Vol.2011-DPS-146 No.7
Vol.2011-CSEC-52 No.7
2011/3/10

カウンタを用いたIPトレースバック方式の評価

唐沢智之[†1,*1] 双紙正和[†2] 宮地充子[†1]

分散サービス不能攻撃（DDoS 攻撃）の対策の一つとして，IP トレースバックが研究されている．本論文では，効率のよいトレースバックを実施するため，パケットサンプリングの相関を攻撃経路全体にわたって高めることができるような IP トレースバック方式を提案する．具体的には，サンプリング回数をパケット上のカウンタに記録し，その値を利用してサンプリング確率を決定する．さらに，本論文では，提案方式の詳細な解析を行い，その結果として，関連研究よりも効率が良いトレースバックが実行できることを示す．

Evaluation of an IP Traceback Scheme with Counters

Tomoyuki Karasawa ,[†1,*1] Masakazu Soshi [†2]
and Atsuko Miyaji[†1]

In this paper we propose an efficient IP traceback scheme, which improves correlation of packet sampling along attack paths in presence DDoS Attacks. For that purpose, first we store the number of sampling on an area designated as a counter on each packet. Then we use the value to determine the probability of sampling the packet. Theoretical and experimental analysis of our scheme show that it is more effective than previous IP traceback schemes.

1. 序　　論

近年，インターネットにおけるセキュリティの脅威が増しているが，その中でも特に深刻な被害をもたらす攻撃の一つとして，サービス不能攻撃（Denial of Service Attacks，DoS 攻撃）が挙げられる．これは，攻撃者が，特定のサーバ（攻撃対象ホスト，victim と呼ばれる）に対し大量のパケットを送ることで，そのサーバの機能を停止させる攻撃である．さらに近年では，複数の攻撃者が DoS 攻撃を行う分散サービス不能攻撃（Distributed Denial of Service Attack，DDoS 攻撃）が大きな問題となっている[9]．

現状では，DDoS 攻撃に対する有効な対策がいまだ確立していない．これは，DDoS 攻撃における攻撃者の数が数千から数万にも及ぶほど大規模であることに加え，パケットの発信元の IP アドレスを記録するフィールドである “Source IP Address” が攻撃者によって偽造が容易なためである．

現在，DDoS 攻撃の有効な対策法の一つとして，IP トレースバック技術[2)–4),6),10),11)] が研究されている．IP トレースバックでは，攻撃者と victim を結ぶ経路上の各ルータが，中継するパケットに関する情報を，そのパケットや自身に記録する．そして，victim とルータがそれらの情報を解析することにより，攻撃の発信元を特定する．このような IP トレースバックは，攻撃者の情報を記録するサンプリングフェーズと，その情報を解析し，攻撃経路を復元するトレースバックフェーズから構成されている．

本研究では，2004 年に Li らによって提案された複合型 IP トレースバック方式[6] に着目し，パケットにカウンタ機能を実装することで，Li らの方式の問題点を克服した IP トレースバック方式を提案する．本論文は，我々が 2008 年に発表した方式[5] に，詳細な理論的解析を行うものである．

2. 関 連 研 究

IP トレースバック方式は，大きく，確率的パケットマーキング法とロギング法に分類される．確率的パケットマーキング法（probabilistic packet marking, 以下 PPM と略す）とは，ルータが，受け取ったパケットのヘッダに確率的に攻撃経路の情報を書き込む（マーキングする）方式である[3),11)]．PPM では，ルータ内部にパケットの情報を保存しておく必要がないため，ルータのストレージ容量に依存しないという利点がある．反面，パケットには断片的な経路の情報しか保存されていないので，攻撃経路を復元するためには，多数のパケットが必要となるという欠点がある．

一方，ロギング法[2),10)] では，ルータがパケットを中継するとき，その情報を自身に保存（ロギング）する．ロギング法は，トレースバックに必要なパケットの数が少なくて済むが，ルータの負荷が高く，かつ膨大な量のログが必要となることが大きな問題となる．

†1 北陸先端科学技術大学院大学
Japan Advanced Institute of Science and Technology (JAIST)
†2 広島市立大学
Hiroshima City University
∗1 現在，(株）インターネットイニシアティブ（IIJ）
Presently with Internet Initiative Japan Inc.

Vol.2011-DPS-146 No.7
Vol.2011-CSEC-52 No.7
2011/3/10

そこで，PPM とロギング法の利点をハイブリッドに備える方式が研究されている．その中でも，Li らの方式[6] は，隣接する 2 ルータ間でのサンプリング相関を高める工夫がなされており，その結果として，各ルータにおけるサンプリング確率が 3%程度でよい．また，その性能について情報理論的な解析が詳細になされている．したがって非常に有望な方式であるといえるが，その一方で，攻撃経路全体でのサンプリング相関については考慮がなされていない，また，パケットにおける 1 ビットしか利用していないといった問題があり，さらなる向上の余地がある．そこで本研究では，これらの問題点を解決する効率の良い IP トレースバック法を提案する．

3. 提 案 方 式

3.1 基本的なアイデア

本研究では，パケットサンプリングの相関を Li らの方式[6] より向上させた，効率のよいトレースバック方式を提案する．このための基本的なアイディアは以下のとおりである．

