ダークネットトラフィックに基づく学習型 DDoS 攻撃監視システムの開発

古谷 暢章† 北園 淳† 小澤 誠一† 班 涛‡ 中里 純二‡ 島村 隼平††

† 神戸大学大学院工学研究科
657-8501 兵庫県神戸市灘区六甲台町 1-1
146t255t@stu.kobe-u.ac.jp
kitazono@eedept.kobe-u.ac.jp
ozawasei@kobe-u.ac.jp

‡ 情報通信研究機構
184-8795 東京都小金井市貫井北町 4-2-1
bantao@nict.go.jp
nakazato@nict.go.jp

†† 株式会社クルウィット
141-0031 東京都品川区西五反田 2-12-15
shimamura@clwit.co.jp

あらまし 本研究では，ダークネット観測網で得られたトラフィック情報から DDoS 攻撃によるバックスキャッタであるかの判定を行う学習型モニタリングシステムを提案する．パケットデータから送信元/送信先ポートや送信元/送信先 IP などに関連した 17 特徴を抽出し，サポートベクトルマシンによる判定を試みる．評価実験では，ルールによる判定が可能な 80/TCP 以外の TCP パケットをモニタリング対象とし，DDoS バックスキャッタの判定と追加学習を行う．また，次元圧縮手法である t 分布型確率的近傍埋め込み法を用いてホストの活動パターンの時間変化を視覚的に表し，機械学習の導入の有効性を示す．

Development of Adaptive Event-Monitoring System for DDoS Attacks

Nobuaki Furutani† Jun Kitazono† Seiichi Ozawa† Tao Ban‡
Junji Nakazato‡ Jumpei Shimamura††

†Guraduate School of Engineering，Kobe University.
1-1, Rokkodai-cho, Nada, Kobe, Hyogo 657-8501, JAPAN
146t255t@stu.kobe-u.ac.jp
kitazono@eedept.kobe-u.ac.jp
ozawasei@kobe-u.ac.jp

‡National Institute of Information and Communications Technology.
4-2-1, Nukui-Kitamachi, Koganei-shi, Tokyo 184-8795, JAPAN
bantao@nict.go.jp
nakazato@nict.go.jp

††clwit Inc.
2-12-12-303, Nishi-Gotannda, Shinagawa, Tokyo, JAPAN
shimamura@clwit.co.jp

Abstract In this work, we propose a learning type monitoring system that discriminate DDoS backscatter packets from those of other traffic observed by darknet monitoring (i.e., backscatter or non-backscatter). Upon the packets that are sent by a host towards the monitored darknet during a short time-window, we define 17 descriptive features, which are then input to an SVM classifier for classification. In the experiments, we use TCP packets sent from except for port 80/TCP as discriminate DDoS backscatter and incremental learning. Its adaptability is further illustrated by visualization of the host activities during the time expanse using dimension reduction techniques.

1　はじめに

様々なサービスがネットワークに依存している現代，これらのサービスの停止は一時的なものでさえ多額の損害が発生するようになった．その結果，金銭的損害を与えることを目的として，ボットに感染した多数のコンピュータを操り，攻撃対象にパケットを大量に送りつける DDoS 攻撃が行われるようになった．一つの例として，2000 年に発生した Yahoo，Amazon を相次ぎ狙った DDoS 攻撃では，サービス停止によって 12 億ドル以上の損害が発生したと推定されている [1].

