Vol.2015-GN-93 No.39
Vol.2015-CDS-12 No.39
Vol.2015-DCC-9 No.39
2015/1/27

Web アクセスログを観察するインタフェースの提案

斉藤　典明

サイバー攻撃に代表されるようにインターネット空間には様々な脅威があり、Web サーバを公開していると様々な種類のアクセスを受けるようになる。Ｗｅｂサーバで観測されるアクセスログには正規のアクセスと不正なアクセスが混在している。Ｗｅｂサーバ管理者は、不正なアクセスに気づき対策を取ってゆくことが必要であるが、不正なアクセスがどのくらいあるのかを把握することが難しい場合がある。そこで、Web アクセスの全体傾向を把握するために、すべて IP`アドレスを時系列で視覚化できるインタフェースによりアクセスログを観察する方式を提案する。

An Observation Interface for Web Access Log

SAITO Noriaki†

There are a lot of threats in the Internet, for example, cyber attack. When we connect a web server to the Internet, many suspicious accesses will be arrived. In the accesses, both normal accesses and illegal accesses are included mixed. To protect their web server, web server administrators have to avoid cyber attacks, however grasping of the tendency of the illegal accesses is difficult. Then, to grasp a tendency of the web server access, to grasp a tendency of the web server access, an observation interface for web access log is proposed which enables to show us all access logs by using full length ip addresses and date information.

1. はじめに

ここ 1～2 年の間に急速にサイバー攻撃に対する脅威が増加している。特に、端末や Web サーバをインターネットに接続すると、外部からのアクセスが発生する。Web サーバにまつわる深刻な脆弱性もいくつか報告されている[1],[2]。そのため、管理が不十分な Web サーバの場合は、Web サーバの脆弱性が攻撃されコンテンツに攻撃コードが埋め込まれることや、Web サイトにマルウェアが設置されるなどの被害を受けることがある。しかしながら、これらの被害は、Web サーバ管理者がすぐに気が付かないこともあり、そのサイトの閲覧者がマルウェア感染する、サーチエンジン業者が検知し検索結果に警告表示が出されるなど該当のサイトへのアクセスへ支障が出てから気づくことになる場合もある[3]。このようなことから、Web サーバの管理を普段から徹底するように Web サーバ管理者への注意喚起[4][5]なども行われている。

Web サーバの管理者は、サイバー攻撃に対して、OS や Web アプリケーションを脆弱性のない最新版にアップデートするか、WAF などのセキュリティ製品を導入することが定番の対策になる。このような対策の実施に対して、実際の脅威に対してどのくらい効果があったのか、危険なアクセスがどの程度あるのかを把握する方法は少ない。そこで、Web サーバの管理者が現在置かれている状況を確認するために、Web サーバのログを全 IP アドレス空間上に時系列でマッピングすることで視覚的に Web サーバの潜在的な脅威を把握できるログ観察インタフェースを提案する。

† NTT　セキュアプラットフォーム研究所
NTT Secure Platform Lab.

2. アクセスログの確認

Web サーバの定番のセキュリティ対策の一つに WAF の導入がある。しかしながら、サイバー攻撃は年々巧妙化しているため WAF では完全には防ぎきれない。そのため Web サーバ管理者は適宜 Web サーバのアクセスログを確認し、不正アクセスの有無を確認する必要がある。

一般的な Web サーバのログを確認すると、コンテンツへのアクセスを目的とした正規のアクセス、サーチエンジンのロボットによるクローリング目的のアクセス、脆弱性への攻撃や偵察を目的とした不正なアクセス、かならずしも攻撃目的や偵察目的と判別できないアクセスに分類できる。

従来、Web サーバのアクセスログの活用は、正常なアクセスを分析し、コンテンツ作成の改善や、ユーザ特性を把握することに使われてきた。ここでは、コンテンツへのアクセスである正規のアクセスが分析対象であり、正規でないアクセスは分析対象外であった。反対に、脅威への対処目的でアクセスログを確認する場合は、正規ではないアクセスを確認することになる。ここでは、目視で確認するか、分析ツールなどで不正なアクセスの痕跡を抽出し確認する方法[6]が定番である。

