Atualização de Modelo baseado em Aumento de Dados e Transferência de Aprendizagem para Detecção de Intrusão em Redes

Pedro Horchulhack[1], **Eduardo K. Viegas**[1], **Altair O. Santin**[1], **Jhonatan Geremias**[1]

[1]Programa de Pós-Graduação em Informática (PPGIa)
Pontifícia Universidade Católica do Paraná (PUCPR)
Curitiba – PR

{pedro.horchulhack,eduardo.viegas,santin,jhonatan}@ppgia.pucpr.br

Resumo. *Neste artigo apresentamos uma abordagem para atualização do modelo de aprendizagem de máquina para detecção de intrusão. Inicialmente, o tráfego de rede é aumentando por uma Redes Adversárias Generativas (GANs). Depois, as atualizações de modelos são realizadas por meio de Transferência de Aprendizagem sobre o conjunto de dados aumentado. O número de instâncias a ser rotuladas e os custos computacionais das atualizações do modelo são diminuídas significativa na proposta. A experimentação foi feita num conjunto de dados de 8TB (tráfego de rede de 1 ano), demonstrando a ineficiência dos trabalhos da literatura para detectar mudanças de comportamento no tráfego na rede. No caso do nosso modelo a taxa de falsos positivos diminuiu em até 18,1% quando atualizações periódicas são realizadas. As atualizações contemplaram somente 2,3% das instâncias, com uma diminuição de 14% no custo computacional.*

Abstract. *In this paper we present an approach for updating the machine learning model for intrusion detection. Initially, the network traffic is augmented by a Generative Adversarial Networks (GANs). Next, model updates are performed by Transfer Learning over the augmented dataset. The number of instances to be labeled and the computational costs of the model updates are decreased significantly in the proposal. The experimentation was done on a dataset of 8TB (1 year network traffic), demonstrating the inefficiency of literature work to detect changes in network traffic behavior. In the case of our model the false positive rate decreased by up to 18.1% when periodic updates are performed. The updates involved only 2.3% of the dataset instances, with a 14% decrease in computational cost.*

1. Introdução

De modo geral, administradores de rede detectam ataques de rede através de sistemas de detecção de intrusão baseados em rede (Network Intrusion Detection Systems, NIDS), implementados fazendo uso de técnicas baseadas em *mau-uso* ou em *comportamento* [Molina-Coronado et al. 2020].

Abordagens baseadas em mal-uso identificam ataques por meio de assinaturas de ataque conhecidos, portanto, apenas são capazes de detectar ameaças previamente conhecidas [Sommer and Paxson 2010]. Abordagens baseadas em comportamento detectam

ataques através da avaliação do comportamento em um evento, sinalizando condutas erradas quando um desvio do comportamento modelado como normal é detectado. Assim, supostamente, esta abordagem é capaz de detectar novos tipos de ataques.

Neste contexto, vários trabalhos foram propostos para a detecção de intrusão por meio de técnicas baseadas em *comportamento*, que, de modo geral, são implementadas fazendo uso de aprendizagem de máquinas (AM) [Molina-Coronado et al. 2020]. Um modelo de AM é obtido através da avaliação de um conjunto de dados de treinamento, composta por uma grande quantidade, normalmente milhões de eventos de tráfego de rede rotulados – com uma classe (normal ou ataque) já associada a cada um [Viegas et al. 2019]. Finalmente, o modelo de AM pode ser implantado em ambiente de produção para a classificação de novos eventos de tráfego de rede.

Apesar da elevada acurácia reportada nos trabalhos relacionados, as técnicas baseadas em *comportamento* dificilmente são implantadas em produção [Horchulhack et al. 2022]. Ambientes em rede apresentam uma grande quantidade de desafios quando comparados com aqueles em que a AM tem sido aplicada com sucesso em outras áreas. O comportamento do tráfego de rede é variável, ao mesmo tempo que evolui ao longo do tempo – situação que pode ser causada pelo provimento de novos serviços de rede, pela descoberta de novos ataques de rede ou mesmo por alterações nos *links de conexão* da rede [Gates and Taylor 2006].

Alterações no tráfego tornam o modelo desatualizado, uma vez que o tráfego de rede avaliado durante o procedimento de construção do modelo já não reflete mais o comportamento do ambiente de produção [Viegas et al. 2019]. Por outro lado, o administrador da rede enfrenta um aumento significativo nas taxas de erro do modelo de AM implantado, o que exige como contramedida a execução de atualizações frequentes, demoradas e dispendiosas de tal modelo [dos Santos et al. 2020].

