スケッチデータ構造を用いた低速DDoS攻撃検出の検討

藤　竜成 [a)]・井 康成 [b)]・山場　久昭 [c)]・油田　健太郎 [d)]・岡崎　直宣 [e)]

An investigation on Detection of Low Rate DDoS Attack Using Sketch Data Structures

Ryusei FUJI, Kosei II, Hisaaki YAMABA, Kentaro ABURADA, Naonobu OKAZAKI

Abstract

The sketch data structure is a technique for extracting characteristics of traffic at high speed by expressing a large amount of data in a small amount of area using a plurality of hash functions. In the DDoS attack detection method using the sketch data structure, the similarity between the sketch generated just before and the new sketch is calculated, and if the similarity is low, the new sketch is determined as an attack. The low-rate DDoS attack attacks a victimized server at the same packet rate as normal traffic. It is difficult to detect with DDoS attack detection system dependent on traffic rate. The purpose of this study is to improve the accuracy by using sketch data structure showing high effectiveness in detection of the DDoS attacks for detecting the Low-rate DDoS attacks. In the existing sketch method, only the source IP address was used as the hash key of the sketch. In the proposed method, we add packet size as a hash key to generate the sketch. In the proposed method, we aim to detect low-rate DDoS attacks using the two sketches generated by the source IP address and the packet size. In the experiment, we evaluated the effectiveness of the sketch data structure in the low-rate DDoS attacks.

Keywords: sketch data structure, low-rate DDoS attack, similarity

1. はじめに

インターネット上での通信を必要とするサービスがますます増加する現代社会において、DDoS(Distributed Denial-of-Service) 攻撃は大きな脅威である。DDoS 攻撃の検出方法の 1 つにアノマリ型がある。これは通常トラフィックの行動パターンや正常なデータをあらかじめ定義しておき、新たなトラフィックが流入した際にそれと比較し、逸脱した場合に異常と判別する方法である [1)]。この方法を用いる際、検知速度の向上のためにシステムに流入したデータを効率的に解析しなければならない。そのため、DDoS 攻撃の大量のトラフィックデータを効率的に処理できるデータ構造が必要とされている。スケッチデータ構造は複数のハッシュ関数を用いて高次元のデータを少次元に集約する。これは大量のデータを少量の領域で表現することでトラフィックの特徴を高速に抽出できる。既存研究で DDoS 攻撃の検知における有効性が確認されている [2)]。

DDoS 攻撃が大量のパケットレートで攻撃するのに対し、低速 DDoS 攻撃は通常トラフィックと同程度のパケットレートで攻撃する。通常トラフィックとの判別が難しく、また、トラフィックレートに依存した DDoS 攻撃の検出システムで検出することも難しい。主な被害は正当なユーザに対するサービスの質が長期的に低下することである。

DDoS 攻撃と低速 DDoS 攻撃をともに高速に検出できるシステムの実現が必要とされている。本研究の目的は、その前段階として DDoS 攻撃の検出における高い有効性が示されているスケッチデータ構造を低速 DDoS 攻撃の検出に用いて精度の向上を図ることである。実験において、DDoS 攻撃におけるスケッチデータ構造の有効性の確認と提案手法の評価を行った。結果、DDoS 攻撃において検出精度が高いこと、提案手法においては低速 DDoS 攻撃を検出できることを示した。

2. スケッチ技術

スケッチ技術とは大量のデータを確率的に要約する技術である [2)]。データにハッシュ関数を適用することでランダムなハッシュ値を求めて、スケッチと呼ばれる限られた枠組みの中に集約する。この技術の利点は 2 つある。1 つ目は高次元のデータストリームを少次元に集約しデータを表現するコストを抑えることで、処理効率を向上する点である。2 つ目はハッシュ関数のランダム化によって、通常のネットワークトラフィックにおけるハッシュテーブル内の値の分布は安定しているため、あらかじめ保存した通常トラフィックの分布と新たな流入トラフィックの分布を比較することで、ネットワークトラフィックの重大な変化を検出できる点である。

2.1 スケッチのデータ構造

スケッチのデータ構造は図 1 のように構成される。スケッチは H 個のハッシュ関数を持ち、各ハッシュ関数は K 個の

[a)] 工学専攻機械・情報系コース大学院生
[b)] 情報システム工学科学部生
[c)] 情報システム工学科助教
[d)] 情報システム工学科准教授
[e)] 情報システム工学科教授

