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

2008－DPS－134 (44)
2008－CSEC－ 40 (44)
2008／3／7

アプリケーションレイヤ由来の情報を用いたトレースバック手法に関する研究

井澤 志充　大島 龍之介　国峯 泰裕

株式会社クルウィット
{izawa, ryu, kunimine}@clwit.co.jp

あらまし 能動的な警戒手段としてのトレースバックプラットフォームの実現に向けてアプリケーションレイヤ由来情報を用いたトレースバックアルゴリズムを提案する。
まず本研究で提案するアプリケーショントレースバックの基本手法について述べ、次に対象とするアプリケーションとしてウイルスメール追跡と DNS 反射攻撃を例にその手法について述べる。

Technique of Trace Back system using derived from application layer's information

Yukimitsu Izawa,　Ryunosuke Ohshima,　Yasuhiro Kunimine

Clwit Inc.
{izawa, ryu, kunimine}@clwit.co.jp

Abstract We propose Technique of Trace Back system using derived from application Layer's information to build actively-actuated Trace Back platform. We summarized some scheme to Trace Back Internet traffic. Next, we describe about our application Trace Back framework and virus mail trace back scheme and DNS reflection trace back.

1. はじめに

近年、インターネットの普及およびネットワーク技術の発展によりインターネットは社会インフラとしての役割を担うようになった。それに伴い、Dos(Denial of Serve)や DDoS(Distributed Dos)、ウイルス発信などのサイバー攻撃は、社会に与える影響を増大させている。

これらサイバー攻撃に対し攻撃元探査を行うことで攻撃元がどのホストであるか、あるいは複数あるネットワーク境界のどれから攻撃パケットが流入してきているのかを明らかにする手法であるトレースバック技術のうち、アプリケーションレイヤ由来の情報を利用する手法について述べる。

本研究では、能動的な警戒手段としてのトレースバックプラットフォームの実現に向けてアプリケーションレイヤ由来情報を用いたトレースバックアルゴリズムを提案する。

まず本研究で提案するアプリケーショントレースバックの基本手法について述べ、次に対象とアプリケーションとしてウイルスメール追跡と DNS 反射攻撃を例にその手法について述べる。

2. 先行・関連研究調査

アプリケーショントレースバックアルゴリズムを開発するにあたって、まず、トレースバックの先行研究分野である IP トレースバックについて文献調査を行った。また、関連分野である踏み台検出手法について文献調査を行い、その手法や特徴についてまとめた。

2.1 IP トレースバック

まず、関連研究である IP トレースバックの概要と長所・短所について整理した。

インターネット上の各通信ノードには郵便における住所に相当する IP アドレスが割り宛てられており、通信に用いられる。しかし、IP パケットの発 IP アドレスに偽のアドレスが書かれていた場合には、発信ノードを特定することができない。このように IP パケットに偽の発 IP アドレスを書き込む行為は IP 詐称(IP spoof)と呼ばれている。IP 詐称を行うと、着信ノードからの返信を受けとることができないため、発信ノードから着信ノードへの単方向の通信しか行えない。しかし、発信ノードの身元を隠蔽することができるため、DoS 攻撃やアイドルスキャンに悪用されることがある。

IP トレースバックは、IP 詐称されたパケットの通信経路を追跡することを目的とした技術である。数多くの先行研究がなされており、逆探知パケット方式[1][2]、マーキング方式[3]、ハッシュベース方式[4]などの手法が提案されている。

どの手法もネットワーク上に観測点を配置して、パケットの通過を検出することによりトレースバックを可能とするものである。それぞれの手法では、あるパケットがある観測点を通過したことを記録/通知する方法に違いがある。各手法の対比を表に示す。

表 1 各トレースバック手法の対比

	逆探知パケット方式	マーキング方式	ハッシュベース方式
パケットへの書き込み	なし	あり	なし
単一パケットトレース	不可能	可能	可能
観測点との通信	なし (通信パケットを利用)	サンプリングごとに発生	トレースごとに発生