インターネット空間上に危険なサイトがどのくらいあり、どのくらいの頻度でどこからどこに攻撃をしているのかは完全には把握できていないが、様々な方法でサイバー攻撃の状況を観測できる。例えば、nicter[8]や DAEDALUS[9]ではダークネットで観測された攻撃情報をリアルタイムに 3D 表示で可視化している。これらの情報によればインターネット上には非常に多くの攻撃があることが実感できる。しかしながらこれらは、正規のアクセスと不正なアクセスが混在する環境において、観測されたアクセスに対してどのように対処してゆくべきかを分析するには

Vol.2015-GN-93 No.39
Vol.2015-CDS-12 No.39
Vol.2015-DCC-9 No.39
2015/1/27

適していないと考えられる。

観測されたアクセスログは、個々のリクエストごとに有害なアクセスなのか、無害なアクセスなのかを確認できるものの、アクセスをフィルタするべきか否かを判断することできない。これらの判断は、全体的な傾向の中で判断されるべきであるからである。そこで、Web サーバ管理者が、正規のアクセスと不正なアクセスが混在する環境で、全体的な傾向を把握するために、IP アドレス空間に静止した状態でアクセスデータをマッピングし、鳥瞰と詳細化を行き来できるようにすることと、関連する情報との視覚的なマッチングすることにより、不審なアクセスログを見つけ出すインタフェースを提案した[7]

ここでは、インターネット上で観測された IP アドレスを、正規のアクセス、不正なアクセスなどのアクセスの種類ごとに、16bit 単位で鳥瞰から詳細化までできるアドレス軸と時間軸で構成される空間上にマッピングした。これにより不審なアクセスの多いアドレス帯と、不審なアクセスの少ないアドレス帯があることなどが視覚的に把握できるようになった。

しかしながら、IP アドレス空間は IPv4 であっても広大であり、PC の画面サイズの中で単純に全体の鳥瞰と個別事象の詳細な確認を両立することは難しかった。そこで、該当のインタフェースを拡張し、IP アドレス空間の中で着目するべきアクセスログを抽出し、抽出したアクセスログを中心に時系列アドレス空間を参照するインタフェースを実現した。本報告は拡張したインタフェースによる分析方法について述べる。

3. インターネット空間の可視化

本章では、はじめに当初提案したインタフェースの詳細について説明する。続いて、当初提案したインタフェースにおける課題と解決した拡張方式について説明する。その後、動作確認を実施した実施例について説明する。

3.1 アクセスログのプロット

アクセスログをプロットする時系列アドレス空間は、縦軸に IP アドレス、横軸に時間軸として日付とした。横軸は固定であるが、縦軸は、16bit 長とし、鳥瞰する際には IP アドレスの第一オクテッドおよび第二オクテッドで集約される空間とし、表示したアドレス空間の中の 8bit 単位で選択し、選択された領域について 16bit 長に拡大して詳細化してゆく（図 1）。

次にアクセスログは、IP アドレスと日付の対でグルーピングし、個々のアクセスログとして時系列アドレス空間にプロットする。個々のアクセスログはアクセス数、アクセスパターンにより類別でき、これらをプロットする点の大きさと色で表現した。

アクセスパターンについては正規のアクセス、サーチエンジン業者のクローラによるもの、脆弱性を探しに来ている不正なアクセス、判別不能のアクセスに分類し、正規のアクセスを青、クローラを緑、不正なアクセスを赤、判別不能なアクセスを黄と色分けした。分類したアクセスの中に含まれるリクエスト数単位のアクセス数の規模で点の大きさを表示した。この時、点の大きさを単純にアクセス数に比例させると、アクセス数が少ない場合は目立たなくなり、アクセスが多い場合は画面に対して領域を取りすぎ障害になることから、アクセス数が少ない場合は実際のアクセス数よりも大きい点で、アクセス数が多い場合は実際のアクセス数よりも小さい点で表示するように大きさを調整した。これらアクセスログの時系列アドレス空間上へのプロットは、Web ブラウザ上の canvas オブジェクトで実現した。

