情報処理学会論文誌 Vol. 52 No. 3 929–939 (Mar. 2011)

DoS 攻撃経路を効率的に再構築するための トポロジ特性を考慮した確率的パケットマーキング手法

金岡 晃[†1] 岡田雅之[†1,†2]
勝野恭治[†3] 岡本栄司[†1]

近年，DoS 攻撃を行っているホストを特定するトレースバック技術の中でも確率的パケットマーキング（Probabilistic Packet Marking，PPM）方式が注目されている．PPM 方式は他のトレースバック方式と比較して様々な利点がある反面，攻撃経路の再構成に必要なパケット数が多く再構成までに時間がかかることと，分散型 DoS 攻撃時などでのハッシュ値衝突による経路の誤った再構成の問題点が指摘されている．さらに，パケットマーキング処理の負荷に関してほとんど議論されていない．本論文では 3 つの改良を PPM に施し，従来方式よりも効率良く被害者側が攻撃経路を再構築できる方式を提案する．そして，提案方式の評価として，数式評価と実装評価を行った．数式評価から従来方式よりも必要とするパケット数の削減に成功し，さらに衝突確率の改善も実現したことを示した．また実装評価からパケットマーキングの処理はほとんど負荷にならないことを示した．

Probabilistic Packet Marking Method Considering Topology Property for Efficiently Re-building DoS Attack Paths

Akira Kanaoka,[†1] Masayuki Okada,[†1,†2]
Yasuharu Katsuno[†3] and Eiji Okamoto[†1]

Recently, there have been much attention to Probabilistic Packet Marking (PPM) in several traceback technologies that specify DoS attack source node, with some advantages as compared with others. But, PPM has three major issues: (i) it takes long time to re-build attack paths because lots of packets are required, (ii) hash value collision in Distributed DoS attacks leads to wrong attack paths and (iii) there are few discussion about workload sizes for packet marking tasks. We propose improved PPM in order to re-build attack paths more efficiently than existing studies, and evaluate it with both mathematical and implementation approaches. A mathematical evaluation shows that proposed PPM can decrease a number of packets for re-building attack paths and improve collision probability. An implementation evaluation shows that packet marking tasks are low workloads.

1. はじめに

サービス妨害（Denial of Service，以後 DoS）攻撃はインターネット上の重大な脅威になっている．DoS 攻撃への対策はこれまで数多くがなされており，対策研究の分類もされてきている[1]．対策技術の 1 つであるトレースバック技術は，DoS 攻撃を行っているホストを特定する技術として数々の研究がされており，いくつかの方式に細分化されている．その中に確率的パケットマーキング（Probabilistic Packet Mariking，PPM）技術がある．

PPM では攻撃元から被害者に至る経路上のルータが，確率的にパケットへ経路に関する情報をマーキングする．被害者側はマーキングされたパケットを収集することで，攻撃経路の再構成を行い，攻撃元の特定をする．PPM は 2000 年に Savege らにより提案された後[2]，複数の方式や応用が提案されてきた[3)–7)]．PPM 方式は他のトレースバック方式と比較して，余分なパケットを発生させない点や別途ストレージを用意する必要がない点など様々な利点がある反面，攻撃経路の再構成に必要なパケット数が多く再構成までに時間がかかることと，分散型 DoS 攻撃時などでのハッシュ値衝突による経路の誤った再構成の問題点が指摘されている．さらに，マーキングで使用するパケットマーキング処理の負荷に関してほとんど議論されていないことも問題である．

本研究では Savage らの方式[2] と Goodrich の方式[7] を基にインターネットのルータトポロジの特徴を考慮した 3 つの改良を PPM に施し，従来方式よりもより効率良く被害者側が攻撃経路を再構成できる方式を提案する．1 つ目の改良で，パケット距離情報に用いるデータサイズを従来の 5 ビットから 4 ビットに削減し，マーキングエリアをより有効に使えるようにする．そして 2 つ目の改良で，検証のために用いられるハッシュ値のサイズを削減し，分割されてマーキングされる情報の分割数を削減する．最後に 3 つ目の改良で，攻撃経

†1 筑波大学
University of Tsukuba
†2 日本ネットワークインフォメーションセンター
Japan Network Information Center (JPNIC)
†3 IBM 東京基礎研究所
IBM Research - Tokyo

路再構成後の検証時にパケット情報とは別に距離情報を確認することでハッシュ衝突による誤った経路再構成への耐性を高める．

本論文では，提案方式の評価として，数式評価と実装評価を行った．数式評価として，攻撃経路の再構成に必要なパケット数の期待値と衝突確率の計算を評価した．その結果，従来方式よりも必要とするパケット数の削減に成功し，さらに衝突確率の改善を実現した．続いて実装評価として，従来の研究ではほとんど行われてこなかった PPM の実装を行い，PPM の負荷を測定した．負荷測定の結果，パケットマーキングの処理はほとんど負荷にならないことが判明し，PPM 方式の優位性を示した．

本論文の構成は以下のとおりである．まず 2 章において IP トレースバック技術の概観と，これまでの PPM 方式についての解説を行う．次に 3 章においてインターネットトポロジの調査とその結果を示す．続く 4 章では前章で得られた結果をもとに新たな PPM 方式を提案する．5 章において提案方式を評価し，最後に 6 章でまとめる．

2. 関 連 研 究

2.1 トレースバック技術

Peng らによる DoS 攻撃対策の分類に “Attack Source Identification” がある[1]．これは攻撃元に関する情報を識別するトレースバック技術についてのものである．

トレースバック技術は複数の提案がされているが，本節では提案方式と同様の事後追跡可能なトレースバック技術のみに焦点を当て既存方式を紹介する．事後追跡可能なトレースバック技術は大きく「ロギング方式」「ICMP 方式」「パケットマーキング方式」に分けることができる．

ロギング方式は，攻撃経路上の中継ルータにおいて記録装置がパケットの記録を行い，被害者が記録された情報との照合を行うことで攻撃元に関する情報を得る方式である[8)–12)]

ICMP 方式は，攻撃経路を再構成するために必要な情報を ICMP パケットとして送信することで被害者による経路再構成を可能にする方針である[13),14)]．

ロギング方式や記録装置の設置や維持のためのコストがかかり，また ICMP 方式では DDoS 中に余分なトラフィックをネットワークに発生させることがそれぞれ欠点としてあげられている[7)]．また双方の方式とも方式を備えたルータや記憶装置に対するサービス妨害となる可能性もある．