本研究で提案するアプリケーショントレースバック手法でも IP トレースバック手法と同様に観測点を配置した構成が必要になると思われる。観測点との通信方式としては、通信パケットへの変更を行う方式やサンプリングごとに通信が発生する方式はデプロイ面で問題がある。ハッシュベース方式のようにトレース時のみ通信を行う方式が導入しやすいものと思われる。

2.2 アプリケーショントレースバック

IP トレースバックはネットワーク層レベルでの経路追跡を目的とした技術であるのに対し、アプリケーショントレースバックは、トランスポート層以上の通信を追跡するための手法をさす造語であり、IP トレースバックのみでは追跡不能なアプリケーションレベルでの通信の追跡を目的として本研究で開発する技術である。関連する先行研究としては、踏み台検出(Stepping-stone detection)がある。

2.3 踏み台検出

IP トレースバックのみでは追跡不能な「踏み台ホスト」を用いた通信の追跡を行う手法として、踏み台検出(stepping-stone detection)が研究されている。踏み台検出とは、複数の中継ホスト(踏み台ホスト)を介した通信の追跡を目的としたものであり、各ホストの入セッションと出セッションの対応付けを行うことにより、順次踏み台ホストを辿ってゆく方法で、通信の経路を追跡する手法である。

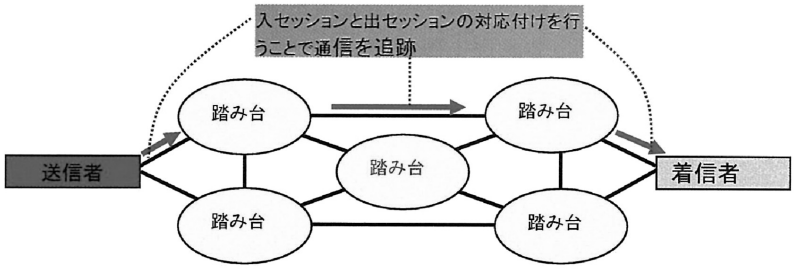

図 1 踏み台検出手法の概要

踏み台検出手法は、ホストベースとネットワークベースに分類される。

2.4 ホストベース

ホストベースは、踏み台になる可能性があるホスト内に踏み台検出のための機構を組み込んでおき、そのホストの入セッションと出セッションの対応付けを行う手法である。OS やアプリケーションの情報を参照できるため、セッション対応付を確定的に行うことができる長所がある。また、ネットワーク上の情報を参照しないので、ネットワーク上での暗号化やネットワークの通信量の影響を受けない。しかし、トレースバックのための機構をホストごとに用意する方式のため、デプロイにコストがかかるという短所もある。先行研究としては、OS 内のプロセスの親子関係の情報を用いる STOP[5]や、セッション確立時に経路情報を交換するように改造された Telnet や ssh のサーバ・クライアントを用いる CIS[6]などがある。

2.5 ネットワークベース

ネットワークベースは、通信経路となる可能性のあるネットワークに観測点を用意しておき、観測されたパケットのタイミングや内容から入セッションと出セッションの対応付けを行う手法である。各ホストへの追加変更が必要なく、OS や環境への依存が少ないことが特長である。しかし、パケットの対応付けに統計的手法を用いるため、対応付けの確度が低いという短所がある。ネットワークベースには、パケットのタイミングを参照してセッションの対応付けを行う時間相関方式(time correlation)と、パケットの内容を参照してセッションの対応付けを行う内容相関方式(content correlation)とがある。

時間相関方式

時間相関方式は、通信経路で観測したパケットのタイミングの特徴を用いて入セッションと出セッションの対応付けを行う方式である。パケットの内容を参照しないので、対象アプリケーションへの依存性が低いことが特長である。また、暗号化通信についても適用できる。しかし、セッションの対応付けに統計的手法を用いるため、単発的な通信には対応できず、対応付けの確度が低いという短所がある。また、通信量の少ないネットワークにしか対応できない。先行研究としては、データが流れている期間と流れていない期間の周期特徴を用いてセッションの対応付けを行う ON/OFF Periods やパケット遅延の偏差を用いてセッションの対応付けを行う Deviation[7]、パケット間の時間間隔を用いてセッションの対応付けを行う Inter-Packet Delay(IPD)[8]、Multi-scale Detection[9]などがある。