Li らの方式は，トレースバックフェーズの成功確率を高めるために，隣接 2 ルータ間における，パケットのサンプリングの相関を高めることを目的とした．そのために，パケットにおける 1 ビットの領域をマーキングに利用している．提案方式では，このパケットサンプリングの相関を，隣接 2 ルータ間ではなく，攻撃経路全体で高めることを考える．

サンプリング相関を経路全体で高めるために，まず，Li らの方式で使用された 1bit のマーク用領域を，4bit の領域へ拡張し，それをカウンタとして利用する．ここで，パケット P のカウンタを P.counter とし，初期値は 0 とする ⋆1．そして，経路上のルータは，パケット P を確率的にサンプリングする．このときサンプリングが行われた場合は，P.counter の値を 1 増やす．すなわち，victim に到着したパケット P の P.counter の値は，経路上の複数のルータによって P がサンプリングされた回数を表すことになる．そこで，直感的には，カウンタの値が大きいパケットほど，経路上の複数のルータに多くの情報が格納されていることになり，トレースバックの際に有用であると考えられる．

以上より，カウンタの値が大きいパケットを優先的に利用することができれば，より効率的なトレースバックが可能となることがいえる．

3.2 サンプリング

カウンタの値が大きいパケットを優先的に利用するためには，直感的にいえば，カウンタの値が大きいパケットをより高い確率でサンプリングし，カウンタの値が小さいパケットはより小さい確率でサンプリングするようにすればよい．このためには，実数 $\alpha \geq 1$, β, M を定数とし，ルータは，カウンタの値 P.counter に依存した確率

$$p = ((P.\text{counter})^{\alpha} + \beta)/M \quad (1)$$

によってサンプリング動作を行えばよい．パケットがサンプリングされた場合，カウンタを +1 インクリメントする．インクリメント後に，パケット P のダイジェストをルータへ保存する．ダイジェストには，Li らの方式と同様に Bloom Filter というデータ構造を用いる．

以上による，提案アルゴリズムを**図 1** に示す．

For each packet P
  with probability $\frac{(P.counter)^{\alpha}+\beta}{M}$
    P.counter $\leftarrow$ P.counter $+ 1$;
    Store digest;

図 1 提案サンプリングアルゴリズム
Fig. 1 Proposed Sampling Algorithm

3.3 トレースバック

Li らの方式と同様に，victim に到着した攻撃パケットを用いて，トレースバックを行う．トレースバックは，victim が攻撃を受けたことを感知した瞬間に開始される．victim は，受信した攻撃パケットの集合 $\mathcal{N}$ から，ある値 threshold について，カウンタ値 P.counter $\geq$ threshold を満たすパケットの集合 $\mathcal{N}'$ を構成する．ここで，カウンタの値が大きいパケットほど，高い相関をもって攻撃経路上のルータにサンプリングされていることに注意せよ．そこで，threshold を高い値とすれば，より有意義な（つまり，高いサンプリング相関をもつ）パケットを利用してトレースバックを行うことができるので，Li らの方式と比べて，トレースバックの際の false positive ⋆2 を格段に小さくすることが可能である．一方，threshold の値をあまり高くしすぎると，トレースバックに利用できるパケットの数は減少し，トレースバックは困難になる．これらの点を考慮して，victim は threshold の値を任意に設定することができる．

提案方式におけるトレースバックアルゴリズムでは，Li らのトレースバック方式と同様

⋆1 カウンタのための領域としては，IP Identification Field を用いることが考えられる．

⋆2 この場合の false positive とは，実際には攻撃経路にそのルータが存在していないにもかかわらず，存在しているとみなすということである．

Vol.2011-DPS-146 No.7
Vol.2011-CSEC-52 No.7
2011/3/10

に，$\mathcal{N}'$ を隣接するルータ R へ送信し，もし $\mathcal{N}'$ の要素が R のサンプリングログに含まれていた場合，R を攻撃経路上であると判断する．次に R は，自身に隣接するルータ R' へ，$\mathcal{N}'$ を送信し，以下同様にして，トレースバック手続きを進めていく．Li らとの相違点は，提案方式では，経路上の各ルータは，カウンタ値 P.counter $\geq$ threshold を満たすパケットの集合を新たに構成する必要がないという点である．これは提案方式の $\mathcal{N}'$ が小さいため，Bloom Filter の false positive が無視できるからである．そこで，提案方式では，計算量の観点からも効率の良いトレースバックを行うことができる．

4. 理論的評価

この節では，提案方式の理論的評価を行う．その際，まず，攻撃者から victim に至る攻撃経路を考える．その経路において攻撃者と隣接するルータを R_1 とし，以下順に $R_2, R_3, \ldots, R_{n-1}, R_n$ とする．このとき，victim と隣接するルータが R_n であり，n は victim と攻撃者の間のルータの数である．

4.1 サンプリングの定式化

次に，提案トレースバック方式におけるサンプリングフェーズについて議論する．パケット P がルータ R_i に到着したとき，P のカウンタ値を表す確率変数を X_{c_i} とする．単純のため R_1 が受け取るパケットのカウンタ値の値を 0 とすると，特に $i = 1, 2$ のとき，

$$\Pr(X_{c_1} = c) = \begin{cases} 1 & c = 0 \text{ のとき} \\ 0 & \text{otherwise} \end{cases}$$

$$\Pr(X_{c_2} = c) = \begin{cases} \beta/M & c = 1 \text{ のとき} \\ 1 - \beta/M & c = 0 \text{ のとき} \\ 0 & \text{otherwise} \end{cases}$$

