社団法人　情報処理学会　研究報告
IPSJ SIG Technical Report

2005－DPS－122（7）
2005－CSEC－28（7）
2005／3／22

TCP に対するポートスキャンの高速検知手法

小原　正芳 †　　堀　良彰 ‡　　櫻井 幸一 ‡

† 九州大学工学部電気情報工学科
812-8581 福岡市東区箱崎 6-10-1
kohara@itslab.csce.kyushu-u.ac.jp

‡ 九州大学大学院システム情報科学研究院
812-8581 福岡市東区箱崎 6-10-1
{hori, sakurai}@csce.kyushu-u.ac.jp

あらまし　近年，インターネット上のエンドホストに対してネットワークを介し無差別に行われる攻撃が増加している．攻撃者は脆弱なエンドホストを探すために TCP に対するポートスキャンを行うため，ポートスキャンは侵入の前兆とみなすことができる．それゆえ，攻撃者からのポートスキャンを早期に検知し必要な対策を行うことは，攻撃を事前に防ぐために重要である．

ポートスキャン検知のために，これまでいくつかのアルゴリズムが考案され，それらはネットワーク侵入検知システムに実装されている．しかしながら，既存のポートスキャンの検知アルゴリズムでは，早期検知よりも精度に重点がおかれているため，精度を損なわず早期検知が可能な新たな手法が求められている．

本稿では，ポートスキャンの特徴に基づく評価基準を用いることでポートスキャンを効率良く検知できる手法を提案し，その評価を行うことで提案手法の有効性を明らかにする．

Fast TCP Portscan Detection Method

Masayoshi KOHARA †　　Yoshiaki HORI ‡　　Kouichi SAKURAI ‡

†School of Information Science and
Electrical Engineering, Kyushu University
6-10-1 Hakozaki, Higashi-ku, Fukuoka,
812-8581 Japan
kohara@itslab.csce.kyushu-u.ac.jp

‡Faculty of Information Science and
Electrical Engineering, Kyushu University
6-10-1 Hakozaki, Higashi-ku, Fukuoka,
812-8581 Japan
{hori,sakurai}@csce.kyushu-u.ac.jp

Abstract　Recently, there are many indiscriminant attacks against end-hosts on the Internet. As attackers carry out TCP portscans to find vulnerable end-hosts, portscans can be considered as the sign of intrusion. Therefore prompt detection of the portscan is significant to prepare protection of end-hosts.

There have already been several work on detecting portscans and have been using their methods in some network intrusion detection systems. Most of them, however, are insufficient for prompt detection of scanners. Consequently, we requires promptness with accuracy.

In this research, I propose an efficient method by which I can detect portscans according to the multiple criteria. In summary, the proposed scheme can achieve prompt detection of portscans with sufficient accuracy.

1　はじめに

近年，インターネット上のエンドホストに対してネットワークを介し無差別に行われる攻撃が増加している．インターネットを介して感染するワームが行う攻撃がこれにあたる．攻撃者は脆弱なエンドホストを探すためにネットワークに対してポートスキャンを行う．したがって，ポートスキャンは侵入の前兆とみなすことができる．また，ポートスキャンの検知は，外部から自組織のネットワークへの攻撃を把握できるだ

けでなく，自組織のネットワークから外部に対する攻撃の把握にも有用であり，ポートスキャンの早期検知によって動的にフィルタリングを行うなど，攻撃に対する対策を事前に講じることができる．さらに，インターネット上でワーム等に感染したホストは攻撃者となり，ネットワークに接続されている他のエンドホストへ攻撃を行うことから，自組織のエンドホストが行うポートスキャンを早期に検知することはワームに感染したホストを検知することにもつながり，動的なフィルタリングなどによって，ワームの感染の拡大を防ぐことにつながる．本稿ではワームなどによって使用される TCP ポートスキャンに着目し，ポートスキャンの特徴に基づく評価基準を用いることでポートスキャンを効率良く検知できる手法を提案し，その評価を行った．評価の結果，早期検知と低誤検知率を達成できたのでで報告する．