内容相関方式

内容方式は、通信経路で観測したパケットの内容の特徴を用いて入セッションと出セッションの対応付けを行う方式である。時間相関方式よりも情報源のエントロピーが高いため確度を高くすることができ、通信量の多いネットワークにも適用できるという長所がある。しかし、セッションの内容を参照するため暗号化通信には適用できないことが短所である。先行研究としては、通信内容の文字出現順序の特徴を用いてセッションの対応付けを行う Thumbprints[10]や、Telnet などのテキスト通信に透かし情報を埋め込むことによりセッションの対応付けを行う Watermarking[11]などがある。

これらの調査から、提案システムを設計するうえで考慮すべき点を次にまとめた。

ネットワークベース手法が有利

ISP のような大規模なネットワークにおいては、多種多数のホストが存在しているため、ホストベースよりもネットワークベースの方がデプロイ面で有利である。提案システムはネットワークベース手法とするのが良い。精度向上等の目的で、ホストベース的な手法を用いる場合には、末端クライアントではなく、ファイアウォール等の中継ノードや、Web、Mail などの主要サーバに限定するべきである。

統計的手法のみに頼らない

統計的手法のみを用いた手法はセッション対応付けの精度が低く実用に難がある。提案システムでは、統計的手法のみに頼ったアルゴリズムは避けるべきである。

セッション切断後のトレースができるものとする

手法によっては、トレース対象のセッションが張られた状態でのみトレースが可能なものもあった。インシデントハンドリングでの応用を考慮すると、セッションが切断された後であってもトレースが可能であるものが望ましい。踏み台検出方式の多くは、Telnet や FTP などの NVT(Network Virtual Terminal)による対話的プロトコルによる踏み台通信の検出を対象にしているものであった。

3. 対象アプリケーションの検討

日本のインターネットにおける総被害のほとんどは、ウイルスによるものである。ウイルスの被害や検出数が増加している。これにはウイルス作者を追跡するのが困難であることが背景にあると考えられる。ウイルス犯罪の多くは、理性が欲望を抑えきれず犯してしまった犯罪とは違い、トレンドマイクロが推測するようにハッカーの悪戯や商業的目的だとすれば、攻撃元を追跡されることはウイルス作者にとって大変脅威なことであると予測できる。総務省「通信利用動向調査」によると、ウイルス感染経路は、MSBlaster が爆発的に蔓延した 2001 年 8 月を除けば、概ね 95%以上はメール経由である。それゆえにウイルスの最大の感染ルートであるメールに焦点を絞り、メールからの感染を確実に追跡できるならば、ウイルス感染の大部分を減少させる可能性をもち、さらにウイルス作成者を検挙に導くことも期待できる。

また近年、インターネットの基幹をなす DNS サーバへの攻撃として DNS 増幅攻撃が脚光を浴びている。

そこで本研究では対象アプリケーションにメールおよびDNSを選択し、ウイルス感染経路の追跡を行う手法およびDNS増幅攻撃の追跡を行う手法についてのアプリケーショントレースバックアルゴリズムを提案する。

4. フレームワーク

これまでの検討をベースに、アプリケーショントレースを実現するためのアルゴリズムとして、ハッシュベースのアプリケーショントレース手法を提案する。

本手法はハッシュベースのIPトレースバックを参考に考案したものであり、次の手順で通信経路の追跡を行う。

通信経路となるネットワークに観測点を配置する。各観測点では、トレース対象アプリケーションの通信パケットから、通信セッションを特定するための特徴情報を抽出し、そのハッシュ値を算出してデータベースに記録する。