である．一般的には，

$$\begin{aligned} \Pr(X_{c_i} = c) &= \Pr(X_{c_i} = c \mid X_{c_{i-1}} = c-1) \times \Pr(X_{c_{i-1}} = c-1) \\ &\quad + \Pr(X_{c_i} = c \mid X_{c_{i-1}} = c) \cdot \Pr(X_{c_{i-1}} = c) \\ &= \frac{(c-1)^\alpha + \beta}{M} \cdot \Pr(X_{c_{i-1}} = c-1) + \left(1 - \frac{c^\alpha + \beta}{M}\right) \cdot \Pr(X_{c_{i-1}} = c) \quad (2) \end{aligned}$$

となる．以下，$\Pr(X_{c_i} = c)$ は，再帰的に求めることができる．ただし，R_i においてとりうるカウンタ値の値は非負で，その最大値は $i-1$ であり，最小値は 0 である．よって，$c < 0$ または $i - 1 < c$ のとき，$\Pr(X_{c_i} = c) = 0$ である．

次に，パケット P に関し，確率変数 X_{p_i} を以下のように定義する．

$$X_{p_i} = \begin{cases} 1 & \text{ルータ } R_i \text{ が } P \text{ をサンプリング} \\ 0 & \text{otherwise} \end{cases}$$

ここで，X_{p_i} は，確率変数 X_{c_i} によって以下のように求められる．

$$\Pr(X_{p_i} = 1 \mid X_{c_i} = c) = (c^\alpha + \beta)/M$$

よって，

$$\Pr(X_{p_i} = 1) = \sum_{c=0}^{i-1} \frac{c^\alpha + \beta}{M} \cdot \Pr(X_{c_i} = c) \quad (3)$$

以上から，式 (2), (3) によって，X_{p_i} の確率分布を求めることができる．

4.2 トレースバックの評価

この節では，提案トレースバック方式における，トレースバックフェーズについて考える．

ルータ R_i から R_{i-1} へのトレースバックすることを考える．3.3 節で議論したように，トレースバックに際しては，R_i のログの，threshold 以上の数のパケットが R_{i-1} のログにサンプリングされているとき，R_{i-1} は攻撃経路上のルータにあると判断する．以下では，単純のため threshold の値は 1 とし，トレースバックの情報理論的評価を行う．

4.2.1 モデル化

最初に，いくつかの記号を定義する．victim がトレースバックに用いる攻撃パケットの集合を $\mathcal{N}$ とし，そのパケット数を N_p とする ($N_p = |\mathcal{N}|$)．さらに，ルータ R_i を通過する攻撃パケットの割合を d_i とおく．通常 d_i の値はルータによって異なるが，単純のため，同一の値 d をとるものとする（$d_i = d$）．また，Bloom filter で用いられるハッシュ関数の数を k とおく．すると，Bloom filter の false positive の確率 f は，$f = 2^{-k}$ と評価することができる[7]．また以下では，二項分布を $\mathrm{Binom}(N, p)$ と表す．ここで，N は試行の総数であり，p は各試行の成功確率を表すものとする．

次に，評価の際に利用する確率変数を以下のように定義する．

- X_{t_i} : R_i にサンプリングされた攻撃パケットの数
- X_{f_i} : $\mathcal{N}$ に属する各パケットを R_i の Bloom filter にかけたときに発生する false positive の数．X_{f_i} は，二項分布 $\mathrm{Binom}(N_p - X_{t_i}, f)$ に従う．
- Y_{t_i} : R_{i-1} へトレースバックする際に使用される攻撃パケットの数．すなわち，R_i と R_{i-1} において，共通してサンプリングされたパケットの数である．
- Y_{f_i} : $X_{t_i} + X_{f_i}$ を R_{i-1} の Bloom filter に通したときに発生する false positive の数．Y_{f_i} は，$\mathrm{Binom}(X_{t_i} + X_{f_i} - Y_{t-1}, f)$ に従う．

Vol.2011-DPS-146 No.7
Vol.2011-CSEC-52 No.7
2011/3/10

以上のパラメータを用いて，さらに以下の確率変数を考える．

- $X_i = X_{t_i} + X_{f_i} = N_p$：$R_{i-1}$ へのトレースバックに使用するパケットの数
  理想的にはトレースバックには R_i にサンプリングされた攻撃パケットのみ使えればよいのだが，Bloom filter によって false positive の可能性が生じ，それらは区別できない．そこで，これらの数を加えたものが，実際にトレースバックに使用されるパケットの数となる．
- $Y_i = Y_{t_i} + Y_{f_i}$：R_{i-1} の Bloom filter に存在する，$\mathcal{N}$ のパケットの数．

以上より，R_i においてトレースバックに使用される攻撃パケットのうち，少なくとも一つのパケットが R_{i-1} によってサンプリングされることを意味する確率変数 Z_i は，以下のように定義できる．

$$Z_i = \begin{cases} 1 & \text{if } X_{t_{i-1}} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