Técnicas tradicionais de atualização do modelo de AM nos NIDS são ainda uma tarefa desafiadora e negligenciada, exigindo tipicamente o fornecimento de um conjunto de dados de treinamento atualizado e a execução de um processo computacionalmente dispendioso de treinamento do modelo [Molina-Coronado et al. 2020]. O administrador deve primeiro coletar grandes quantidades de tráfego de rede, ao mesmo tempo que rotula cada evento como sendo normal ou ataque. Se esta atividade for manual, rotular cada evento pode exigir várias semanas ou mesmo meses de trabalho [Abreu et al. 2020].

As técnicas tradicionais de atualização do modelo descartam o modelo antigo e constroem um novo desde o início, aumentando ainda mais o tempo necessário e os custos computacionais para realizar tal tarefa. As atualizações de modelos em NIDS continuam a ser uma tarefa negligenciada na literatura. Quase sempre os autores assumem que as atualizações periódicas de modelos podem ser facilmente realizadas ou nem as consideram.

Este trabalho propõe um novo método de atualização de modelos de AM adequado para os NIDS, visando facilitar a atualização do modelo. A proposta é implementada em três etapas. Por primeiro as atualizações de modelos são realizadas através de um mecanismo de janela deslizante, que é regida por uma técnica de amostragem de dados, sendo que o objetivo é diminuir o número de amostras que devem ser rotuladas; ao mesmo tempo que se mantem uma coleta adequada do tráfego da rede ao longo do

tempo. Na segunda etapa aplicamos Redes Adversárias Generativas (Generative Adversarial Networks, GAN) visando o aumento de dados, reconstruindo assim a distribuição original do tráfego da rede sem exigir que seja fornecido um rótulo de evento adicional e facilitando o processo de atualização.

Por último realizamos atualizações de modelos por meio de uma abordagem de Transferência de Aprendizagem, objetivando a diminuição dos custos computacionais, ao mesmo tempo que aproveitamos o conhecimento acumulado no modelo desatualizado. O principal objetivo da nossa proposta é aproveitar o aumento de dados com base na GAN para diminuir o número de amostras que devem ser rotuladas durante as atualizações do modelo, ao mesmo tempo que fazemos uso de Transferência de Aprendizagem para diminuir os custos computacionais da atualização do modelo.

2. Fundamentação

O comportamento do tráfego de rede muda com o tempo, uma situação que pode ser causada pela prestação de novos serviços, descoberta de novos ataques, ou mesmo mudanças nos *links* de comunicação da rede [Horchulhack et al. 2022]. Tais mudanças, afetam o modelo de AM implantado na detecção de tentativas de intrusão [Molina-Coronado et al. 2020]. Assim, o comportamento do ambiente de produção não reflete o comportamento apreendido durante a fase de *treinamento*, tal como representado no conjunto de dados de treinamento. Se fazem necessárias atualizações periódicas do modelo, atividade que devem ser conduzidas pelo administrador de rede, porém está tarefa apresenta dois principais desafios. O primeiro é fornecer um conjunto de dados de treinamento atualizada.

O desafio para administrador de rede é coletar e rotular o tráfego da rede, tarefa que em geral, só possível com a assistência de especialistas, o que implica custos elevados a para efetivação [Viegas et al. 2020]. Posteriormente, deve ser realizado um processo de treinamento do modelo, processo computacionalmente custoso. Contudo, as abordagens atuais da literatura ignoram os desafios relacionados as atualizações de modelos, negligenciando a forma como as mudanças de comportamento no tráfego da rede podem afetar o esquema proposto, e como as atualizações do modelo podem ser realizadas de forma facilitada [Ramos et al. 2021].

3. Trabalhos Relacionados

Nas últimas décadas, foram propostos diversos trabalhos para a detecção de intrusão baseada em rede usando técnicas baseadas em AM [Molina-Coronado et al. 2020]. De modo geral, as abordagens propostas concentram-se em aumentar a acurácia do modelo proposto, não tratando da forma como a sua proposta irá funcionar ao lidar com alterações no comportamento do tráfego de rede à medida que o tempo passa. Por exemplo, Y. Yuan *et al.* [Yuan et al. 2017] faz uso de um conjunto de classificadores para aumentar a precisão em um conjunto de dados de detecção de intrusão amplamente utilizada, ignorando como as atualizações e alterações de modelo no tráfego de rede afetam a sua técnica.

Outra abordagem baseada em um conjunto de classificadores, visando maior precisão, foi proposta por X. Gao *et al.* [Gao et al. 2019], considerando um conjunto de dados estáticos através da aplicação de vários classificadores baseados em árvores; do mesmo

modo, as alterações do comportamento do tráfego na rede e as atualizações do modelo não foram abordadas.