パケットマーキング方式はパケットを中継するルータが中継するパケットに対して情報を付加する方式である．本論文はパケットマーキング方式の中でも確率的にパケットマーキングを行う確率的パケットマーキング（Probabilistic Packet Marking，PPM）方式の改良を提案するため，PPM 方式について次節で詳細に解説する．

2.2 確率的パケットマーキング

PPM 方式は Savage らにより提案された[2)]．パケットを中継するルータが確率的にルータ情報をパケットのヘッダに書き込む．被害者は，攻撃発生の際の攻撃パケットを収集することで攻撃者までの経路を復元することが可能である．Savage らの方式では，ルータの IP アドレス（32 ビット）とそのハッシュ値（32 ビット）を合わせたデータを 8 分割し，オフセットや距離情報とともに IPv4 ヘッダの Identification フィールド（16 ビット）に書き込む．Savage らの方式は本論文で提案する方式の基になる方式であるため次項で詳述する．

Savage らの方式が提案された後，様々な改良方式が提案された．Song らの方式[3)] では，書き込むデータを IP アドレスそのものではなく，IP アドレスのハッシュだけを書き込むことにより経路再構成時に必要なパケット数の削減を実現している．しかし，IP アドレスそのものはデータに含まれないため，再構成を行う側では IP アドレスを含むルータトポロジの情報すべてを知っておく必要がある．

Dean らの方式は，攻撃経路上のルータ IP アドレスを多項式の係数として扱い，攻撃経路の再構成時には多項式の連立方程式を解くことにより各ルータ IP アドレスを計算する方式である[4)]．

Savage らの方式が IPv4 ヘッダ上の Identification フィールドの 16 ビットを使うことに対し，岡崎らは Type of Service（ToS）フィールドの 8 ビットと合わせた 24 ビットを使うことでパケット再構成に必要なパケット数を削減する方式を提案した[5)]．マーキングのアルゴリズムや再構成アルゴリズムは Savage らの方式と変わらない．

Law らは，PPM 方式を応用し，各ローカルネットワークに流れるトラフィック量データを書き込み，被害者側が各ルータのトラフィック量変化を調べることで攻撃元を特定できる方式を提案した[6)]．この方式は，これまでの PPM 方式と異なり攻撃経路の再構成ではなく，攻撃元の特定を行っている．

Goodrich は，マーキングされるデータ長が高い柔軟性を持つ方式を提案した[7)]．この方式では，これまでのように IP アドレスに限定したデータではなく，様々なデータをマークすることができることから，攻撃経路の再構成や攻撃元のみの特定など，マーキング内容により用途が変えられる．Goodrich の方式はより広い PPM フレームワークの提案であるといえよう．Goodrich の方式についても詳述する．

PPM 方式は，確率的なマーキングであるために情報の再構成を行う際に必要となるパ

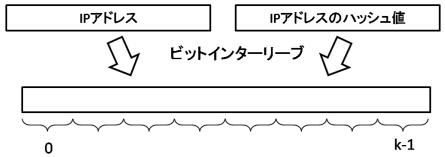

図 1　Savage らの方式におけるビットインタリーブ
Fig. 1　BitInterleave method by Savage, et al.

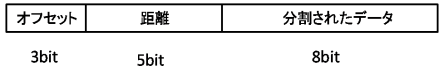

図 2　Savage らの方式におけるマーキングデータ構造
Fig. 2　Structure of marking data by Savage, et al.

ケット数が多くなることが欠点とされており，各方式では再構成に必要な平均パケット数が評価指標の 1 つとして用いられている．たとえば Savage らの方式では，1/25 の確率でマーキングが行われているときに 10 ホップ先にいる攻撃者の情報を再構成するために必要なパケット数は 1,300 パケットであることが示されている[2)]．

2.3　Savage らの方式

Savage らの方式は，被害者から攻撃者までに中継されてきたルータの列を攻撃経路として再構成することを可能にする方式である．その基本アイデアは，マーキングを開始したルータの IP 情報とその次のルータの IP 情報の排他的論理和（XOR）を情報として運ぶことにある．再構成時には，XOR された IP 情報から順次ルータ情報を復元していく．

2.3.1　マーキング手法

ルータ X は自身の IP アドレス A_X（32 ビット）に IP アドレスをハッシュした $H(A_X)$（32 ビット）を 1 ビットずつ挟み込む（ビットインタリーブ）．そしてインタリーブされた 64 ビットデータを 8 ビットごとに分割し，あらかじめ保持しておく（**図 1**）．

ルータがマーキングするデータは，分割したデータ（8 ビット）とどの分割データがマーキングされたかを示すオフセット（3 ビット）と距離情報（5 ビット）の計 16 ビットであり（**図 2**），IPv4 ヘッダの Identification フィールドを利用する．

ルータ X は中継するパケットを受け取ると，確率 p でマーキングを行う．マーキングは，まず 8 つの分割データをランダムに選択し，そのオフセット番号を書き込む．そして距離情報 d を 0 に設定する．マーキングを行わない場合，$d = 0$ の場合にはオフセット番号に従って自分の分割データを XOR する．そして d の値にかかわらず d に 1 を加える．

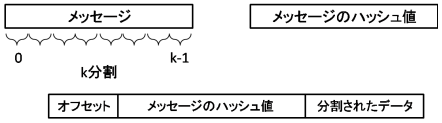

図 3　Goodrich の方式におけるマーキングデータ構造
Fig. 3　Structure of marking data by Goodrich.