4.2.2 攻撃パケット

ルータにおけるサンプリングは，パケットごとに独立に行われる．よって，すべてのパケットにおいて X_{p_i}，X_{c_i} の確率分布は等しくなる．

X_{t_i} は，二項分布 $\mathrm{Binom}(N_p d_i, \Pr(X_{p_i} = 1))$ に従うと考えることができるので，

$$\Pr(X_{t_i} = j) = \binom{N_p d_i}{j} \Pr(X_{p_i} = 1)^j \times (1 - \Pr(X_{p_i} = 1))^{N_p d_i - j} \tag{5}$$

である．そこで，X_{t_i} の確率分布は，式 (3), (5) によって求められる．

victim は攻撃パケットの集合 $\mathcal{N}$ を用いて R_i から R_{i-1} へトレースバックを行う．X_{f_i} は，R_i の Bloom Filter へ $\mathcal{N}$ を通したときの false positive の数を表すので，

$$\Pr(X_{f_i} = \ell) = \sum_{j=0}^{N_p d_i} \Pr(X_{t_i} = j) \binom{N_p - j}{\ell} f^\ell (1-f)^{N_p - j - \ell} \tag{6}$$

これより，X_{f_i} の確率分布は，式 (5), (6) によって求められる．

さらに，$X_i = X_{t_i} + X_{f_i}$ であるから，式 (5), (6) を用いて，X_i の確率分布は次のように与えられる．

$$\Pr(X_i = j) = \sum_{\ell=0}^{\min(j, N_p d_i)} \Pr(X_{t_i} = \ell) \cdot \Pr(X_{f_i} = j - \ell) \tag{7}$$

次に，R_i と R_{i-1} のログに共通するパケットの数である Y_{t_i} について考察する．R_{i-1} が受け取ったパケットのカウンタの値が c であるとき，そのパケットを R_{i-1} と R_i が順にサンプリングする確率は，以下のように表せる：

$$\Pr(X_{p_i} = 1, X_{p_{i-1}} = 1 \mid X_{c_{i-1}} = c) = \frac{(c+1)^\alpha + \beta}{M} \cdot \frac{c^\alpha + \beta}{M}$$

これから，

$$\Pr(X_{p_i} = 1, X_{p_{i-1}} = 1) = \sum_{c=0}^{i-2} \frac{(c+1)^\alpha + \beta}{M} \cdot \frac{c^\alpha + \beta}{M} \cdot \Pr(X_{c_{i-1}} = c) \tag{8}$$

とできるので，式 (3), (8) より

$$\Pr(X_{p_i} = 1 \mid X_{p_{i-1}} = 1) \tag{9}$$

を計算できる．そして，6) の解析と同様に，Y_{t_i} は二項分布 $\mathrm{Binom}(X_{t_{i-1}}, \Pr(X_{p_i} = 1 \mid X_{p_{i-1}} = 1))$ に従っているとすることができる．

4.2.3 Z_i の条件付きエントロピーの導出

4.2.2 節で議論した確率変数を用いて，条件付きエントロピーの定義より，$H(Z_i \mid X_i, Y_i)$ は以下のように計算できる．

$$\begin{aligned} &H(Z_i \mid X_i, Y_i) = \\ &-\sum_{j=0}\sum_{m=0} \Pr(X_i = j, Y_i = m, Z_i = 1) \times \log_2 \frac{\Pr(X_i = j, Y_i = m, Z_i = 1)}{\Pr(X_i = j, Y_i = m)} \\ &-\sum_{j=0}\sum_{m=0} \Pr(X_i = j, Y_i = m, Z_i = 0) \times \log_2 \frac{\Pr(X_i = j, Y_i = m, Z_i = 0)}{\Pr(X_i = j, Y_i = m)} \end{aligned} \tag{10}$$

$H(Z_i \mid X_i, Y_i)$ が低い値ほど，トレースバックの成功確率が高くなる．そこで，式 (10) を求めることが，提案トレースバック方式における情報理論的評価の最終的な目標である．そのためにまず，

$$\Pr(X_i = j, Y_i = m, Z_i = a) = \Pr(X_i = j, Y_i = m \mid Z_i = a) \Pr(Z_i = a) \tag{11}$$

であることに注意する．ここで，Z_i の取り得る値は，$Z_i = \{0, 1\}$ であるが，6) と同様に，

$$\Pr(Z_i = 0) = \Pr(Z_i = 1) = 1/2 \tag{12}$$

と仮定する．

以下では，$Z_i = 1$ のときと $Z_i = 0$ のときに場合を分けて，$H(Z_i \mid X_i, Y_i)$ を計算していく．

4.2.3.1 $Z_i = 1$ のとき

$Z_i = 1$ のとき，式 (11) の右辺第一項はさらに，

$$\Pr(X_i = j, Y_i = m \mid Z_i = 1) = \Pr(X_i = j \mid Z_i = 1) \cdot \Pr(Y_i = m \mid X_i = j, Z_i = 1) \tag{13}$$