このような攻撃者は自身の特定を困難とするために送信元の IP アドレスを詐称することが多い．特に DDoS 攻撃では，攻撃者が多数のコンピュータを操り，IP アドレスを詐称させてパケットを攻撃対象に送りつける．この時，攻撃を受けたサーバは，正常な通信と DDoS 攻撃による通信を容易には区別できず，また攻撃者の身元を特定することも困難である．そのため，詐称した IP アドレスを用いた DDoS 攻撃をいち早く発見し，サービス停止に追い込まれる前に対応することが課題となっている．

攻撃を受けたサーバは，通常通信と攻撃の通信が区別できないため，すべての通信に対して返答を行う．そのため，詐称された送信元に対しても返答を行い，コンピュータ等が存在しない未使用 IP アドレスのネットワーク（ダークネット [2]）に対しても返信パケットが送信される．この返信パケットはバックスキャッタと呼ばれる．ダークネットに届くバックスキャッタを観測することで DDoS 攻撃の特徴や傾向などを分析可能になる．そのため，早期に DDoS 攻撃を発見し，DDoS 攻撃によってサーバがサービス不能になる前にその対策を行うことが可能となる．ダークネットは，ネットワークサービスの提供などは一切行なっていないネットワークのため目的を持った通信は発生しない．そのため，ダークネット観測ではマルウェアのスキャンや DDoS 攻撃によるバックスキャッタなど不特定多数に対して行なう通信が観測可能である．また，無応答なパッシブモニタリングを行なうため容易に広範囲な観測を可能とし，グローバルな傾向をいち早く捉えることができる．特に，広範囲なダークネットモニタリングを行うことでいち早くバックスキャッタの検知，対策を行うことが期待される [3].

著者らは，ダークネットで観測された短時間のパケットデータから DDoS 攻撃によるバックスキャッタを判別する手法を提案した [4, 5]．判別に用いる特徴は各パラメータ（送信元／送信先ポート番号や送信先 IP アドレスなど）の情報は直接用いず，各パラメータの統計情報，ペイロードに関するものなど，通信プロトコルに依らないものを使用した．本稿では，全てのダークネットパケットに対してクラスラベルを与えることは困難であるため，専門的な知識による判断に基づいたラベルの付加が容易な TCP パケットを用いて評価実験を行い，提案手法の精度を示す．その結果から，様々なポートやプロトコルを用いて送信されたダークネットパケットに対して提案手法が適用可能であるかの考察を行う．また，DDoS 攻撃のパターンの変化や新たな不正通信形態の登場に伴い，観測されるダークネットトラフィックのパターンも多様化すると考えられる．この多様化に対応するため，追加学習を導入する．追加学習では，判別器による判別の信頼度が低かったデータについて，監視者が，目視による確認や総合的知見に基づき，バックスキャッタか否かの判定を行う．この判定を行ったデータを訓練データに新たに追加し，判別器を再度学習させる．この手法によって高い精度でバックスキャッタを判別できることを示す．

第 2 章で提案したバックスキャッタの判別手法について紹介する．第 3 章では性能評価実験と実験結果への評価を行う．第 4 章では実験結果に関する考察を行う．t 分布型確率的近傍埋め込み法 (t-distributed stochastic neighbor embedding:t-SNE[6]) を用いて，ダークネットパケット送信元ホストの活動パターンの時間変化を可視化し，機械学習と追加学習の有効性を示す．第 5 章では本稿の結論を述べる．

2　提案システム

2.1　システム構成

著者らは，バックスキャッタの判別手法として，特徴生成部でホストごとの短時間のパケットを特徴ベクトルに変換し，変換した特徴ベクトルを用いて判別部で学習と判別を行う手法を提案してきた [4, 5]．本稿では，前述の手法に加え，判別·評価を行ったデータを新たに訓練データに加える追加学習を導入する．追加学習を導入した DDoS バックスキャッタの判別フローを図 1 に示す．図 1 上部に示したように，まずダークネットで観測されたパケットデータをホストごとに分割する．つぎに，最初のパケットから設定した時間間隔内のパケットを抽出し，ホストの活動を表す特徴ベクトルを生成する．図 1 下部に示した DDoS バックスキャッタ判別では，訓練データを学習させた判別器に特徴生成で作成した入力データ与えて判別を行う．判別結果のうち，判別器によって出力された判別信頼度が高いものは正しく判別が行えているとして信頼する．そのうち，DDoS バックスキャッタであると判別されたものについては，当該のホストが DDoS 攻撃を受けているものとして監視者に報告する．一方，判別信頼度が低いものに関しては，判別器での判別は難しいと考え，監視者が目視判別を行う．その際，次回の判別で同様のデータが正しく判別できるように，評価後のデータを訓練データに加え，再度判別器を学習させる．判別に用いる判別器には，教師あり学習において未学習データに対して高い判別性能が得られるサポートベクトルマシン (SVM)[7] を用いる．

