THE IEICE TRANSACTIONS ON COMMUNICATIONS (JAPANESE EDITION)

電子情報通信学会 B 論文誌

通　信

VOL. J103-B NO. 6
JUNE 2020

通信ソサイエティ

一般社団法人 電子情報通信学会

THE COMMUNICATIONS SOCIETY

THE INSTITUTE OF ELECTRONICS, INFORMATION AND COMMUNICATION ENGINEERS

論　文

大規模IPv6アドレスの収集・分析*

新津　雄大[†a)]　小林　諭[††b)]　福田　健介[††c)]　江崎　浩[†d)]

Large-scale Measurement and Analysis of Active IPv6 Addresses*

Yudai ARATSU[†a)], Satoru KOBAYASHI[††b)], Kensuke FUKUDA[††c)], and Hiroshi ESAKI[†d)]

あらまし　IPv6でのネットワークスキャンは非効率的であると言われているが，近年ではホストに割り当てられたIPv6アドレスを収集して効率的なスキャンのためのアドレスリストを生成する研究が行われている．しかし，それらの研究の多くは収集できるアドレス数の議論に終始しており，アドレス収集に要した時間やどのようなアドレスが集まったかという議論が十分になされていない．本研究では様々な手法で3,800万のアドレスを収集し，その結果から各収集手法の特徴を議論する．各手法で得られたアドレス数と時間の関係を示し，また集まったアドレスをASやホストのタイプで分類を行うことで，手法ごとに集まるアドレスに違いが存在することを明らかにする．

キーワード　IPv6アドレス，ネットワークスキャン，セキュリティ

1. まえがき

IPv6はアドレス空間が非常に広大であることから，その全体スキャンは非効率的であると言われている[1]．32ビットで表現されるIPv4アドレスの理論上の全アドレス数は約4.3×10^9個であり，Zmap [2]やMASSCAN [3]などの高速なツールを用いると，全アドレスのスキャンを一時間以内で終えることができる．一方，IPv6アドレスは128ビットの空間をもつため，約3.4×10^{38}個ものアドレスが存在する．このため全アドレスに対してスキャンを行うには膨大な時間が必要となり，現在の技術では40億年かかってもスキャンは終了しない．また，IPv6にはホストをネットワークに接続するとアドレスが付与されるステートレスアドレス自動設定機能（Stateless Address Autoconfiguration; SLAAC）[4]が備わっているが，これもIPv6におけるスキャンを困難にする一因である．これは，SLAACで生成されるアドレスはインタフェースのMACアドレスや疑似乱数をもとに生成されることから，割り当てられるアドレスを予測することが難しいためである．先行研究でIPv4に比べてIPv6のセキュリティに問題があるホストが多く存在することが示されているように[5], [6]，スキャンが不可能であれば管理すべき対象の把握は難しくなり，セキュリティ管理が困難になる．

このような現状においてIPv6ネットワークの把握やスキャンを可能にするために，インタフェースに割り当てられたアクティブなアドレスを収集する研究がなされている[7]～[11]．しかし，それらの研究では収集できるアドレス数に主眼が置かれており，収集に要する時間や集めたアドレスがどのようなホストのものであるかを考慮していない．そのため先行研究の内容を参考にしてアドレス収集を行う場合，要する期間が予測できない，あるいはスキャンを目的としてアドレスを収集したはずが，サーバ以外のアドレスが大量に集まってしまうといったことが起こる．そこで本研究では様々な手法でアドレス収集を行い，先行研究では重視されていない収集できるアドレス数の経時変化の観測や集めたアドレスの分類を通じて，アドレス収集

† 東京大学大学院情報理工学系研究科，東京都
Guraduate School of Information Science and Technology, The University of Tokyo, 7–3–1 Hongo, Bunkyo-ku, Tokyo, 113–8654 Japan
†† 国立情報学研究所，東京都
National Institute of Informatics, 2–1–2 Hitotsubashi, Chiyoda-ku, Tokyo, 101–8430 Japan
a) E-mail: aratsu@hongo.wide.ad.jp
b) E-mail: sat@nii.ac.jp
c) E-mail: kensuke@nii.ac.jp
d) E-mail: hiroshi@wide.ad.jp
* 本論文は，インターネットアーキテクチャ研究専門委員会推薦論文である．
DOI:10.14923/transcomj.2019JBT0002

手法ごとの比較・分析を行う．また，それぞれの手法で IPv4 と IPv6 のアドレスを並行して集めることで，両者の現在の収集速度の比較結果を示すとともに，更なる IPv6 アドレスの収集について考察を行う．

本研究の貢献は以下のとおりである．(a) アドレス収集手法ごとのアドレス数の経時変化を明らかにする．(b) 集めたアドレスに関して国・AS・アドレスの種類の観点で分析を行う．(c) IPv4 と IPv6 でのアドレス収集の比較を行う．また，集めたアドレスは最終的には倫理に反しない範囲で公開することを考えており，セキュリティ以外の目的にも使用されることで広く貢献できることを期待する．

2. 背景と関連研究

2.1 IPv6 アドレス

IPv6 アドレスの割り当てについて簡単に説明する．IPv6 アドレスのサイズは 128 ビットであり，ほとんどの IPv6 ユニキャストアドレスは上位 64 ビットをプレフィックス，下位 64 ビットをインタフェース識別子 (Interface IDentifier; IID) としてもつ [12], [13]．プレフィックスは基本的にネットワーク管理者がネットワークに対して割り当てる．一方，IID の割り当て方法は大きく分けて次の 3 種類が存在する．(1) ユーザが手動で設定，(2) インタフェースの MAC アドレスをもとに自動生成 (64-bit Extended Unique Identifier; EUI-64 [12])，(3) その他 (Privacy Extensions [14] や Dynamic Host Configuration Protocol for IPv6 (DHCPv6) [15] などの疑似乱数をもとに自動生成されたものを含む)．これらの方法で割り当てられた IID のうち，(1) ユーザが設定した IID と (2) SLAAC により MAC アドレスを基に生成された EUI-64 形式の IID は時間経過により変化することはない．しかし，Privacy Extensions や DHCPv6 により疑似乱数をもとに生成される IID は時間とともに変わりうる．また MAC アドレスを基にした EUI-64 形式の IID は生成の方法から必ず 32 ビット目からの 16 ビットが 16 進数表記で *fffe* となる．