Liang e Ma [Liang and Ma 2021] mencionaram o problema das taxas de detecção de NIDSs decaírem gradualmente com o aparecimento de novos ataques. Os autores abordam o problema adquirindo um conjunto de dados mais recente e retreinando todo o modelo, sem tirar proveito do conhecimento prévio da modelagem, aumentando assim o custo computacional da sua abordagem.

O comportamento não estático do tráfego de rede foi também considerado por Ying-Feng Hsu e Morito Matsuoka [Benaddi et al. 2020]. Os autores criaram uma mudança de comportamento do tráfego de rede através da aplicação de vários lotes do conjunto de dados de detecção de intrusão enquanto avaliavam a sua técnica de acordo com cada lote. Neste caso, as alterações no comportamento do tráfego de rede foram criadas em um cenário controlado, não representando o comportamento dos ambientes do mundo real.

Da mesma forma, a facilidade de atualização de modelos também continua a ser negligenciada na maior parte dos NIDS. N. Martindale *et al.* [Martindale et al. 2020] propôs uma abordagem de detecção de intrusão por aprendizagem de fluxo para baixar os custos computacionais da atualização do modelo. No entanto, os autores não abordam a questão da rotulação dos eventos, e assumindo que o rótulo do tráfego de rede pode ser facilmente solicitado quando necessário. Em contraste, X. Li *et al.* [Li et al. 2021] propõe a aplicação da abordagem de transferência de aprendizagem para facilitar o processo de treinamento do modelo em um ambiente distribuído.

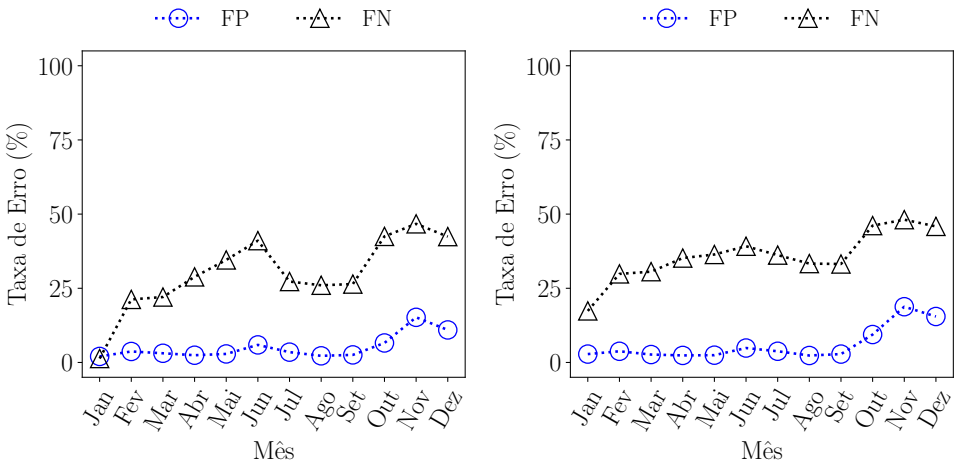
Os autores foram capazes de diminuir os custos computacionais do treinamento de modelos, porém, negligenciam as dificuldades relacionadas com a tarefa de rotulação. Nos últimos anos, as técnicas de aumento de dados têm sido cada vez mais utilizadas para fins de construção de modelos de AM. Por exemplo, U. Otokwala *et al.* [Otokwala et al. 2021] propuseram uma abordagem de aumento de dados para equilibrar a ocorrência de tráfego de rede durante a construção do modelo, aumentando a taxa de detecção, mas desconsiderando a sua aplicação para atualizações do modelo. Um aumento do conjunto de dados baseado em GAN foi proposto por G. Andresini *et al.* [Andresini et al. 2021] para lidar com o desequilíbrio de classes. O modelo proposto melhora a detecção de classe sub-representadas mas não considera o desafio da atualização do modelo.

4. Definição do Problema

Nesta seção são investigadas as mudanças de comportamento no tráfego de rede e os seus impactos no desempenho de classificação das técnicas tradicionais de NIDS baseadas em AM. Mais especificamente, introduzimos primeiramente o conjunto de dados utilizado, depois, avaliamos várias técnicas baseadas em AM em relação a degradação da sua acurácia ao longo do tempo.

