3-3 Spoofing対策技術の研究―IPアドレス詐称とフィッシングサイトの事例について―

3-3 Studies on Countermeasures for Spoofing Attacks—Cases of IP Address Spoofing and Web Spoofing—

宮本大輔　櫨山寛章　門林雄基
MIYAMOTO Daisuke, HAZEYAMA Hiroaki, and KADOBAYASHI Youki

要旨

本稿ではSpoofingとよばれる一連の攻撃技術に対する研究事例を報告する。現在、サイバー犯罪を行う攻撃者たちは、その犯罪を成功させるためにSpoofing技術を広く用いている。例えばDoS（サービス運用妨害）攻撃の場合には、IP Address Spoofing（アドレス詐称）が用いられる。アドレス詐称とは、IPパケットの送信元IPアドレスを詐称して別のIPアドレスになりすまし、そのパケットの真の発信源を隠蔽する行為である。また、ソーシャル・エンジニアリングの場合では、Web Spoofingと呼ばれる、本物そっくりに作った偽のウェブサイトを用いた攻撃が行われる。このような偽のウェブサイトはフィッシングサイトと呼ばれ、犯罪者はおびき寄せられた被害者に対し、個人情報を入力させるよう促し、入力された情報を盗み取る。

我々の研究グループでは、このようなSpoofing技術を用いたサイバー犯罪を分析し、これに対するシステムやアルゴリズムを開発している。DoS攻撃の対策として、IPトレースバックシステムの普及を目的とし、その導入シナリオについて考察を行う。IPトレースバックシステムとは、たとえアドレス詐称を行っていたとしても、問題のパケットの発信源を特定する仕組みである。ただし、本システムの追跡の成功率を高めるためには、システムがネットワークに広く導入される必要がある。そこで、日中韓3カ国のインターネットのトポロジについて、どのような導入シナリオが効果的であるかを議論する。

また、フィッシングサイト対策として、HumanBoostと名付けた検知方式を提案する。フィッシングサイトの判別の課題は、検知精度を高めることにある。HumanBoost方式は、機械学習の一手法であるAdaBoostを用いた検知方式を拡張したものであり、ユーザがこれまで行ってきた過去の判断履歴をフィッシングサイト検知に活用するという特徴がある。この方式の有効性を調査するため、被験者を募ったフィッシングサイト判別実験を行い、HumanBoost方式、AdaBoost方式、そして被験者毎の判別誤り率の平均値を比較評価する。

This article intends to give case studies of mitigating spoofing techniques. Spoofing techniques are widely used while attackers intend to increase the successful rate of their cyber crimes. In the context of Denial of Service attacks, IP address spoofing is maneuver for camouflaging the attackers' location. It the context of social engineering attacks, Web spoofing is used to persuade victims into giving away personal information; Web spoofing is such deception technique that an attacker creates a convincing but false copy of the legitimate enterprises' web sites. These websites are also known as phishing sites.

Our research group developed the algorithms, systems, and practices all of which analyze cyber crimes based on spoofing techniques. To thwart DoS attacks, we show the deployment scenario for IP traceback systems. IP traceback aims to locate attacker source, regardless of the spoofed source IP addresses. Unfortunately, IP traceback requires large-scale deployments. We argue the practical deployment scenario for the Internet of China, Japan, and South Korea.

We also develop a detection method for phishing sites, named HumanBoost. A current challenge of the detection methods is increasing the detection accuracy. HumanBoost aims at

improving AdaBoost-based detection methods by utilizing Web users' past trust decisions. Based on our subject experiments, we compared the average of the detection accuracy for HumanBoost, AdaBoost-based detection method, and the cases of each participant.

[キーワード]
IPアドレス詐称，フィッシングサイト，模倣インターネット，IPトレースバック，機械学習
IP spoofing, Web spoofing, Internet emulation, IP traceback, Machine learning

1 はじめに

DoS（サービス運用妨害）攻撃は、ホストやネットワークの資源を枯渇させ、正規のユーザがこのような資源に接続することを妨害する攻撃である。とりわけ、Floodingと呼ばれるDoS攻撃が猛威をふるっているが、この攻撃ではIP Address Spoofing（IPアドレス詐称）技術が用いられることが多い。IPアドレス詐称とは、IPパケットの送信者が本来用いるべきではないアドレスを送信元IPアドレスとしてパケットに設定する技術である。この技術により、当該パケットの真の送信元が隠蔽されるため、探知することが困難とされていた。IPトレースバック技術は、IPアドレス詐称をされていたとしても、パケットの真の送信元を探知する技術である。過去にいくつかのIPトレースバック技術[1]-[3]が提案されているが、とりわけSource Path Isolation Engine（SPIE）[3]がパケットを追跡する際に最も有力ではないかと考えられている。ただし、SPIEは追跡の達成率（追跡性）を高める際に、ネットワークの広範囲にSPIEの導入を必要としている。

一方、フィッシング攻撃は、オンライン詐欺の一種であり、攻撃の対象がコンピュータ機器ではなくコンピュータユーザであるという点に特徴がある。フィッシング攻撃者は、エンドユーザを本物そっくりに作った偽サイトに誘導し、そのウェブサイトに個人情報を入力するよう促す。騙されたエンドユーザが入力したクレジットカードなどの個人情報を攻撃者が盗み取る、というのが攻撃の骨子である。フィッシングサイトの対策技術の1つに、ユーザが閲覧しているウェブサイトがフィッシングサイトであるかどうか検知する技術があり、代表的な検知手法としてヒューリスティクス方式が知られている。この方式は、ウェブサイトのURLやドメイン名を分析してフィッシングサイトらしさを計算し、そのスコアによってフィッシングサイトの検知を行う。ヒューリスティクス方式の課題は検知精度であり、新しいヒューリスティクスの開発や、複数のヒューリスティクスの組み合わせ手法などにより、精度の向上を目的とした研究がなされている。

本論文ではIP Address Spoofing（IPアドレス偽装）、Web Spoofing（フィッシング攻撃）の2つのSpoofingを用いた攻撃の対策について説明する。まず、**2**ではIPトレースバックシステムの普及を目的とし、その配備シナリオについて説明する。**3**では、フィッシングサイトの検知率の向上について解説する。最後に、**4**に本研究から得られた知見をまとめる。

2 IPトレースバックシステムの導入シナリオ

2.1 背景

前述のように、IPトレースバック技術[1]-[3]が提案されているが、最も有力であるSource Path Isolation Engine（SPIE）[3]について考察する。SPIEの問題点は、追跡性を高める際には、ネットワークの広範囲にSPIEの導入が必要である点である。