IPv6 アドレスに割り当てられる IID の傾向は，IP アドレスが付与されるホストの用途により傾向が異なる [1]．例えば，ラップトップ PC やスマートフォンのように主にクライアントとして用いられるホストでは MAC アドレスや疑似乱数をもとに自動的に生成された IID が用いられることが多いが，DNS に IP アドレスが登録されるようなサーバ用途のホストではユーザが手動で固定した IID が用いられやすい．

2.2 IPv6 アドレス収集手法

全 IPv6 アドレスに対してスキャンする代わりに，アクティブな IPv6 アドレスを収集して効率的にアドレスのリストを生成する研究がなされている．それらの研究で用いられる手法は大きく分けて 2 種類に分類される．一つ目は IPv6 アドレスを一切もたない状態から，トラヒックデータやインターネット上の公開データ，サーバが提供する情報を用いてアドレスを収集する一次収集手法である．二つ目は事前にシードと呼ばれるアドレスセットを用意しておき，それをもとにして効率良くスキャンや，尤もらしいアドレス候補の生成を行う二次収集手法である．表 1 にそれぞれの代表的な手法をまとめた．以下ではこれらの IPv6 アドレス収集手法について説明する．

2.2.1 一 次 収 集

Gasser [11] らは受動的手法と能動的手法を組み合わせて IPv6 アドレスの収集を行った．受動的手法ではヨーロッパのあるインターネットエクスチェンジポイントとネットワークで観測したトラヒックからアドレスを抽出した．能動的手法では Alexa Top 1 Million domains [18] や Rapid7 の Project Sornar [19] と CAIDA [20] が公開している DNS データセット，トップレベルドメインの DNS ゾーンファイルなどにおける AAAA レコードや PTR レコードを参照することで IPv6 アドレスを得た．またそれに加えて，A レコードなどから得られた IPv4 アドレスから DNS の逆引き・IPv6 の正引きをすることで，IPv4 アドレス，PTR レコード，AAAA レコード，IPv6 アドレスの順でアドレスを収集することも行った．更に traceroute 測定の結果も用いることで，4 週間で 1.5×10^8 個の IPv6 アドレスを収集した．これらのアドレスは，IPv6 の広告されているプレフィックスの 72%，AS の 84%をカバーしていた．

表 1　先行研究の収集対象と集まった IPv6 アドレスの数
Table 1　Summary of past literature.

	Data source	#Addr	
Gasser [11]	traffic, traceroute, DNS AAAA/PTR	150M	
Defeche [16]	BitTorrent peers	1.5M	
	Method	**#Addr (seeds)**	**Alias detection**
Fiebig [17]	rDNS scan	5.8M (73K)	✓
Foremski [7]	Entropy	770K (10K)	
Murdock [10]	Clustering	55M (3.0M)	✓

Defeche [16] らは BitTorrent の IPv6 ネットワークを調査する過程で，ピアの IPv6 アドレスを収集した．2009 年 10 月時点では 1 日で 100,000 ピアを発見し，2011 年 10 月には 1.8×10^5 ピア，2012 年 2 月には 2 日で 6.6×10^5 ピアを発見している．また 2009 年の 5 月から 7 月の間に 5.0×10^6 個のアドレスを発見したが，実際に BitTorrent の接続を確立できたのは 1.5×10^6 個のアドレスであったと報告している．

2.2.2 二次収集

Fiebig [17] らは DNS の逆引きの応答に含まれるステータスを利用することで，シードの IPv6 アドレスをもとに DNS 権威サーバに PTR レコードが登録されている IPv6 アドレスを収集した．DNS ではキャッシュサーバからの応答に NoError や ServFail などのステータスが含まれる [21]．その中で NXDomain ステータスは問い合わせたドメイン名は存在せず，そのサブドメインも存在しないことを表す [22]．IPv6 アドレスの逆引きドメインは *ip6.arpa.* を根とする深さ 32 の木構造をしており，各ドメインは 0 から f までの 16 個のサブドメインをもつ．Fiebig らはこの木構造を上位ドメインから下位ドメインへと再帰的に逆引きを行い探索した．その過程で応答ステータスが NXDomain であったドメインについては，それより下位のドメインは探索しないことで全探索する場合と比較して少ない回数の名前解決で探索した．ただし権威 DNS サーバの設定ミスなどでサブドメインが存在するにもかかわらず NXDomain を返すドメインが存在するため，最上位ドメイン（*ip6.arpa.*）から探索するだけではなく，事前に収集していたシードアドレスの/32，/48，/64 の値を用いて中間ドメイン（/32 の場合は *8.b.d.0.1.0.0.2.ip6.arpa.* など）から探索を行った．この手法により，シードに Alexa Top 1 Million domains [18] や traceroute のデータセットなど一般に公開されているアドレス 7.3×10^4 個を使用し，DNS の逆引きを繰り返すことで最終的に 5.8×10^6 個のアドレスを得た．

機械学習を用いてシードとなるアドレスからホスト割り当てられていそうなアドレスの候補を生成する研究も行われている．Entropy/IP [7] では IPv6 アドレスを 4 ビットごとに区切ったニブルに分割し，各ニブルのエントロピーを計算，更にベイジアンネットワークによりニブル間の依存関係をモデル化することで，シードから尤もらしいアドレスを生成する．この手法により，1000 個のアドレスリスト 10 セットのシードから 1.0×10^7 個の IPv6 アドレスを生成した．生成したアドレスの中には実際に ICMPv6 に応答するアドレスが 7.7×10^5 個含まれていた．

6Gen [10] ではシードからアドレスが密に存在する範囲をスキャンの候補アドレスとすることで，約 3.0×10^6 のシードアドレスから 5.5×10^7 個の応答がある IPv6 アドレスを得た．後述するエイリアス空間を検出する方法を提案した．

以上に紹介した先行研究の中には，非公開のデータが必要なため他者が再現できないものも含まれる．本研究のアドレス収集では，それらの非公開のデータが必要な収集手法を除いた範囲で，先行研究を参考にした手法を用い，収集したアドレスに対して分析することで各手法の比較・検討を行う．

2.3 エイリアス空間

IPv6 では，実際にはホストに割り当てられていない IPv6 アドレスが，外部からはあたかも使用中であるかのように見える領域が存在する．そのような領域をエイリアス空間と呼ぶ．このエイリアス空間は未使用の IPv6 アドレス宛の通信にルータなどがリプライを返したり，領域内の全てのアドレスの DNS の逆引きに対して自動で生成した名前を返すことによって生じる．二次収集では生成したアドレスが使用中であるかを確認するために，そのアドレス宛に ping などの応答を確認する必要があるが，このときに生成したアドレスがエイリアス空間内に存在するものであるか否かを判別する必要がある．