2.2　特徴生成

図 1 上部に示した特徴生成では，ダークネットで観測されたパケットデータを用いてホストごとに観測継続時間 t の特徴ベクトルを作成する．短い観測継続時間 t のデータを使用することで，サーバが処理不能に陥る前に DDoS 攻撃を検知できると考える．また，特徴ベクトルを作成する際にあまりにもパケット数が少ないと

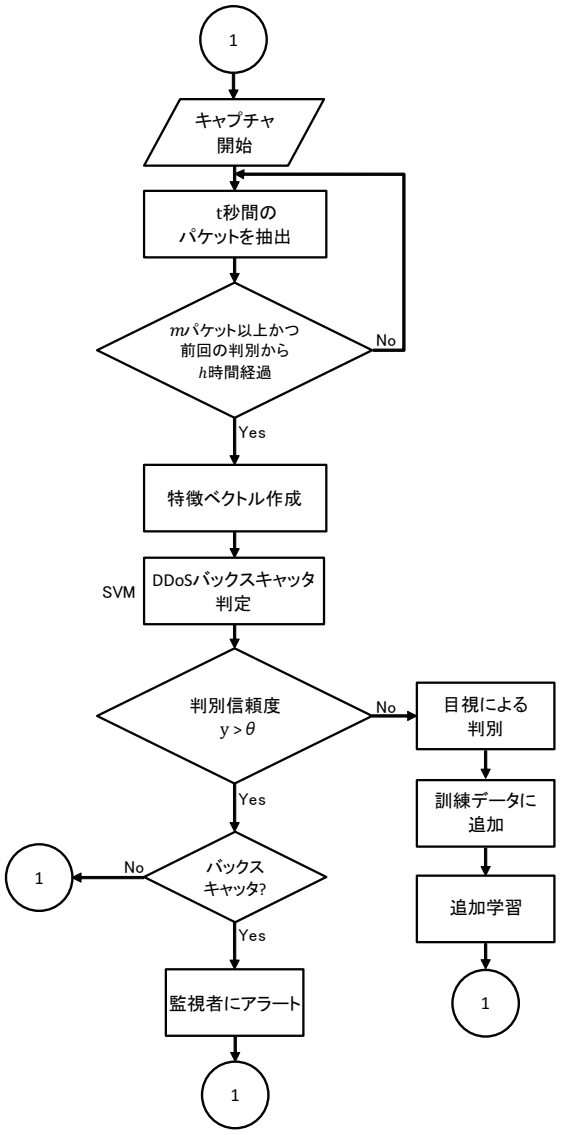


図 1: 提案システムの DDoS バックスキャッタ判別フロー

判別に十分な情報が得られないため，m パケット未満だった場合は，特徴ベクトルは作成せず m パケット取得するで観測を継続する．また，h 時間パケットが観測されなかった場合，新たな攻撃・活動が行われていると考え，新たにデータを抽出し，特徴ベクトルを作成する．

本システムでは，特定のポートやプロトコルを用いることが知られている既知の攻撃でなく，新たな攻撃手法に対応できることを目標とする．そのため，ポート番号やプロトコルを特徴に用いることは行なわない．そこで，特徴ベクトル作成の際に用いる特徴には，TCP 通信と UDP