2.3.2　再構成手法

再構成時には，まず同じオフセット番号を持つ $d = 0$ の分割データと $d = 1$ の分割データを XOR する．$d = 0$ の分割データは直近のルータ情報であり，マーキング時に XOR されていないことから，$d = 1$ のルータ情報が復元できる．このデータを用いて $d = 2$ の分割データと XOR することで $d = 2$ のルータ情報を復元できる．このように再帰的に各ルータの情報を復元し，分割されたデータを結合しビットインタリーブを解くことで各ルータの IP 情報と，攻撃者までの経路を得ることができる．ビットインタリーブを解いた後にハッシュ値を再計算し照合することでデータの検証を行う．

2.4　Goodrich の方式

Goodrich の方式も，情報を分割し確率的にマーキングを行うものであるが，Savage らの方式と異なりハッシュ値はビットインタリーブせず，各パケットに分割せずに書き込む．Goodrich の方式の特長は，各データサイズやハッシュ長については規定せず，可変としていることにある．また Savage らと方式と違い，距離情報をマーキングデータに含んでいない．

2.4.1　マーキング手法

ルータ X は書き込むべき情報のハッシュ値を計算し，情報だけを分割する．そして分割したデータとそのオフセット番号，そしてハッシュ値をマーキングする（**図 3**）．

2.4.2　再構成手法

同じハッシュ値を持つパケットを集め，オフセットに合わせて結合することで情報を再構成する．再構成時にはハッシュ値を再計算し照合することでデータの検証を行う．

3.　ルータトポロジのスモールワールド性

3.1　インターネットトポロジ

インターネットの普及にともない，情報科学や科学技術に活かすことを目的としたネットワークトポロジの研究・調査がさかんに行われ，インターネットの構造が明らかとなってき

ている．調査の結果，インターネットはスケールフリー性とスモールワールド性という性質を持っているということが明らかになっている．

3.1.1　スケールフリー性

ノードとリンクから構成されるグラフにおいて，ノードに接続しているリンク数（次数と呼ばれる）の分布がベキ乗則に従う性質のことスケールフリー性という．この性質を持つネットワークでは，多数のリンクを持つ少数ノード（ハブノード）と少数のリンクを持つ多数のノードが存在している．Faloutsos らはインターネット上のルータや Autonomous System（AS）のつながりからなるトポロジを調査し，これらのトポロジがスケールフリー性を持つことを明らかにした[15]．

3.1.2　スモールワールド性

ネットワーク上の任意の 2 つのノードが，間にわずかな数のノードを経由するだけで接続することができる性質をスモールワールド性という．CAIDA[16] や DIMES[17] ではインターネットのトポロジ調査を行っており，それらのデータを公開している．これらのデータを使った調査によると，AS レベルでは 4AS，Internet Router レベルでは 7 ホップ進むことで任意のノードから全体の 90%以上のノードにたどり着くことができるとされている[18]．

3.2　インターネット上のルータ間距離分布

インターネットのルータトポロジはスモールワールド性を持つことは示されているが，その分布の数値は与えられていない[18]．

この数値は，PPM 方式を評価するうえで重要な数値となる．PPM 方式の評価は，再構成に必要となるパケット数の期待値を求めることで行われていることは 2.2 節で述べたが，それらの評価は距離別の必要パケット数の期待値をグラフ化して行われたものであった．

攻撃元がインターネット上にランダムに存在すると仮定した場合，被害者から攻撃元までの距離の分布はトポロジ上の 2 点間距離の分布に従う．このことから距離分布が分かることで，距離数に依存しない必要パケット数の期待値が計算可能になる．しかし Caldarelli らの文献 18) では分布の具体的数値が与えられていないため，計算ができない．

本節では PPM 方式本来の必要パケット数期待値を計算可能にするため，CAIDA の 2003 年におけるルータトポロジデータ[16] を用いて，その分布を求める．

CAIDA データは 192,244 のルータノードと 609,066 のリンクからなる．調査は，ランダムに 2 つのノードペアを選択し，CAIDA データ上に対して Dijkstra 法を用いてノードペア間の最短距離を計測した．調査したノードペアは 1,959,065,064 個であり，これは全体の 10.6%にあたる．

表 1　CAIDA データの 2 点間距離分布
Table 1　Distance distribution on CAIDA data.

# of hop d	f_d (%)	$\sum_{i<d} f_i$ (%)	# of hop d	f_d (%)	$\sum_{i<d} f_i$ (%)
1	0.003%	0.003%	11	1.864%	98.407%
2	0.055%	0.062%	12	0.930%	99.333%
3	0.572%	0.698%	13	0.417%	99.744%
4	3.772%	4.716%	14	0.164%	99.903%
5	13.216%	18.070%	15	0.075%	99.964%
6	24.134%	41.887%	16	0.035%	99.988%
7	25.061%	66.617%	17	0.012%	99.996%
8	17.036%	83.615%	18	0.003%	99.999%
9	8.722%	92.490%	19	0.001%	100.000%
10	3.928%	96.520%	20	0.000%	100.000%

調査結果から得た距離分布を**表 1** に示す．ここから，9 ホップでインターネット上の 90%以上のノードに到達できることが分かる．また 15 ホップでインターネット上の 99.964%のノードに到達できることが分かる．ここで，f_d はノード間距離が d であるノード対の割合である．

4.　トポロジ特性を利用した確率的パケットマーキング方式

Savage らが提案した PPM 方式は距離 d を示す情報に 5 ビットを用いており $d < 32$ に対応していた．5 ビットの選択理由は，ほとんとすべてのノード間距離に対応しているためとしていた．しかし 3 章の結果を見ると，$d < 16$ で 99.964%のノードに到達可能であることから 4 ビットの距離情報で十分であるといえる．PPM でマーキングするデータは非常に少ないビット数であるためマーキング情報の 1 ビットの差はそのまま再構成に必要なパケット数期待値に関わる．

そこで，本章では距離情報を 4 ビットとし，マーキング可能なデータサイズを 1 ビット増やすことで必要なパケット数を減らす方式を提案する．提案方式は Savage らの方式を改良した改良 Savage 法と Goodrich らの方式を改良した改良 Goodrich 法の 2 つからなる．また 2 つの改良方式に加え，実運用時のハッシュ値衝突対策としてホップ数の照合作業も提案する．

4.1　改良 Savage 法

Savage らの方式では，64 ビットのインタリーブされたデータを 8 分割しマーキングする

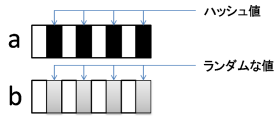

図 4　改良 Savage 法における $d > 15$ へのマーキング対応
Fig. 4　Extention for $d > 15$ case on improved Savage method.

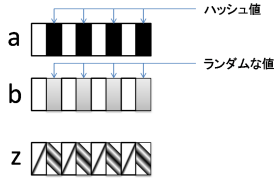

図 5　改良 Savage 法における $d > 15$ の識別
Fig. 5　Identifacation for $d > 15$ case on improved Savage method.

