機械学習を利用したDRDoS攻撃検知の提案とその性能実証

高, 宇軒
九州大学

馮, 尭鍇
九州大学

川本, 淳平
九州大学

櫻井, 幸一
九州大学

出版情報：暗号と情報セキュリティシンポジウム. 2016, pp.4B2-4-, 2016-01-19. 電子情報通信学会情報セキュリティ研究専門委員会(ISEC研)
バージョン：
権利関係：

SCIS 2016 2016 Symposium on Cryptography and Information Security
Kumamoto, Japan, Jan. 19- 22, 2016
The Institute of Electronics, Information and Communication Engineers

機械学習を利用したDRDoS攻撃検知の提案とその性能実証
A Machine Learning Based Approach for Detection of DRDoS Attacks and Its Performance Evaluation

高 宇軒* フォン ヤオカイ*† 川本 淳平*† 櫻井 幸一*†
Yuxuan Gao Yaokai Feng Junpei Kawamoto Kouichi Sakurai

あらまし DDoS攻撃の一種であるDRDoS攻撃は，問合せに対して増幅された応答を返す端末，即ちリフレクターを利用するDoS攻撃のことを指す．このDRDoS攻撃の影響によりDDoS攻撃のトラフィック量が増大している．典型的なDRDoS攻撃の検知手法は，プロトコル別に既知の攻撃を基に設計されたものが多く，新しい攻撃の検知は難しい．また，プロトコルに依存しない検知手法では，典型的な方法として，同じ送信元と宛先を持つパケットをフローという単位で考える．攻撃時にリフレクターから被害者までのフロー，つまり異常なフローの2つでは，定められた時間単位に到着したパケット数の間に強い線形関係があるという特徴を用いて検知を行っている．しかし，この方法では、攻撃時リフレクターから送信されるパケットは全て攻撃パケットであると仮定し，一部の攻撃を検知しにくいという欠点がある．本論文では，このような欠点を回避するために，DRDoS攻撃に共通な5種類の特徴を発見し，その特徴を基に用いてDRDoS攻撃のパターンを学習する方法を提案する．また，攻撃のシミュレーションから得たパケットデータを用いて提案手法の攻撃検知精度を評価する．

キーワード DRDoS，機械学習，攻撃検知

1 はじめに

DoS（Denial of Service）攻撃とは，一つあるいは複数の目標をサービス提供不能な状態に陥れる攻撃である[1]．DoS攻撃が観測された初期には，一つの端末からの攻撃が主であった．しかし次第に，攻撃の効果を向上させるために，複数の端末を乗っ取り目標に対して一斉攻撃が行われるようになった．そのような攻撃を DDoS（Distributed Denial of Service）攻撃と呼ぶ．その後，攻撃者は目標を直接攻撃するのではなく，パケット送信元のIPアドレスを攻撃目標のIPに偽装した要求パケットをネットワーク上のリフレクターに大量に送信する攻撃が観測されている．リフレクターとは，送信元からの問合せに対して送信元にそれなりの応答，即ち反射的な応答を返すように動作するものである．その結果，攻撃目標は大量の応答パケットを受け取りサービス提供不能な状況になってしまう．このような攻撃は DRDoS（Distributed Reflector Denial of Service）攻撃と呼ばれる．原理を考えると，DRDoS 攻撃のパケットに記述されたパケットの送信元は必ず偽装されているため，攻撃元の特定やリフレクターにおいてDRDoS攻撃を検知することは難しく，大きな課題になっている．

DRDoS 攻撃の対策は，主に攻撃の検知とパケットのフィルタリングの2つからなる．本論文では，攻撃の検知に着目し，既存の検知手法の改良案を提案する．

DRDoS 攻撃の一般的な検知手法は要求パケットと応答パケットのペアを統計的に解析することによる．大量なリクエストに対し対応するレスポンスが見つからない場合，即ちリクエストを送っていないにも関わらず，レスポンスを大量に受け取る場合, 攻撃されたと判断する．しかし，このような手法はプロトコル別に設計されたもので, 新しい種類のDRDoS攻撃の検知は難しい. 一方，プロトコルに依存しない検知手法 [2]は提案されているものの，攻撃時にリフレクターからのパケットを全て攻撃パケットであると仮定しているため，一部のトラフィックをしか取らない攻撃，即ち小規模な攻撃を検知しにくいという欠点がある．そして，検知手法としては，パケット数だけ数えるため，全部のトラフィックを占める通常な通信まで攻撃として検知されてしまう恐れがある．