Fiebig [17] らは DNS の逆引きを上位ドメインから下位ドメインに向かって探索していく過程で，/32，/48，/64 に対応するドメインにおいて，それ以下のサブドメインが全て同じであるような，/128 に対応する深さ 32 の最下位ドメインの PTR レコード 16 個（例えば/32 のドメイン *8.b.d.0.1.0.0.2.ip6.arpa.* に対しては *0.0.⋯.0.8.b.d.0.1.0.0.2.ip6.arpa.*，⋯，*f.f.⋯.f.8.b.d.0.1.0.0.2.ip6.arpa.*）の名前解決を行い，三つの PTR レコードについて NoError を受信した場合はエイリアス空間である，というヒューリスティックな手法を用いた．

Gasser [8] らは生成したアドレスを含むある一定のプレフィックスに含まれるランダムな IPv6 アドレスをスキャンすることで，その領域がエイリアス空間であるかの判別手法を提案した．

本研究の二次収集においてもエイリアス空間の検出を行う．

3. アドレス収集手法

本研究で用いた七つの一次収集手法と一つの二次収集手法について説明する．なお，本研究ではインターネット上に公開されている DNS のデータや traceroute の結果などの，容易に取得できて長期収集の必要がないものは用いていない．

3.1 一次収集手法

a） 権威 DNS サーバによる収集

権威 DNS サーバを構築し，問合せ元キャッシュサーバの IPv6 アドレスを収集した．DNS サーバには実験用に取得したドメイン名と IPv6 アドレスの対応が登録してあり，正引き・逆引き共に可能になるように設定した．この状態で他手法のアドレス収集を行うことで，ドメイン名と IP アドレスを外部に知らしめ，名前解決のクエリが送られてくるようにする．また，IPv4 アドレスの逆引きを一つ同じ名前で登録しておくことで，IPv4 アドレスから FQDN を経由して IPv6 アドレスへ到達できるようにした．

b） Web サーバによる収集

実験用サーバに作成した Web ページにアクセスするクライアントの IPv6 アドレスを収集した．また，インターネット上の幾つかの Web ページに，本サーバの Web ページへのリンクを貼り，そこから辿ってくるクローラの IPv6 アドレスの収集も行った．

c） Mail サーバによる収集

Mail サーバを新たに構築し，送られてくるメールのヘッダから送信元の IPv6 アドレスを収集した．より多くのメールが送られてくるように，特にスパムメールを集めることを目指して，メールアドレスを次の三つの方法で広報した．(1) Web ページにメールアドレスを貼付 (4 件)，(2) メーリングリストに登録 (20 件)，(3) スパムメールへ返信 (3 件)．また，実験用サーバのドメイン名を知ったスパムメール送信者が，こちらが想定していないメールアドレス宛にスパムメールを送信してくる可能性も考慮し，実験用ドメイン宛のメールは全て受信するようにエイリアスを設定した．

d） NTP サーバによる収集

実験用サーバで時刻同期用の NTP サーバを構築して，問い合わせ元の IPv6 アドレスを収集した．そして，より多くのクライアントから問い合わせを受信するために，NTP Pool Project [23] にサーバの IP アドレスを登録した．NTP Pool Project は世界各地の NTP サーバをプールした大規模仮想クラスタを提供しており，世界中の数百万から数千万の機器で使用されている．NTP Pool Project の特徴として DNS による分散処理を行っており，またクライアントに地理的に近いサーバが採用されやすい，といった点が挙げられる．この NTP Pool Project に登録することにより構築したばかりの NTP サーバでも多くの問い合わせを受信することができる．ただし，NTP サーバの応答時間が安定しない際にはその NTP サーバは問い合わせ候補から一時的に除外されるため，サーバに届くクエリ数は安定しない．

e） BitTorrent ネットワークの探索による収集

BitTorrent の Peer to Peer (P2P) ネットワークを探索することで，ネットワーク上のピアの IPv6 アドレスを収集した．Defeche [16] らの手法にならい，本研究でもファイルをダウンロード・アップロードすることなくネットワークを探索し，アドレスの収集のみを行った．具体的には各ピアが他のピアの情報を共有する DHT ネットワーク内において，一つのピアを起点にピア情報の要求メッセージを再帰的に送信することでネットワーク上に存在するピアのアドレスを収集した．

f） Bitcoin ネットワーク上のアドレス収集

Bitcoin のネットワーク上のノードの IPv6 アドレスを収集した．Bitcoin ネットワークを実際に探索することはせず，Bitnodes [24] が公開している API を用いた．Bitnodes は Bitcoin ネットワーク上のノードに，他のノードのアドレスを要求するメッセージを再帰的に送るクローラを動かしている．公開されている API はこのクローラによって得られたデータを取得することができる．これを定期的に取得することで Bitcoin ネットワーク上のノードの IPv6 アドレスを取得した．

g） トラヒック観測による収集

データ収集の比較のために，トランジットトラヒックからも IPv6 アドレスの収集を行った．トラヒックは WIDE プロジェクト MAWI traffic repository [25] のデータを利用した．これには毎日 14:00-14:15 の 15 分間のトラヒックが含まれる．

3.2 二次収集手法

a） DNS 逆引き探索による収集

DNS の PTR レコードが登録されている IPv6 アドレスを **2.2.2** で述べた Fiebig [17] らの手法にならい収集した．またエイリアス空間の検知・排除についても **2.3** で述べた Fiebig らの手法に従った．

4. 実験・結果

前節で述べた八つのアドレス収集手法を用い，以下の三つの実験を行った．

（1） 同一環境でのアドレス収集手法の比較
（2） 地理的に離れたサーバにおける収集
（3） IPv4 と IPv6 におけるアドレス収集の比較

本節ではそれぞれの実験に関して，その目的・内容を述べ結果を示していく．

4.1 アドレス収集手法の比較

まず，前節で述べた各アドレス収集手法を比較するために，一つの仮想マシン上に全ての手法を実装し，アドレスを収集した．ただし，Mail サーバでの収集できたアドレスは Gmail による Google のアドレス四つのみであったため，以降では特に言及しない．アドレス収集は 2018 年 6 月より開始した．

4.1.1 実験結果

各手法で収集したユニークな IPv6 アドレス，/64 プレフィックス，AS 番号の数を表 2 に示す．各手法でアドレス収集期間が異なるため，単純に数値比較をすべきでないことに注意が必要である．約 400 日前後の収集により，一つの仮想マシンでの収集で 3,800 万の IPv6 アドレスを集めることができた．またこれらのアドレスは 7,369 の AS をカバーしている．2019 年 6 月時点で IPv6 を広告をしている AS 数は 17,119 であるため，この結果はそれらの AS の 4 割以上をカバーしていることになる．

表 2 各手法で収集した IPv6 アドレス，/64 プレフィックス，AS 数
Table 2 Collected IPv6 addresses, /64 prefixes and ASes.