通信に共通してみられ，ホストの活動を表す送信元/送信先ポートや送信先 IP の統計情報，ペイロードに関する情報を用いる．これらのポート番号や通信形態に依らない特徴量を用いることで，生成された特徴ベクトルではプロトコルやポートによる区別はなくなる．つぎに，各特徴の最大値が 1，最小値が 0 となるように値の正規化を行った．これによりすべての特徴を偏りなく用いて判定を行えるようになる．

これらの工程により，ホストあたりの短時間の振る舞いごとに 1 つの特徴ベクトルを生成する．

2.3 学習

2.3.1 初期学習とパラメータ

バックスキャッタ判定に用いる判別器には SVM を用いる．判別器の初期学習では，まず監視者の目視による解析によってバックスキャッタの有無のラベルの付加された初期データを SVM に学習させる．SVM のカーネル関数には RBF(Radial Basis Function) カーネルを用いる．SVM のペナルティパラメータ C とカーネルパラメータ γ の決定には，グリッドサーチ [9] と交差検定 [10] を用いる．グリッドサーチでは，(C, γ) を指数関数的に変化させ，すべての値の組合せに対して交差検定を行い，最も判別率の良いパラメータの値の組を求める．

また，バックスキャッタであるか否かの 2 クラス間でデータ数が大きく異なるため，コスト考慮型学習 [11] を適用する．コスト考慮型学習は，正例と負例とで異なる重みを与えることで誤判別した際のペナルティの量を変え，データの不均衡に応じて影響力を変える学習法である．SVM では，数の少ない方のクラスの重みにサンプル数の比をかけ，重みを調整し，データ数の少ない方のクラスの影響力を強める．

2.3.2 追加学習

ホストの活動パターンは時間経過とともに変化していき，初期学習のみでは新たな攻撃や活動パターンには対応することは難しく，正しく判別を行うことは困難であると考える．そこで，判別器で出力された判別信頼度が低いものに関しては初期データの学習では十分に判別が行えなかったと考え，判別を行ったデータを監視者に送信する．次に，監視者の目視によってホストの活動の再評価を行うことで，バックスキャッタか否かのラベルを付け直す．再評価後，新たにラベルの付加されたデータを訓練データに加え，判別器の再学習を行う．追加学習で行われるこれらの手順により，元々の訓練データには現れていなかったホストの活動パターンを新たに判別器に学習させることが可能となる．よって，新たな攻撃手法にも対応できるようになり，判別性能の向上が期待できると考える．これらの工程によって判別器の学習と判別，追加学習を行う．

3 性能評価実験

3.1 実験設定

実験に用いるデータセットには，NICT のダークネットで観測されたパケットデータ [12] のうち，2014 年 1 月 1 日から 2 月 28 日までの 8 週間のデータを用いる．提案手法は特定のポート番号などの既知のルールに基づいたラベル付けの行えないダークネットパケットに対しての適用を目的としており，80 番ポートから送信された TCP パケットや，53 番ポートから送信された UDP パケットは送信元のポート番号とパケットのヘッダ内部の情報を確認することでバックスキャッタか否かの判別が容易であるため，本実験には 80 番ポート以外から送信された TCP パケットデータを用いて判別・評価を行い，既知のルールに基づいたラベル付けの不可能なデータに対しての提案手法の判別性能を示す．また，SYN パケットは返信パケットとして存在しないため，SYN パケットのみを用いて作成された特徴ベクトルはあらかじめ削除した．

評価実験を行うために，まずはじめにダークネットパケットから作成した特徴ベクトルにバックスキャッタか否かのクラスラベルを付加する．クラスラベルは特徴ベクトルごとに可視化されたパケットの送信元のホストから送信先のダー