本稿の構成は以下のようになっている．2 節で関連研究について説明する．3 節では提案手法の説明を行う．4 節では今回使用したデータセットに関する説明をする．5 節では提案方式の評価を行う．6 節ではまとめと今後の課題について述べる．

2 関連研究

2.1 特定のホストからの大量のパケットを検知する手法

ポートスキャンが行われると，特定のホストからの大量のパケットが観測される．この手法は特定のホストから一定数以上のパケットをポートスキャンとして検知する手法である．Snort [1] は特定のホストから t 秒間以内に p 以上のポートに対して TCP パケット，もしくは UDP パケットが観測されたときにポートスキャンとして検知する．しかし，この手法によるポートスキャンの検知では，TCP SYN パケットを大量に送出するホストはすべてスキャナーとなってしまう．したがって，見逃し率は小さいものの誤検知率が非常に大きい．

2.2 特定のホストからの大量のコネクションの失敗を検知する手法

スキャナーは対象ネットワークに対する情報を持っていない場合，存在しないホストや，サービスの行われていないポートに対してもコネクションリクエストを送出する．したがって大量のコネクションの失敗が検知されるため，これを利用してポートスキャンを検知することができる．これはポートスキャンの検知手法として Bro [2] で使用されている．しかし，アプリケーションによるポートスキャンを区別していないため，P2P アプリケーション等による良性のポートスキャンも検知してしまう．したがって，悪意あるポートスキャン検知には誤検知率が大きくなる．

2.3 TCP RST パケットをカウントする手法

参考文献 [3] で述べられている手法は TCP RST パケットを使ってワームを検知する手法である．TCP RST パケットはエンドホストが稼動中であるが，ポートが閉じている時に返されるパケットである．そのようなポートに対してコネクションリクエストを行うホストは，スキャナーである可能性が高い．しかし，IP アドレスに基づくフィルタリングを行っている場合等においては，終端ポートが閉じていても TCP RST が返ってくるわけではないし，また，ファイアウォールによっては許可されていないネットワークからのコネクションリクエストに対して TCP RST を返すように設定されているものも存在する．したがって，TCP RST だけを使って全てのポートスキャンを検知することは難しい．

2.4 シグネチャベースの検知手法

TCP SYN-FIN, Xmastree といったプロトコルの規格外のパケットを用いたポートスキャンがあるが，これらのスキャンを検知するにはシグネチャベースの検知手法が非常に有効である．Snort ではシグネチャによってこれらのパケットを検知する．しかし，この手法で検知できるのはプロトコル規定外のパケットを用いたポートスキャンだけである．TCP SYN スキャンのような TCP の通常の動作を利用したポートスキャンの検知はできない．

2.5 TRW (Threshold Random Walk)

TRW [4] は正当なユーザのコネクション成功確率およびスキャナーのコネクション成功確率を予め設定し，逐次検定 (Sequential Hypothesis Testing) を行うことで，スキャナーを早期に検知する手法である．これまでの手法と比べ，検知までのパケット数が少なく，コネクションの成功までカウントすることで誤検知率を低減させている．

2.6 TRW の改良方式

前述した TRW パラメータの設定や誤検知率が大きいといった問題点がある．参考文献 [5] ではそれらの問題を解決する手法を提案している．TRW の改良方式ではパラメータの設定をヒューリスティックに行い，また，ポートスキャン検知までの履歴情報を格納しておくことで，ホストの停止やネットワークの異常の際に現れる，正常なホストからの TCP SYN 再送をポートスキャンとして判定することを避けることで誤検知率を低減させている．

3 提案手法

本研究では複数の評価基準を用いてポートスキャンを検知する手法を提案する．ポートスキャンの検知に基づいた攻撃への対策の 1 つは検知したポートスキャンを遮断することであるが，誤検知率が大きくては使用することができない．既存の手法のほとんどが誤検知率が大きく，そのような手法を用いたポートスキャンの検知では正常なホストをスキャナーと検知してしまい，フィルタ装置として使用することはできない．見逃し率があっても誤検知率の少ない手法が望ましい．したがって提案方式では誤検知率を少なくするために複数の評価基準を用いている．