	Period	#Address	#prefix/64	#AS
DNS	425 days	12.8K	1.7K	786
Mail	354 days	4	3	1
Web	352 days	984	793	152
NTP	393 days	1.8M	1.8M	576
BitTorrent	410 days	28M	16M	2,963
Bitcoin	385 days	27K	16K	618
Traffic	420 days	2.2M	1.3M	6,695
rDNS	55 days	7.5M	118K	582
Total		40M	20M	12,370
Unique		38M	19M	7,369

次に，表 3，表 4 は各手法で得られたアドレスが属する国，AS の上位五つとその割合を示したものであり，それぞれ一次収集，二次収集の結果を表している．

そして表 5 は収集したアドレスを IID の種類ごとの割合である．ここでは IID の種類として (1) 16 進数表記にして 4 桁ずつ区切った際に “0000” を含むもの，(2) 25 ビット目から 42 ビット目が “fffe” であるもの，(3) それ以外，の 3 種類を考える．これらは **2.1.1** で述べた (1) ユーザが手動で設定した IID，(2) MAC アドレスをもとに生成されたの IID，(3) その他の方法で生成された IID に対応している．前者二つは基本的に短時間に値が変化することはなく，その他の方法で生成された IID に時間とともに値が変化するものが含まれる．

最後に，各手法での収集したユニークなアドレス数と，それらのアドレスを/64，/48，/32 のプレフィックスで区切ったもののユニークな個数の累積数を表したものを図 1 に示す．この図は横軸が収集開始日からの経過日数，縦軸がその日までに収集したアドレス・プレフィックスの累積数である．ただし DNS 逆引き探索による収集手法は，用いたシードの数・種類や実装の並列数によって，アドレスの収集速度が大きく異なるため，図 1 には含まない．

表 3 一次収集により集まったアドレスの国・AS 上位 5 位
Table 3 Top 5 countries and ASes of addresses. (primary collection)

DNS

Country		AS	
US	47.0%	Cloudflare (AS13335)	23.9%
TW	18.3%	HINET (AS34625)	17.9%
JP	8.1%	Google (AS15169)	16.6%
CN	3.0%	OpenDNS (AS36692)	3.1%
IE	2.5%	AT&T (AS7018)	2.1%

Web

Country		AS	
US	31.4%	Amazon (AS14618)	10.5%
BR	8.3%	Google (AS15169)	7.6%
FR	6.1%	Comcast (AS7922)	5.0%
MY	5.2%	CLARO (AS28573)	4.3%
JP	5.1%	TM Net (AS4788)	3.8%

NTP

Country		AS	
IN	52.7%	RelianceJio (AS55836)	37.6%
SA	15.8%	Saudinetsrc (AS25019)	15.8%
JP	10.1%	KDDI (AS2516)	7.1%
VN	5.7%	VodafoneIndia (AS38266)	4.9%
CN	5.3%	VNPT (AS45899)	4.8%

BitTorrent

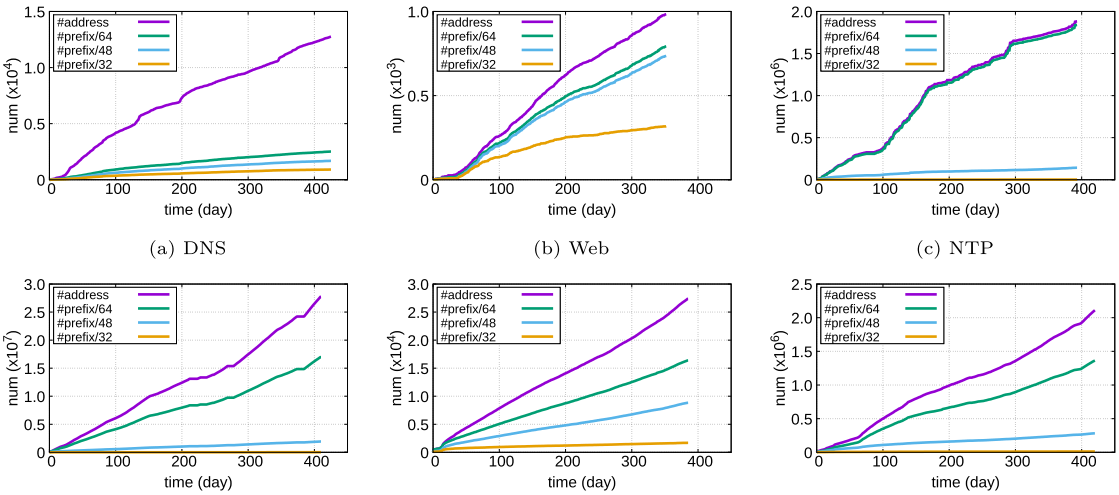
Country		AS	
US	39.3%	T-mobile (AS21928)	13.8%
IN	11.6%	RelianceJio (AS55836)	9.5%
CN	8.4%	Comcast (AS7922)	6.8%
JP	7.5%	HINET (AS3462)	4.5%
RU	5.3%	Comcast (AS7725)	3.7%

Bitcoin

Country		AS	
US	16.0%	Proxad (AS12322)	6.3%
DE	15.4%	Comcast (AS7922)	5.7%
CN	10.3%	Swisscom (AS3303)	4.5%
FR	8.7%	Hetzner (AS24940)	3.8%
BR	5.3%	DTAG (AS3320)	3.6%

Traffic

Country		AS	
US	35.5%	T-mobile (AS21928)	11.5%
JP	11.0%	HINET (AS3462)	9.9%
TW	10.5%	Softbank (AS17676)	5.8%
CN	8.5%	Comcast (AS7922)	4.7%
IN	5.2%	RelianceJio (AS55836)	4.6%



(a) DNS　(b) Web　(c) NTP

(d) BitTorrent　(e) Bitcoin　(f) Traffic

図 1　手法ごとの収集したアドレス数の時間的推移
Fig. 1　Growth of the number of collected addresses.