となる．ここで式 (13) 右辺第二項は，$Y_i = Y_{t_i} + Y_{f_i}$ に注意すれば，

$$\begin{aligned}&\Pr(Y_{t_i}+Y_{f_i}=m \mid X_i=j, Z_i=1)\\ &=\sum_{r=0}^{\min(m,N_pd_i)} \Pr(Y_{t_i}=r \mid X_i=j, Z_i=1)\\ &\quad\times \Pr(Y_{f_i}=m-r \mid X_i=j, Y_{t_i}=r, Z_i=1)\end{aligned} \tag{14}$$

特に，式 (14) において，

$$\Pr(Y_{f_i}=m-r \mid X_i=j, Y_{t_i}=r, Z_i=1)=\binom{j-r}{m-r}f^{m-r}(1-f)^{j-m} \tag{15}$$

であることに注意せよ．

さらに，式 (14) を求めるため，確率

$$\Pr(Y_{t_i}=r \mid X_i=j, Z_i=1) \tag{16}$$

について考える．今，確率変数 X_i $(=X_{t_i}+X_{f_i})$ と Y_{t_i} について，次の条件を満たす確率変数 W_i を考える．

$$X_{t_i}=Y_{t_i}+W_i \tag{17}$$

Y_{t_i} は，R_{i-1}，R_i の両方にサンプリングされたパケットの数を表すので，W_i は，R_i でサンプリングされたが，R_{i-1} でサンプルされなかった攻撃パケットの数に他ならない．そこで W_i は以下のように表せる：

$$\Pr(X_{p_i}=1, X_{p_{i-1}}=0 \mid X_{c_{i-1}}=c)=\frac{c^\alpha+\beta}{M}\cdot\left(1-\frac{c^\alpha+\beta}{M}\right) \tag{18}$$

式 (2), (18) から，

$$\begin{aligned}&\Pr(X_{p_i}=1, X_{p_{i-1}}=0)=\\ &\sum_{c=0}^{i-2}\Pr(X_{p_i}=1, X_{p_{i-1}}=0 \mid X_{c_{i-1}}=c)\cdot\Pr(X_{c_{i-1}}=c)\end{aligned} \tag{19}$$

を計算できるので，式 (3), (19) により，

$$\Pr(X_{p_i}=1 \mid X_{p_{i-1}}=0) \tag{20}$$

が求まる．W_i は二項分布 $\mathrm{Binom}(N_pd_i-X_{t_{i-1}}, \Pr(X_{p_i}=1 \mid X_{p_{i-1}}=0))$ に従っていると考えることができるので，式 (20) により，W_i の確率分布は計算できる．

以上の議論に基づき，式 (16) を求める．まず，

$$\begin{aligned}&\Pr(Y_{t_i}=r \mid X_i=j, Z_i=1)\\ &=\sum_{\ell=0}^{j}\Pr(X_{f_i}=\ell)\times\Pr(Y_{t_i}=r \mid X_{t_i}=j-\ell, X_{f_i}=\ell, Z_i=1)\end{aligned}$$

よって，式 (5) および，Y_{t_i}, W_i のそれぞれの二項分布より，6) の解析と同様にすることで，

$$\begin{aligned}&\Pr(Y_{t_i}=r \mid X_{t_i}=j-\ell, X_{f_i}=\ell, Z_i=1)\\ &=\sum_{g=r}^{N_pd_i}\Pr(X_{t_{i-1}}=g)\\ &\quad\times\Pr(Y_{t_i}=r, W_i=j-\ell-r \mid X_{t_i}=j-\ell,\\ &\quad X_{f_i}=\ell, X_{t_{i-1}}=g, Z_i=1)\\ &=\sum_{g=r}^{N_pd_i}\binom{N_pd_i}{g}\Pr(X_{p_{i-1}}=1)^g\cdot(1-\Pr(X_{p_{i-1}}=1))^{N_pd_i-g}\\ &\quad\times\binom{g}{r}\Pr(X_{p_i}=1 \mid X_{p_{i-1}}=1)^r\cdot(1-\Pr(X_{p_i}=1 \mid X_{p_{i-1}}=1))^{g-r}\\ &\quad\times\binom{N_pd_i-g}{j-\ell-r}\Pr(X_{p_i}=1 \mid X_{p_{i-1}}=0)^{j-\ell-r}\cdot\\ &\quad(1-\Pr(X_{p_i}=1 \mid X_{p_{i-1}}=0))^{N_pd_i-g-j+\ell+r}\end{aligned} \tag{21}$$

となる．式 (21) は，式 (3), (9), (20) によって計算できる．

以上より式 (16) が求まったので，式 (15) と合わせて，式 (14) を計算できた．そこで，式 (7) を用いて，結局式 (13) を計算できたことになる．

以上より，式 (12) を利用することにより，式 (11), (13) から

$$\Pr(X_i=j, Y_i=m, Z_i=1) \tag{22}$$

を求めることができた．

4.2.3.2 $Z_i=0$ のとき

この場合は容易に

$$\Pr(X_i=j, Y_i=m \mid Z_i=0)=\Pr(X_i=j)\binom{j}{m}f^m(1-f)^{j-m} \tag{23}$$

と求められ，式 (7), (12), (23) より，

$$\Pr(X_i=j, Y_i=m, Z_i=0) \tag{24}$$

を計算できる．

Vol.2011-DPS-146 No.7
Vol.2011-CSEC-52 No.7
2011/3/10

4.2.3.3 $H(Z_i \mid X_i, Y_i)$ の計算

$\Pr(X_i = j, Y_i = m)$ は，以下のように計算できる．

$$\Pr(X_i = j, Y_i = m) = \Pr(X_{t_{i-1}} = 0) \Pr(X_i = j, Y_i = m \mid Z_i = 0) + \Pr(X_{t_{i-1}} > 0) \Pr(X_i = j, Y_i = m \mid Z_i = 1) \quad (25)$$