先行研究には、SPIEのようなIPトレースバックシステム（以下、IP-TBSと略す）の導入を容易にするため、各組織を自律分散システム（Autonomous System、以下ASと略す）単位で集約し、全ての中継装置ではなく、全てのASへの導入の検討を行っているものがある。IP-TBSがAS境界ルータを通過するパケットを監視し、問題のパケットを追跡するのに必要な情報を相互交換することにより、問題のパケットの送信元ASを特定することが可能となる。この場合、問題のパケットの送信者のホストを特定するのではなく、

問題のパケットの送信者が属するASが追跡の対象となり、追跡の粒度が下がる。しかしIP-TBSをネットワーク上の全ての中継装置に実装する必要がなくなり、1つのASに導入されれば、そのASについては追跡可能と見なすことが出来るため、コスト的に現実的ではないかと考え得る。

しかし追跡性は、ネットワークのトポロジや導入シナリオの影響を受ける。Gongらは3種類の人工的なネットワークを用いてシミュレーションを実施した[4]が、彼らの導入シナリオはランダム配備であった。すなわち、彼らはネットワークを構成するASからランダムに選択し、IP-TBSを順次導入していた。Castelucioは、IP-TBSの導入は、BGPの隣接ルータの多い順に沿って行われるべき[5]であると述べている。櫨山らは、模倣インターネット[6]を作成し、この上で導入シナリオを考察している。模倣インターネットはCAIDA[7]が観測しているインターネットのトポロジを元に、インターネットを模倣した環境である。櫨山らは、4通りの導入シナリオについて、日本国内のインターネットのトポロジを模倣した環境でシミュレーションを行った[8]。

本研究では、日中韓の3カ国のインターネットにおけるASへの導入を、大規模AS優先、小規模AS優先、中規模AS優先、そしてランダムといった4通りの導入シナリオを用いて行った場合の追跡性について、それぞれ評価を行う。

2.2 導入シナリオのシミュレーション

後の議論を分かりやすくするため、まずIP-TBSの導入シナリオとネットワークトポロジの関連性について説明する。ここでは、導入シナリオのシミュレーションにより追跡性を計測する。まず、追跡性についてシミュレーションで用いる2つの指標について説明する。次に、日中韓の3カ国のインターネットのトポロジを作成する手順を紹介する。最後に、4通りの導入シナリオについて説明する。

2.2.1 追跡性の指標

本研究では追跡性を性能の指標として用いる。追跡性には、問題のパケットの送信元ASを特定するという意味における追跡性と、問題のパケットが伝搬されてきたASパスを特定するという意味における追跡性の2通りがあり、前者をパケット追跡性、後者をパス追跡性と呼ぶこととする。

追跡性の計算には、先行研究[8]を参照することとした。先行研究ではパケット追跡性は式1で表現される。ここで、N_SはStrict ASの数であり、N_LはLoose ASの数であり、NはネットワークトポロジにおけるASの数である。

$$T_{packet} = \frac{(N_S + N_L)}{N} \quad (1)$$

Strict ASとは、IP-TBSが導入されたASを意味する。また、Loose ASとは、IP-TBSは導入されていないものの、近隣ASがIP-TBSを導入しているASを意味する。IP-TBS[9]のAS海峡を追跡できる特性により、櫨山らはLoose ASも同様に追跡性があると認識した。

また、パス追跡性は式2で表現される。ここで、L_SはStrict AS境界リンクの数であり、L_LはLoose AS境界リンクの数であり、LはネットワークトポロジにおけるAS境界リンクの数である。

$$T_{link} = \frac{(L_S + L_L)}{L} \quad (2)$$

Strict AS境界リンクとは、このリンク上で接続する2つのASが共にIP-TBSを導入しているという場合のAS境界のデータリンクである。また、Loose AS境界リンクとは、片側のASのみがIP-TBSを導入し、もう片側が未導入という場合のAS境界のデータリンクである。

2.2.2 ネットワークトポロジ

我々は、いくつかの国のインターネットトポロジを模倣したトポロジを用いて実験を行った。一般に、国家を越えた追跡は難しいとされている。これは技術的な問題ではなく、法律上の問題である。全てのトレースバック方式は攻撃パスを特定することが可能であるが、こうした技術は通信の秘密などの法律的な問題と干渉することが多い。そして、国家毎に通信の秘密の取り扱いが異なるためである。

そこで、本研究では日中韓それぞれ独立したネットワークトポロジを取り上げ、分析を行うこととした。

シミュレーションでは、我々の先行研究である模倣インターネット技術[6]を用いた。模倣インターネットでは、膨大な数のASを、限られたテ

ストベッド上のノードで再現するため、いくつかの枝狩りを行ったトポロジを用いている。本研究ではRegion Based Filtering（RBF）と名付けた枝狩りを用いて、各国のネットワークを、全インターネットのトポロジより分割した。

本実験では、CAIDAの提供する2008年11月22日におけるAS間の隣接構造データセットを用いた。本データセットの概要を表1に示す。ただしLoose ASが各国の外に含まれてしまうことも考慮し、導入対象のASの数と、パケットの追跡可能性を測定する対象のASの数は異なっている。なお、CAIDAは大規模にAS間の隣接構造の計測を測定しているが、プライベート・ピアリングなど隣接関係にあることを特定することが困難な場合もあるため、こうした情報は含まれていない。

表1 導入対象と追跡対象

	日本	中国	韓国
導入対象（ASの数）	500	196	640
パケットの追跡対象（ASの数）	768	308	755
パスの追跡対象（AS境界リンクの数）	1,589	529	1,375

2.2.3 導入シナリオ

我々の考察する4つの導入シナリオを以下に示す。

S1：大規模ASを優先するシナリオ

大規模ASから順に導入を狙うシナリオである。多数のASと隣接関係にある大規模ASに導入すると、IP-TBSも同様に多数のASとAS境界リンクを追跡できる。従って、導入したASの数が少なくても、それらASが大規模ASである場合は、網全体の追跡性が高いと思われる。ネットワークにおける大規模ASと特定するには様々な手法があるが、我々はBGPにおける隣接関係を持つASが最も多いASを基準に選定することとした。従って、このシナリオではBGPの隣接ASの多い順に沿って、IP-TBSが導入される。

S2：小規模ASを優先するシナリオ

このシナリオでは、IP-TBSはBGP隣接ASの少ない順に沿ってIP-TBSが導入される。もちろん限られた範囲のASとAS境界リンクしか追跡できないため、追跡性は限定的となる。しかし、攻撃を行うノードはしばしば、小規模ASに存在することが知られている。とりわけ、大手ISPのような大規模ASでは、Ingress Filtering[10]のようなDoS対策装置を導入しており、自らが攻撃者にならないような対策を行っているとされる。従って、小規模ASから導入した場合についても考慮しておく必要があると考えられる。

S3：中規模ASを優先するシナリオ

櫨山らの研究[8]では、大規模AS優先のシナリオが最も追跡可能性が高いと報告されていたが、実際の所、大規模ASには多くのAS境界が存在しているため、これら全てを監視するIP-TBSの導入は金銭的にコストが高くなることが予想される。