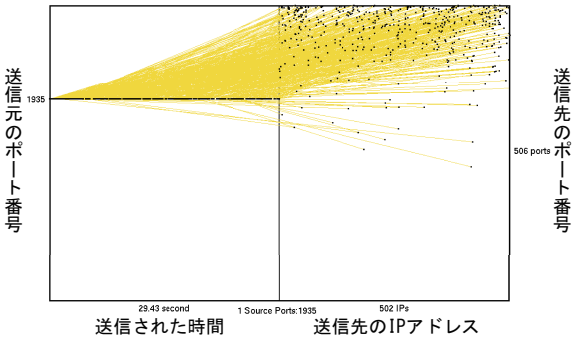


図 2: ホストの活動の一例

クネットの IP やポート等の情報を含む図に基づいて，専門的な知識による判断によって与えられる [4, 5]．ラベル付けに用いたホストの活動を表す図の一例を図 2 に示す．図の左半分の縦軸は送信元のポート番号，横軸はパケットの送信された時間を表す．右半分の縦軸は送信先のポート番号，横軸は送信先の IP アドレスを表す．この 2 点を結ぶことでパケットごとの送信された時間，送信元のポート番号，送信先のポート番号，送信先の IP アドレスを表す．線の色はパケットのプロトコル，TCP 制御フラグを表わす．黄色は SYN-ACK を表わす．

特徴ベクトル作成の際，パケットを抽出する時間間隔 $t = 30$ 秒，パケット数の閾値 $m = 20$，前回の判別からの経過時間 $h = 1$ 時間としてパケットを抽出する．これらの理由として，経験的に 30 秒間に 20 パケット以上あれば判別に十分な情報が得られる上に十分なデータ数が確保でき，1 時間パケットが観測されなければ，新たな攻撃・活動が行われている考えられるためである．特徴ベクトル作成に用いる特徴には以下の $n = 17$ 個を使用した．

- パケット総数
- パケット間の時間間隔の平均・分散
- 送信元ポートの総数
- 送信元ポートごとの送信されたパケット数の平均・分散
- プロトコルの種類数
- 送信先 IP アドレスの総数
- 送信先 IP アドレスごとの送信されたパケット数の平均・分散
- パケット間の送信先 IP アドレスの差分の平均・分散

表 1: 評価実験に用いた特徴ベクトル数

	バックスキャッタ	非バックスキャッタ	合計
第 1, 2 週	2,901	378	3,279
第 3 週	842	201	1,043
第 4 週	691	189	880
第 5 週	622	232	854
第 6 週	666	160	826
第 7 週	741	175	916
第 8 週	502	264	766

- 送信先ポートの総数
- 送信先ポートごとの送信されたパケット数の平均・分散
- ペイロードの平均・分散

評価実験に用いる 2014 年 1 月 1 日から 8 週間の作成した特徴ベクトル数を表 1 に示す．はじめの 2 週間の特徴ベクトルは判別器の初期学習用の初期データとして用いる．そして，残りの 6 週間の特徴ベクトルを評価と追加学習に用いる．初期学習後，第 3 週の 1,043 個の特徴ベクトルを判別し評価を行い，第 1 週から第 3 週までのすべてのデータを用いて判別器の再学習を行う．つぎに，第 4 週の 880 個の特徴ベクトルを判別し評価を行い再学習を行う．同様の手順を第 8 週まで行う．また，追加学習の有効性を検証するために，追加学習を行わずに，初期学習によって得られた判別器を用いて第 3 週以降のデータを判別した結果と比較する．

判別性能の評価は，バックスキャッタをポジティブなクラスとして，適合率，再現率，F 値の 3 つの指標と判別器の学習にかかった時間を用いて評価を行う．判別器に使用する L2-SVM には LibSVM[13] を利用し，グリッドサーチでは，$C = 2^{-5}, 2^{-3}, \cdots, 2^{15}$，$\gamma = 2^{-7}, 2^{-5}, \cdots, 2^{9}$ と変化させ，5-fold 交差検定を用いた．

3.2 実験結果

表 2(a) は，第 1 週と第 2 週のデータを初期データとして判別器に学習させ，追加学習を行わずに第 3 週から第 8 週までのデータを与え，判別を行った実験結果を示す．表 2(b) は，第 1 週と第 2 週のデータを初期データとして判別器に学習させ，第 3 週から第 8 週までのデータを