式 (5), (13), (23) より，式 (25) を求められる．

以上より，式 (22), (24), (25) を計算することができたので，式 (10) によって，$H(Z_i \mid X_i, Y_i)$ を求めることができた．

4.3 理論的評価の数値計算

この節では，提案方式の挙動を理解するために，今まで行ってきた理論的評価の数値計算を行う．

4.3.1 α, β, M の決定

各ルータのサンプリング確率 p は，式 (1) で与えられる．そこで，$H(Z_i \mid X_i, Y_i)$ の値を計算したり，シミュレーションによる実験（5 節参照）を行ったりするためには，定数 α, β, M の値を定め，サンプリング確率を決定する必要がある．

最初に，パケット P のカウンタ P.counter における最大値と最小値について考察する．P.counter の値が最後に利用されるのは，経路上の最後のルータ R_n のサンプリング確率を決定する際である（R_n がサンプリングしてカウンタの値が 1 インクリメントされたとしても，その値は以後使われない）．R_n が参照する P.counter の値は，R_1 から R_{n-1} までのルータがサンプリングしたかどうかによって決まる．すなわち，カウンタの最大値としては，$n-1$ を考えればよい．

ここで，ネットワークを流れるパケットが経由する平均ホップ数は 16 と考えてよいが[8]，今後の解析のため，1 ホップ分だけ余裕をもって $n = 17$ と仮定する．カウンタの値は単調増加するので，サンプリング確率も単調増加する．そこで，R_{17} が受け取るパケットのカウンタが最大になったときのサンプリング確率 p が 1 以下となればよいので，p は，$p = (16^\alpha + \beta)/M \leq 1$ とできる．一方，P.counter の最小値は，カウンタの値が 0 のときであり，サンプリングされた回数が 0 回のときの確率である．この確率を初期確率 p_I とすると，$\beta/M = p_I$ とできる．

一般的に言って，サンプリング確率が高ければ高いほどトレースバックの性能は向上するが，同時にルータの負荷は高くなる．そこで，サンプリング確率が低く，かつトレースバックにおける効率が良い方式が望ましい．ここではトレースバックの評価のみ考えるので，公正を期すため，Li らの方式[6] における標準的なサンプリング確率 $p = 0.03$ よりも低い確率を，すなわち，例として $p_I = 0.01$ と設定する．以上の考察から，本論文の以降では，

- $\alpha = 2, \beta = 2.58586, M = 258.586$
- $\alpha = 3, \beta = 41.3737, M = 4137.37$
- $\alpha = 4, \beta = 661.98, M = 66198$

の 3 つの場合をここでは例として考える．

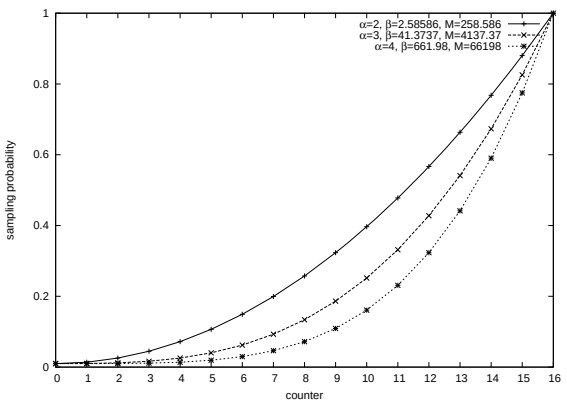

図 2 サンプリング確率
Fig. 2 Sampling Probability

4.3.2 サンプリング確率とカウンタ数

4.3.1 節で議論したパラメータの上で，式 (1) に従ってサンプリング確率を計算すると，**図 2** のようになる． 4.3.1 節で述べたように，カウンタの値が 0, 16 のとき，それぞれサンプリング確率は 0.01, 1 となる．図 2 から明らかなように，サンプリング確率は，$\alpha = 2, 3, 4$ の順で，急激に増加する．これは，$\alpha = 2, 3, 4$ の順で，カウンタの値が大きいパケットを優先してサンプリングするということを意味する．

さらに，ルータ R_{16} （ホップカウント 16）におけるカウンタ数の確率分布を求めると，**図 3** のようになる． 図 3 から分かるように，4.3.1 節で述べたようなパラメータ設定では，一般にサンプリング確率が非常に低く抑えられるために，ほとんどのパケットのカウンタ値が 0,1,2 となることがいえる．これによって，各ルータの負荷は非常に小さいことが分かる．さらに我々は，このような低い負荷の上で，非常に効率の良いトレースバックが行えることを，4.3.3 節, 5 節で示す．