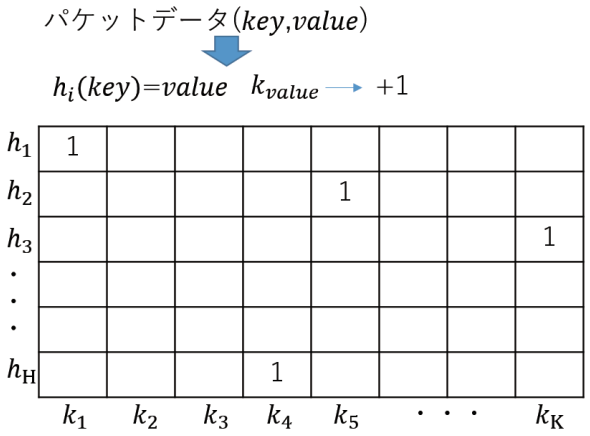

図 1. 1 パケット流入時のスケッチデータ構造

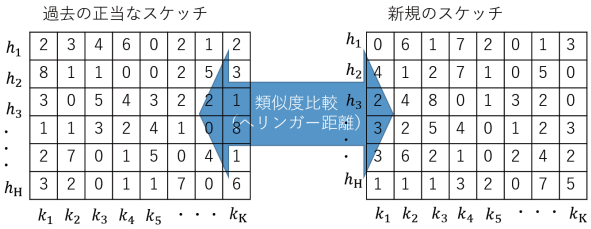

図 2. スケッチによる異常検出

サイズ列を持つ。したがって $K \times H$ のハッシュテーブルである。パケットデータは (*key*:ハッシュキー、*value*:ハッシュ値) の項目を持つ [3]。スケッチのサイズ列の初期値は 0 である。パケットデータが流入するたびにハッシュ関数 $h_i : i = 1 \cdots H$ が *key* をもとに *value* を求めて、h_i に対応したサイズ列 $k_j : j$=*value* の値に 1 を加算する。例として、図 1 では h_1 (*key*)=1、h_2(*key*)=5、h_3(*key*)=K、h_H(*key*)=4 は対応するサイズ列 k_1、k_5、k_K、k_4 の値に 1 を加算している。

2.2　スケッチを用いた DDoS 攻撃検出

トラフィックの流入時にスケッチを作成する。図 2 のように通常トラフィックで作成された過去の正当なスケッチと新たなトラフィックで作成された新規のスケッチの類似度をヘリンガー距離を用いて計算する [2]。類似度が低い場合、DDoS 攻撃トラフィックとみなす。ヘリンガー距離については 2.3 節で説明する。スケッチにおいて、ハッシュキーは主にパケットの送信元 IP アドレスを用いるため、スケッチには IP アドレスのばらつき具合の情報が現れる。また、トラフィックのパケット量が多いほどサイズ列に格納された値の平均は大きくなる。したがって、スケッチの IP アドレスの分散具合やパケット量の変化を検知することで DDoS 攻撃を検出できる。

2.3　ヘリンガー距離

ヘリンガー距離は 2 つの確率分布間の距離を測定する。2 つの確率分布 P=(p_1,p_2,p_3,・・・,p_n) と Q(q_1,q_2,q_3,・・・,q_n) があるとき式 (1) のように、確率分布間の類似度を求める。