このシナリオでは、大規模ASを除いた中規模のASについて、隣接AS数の高い順に沿って、IP-TBSを導入する。大規模ASの特定は、ネットワークトポロジにおいてよく知られている、べき乗則[11]により行う。例えば表1に示した通り、日本には500個のASが存在する場合、大規模ASの数は22（$\approx\sqrt{500}$）として計算する。従って、BGPの隣接AS数が23番目から順番に、隣接AS数が多い順番に導入する。

S4：ランダムに導入するシナリオ

Gongらの研究[4]と同じく、このシナリオではIP-TBSをランダムに配備する。なお、偏りの影響を弱めるため、このシミュレーションは10回行い、その平均を求めることとした。

2.3 実験結果

最初に、日本のネットワークにおけるパケット追跡性を図1（a）に示す。ここでx軸はIP-TBSを導入したASの数、y軸はパケット追跡性（T_{packet}）をそれぞれ示す。もし15 ASに導入された場合、最も高いT_{packet}はS1の場合に観測され（86.3%）、次いでS3（18.5%）、S4（15.4%）、そしてS2（3.4%）の順であった。同様にパス追跡性について図1（b）に示す。ここでx軸はIP-TBSを導入したASの数、y軸はパス追跡性（T_{path}）をそれぞれ示す。

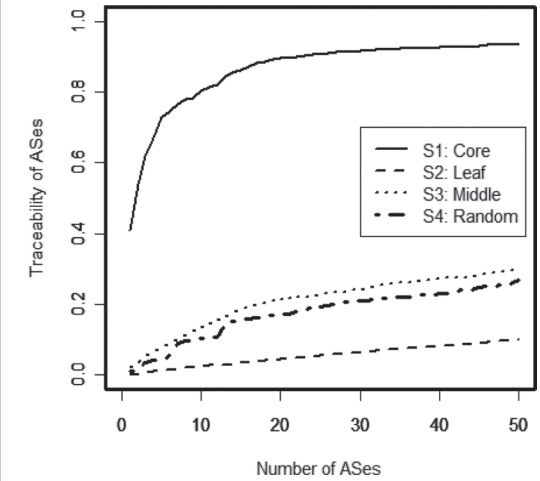

図1(a) 日本におけるパケット追跡性

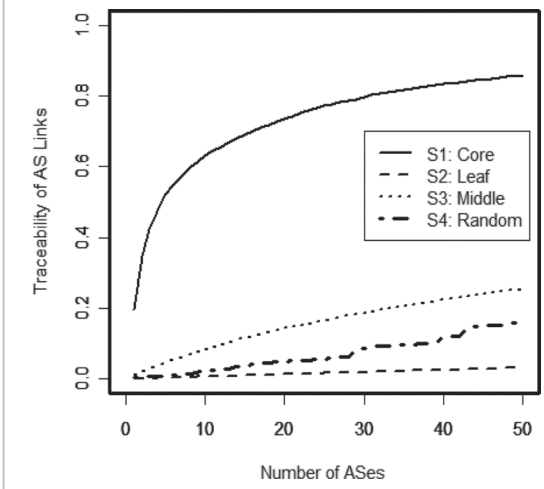

図1(b) 日本におけるパス追跡性

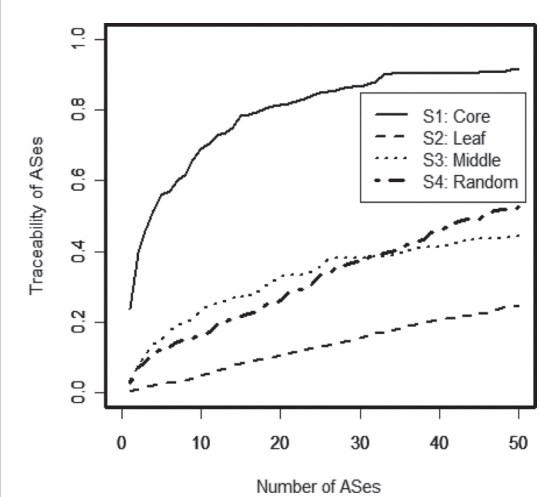

図2(a) 中国におけるパケット追跡性

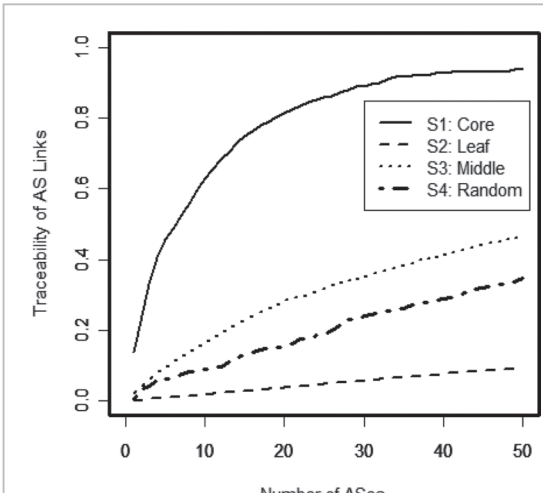

図2(b) 中国におけるパス追跡性

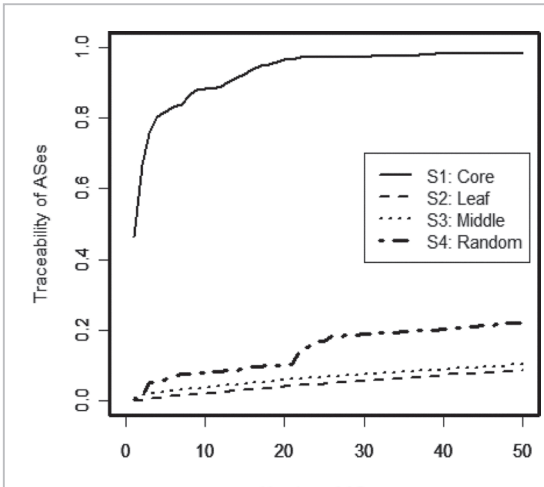

図3(a) 韓国におけるパケット追跡性

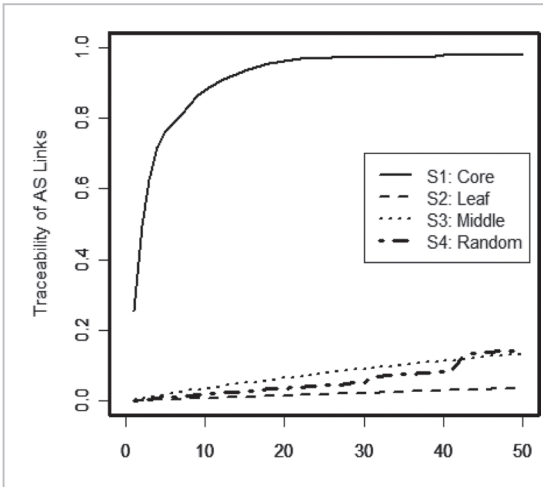

図3(b) 韓国におけるパス追跡性

15ASに導入された場合、最も高いT_{path}はS1の場合に観測され(69.0%)、次いでS3(11.6%)、S4(3.6%)、そしてS2(0.9%)の順であった。