データを 8 ビットとしていたが，距離情報 5 ビットを 4 ビットに削減することでマーキングエリアを 9 ビットにすることが可能である．

しかし，64 ビットデータに対して，マーキングエリアが 9 ビットであると，結局分割数は 8 のままになる．そこで，ハッシュ値のサイズを 1 ビット削減し 31 ビットにし，63 ビットとすることで分割数を 7 に削減する．ハッシュ値のサイズを削減することで，ハッシュ値の衝突確率が高まり精度が低下することが考えられるが，それらについては次章で評価する．

4.1.1　$d > 15$ への対応

インターネット上のルータの 99.964%に 15 ホップで到達可能とはいえ，16 ホップ以上先にあるノードも低確率ではあるが存在する．本項ではその対応を示す．

ルータ X が受けたパケットに対して確率 $1 - p$ で距離情報を加算する際，もし距離情報 $d = 15$ であった場合，まず $d = 0$ に設定する．このパケットを a とする．次に分割されたデータの偶数番目のビットをランダムな値に入れ替えた b を用意する．そして，確率 1/2 で a または b を転送する．

再構成を行う場合を，距離 $d = 16$（パケットには $d = 0$ とマーキングされている）のケースで考える．被害者が受け取ったパケットのうち，$d = 0$ と書かれたパケットは 3 種類存在する．1 つは正しく距離が 0 であるもの，つまり直近ルータからのマーキングパケット（z）である．そしてもう 2 つは距離 16 のルータからのパケット 2 種類（a，b）である．この 3 種類のうち，2 種類は分割されたデータの奇数ビット部（IP アドレス部）が同じになり，偶数ビット部（ハッシュ部）が異なることになり，それぞれ識別が可能になる．z のパケットにおける IP アドレス部（4 ビット）は a，b の IP アドレス部と同一になってしまうことが 1/16 の確率で発生するが，$d > 15$ となることは 0.036% の確率で発生し，さらにその 1/16 であるため発生確率は小さいことからここでは a，b，z の IP アドレス部が一致する可能性を無視した．

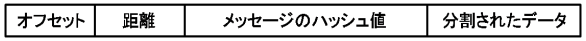

図 6　改良 Goodrich 法のマーキングデータ構造
Fig. 6　Structure of marking data on improved Goodrich method.

4.2　改良 Goodrich 法

Savage らの方式はルータの IP アドレスを基にした情報をマーキングし，攻撃元から被害者までの攻撃経路を再構築することを目的としている．一方で，Goodrich の方式は書き込まれるデータについては規定しておらず，Savage らの方式と比較してより広い概念となっている．また，そのデータ構造に距離情報を持っていない．

攻撃元のみを特定する場合においては距離情報は必要とならず，Goodrich の方式は有効であるが Savege らの方式や岡崎らの方式，Song らの方式，Dean らの方式などのように攻撃経路を明らかにする方式の場合，Goodrich の方式は必ずしも有効ではない．距離情報は通過するルータごとに加算されていくため，ハッシュされる情報とは独立して存在することが必要であるためである．

そこで本論文では Goodrich の方式を，距離情報を独立に持ったデータ構造に改良する．データ構造を**図 6** に示す．

4.2.1　$d > 15$ への対応

改良 Goodrich 方式においても，改良 Savage 法と同じく $d > 15$ への対応を行う．

ルータ X が受けたパケットに対して確率 $1 - p$ で距離情報を加算する際，もし距離情報 $d = 15$ であった場合，まず $d = 0$ に設定する．このパケットを a とする．ハッシュ値を入れ替えるためのランダムな値を生成し，a のハッシュ値とともに記憶しておく．そしてランダム値にハッシュ値入れ替えた b を用意し，確率 1/2 で a または b を転送する．距離情報 $d = 15$ のパケットが来た場合，ハッシュ値がすでに記憶している情報と一緒であるばあいは，記憶していたランダム値を用い a と b を用意し，1/2 の確率でどちらかを転送する．

4.3　ホップ数照合

これまでの PPM 方式におけるハッシュ値のサイズは Savage らや岡崎らの方式では 32 ビットであったが改良 Savage 法では 31 ビットに削減している．また改良 Goodrich 法でも距離情報を独立して設けていることから同じデータサイズを書き込む場合にはハッシュサイズを削減する必要がでてくる．この 2 つのことから，ハッシュ値の衝突がこれまでの方式よりも高くなることが予想される．

そこで，その対応として実運用時でのホップ数照合を行う．マーキングされたデータの再構成時に，再構成により得られたルータ IP アドレスに対し Tracepath コマンドを使いそのホップ数を確認する．そのホップ数がマーキングされている距離情報と一致するかを調べる．Tracepath コマンドでは調査する経路の往路と復路が異なる場合にその旨を表示し，さらに復路のホップ数を表示する機能を持つ．復路の経路は攻撃者にとっての攻撃経路と考えることができるため，往路と復路が異なる場合は復路のホップ数を用いて照合を行う．これにより，ハッシュ衝突が起きた際の誤った再構成の回避率を高める．

5. 提案方式の評価

本章では，前章で提案した改良 Savage 法と改良 Goodrich 法の評価を行う．提案方式の評価には必要パケット数の期待値とハッシュ値の衝突確率の 2 つを行う．これらの評価はこれまでの PPM 方式でも評価されてきた項目であり，PPM 方式の評価として一般的である．

提案方式の評価は計算式とトポロジの分布データからの計算により行われるが，さらに実装の評価も行う．実装評価では実際に Linux サーバに提案機構を実装し，その負荷を測定する．

5.1 提案方式の具体的構成

評価に用いるそれぞれの方式の具体的構成は，まず書き込む IPv4 ヘッダ領域により 2 つに分ける．1 つは Identification フィールドの 16 ビットを使ったもの（以後，16 ビット版）であり，もう 1 つは Identification16 ビットと Type of Service（TOS）フィールドの合わせて 24 ビットを使ったもの（以後，24 ビット版）である．いずれもマーキングされる情報はルータの IP アドレス（32 ビット）である．