このインタフェースに手元の Web サーバのログデータをプロットし、動作確認を行った（図 3）。この方式では、全体の傾向からアクセスのあるアドレス空間とアクセスの少ないアドレス空間があることなどがわかった。

一方で、このインタフェースは Ajax 方式で画面を遷移していたが、データ量が膨大になると動作が不安定になる問題があった。さらに、鳥瞰した段階では複数のアクセスログが関係していると想定されたが、該当部分を詳細化してゆくと、それらが関係しないことが判明することがほとんどであった。そのため、膨大な時系列アドレス空間を探索するというよりは漂流することになる問題があった。

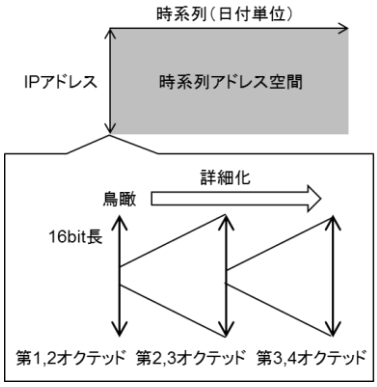

図 1 アドレス空間の鳥瞰と詳細化

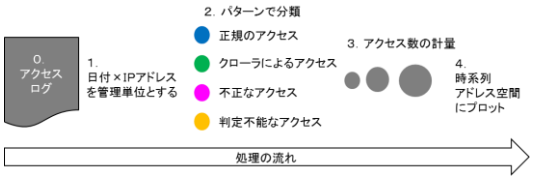

図 2 アクセスログのプロットまでの流れ

Vol.2015-GN-93 No.39
Vol.2015-CDS-12 No.39
Vol.2015-DCC-9 No.39
2015/1/27

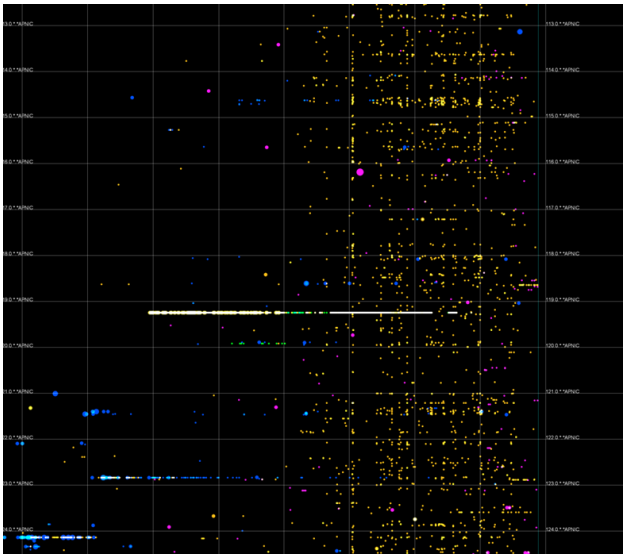

図 3 アクセスログの表示例

3.2 アクセスログ観察インタフェースの拡張

そこで、前節の問題を解決するために、次の方法で拡張した。動作が不安定になることに対しては、canvas オブジェクトのみで表示を切り替える方式とした。これにより Ajax よりも軽量に画面遷移なしで表示を切り替えることが可能になった。

時系列アドレス空間を漂流する問題に対しては、システムがナビゲーションする方針とし、注目するべきアクセスログを抽出しユーザに提示する方法で解決することとした。ここでは、注目するアドレスは IP アドレスの第 4 オクテッドの 8bit のアドレスブロック単位で抽出し、抽出したブロックを並列化し最大 16bit 長で鳥瞰する方式とした(図 4)。