* 九州大学，福岡市西区元岡744，Kyushu University，Motooka 744, Nishi-ku，Fukuoka
†（公財）九州先端科学技術研究所，福岡市早良区百道浜 2-1-22, Momochihama2-1-22, Sawara-ku, Fukuoka

このような欠点を解決するために，本論文ではプロトコルに依存しない検知手法の改良案を提案する．

プロトコルに依存せずDRDoSを検知するためには，DRDoS 攻撃におけるプロトコルに依存しない共通な特徴を探す必要がある．従来の手法では，リフレクターからのパケットは全て攻撃パケットであると仮定して，同じ送信元と宛先を持つパケットをフローという単位で検知を行う．受け取ったフローと既知のDRDoS攻撃フローを比較して，もし受け取ったフローが同じく DRDoS 攻撃フローである場合，即ち二つのフローがともに攻撃フローの際，たとえ違う種類の攻撃でも，フローの間に時間単位でのパケットの数の累積和が強い線形関係を持つという共通な特性を用いて検知を行う．

我々は，DRDoS 攻撃に関わるパケットには，パケットヘッダーに IP ヘッダーしか含まないパケットの数，端末宛て UDP パケットのサイズ，端末宛ての UDP パケットの数，端末からの UDP パケットの数，そして，ポートごとの端末宛てのパケットの数の最大値の5種類の値が他の通常のパケットと異なっていることを発見した．本論文では，それらを用いてDRDoS攻撃のパターンを学習し検知を行う方法を提案する．

本論文の構成を紹介する．第2節ではDRDoS攻撃に関する概念と種類について紹介する．第3節ではDRDoS攻撃に対する検知の関連研究を紹介する．第4節と第5節では我々の提案とそれに対する実験および検証について紹介する．最後の第6節はまとめである．

2　DRDoS 攻撃

2.1 拡大倍数

第1節に言ったように，DRDoS攻撃は反射攻撃のことを指す．そして，攻撃者にとって，トラフィックを反射することが唯一の目的ではなく，攻撃のトラフィックを拡大させることも目的である．それで，DRDOS攻撃の効果を測るために，拡大倍数という概念が導入された[3]. 拡大倍数はBAF（bandwidth amplification factor）とPAF（packet amplification factor）2種類に分け，それぞれの定義は式1と式2に示される．

$$\mathrm{BAF}=\frac{\text{リフレクターから目標までのペイロード}}{\text{攻撃者からリフレクターまでのペイロード}} \quad (1)$$

$$\mathrm{PAF}=\frac{\text{リフレクターから目標までのパケットの数}}{\text{攻撃者からリフレクターまでのパケットの数}} \quad (2)$$

BAF はペイロード的な拡大の効果，PAF はパケットの数的な効果を測る．攻撃の実際の影響を測るのはBAFである．BAFが大きいほど，攻撃のターゲットに対する影響が大きくなる．

2.2 DRDoS 攻撃の種類

利用されたプロトコル上に，DRDoS攻撃は2種類に分ける．一つはTCPを利用する攻撃であり，TCP-basedと呼ばれる．そして，TCPのハンドシェイクを完了した後，即ちセッションはすでに立たれたということで，サーバーなどを利用する反射的な攻撃を起こすことは不可能となる．そのため，SYN/SYN-ACK のペアしか利用できない．しかし，攻撃者からリフレクターあてのSYNパケットより，リフレクターから目標あてのSYN-ACKパケットのほうが大きいとは限らないため，BAFは常に1ぐらいしかない．そのため，攻撃者はTCP-based攻撃を基本的に使わない．もう一つはUDPを利用する攻撃であり，UDP-basedと呼ばれる．UDPで通信をする時，ハンドシェイクを行わないため，攻撃者に利用されることが可能なリクエスト/レスポンスペアの数が大幅に上がる．Handleyたちは14種類のプロトコルについて実験し，それぞれのBAFとPAFを計算した．[3]その結果は表1に示される．その結果のように，UDP-based攻撃には幾つかの高い BAF を達することができるプロトコルがある．そのため，UDP-based は DRDoS 攻撃の主流である．