表 4　二次収集により集まったアドレスの国・AS 上位 5 位
Table 4　Top 5 countries and ASes of addresses. (secondary collection)

rDNS			
Country		AS	
CZ	55.3%	Casablanca (AS15685)	55.0%
US	22.2%	Comcast (AS7922)	18.4%
BY	7.4%	BELPAK (AS6697)	7.4%
JP	3.8%	interQ (AS63949)	3.7%
DE	3.6%	Linode (AS13030)	3.6%

表 5　収集したアドレスの IID タイプ別割合
Table 5　IID types of collected addresses.

Type	DNS	Web	NTP	BT	BC	Traffic	rDNS
(1) “0000”	90.4%	19.1%	21.9%	13.0%	20.2%	19.3%	91.3%
(2) “fffe”	2.8%	0.8%	3.1%	6.5%	11.8%	12.4%	1.1%
(3) Others	6.9%	80.1%	75.0%	80.5%	68.1%	68.4%	7.5%

以下ではこれらの収集結果に関して，各アドレス収集手法特有の経時変化の傾向や収集したアドレスの特徴に重点を置いて考察する．本研究では各アドレス収集手法間の特徴を比較し，手法ごとに特有の傾向を観測すること目的としているため，全ての手法に共通して見られる傾向については図・表に示すのみに留め，詳しい言及はしないものとする．そのため，各手法の考察に置いて言及する観点が異なることに注意されたい．

a）　権威 DNS サーバによる収集

権威 DNS サーバへの問い合わせ元の IPv6 アドレスを集めたが，他の手法の結果と比べると，集まったアドレスに対する /64 プレフィックス数と AS 番号数が少ない．これは集まったアドレスの多くが自動的に DNS の問い合わせを行っているクローラであり，幾つかのクローラが一定時間ごとに IID を変化させながら問い合わせを行っていたためである．表 5 の “0000” を含む IID が多いのも同様に，クローラのアドレスが “0000” をもっていたためである．また図 1 (a) では 20 日を過ぎたところでアドレス数の増加率が大きくなっているが，これはこの日から新たなクローラに巡回されるようになったためである．これと同様に他の今後も他のクローラが問い合わせをしてくるようになると，更にアドレス収集速度が上がることが期待される．

b）　Web サーバによる収集

Web サーバではアクセスしてくるクライアントのアドレスを収集したが，アドレスは権威 DNS サーバと同様に Web サイトを自動的に巡回しているクローラのものが主であった．しかし Web サーバのアクセスログを確認すると，存在しない “wp-login.php” ページへのスキャンと見られるアクセスが確認でき，DNS サーバとは異なり企業が運用するクローラだけではなく ISP が顧客に払い出していると思われるアドレスが多く収集された．それらのアドレスの IID は疑似乱数をもとに生成されたように見え，それが表 5 の DNS と Web のアドレスタイプの構成割合の違いに表れた．

クローラからのアクセスはドメイン名を指定してア

クセスしてくるもの，IP アドレスからアクセスしてくるものの双方が観測された．また，web 上の様々な場所に設置したリンクを辿ってアクセスしてくるクローラも現れた．図 1 (b) に示すように，収集開始直後は IPv6 アドレスの集まりが少ないが，1ヶ月程度経った頃からアドレス数が伸び始めるようになった．これは DNS サーバと同様に新たなクローラからアクセスを受けるようになったためであり，同じクローラがアドレスを変えてアクセスしてくることも考慮すると，今後もアドレス数は増加していくと予想できる．

c） NTP サーバによる収集

NTP サーバでは時刻の問い合わせをしてくるクライアントのアドレスを収集したが，表 3 を見ると，収集したアドレスの半数以上がインドのものであった．特に AS55836 (RelianceJio) に割り当てられたアドレスが多く集まっている．この RelianceJio はインドの通信会社で，スマートフォンなどのモバイル回線を運用している企業である．APNIC の調査 [26] によると，2019 年 8 月時点でこの AS は全 AS の中で二番目にユーザ数が多いとされている．そのため，それらのスマートフォンのアドレスが多く集まったのではないかと思われる．

また，NTP サーバでは NTP Pool Project の仕様により，サーバが存在する日本に近いアジア圏の国のアドレスが多く集まっている．サーバが存在している国に地理的に近接している国のアドレスが収集できると考えると，国や AS 番号の観点でより多様なアドレスを集めるためには，世界各地にサーバを構築し，それぞれのサーバでアドレスを集めれば良いと言える．

d） BitTorrent ネットワークの探索による収集

BitTorrent ネットワークのクロールは今回実装した手法の中で一番多くアドレスを収集することができた．図 1 (d) を見るとこの手法で得られる収集アドレス数はおおむね収集時間に比例していることがわかる．また，アドレスだけでなく /64，/48 プレフィックスの数も，一年以上経過した現在でも増加し続けている．しかし BitTorrent ネットワークから得られたアドレスに対して ICMPv6 echo request を送信してみると，reply を返すピアの数は，常にほぼ一定のままであった．これはノードが IPv6 アドレスを一定期間ごとに変更していることによるためと考えられる．このことは表 5 において，時間とともに変化するアドレスが含まれる Others が多いことからも確認できる．

また，アドレス数の増加に伴いプレフィックス数も増加を続けていることから，BitTorrent のピアは IID だけではなく，プレフィックスも変化しているものと考えられる．これは，集めたアドレスの中に IID が共通の EUI-64 形式で，プレフィックスが異なるアドレスが多数存在していたことからも，確認することができた．アドレスのプレフィックスが変化する原因として，家庭などにプレフィックスを割り当てている ISP が一定期間ごとにその割り振りを変更していることや，ピアがラップトップ PC などのもち運び可能なホストで，様々なネットワークで BitTorrent ネットワークに参加していることが考えられる．