次に、中国の場合についての結果を図2(a)及び図2(b)について示す。同様に15ASに導入された場合、最も高いT_{packet}はS1の場合に観測され(74.8%)、次いでS3(27.3%)、S4(21.7%)、そしてS2(8.1%)の順番であった。また、最も高いT_{path}もS1の場合に観測され(74.9%)、次いでS3(22.9%)、S4(13.0%)、そしてS2(2.8%)の順であった。

最後に、韓国の場合についての結果を図3(a)及び図3(b)について示す。同様に15ASに導入された場合、最も高いT_{packet}はS1の場合に観測され(92.6%)、次いでS4(8.8%)、S3(4.5%)、そしてS2(3.1%)の順番であった。また、最も高いT_{path}もS1の場合に観測され(93.4%)、次いでS3(5.0%)、S4(2.7%)、そしてS2(1.1%)の順であった。

全ての事例において、最も高いT_{packet}及びT_{path}は、S1の導入シナリオの場合において観測されたこととなる。とりわけ、S1における韓国の追跡性は、他国と比較しても高くなっている。これは、韓国の場合、少数の大規模ASに多数のASが相互接続されているからではないかと考える。

S1の場合において、中国における追跡性は、他国と比較して低いと言える。しかし、15ASに導入したパケット追跡性、パス追跡性はそれぞれ27.3%、22.9%と他国より高くなっている。これには2つの理由があると推測する。1つ目の理由は、導入対象とするASの数である。表1に示す通り、中国の導入対象は196ASと少ないため、大規模ASの影響が限定的であると考えられる。2つ目の理由は、中国が地理的に広大な国土を持つ国であり、ネットワークが広域に分散していると考えられるためである。従って、ごく少数の大規模ASではなく、BGPの隣接AS数が他国より少ない大規模ASが多数存在しているのではなかろうかと考える。

これらの知見により、我々は韓国のネットワークはいわゆる「集中型」であると捉えており、中国のネットワークは「分散型」であるとして捉えている。日本のネットワークの特徴は、中国と韓国のネットワークの中間である。しかし、パス追跡性は中国、韓国と比べると低くなっている。これは、表1に示した通り、AS境界リンクの数が多いことによると考えられる。日本におけるAS境界リンクの数は1,589であり、韓国(1,378)より多いが、日本におけるASの数は500であり、韓国(640)より少ない。これは、中小規模のASが、大規模ASと相互接続されているだけでなく、中小規模AS同士でも多数に接続されているため、ASリンク数が増加したからではないかと考える。

3 過去の判断履歴を活かしたフィッシングサイトの検知

3.1 背景

フィッシングサイトの検知を行う方式としては、URLフィルタリング方式とヒューリスティクス方式がある。URLフィルタリング方式は、ユーザが閲覧しているウェブサイトのURLを、フィッシングサイトのURLデータベースと照合することによって、フィッシングサイトであることを検知する。カーネギーメロン大学においてZhangらが行った2007年の調査研究[12]では、攻撃の初期段階においてフィッシング検知精度は約70%であることが示されていた。しかし、同大学で2009年に行われた、様々なフィッシングサイトのデータベースを対象とした調査[13]では、攻撃が行われて間もないフィッシングサイトは、その20%未満しかデータベースに登録されていないことが報告された。

ヒューリスティクス方式はウェブサイトのURLやコンテンツなどからフィッシングサイトらしさを計算する方式である。有名なヒューリスティクスの例としては、ドメイン名の取得期間の長さという手法がある。フィッシングサイトは発生してから消滅するまでの期間が短い。従って、ドメイン名が登録されてから現在までの期間が短い場合はフィッシングサイト、そうでない場合は正規サイトというように判別することができる。こうしたヒューリスティクスは必ずしも正確ではないため、複数の異なるヒューリスティクスを組み合わせる必要がある。ヒューリスティクス方式の課題は検知精度にある。前述のZhangらの調査研究[12]では、ヒューリスティクス方式のSpoofGuard[14]は約94%のフィッシングサイトを正しく判別できるものの、約42%の正規サイトを誤ってフィッシングサイトと判別する問題が報告された。我々の先

行研究では機械学習を用いた検知手法を提案し[15]、既存方式と9種類の機械学習を用いた方式の比較を行い、ほとんどの場合において機械学習を用いることにより判定精度が高まることを観測した。中でも、機械学習の一手法であるAdaBoostの場合おいて最も高い検知精度が観測された。

現在においても、ヒューリスティクス方式の検知精度を高める研究は重要な課題である。本論文では、我々はHumanBoostという、エンドユーザがウェブサイトを信頼できる、信頼できないといった判断を行った結果（Past Trust Decision, PTD）を、機械学習であるAdaBoostによって既存のヒューリスティクスと組み合わせて用いる検知方式を提案する。一般的に、ヒューリスティクスが動作しないような場合においても、エンドユーザは潜在的にフィッシングサイトを発見できる何らかの能力を持っていると考えられる。もし仮にPTDの履歴をエンドユーザ毎に構築し、意志決定結果を保存できるとすれば、それは既存のヒューリスティクスのように「フィッシングである」「フィッシングでない」という結果を出力する判定器として用いることが出来る。

さらに、HumanBoost方式では、フィッシングの検知をエンドユーザに合わせて調節することについて考慮する。もしユーザがセキュリティの専門家であった場合、当人の意志決定は、検知においてもっとも支配的な用途となるであろう。反対に、既存のヒューリスティクスが強い影響力を持つ場合には、ユーザが初心者であるなどの理由で、意志決定が頻繁に誤っていると考えられる。

3.2 被験者実験とその結果

PTDの実用性を調べるため、我々は被験者を募ってフィッシングサイト判定実験からPTDを作成するという試みを2007年11月、2010年3月、2010年7月に行った。ここでは、最初に行われた試験について、フィッシングIQで用いたデータセットの内容や用いたヒューリスティクス、実験のデザイン及び実験結果について説明する。

3.2.1 データセットについて

代表的なフィッシングサイト判別実験がDhamija[16]によって行われており、我々はこの方式を参照することとした。我々が実験に用いたウェブサイトのリストを表2に示す。用意したウェブサイトのうち14の模擬的に作成したフィッシングサイトと6の正規サイトが含まれる。全てのウェブサイトは、ユーザ名やパスワードなどの個人情報を入力するフォームを持っている。より詳細な説明は文献[17]を参照されたい。