この方式では 3 つのモジュールを使ってポートスキャンの検知システムを実現する．本手法の構成図を図 1 に示す．1 つ目はコネクションの成功，失敗を判断するためのモジュール (モジュール A)．2 つ目はコネクションの成功，失敗の情報を使って，イベントを生成するモジュール (モジュール B)．3 つ目はイベントがポートスキャンかどうかを判断するモジュール (モジュール C) である．さらに，これらのモジュールの動作を詳しく説明する．

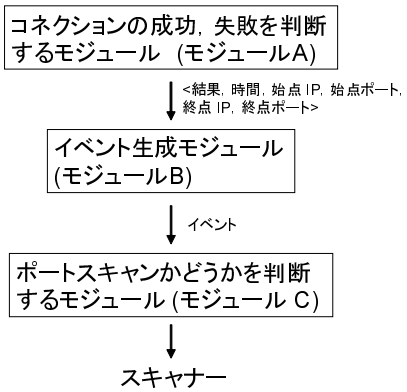

図 1: ポートスキャン検知システムの構成

3.1 コネクションの成功，失敗を判断するモジュール (モジュール A)

このモジュール A はコネクションの成功，失敗を判断するためのモジュールである．入力は TCP SYN, TCP SYN+ACK, TCP RST+ACK パケットである．モジュール A では TCP SYN パケットが観測された際，コネクションが成功したか失敗したかを判断する．これは，TCP SYN+ACK が返されると正常にコネクションの確立に成功したと判断し，TCP RST+ACK が返されるか，t 時間経過しても何も返ってこなかった場合にはコネクションの確立に失敗したと判断する．モジュール A における処理を図 2 に示す．モジュール A はこうして得られたコネクションの成功または失敗の情報を，観測された時間，送信元 IP アドレス，送信元ポート，送信先 IP アドレス，送信先ポートとともに出力する．

3.2 イベントを生成するモジュール (モジュール B)

モジュール B ではモジュール A の出力情報を入力とし，イベントを出力とする．イベントはコネクションの失敗をカウントすることで生成される．このモジュールは各 IP アドレスごとにコネクションの成功，失敗の情報，観測された時間，送信元 IP アドレス，送信元ポート，送信先 IP アドレス，送信先ポートの情報をすべて保持しておく．そのようにして特定の送信元

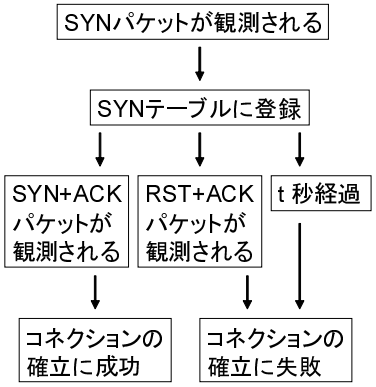

図 2: コネクションの成功，失敗を判断するモジュール (モジュール A)

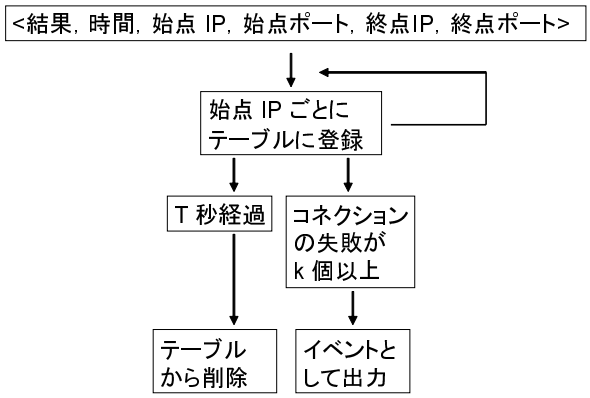

図 3: イベントを生成するモジュール (モジュール B)

IP アドレスから k 個のコネクションの失敗がカウントされるとそれまで観測された情報と共に出力する．これをイベントと呼ぶ．モジュール B における処理を図 3 に示す.

