DNS over HTTPS によるオープンリゾルバにおいて Slow HTTP DoS 攻撃のトラフィックを遮断し正規のトラフィックを通過させるための最適な閾値の分析と評価

笹川 尋翔

2023 年 1 月 10 日

概要

ユーザのプライバシーの保護やセキュリティの強化のために DNS over HTTPS(DoH) の普及が図られている．そのためには，HTTP/HTTPS の通信を悪用した攻撃に対する対策が必要である．その攻撃の 1 つである Slow HTTP DoS 攻撃は，少ないパケットで攻撃を行うことにより，トラフィックがほとんど増加しないため対策が難しい．そこで，DoH のデータセットを用いて Slow HTTP DoS 攻撃のトラフィックを遮断し，正規のトラフィックを通過させるための閾値を分析する．その後に，閾値に基づいてトラフィックを遮断するシステムを実際に実装して評価する．

1 はじめに

UDP の 53 番ポートで送受信される DNS トラフィックは平文であるため，悪意のある第三者が盗聴したり改ざんしたりすることができる．この問題の解決策の 1 つとして，DNS トラフィックを暗号化する技術である DNS over HTTPS(DoH) の普及が図られている．

DoH は，広く普及している仕組みである HTTPS を利用するため，多くのシステムに組み込みやすく，ファイアウォールの設定の変更が不要であることが多いという利点がある．その一方で，HTTP/HTTPS を用いた攻撃を受ける可能性があるという欠点も存在する．その攻撃の 1 つである Slow HTTP DoS 攻撃は，大量のコネクションを確立し，HTTP のコネクションが切断されない程度の少ないパケットを送信し続けることで Web サーバに負荷を掛ける攻撃である．Slow HTTP DoS 攻撃は，前述の攻撃に比べてトラフィックの増加が少ないため，対処が比較的難しいという特徴がある．また，パブリック DNS のようなオープンリゾルバでは，アクセス元の IP アドレスによる制限を掛けることができないため，タイムアウトの時間を短縮するなどの他の手段で対処する必要がある．

2 関連研究

2021 年に公開された，DoH トラフィックを含まない HTTP/HTTPS 通信のデータセットを用いた Slow HTTP DoS 攻撃の分類の研究 [1] では，Muraleedharan N らがディープラーニングによる多クラス分類を行い，99.61%の正解率で Slow HTTP DoS 攻撃のトラフィックを検出するモデルを構築した．

3 研究内容

初めに，2022 年に公開された実世界の DoH のデータセット [2] を用いて，Slow HTTP DoS 攻撃のトラフィックと正規のトラフィックを分類するための閾値を分析する．Slow HTTP DoS 攻撃のトラフィックの特徴としては以下がある．

- 単位時間当たりに送信されるパケットのサイズが非常に小さい
- TCP ウィンドウのサイズが非常に小さい
- HTTP リクエストボディのサイズが非常に大きい
- 同じ IP アドレスからの同時セッション数が非常に大きい

そのため，これらの特徴を持つグループとそれ以外のグループに，データを分類する必要がある．データセットのサイズは 100GB 以上あるが，そのデータセットに Slow HTTP DoS 攻撃のトラフィックが含まれていない，あるいは 2 クラスに分類する上で不十分なトラフィックしか含まれていない可能性がある．その場合は，自分でフルサービスリゾルバを構築し，そのフルサービスリゾルバに対して Slow HTTP DoS 攻撃を行うことで，不足しているトラフィックを収集する．

次に，Bind と dnscrypt-proxy を用いて DoH を実装する．この際に，先ほどの分析で求めた閾値を Bind の設定ファイルに書き込む．dnscrypt-proxy が動作しているクライアントで Slow HTTP DoS 攻撃を行い，Bind のオープンリゾルバがトラフィックをどの程度正確に遮断するかを検証する．

3.1 Bind と dnscrypt-proxy による DoH の実装

Bind は権威 DNS サーバやフルサービスリゾルバとして動作し，dnscrypt-proxy はスタブリゾルバとして動作する．また，Bind と dnscrypt-proxy のどちらも，DoH に対応している．これらを用いて，オープンリゾルバとスタブリゾルバを構築することで，DoH を実装する．

3.2 Wireshark によるトラフィックの収集，ファイルの結合，ファイル形式の変換

1 つのデータセットを作成して機械学習の入力データとして用いるために，以下の手順でファイルを加工する．

1. tshark コマンドを用いてトラフィックを収集する
2. mergecap コマンドを用いて複数の pcapng ファイルを 1 つのファイルに結合する
3. thsark コマンドを用いて csv ファイルに変換する

3.3 不均衡データの対処，分類手法の選択

正規のトラフィックと比べて Slow HTTP DoS 攻撃のトラフィックは非常に少ないため，ここまでは少数派である正規のトラフィックの特徴が閾値にあまり反映されない．また，正規のトラフィックが非常に多いため，モデルの訓練に余分な時間がかかる．そのため，データの重み付け，アンダーサンプリング，オーバーサンプリングなどを組み合わせてこれらの問題を修正する．

分類手法に関しては，ロジスティック回帰分析，k 近傍法，サポートベクターマシン，決定木は，データの散らばりが大きい場合でも比較的高い精度で分類できる．この研究では，モデルの説明のしやすさと実装のしやすさを重視し，決定木のアンサンブル学習の手法であるランダムフォレストと勾配ブースティング決定木を用いる．

4 評価方法

分類の性能を評価するために用いる指標を決める際には，偽陽性率と偽陰性率の中からどの比率を重視するかを考慮する．Slow HTTP DoS のデータが少ないため，評価指標として適合率や再現率を用いる．それに加えて，予測の確実性を算出するために LogLoss を用い，分類の正確さを算出するために正解率を用いる．

実装したオープンリゾルバの評価には，Slow HTTP Dos 攻撃のトラフィックを遮断した割合を用いる．

5 研究結果

Slow HTTP DoS 攻撃にはいくつかの種類があるため，それらをまとめた概念である Slow HTTP DoS 攻撃のトラフィックと正規のトラフィックの 2 クラス分類では，モデルの性能が低くなる可能性がある．その場合には，その攻撃の種類ごとにクラスを作成してモデルを訓練することで，モデルの性能を向上させることを目指す．

6 まとめ

機械学習によりリアルタイムにトラフィックを分析して不正なトラフィックを遮断するシステムと比べ，安価に実装できるようになることが期待される．既存のソフトウェアの設定を変更するだけで Slow HTTPS DoS 攻撃に対処できるため，簡単に導入できるようになる．

参考文献

[1]Muraleedharan N and Janet B. “A deep learning based HTTP slow DoS classification approach using flow data”. In: *ICT Express* 7 (2021). DOI: https://doi.org/10.1016/j.icte.2020.08.005. URL: https://www.sciencedirect.com/science/article/pii/S2405959520300965 (visited on 2022-01-08).

[2]Kamil Jeřábek et al. “Collection of datasets with DNS over HTTPS traffic”. In: *Data in Brief* 42 (2022). DOI: https://doi.org/10.1016/j.dib.2022.108310. URL: https://www.sciencedirect.com/science/article/pii/S2352340922005121 (visited on 2022-01-07).