3.2.2 ヒューリスティクス

我々の実験ではCANTINA[18]に採用されている8種類のヒューリスティクスを用いる事とした。当時、我々の知る限りCANTINAはURLフィルタリングのデータベースを用いない中では、最も検知精度の高い検知ツールであった。

3.2.3 実験のデザイン

我々はwith-subject design、すなわち全ての被験者が同一のウェブサイトを閲覧するような実験計画を採用した。実験では、10人の被験者に対し自由にウェブサイトを閲覧するよう促した。各被験者はWindows XPとInternet Explorer（IE）6.0がインストールされたPCを用いてウェブサイトを閲覧する。IEを多国語ドメイン対応に設定した以外、セキュリティソフトやフィッシング対策ソフトは導入していない。また、我々は被験者に対し表2で指定したウェブサイト以外のサイトを閲覧することを禁止しなかった。被験者によっては、Googleなどのサーチエンジンにいくつかのキーワードを入力し、検索結果に表れたURLとブラウザに表示されているURLの確認を行う者もいた。

なお、この実験では、判別誤り率を指標として用いる。また、偏りを減らすため、4分割交差検定を10回繰り返して平均値を算出することとする。

3.2.4 実験結果

10人の被験者については、全員が日本人の男性であり、奈良先端科学技術大学院大学に属していた。うち過去5年以内に修士課程を卒業しており、残りは修士課程の学生である。被験者は表2に示したウェブサイトについて、それぞれ真（正規サイトである）偽（フィッシングサイトである）の判定を行った。

次にAdaBoostによる検知手法の精度を計測する。実験で用いた8つのヒューリスティクスは「フィッシングサイトである」「フィッシングサイトでない」のどちらかを出力する。この結果を基に、AdaBoostにより組み合わせ、判別誤り率の

表2 各ウェブサイトの状況

#	ウェブサイト	真／偽	言語	備考
1	Live.com	真	英語	URL（*login.live.com*）
2	東京三菱 UFJ 銀行	偽	日本語	URL（*www-bk-mufg.jp*）， 正規サイトは URL（*www.bk.mufg.jp*）
3	PayPal	偽	英語	URL（www.paypal.com.%73%69 ... %6f%6d） （URL Encoding Abuse）
4	Goldman Sachs	真	英語	URL（*webid2.gs.com*），SSL
5	Natwest Bank	偽	英語	URL（*onlinesession-0815.natwest.com.esb6eyond.gz.cn*）， PhishTank.com に報告されたサイト
6	Bank of the West	偽	英語	URL（*www.bankofthevvest.com*）， 正規サイトは URL（*www.bankofthewest.com*）
7	南都銀行	真	日本語	URL（*www2.paweb.anser.or.jp*），SSL 南都銀行を連想させない URL
8	Bank of America	偽	英語	URL（*bankofamerica.com@index.jsp-login-page.com*） （URL Scheme Abuse）
9	PayPal	偽	英語	URL（*www.paypal.com*），最初の“a”がキリル語の“а” （U+430）（国際ドメイン名の悪用）
10	Citibank	偽	英語	URL（IP address）（IP Address Abuse）
11	Amazon	偽	英語	URL（*www.importen.se*），path 部に amzon が含まれて いる PhishTank.com に報告されたサイト
12	Xanga	真	英語	URL（*www.xanga.com*）
13	Morgan Stanley	真	英語	URL（*www.morganstanleyclientserv.com*），SSL
14	Yahoo	偽	英語	URL（IP address）（IP Address Abuse）
15	U.S.D. of the Treasury	偽	英語	URL（*www.tarekfayed.com*） PhishTank.com に報告されたサイト
16	三井住友カード	偽	日本語	URL（*www.smcb-card.com*） 正規サイトは URL（*www.smbc-card.com*）
17	eBay	偽	英語	URL（*secuirty.ebayonlineregist.com*）
18	Citibank	偽	英語	URL（*シテイバンク.com*），日本語で Citibank に見える偽 サイト（国際ドメイン名の悪用）
19	Apple	真	英語	URL（*connect.apple.com*），SSL，画面に SSL 未使用のコ ンテンツにアクセスしているという警告
20	PayPal	偽	英語	URL（*www.paypal.com@verisign-registered.com*）， （URL Scheme Abuse）

平均値を 3.2.3 に述べた方法に沿って計算した。なお、各ヒューリスティクスの検知結果は文献 [17] を参照されたい。

最後に、HumanBoost の検知精度を計測する。まず 10 人の PTD のデータベースを作成し、これと既存の検知結果を組み合わせた行列を作成した。言い換えると、被験者毎に異なる 10 個の 20＊9 のバイナリ行列を作成したことになる。このバイナリ行列を用い、AdaBoost の時と同じ条件の下で、判別誤り率の平均値を被験者毎に計算した。

結果を図 4 に示す。図の灰色の棒グラフは各被

験者の判別誤り率、白色の棒グラフはAdaBoostの場合における判別誤り率の平均値、黒色の棒グラフはHumanBoostの場合における判別誤り率の平均値である。全体を通して見ると、HumanBoostの場合の判定誤り率の平均は13.4%であり、AdaBoost（20.0%）、被験者（19.0%）の場合よりも誤り率が低い。正規サイトをフィッシングサイトであると誤判定する割合について、HumanBoost（19.6%）は、AdaBoost（28.1%）、被験者（29.7%）の場合を下回った。また、フィッシングサイトを正規サイトであると誤判定する割合についても、HumanBoost（8.5%）がAdaBoost（13.5%）、被験者（14.0%）の場合を下回った。

これらの結果より、HumanBoostは平均誤り率を低下するのに効果があったと考える。しかし、特定の被験者の場合にHumanBoostを用いることにより誤り率が増加している場合も見受けられる。例えば図4では被験者9の場合である。この被験者9は3つの正規サイトをフィッシングサイトであると判断したが、このサイトを判断できるヒューリスティクスに高い重みが割り当てられることになった。このため、彼のPTDはヒューリスティクスと比較して重要視されず、誤り率が高まる遠因となったのではないかと思われる。

3.3 被験者実験の追試

実験に参加する被験者の数を増やすことは、HumanBoostの実験結果をより一般的なものとする上で重要である。ここでは、2010年に行った2つの被験者実験の追試験について説明する。なお、初回の実験と単純に比較できるものではないことについて明記しておく。これは、初回の実験から3年経過しており、この間、被験者のフィッシングに対する意識が変化している可能性がある