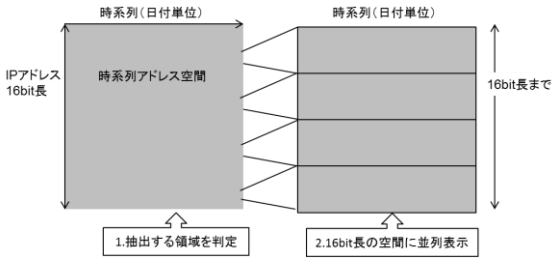

図 4　着目する部分の並列表示

3.3 Web サーバログでの実施例

拡張した提案方式について、手元の Web サーバのログに対して実施した。対象とした Web サーバのログは一般家庭に設置された Web サーバで、特に大きく公開したものではなく、仲間内での情報共有に活用している程度のサーバである。ここには 2006 年 8 月からのアクセスログがリクエスト数単位で 17 万件程度記録されている。これを図 2 の流れにそって処理するとアクセスログ数は約 29,300 件となる。分類結果は、正常なアクセスが約 11,700 件(40.0%)、クローラによるものが約 1,800 件(6.1%)、不正なアクセスが約 1,300 件(4.5%)、判別不能が約 14,400 件(49.4%)であった。これを時系列アドレス空間にプロットすると図 3 のようになる。図において白の横線が 8bit 単位でのアドレス軸であり、白の縦線が 1 年単位での時間軸である。プロットされた点のインターネット空間における分布をみることで、不審なアクセスの多いアドレスブロックと不審なアクセスの少ないアドレスブロックなど、アドレスブロックごとの傾向が把握できる。また、横軸を中心に見ることで、不審なアクセスが増加した時期などの傾向がつかめる。

しかしながらこのままでは、鳥瞰した画面から詳細化してゆき個々のアクセスを分析することは困難であった。これは、鳥瞰した際のアドレスブロックの中身には、傾向の異なる所有者が混在している。そのため鳥瞰した時系列アドレス空間から詳細化してゆくと傾向がつかめなくなってゆくことがある。

これに対して、IP アドレスの所有者情報を基に時系列アドレス空間を並べ直すことも一つの方法ではあるが、これらの情報を確実に取得することは容易でないことから別な方法を検討した。

多数のアクセスログの中から注目するべきアクセスログを抽出する方法とし、機械的な分析処理を施したのちに、アクセスログ観察インタフェースで表示する方法とした。機械的な分析は、例題として一定の文字パターンを含むアクセスログ、一連のアクセスが同じパターンのもの、他の情報とのマッチングの 3 種類で実施した。このうち、他の情報とマッチングでは、マッチング情報が大量にある場合と、少ない場合で表示方法を変えた。実施結果について述べる。

3.3.1 文字パターンによる分類

不正と言われるアクセスは、有名なものとしてクロスサイトスクリプティングや、リモートファイルインクルード、SQL インジェクション、OS コマンドインジェクション、ディレクトリトラバーサルの他に、2014 年から出現したものとして shellshock などがある。これらはアクセスログに残っている一定の文字列パターンから判別できる。

このうち、2014 年 9 月 24 日に発表された bash の脆弱性に対する攻撃である shellshock の攻撃を抽出し、発覚日とセットでプロットした結果を図 5 に示す。図の縦の黄色い線が shellshock の発覚した日付である。青の縦線が解析した日である。赤の横軸は検知した IP アドレスを示す線である。時系列アドレス空間上にプロットされたアクセスログを見ると shellshock の発覚前には該当の不正アクセスは観測されなかったが、発覚後すぐに不正アクセスが試されていることがわかる。さらには、大量にアクセスを仕掛けてきているログも観測された。発覚した不正アクセスには迅速に対応する必要があることがわかる。