Vol.2011-DPS-146 No.7
Vol.2011-CSEC-52 No.7
2011/3/10

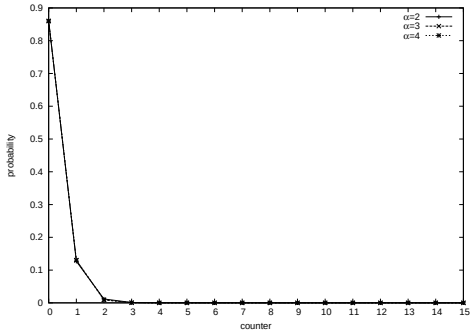

図 3 カウンタ数の確率分布（ホップカウント 16）
Fig. 3 Distribution of the number of counters (hop count = 16)

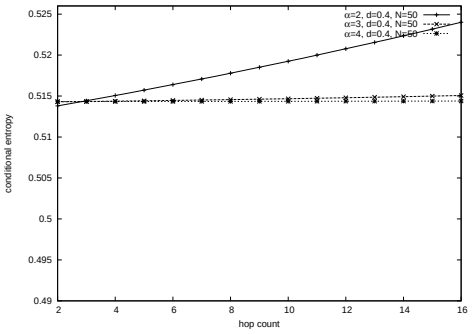

図 5 $H(Z_i \mid X_i, Y_i)$ の値 （$d = 0.4, N_p = 50$）
Fig. 5 $H(Z_i \mid X_i, Y_i)$ ($d = 0.4, N_p = 50$)

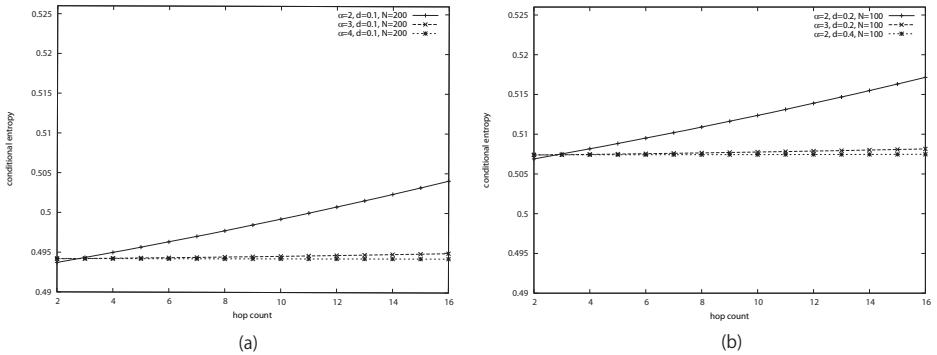

図 4 $H(Z_i \mid X_i, Y_i)$ の値： 図 (a) $d = 0.1, N_p = 200$，図 (b) $d = 0.2, N_p = 100$
Fig. 4 $H(Z_i \mid X_i, Y_i)$: Fig. (a) with $d = 0.1, N_p = 200$, Fig. (b) with $d = 0.2, N_p = 100$

4.3.3 $H(Z_i \mid X_i, Y_i)$ の値

4.3.1 節で述べたパラメータもとで，以下の条件で条件付きエントロピー $H(Z_i \mid X_i, Y_i)$（式 (10)）の値を計算した．

- $d = 0.1, N_p = 200$ （**図 4** (a)）
- $d = 0.2, N_p = 100$ （図 4 (b)）
- $d = 0.4, N_p = 50$ （**図 5**）

すなわち，サンプリングの対象となるパケット数を，$N_p d = 20$ として一定にしたものである．ここで，グラフの横軸はホップカウント i であり，縦軸に対応する $H(Z_i \mid X_i, Y_i)$ の値をプロットしている．また，評価を公正にするため，Bloom filter に用いられるハッシュ関数の数 k は，Li の方式で最適となる 12 とした．なお，4.2.3 節で述べたように，条件付きエントロピー $H(Z_i \mid X_i, Y_i)$ の値が小さければ小さいほど，トレースバックに必要となるパケット数も少なくてすむ．

ここで，Li の方式における理論的評価の数値的パラメータでは，たとえばサンプリング確率 $p = 0.03$ や, $N_p d = 50$ となっているが，一方本論文の評価では,（初期）サンプリング確率 $p = 0.01$, $N_p d = 20$ となっている．このような設定をはじめとして，評価に公正を期すため，本論文の評価では意図的に本方式にとって不利となるようなパラメータ設定を行ってきた．そこで，Li らの方式の評価に，本論文で使用したパラメータを使ったとすれば，彼らの論文[6] で示された値よりはるかに悪い結果になると予想できる．その議論のもとで，本論文のパラメータ設定と最も近い設定のものは，Li らの論文では，$d = 1/1000, N_p = 50,000$ のものであり，このとき条件付きエントロピーは，最適値で 0.63 程度となる．

一方で，我々の方式では，$H(Z_i \mid X_i, Y_i)$ の値は最悪でも 0.525 程度であり，最も良い結果は 0.495 程度となっている．以上から，我々の方式の優位性が結論できる．