16 ビット版では，比較対象として Savage らの方式をあげる．他の方式については，たとえば Song らの方式は必要数パケット数の期待値は低いが前提条件としてルータトポロジの知識を必要であるなどの点や，Law らの方式は PPM 方式そのものではなく応用方式である点から比較対象外とした．また Dean らの方式は PPM 方式の提案であるものの，必要とするパケット数の期待値は Savege らの方式と比較してはるかに多いものとなっているため，性能として Savage らの方式が上であることから比較対象から外した．そして岡崎らの方式は 24 ビットを使ったものであることから，24 ビット版の比較として用いる．

16 ビット版では，改良 Savage 法はハッシュ値サイズ（以後 H）は 31 ビットを用いて，距離情報のサイズ（以後 D）は 4 ビット，オフセットのサイズ（以後 O）は 3 ビット，データ領域は（L）9 ビットとした．この場合，分割数（以後 F）は 7 になる．この条件の表記を (H, D, O, L, F) の形式で表すこととする．16 ビット版改良 Savage 法は $S'[16] = (31, 4, 3, 9, 7)$，Savege らの方式は $S = (32, 5, 3, 8, 8)$ で表される．また改良 Goodrich 法は分割数を複数用意し，それぞれ $G[16]_1 = (5, 4, 2, 8, 4)$，$G[16]_2 = (9, 4, 3, 4, 8)$ とした．

24 ビット版では，岡崎らの方式は $O = (32, 5, 2, 16, 4)$，改良 Savage 法は $S'[24] = (22, 4, 2, 18, 3)$，また改良 Goodrich 法は分割数を複数用意し，それぞれ $G[24]_1 = (3, 4, 1, 16, 2)$，$G[24]_2 = (10, 4, 2, 8, 4)$ とした．

5.2 必要パケット数の期待値

期待値を求めるにあたり，インターネットトポロジ上の全ルータが各 PPM 方式を採用しているとし，また同じ確率 p でマーキングを施すものとする．

距離 d においてマーキングされたパケットが被害者のもとに届く確率は，$p(1-p)^{d-1}$ となる．攻撃者が距離 d にいる場合での攻撃経路の再構築には，距離が $0, \ldots, d$ のすべてのパケットを収集しなければならず，それぞれの距離によって被害者のもとに届く確率は異なる．しかしこれまでの PPM 方式の論文では評価に用いるこの確率をすべて距離 d からの確率として扱ってきた．本論文でもこの確率を採用する．

各距離からマーキングされたパケットが $p(1-p)^{d-1}$ である場合，距離 d から攻撃されている場合における再構成に必要なパケット数は，「集めなければならないマークされたパケットの数の期待値」をこの確率で割ったものとなる．ここで，「集めなければならないマークされたパケット数の期待値」は単純に「集めるべきマークされたパケットの種類数」ではない．これはクーポンコレクタ問題（the coupon collectors problem）と呼ばれるものに帰着する．

クーポンコレクタ問題は，M 種類あるクーポンがランダムに与えられる場合に M 種類すべてを集めるのに必要なクーポン数 n の期待値を求めるものである．このとき n の期待値は

$$nH_n \tag{1}$$

で与えられる．ここで H_n はオレの調和数（Ore's Harmonic Number）であり，

$$H_n = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$$

で表される．オレの調和数 H_n は次の式で近似される．

$$H_n < \ln n + \gamma + \frac{1}{2n} \tag{2}$$

ここで γ はオイラーの定数であり，$\gamma = 0.5772156649\cdots$ である．

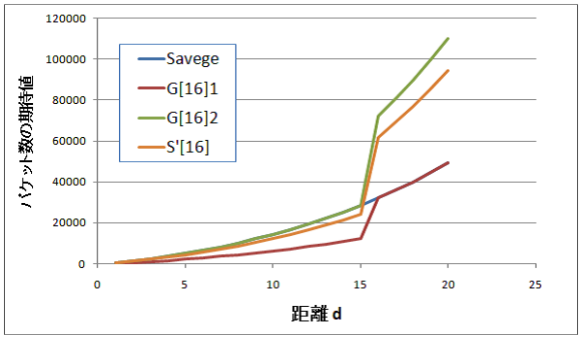

図 7　16 ビット版での $E_d[n]$
Fig. 7　$E_d[n]$ on 16 bit version.

この結果，クーポンコレクタ問題の期待値とオレの調和数の上界を用いることで距離 d からの攻撃者の経路再構成に必要なパケット数の期待値 $E_d[n]$ は以下の式になる．

$$E_d[n] = \frac{ml\ln(ml) + \gamma ml + 1/2}{p(1-p)^{d-1}} \tag{3}$$

ここで m は攻撃者の数，l はマーキングされるデータの分割数である．なお改良 Savage 法と改良 Goodrich 法では $d > 15$ のときにパケット種類数が 2 倍になる．また，距離に依存しない必要パケット数期待値 $E[n]$ は，3 章で得られた分布 f_d を用いて

$$E[n] = \sum_{d=1}^{\infty} f_d E_d[n] \tag{4}$$

で求められる．

5.3　16 ビット版の評価

S，$S'[16]$，$G[16]_1$，$G[16]_2$ の評価を行う．評価にあたり，$m = 1$，$p = 1/25$ とした．各方式の $E_d[n]$ を距離ごとに表したもの**図 7** に示し，各方式の $E[n]$ を**表 2** に示した．

この結果，$S'[16]$ は Savega らの方式（S）と比較して再構成に必要なパケット数を 14.83%，$G[16]_1$ は 57.27%減少させた．また $G[16]_2$ は 0.18%の増加となった．

5.4　24 ビット版の評価

O，$S'[24]$，$G[24]_1$，$G[24]_2$ の評価を行う．評価にあたり，$m = 1$，$p = 1/25$ とした．各方式の $E_d[n]$ を距離ごとに表したもの**図 8** に示し，各方式の $E[n]$ を**表 3** に示した．

表 2　16 ビット版での $E[n]$
Table 2　$E[n]$ on 16 bit version.