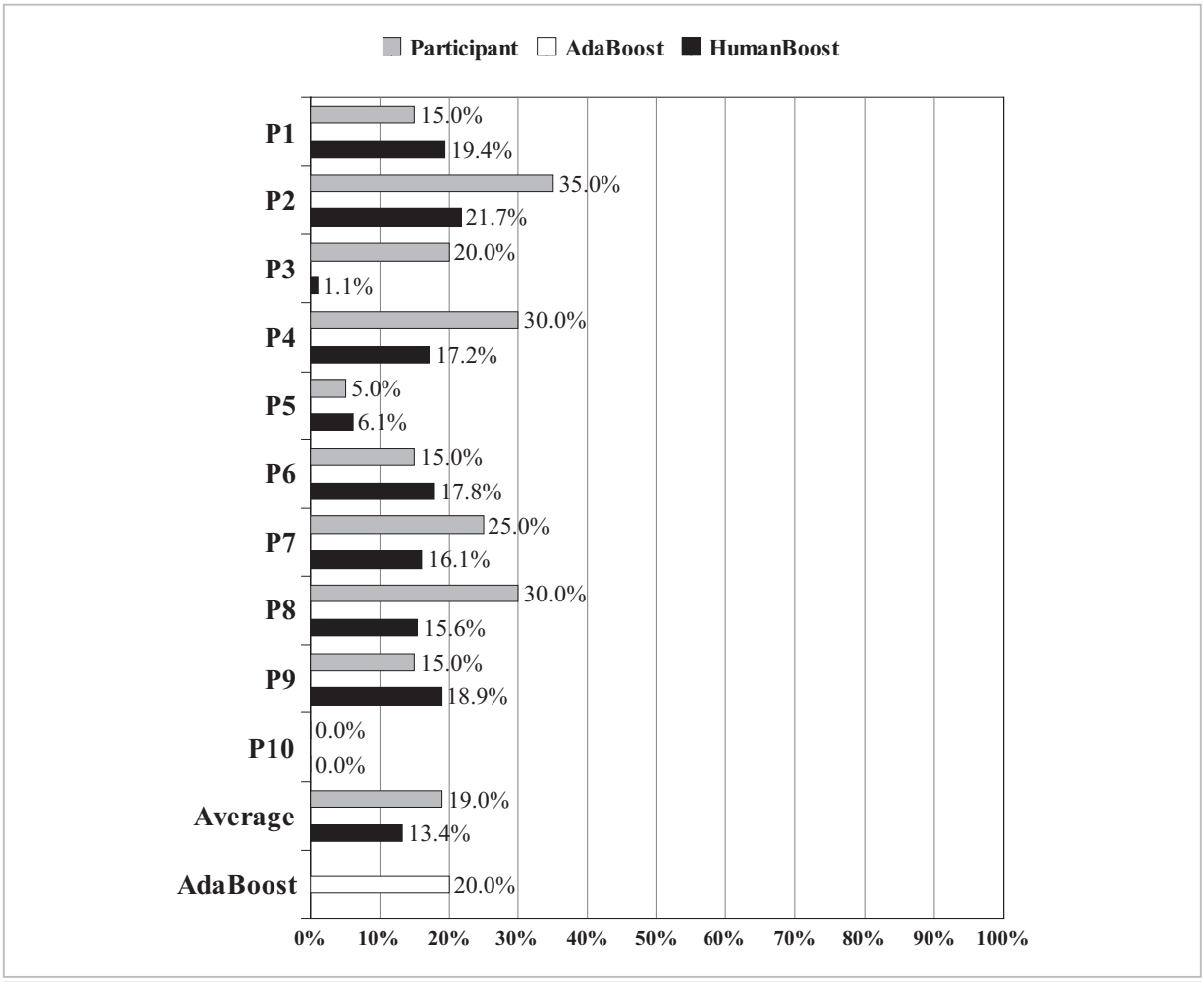

図4 各被験者、AdaBoost、HumanBoostの場合における平均判別誤り率

ためである。

3.3.1 2010年3月に実施した被験者実験

23歳から30歳まで11人の被験者を募り追試を行った。被験者は全員、北陸先端科学技術大学院大学に所属する日本人である。うち2人が過去5年以内に修士課程を卒業しており、残りは修士課程の学生である。

追試にあたっては、表2に記載したウェブサイトの更新を行った。まず、2007年から2010年に当たってPayPalのトップページのデザインが変更されたため、実験当時である2010年のPayPalのトップページを用いることとした。また、南都銀行（表2におけるウェブサイト6）のトップページはURLが変更された。南都銀行は、今回の被験者の多くが住む石川県ではさほど有名ではないこともあり、北陸銀行のサイトを用いるよう変更した。北陸銀行は、南都銀行と同じく日本における地方銀行の1つであり、ドメイン名は*www2.paweb.answer.or.jp*と、2007年における南都銀行と同一である。

2010年3月に、11人の被験者は20ウェブサイトを閲覧し、フィッシングと思うか否かについて判定を行った。**3.2**に述べた初期実験とは異なり、各被験者はブラウザを操作するのではなく、各ウェブサイトを表示したブラウザのスクリーンショットをカラー印刷したものを閲覧した。2010年にはWindows XPとIE 6.0は旧式となっていたため、Windows VistaとIE 8.0を用いることとした。なお、ブラウザのスクリーンショットにより判定をさせるという方式は、しばしばフィッシングサイト判別実験で用いられる手法である。