他の不正アクセスの抽出結果については 4 章で述べる。

Vol.2015-GN-93 No.39
Vol.2015-CDS-12 No.39
Vol.2015-DCC-9 No.39
2015/1/27

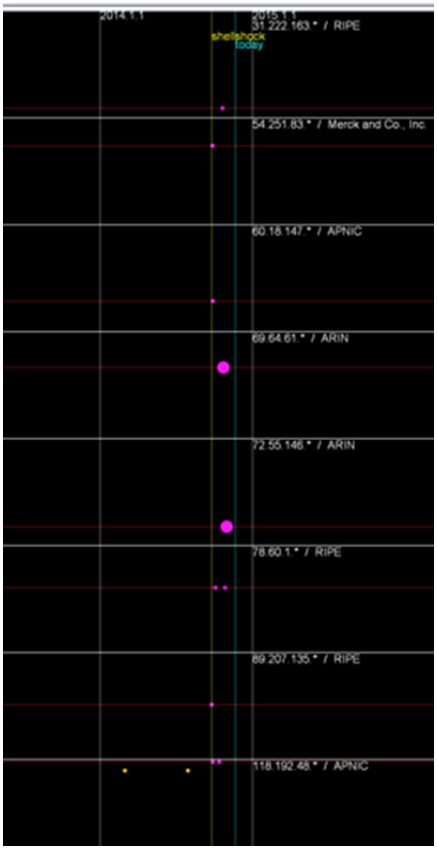

図 5 shellshock の検出例

3.3.2 同一アクセスパターンによる分類

不正アクセスは、ブラックマーケットで出回っているツールを使って実施していると考えられる。そのため、同じツールの利用であれば同一のログパターンになることが想定される。そこで、不正なアクセスとして判定したアクセスログに対して同一性を分類した。その結果 137 個のパターンを抽出した。これらを順に確認した結果、同じアドレスから連続した同一のアクセス(図 6)や、同じ時期に複数のサイトから同一のアクセス(図 7)など、いくつかの特徴的なアクセスを観測できた。

例えば、図 6 のアクセスは約 1 年に渡って不正なアクセスを繰り返していることがわかる。ここのサイトからのアクセスは止めるべきであると判断できる。図 7 は、複数のサイトからほぼ同時にアクセスされている様子である。これらのサイトはなんらかの関係があると思われる。

このように、システムが注目するべきログを提示することで、鳥瞰した時系列アドレス空間の中から確認するべきポイントがいくつか明らかになってゆく。これらの分析や結果の活用方法は今後の課題である。

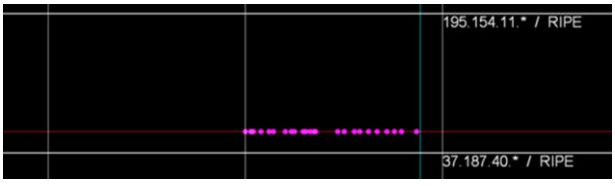

図 6　同じアドレスからの連続したアクセス例

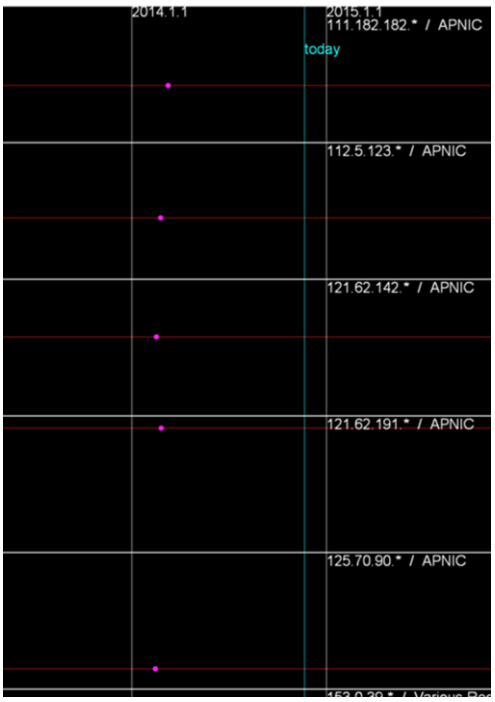

図 7 複数アドレスからのアクセス例

3.3.3 他の情報リストとの比較

不正なアクセスを行うサイトの情報はブラックリストとして出回っている場合がある。これらと手元の Web サーバのログを突合することにより、自サイトが危険にさらされているかどうかを確認するために有用である。ブラックリストの規模は数万件に上るものから、数件のものまで様々なものがある。提案インタフェースで現在表示できる時系列アドレス空間は 16bit に限定しているため、情報リストが多い時と少ない時で表示方法のアプローチを変えた。