セッション観測の補助手段として、通信経路となるルータやサーバ、ホストなどからのログ情報を利用することも検討する。必要に応じてログ情報からセッションを特定するための特徴情報を抽出し、そのハッシュ値を算出してデータベースに記録する。

トレース実行時には、まず、トレース対象通信の特徴情報を抽出し、そのハッシュ値を算出する。このハッシュ値を持つ通信の通過/非通過を各センサに問い合わせて回答を得る。通過した観測点をつなぎあわせることで、当該セッションの通信経路を特定する。

提案アルゴリズムで用いるシステムの構成を図2に示す。

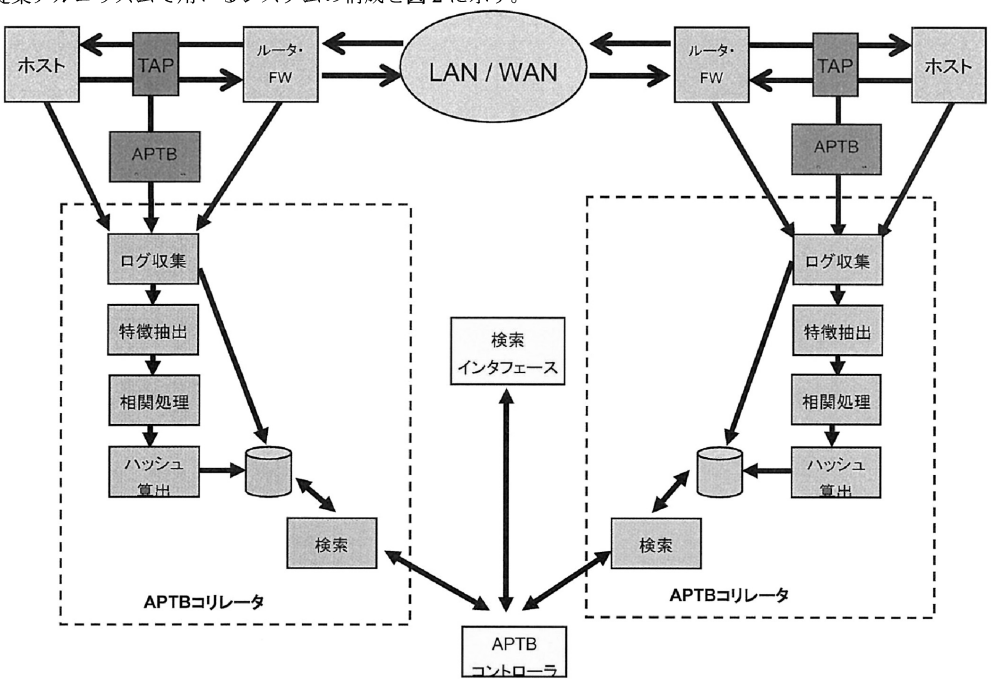

図2 ハッシュベースAPTBのシステム構成

4.1 機能構成

アプリケーショントレースを実現するAPTBシステムは、APTBプローブ, APTBコリレータ, APTBコントーラの各機能要素で構成される。

APTBプローブは、通信経路となるネットワークを観測してトレース対象アプリケーションの通信に関する情報を取り出し、ログを生成する機能である。各ASやサイトの境界ネットワークに設置し、入セッションと出セッションをそれぞれ監視する。

APTBコリレータは、APTBプローブや各種サーバ、ファイアウォール、ルータ,IDS等から受信したログ情報からセッションを特定するための特徴情報を抽出する機能である。各サイトやASごとに配置する。どの情報源からログを収集するのかについては、トレース対象とするアプリケーションごとに検討する必要がある。抽出した特徴情報から、入セッションと出セッションの相関を取ることにより、サイトやAS単位での踏み台検出を行う。

抽出した特徴情報は、ハッシュアルゴリズムにより匿名化したうえでデータベースに格納する。このデータベースを用いてAPTBコントローラからのトレース要求に対して、当該セッションの通過/非通過を返答する。
APTB コントローラトレースバック管理者や、上位トレースバックシステムからのアプリケーショントレース要求を受け付けて、各APTBコリレータにトレース要求を発行し、APTBコリレータからの返答を集約して、トレース結果を提示する機能である。