方式	$E[n]$	S との比較
S	8466.506	1.0
$S'[16]$	7210.663	0.8517
$G[16]_1$	3617.853	0.4273
$G[16]_2$	8481.562	1.0018

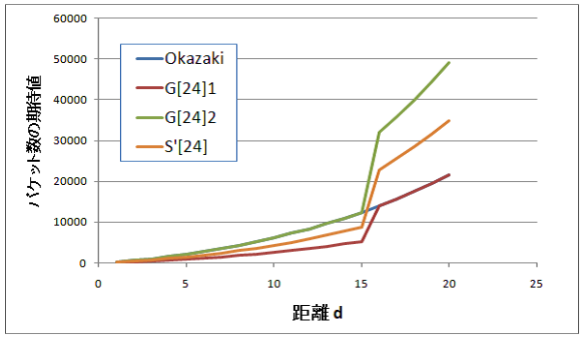

図 8　24 ビット版での $E_d[n]$
Fig. 8　$E_d[n]$ on 24 bit version.

表 3　24 ビット版での $E[n]$
Table 3　$E[n]$ on 24 bit version.

方式	$E[n]$	O との比較
O	3611.087	1.0
$S'[24]$	2520.995	0.6981
$G[24]_1$	1501.462	0.4158
$G[24]_2$	3617.853	1.0018

この結果，$S'[24]$ は岡崎らの方式（O）と比較して再構成に必要なパケット数を 30.19%，$G[24]_1$ は 58.42%減少させた．また $G[24]_2$ は 0.18%の増加となった．

5.5　衝 突 耐 性

PPM 方式のほとんどは，マークされるデータはマーキングエリアには収まらないため分割されてマーキングされる．再構成を行ったときのデータ正確性の検証用としてハッシュ値

が付与されていることが一般的である．本論文での提案方式もハッシュ値を検証に利用しており，Savege らの方式と Goodrich らの方式もハッシュ値を検証に採用している．

しかし，ハッシュ値のサイズ自体もマーキングエリアの制限により少ないものとなっており，衝突の危険性は高い．ハッシュ値の衝突が発生すると，再構成時に正しくない組合せを正しいと誤認識することになり，攻撃経路の復元が困難になることや間違った攻撃経路の導出がされてしまう．たとえば Savage らの方式では，経路の復元を XOR により距離 $d = 1$ から順に求めていくため，ハッシュ値の衝突が発生すると，衝突が起こった場所以降の経路がすべて間違った経路となってしまう．

本節では各方式の衝突耐性の評価を行う．衝突耐性の評価はハッシュ値が衝突する確率を求めて行う．

Savage らの方式や改良 Savage 法では，ハッシュ値も分割されるため，再構成にあたり同じ距離情報をもとにすべてのオフセットを集めて，ビットインタリーブを解いた後に検証を行う．Savage らの方式や改良 Savage 法のハッシュ値の衝突確率 P_S は攻撃者が複数の場合以下の式で求められる．

$$P_S = 1 - \left(1 - \frac{1}{2^h}\right)^{m^l} \tag{5}$$

ここで h はハッシュ値のサイズ（ビット），m は攻撃者数，l は分割数である．

一方，Goodrich の方式は改良 Goorich 法では，ハッシュ値は分解されない．再構成にあたっては，ハッシュ値（改良 Goodrich 法では距離情報も）が同じになるパケットからすべてのオフセットを集めて，検証を行う．

Goodrich の方式や改良 Goodich 法のハッシュ値の衝突確率 P_G は以下の式で求められる．

$$P_G = 1 - \left(1 - \frac{1}{2^h}\right)^{m} \tag{6}$$

P_G は P_S と比較すると，分割数 l に依存した衝突確率ではなくなっている．これは Goodrich の方式らはハッシュ値が分割されていないことから，ハッシュ値を再計算する前に同一ハッシュのパケットだけを集めることができるため，衝突の発生が分割による影響を受けないためである．

3.3 節で提案したホップ数照合を用いると，衝突を一定の確率で回避可能である．ハッシュ衝突が起こった際の距離情報と，traceroute コマンドを用いて距離情報を測定した際に距離情報が一致する確率は $\sum_d {f_d}^2$ と考えることができるためホップ数照合適用後の P_S，P_G をそれぞれ P'_S，P'_G とすると

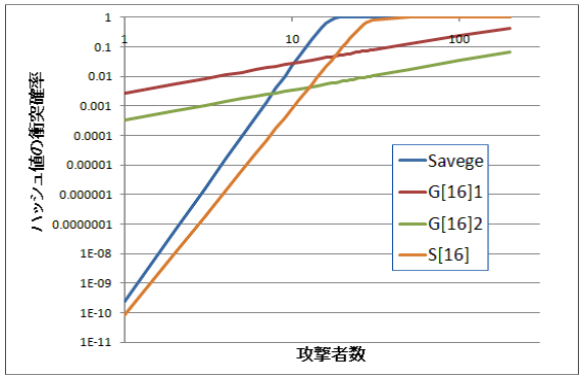

図 9　16 ビット版での衝突確率
Fig. 9　Collision probability on 16 bit version.

$$P'_S = 1 - \left\{\left(1 - \frac{1}{2^h}\right) + \frac{1}{2^h}\left(1 - \sum_d {f_d}^2\right)\right\}^{m^l} \tag{7}$$

$$P'_G = 1 - \left\{\left(1 - \frac{1}{2^h}\right) + \frac{1}{2^h}\left(1 - \sum_d {f_d}^2\right)\right\}^{m} \tag{8}$$

と求められる．なお，3 章でもとめた分布の結果，$\sum_d {f_d}^2 = 0.1762$ であった．

16 ビット版の衝突確率と 24 ビット版の衝突確率をそれぞれ**図 9**，**図 10** に示す．Savage らの方式，岡崎らの方式については P_S で求め，その他の提案方式は P'_S または P'_G で求めた．

16 ビット版では $S[16]$ は攻撃者数にかかわらず S より衝突確率が低いことが分かる．また $G[16]_1$，$G[16]_2$ は攻撃者数 20 未満という少ない状況では S より衝突確率が劣るが攻撃者数が増加するにつれ S より衝突確率が低くなることが分かる．

24 ビット版では $S[24]$ は攻撃者数が 200 を超える程度まで O より衝突確率が高いことが分かる．また $G[24]_1$，$G[24]_2$ は攻撃者数 100 未満という少ない状況では O より衝突確率が劣るが攻撃者数が増加するにつれ O より衝突確率が低くなることが分かる．

5.6　実装による負荷の評価

PPM 方式は，ロギング手法や ICMP 手法と比較してストレージが不要な点や余計なトラフィックを生まない点が利点であった．一方で，PPM 方式にはパケット再構成のために

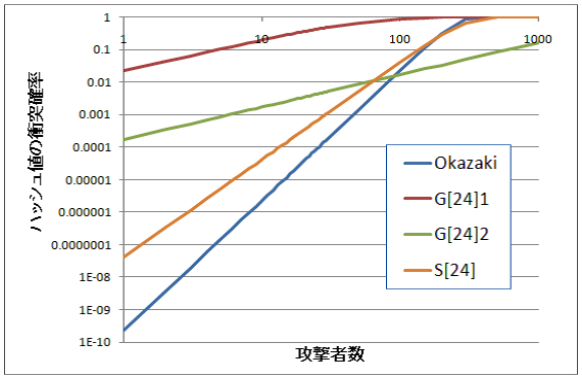

図 10　24 ビット版での衝突確率
Fig. 10　Collision probability on 24 bit version.