情報リストが多い時は、閲覧中の時系列アドレス空間に情報リストの情報をオーバーレイして見せる方式とした(図 8)。反対に 16bit のアドレス軸に収まる程度に少ない場合は、該当のアドレスブロックを抽出し表示する方式とした(図の例示は図 5,6,7 とほぼ同じなため省略する)。

両者の場合も、世の中のブラックリストに該当するアクセスログが手元の Web サーバで観測できたかを把握することが可能になる。

図 8 において黄色の横線が他の情報リストに掲載されている IP アドレスである。これは例えば、他のサイトで観測された攻撃が観測された IP アドレス、特定の利用法の IP アドレス、特定の利用者の IP アドレスなどが考えられる。時系列アドレス空間上でリストアップされた IP アドレスごとにアクセスログとのマッピングを行う。これにより特に判定不明のアクセスに対して、アクセスの意味づけや判定が容易になる。

Vol.2015-GN-93 No.39
Vol.2015-CDS-12 No.39
Vol.2015-DCC-9 No.39
2015/1/27

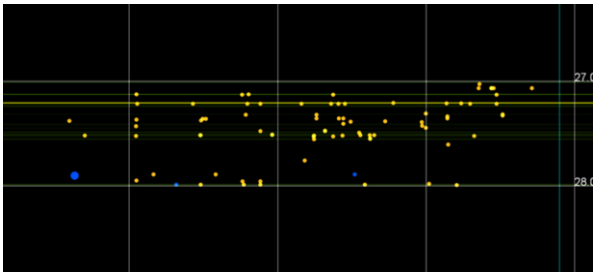

図 8 他の情報リストとの比較例

3.4 実施結果の考察

図 3 のようにアクセスログを単純に時系列アドレス空間にマッピングしただけでは、どのように詳細化してよいのかがわからなかった。これに対して拡張したインタフェースでは、機械的な分析と組み合わせ、注目するべき時系列アドレス空間を並列表示した。これにより、特定の不正アクセスがいつごろからどのくらい発生したのか、一定のパターンの不正アクセスがどのくらいあり、どのようなパターンがあるのかが観測できた。このことから鳥瞰から詳細化に移行するにはナビゲーションする機能が有効であることが確認できた。また、他の情報リストとの比較により不正アクセスの意味づけなども可能になるが、今回実施したいくつかの情報リストは、ほとんど一致がなかったため、他で観測された情報が自サーバのアクセスログの分析にはあまり役に立たないことも判明した。

4. 既存ツールとの比較・組み合わせ

提案インタフェースによるアクセスログ解析の利点を、既存のアクセスログ解析ツールと比較することで考察を行った。

4.1 既存ツールによる抽出

アクセスログを観察し危険なアクセスがどのくらいあったかを把握するために Web のログファイルを解析するツールを使う方法がある。例えば、IPA の iLogScanner である[6]。これは Web サーバのアクセスログから攻撃と思われる痕跡を検出するためのツールである。利用方法は、ログファイルを指定すると自動的にいくつかの不正アクセスパターンを抽出し該当のログの件数をカウントする。これを使うことによって Web サーバ管理者は、自サイトがどの程度危険にさらされているかを知ることができる。

提案インタフェースの動作確認に用いたログデータをこのツールにかけると図 9 のように出力される。出力レポートは、件数の他に該当のログ部分についても抽出される。

検出したWebサイトへの攻撃について、下記に詳細を記述します。

検出対象脆弱性	攻撃があったと思われる件数	攻撃が成功した可能性の高い件数
SQLインジェクション	10	0
OSコマンドインジェクション	81	-
ディレクトリトラバーサル	268	-
クロスサイトスクリプティング	239	-
その他	2	-