5. メールトレースアルゴリズム

提案手法では、サイト境界に設置したAPTBプローブでSMTPチャットから電子メールの添付ファイルに関する情報を取り出し、添付ファイルの特徴を抽出して蓄積することにより、ウイルス感染発生時にウイルスメールの感染経路を追跡する。

提案手法を実現するためのシステムは、APTBプローブ、NIDS、APTBログサーバ、APTBコリレータから構成される。

APTB プローブ

サイト内のSMTPサーバの通信をキャプチャし、SMTPチャットから添付ファイルの送受信に関する情報を取り出してログを出力する装置である。SMTPサーバの通信が監視できる場所(SMTPサーバが接続されているスイッチのミラーポートなど)に設置する。SMTPチャットから取り出す情報としては、添付ファイル名、添付ファイルサイズ、添付ファイルから算出したハッシュ値、HELOドメイン名、MAIL From:アドレス、RCPT To:アドレス、Subject:、などである。

NIDS

サイト内のウイルス感染ホストを検出するための NIDS。既存のウイルス検知機能を持つ NIDS を使用する。主にサイト内のホストからサイト外への通信を監視し、ウイルス起因と思われる非定常な通信を検知してアラートログを生成する。このログは、後述のAPTBコリレータにおいてウイルス感染疑いホストの検出に使用する。

APTB ログサーバ

APTB プローブや NIDS からログを受け取り、ログに含まれるプライバシ情報の匿名化処理を施して、検索可能な形で蓄積する機能。APTB コリレータからの要求に応じてログの検索を行う。ログデータベースとしての機能と同時にAPTBコリレータへの入力バッファとしての機能を持つ。

APTB コリレータ

特徴抽出機能とトレース機能の2つの機能で構成される装置である。

特徴抽出機能は、APTB ログサーバから受信したログからセッション特定のための特徴情報を抽出して検索可能な形式で蓄積する機能である。特徴情報としては、APTB プローブで観測された電子メール添付ファイルのハッシュ値を用いる。

トレース機能は、管理者や上位トレースバックシステムからのトレース要求を受け付け、与えられた特徴量を持つメールが過去に通過したかを検索し、入セッションと出セッションの対応付けを調査することで当該メールがサイト内のホストで中継された可能性があるかどうかを返答する機能である。また、IDS でウイルス感染疑いのアラート検出されたホストが送受信した添付ファイルの記録を照会することで、ウイルス疑いファイルを推測し、そのウイルス疑いファイルを送受信したホストを列挙することで、ウイルス疑いホストのリストを作成する機能についても検討中である。

6. DNS トレースアルゴリズム

DNSトレースアルゴリズムが対象とする攻撃をDNS反射攻撃あるいはその一種であるDNS増幅攻撃である。DNS反射攻撃について図3に示す。

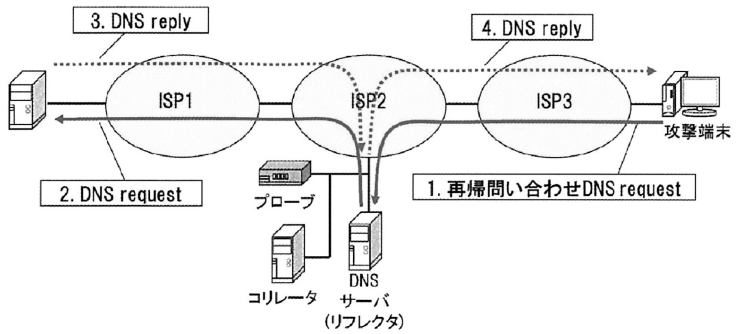

図3　DNS反射攻撃

APTB プローブ

サイト内の再帰問い合わせ DNS Request パケットおよび DNS Request パケットをキャプチャし、これらのパケットから IP ヘッダの情報および QNAME のハッシュ・QCLASS・QTYPE に関する情報を取り出してログを出力する装置である。DNS クエリの通信が監視できる場所(SMTP サーバが接続されているスイッチのミラーポートなど)に設置する。