表 1　各プロトコルの拡大倍数

Protocol	BAF	PAF	Scenario
SNMP v2	6.3	1.00	GetBulk request
NTP	556.9	3.84	Request client statistics
$\mathrm{DNS_{NS}}$	54.6	2.08	ANY lookup at author. NS
$\mathrm{DNS_{OR}}$	28.7	1.32	ANY lookup at open resolv.
NetBios	3.8	1.00	Name resoluteion
SSDP	30.8	9.92	SEARCH request
CharGen	358.8	1.00	Character generation request
QOTD	140.3	1.00	Quote request
BitTorrent	3.8	1.58	File search
Kad	16.3	1.00	Peer list exchange
Quake 3	63.9	1.01	Server info exchange
Steam	5.5	1.12	Server info exchange
ZAv2	36.0	1.02	Peer list and cmd exchange
Sality	37.3	1.00	URL list exchange
Gameover	45.4	5.39	Peer and proxy exchange

2.3DRDoS 攻撃の現状

図1には2013年と2014年の典型的なDRDoS攻撃を示している[4]．x軸は攻撃が発生する時間で，y軸は攻撃のトラフィックの量である．図から実際の DRDoS でも，UDP-based は主流であることを見える．

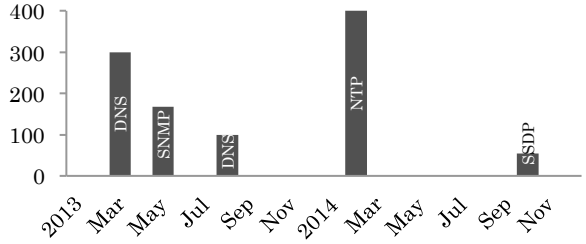

図 1　近年発生した DRDoS 攻撃

3　既存の DRDoS 検知手法

当初，DoS 攻撃の検知は，攻撃者と目標の中で，攻撃対象が属するネットワーク，攻撃対象が利用している ISP のネットワーク並びに上流のネットワーク及び攻撃元のネットワークの 4 地点で検知できると言われた[5]. しかし，DRDoS 攻撃に関しては検知が難しく，様々な手法が提案されている．その中で，攻撃が属するネットワークで検知する手法が多いので，それを更にわけられる．図 2 は DRDoS 攻撃の検知を行う地点別に分類したものである[3]. 我々は一つのルーターでプロトコルに依存しない検知手法を基に改良するために，検知する場所もルーターにする．

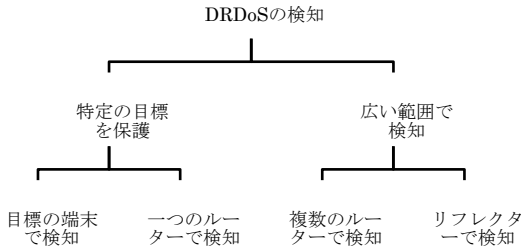

図 2　DRDoS の検知手法の関係図

一つのルーターで検知する手法として，当初，Tsunoda たちはリフレクターと目標の間のあるポイントでリクエストとレスポンスをマッチングし，攻撃を検知するという手法を提出した[6]．この手法では，リクエストを記録し，そしてリクエストにより可能なレスポンスを推測して記録した．その後，きたレスポンスを記録した推測のレスポンスとマッチングし，一致でないものを標記する．しかし，正しく推測できるために，一つのパケットに対して，ソースアドレス，デスティネーションアドレス，プロトコル，ヘッダーなどを記録しなければならない．その上，この手法では，既知の攻撃手段しか検知できない．

その次に，Wei たちはフロー単位で検知する手法を提出した[2]．フローは同じ送信元と宛先の全てのパケットのことを指す．この手法では，攻撃時にリフレクターからのパケットを全て攻撃パケットであると仮定し，二つの異常なフローの間に強い線形関連性があるということが計算できる．こういう特性を元に，来たフローを既知の異常フローと比較することで，線形関連性の強さにより攻撃を検知できる．そして，この手法はプロトコルを依存しないため，未知の攻撃でも検知できる．一方，欠点としては，攻撃された時リフレクターからのパケットは全て攻撃パケットを仮定したため，汎用性的には問題がある．そして，検知手法としては，パケット数だけ数えるため，全部のトラフィックを占める通常な通信まで攻撃として検知されてしまう恐れがある．