表 2: 実験結果

(a) 追加学習を行わなかった実験結果

データ	適合率	再現率	F 値	学習時間 (s)
第 3 週	0.966	1	0.982	120
第 4 週	0.969	0.998	0.983	−
第 5 週	0.987	1	0.992	−
第 6 週	0.963	1	0.981	−
第 7 週	0.983	1	0.991	−
第 8 週	0.967	0.998	0.982	−

(b) 追加学習を行った実験結果

データ	適合率	再現率	F 値	学習時間 (秒)
第 3 週	0.966	1	0.983	120
第 4 週	0.974	0.998	0.986	237
第 5 週	0.987	1	0.992	368
第 6 週	0.964	1	0.982	531
第 7 週	0.984	1	0.992	676
第 8 週	0.968	0.998	0.983	880

一週間ごとに判別・追加学習を行った実験結果を示す．

表 2(a) より，追加学習を行わず，初期学習のみでも高い性能でバックスキャッタの判別が行えることがわかる．特に再現率は 1 に近い値となっており，ほぼ見逃すことなくバックスキャッタを検知できていると言える．これにより，作成した 17 次元の特徴ベクトルと判別器によってバックスキャッタの判別を行う提案手法の有効性が示せた．

さらに，表 2(b) より，追加学習を行うことで，第 5 週を除くすべての週で追加学習を行わなかった実験結果より判別性能の向上が見られた．元々の性能が非常に高かったため，大きな性能向上は見られなかったが，この結果より，追加学習を行うことで，ホストの活動パターンの変化に対応し，新たな攻撃パターンに対しても判別が可能となっていると考えられる．

また，学習にかかった時間については，表 2(b) で示された時間については実用運用可能な範囲となっているが，現在の追加学習の手法は新たなデータが入ってきたときにすべてを再学習する手法であるため，いずれ処理不可能な時間になると考えられる．オンライン学習の手法の導入が今後の課題となっている．

4 考察

3.2 節より，追加学習を行わず，初期学習のみでも高い性能でバックスキャッタの判別が行えることを示せたが，追加学習を行うことで，判別性能の更なる向上が見られた．ここれにより，追加学習を行うことで，初期学習では判別できなかった活動パターンを学習することができるようになり，ホストの活動や攻撃パターンの多様化に対応可能となっていくと考えられる．これらの結果から，日を追うごとに，訓練データにはない新たな活動パターンが存在することが確認できた．そこで，高次元データを視覚可能な低次元に埋め込む手法である t-SNE[6] を用いて，特徴生成で作成した 17 次元の特徴ベクトルを 2 次元に圧縮して可視化を行い，時間の経過におけるホストの活動パターンの時間変化をを図 3 で視覚的に示した．

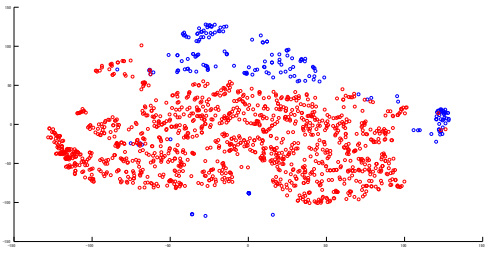
図 3(a) は 1 月 1 日から 1 月 7 日までの第 1 週のデータ，図 3(b) は 1 月 1 日から 2 月 28 日までの 8 週間のデータ，図 3(c) は 1 月 1 日から 6 月 31 日までの 6ヶ月間のデータを可視化した図である．また，赤色はバックスキャッタ，青色は非バックスキャッタ，緑色は専門家の目視による解析を行っておらずラベルの付加できていない未知のデータを表す．