7. まとめ

アプリケーショントレースアルゴリズムを開発するにあたりまず、関連研究に関する文献調査を行い関連分野である IP トレースバックと踏み台検出手法についてまとめた。何れも研究段階の技術であることや、そのままの形では、現実的な問題を解決する手段として応用しづらいことを把握した。

これらの調査結果をふまえて、踏み台ホストを悪用した不正行為の抑止と、インシデント対応の支援を目的としたアプリケーショントレースバックアルゴリズムについて検討し、ハッシュベース APTB 方式を提案した。提案システムでは、ネットワークベースとホストベースを組み合わせた方式を採用することで、ネットワークベース手法の特長であるデプロイ性の良さと、ホストベース手法の特長である精度の高さを持ち合わせた方式とした。提案手法の応用として、メールによるウイルス感染の経路追跡を行うシステムおよび DNS 反射攻撃の経路追跡を行うシステムの構成について考察した。

現在、APTB コリレータおよび APTB プローブの実装を行っており基礎部分での接続性実験を終え、リファイン作業のフェーズおよび IPTB システムとの連携、本システムを包含するトレースバック機構との連携のフェーズに移行している。

謝辞

本研究は、独立行政法人情報通信研究機構の平成 17 年度からの研究案件「インターネットにおけるトレースバック技術に関する研究開発」の一部である。

参考文献

[1] Steve Bellovin, Marcus Leech, Tom Taylor: ICMP Traceback Messages, Internet draft. Document: draft-IETF-iTrace-04.txt, (2003).

[2] 甲斐, 中谷, 清水, 塚本, "不正アクセスに対する高性能発信源探査方式の提案", 情報通信研究季報, Vol51, pp.41-49, (2005).

[3] S. Savage, D.Wetherall, A. Kerlin, T. Anderson, "Practical network support for IP traceback", Proc. of ACM 2000 SIGCOMM Conference, pp295-306, (2000).

[4] A. C. Snoeren, C. Partridge, L. A. Sanchez, C. E. Jones, F.Tchakountio, S. T. Kent and W. T. Strayer, "Hash-based IP Traceback", In Proceedings of SIGCOMM '01 (2001).

[5] CARRIER, B., AND SHIELDS, C. A recursive session token protocol for use in computer forensics and tcp traceback. In Proc. IEEE Infocom '02 (June 2002), (2002).

[6] JUNG, H. T., KIM, H. L., SEO, Y. M., CHOE, G., MIN, S. L., AND KIM, C. S. ,"Caller identification system in the internet environment", In Proc. USENIX Security Symposium '93 (Oct. 1993), (1993).

[7] YODA, K., AND ETOH, H., "Finding a connection chain for tracing intruders", In Proc. European Symposium on Research in Computer Security (Oct. 2000), pp. 191--205, (2000).

[8] WANG, X., REEVES, D. S., AND WU, S. F.,"Inter-packet delay based correlation for tracing encrypted connections through stepping stones", In Proc. European Symposium on Research in Computer Security (Oct. 2002), pp. 244--263, (2002).

[9] DONOHO, D. L., FLESIA, A. G., SHANKAR, U., PAXSON, V., COIT, J., AND STANIFORD, S., "Multiscale stepping-stone detection: Detecting pairs of jittered interactive streams by exploiting maximum tolerable delay", In Proc. International Symposium on Recent Advances in Intrusion Detection (Oct. 2002), pp. 17--35, (2002).

[10] STANIFORD-CHEN, S., AND HEBERLEIN, L. T.,"Holding intruders accountable on the internet", In Proc. IEEE Symposium on Security, (1995).

[11] WANG, X., REEVES, D. S., WU, S. F., AND YUILL, J., "Sleepy watermark tracing: An active network-based intrusion response framework", In Proc. International Conference on Information Security (June 2001), pp. 369--384, (2001).