なお，一つのルーターでリクエストとレスポンス両方を監視することは条件がある．インタネット上の二つの端末の間の通信は必ず同じルート経由することに限らないため，このような方法では監視する端末の近いルーターでしか働けない．

4　提案

既存のプロトコルに依存しない検知手法の欠点を解決するために，我々は改良案を提案する．近年の典型的な DRDoS 攻撃はすべて UDP によるものである．そのため，本論文でも UDP 通信における DRDoS 攻撃を対象とする．

我々の提案手法では，検知器は特定の端末を保護するために，当該端末と外部ネットワークの境界であるルーターで動作する．提案手法は，通信を監視し，それらの中から特徴を学習する．そして，定められた時間単位に到着したパケットの統計的な特徴を基に攻撃の有無を判断する．

本提案手法では，次の 5 つの特徴量を用いて学習を行う．それぞれの選び理由を以下で述べる．

1. 端末あてのヘッダーの中で IP ヘッダーしか含まないパケットの数

UDP データグラムのバイト数の上限は IP パケットのバイト数の上限より大きいので，図 3 に示すように，UDP データグラムが大きいと複数の IP パケットに分割される．つまり，このようなパケットが大量検知されたら，大量なサイズの大きい UDP パケットが来るという意味である．そのため，この特徴を選ぶ．

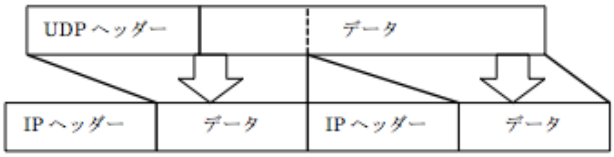

図 3　UDP 分割の模式図

2. 端末あての UDP パケットのサイズ

時間単位に端末で受け取ったパケットのサイズの和を特徴として選ぶ．UDP-based 攻撃を仮定するため，攻撃が発生する時，それは必ず遥かに大きくなる．

3. 端末宛ての UDP パケットの数

BAF の小さいが PAF の大きい攻撃に対して，サイズより数の変化のほうが激しい．そのため，サイズだけでこのような攻撃を検知しにくく，端末宛て UDP パケットの数を特徴にする．

4. 端末からの UDP パケットの数

以上の特徴だけで，通常の大量な UDP 通信と区別しにくいので，端末からの UDP のパケットの数も特徴にする. 特徴量を計算する時, これを直接使うのではなく，端末宛ての UDP パケットの数と端末からの UDP パケットの数の差を求め，それを特徴量とする．反射的な攻撃のため，攻撃されるとレスポンスがリクエストより多くなる．つまり，差も通常より大きくなる．

5. ポートごとに端末あてのパケットの最大値

一般的に，DRDoS 攻撃は一つのプロトコルを利用する．つまり，たとえ分散型攻撃でも，全部の攻撃パケットは同じポートからのものである．即ち，攻撃された時，あるポートからのパケットの数が大きくなる．そのため，この特性を利用し，時間単位にどのポートからのパケット数が一番大きいと確定し，そのパケット数を特徴にする．さらに，ポートを確定することで，攻撃の利用するプロトコルも確定できる．

これらの特徴量を用いて，図 4 に示すようなシステムを用いて検知を行う．提案のシステムはシステムの実装，特徴量の計算と検知の 3 つの部分から構成される．

その中，システムの実装の部分にはシステム初期化の流れを示す．保護したい端末の過去データを取った後，それぞれの特徴量を計算する．そして，その結果をサンプルとして保存する．

特徴量の計算の部分はパケットの統計と特徴量の抽出の流れを示す．パケットが来ると，そのパケットにより 5 つの特徴量を更新し，その後更新した特徴量を保存する．

検知の部分は時間単位に攻撃の有無を判断する部分の流れを示す．単位時間を計るために，タイマーを使う．タイマーをスタートした後，あらかじめ設定された時間に達すると，以下の一連の処理を行う．まずはシステム初期化する時に保存したサンプルと今まで保存した特徴量により分類する．具体的なのアルゴリズムについて，過去データにより違う可能性がある．それについて次の節で実験結果を元に検討を行う．分類した後，保存した特徴量をクリアーする．即ち，特徴量として保存したものは，この時間単位の統計結果だけである．その後，もし攻撃に分類されたら，警告した後タイマーをリセットする，そうでなければ直接タイマーをリセットする．