図 3(a), 3(b) より，両クラスの境目ははっきりしており，現在の 17 個の特徴でクラス分離が正しく行えていることが確認できる．さらに，同じクラス内でも複数のクラスタに分かれており，それらのクラスタごとに何らかの攻撃や活動など固有の振る舞いが現れていると考えられる．また，これらの図から，評価実験を行ったデータのうち，バックスキャッタのデータの分布は第 1 週目から大きく変化していないことが確認できる．よって，表 2 において追加学習の有無にかかわらず再現率の値に変化がなかった理由については，実験期間においては初期学習のみの状態でも，バックスキャッタのデータに大きな変化がなく，正しく判別が行えたためである考えられる．一方，非バックスキャッタのクラスにおいては，図 3(a) で示された第 1 週目のデータの分布には現れていなかった場所にデータが分布し，新たなクラスタを形成していたり，

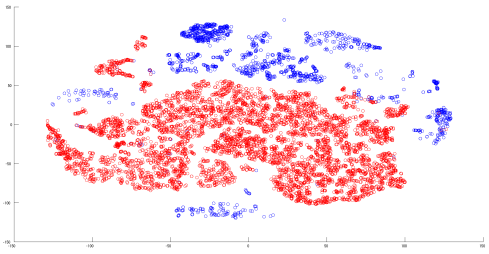
既存の分布から広がっていたりしていることが図 3(b) より確認できる．よって，表 2 の第 6 週目と第 8 週目で適合率の値が下がった理由については，第 6 週目と第 8 週目で今までの分布には現れていなかったデータが観測され，うまく判別が行えなかったためと考えられる．第 7 週目の判別結果では追加学習の有無によって大きく差がついていないため，これらのデータが多くは観測されなかったと考えられるが，追加学習によって判別性能が向上していることが確認できる．

図 3(c) より，解析の行えていない未知データの分布は，ラベルの付加できている既存の分布から少しずつ変化していることがわかる．大部分は既存の分布とほぼ同じであり，現在の学習でほぼ判別は行えると考えられる．しかしながら，3(c) の差し込み図に示したように，一部では，今までにない新しいクラスタを形成していたり，バックスキャッタと非バックスキャッタの間に分布しているなど，現在の学習だけでは正しく判別が行えない可能性のあるデータも存在していることが確認できる．これらのデータを正しく判別できるようになるためには，追加学習を行うことによってデータの時間変化対応していくことが必要である．

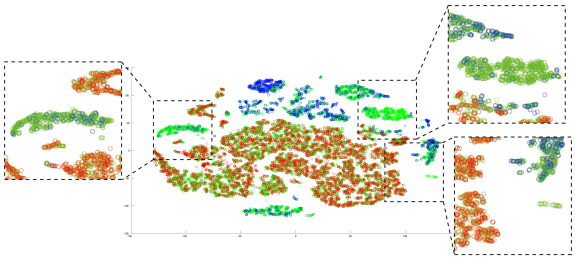
t-SNE を用いた可視化により，日が経過するごとにホストの活動パターンは変化していっており，その変化に対応するためには追加学習が必要であることが確認できた．



(a) 第 1 週のデータ分布



(b) 8 週間のデータ分布



(c) 6ヶ月間のデータ分布

図 3: t-SNE を用いたホストの活動の可視化

5　まとめ

本稿では，ダークネットで観測されたパケットを用いて DDoS バックスキャッタの判別を行うシステムを提案した．評価を行うデータには，ポート番号やヘッダ内部の情報に基づいてバックスキャッタか否かの判別が可能なデータ以外の TCP パケットを用いることで，様々なポートやプロトコルを用いて送信されたダークネットパケットに提案手法が適用できる可能性を示した．提案手法では，ダークネットで観測されたパケットデータをホストごとに分割し，最初のパケットから短時間のパケットを抽出して，特徴ベクトルを生成する．そして，生成した特徴ベクトルを SVM に学習させて，バックスキャッタか否かの判別を行う．さらに，新たな攻撃パターンに対応するため，評価を行ったデータを新たに訓練データに加え，判別器の再学習を行う追加学習を実装した．