実験結果を図5に示す。図の灰色の棒グラフは各被験者の判別誤り率、白色の棒グラフはAdaBoostの場合における判別誤り率の平均値、

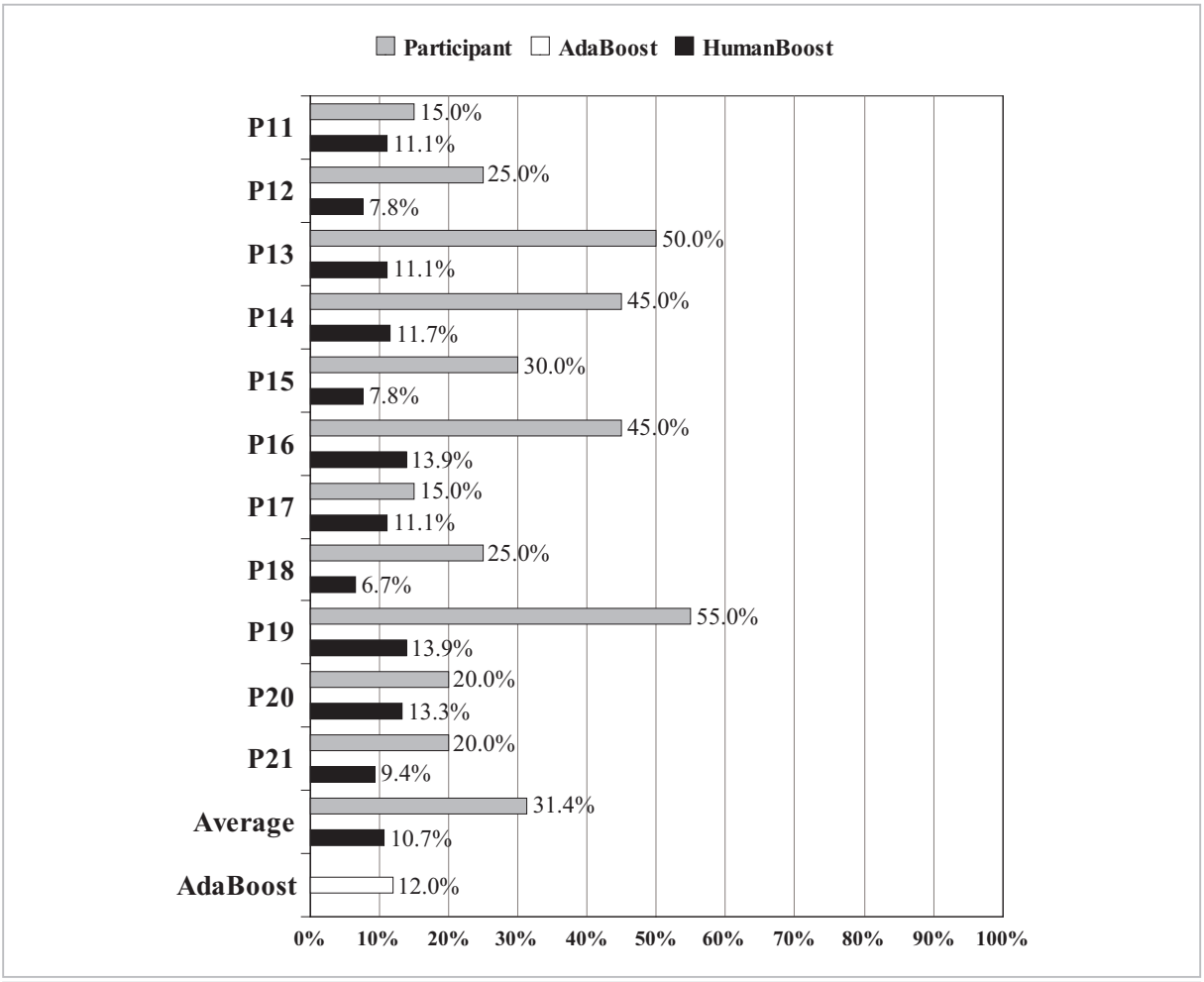

図5 2010年3月の追試における各被験者、AdaBoost、HumanBoostの場合における平均判別誤り率

黒色の棒グラフはHumanBoostの場合における判別誤り率の平均値である。全体を通して見ると、HumanBoostの場合の判定誤り率の平均は10.7%であり、AdaBoost（12.0%）、被験者（31.4%）の場合よりも誤り率が低い。正規サイトをフィッシングサイトであると誤判定する割合について、HumanBoost（15.4%）は、AdaBoost（18.1%）、被験者（39.9%）の場合を下回った。また、フィッシングサイトを正規サイトであると誤判定する割合についても、HumanBoost（6.1%）がAdaBoost（8.4%）、被験者（25.9%）の場合を下回った。

初期実験とは条件も異なる上、トップページのデザインも違うので単純には比較できない。ただし、少なくともHumanBoostにより検知精度が高まるという傾向は観測された。

3.3.2　2010年7月に実施した被験者実験

より多くのPTDを採取するため、インターネット市場調査企業を通じて被験者を集め実験することにした。ここではその結果を報告する。

集めた309人の被験者のうち、42.4%（131人）が男性、57.6%（178人）が女性であり、年齢は16歳から77歳までであった。48.2%（149人）が会社員であり、19.7%（61人）が主婦、5.8%（18人）が学生であった。学生のうち66.7%（12人）は大学生、11.1%（2名）は高校生、5.6%（1名）は修士課程に在学中であった。彼らの多くは東京在住であり、表2におけるウェブサイト6は東京都民銀行に変更した。東京都民銀行も同じく日本の地方銀行の1つであり、同様にドメイン名が*www2.paweb.answer.or.jp*である。その他の条件は**3.3.1**で述べた通りである。

各被験者の判断結果に基づき、我々は被験者毎の場合、AdaBoostの場合、HumanBoostの場合についてそれぞれ判別誤り率の平均値を計測したところ、HumanBoostは9.7%であり、AdaBoost（10.5%）、被験者（40.5%）の場合を下回った。正規サイトをフィッシングサイトであると誤判定する割合について、AdaBoost（18.3%）がHumanBoost（19.5%）、被験者（57.4%）の場合を下回った。また、フィッシングサイトを正規サイトであると誤判定する割合については、HumanBoost（5.5%）がAdaBoost（7.1%）、被験者（33.2%）の場合を下回った。

4　おわりに

我々は2つのSpoofing技術、すなわちIPアドレス詐称とフィッシングサイト対策に取り組んだ。詐称されたIPアドレスの発信源を突き止めるには、IPトレースバックシステム（IP-TBS）の効果的な導入が重要である。そこで、自律分散システム（AS）に対し、いくつかの導入戦略に沿って追跡性を調べた。ここで用いた追跡性とは、パケット追跡性とパス追跡性である。また、追跡性は導入戦略だけでなくネットワークトポロジによっても異なる。そこで本研究では、効果的なシミュレーションを行うため、日中韓のインターネットトポロジを模倣したネットワークトポロジを取り上げて考察した。また、導入戦略も大規模AS優先、小規模AS優先、中規模AS優先及びランダムに導入する4通りのシナリオを取り上げ、それぞれ考察を行った。

結果として、限られたASに導入する場合においては、大規模ASから順番に導入するシナリオがネットワークトポロジに関係なく追跡性を高められることを確認した。仮に15ASに導入する場合、パケット追跡性とパス追跡性のペアは、日本の場合はそれぞれ86.3%、69.0%であり、中国の場合は74.8%、74.9%であり、韓国の場合は92.6%、93.4%であった。中規模ASから導入するシナリオは、大規模ASから導入するシナリオに次いで追跡性が高かったが、日本の場合のパケット追跡性、パス追跡性はそれぞれ18.5%、11.6%であり、中国の場合は27.3%、22.9%であり、韓国の場合は4.5%、5.0%とばらつきが見られた。

この結果は、3カ国のネットトワークトポロジの特徴を浮き彫りにしている。中国のネットワークトポロジは中規模ASや小規模ASの結びつきが強く、韓国のネットワークトポロジは大規模ASとそれ以外の結びつきが強い。これら2カ国と比べると、日本のネットワークトポロジは中国と韓国の中間であると言えよう。

翻って、フィッシングサイト対策については効果的な検知アルゴリズムが求められている。本研究ではHumanBoostと名付けた、エンドユーザの過去の判断履歴を活用することにより検知精度を高める手法について提案した。エンドユーザは自身の個人情報を入力する際には何らかの意志決定

をウェブサイトに対して行っていると考えられる。このように過去の意志決定（Past Trust Decision; PTD）は「フィッシングサイトである」「フィッシングサイトではない」といった既存のヒューリスティクスと同様に2値を出力する判別器として捉えることができる。そこで、意志決定の履歴と既存のヒューリスティクスを機械学習によって組み合わせできると考えた。