5　実験と検証

提案手法の検知率を検証するために，攻撃をシミュレートしてデータを収集した．そのデータを元に提案した 5 つの特徴により分類を行う．本節では，まずデータについて説明し，特徴ごとの検証と検知率の考察について述べる．

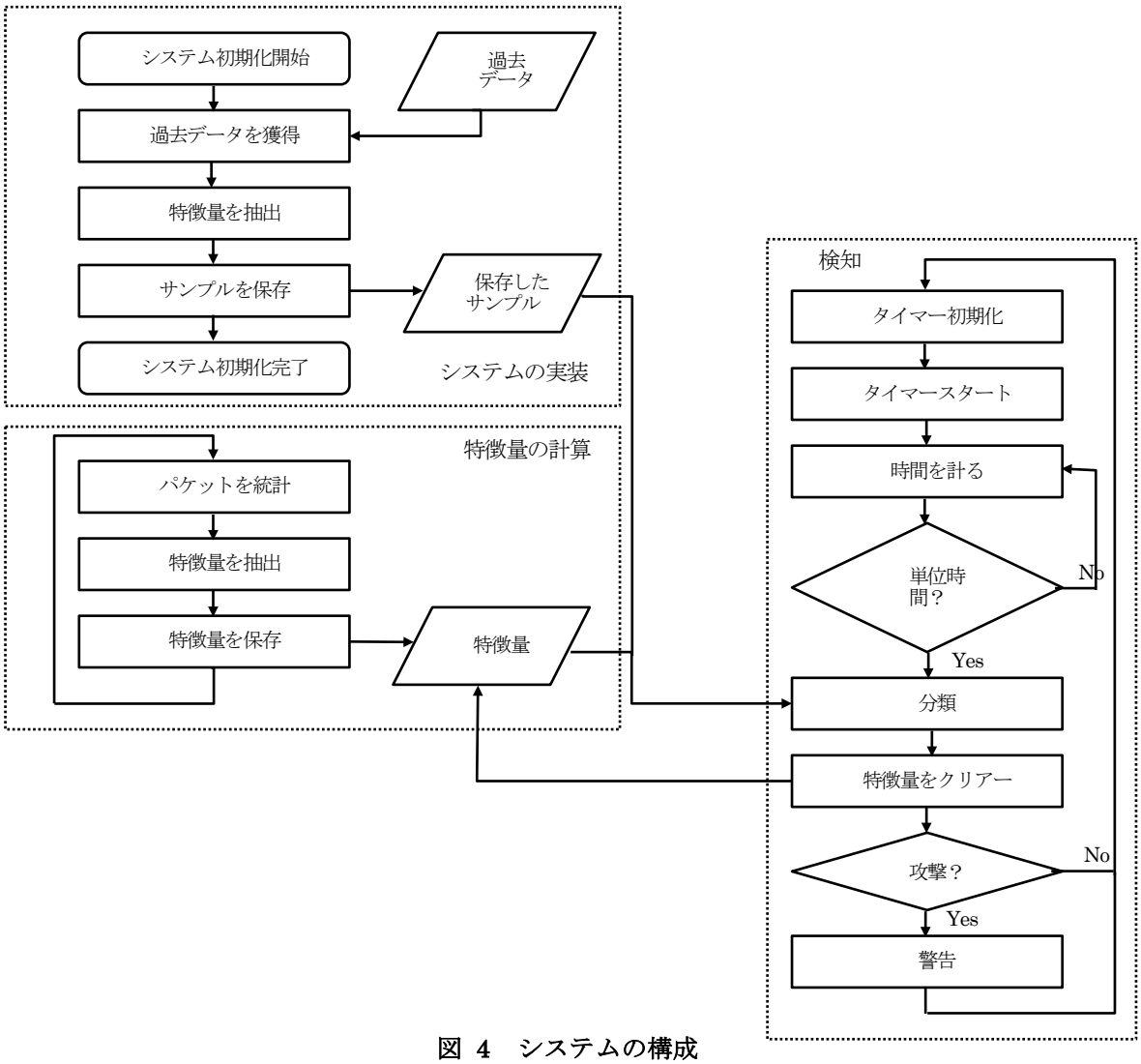

図 4　システムの構成

5.1 データの説明

攻撃のシミュレーションから 5 つのデータセットを収集した．攻撃のプロトコルを Chargen にし，リフレクターの台数は 2 台にする．PC-C と PC-D の Chargen サービスを起動し，リフレクターにする．そして攻撃者に模擬する PC-A は PC-C と PC-D の 19 番ポートにランダムの UDP パケットを送る．発送した UDP パケットの送信元の IP アドレスを PC-B の IP にする．それで，PC-B が攻撃の目標になり，リフレクターからの応答パケットは攻撃パケットになる．最後に PC-B のところでパケットを収集する．模式図は図 5 のようになる．

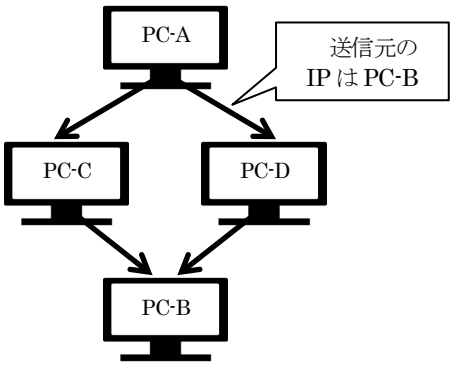

図 5　実験の模式図

各データセットの特性は表 2 に示される．収集時間と攻撃時間の単位は秒である．

表 2　各特徴セットの特性

データセット番号	収集時間	攻撃時間	攻撃種類	通常の UDP 通信	通常の通信の量	
1		81	26	Chargen	少ない	普通
2		162	23	Chargen	多い	多い
3		173	21	Chargen	少ない	少ない
4		252	47	Chargen	普通	多い
5		212	52	Chargen	多い	普通

5.2 特徴ごとの検証

図 6 にはデータセット 5 の中で各特徴量の変化を示している．x 軸は時間で，単位は秒であり，y 軸は上から端末あてのヘッダーの中で IP ヘッダーしか含まないパケットの数，端末宛ての UDP パケットのサイズの和，端末宛ての UDP パケットの数，端末宛ての UDP パケットの数とからの数の差，ポートごとに端末あてのパケットの最大値の 5 つの特徴量の値である．攻撃時間は 120 秒から 171 秒である．図の結果によると，各特徴が有効であることをわかられる．