特殊な例として，同一の EUI-64 形式の IID をもつアドレスだが異なるネットワークに割り当てられていたアドレスが観測された．これは MAC アドレスが固定された仮想マシンのイメージを複数のホストマシンで可動した場合などに見られる．このように IID が同一のアドレスがあっても，それが全て一つのホストのアドレスであるとは限らない．

e） Bitcoin ネットワーク上のアドレス収集

本研究では Bitcoin については既に動いているクローラのデータを API で参照したため，収集開始当初からまとまった IPv6 アドレスを手に入れることができた．しかし，その後のアドレス収集速度は他の手法と比べると遅い．これは同じ P2P 通信のクライアントである BitTorrent のピアとは違い，Bitcoin ネットワークのピアは頻繁に IP アドレスを変えないためと考えられる．収集の過程から得られた情報ではこのことを裏付ける情報を得られなかったため，BitTorrent と Bitcoin で集めたアドレスについては，追加として ICMPv6 echo request の送信を試みた．その結果，ある日に応答があったアドレスのうち一週間後も応答があったアドレスの割合は，BitTorrent のピアのアドレスでは 70%であったのに対して，Bitcoin のノードのアドレスでは 92%であった．また，表 5 をもとに比較すると，BitTorrent のピアのアドレスは約 80%がその他の IID に分類されたが，BitCoin で得られたアドレスは 20%以上が “0000” を含む，手動で設定したと思われる IID であった．これらのことから，上述のとおり Bitcoin のネットワーク上に存在しているアドレスは BitTorrent ネットワークのピアのアドレスと比較して変化しにくいことが推測される．同一ホストの異なるアドレスをどのように解釈するかは本研究の焦点からは外れるため，ここでは別のアドレスとして数え，そのような傾向が見られたと述べるに留める．

f） トラヒック観測による収集

トラヒックで得られたIPv6アドレスは表2からわかるように，数量では他の手法で得られたものより少ないものの，多くのASのアドレスを集めることができた．更に表3のように，特定のASのアドレスに偏ること無く収集することができた．

g） DNS逆引き探索による収集

今回の実験ではBitTorrentネットワークの探索によって得られた/32プレフィックス2,526個を含む40Kのアドレスをシードとして用いた．そこから本手法で得られたアドレス数が5.5Mであるから，シードから約137倍のアドレスを収集することができた．シードアドレスの個数，集まったアドレスの個数とその過程での集まったプレフィックスの個数を，Fiebig [17] らの結果の一つと比較した結果が表6である．本研究で用いたシードの個数はFiebigら用いたシードの30分の1程度でありながら，Fiebigらと同程度の数のアドレスを収集することができた．これはFiebigらが公開サーバのアドレスを用いているのに対し，本研究では主に家庭のクライアントPCのアドレスを用いているという違いに由来する．

この収集手法においては **2.3** で述べたエイリアス空間が観測された．表6においてFiebigらの収集結果では/48プレフィックス数に対し/64プレフィックス数が少ない．本研究の結果においても/32プレフィックス数に対して/48プレフィックス数が17倍であるのに対し，/48プレフィックス数に対して/64プレフィックス数が2.5倍であり，比率に変化が生じている．これらはエイリアス空間を除去したことに由来する．

本研究におけるDNSの逆引きで収集したアドレスのICMPv6応答率はごくわずかであった．このことから，収集したアドレスには未検出のエイリアス空間に属するものが存在していると考えた．収集したアドレスのうち，表4において多数を占めているAS15685 (Casablanca) のアドレスについて調査すると，特定のアドレス空間において機械的な逆引きを行っている挙動が見られた．よって，DNSの逆引きによる収集ではFiebigらの手法では検出できないエイリアス空間が存在し，本研究で収集アドレスにはそれらが含まれる結果となった．

収集時間については，実装したFiebigらの手法は幅優先探索と深さ優先探索を組み合わせた手法が使われており，シードの数にも依るが実際にIPv6アドレスが得られ始めるまでに，今回は20日間近くを要した．そのため本手法は短期間でアドレスを収集したい場合は不向きな手法であるといえる．

4.1.2 各収集手法の比較

各手法で集まるアドレスの種類を整理する．収集したアドレスのIIDを示した表5と各手法で収集対象とするホストの性質を考慮すると，各手法で主として集まるアドレスの種類は表7に示すとおりとなる．ここでの分類は **2.1.1** での分類に対応しており，(3) Othersには Privacy Extensions や DHCPv6 など疑似乱数をもとに生成されたアドレスを含む．これを踏まえて幾つかのケースにおける最適なアドレス収集手法について考える．

まず，セキュリティ向上を目的として，スキャン対象とするアドレスリストを生成したい場合，定期的にアドレスが変化しないサーバのアドレスを集めることを考える．BitTorrentネットワークの探索などのクライアントPCのアドレスが集まる手法は適しておらず，今回試した手法の中ではそのため表7を参考にして，DNSサーバやBitcoinネットワークの探索，DNSの逆引き探索による収集が適していることがわかる．次に，二次収集手法の開発など実験に用いるために様々なASに含まれるアドレスリストが必要な場合は，表3を参考に，BitTorrent・Bitcoinネットワークの探索やトラヒックからの抽出による収集を行えば良い．他にも，種別を問わず多くのIPv6アドレスが必要な場合は図1より，BitTorrentネットワークの探索を行うのが最適であるといえる．このようにアドレス収集の目的によって異なる収集手法を用いることで，効率良くアドレスを集めることができる．

4.2 海外サーバにおけるアドレス収集

次に，地理的に離れた複数のサーバで収集を行った際の収集できるIPアドレスの数や重複具合を確認する

表6 rDNS enumeratoinの先行研究との比較
Table 6 Comparison of rDNS enumeration.

	Seed			Address
	# /32	# /48	# /64	# /128
Fiebig [17] (80 Parallel)	73k	856k	582k	5.3M
this study (10 Parallel)	2.5k	42k	144k	5.5M

表7 手法ごとに集まるアドレスの種類
Table 7 Collected Address type.

	DNS	Web	NTP	BT	BC	Traffic	rDNS
(1) **Manual**	✓	✓	✓		✓	✓	✓
(2) **SLAAC**		✓	✓	✓	✓	✓	
(3) **Others**		✓	✓	✓	✓	✓	

ために，日本の他にイギリス・アメリカに設置した仮想マシンに NTP サーバでの収集環境と BitTorrent の探索環境を構築し，期間を合わせて各国で同時に収集を行った．実験期間は NTP サーバでの収集が 2019 年 5 月 10 日の 1 日間，BitTorrent の探索が 2018 年 12 月 14 日からの 28 日間である．なお，NTP と BitTorrent 以外の手法については，DNS と Web ではクライアントが IP アドレスまたはドメイン名をもとにアクセスしてくるため地理的な影響がなく，Bitcoin に関してはクローラを運用している組織の探索結果を利用しており，いずれもサーバのロケーションによる結果の変化は考えにくいため，海外サーバでの収集は行わなかった．