評価実験では，NICT のダークネットで観測されたパケットデータ [12] を用い，0.98 を超える F 値が得られ，追加学習を行うことでさらなる判別性能の向上がみられた，この結果より，提案手法でダークネットパケットを用いて DDoS 攻撃の判別を行うことができ，さらに，追加学習を行うことで，今まで判別できなかったデータに対しても判別が行えるようになり，新たな攻撃パターンにも対応可能となっているといえる．

また，ホストの活動パターンの時間変化を視覚的に確認するため，t-SNE[6] を用いて評価実験に使用した特徴ベクトルの可視化を行った．可視化により，時間変化による攻撃や活動パターンの多様化が確認でき，その多様化に対応していくためには，追加学習が必要であると示すことができた．

今回の実験では，提案手法を TCP パケットのみに適用したが，UDP や他のパケットにも適用することですべてのダークネットパケットに対してバックスキャッタの判別を行う必要がある．今後は，計算時間の削減のため，新たなデータが入ってきたときにすべてを学習しなおす再学習でなく，逐次的に判別器を修正できるオンライン学習を実装することと，提案手法の実装化に向けて，リアルタイムの判別を行えることを目標とする．

参考文献

[1] P. A. R. Kumar, S. Selvakumar, “Distributed Denial-of-Service (DDoS) Threat in Collaborative Environment - A Survey on DDoS Attack Tools and Traceback Mechanisms,” *Proc.of Advance Computing Conference*, pp. 1275–1280, 2009.

[2] D. Moore, C. Shannon, G. Voelker, and S. Savage, “Network Telescopes: Techical Report,” *Technical report, Department of Computer Science and Engineering, University of California, San Diego*, 2004.

[3] K. Nakao, D. Inoue, M. Eto, K. Yoshioka, “Practical correlation analysis between scan and malware profiles against zero-day attacks based ondarknet monitoring,” *IEICE Trans. on Information and Systems*, vol.E92-D, no.5, pp. 787–798, 2009.

[4] N. Furutani, T. Ban, J. Nakazato, J. Shimamura, J. Kitazono, and S. Ozawa,“Detection of DDoS Backscatter Based on Traffic Features of Darknet TCP Packets, ” *2014 Ninth Asia Joint Conference on Information Security*, pp. 39–43, 2014

[5] 古谷暢章，班涛，中里純二，島村隼平，北園淳，小澤誠一，“ダークネットトラフィック観測による DDoS バックスキャッタ判定,” *Information and Communication System Security*, vol.114, no.340, pp. 49–53, 2014.

[6] L Van der Maaten, G Hinton , “Visualizing data using t-SNE,” *Journal of Machine Learning Research*, vol.9, pp. 2579–2605, 2008.

[7] V. N. Vapnik, *Statistical learning theory,* vol.1, New York, Wiley, 1998.

[8] 中里純二，島村隼平，衛藤将史，井上大介，中尾康二，“DoS 攻撃の分類に向けたバックスキャッタ分析,” *The 30th Symposium on Cryptography and Information Security*, 2013.

[9] C. W. Hsu, C. C. Chang, C. J. Lin, “A practical guide to support vector classification,” *Technical report, Department of Computer Science, National Taiwan University*, 2003.

[10] R. Kohavi, “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection,” *International Joint Conference on Artificial Intelligence*, vol.14, No. 2. 1995.

[11] R. Akbani, S. Kwek, and N. Japkowicz, “Applying support vector machinesto imbalanced data sets,” *Lecture Notes in Computer Science*, vol.3201, pp. 39–50, 2004.

[12] 神薗雅紀，“マルウェア対策のための研究用データセット (MWS Datasets 2014),” 2014

[13] C. C. Chang, C. J. Lin, “LIBSVM : A library for support vector machines,” *ACM Transactions on Intelligent Systems and Technology*, vol.2, 2011.