5. シミュレーション実験

この節では，提案方式をシミュレーションによって評価する．

シミュレーションに用いたネットワークトポロジーは，Li らの研究[6] のものと同じもの

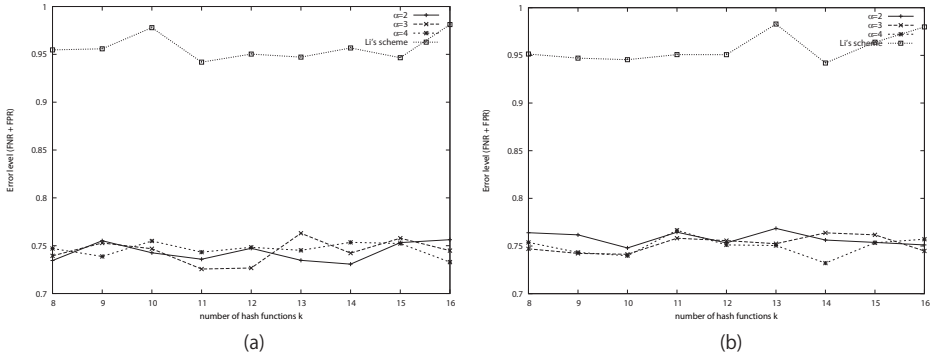

図 6 シミュレーション結果
Fig. 6 Results of Simulation

を用いた．すなわち，Skitter プロジェクト[1] による二つのトポロジ：

- a-root: a-root.skitter.caida.org による，2001 年 11 月 28 日のデータ
- e-root: e-root.skitter.caida.org による，2001 年 11 月 28 日のデータ

である．さらに，攻撃者の数 1000, 攻撃パケットの数 $N_p = 50,000$ とした．そのシミュレーション結果を，**図 6** (a) (a-root)，図 6 (b) (e-root) に示す．なお，ここでは Bloom filter におけるハッシュ関数の効果を調べるため，グラフの横軸はハッシュ関数の数 k としている．また，縦軸は，トレースバックの際の error level （すなわち，false positive と false negative の和 $^{\star 1}$ ）である． 図 6 の結果から明らかなように，いずれのパラメータ設定においても，提案方式は，Li らの方式よりも優れていることが分かる．また，このシミュレーションにおけるパラメータ設定では，ハッシュ関数の数による顕著な差は見られなかった．これは，サンプリング確率や，$N_p d$ の値が比較的小さなことから，Bloom filter における false positive の影響がほとんど出なかったためであると考えられる．

6. 結　　論

本論文では，効率のよいトレースバックを実施するため，パケットサンプリングの相関を攻撃経路全体にわたって高めることができるような IP トレースバック方式を提案した．提案方式では，サンプリング回数をパケット上のカウンタに記録し，その値を利用してサンプリング確率を決定する．本論文では，提案方式の詳細な理論的解析および実験的解析を行い，その結果として，Li らの方式[6] などよりも効率が良いトレースバックが実現されていることを示した．

$^{\star 1}$ この場合の false negative とは，実際には攻撃経路にそのルータが存在するにもかかわらず，存在していないとみなすことである．Bloom filter における false positive/negative との違いに注意．

参 考 文 献

1) CAIDA: Skitter project, http://www.caida.org/tools/measurement/skitter/.
2) Gong, C. and Sarac, K.: A More Practical Approach for Single-Packet IP Traceback using Packet Logging and Marking, *IEEE Transactions on Parallel and Distributed Systems*, Vol.19, No.10, pp.1310–1324 (2008).
3) Gong, C. and Sarac, K.: Toward a Practical Packet Marking Approach for IP Traceback, *International Journal of Network Security*, Vol. 8, No. 3, pp. 271–281 (2009).
4) 門林雄基, 大江将史：IP トレースバック技術, 情報処理, Vol.42, No.12, pp.1175–1180 (2001).
5) 唐沢智之，双紙正和，宮地充子：サンプリング確率を変動させた IP トレースバック方式の考察，情報処理学会研究報告，2008-CSEC-40, pp.67–72 (2008).
6) Li, J., Sung, M., Xu, J. and Li, L.: Large-Scale IP Traceback in High-Speed Internet: Practical Techniques and Theoretical Foundation, *Proceedings of the IEEE Symposium on Security and Privacy*, pp.115–129 (2004).
7) Mitzenmacher, M. and Upfal, E.: *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*, Cambridge University Press (2005).
8) Muthuprasanna, M., Manimaran, G., Alicherry, M. and Kumar, V.: Coloring the Internet: IP Traceback, *Proceedings of the 12th International Conference on Parallel and Distributed Systems (ICPADS)*, pp.589–598 (2006).
9) Peng, T., Leckie, C. and Ramamohanarao, K.: Survey of Network-Based Defense Mechanisms Countering the DoS and DDoS problems, *ACM Computing Surveys*, Vol.39, No.1, p.3 (2007).
10) Snoren, A.C., Partridge, C., Sanchez, L.A., Jones, C.E., Tchakountio, F., Kent, S.T. and Strayer, W.T.: Hash-Based IP Traceback, *Proceedings of the ACM SIGCOMM* (2001).
11) Xiang, Y., Zhou, W. and Guo, M.: Flexible Deterministic Packet Marking: An IP Traceback System to Find the Real Source of Attacks, *IEEE Transactions on Parallel and Distributed Systems*, Vol.20, No.4, pp.567–580 (2009).