2ZC-06

サーバアクセスログの分析支援技術

小坂 暁洋 †　岡本 秀輔 ‡　坂本 真仁 §
成蹊大学 †　成蹊大学 ‡　金沢工業大学 §

概要

インターネットの発展に伴い Web サーバへのアクセスは多様化している。そのため、ログに記録される情報も増加している。しかし、ログのフォーマットは変化がなく、多様化するアクセスに対応ができないことが多い。

この発表では Web アクセスの分析支援の一つの方法を提案する。アクセスログの各行から特徴を取得後にラベルを付け、ラベルを条件別に使用することでログを絞り込みを可能とする。その後、CVE(Common Vulnerabilities and Exposures)[1] の提供するデータベースに検索をかけて Web サーバの脆弱性を探すアクセスを試みているログのみを出力する。評価方法としては、効率よく検索が行われているかを実行時間と絞り込み条件の点で評価する。

1　はじめに

Web サーバには、日々ログの書き込みが増えている。実際に、2020 年 7 月 1 日から 7 月 18 日の間には、282,128 件のログが大学研究室のサーバに記録されている。また、ログには IP アドレス・時間帯・HTTP 要求など多くの点に注目する必要がある。さらに複数行のログをまとめて見ることで、規則的なアクセスパターンが見出されることもあるため、注目対象となる。

本研究では、ログの特徴抽出として 1 行に書き込まれた情報のみを使うだけではなく複数行のログを 1 つのグループとして扱い、ログの絞り込みを行う。

2　ログの特徴抽出

本研究のアクセスログの特徴抽出においては、以下の 3 つの方法を使用した。

- ログに書き込まれている情報を取得
- IP アドレス発信元の国を取得
- HTTP 要求を解析

ログに書き込まれている情報は、そのまま特徴として抽出を行う。また、曜日を時間帯から特定を行い特徴付けした。

2 つ目の IP アドレス発信元の特定に関しては、GeolocationAPI[2] を使用する。この API では、IP アドレスの入力で位置情報を取得できる。

3 つ目の HTTP 要求の解析においては、HTTP メソッド・リクエストパス・HTTP バージョンに分けて扱う。ここで、記録された時刻を Unix 時間に変換し、同一 IP アドレスの際に次のアクセスまでの時間が x 秒以内の場合をグループ化してセッションとして扱う。同一 IP アドレスにおいても、次のアクセスまでの時間が x 秒を超えた場合は別のセッションとした。このセッションは、アクセスログにおいて複数行による規則性を持つ HTTP 要求を見つけるために使用する。

A Support Technology to Analyze Server Access Log

† Akihiro Kosaka, Seikei University
‡ Shusuke Okamoto, Seikei University
§ Shinji Sakamoto, Kanazawa Institute of Technology

3 脆弱性に関連しそうなログの出力

脆弱性に繋がるログの確認は、絞り込んだログのリクエストパスを CVE の外部データベースに検索をかけて判断を行う。ここではリクエストパスの文字を分解し、始めのディレクトリの文字の検索から始める。その後、CVE が提供するデータベースから検索文字列に関連する CVE 識別子と CVSS[3] の評価値を取得する。

この CVSS の値は攻撃元の区分や複雑さなどを評価した値であるが、脆弱性の深刻度の基本評価基準のレベル分けされた値を使用する。この値は影響度・攻撃容易性・基本値の 3 つの基本評価基準によって決められる。また、この値は 0 から 10 までの値に分けられ、7.0 ～ 8.9 の値は深刻度が重要、9.0 ～ 10.0 が深刻度が緊急である。検索した文字列に関連する CVE 識別子の基本評価基準の値が一度でも 7.0 以上表示されたものは脆弱性をつく可能性があるログの出力とする。

2021 年 11 月 1 日から 28 日までの 53,367 件のログを絞り込んだ例を表 1 に示す。これは、表の上部から段階的な絞り込みを行っている。

TIME ラベルは、指定した時間を除外する絞り込みであり 2 行目の 2 つの数字の範囲内の時間を除外したため、この場では 0:00 から 5:59 までの 6,668 件となった。PATH ラベルは、指定した文字列を含むリクエストパスを除外するものであり、研究室のサーバへのアクセスの際に使われる study のパスを除外した。同様に-(ハイフン) の除外も行った。SESSION ラベルは、直前の結果から同一 IP アドレス下において、次のアクセスまでの時間が 10 秒以内のセッションを作り 3 行以上ある場合にセッション数と総ログ数を出力した。最後に、CVE ラベルによってセッション内の全ての分解したリクエストパスを検索にかけ既知の脆弱性と関連があると予想されるものを基本評価基準から出力した。

表 1　ログを絞り込む流れ

ラベルと指定	検索結果
TIME	
6 24(6 時から 24 時を除外)	結果:6668(12.489%)
PATH	
/study/	結果:685(1.28%)
PATH	
-(ハイフン)	結果:460(0.86%)
SESSION	セッション数:26 結果:410(0.77%)
CVE (基本評価基準が 7.0 以上)	結果:33(0.06%)

4 まとめ

膨大なログデータの中から Web サーバの脆弱性に繋がるログを見つけることで、アクセスログの分析支援を行う際のログの扱い方を提案した。事前にログを特徴別に分類を行い、その後既知の脆弱性をつくアクセスログとの比較を行う方法である。

今後は、ログの絞り込み条件を組み合わせることで、脆弱性に関連するログを見つけるまでの効率について実行時間から評価を行う。

参考文献

[1] CVE - The MITRE Corporation https://cve.mitre.org/ (Accessed on 26 Dec, 2021)

[2] Geolocation API - W3 https://www.w3.org/TR/geolocation/ (Accessed on 26 Dec, 2021)

[3] Common Vulnerability Scoring System SIG - FIRST.org https://www.first.org/cvss/ (Accessed on 26 Dec, 2021)