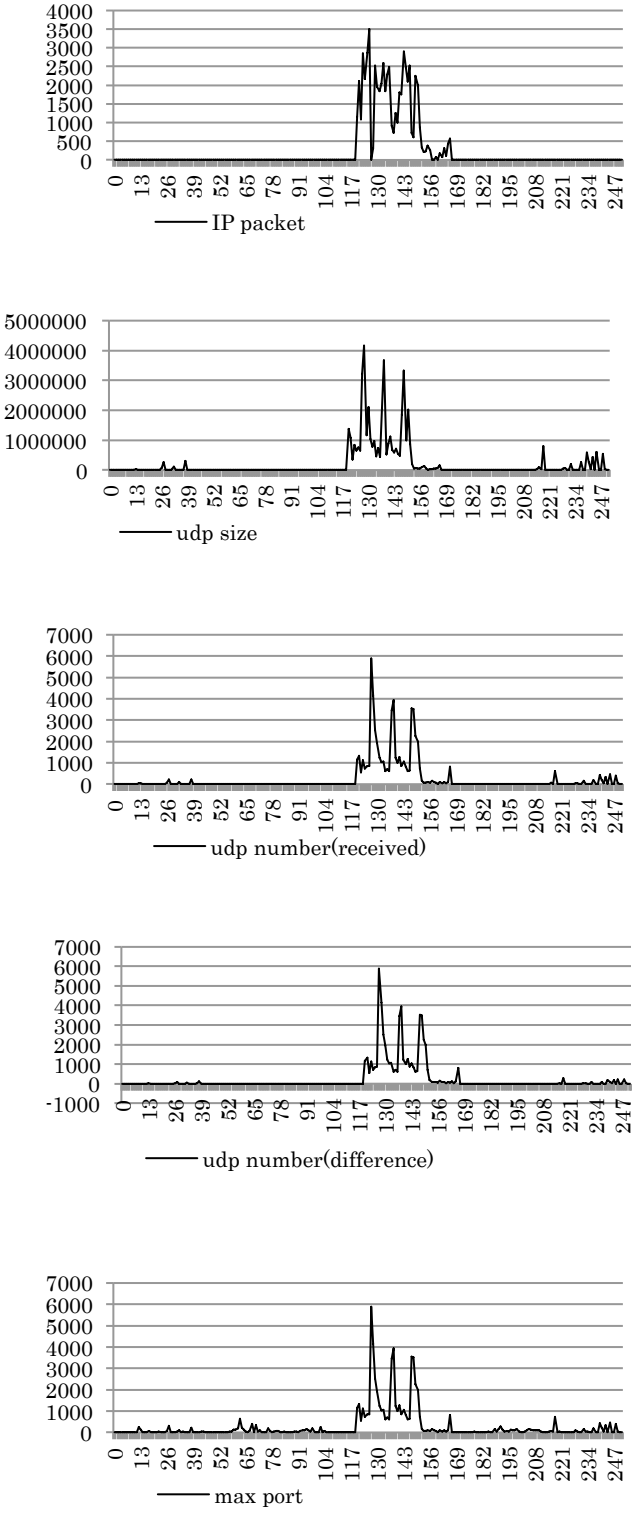

図 6　各特徴量の変化

5.3 検知率の考察

5 つのデータセットを元に，それぞれをサンプルとテストにして，実験を行う．結果について，総成功率，検知率と誤検知率の 3 つの評価基準によって評価されていた．それぞれの定義は以下のように示される．

$$\text{総成功率}=\frac{\text{正しく分類されたデータ数}}{\text{総データ数}}$$

$$\text{検知率}=\frac{\text{正しく分類された攻撃データ数}}{\text{総攻撃データ数}}$$

$$\text{誤検知率}=\frac{\text{正しく分類されていない通常データ数}}{\text{総通常データ数}}$$

3 つの評価基準が以上のように定義されている．総検知率は通常データと攻撃された時のデータともに分類の正確さについて考察でき，検知率は攻撃データの中で正しく検知できる割合について考察でき，誤検知率は通常であるが攻撃に分類されてしまうデータの割合について考察できる．

分類アルゴリズムについて，実験では KNN(k=1), KNN(k=3)，単純ベイズの 3 つのアルゴリズムを利用した.それぞれの結果と考察は5.3.1 から 5.3.3 で紹介する．

5.3.1 KNN(k=1)

k=1 と時の KNN 法，即ち最近傍アルゴリズムを用いて実験を行う．実験の結果は表 3 に示されている．

データセット 5 に含まれた攻撃シチュエーションが一番豊かなので，テストにする場合の総成功率と検知率は明らかに低くなった．一方，データセット 5 をサンプルにする場合，総成功率と検知率は高いレベルにおける．ほかのアルゴリズムを用いる時の実験の結果も同様である．

表 3　KNN(k=1)を用いる結果

総成功率(%)					
test / sample	1	2	3	4	5
1	100.0	99.4	99.4	96.0	95.8
2	96.2	100.0	98.2	98.4	95.2
3	92.6	97.5	100.0	93.7	84.4
4	95.1	96.9	97.7	100.0	95.3
5	97.5	98.1	97.7	94.4	100.0
検知率(%)					
test / sample	1	2	3	4	5
1	100.0	100.0	95.2	95.7	86.5
2	88.5	100.0	85.7	93.6	80.8
3	76.9	91.3	100.0	89.4	42.3
4	84.6	78.3	81.0	100.0	80.8
5	100.0	100.0	95.2	100.0	100.0
誤検知率(%)					
test / sample	1	2	3	4	5
1	0.0	0.7	0.0	3.9	1.3
2	0.0	0.0	0.0	0.5	0.0
3	0.0	1.4	0.0	5.4	1.9
4	0.0	0.0	0.0	0.0	0.0
5	3.6	2.2	2.0	6.8	0.0