4.2.1 海外での NTP サーバでの収集結果

各国で収集できた IPv6 アドレス数・/32 プレフィックス数と，その国のサーバのみで集まったそれぞれの個数を表 8 (a) に示す．

NTP サーバによる収集では，サーバが位置する国ごとに異なるアドレスが集まり，/32 プレフィックスの重複は少数であった．この結果は，NTP Pool ではクライアントが明示的にどの国の NTP サーバを利用するか指定しない限り地理的に近い NTP サーバが使用される，という仕様の影響が大きいと考えられる．それを確認したのが，表 8 (b) である．この表は NTP サーバを構築した国ごとに，集まったアドレスが属する国の上位 5 カ国を示したものである．これにより NTP クライアントは地理的に距離が近いサーバを利用していることがわかる．このように NTP Pool では収集されるアドレスが収集を行う国に依存することから，より多くの国のアドレスを収集するには複数の国での収集が必要となる．

4.2.2 海外での BitTorrent 探索による収集結果

BitTorrent のクローラをアメリカ・イギリスに用意し，それぞれのクローラで日本でおこなったものと同様の DHT ネットワーク上のピアのアドレス収集を行った．その結果を表 9 に示す．

複数サーバでの BitTorrent の探索で集まったアドレスは，NTP での結果と比べて特に/32 での重複の割合が多い結果となった．また集まったアドレスの国ごとの内訳を見ると，どの国で集めたアドレスも類似の内訳となっており，海外にクローラを設置したことによる集まるアドレスの多様性の向上はみられないと考えられる．ただし，ユニークなアドレスの数は少なくないため，国や AS の多様性が重要でない場合には複数のクローラを用意することは効果的である．

4.3 IPv4 での収集との比較

IPv6 と IPv4 でのアドレス収集の比較を行うために，幾つかのアドレス収集手法を期間を合わせて IPv4 と IPv6 で同時に行った．用いた収集方法は DNS・Web・NTP サーバにおける収集，BitTorrent・Bitcoin ネットワークの探索による収集である．収集は全て国内で行い，各手法で IPv4 と IPv6 の収集は同一ホスト上の同一環境で行った．

4.3.1 実 験 結 果

集まったアドレス数と AS 数を表 10 に示す．なお，2019 年 6 月時点で経路広告をしている AS 数は 65,036 であり，そのうち IPv6 の経路広告している AS 数は 17,119 である [27]．IPv4 と IPv6 を比較したときに，どの手法でも収集できたアドレス数が IPv4 のほうが多くなっている．これは，インターネットプロトコルの主流が未だ IPv4 であり，ユーザ数以外にもクロー

表 8 海外 NTP サーバにでの収集結果
Table 8 IPv6 address collection on NTP servers in three countries.

(a) The number of collected addresses

	#Addr	#uniqAddr	# /32	#uniq /32
JP	1.8K	1.8K	92	51 (55.4%)
UK	193	187	49	34 (69.4%)
US	8.3K	8.3K	398	360 (90.5%)

(b) Top 5 break down of source countries

JP		UK		US	
CN	63.97%	DE	41.97%	US	99.28%
JP	15.28%	GB	36.79%	BR	0.35%
OM	6.56%	IN	6.22%	PA	0.14%
TH	5.82%	BR	5.70%	IN	0.13%
VN	5.07%	US	4.66%	GT	0.02%

表 9 海外 BitTorrent クローラでの収集結果
Table 9 IPv6 address collection by BitTorrent in three countries.

(a) The number of collected addresses

	#Addr	#uniqAddr	# /32	#uniq /32
JP	1.8M	1.2M	6,515	991 (15.2%)
UK	1.1M	0.5M	6,070	336 (5.5%)
US	136K	21K	3,967	17 (0.4%)

(b) Top 5 break down of source countries

JP		UK		US	
US	48.49%	US	52.76%	US	50.19%
JP	8.26%	JP	8.20%	JP	7.32%
IN	7.48%	TW	6.39%	FR	6.41%
TW	5.87%	IN	4.50%	TW	6.18%
RU	5.40%	RU	4.21%	IN	4.47%

表 10　IPv4・IPv6 の収集結果
Table 10　Collected IPv4 and IPv6 addresses.

		IPv4		IPv6	
	Period	#Address	#AS	#Address	#AS
DNS	336 days	31.4K	3,053	12.0K	735
Web	352 days	20.3K	3,639	1.0K	154
NTP	20 days	1.3M	17,640	40.1K	195
BitTorrent	25 days	7.4M	23,176	2.6M	1,968
Bitcoin	289 days	118K	2,991	19.6K	548
Traffic	20 days	70.7M	41,606	0.2M	3,346

ラやスキャナの数が IPv4 の方が多いからであると考えられる．AS 数を考慮すると，例えば BitTorrent の探索で集めたアドレスの AS 数が IPv4 では IPv6 の 10 倍以上であるのに対し，アドレス数は 3 倍程度であるなど，IPv4 では AS 数の割に IP アドレスが集まっていない．これは，単に IPv6 に対応している AS では IPv6 ホストが多いというだけでなく，IPv6 アドレスが変化するホストの影響もあると考えられる．そのため，IPv6 ではアドレス数に対してホスト数が少ない可能性があり，逆に IPv4 では NAT の存在などによりアドレス数に対してホスト数が多い可能性がある．

また，収集したアドレス数の時間的変化を確認すると IPv4 でも IPv6 と同様にどの手法でも頭打ちは確認できず，ホスト数を考慮しない単純なアドレス数は増加を続けていくことが推測でき，更に長期の観測が必要である．

5. 実験に関する倫理

一般に，ネットワークに対するスキャン行為は攻撃と判断されることがあり，本研究で行う実験にも，外部のサーバに対して大量のパケットを送信するものが含まれる．そこで，本研究では外部のサーバに対して高い負荷をかけないために，先行研究で配慮されている手法はそれに倣い，その他の手法に関しても最新の注意を心掛けた．実験を行うホストには Web サーバを構築し，スキャンを行っている目的と連絡先を明記し，もしスキャン対象から除外する要求があれば，適切に対応を行う．また，集めたアドレスを研究目的以外には使用しない．

6. むすび

本研究では，まず先行研究を含む八つの IPv6 アドレス収集手法を国内の一つの仮想マシン上に実装してアドレス収集を行い，3,800 万のアドレスを得た．その際の手法ごとのアドレス数の増加を示すことで，それぞれの手法では集まるアドレスの速度が大きく異なること，いずれの一次収集手法でも一年間の収集ではアドレス数が頭打ちにならないことを示した．そして，収集したアドレスの分析を行い，手法ごとの集まるアドレスの種類や国・AS に違いがあることを明らかにし，海外でのアドレス収集が有用であることを確認した．また IPv4 で同様の収集を行い，収集できるアドレス数が現在では IPv4 が勝っていること，そして IPv4 においてもアドレス数が収束していないことから，IPv6 における更なる収集の余地を確認した．