必要となるパケット数の収集やハッシュ値の衝突に難点があった．提案方式ではそれらについて従来方式より少ないパケット数と高い衝突耐性を実現するものであった．

しかし，PPM をすることでルータの負荷が高くなるようでは，PPM 方式の利点が少なくなり問題である．これまでの PPM 方式の研究では，パケット数期待値の計算や衝突耐性の計算などは行われてきたが，実装による負荷の調査は行われていなかった．本節では，PPM を実環境で運用した場合のルータの負荷を調査する．

実装は Linux サーバマシンに Network Interface Card（NIC）を 2 枚用いて Linux ルータとして稼働させることで実現した．PPM 方式の実装には，カーネル中の IP パケット出力に関するプログラムを修正し，確率による条件分岐と IPv4 ヘッダの Identification エリアの書き換え処理を追加し，カーネルを再構築した．なお，ハッシュ値の計算や隣接ルータの IP アドレスとの XOR や分割は事前にすませていることを仮定している．PPM 対応ルータの実装環境を**表 4** に示す．

ルータのスループットと CPU 使用率の 2 点から負荷を評価する．実験は，PPM 対応ルータを挟みクライアントとサーバを用意し，サーバからクライアントに対し netperf を用いてペイロードサイズを 1 バイトから 10001 バイトまで 500 バイトごとに 10 秒間送信を行う．その際 PPM ルータ上で送信中のスループットと CPU 使用率を取得する．

PPM は確率 p でマーキングを施し，確率 $1-p$ では距離情報を加算する．つまり，PPM 方式では確率 p をいかに設定しようと，すべてのパケットに対しなんらかの情報付加を行う作業が発生する．そこで，加算の負荷を測る目的として余分な処理をせず単に IP ヘッダの Identification フィールドに定数を書き込む処理を行うカーネルを用意した．また，マーキング処理の負荷を測るために，マーキング確率を変えた 2 つのカーネルを用意した．

表 4　PPM 実装環境の詳細
Table 4　PPM implementation environment detail.

CPU:	Intel(R) Core 2 Duo E8400　3.00 GHz
メモリ：	2 GB
NIC:	Marvell Yukon 88E8057 Intel(R) PRO/1000 PT Desktop Adapter
Linux Kernel:	2.6.18
Linux Distribution:	Cent OS 5.3

表 5　PPM 負荷測定の結果
Table 5　Results of performance test for PPM.

カーネル種類	スループット (Mbps)	CPU 使用率 (%)
Default	877.29	2.94
ID only	878.59	2.83
PPM004	873.70	2.96
PPM050	879.05	3.07

準備したカーネルは以下の 4 つである．

(1)　通常の Linux（以後 Default）
(2)　Identity フィールドに定数を書き込む（以後 ID only）
(3)　PPM（マーキング確率 4%，以後 PPM004）
(4)　PPM（マーキング確率 50%，以後 PPM050）

結果を**表 5** に示す．この表から，スループットと CPU 使用率双方において Default のものと有意な差は認められなかった．このことから，マーキング処理も加算処理もほとんどルータの処理に負荷をかけるものではないことが分かった．Default での処理に比較して，PPM004 の処理が高いスループットを出す結果となっているがこれは誤差の範囲と考える．

実験は Linux ルータにおいて行ったが，ルータやスイッチのアプライアンスにも Linux ベースで稼働しているものもあることから PPM の負荷については同様の結果が得られることが予想される．

5.7　実運用に向けた考察

PPM 方式では IP パケットの Identification フィールドを利用してマーキングを行うこ

とが共通して行われており，さらには Type Of Service フィールドを使われたものも存在する[5),7)]．PPM 方式ではこれらのフィールドが現在ほとんど利用されていないことを前提にマーキングが行われている[2)]．しかし，これらのフィールドは PPM 方式のマーキングを行うことが本来の用途ではなく実際に運用を行った場合に，相互に影響を及ぼし合う可能性がある．

特に Type Of Service フィールドは，RFC 2417 で QoS を行うための DiffServ に用いるフィールドとしてあらたに定義が追加されており，Type Of Service フィールドを用いた PPM 方式は DiffServ で利用されている QoS を阻害する可能性が大きい．また，利用されている可能性が少ない Identification フィールドにおいても，PPM 方式と同様に独自方式としてデータが書き込まれている可能性もある．

ハッシュ値による検証が含まれることから，衝突が起きない限り本来の用途や他手法での Identification フィールドの利用などは検証により排除することができる．しかし他の手法が PPM 方式を上回る規模で広範囲に適用されている場合は可能性としては考えられ，その場合においては，PPM 方式と他の手法が互いに影響を及ぼす可能性が出てくるため，広範囲での利用にあたっては他手法との関係を考慮する必要がある．

6. ま と め

IP トレースバック技術の 1 つである確率的パケットマーキング（Probabilistic Packet Marking，PPM）方式は，これまで多くの研究が行われてきた．しかしこれまでの研究では，ネットワークのトポロジを特長を利用している PPM 方式はなかった．

そこで本論文では，最初にインターネットのトポロジの特長であるスモールワールド性に着目しその到達性を調査した．その結果，任意のノードから 16 ホップ未満で到達できるノードの割合が 99.964%にのぼるという結果を得た．そしてこの結果を利用し，従来の PPM 方式では 5 ビット利用されていた距離情報を 4 ビットにして書き込めるデータ量を 1 ビット増加させる方式を提案した．提案方式は 4 ビットでは表現できないごくわずかなノードに対しても対応可能な方式である．さらに，運用面での対応としてトレースバックに必要な情報を再構成後にホップ数照合を行うことで，PPM 方式の欠点でもある衝突確率の高さを削減させる方式を提案した．

