Classificação de requisições HTTP maliciosas por meio de aprendizagem de máquina

Aluno: Diego Zurita
Orientador: Prof. Dr. Daniel Macêdo Batista

Abril 2021

Com a variedade de serviços oferecidos pela internet, a quantidade de requisições que são feitas a um servidor web aumenta. De acordo com [1], o Facebook em 2014 recebia uma quatidade de consultas da ordem de bilhões por segundo. Nesse sentido é esperado que também aumente o número de requisições maliciosas, como por exemplo SQL injection que está no top 10 da OWASP [2], uma organização sem fins lucrativos que visa melhorar a segurança em softwares, através da publicação de artigos, metodologias e documentação de maneira gratuita.

Em tal cenário se faz necessário um sistema automático de detecção de requisições maliciosas, pois analisar cada uma delas manualmente pode ser inviável ou bastante custoso. Uma possível solução é criar modelos de aprendizagem de máquina para encontrar padrões de requisições maliciosas e então tomar decisões automatizadas com base no resultado desses modelos.

Além disso, é desejável que esse projeto tenha um custo baixo em termos financeiros e em termos de consumo de energia. Para isso, se tem proposto o uso de hardware de baixo custo para implantação desse tipo de sistema. Em [3] há uma análise que conclui que é possível usar Raspberry Pi como nó de um cluster de processamento de dados.

Este trabalho de conclusão de curso destina-se a construir modelos de aprendizagem de máquina para classificar requisições maliciosas. Tais modelos utilizarão principalmente logs de servidores HTTP para detectar potenciais ataques de SQL Injection, XSS Refletido e XSS Persistido. E por fim, rodar esses modelos em Raspberry Pi de maneira distribuída.

Portanto, este trabalho pretende avançar o que foi realizado no TCC "Análise de Desempenho de Computadores de Baixo Custo em um Sistema de Detecção de Intrusão" de Lucas Seiki Oshiro em 2019 [4]. O avanço será no sentido da construção dos modelos de aprendizagem de máquina utilizando as características ali encontradas, como também novas que venham a ser descobertas.

Referências

[1] Janet Wiener, e Nathan Bronson. Facebook's Top Open Data Problems, 2014.
https://research.fb.com/blog/2014/10/facebook-s-top-open-data-problems/

[2] Andrew van der Stock, Brian Glas, Neil Smithline, e Torsten Gigler. OWASP Top 10, 2017.
https://owasp.org/www-project-top-ten/2017/Top_10.html

[3] Luscas Seiki Oshiro, e Daniel Macêdo Batista. Análise Preliminar de Detecção de Ataques Ofuscados e do Uso de Hardware de Baixo Custo em um Sistema para Detecção de Ameaças, 2019.
https://sol.sbc.org.br/index.php/sbrc_estendido/article/view/7792/7666

[4] Lucas Seiki Oshiro. Análise de Desempenho de Computadores de Baixo Custo em um Sistema de Detecção de Intrusão, 2019.
https://linux.ime.usp.br/~lucasoshiro/assets/pdf/mac0499/monografia.pdf