図 9 iLogScanner による分析例

4.2 比較における考察

既存ツールでは、明確に不正アクセスの数をカウントする。良い意味では定量的に判定されるので、Web 管理者は安心できる。しかしながら、カウントできたものがすべてとは限らない。例えば、既存ツールで検出した不正アクセスは 600 件であるが、これを提案方式の計数方法で数えると 17 件になる。この数は全体の 0.1%ほどである。提案インタフェースの分類では約 1,300 件の不正アクセスがあるため、これらの扱いをどのように理解するかである(表 1)。提案方式における誤検知とするか、既存ツールにおける検知漏れとするかである。

既存ツールはあらかじめ検知する事象を定義した上で計数している。そのため、その他の事象に対しては何も触れていないため、大胆に判断すると正常なアクセスという分類になる。一方、提案インタフェースは、すべてをプロットするので、一定の閾値でいずれかに分類することになる。全体の動向を表示し、その中からユーザの感性に従って、意味づけをしてゆくものである。両者の特徴の違いを表 2 に示す。

今回、Shellshock の攻撃手法が発覚してすぐに該当の攻撃が発生していることが観測できた。このことから未定義のアクセスを正規のアクセスとして位置付けてしまうことは危険であることがわかる。この場合、Web サーバの管理者は正確な理由はわからなくても不正アクセスが来ているということを把握する必要があり、このような目的のために提案インタフェースを活用することは有益と考えられる。

表 1 検出量の比較

種類	既存方式	提案方式
正規のアクセス	99.9%	40.0%
クローラによるアクセス	—	6.1%
判定不明のアクセス	—	49.4%
不正なアクセス	0.1%	4.5%

表 2 既存ツールとの比較

	既存ツール	提案インタフェース
把握方法	定量的	定性的
計数方法	リクエスト単位	日付×アドレス単位
確認方法	数値	2 次元
分析対象	特定の不正アクセス	すべて

4.3 既存ツールとの組み合わせ

既存ツールも提案インタフェースも、Web サーバの管理者が Web サーバの現状を知るための仕組みであり、同じ目的で作られている。そこで、両者を組み合わせ、既存ツールで検出した不正アクセスについて提案インタフェースによる可視化を試行した。

その結果、OS コマンドインジェクションやディレクトリトラバーサル、その他に関しては特に目立った傾向はみられなかった(図 11)。SQL インジェクションとクロスサイトスクリプティングは、既存ツールにおける不正アクセス数

Vol.2015-GN-93 No.39
Vol.2015-CDS-12 No.39
Vol.2015-DCC-9 No.39
2015/1/27

はかなりの数であるが、提案インタフェースで見るとそれぞれ 1 件であり、実際は特定の日に特定のアドレスから大量の不正アクセスがあったものである。さらには該当の攻撃者は SQL インジェクション、クロスサイトスクリプティングその他の手法も組み合わせているため、1 回の不正アクセスで実施されたものであった(図 12)。

