フォームの入力制限に対して長さを最適化した敵対的XSS に関する考察

A Study on Adversarial XSS Optimized for Form Input Length Constraints

四方隆之介・ネットワーク分科会
情報セキュリティ大学院大学

Abstract:As deep learning-based XSS detection advances, adversarial generation techniques are also evolving but often yield excessively long payloads. We propose a post-processing optimization to remove redundant tags. Experiments across CNN, MLP, and LSTM models achieved an average 62% length reduction while retaining an average 86.2% attack validity.

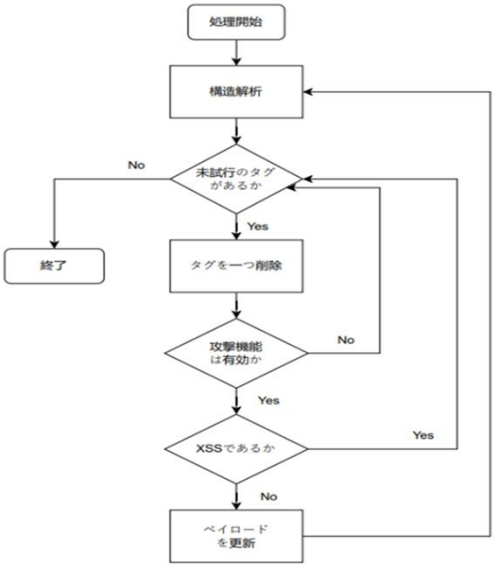
背景

- 現状：敵対的XSS生成技術の発展。
- 課題：ペイロードの長大化によりWAF等のリクエスト容量制限に抵触し、攻撃が実行される前に遮断される可能性が高い。
- 目的：冗長なタグを削除する事後最適化。

提案手法

既存の生成手法に対する事後的な最適化処理を提案。

1. 貪欲法を用い、外側のタグから削除を試行。
2. タグの欠損による不発を防ぐため。
攻撃機能の検証を行う。
3. 終了したペイロードに対して、攻撃の有効性を確認する。



実験と評価

- 攻撃対象モデルはCNN,MLP,LSTM。
- 4,487件のXSS敵対的サンプルの内、508件のペイロード長削減に成功。約62%の削減率。

モデル	短縮成功数 (件)	平均ペイロード長 (byte) 短縮前	平均ペイロード長 (byte) 短縮後	平均削減率
LSTM	310	1,246	449	64%
CNN	93	1,178	471	60%
MLP	105	1,189	487	59%
全体	508	1,222	461	62%

- 508件に対して、攻撃有効性を検証した結果、438件が短縮後もXSSとして成立していることが確認できた。

モデル	短縮成功数	XSS 有効数	攻撃有効率
CNN	93	72	77.4%
MLP	105	75	71.4%
LSTM	310	291	93.9%
全体	**508**	**438**	**86.2%**

短縮前（Original Payload）

http://example.com/
<h1><h1><h1><marquee><marquee><script>
<script>alert(1)</script></script>
<h1><h1><h1>

最適化

短縮後
http://example.com/<h1><script>alert(1)</script><h1>

今後の課題

- 本研究では事後的な短縮手法を提案したが、今後は強化学習の報酬関数にペイロード長へのペナルティを設計することで、生成段階から長さを抑制し、より効率的な探索を実現することが課題である。