$$HD(P,Q) = \frac{1}{2}\sum_{i=1}^{n}(\sqrt{p_i} - \sqrt{q_i})^2 \qquad (1)$$

ヘリンガー距離の値は 2 つの確率分布が完全に同一であれば 0 になり、完全に異なれば 1 になる [2]。ヘリンガー距離の値

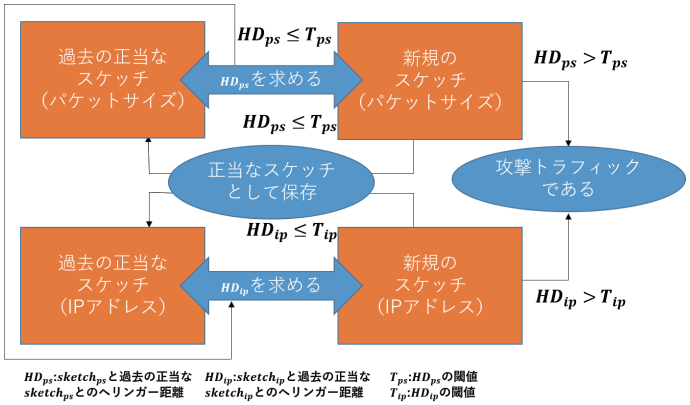

図 3. 提案手法の攻撃検知処理の流れ

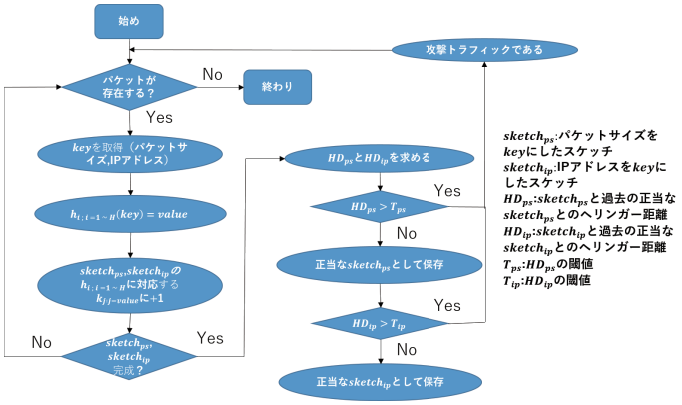

図 4. 提案手法の全体処理の流れ

が低いほど現在のトラフィックが過去のトラフィックと比べて変化が小さいことを意味し、高いほど異常が発生した可能性が高いことを意味する。

3.　提案手法

本研究ではスケッチ技術を用いて低速 DDoS 攻撃を検出する手法を検討した。既存のスケッチの手法では、スケッチ作成の際のハッシュキーに送信元 IP アドレスのみを用いる。まず低速 DDoS 攻撃の検出を既存のスケッチの手法で行ったが検出精度は低かった。検出結果の詳細は 4.3.1 に記す。低速 DDoS 攻撃はパケット量が通常トラフィックと同程度のため、トラフィック量の変化による検知は難しい。IP アドレスの分散具合の他に攻撃と通常トラフィックを判別する情報を追加する必要がある。そこで、本研究では既存手法にパケットサイズをハッシュキーにしたスケッチの作成を追加する。

3.1　攻撃と正当なトラフィックの判別

通常トラフィックは攻撃トラフィックに比べて、パケットサイズのばらつきが大きい [4]。これは通常トラフィックのパケットはクライアントの要求するデータサイズに従うのに対して、攻撃トラフィックは同じデータサイズのパケットを自動的に生成する傾向があるためである。この特徴を考慮し、パケットサイズをスケッチ作成の際のハッシュキーとして利用する。したがって、スケッチにはパケットサイズの分散具合の情報が現れる。本研究では既存手法にパケットサイズをハッ

シュキーにしたスケッチの作成を追加する。提案手法による低速 DDoS 攻撃検出の流れは図 3 の通りである。IP アドレス・パケットサイズを用いた 2 つのスケッチがそれぞれ、過去の正当なスケッチと比較される。ヘリンガー距離を用いて類似度を計算し、閾値以降の場合に低速 DDoS 攻撃と判別する。図 4 は提案手法の全体的な処理の流れを示す。

3.2 ハッシュ関数 LSH

本提案手法ではハッシュ関数として LSH(Locality-sensitive Hash) を用いる。LSH は NN 探索問題において厳密解ではなく近似解を求めることで高速化を図る。LSH はある 2 点において、距離が近いほどハッシュ値が同じになる確率が高くなるハッシュ関数である [5]。スケッチ技術に用いることで類似したトラフィックのスケッチ分布は類似しやすくなる。よって、スケッチを比較する際に正当なスケッチを誤って異常なスケッチとして判別することを防ぐことができると考えられる。

$$h(p) = \lfloor \frac{a \cdot p + b}{W} \rfloor \tag{2}$$

式 (2) において a は安定分布 (本手法はガウス分布を用いる) から選ばれた確率変数であり、b は閉区間 $[0,W](W > 0)$ からランダムに選ばれた実数である。式 (2) を利用して本提案手法のハッシュ関数を式 (3) とする。

$$h(key) = \frac{a(key) + b}{W} \bmod K \tag{3}$$

4. 評価実験

DDoS 攻撃検出におけるスケッチデータ構造の有効性の確認と、提案手法の検知精度の評価を行う。すべての実験においてパラメータは H=8、K=16384 とする。これは [3] で用いられたパラメータを引用している。また、スケッチの作成は 60 秒ごとにトラフィックを収集し行う。閾値設定は次の式 (4) で行う。

$$T = \mu + x\ \sigma \tag{4}$$

μ は正当なパケットのみを含むデータセットのキャプチャデータを学習させて計算したヘリンガー距離の平均値であり、σ は同様に求めた標準偏差である。x は任意の値である。ただし、通常トラフィックのみのデータセットに式 (4) を用いて実験し通常トラフィックを攻撃トラフィックとして検知しない値に設定する。

4.1 実験環境

実験環境を表 1 に示す。Windows10 が稼働する PC 上で、VMware Workstation 14 Player がゲスト OS として動作する仮想環境を構築した。このゲスト OS 上で実験を行った。