5.3.2 KNN(k=3)

k=3 の時の KNN 法を用いて，実験を行う．その結果は表 4 に示されている．k=1 の時と比べ，各結果がほぼ変わらない．しかし，検知率は少し下げる．一方，誤検知率も少し下げる．

表 4　KNN(k=3)を用いる結果

総成功率(%)					
test / sample	1	2	3	4	5
1	100.0	100.0	99.4	96.0	94.8
2	95.1	100.0	98.3	98.0	93.4
3	91.3	97.5	98.8	93.7	82.5
4	92.6	97.5	97.7	99.2	93.4
5	97.5	98.1	97.7	94.4	99.5
検知率(%)					
test / sample	1	2	3	4	5
1	100.0	100.0	95.2	93.6	80.8
2	84.6	100.0	85.7	93.6	73.1
3	73.1	91.3	90.5	85.1	28.8
4	76.9	82.6	81.0	95.8	73.1
5	100.0	100.0	95.2	100.0	98.1
誤検知率(%)					
test / sample	1	2	3	4	5
1	0.0	0.0	0.0	3.4	0.6
2	0.0	0.0	0.0	1.0	0.0
3	0.0	1.4	0.0	4.4	0.0
4	0.0	0.0	0.0	0.0	0.0
5	3.6	2.2	2.0	6.8	0.0

5.3.3 単純ベイズ

単純ベイズアルゴリズムを用いる実験結果は表 5 に示されている．KNN と比べ，全体的な成功率と検知率の下げ誤検知率の上昇が見える．即ち，この過去データに対し，単純ベイズのほうが向いていないという結論を出られる．

表 5　単純ベイズを用いる結果

総成功率(%)					
test / sample	1	2	3	4	5
1	95.1	98.1	98.3	93.7	96.7
2	87.7	98.8	98.3	94.0	95.3
3	88.9	97.5	97.7	92.9	95.8
4	90.1	98.8	98.8	95.6	85.8
5	87.7	95.7	94.2	85.7	98.1
検知率(%)					
test / sample	1	2	3	4	5
1	92.3	100.0	100.0	95.7	92.3
2	88.5	100.0	100.0	97.9	82.7
3	88.5	95.7	100.0	95.7	88.5
4	76.9	91.3	90.5	87.2	42.3
5	96.2	95.7	100.0	87.2	98.1
誤検知率(%)					
test / sample	1	2	3	4	5
1	3.6	2.2	2.0	6.8	1.9
2	12.7	1.4	2.0	6.8	0.6
3	10.9	2.2	2.6	7.8	1.9
4	3.6	0.0	0.0	2.4	0.0
5	16.4	4.3	6.6	14.6	1.9

5.3.1 から 5.3.3 の実験結果から見ると，検知の正確率は基本的に高いレベルにおける．そのため，この特徴ベクトルは有効であることを言える．

6　まとめ

本文では，背景と既存研究を紹介した後，５つの特徴を利用した DRDoS 攻撃の検知手法を提案して，その検知性能を実証した．

今後の改良点としては，特徴ベクトルおよび実験のデータセットの充実が必要である．そして，システム的には，自動的に学習サンプルの拡張も考えるべきである．

謝辞

この研究の一部は，総務省による「国際連携によるサイバー攻撃の予知技術の研究開発」および科学研究費(基盤研究 No. 25330131)の支援を受けています．ここに記して謝意を表すものとします．

参考文献

[1] Handley, M., and Eric Rescorla. "RFC 4732: Internet Denial-of-Service Considerations." (2006).

[2] Wei, Wei, et al. "A rank correlation based detection against distributed reflection DoS attacks." Communications Letters, IEEE 17.1 (2013): 173-175.

[3] Rossow, Christian. "Amplification hell: Revisiting network protocols for DDoS abuse." Symposium on Network and Distributed System Security (NDSS). 2014.

[4] Ryba, Fabrice J., et al. "Amplification and DRDoS Attack Defense--A Survey and New Perspectives." arXiv preprint arXiv:1505.07892 (2015).

[5] Chang , Rocky KC. "Defending against flooding-based distributed denial-of-service attacks: a tutorial." Communications Magazine, IEEE 40.10 (2002): 42-51.

[6] Tsunoda, Hiroshi, et al. "Detecting DRDoS attacks by a simple response packet confirmation mechanism." Computer Communications 31.14 (2008): 3299-3306.