今後は今回用いたアドレス収集手法の更なる長期的な結果を調べるとともに，他の収集手法についても比較・検討をしていく．海外での収集を効率的に行うに当たり，効率的なアドレス収集環境の構築を行うためには，実験に用いるサービス・クローラを一つの仮想マシンやコンテナにパッケージ化することが必要となるが，これも今後の課題である．また，集めたアドレスを用いて IPv6 ネットワークに関する詳細な測定・実験や IPv4 と IPv6 のアドレス収集割合の変化についても検討していく．

謝辞　本研究は JSPS 科研費 18H03237 の助成を受けたものである．

文　献

[1] F. Gont and T. Chown, “Network Reconnaissance in IPv6 networks,” RFC7707, 2016.
[2] Z. Durumeric, E. Wustrow, and J.A. Halderman, “ZMap: Fast internet-wide scanning and its security applications,” Proc. USENIX Security Symposium, pp.605–619, 2013.
[3] R.D. Graham, “MASSCAN: mass IP port scanner,” https://github.com/robertdavidgraham/masscan. accessed Sept. 5, 2019.
[4] S. Thomson, T. Narten, and T. Jinmei, “IPv6 stateless address autoconfiguration,” RFC4862, 2007.
[5] J. Czyz, M. Luckie, M. Allman, and M. Bailey, “Don’t forget to lock the back door! A Characterization of IPv6 network security policy,” Proc. ISOC NDSS’16, pp.21–24, 2016.
[6] K. Fukuda and J. Heidemann, “Who knocks at the IPv6 door? Detecting IPv6 scanning,” Proc. ACM IMC’18, pp.231–237, 2018.
[7] P. Foremski, D. Plonka, and A. Berger, “Entropy/IP: Uncovering structure in IPv6 addresses,” Proc. ACM IMC’16, pp.167–181, 2016.
[8] O. Gasser, Q. Lone, Q. Scheitle, M. Korczyński, P. Foremski, S.D. Strowes, L. Hendriks, and G. Carle, “Clusters in the expanse: Understanding and unbiasing IPv6 hitlists,” Proc. ACM IMC’18, pp.364–378, 2018.

[9] K. Borgolte, S. Hao, T. Fiebig, and G. Vigna, “Enumerating active IPv6 hosts for large-scale security scans via DNSSEC-signed reverse zones,” Proc. IEEE SSP’18, pp.770–784, 2018.
[10] A. Murdock, F. Li, P. Bramsen, Z. Durumeric, and V. Paxson, “Target generation for internet-wide IPv6 scanning,” Proc. ACM IMC’17, pp.242–253, 2017.
[11] O. Gasser, Q. Scheitle, S. Gebhard, and G. Carle, “Scanning the IPv6 internet: Towards a comprehensive hitlist,” Proc. IFIP TMA’16, p.9, 2016.
[12] R. Hinden and S. Deering, “IP version 6 addressing architecture,” RFC4291, 2006.
[13] B. Carpenter, T. Chown, F. Gont, S. Jiang, A. Petrescu, and A. Yourtchenko, “Analysis of the 64-bit boundary in IPv6 addressing,” RFC7421, 2015.
[14] T. Narten, R. Draves, and S. Krishnan, “Privacy extensions for stateless address autoconfiguration in IPv6,” RFC4941, 2007.
[15] T. Mrugalski, M. Siodelski, B. Volz, A. Yourtchenko, M. Richardson, S. Jiang, T. Lemon, and T. Winters, “Dynamic host configuration protocol for IPv6 (DHCPv6),” RFC8415, 2018.
[16] M. Defeche and E. Vyncke, “Measuring IPv6 traffic in BitTorrent networks,” Internet-Draft, 2012.
[17] T. Fiebig, K. Borgolte, S. Hao, C. Kruegel, and G. Vigna, “Something from nothing (There): Collecting global IPv6 datasets from DNS,” Proc. PAM’17, pp.30–43, 2017.
[18] “Alexa Top 1,000,000 sites,” http://www.alexa.com/topsites. accessed Sept. 5, 2019.
[19] “Rapid7 Soner Project: Open Data,” https://opendata.rapid7.com. accessed Sept. 5, 2019.
[20] Center for Applied Internet Data Analysis, “CAIDA IPv6 DNS Names Dataset,” http://www.caida.org/data/active/ipv6_dnsnames_dataset.xml.
[21] D. Eastlake, “Domain name system (DNS) iana considerations,” RFC6895, 2013.
[22] S. Bortzmeyer and S. Huque, “NXDOMAIN: There Really is nothing underneath,” RFC8020, 2016.
[23] “pool.ntp.org: the internet cluster of ntp servers,” https://www.pool.ntp.org. accessed Sept. 5, 2019.
[24] “Global Bitcoin Nodes Distribution,” https://bitnodes.earn.com. accessed Sept. 5, 2019.
[25] “WIDE MAWI WorkingGroup,” http://mawi.wide.ad.jp. accessed Sept. 5, 2019.
[26] APNIC, “Customers per AS Measurements,” https://stats.labs.apnic.net/aspop/. accessed Sept. 5, 2019.
[27] RIPE NCC, “IPv6 Enabled Networks,” http://v6asns.ripe.net. accessed Sept. 5, 2019.

（2019 年 9 月 13 日受付，12 月 3 日再受付，
2020 年 2 月 21 日早期公開）

新津　雄大

2018 東大・工・電子情報卒．現在，東大大学院情報理工学系研究科修士課程．IPv6 に関する研究に従事．

小林　　諭

2018 東大大学院情報理工学系研究科博士課程了．博士 (情報理工学)．現在，国立情報学研究所特任研究員．ネットワーク運用情報からの知識抽出に関する研究に従事．

福田　健介　（正員）

1999 慶大大学院理工学研究科計算機科学専攻博士課程了．博士 (工学)．現在，国立情報学研究所アーキテクチャ科学研究系准教授．インターネット及びネットワークセキュリティに関する研究に従事．

江崎　　浩　（正員）

1987 九州大学大学院・工・電子修士課程了．同年 (株) 東芝入社．1990 米国ニュージャージ州ベルコア社．1994 コロンビア大学・客員研究員．1998 東京大学大型計算機センター・助教授．2001 同大学大学院・情報理工学系研究科・助教授．2015 同大学大学院・同研究科・教授，現在に至る．博士 (工学，東京大学)．MPLSJAPAN 代表，IPv6 普及・高度化推進協議会専務理事，WIDE プロジェクト代表，JPNIC 副理事長．