表 1. 実験環境

CPU	Intel(R)Core(TM)i7-7700CPU@3.60GHz
メモリ	16GB
OS	ホスト OS:Windows 10, ゲスト OS:VMware Workstation 14 Player (メモリ:8GB CPU2 コア)
開発環境	C++

4.2 評価方法

本研究の検知精度の評価には Precision(適合率)、Recall (再現率)、F 尺度を利用する。Precision は攻撃として検出したパケットのうち本当に攻撃パケットであった割合を表す。Recall は攻撃パケットを正しく攻撃パケットとして検出できた割合を表す。F 尺度は Precision と Recall を総合的に評価する指標である。いずれも値が大きいほど検出精度が高いと評価する。Precision、Recall、F 尺度を求める式を (5)、(6)、(7) に示す。tp は攻撃パケットを攻撃パケットとして検出できた数、fp は正当パケットを誤って攻撃パケットとして検出した数、fn は攻撃パケットを誤って正当パケットとして検出した数を表す。

$$Precision = \frac{tp}{tp + fp} \tag{5}$$

$$Recall = \frac{tp}{tp + fn} \tag{6}$$

$$F = \frac{2Precision * Recall}{Precision + Recall} \tag{7}$$

4.3 DDoS 攻撃の検出

DDoS 攻撃実験用の攻撃パケットとして、MIT Lincoln Laboratory が作成した DARPA2000 のデータセットを使用した。DARPA2000 が DDoS 攻撃を行うシナリオは次の 5 段階である [6]。ターゲット組織に IP スキャンを行い使用中の IP アドレスを調査する。使用中の IP に対して sadmind デーモンの有無を調査する。telnet により sadmind の脆弱性を利用したシステムの侵入を試みる。DDoS 攻撃ソフトの mstream を 3 台のホストにインストールしボット化する。攻撃者がボットに攻撃開始を指示する。

本研究では DARPA2000 においてファイアウォールの内部で観測された inside データを利用する。inside データの DDoS 攻撃部分のみを攻撃とみなす. 閾値設定には DARPA2000 と同程度のトラフィックトレースを持ち、通常パケットのみを含む DARPA1999 の火曜日のデータを利用した。

4.3.1 DDoS 攻撃における既存手法の検出精度

IP アドレスをキーにしたスケッチについて閾値を T_{ip}=μ_{ip}+2.5 σ_{ip} と定めた。結果を表 2 に示す。Precision と Recall

表 2. DDoS 攻撃における既存手法の検出精度

Precision	0.963783
Recall	0.996273
F 尺度	0.979759

がともに 0.96 を超えており、他の手法と同程度の十分な検知精度である [7]。スケッチデータ構造は DDoS 攻撃検出において有用であると言える。

4.4 低速 DDoS 攻撃の検出

低速 DDoS 攻撃の実験のための攻撃パケットとして、CAIDA2007[8] の最初の 25 分間のトラフィックトレースを使用した。CAIDA2007 には約 1 時間の DDoS 攻撃トラフィックトレースが含まれる。トラフィックトレースは被害者への攻撃トラフィックと被害者からの攻撃への応答のみを含み、単一のキーを用いて CryptoPAn プレフィックス保存を行うことで匿名化されている。

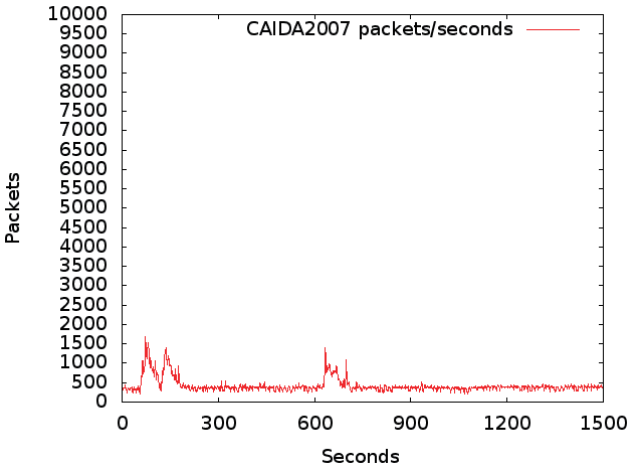

図 5. CAIDA の最初の 25 分 (1500 秒) 間のパケットレート

低速 DDoS 攻撃のパケットレートは 10000pps 未満である[4]。CAIDA2007 の最初の 25 分間のトラフィックトレースは図 5 に示すように、これを満たすため低速 DDoS 攻撃とみなされる。閾値設定には CAIDA2007 に含まれる通常トラフィック部分を結合し作成した通常パケットのみを含むデータセットを用いる。パケットレートは攻撃データセットと同程度にした。

