Uma proposta para Federated Learning em Cibersegurança

Bruno H. Meyer[1], **Aurora Pozo**[1], **Michele Nogueira**[2], **Wagner M. Nunan Zola**[1]

[1]Departamento de Informática – Universidade Federal Paraná (UFPR)

[2]Depto. de Ciência da Computação – Universidade Federal de Minas Gerais (UFMG)

Resumo. *Neste artigo, propõem-se dois métodos de votação para Ensemble Learning em Federated Learning e Cibersegurança. As propostas consideram características em dados de cibersegurança e aumentam a eficiência de sistemas que classificam tráfego de rede.*

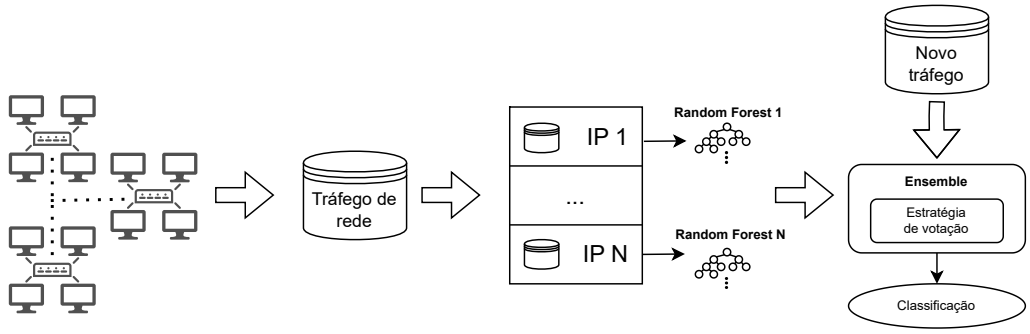
1. Introdução

Federated Learning (FL) é uma técnica de inteligência artificial emergente que tem sido explorada para resolver problemas de cibersegurança [Ghimire and Rawat 2022] usando dados e processamento distribuído. Um problema de FL consiste em dados distribuídos entre vários nós denominados clientes que devem criar um modelo que aprende com seus dados. Em seguida, apenas os modelos são enviados para um servidor onde são agregados para compor um modelo global robusto que será usado para tarefas como classificação, agrupamento, entre outras tarefas presentes no contexto de inteligência artificial. Dentre os trabalhos anteriores que usam FL para cibersegurança, é comum a aplicação do algoritmo FedAvg ou variações do mesmo. O FedAvg consiste em utilizar um modelo de aprendizado paramétrico que é treinado por várias iterações denominadas *rounds*. Em um *round*, os modelos locais são agregados no servidor que devolve o modelo global aos clientes. Por fim, o modelo global é enviado aos clientes para substituir seus modelos locais e continuar o treinamento em seus dados em *rounds* posteriores. Atualmente existem técnicas alternativas de FL como o *One-Shot* FL (OSFL) [Zhou et al. 2020] que consiste em um *framework* que gera o modelo global final em apenas um *round*, e é mais adequado em sistemas distribuídos que possuem dispositivos com capacidades computacionais heterogêneas. Contudo, o modelo originalmente proposto para OSFL utiliza apenas um modelo de destilação de redes neurais para agregar os modelos, e é um consenso entre as pesquisas que modelos clássicos de aprendizado de máquina ainda não são efetivos em cibersegurança [Khraisat et al. 2019]. Neste trabalho é proposta uma estratégia de OSFL usando ensemble de classificadores que não depende do uso de redes neurais. A estratégia é focada em características observadas em problemas de cibersegurança para detecção de ataques de segurança, e usa algoritmos de Ensemble para combinar o resultados de diversos classificadores usando duas novas estratégias de votação (FLENV e FLEWNV).

2. Proposta

A Figura 1 ilustra o *framework* proposto onde serão avaliadas os métodos descritos na proposta a seguir. Serão utilizadas bases de dados conhecidas como TonIoT e BotIoT que contêm dados de cenários que contém tráfego de rede normal e ataques de cibersegurança. Esses dados podem ser utilizados para treinar modelos que classificam conexões de rede como tráfego normal ou algum tipo de ataque. Para representar fielmente cenários de FL reais, os dados serão particionados por IP, representando um conjunto de redes locais, que

Figura 1. Aplicação de Ensemble learning em FL e Cibersegurança



serão os clientes. Cada cliente irá treinar um modelo Random Forest para classificar seus dados. Esses modelos são enviados para um servidor onde um modelo de Ensemble baseado em votação é usado. Vários modelos dos clientes são agregados para rotular novas conexões de tráfego como normal ou ataque. Um problema em realizar uma votação simples, como uma votação majoritária, é o fato que clientes podem gerar modelos enviesados devido a falta de dados ou a dominância de um ou poucos tipos de tráfego, característica comum em cenários de cibersegurança [Ghimire and Rawat 2022]. Dois novos métodos de votação são propostos neste trabalho para atribuir uma probabilidade a cada classe. Essas probabilidades consideram pesos diferentes para cada cliente e são usadas para gerar um voto final. O primeiro método de votação denominado FLENV (*FL Ensemble Normalized Voting*) é descrito na Equação 1, que normaliza a votação usando o número de classes (tipos de ataques) observadas nos dados de cada cliente. Dessa forma, clientes com pouca diversidade de classes terão um menor peso na votação. O segundo método proposto neste trabalho é o FLEWNV (*FL Ensemble Weighted Normalized Voting*), representado na Equação 2 e que adiciona um fator de peso que representa a quantidade de dados usados no treinamento de cada modelo dos clientes. Nas equações, C_i representa o conjunto de classes observada nos dados do cliente i, V_i o voto (classificação) do modelo do cliente i, $W_{c,i}$ a quantidade de instâncias da classe c nos dados do cliente i, $X_{i,j}$ a instância j usada no treinamento do cliente i e Y_{i_j} o correspondente rótulo a essa instância.

$$\text{FLENV P}_c \;=\; \frac{\mid \{i : V_i = c\} \mid}{\mid \{i : c \in C_i\} \mid} \tag{1}$$

$$\begin{aligned} \text{TCWN}_c \;&=\; \sum_i \sum_c \mid \{X_{i,j} : X_{i,j} \in X_i,\; Y_{i,j} = c\} \mid \\ \text{TWV}_c \;&=\; \sum \mid \{\text{TCW}_{c,i} : \text{V}_i = c\} \mid \qquad \text{FLEWNV P}c \;=\; \frac{\text{TWV}_c}{\text{TCWN}_c} \end{aligned} \tag{2}$$

Referências

Ghimire, B. and Rawat, D. B. (2022). Recent advances on federated learning for cybersecurity and cybersecurity for federated learning for internet of things. *IEEE Internet of Things Journal*.

Khraisat, A., Gondal, I., Vamplew, P., and Kamruzzaman, J. (2019). Survey of intrusion detection systems: techniques, datasets and challenges. *Cybersecurity*, 2(1):1–22.

Zhou, Y., Pu, G., Ma, X., Li, X., and Wu, D. (2020). Distilled one-shot federated learning. *arXiv preprint arXiv:2009.07999*.