2007年11月には、10人の被験者を募った最初の実験が行われた。被験者は14のフィッシングサイトと6の正規サイトを閲覧し、フィッシングサイトか否かを判定した。これの判定結果から10人分のPTDを作成し、機械学習の一手法であるAdaBoostを用いて既存の8つのヒューリスティクスと組み合わせた検知を行った。

この結果、HumanBoostの場合における判別誤り率の平均は13.4%であり、AdaBoost（20.0%）、被験者（19.0%）の場合を下回った。追試として、11人の被験者を2010年3月に募って同様の実験を行ったところ、HumanBoostの場合における判別誤り率の平均は10.7%であり、AdaBoost（12.0%）、被験者（31.4%）の場合を下回った。また、2010年7月に309人の被験者に対して追試したところ、HumanBoostの場合における判別誤り率の平均は9.7%であり、AdaBoost（10.5%）、被験者（40.5%）の場合を下回った。従って、PTDを用いてフィッシングサイト検知を行うことにより、検知精度は高まるものと考えられる。

今後の課題についても、IPトレースバック、HumanBoostの場合について示す。まず、IPトレースバックの場合では、日中韓が互いに相互接続された場合の追跡性について評価する。通信の秘密といった法律の取り扱いが3カ国で異なるが、DoS攻撃は国境を越えて行われるのであり、IP-TBSの導入シナリオに基づいた評価を実施する必要があると考える。また、他の国についても同様にシミュレーションを行う予定である。

HumanBoostの課題としては、より多くの被験者を相手とした実験を行うフィールドテストを検討している。被験者実験では、データセットの偏りの影響を減らす事は重要であり、交差検定を行ったとしても偏ったデータセット、偏った被験者を用いていたという可能性は否定できない。エンドユーザの許可は必要となるが、PTDを採取可能な機能を組み込んだブラウザを用いたフィールドテストを行うことにより、HumanBoost実験の偏りを減らすよう試みる予定である。

参考文献

1 S. Bellovin, M. Leech, and T. Taylor, “ICMP Traceback Message,” IETF Internet Draft, draft-ietf-itrace-04.txt, Feb. 2003.

2 S. Savage, D. Wetherall, A. R. Karlin, and T. E. Anderson, “Practical network support for IP traceback,” Proceedings of the ACM SIGCOMM 2000 Conference on Applications, Technologies, Architectures, and Protocols for computer communications, pp. 295–306, Aug. 2000.

3 A. C. Snoeren, C. Partridge, L. A. Sanches, C. E. Jones, F. Tchakountio, S. T. Kent, and W. T. Stayer, “Hash-based IP traceback,” Proceedings of the ACM SIGCOMM 2001 Conference on Applications, Technologies, Architectures, and Protocols for computer communications, pp. 3–14, Aug. 2001.

4 C. Gong, T. Le, T. Korkmaz, and K. Sarac, “Single Packet IP Traceback in ASlevel Partial Deployment Scenario,” Proceedings of the IEEE Global Telecommunications Conference, Nov. 2005.

5 A. O. Castelucio, Ronaldo M. Salles, and A. Ziviani, “Evaluating the partial deployment of an AS-level IP traceback system,” Proceedings of the ACM symposium on Applied computing, pp. 2069–2073, Mar. 2008.

6 H. Hazeyama, M. Suzuki, S. Miwa, D. Miyamoto, and Y. Kadobayashi, “Outfitting an Inter-AS Topology to A Network Emulation TestBed for Realistic Performance Tests of DDoS Countermeasures,” Proceedings of the Workshop on Cyber Security and Test, Jul. 2008.

7 The CAIDA Web Site,“CAIDA: cooperative association for internet data analysis,”Available at: http://www.caida.org/
8 櫨山寛章，若狭賢，門林雄基，“実証実験に向けたIPトレースバックシステム導入シナリオに関する一考察,”電子情報通信学会技術研究報告，IA2008-14, pp. 25–30, Jul. 2008.
9 H. Hazeyama, Y. Kadobayashi, D. Miyamoto, and M. Oe,“An Autonomous Architecture for Inter-Domain Traceback across the Borders of Network Operation,”Proceedings of the IEEE Symposium on Computers and Communications, Jun. 2006.
10 B. R. Greene, C. Morrow, and B. W. Gemberling,“Tutorial: ISP Security - Real World Techniques II,”Available at: http://www.nanog.org/meetings/nanog23/
11 A. Barabasi and R. Albert,“Emergence of Scaling in Random Networks,”Science, Vol. 286, pp. 509–512, 1999.
12 Y. Zhang, S. Egelman, L. Cranor, and J. Hong,“Phinding Phish: Evaluating Anti-Phishing Tools,”Proceedings of the 14th Annual Network and Distributed System Security Symposium, Feb. 2007.
13 S. Sheng, B. Wardman, G. Warner, L. Cranor, J. Hong, and C. Zhang,“An Empirical Analysis of Phishing Blacklists,”Proceedings of the 6th Conference on Email and Anti-Spam, Jul. 2009.
14 N. Chou, R. Ledesma, Y. Teraguchi, D. Boneh, and J. C. Mitchell,“Client-side defense against web-based identity theft,”Proceedings of the 11th Annual Network and Distributed System Security Symposium, Feb. 2004.
15 D. Miyamoto, H. Hazeyama, and Y. Kadobayashi,“An Evaluation of Machine Learning-based Methods for Detection of Phishing Sites,”Australian Journal of Intelligent Information Processing Systems, Vol. 10, No. 2, pp. 54–63, Nov. 2008.
16 R. Dhamija, J. D. Tygar, and M. A. Hearst,“Why Phishing Works,”Proceedings of Conference On Human Factors In Computing Systems, Apr. 2006.
17 D. Miyamoto, H. Hazeyama, and Y. Kadobayashi,“HumanBoost: Utilization of Users' Past Trust Decision for Identifying Fraudulent Websites,”Journal of Intelligent Learning Systems and Applications, Vol. 2, No. 4, pp. 190–199, Scientific Research Publishing, Dec. 2010.
18 Y. Zhang, J. Hong, and L. Cranor,“CANTINA: A Con-tent-Based Approach to Detect Phishing Web Sites,”Proceedings of the 16th World Wide Web Conference, May 2007.

（平成23年6月15日 採録）

みやもとだいすけ
宮本大輔
東京大学情報基盤センター
助教　博士（工学）
ネットワークセキュリティ

はぜやまひろあき
櫨山寛章
奈良先端科学技術大学院大学
情報科学研究科助教　博士（工学）
IP トレースバック、テストベッド

かどばやしゆうき
門林雄基
ネットワークセキュリティ研究所
専攻研究員／
奈良先端科学技術大学院大学
情報科学研究科准教授　工学博士
IP トレースバック、サイバーセキュリティ