4.4.1 低速 DDoS 攻撃の既存手法における検出（キーが IP のみ）

IP アドレスをキーにしたスケッチについて閾値を $T_{ip}=\mu_{ip}+2.5\ \sigma_{ip}$ と定めた。結果を表 3 に示す。Recall が 0.1 に

表 3. 低速 DDoS 攻撃における既存手法の検出精度

Precision	0.193919
Recall	0.0413923
F 尺度	0.0682224

も満たず、ほとんどの攻撃を検出できなかった。トラフィック量の変化が小さいため、IP アドレスの分散も大きな変化が現れなかったと考えられる。

4.4.2 低速 DDoS 攻撃の提案手法における検出（キーが IP とパケットサイズ）

パケットサイズと IP アドレスについての閾値 $T_{ip}=\mu_{ip}+x_{ip}\,\sigma_{ip}$、$T_{ps}=\mu_{ps}+x_{ps}\,\sigma_{ps}$ ごとの結果を表 4 に示す. Precision、Recall ともに既存手法より大幅に向上した。Recall

表 4. 低速 DDoS 攻撃における提案手法の検出精度

x_{ip}	x_{ps}	Precision	recall	F 尺度
2.5	2	0.71574	0.968947	0.823315
2.5	2.5	0.801055	0.912517	0.853161
3.5	2	0.757394	1	0.861952
3.5	2.5	0.849536	0.911683	0.879513

は 0.9 を超えており、閾値の設定次第で低速 DDoS 攻撃を攻撃として判別することはできる。しかし、Precision は 0.7 や 0.8 付近であり十分な精度とは言えない。これは低速 DDoS 攻撃トラフィックのなかに正当なパケットが含まれているため、攻撃トラフィックを検出する際に混入した正当パケットも同時に検出してしまうからだと考えられる。スケッチの中の攻撃パケットと正当なパケットを判別することが今後の課題である。

5. まとめ

本研究ではスケッチデータ構造を用いた低速 DDoS 攻撃検出手法を検討した。既存手法のハッシュキーに IP アドレスのみを用いる方法で試みたところ、検出精度は非常に低かった。トラフィックごとのパケット量の変化が小さいためである。そこで本研究では、通常トラフィックが攻撃トラフィックに比べてパケットサイズのばらつきが大きいという特徴から、パケットサイズをハッシュキーにしたスケッチの作成を追加したところ、検出精度は大幅に向上した。Recall は 0.91 を超え、低速 DDoS 攻撃を攻撃として検出できることが確認できたが、Precision が 0.8 程度であり十分な精度とは言えなかった。これは攻撃トラフィックのスケッチを検出する際にその中に含まれる正当なパケットも検出してしまうためである。今後の課題として、スケッチ内の攻撃と正当パケットの判別方法の検討が必要である。また、今回の実験は固定された閾値で行った。実際のネットワークトラフィックは流動的なため、トラフィックの状況に応じて変動する、閾値の設定方法を検討する必要がある。また、既存の低速 DDoS 攻撃検出方法との精度比較や処理速度等の性能比較を行うことも考えている。

参考文献

1) https://www.nttpc.co.jp/yougo/アノマリ型.html, (2019/01/28 閲覧)
2) J.Tang, Y.Cheng, Y.Hao, and W.Song: SIP Flooding Attack Detection with a Multi-Dimensional Sketch Design, IEEE Transactions on Dependable and Secure Computing, pp.1-14, 2014.
3) C.Wang, T.T.N.Miu, X.Luo, and J.Wang: SkyShield: A Sketch-Based Defense System Against Application Layer DDoS Attacks, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, Vol.13, pp.559-573, 2018.
4) K.Bhushan and B.B.Gupta: Hypothesis Test for Low-rate DDoS Attack Detection in Cloud Computing Environment, ScienceDirect Procedia Computer Science 132, pp.947-955, 2018.
5) 古賀 久志: ハッシュを用いた類似検索技術とその応用, IEICE Fundamentals Review, Vol.7, pp.256-268, 2014.
6) 小島 俊輔, 中嶋 卓雄, 末吉 敏則: エントロピーベースのマハラノビス距離による高速な異常検知方法, 情報処理学会論文誌, Vol.52, pp.656-668, 2011.
7) N.Hoque, H.Kashyap, and D.K.Bhattacharyya: Real-time DDoS attack detection using FPGA, Sicience Direct Computer Communications, pp.48-58, 2017.
8) http://www.caida.org/data/passive/ddos-20070804_dataset.xml, (Accessed on:2019/01/29)