りの機器推定不可の件数は 2,910 件とした．1 IP アドレスあたりのスキャンポート数の詳細については第 4 章にて説明した．

表 5　スキャンポート数の試算

スキャン方式	1 IP 当りのポート数	IP アドレス数	総ポート数
標準的なスキャン方式	11,000	100,000	1,100,000,000
応答がある IP に対してのスキャン（機器推定可）①	平均 2.2	90	198
応答がある IP に対してのスキャン（機器推定不可）②	5,100	2,910	14,841,000
応答がない IP に対してのスキャン③	500	97,000	48,500,000
効率的なスキャン方式　合計①＋②＋③	—	—	63,341,198

2 週間のスキャンにおいて，効率的なスキャン方式（応答がある IP に対してのスキャンと応答がない IP に対してのスキャン）の総ポート数は 63,341,198 ポートであり，標準的なスキャン方式の総ポート数 1,100,000,000 の 5.8%であった．スキャンの応答有無，機器の推定可否によってスキャンポートを変えることにより，ポート数を大きく削減することが可能である.そのことは通信量の削減に大きな効果を表している．

効率的なスキャン頻度方式を用い，10 万件の IP アドレスに対してスキャンを行い，8 週間に渡る延伸の状況を確認した．図 6 は，縦軸は延伸日数の割合，横軸はスキャン日数の延伸が発生した時を表している．最長延伸の割合はそれぞれ約 70%程度となっている．この割合は 3 週間の定点観測において，期間中変化なく同じ応答を返した割合と同程度であった．そのため，継続して同じ応答を返した約 70%の IP アドレスは，機器に固定的に設定された IP アドレスを使用しており変化はないものと考えられる．効率的なスキャン方式では，変化がない IP アドレスに対するスキャンは削減することになり，そのことが通信量の削減に繋がっている．

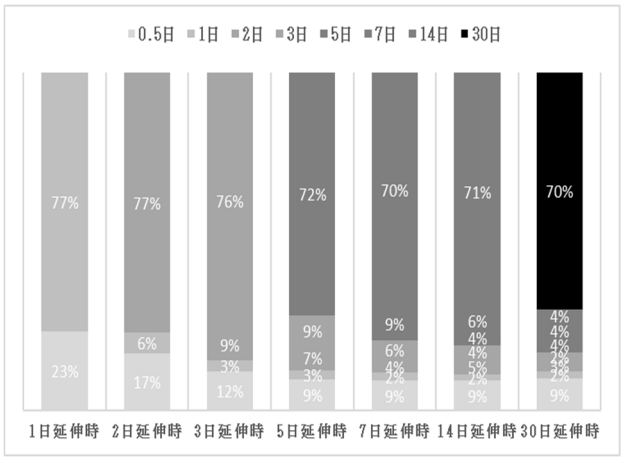


図 6 効率的なスキャンにおける延伸の状況

標準的なスキャンと効率的なスキャンにおいて，スキャンの試行 IP アドレス数で頻度削減の効果を確認した．図 7 は，2 週間のスキャンでのスキャン試行回数の削減効果を示す．効率的なスキャン方式ではスキャンをスキップすることでスキャン回数が削減でき，応答がない IP アドレスに対するスキャンは 3.5 日に 1 回としているため，標準的なスキャン方式に比べて効率的なスキャン方式の試行 IP アドレス数は 2 週間のスキャンで約 28%に削減でき，8 週間のスキャンでは約 21%に削減できていた．