3.3 ポートスキャンを判定するモジュール (モジュール C)

ポートスキャンの判定は，入力された情報をいくつかの基準に基づいて評価し，それらの評価結果に基づいて行われる．本稿で採用した基準は以下の 3 つである．それぞれについて採用した理由を詳しく述べる．

3.3.1 コネクションの成功率

コネクションの成功の情報は重要である．コネクションの失敗ばかりを見ていると，アプリケーションによるポートスキャン等も同時に検知してしまう．したがってコネクションの成功も考慮することで，そのような悪意の無いポートスキャンの検知を少なくすることができる．モジュール C では同じ送信元からのイベント中のコネクション成功率が一定以上であるとそのイベントを破棄する．

3.3.2 同じ送信先 IP アドレスで同じ送信先ポートのものが含まれている割合

同じ送信先 IP アドレスで同じ送信先ポートに対してコネクションの確立に失敗するのは，サーバのダウンやネットワークの異常，フィルタリングによるのが主な原因である．このようなコネクションの失敗はポートスキャンの動向とは異なっている．ポートスキャンは主に情報を収集するために行うものであり，同じ送信先 IP アドレスの同じポートに対して何度もコネクションリクエストを送出することは無い.

3.3.3 送信先 IP アドレスの上位 16 bit が固定されているか

ポートスキャンの対象 IP アドレスの決定に当たっては，送信先 IP アドレスの上位 16 bit ないしは 24 bit を固定してポートスキャンを行う方法が多くのワームで用いられている．これは，効率的にポートスキャンを行うためにこのような手法を用いていると考えられる．したがってこのような特徴を検知することで，精度のよいポートスキャンの検知ができる.

4 評価用データセット

ここでは評価に用いるデータセットについて説明する．観測対象ネットワークは /20 のアドレス空間が割り当てられておりピーク時において 2000 程度のアドレスが利用されている．表 1 に観測期間およびネットワークアドレス空間を示す.

スキャナーおよびワームの判断は，解析者の主観により行った．スキャナーについては，送信先 IP アドレスを変化させながら，10 以上の

表 1: 対象ネットワーク

対象ネットワーク	九州大学キャンパスネットワークの一部
観測期間	2004/11/10〜2004/11/20
アドレス空間 (キャンパスネットワーク)	65,536
アドレス空間 (対象ネットワーク)	4,096

コネクションの失敗が観測された場合で且つ，アプリケーションによるポートスキャンで無い場合にスキャナーであると判断した．その中で，送信先ポートに 135 か 445 が含まれる場合は，ワームと判断している．また，表中のデータは 11 日間の観測期間における 1 日当たりの平均値である．

データ解析の結果を表 2 に示す．表 2 からスキャナーの送出する SYN パケットの数が全体の約半分を占めていることが分かる．また, 表 2 ではポートスキャンとワームによるポートスキャンを区別してあるが，これ以降は特に区別はしない．

5 提案方式の評価

本研究では，提案方式の評価を行うためにキャンパスネットワークにおけるトラフィック観測によって得られたデータセットを用いる．提案方式によるポートスキャン検知の評価結果を表 3 に示す．表 3 ではコネクションの失敗をカウントする手法 (2.2 節参照)，TRW (2.5 節参照), 提案方式の評価結果を示す．複数の評価基準を用いる手法では誤検知率, 見逃し率が他の手法と比べても小さいことが分かる．

しかし，提案方式では検知できないポートスキャンや，逆に誤検知してしまうものも存在する．したがって，提案方式ごとに誤検知率の原因と見逃し率の原因を考察する．また，提案方式では履歴情報を格納するために既存の手法と比べてメモリの使用量が大きい．これについても考察を行う．

5.1 誤検知の原因

誤検知率 の原因はアプリケーションによるポートスキャンである．アプリケーションによるポートスキャンはコネクションの成功を見ることでほとんどの場合は検知されないが，まれに検知されることがある．対策としては特定のアプリケーションとして使われるポートに関してそのコネクションの失敗はカウントしない等が挙げられる．