以上のことから、機械的な抽出結果について事象の起こった状況を把握し、脅威に対する理解を深めることに提案インタフェースが有効であることを確認できた。

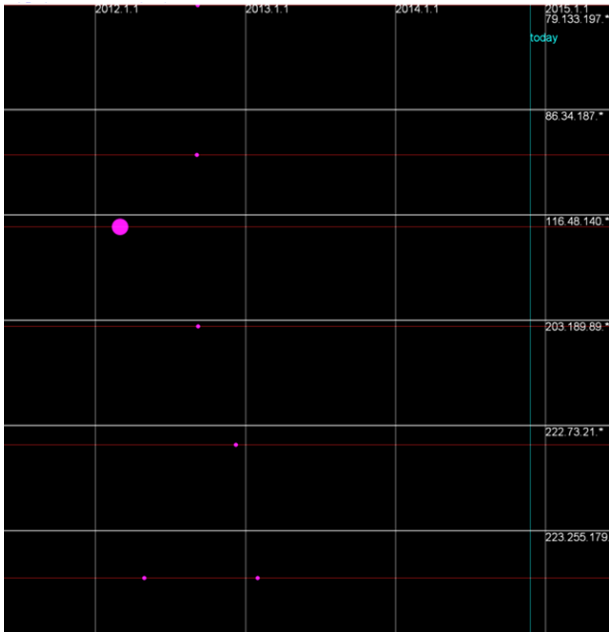

図 11 検知したディレクトリトラバーサルの例

図 12 検知したクロスサイトスクリプティングの例

5. 考察

機械的な分析では、定量的であり一見各自なデータ分析に見える。しかしながら、そもそもの不正アクセスが、正規のアクセスに紛れてくることから、機械的な分析の精度は疑わしい事象がある。一方で、初めから定性的にログ分析をするには、IP アドレス空間は膨大なため、鳥瞰から詳細化してゆくと殆どの分析が的外れになる。そこで、両者を適度におりまぜて確認してゆくことが望ましい。

提案インタフェースの想定利用シーンの一つは、Web サーバのアクセスログを分析し、脅威の傾向をつかみアクセス制限などのセキュア化を実施する場合である。

もう一つは、ネットワークの運用において調べたいアドレスが発生した際に、普段観測されているアドレスと比較することで判断する場合である。通常このようなシーンでは、外部の検査するサイト[10]を利用して、個別のアドレスの危険性を判断することが多い。しかしながら、これらの判定は、攻撃者が観測者を攻撃したことによる観測結果であり、自ネットワークが直面している不審なアクセスの脅威を判定できるとは限らない。そこで、普段の観測状況や世の中の動向と比較して、調べたいアクセスの脅威性を判定するような仕組みが必要になる。それには、提案方式のようなアクセスログ全体を時系列で鳥瞰できる仕組みの他に、図 5 で shellshock の発覚日をプロットしたように、普段から観測される知見を逐次反映できる仕組みが今後必要になる。

6. おわりに

正規のアクセスと不正なアクセスが混在するアクセスログを時系列のアドレス空間上にマッピングし、注目するべきアクセスログを抽出したのちに、アクセスログを観察することで、Web サーバ管理者が自サーバの状況を確認するインタフェースを提案した。ここに、実際の Web サーバのログの投入と、例題として既存ツール iLogScanner との比較と組み合わせを行った。今回の可視化ツールは小規模な Web サーバのログを用いて検証をした。今後の課題として、いくつかの Web サーバでの検証などを通して、提案方式の有効性や拡張性を検討する。

参考文献

1) Heartbleed(オンライン),入手先 http://www.symantec.com/content/ja/jp/enterprise/images/outbreak/Heartbleed_vulnerability.pdf(参照 2014-11-20).
2) Shellshock(オンライン),入手先 http://www.trendmicro.co.jp/cloud-content/jp/pdfs/security-intelligence/vulnerability/shellshock/techbrief-bash-20141003.pdf(参照 2014-11-20).
3) ウェブの安全性を高める(オンライン),入手先 http://www.google.com/transparencyreport/safebrowsing/?hl=ja(参照 2014-11-20).
4) ウェブサイトが改ざんされないように対策を！(オンライン),入手先 http://www.ipa.go.jp/files/000029085.pdf (参照 2014-11-20).
5) 止まらないウェブ改ざん！(オンライン),入手先 http://www.ipa.go.jp/files/000031486.pdf (参照 2014-11-20).
6) ウェブサイトの攻撃兆候検出ツール iLogScanner(オンライン),入手先 http://www.ipa.go.jp/security/vuln/iLogScanner/index.html(参照 2014-11-20)
7) 斉藤,”インターネット空間の汚れ具合を観察するインタフェースの提案”, 情報処理学会研究報告 Vol.2013-GN-89 No.29, 2013.
8) nicter web(オンライン),入手先 http://www.nicter.jp/nw_public/scripts/ (参照 2014-11-20).
9) 対サイバー攻撃アラートシステム”DAEDALUS”の外部展開を開始！(オンライン),入手先 http://www.nict.go.jp/press/2012/06/06-1.html(参照 2014-12-2).
10) IP Address Blacklist Checker Tool(オンライン),入手先 http://www.ipvoid.com/ (参照 2014-11-27).