提案方式の評価の結果，従来方式よりも少ないパケット数でトレースバックに必要な情報を再構成可能であることが示された．また衝突耐性についても従来方式よりも DDoS に強い方式であることが確認された．評価は期待値や確率の計算だけではく，実装についても行った．これまでの PPM に関する研究では実装による負荷評価を行ったものはなかった．提案方式を Linux カーネルに適用し，スループットと CPU 使用時間の負荷を調査し，提案方式がマーキング確率にかかわらず通常のルーティング動作への影響がほぼ認められないことが示された．

本論文の提案方式により，より効果的な PPM 方式が示されたことに加え，実装における負荷は十分に低いものであることが示された．

参 考 文 献

1) Peng, T., Leckie, C. and Ramamohanarao, K.: Survey of Network-Based Defense Mechanisms Countering the DoS and DDoS Problems, *ACM Computing Surveys*, Vol.39, Issue 1 (2007).
2) Savage, S., Wetherall, D., Karlin, A.R. and Anderson, T.: Practical network support for IP Traceback, *Proc. ACM SIGCOMM*, pp.295–306 (2000).
3) Song, D. and Perrig, A.: Advanced and Authenticated Marking Schemes for IP Traceback, *Proc. IEEE INFOCOM*, pp.876–886 (2001).
4) Dean, D., Franklin, M. and Stubblefield, A.: An algebraic approach to ip traceback, *Proc. Network and Distributed System Security Symp* (*NDSS*), pp.3–12 (2001).
5) 岡崎直宣，河村栄寿，林　美娘：サービス不能攻撃の経路追跡手法の効率化に関する検討，情報処理学会論文誌，Vol.44, No.12, pp.3197–3201 (2003).
6) Law, T.K.T., Yau, D.K.Y. and Lui, J.C.S.: An effective statistical methodology to trace back DDOS attackers, *IEEE Trans. Parallel Distrib. Syst.*, Vol.16, No.9, pp.799–813 (2005).
7) Goodrich, M.T.: Probabilistic Packet Marking for Large-Scale IP Traceback, *IEEE/ACM Trans. Networking*, Vol.16, No.1, pp.15–24 (2008).
8) Stone, R.: CenterTrack: An IP Overlay Network for Tracking DoS Floods, *Proc. 9th Conference on USENIX Security Symposium* (2000).
9) Snoeren, A.C., Partidge, C., Sanchez, L.A., Jones, C.E., Tchakountio, F., Kent, S.T. and Strayer, W.T.: Hash-Based IP Traceback, *Proc. 2001 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications* (*SIGCOMM 2001*) (2001).
10) Baba, T. and Matsuda, S.: Tracing Network Attacks to Their Sources, *IEEE Internet Computing*, Vol.6, No.2, pp.20–26 (2002).
11) Hazeyama, H., Oe, M. and Kadobayashi, Y.: A Layer-2 Extension to Hash-Based IP Traceback, *IEICE Trans. Information and Systems*, Vol.E86, No.11 (2003).
12) Andreou, M.S. and Moorsel, A.V.: Logging Based IP Traceback in Switched Ethernets, *Proc. 1st European Workshop on System Security* (2008).

13) Bellovin, S.M., Leech, M. and Taylor, T.: ICMP traceback messages, Obsolete Internet draft (2003).
14) Izaddoost, A., Othman, M. and Rasid, M.F.A.: Accurate ICMP TraceBack Model under DoS/DDoS Attack, *Proc. 15th International Conference on Advanced Computing and Communications* (2007).
15) Faloutsos, C., Faloutsos, P. and Faloutsos, M.: On Power-Law Relationships of the Internet Topology, *Proc. ACM SIGCOMM* (Sep. 1999).
16) CAIDA: CAIDA's Router-Level Topology Measurements. http://www.caida.org/tools/measurement/skitter/router_topology/
17) The DIMES project. http://www.netdimes.org/new/
18) Caldarelli, G. and Vesprignani, A.: *Large scale structure and dynamics of complex networks*, World Scientific Publishing Company (2007).
19) Belenky, S. and Ansari, N.: IP Traceback With Deterministic Packet Marking, *IEEE Communications Letters*, Vol.7, No.4, pp.162–164 (2003).

(平成 22 年 5 月 17 日受付)
(平成 22 年 11 月 5 日採録)

金岡　晃（正会員）

2004 年年筑波大学大学院博士課程システム情報工学研究科修了．同年セコム株式会社入社．筑波大学大学院システム情報工学研究科研究員を経て 2008 年より筑波大学大学院システム情報工学研究科助教．ネットワークシステムの安全設計方式，電子認証に関する研究に従事．博士（工学）．IEEE, ACM, 電子情報通信学会各会員．

岡田　雅之

2000 年東邦大学大学院理学研究科情報科学専攻修士課程修了．同年 ISP，通信事業者においてネットワーク設計，運用に従事．2004 年社団法人日本ネットワークインフォメーションセンター入社，2008 年筑波大学大学院システム情報工学研究科リスク工学専攻博士後期課程．インターネット経路制御，IP アドレス管理・運用に関するシステムの研究と開発に従事．

勝野　恭治（正会員）

1998 年慶應義塾大学大学院理工学研究科計算機科学専攻修士課程修了．同年日本アイ・ビー・エム株式会社入社．東京基礎研究所主任研究員．2009 年筑波大学大学院システム情報工学研究科リスク工学専攻後期博士課程修了．2003 年ソフトウェア科学会高橋奨励賞受賞．情報セキュリティ，コンピュータ・ネットワーク，エージェント技術に関する研究開発に従事．博士（工学）．日本ソフトウェア科学会会員．

岡本　栄司（正会員）

1973 年東京工業大学工学部電子工学科卒業．1978 年同大学院博士課程修了．工学博士．同年日本電気中央研究所入社．その後，北陸先端科学技術大学院大学，東邦大学をへて 2002 年より筑波大学教授．情報セキュリティの教育・研究に従事．1990 年電子情報通信学会論文賞，1993 年本会ベストオーサ賞受賞．著書『暗号理論入門』（共立出版），『電子マネー』（岩波書店）等．