5.2 見逃しの原因

見逃し率 の原因となるのは長期間かけてゆっくりポートスキャンを行うものである．長期間かけてゆっくりポートスキャンを行うものは検知するが非常に困難である．この原因は明らかで，モジュール B において T 時間しか状態を保持していないからであり，同じホストから T 時間以上何もアクセスがないと状態は失われてしまう．しかし，長時間にわたってアクセスのないホストの情報を保持しておくためには多くのメモリを使用することになり，それに伴う性能低下も考えられる．

5.3 メモリ使用量

提案手法ではメモリの使用量が既存の手法に比べて大きい．これは複数の評価基準を利用してポートスキャンであるかを判断する際にイベントが生成されるまでの情報を保持しておかなくてはならないためである．保持している情報は，k 個のコネクションの失敗がカウントされるまでの < コネクションリクエストの結果，コネクションリクエストが観測された時間，始点 IP アドレス，始点ポート，終点 IP アドレス，終点ポート > である．これらの情報を各リモートホストごとに保持する必要があるので，リモートホストが増えるにつれてメモリの使用量は増大する．今回対象としたネットワークではリモートホストが 30,000 台ぐらいのネットワークであり，K = 10 として評価したので常時必要なメモリは約 5 M バイトほどであった．しかし，対象とするネットワークの規模が大きくなり，リモートホストの数が増えると必要なメモリはそれに比例して増える．さらに，空間が拡大する

表 2: データ解析 (1 日平均)

SYN パケットの数	1,364,243
リモートホストの数	29,243
スキャナーの数	111
ワームの数	34
スキャナーによる SYN パケットの数	96,828
ワームによる SYN パケットの数	535,127

表 3: 誤検知率，見逃し率，ポートスキャン検知までのパケット数

	誤検知率	見逃し率	ポートスキャン検知までのパケット数
コネクションの失敗をカウントする手法	0.418	0.543	約 200
TRW	0.880	0.001	約 4
複数の評価基準を用いる手法	0.02	0.08	約 10

と，検索に必要な時間も増加するため，どれぐらいのネットワークを対象とすればよいのかさらに考察が必要である．

6 まとめ

今回我々は新たにポートスキャン検知手法として複数の評価基準を用いる手法を提案した．本稿ではこの手法に関する誤検知率，見逃し率，メモリの使用量について評価も行った．さらに，ポートスキャン検知までの時間に関する考察を行う必要があると考えている．

また，今後の改良の方針として，アプリケーションによるコネクションの失敗を送信先ポートにより区別する手法を導入することで，failse positive を改善できると考えている．ネットワークを利用するアプリケーションは主に，特定のポートを使用することでサービスを区別している．したがって，コネクションの確立に失敗しやすいアプリケーションの使用するポートに関しては独立に評価する手法についても検討したい．さらに，他にもポートスキャンの検知精度や検知速度を向上させる情報があればそれを用いることでさらに改良を加えていきたい．

本稿で扱った TCP に対するポートスキャンの高速検知手法の応用として，提案方式を利用したポートスキャン遮断フィルタを実装したいと考えている．今回提案した手法はオフライン評価しか行っていない．しかし，提案方式では動的な解析も可能である．したがって，オンライン評価を行って，パケットの取りこぼしなどの問題についても考える必要がある．

参考文献

[1] SnortUsers Manual2.2.0
http://www.snort.org/docs/snort_manual/

[2] Bro User Manual
http://bro-ids.org/manuals.html

[3] Shigang Chen, Sanjay Ranka, “An Internet-Worm Early Warning System” In Proceedings of IEEE Globecom 2004 - Security and Network Management, Dallas Texas, USA, November/December 2004.

[4] Jaeyeon Jung, Vern Paxson, Arthur W. Berger, and Hari Balakrishnan, “First Portscan Detection Using Sequential Hypothesis Testing” In Proceedings of the IEEE Symposium on Security and Privacy, May 9-12, 2004

[5] 小原 正芳, 堀 良彰, 櫻井 幸一, “キャンパスネットワークにおけるポートスキャンの現状およびその自動検知手法” 2005 年暗号と情報セキュリティシンポジウム（SCIS2005）予稿集, pp.1543–1548，2005 年 1 月．