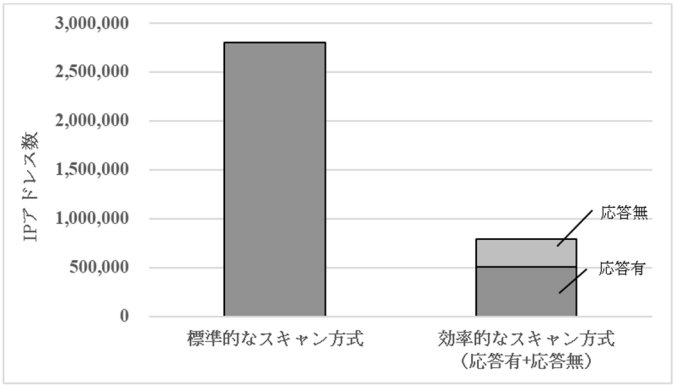


図 7 スキャン試行回数の削減効果

5.3 今後の課題

実ネットワークのスキャンでは，標準的なスキャン方式と効率的なスキャン方式を同時に実施しているにも関わらず，12 時間内でそれぞれ非同期にスキャンを実施しており，全く同じタイミングでスキャンしていないことが原因で取得したポートに差異があった．また，ネットワークの輻輳等によるスキャンの失敗が原因の差異もあった．これらの課題に対し，より理論的に効率的なスキャン方式を評価するためにシミュレータを作成し，上記の課題を排除した評価を検討している．

スキャン頻度に関しては，スキャンを延伸させることにより十分な通信量の削減が実現できたものの，スキャンをスキップすることで脆弱性の検知が遅れるといった問題

があり検討が必要である．過去に発生した脆弱性を狙ったマルウェアなどが脆弱性公表から攻撃までに要した日数等を参考に，最長延伸日数について再考していきたい．

5.4 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 [11]．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が最小限となるよう調整している．

また，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，今後爆発的に増加するであろう IoT 機器に対して，通信量を削減しつつ効率的にスキャンする手法について紹介した．最初に基礎データとして広域ネットワークを対象にした TCP のフルポートスキャンや国内 1.5 億の IP アドレスを対象に応答状況の調査を行った．取得したデータから網羅的にスキャンするための標準的なスキャンポートと標準的なスキャン頻度を検討した．さらに標準的なスキャン方式によるスキャン品質を維持しながら，通信量を削減した効率的なスキャンについて検討し，効率的なスキャンポートと効率的なスキャン頻度を策定した．これらを用いた効率的なスキャン方式を実施したところ，ポート検出率 95%を保ちつつ，8 週間のスキャンで通信量を 3.95%に削減したスキャンを実施することができた．実ネットワークによる課題については今後シミュレータ等を用いながら理論的な検討を進めていきたい．

謝辞

本研究は総務省の「電波資源拡大のための研究開発 (JPJ000254)」における委託研究「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技術の研究開発」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] サイバー攻撃に悪用されるおそれのある IoT 機器の調査、注意喚起を行う NOTICE プロジェクト(https://notice.go.jp/)

[3] 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” Security management 32(1), 39-48, 2018-05

[4] 笠間貴弘，井上大介，“大規模ダークネット観測と能動的スキャンによるマルウェア感染 IoT 機器の分類” 情報処理学会論文誌 58(9), 1388-1398, 2017-09-15

[5] 藤田彬，内田佳介，森博志，吉岡克成，松本勉，“WebUI の画像的特徴に基づく IoT 機器判別手法” 情報処理学会論文誌 60(3), 849-858, 2019-03-15

[6] Masscan (https://github.com/robertdavidgraham/masscan)

[7] 石岡裕,和氣弘明,松下一仁,大崎光洋 ”国内を対象とした広域ネットワークスキャンによる応答特性の調査,”2019 信学技報, vol. 119, no. 297, NS2019-128, pp. 45-49, Nov.2019

[8] 松下一仁，石岡裕，和氣弘明，大崎光洋，“IoT 機器に対する広域ネットワークスキャンのポート最適化技術”，電子情報通信学会 2020 年総合大会講演論文集 B-6-93

[9] 石岡裕，松下一仁，和氣弘明，大崎光洋，“IoT 機器に対する広域ネットワークスキャンの頻度最適化技術”，電子情報通信学会 2020 年総合大会講演論文集 B-6-94

[10] 大崎光洋，種茂文之，和氣弘明，石岡裕，松下一仁，”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，“2019 情報処理学会論文誌，Oct.2019

[11] 広域ネットワークスキャンによる調査について，NTT アドバンステクノロジ株式会社(https://www.ntt-at.co.jp/news/2019/detail/release190709.html)