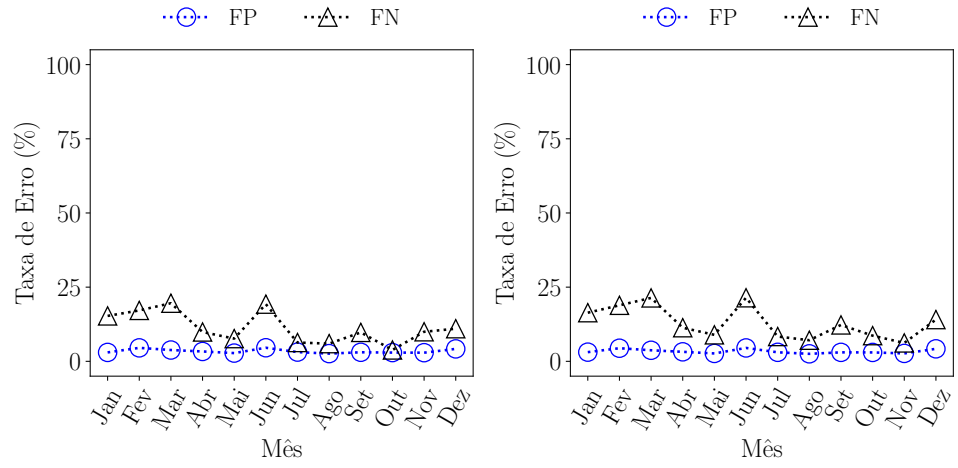
4.1. Conjunto de Dados MAWIFlow

Atualmente, os trabalhos relacionados assumem que o comportamento do tráfego da rede não é alterado à medida que o tempo passa, uma vez que o conjunto de dados utilizado



(a) Random Forest **(b)** Decision Tree

Figura 1. Comportamento das taxas de erro dos modelos de AM ao longo do tempo no ocnjunto de dados *MAWIFlow*. Os classificadores foram treinados com eventos de Janeiro e não foram atualizados ao longo do tempo.



(a) Random Forest **(b)** Decision Tree

Figura 2. Comportamento das taxas de erro dos modelos de AM ao longo do tempo no conjunto de dados *MAWIFlow*. Os classificadores foram treinados todos os meses com eventos de um mês de atraso.

não consideram longos períodos de registro. Como resultado, os modelos propostos construídos sobre tais conjunto de dados são incapazes de avaliar o desempenho das suas propostas quando o comportamento do tráfego da rede se altera.

Este trabalho faz uso do conjunto de dados *MAWIFlow* [Viegas et al. 2019]. O conjunto de dados utilizado foi criado utilizando o Samplepoint-F do arquivo MAWI [MAWI 2021]. Como resultado, foi feito a partir de tráfego de rede real e válido, que foi coletado diariamente em um intervalo de 15 minutos numa conexão intercontinental entre o Japão e os EUA.

Para efeitos de avaliação, foi utilizado todo o tráfego de rede coletado no ano de 2014. O conjunto de dados construído é composto de mais de 8 terabytes de dados, incluindo assim aproximadamente 4 bilhões de fluxos de rede. Para a rotulação prévia dos eventos, aplicamos uma técnica de AM não-supervisionada do MAWILab [Fontugne et al. 2010], que rotula automaticamente os registos de entrada como *normal* ou *ataque*. O MAWILab emprega vários algoritmos de AM não supervisionada para encontrar anomalias nos dados do MAWI, sem assistência humana para a tarefa de rotulação de cada evento. As anomalias encontradas são rotuladas como *ataque*, enquanto os restantes dados são assumidos como eventos *normais*.

4.2. As Mudanças de Comportamento no Tráfego de Rede

Dois classificadores de AM amplamente utilizados foram avaliados, denominados Árvore de Decisão (DT – Decision Tree), e Florestas Aleatórias (RF – Random Forest). O classificador DT foi implementado como métrica de qualidade de divisão do nó através da fórmula de *gini*.

Os classificadores foram implementados através da API *scikit-learn* v0.24. O classificador RF foi implementado com 100 árvores de decisão como seu modelo-base, onde cada uma delas também usa $gini$ como métrica de qualidade de divisão do nó. Uma subamostragem aleatória sem substituição foi utilizada no procedimento de treinamento para balancear a ocorrência entre as classes.

Os classificadores foram avaliados de acordo com as suas taxas de Falso-Positivo (FP) e Falso-Negativo (FN). O FP denota a taxa de instâncias de *normal* incorretamente classificadas como *ataque*, enquanto o FN denota a taxa de instâncias de *ataque* incorretamente classificadas como *normal*.

O primeiro experimento avalia o desempenho da classificação dos classificadores selecionados quando não são realizadas atualizações do modelo ao longo do tempo. Neste caso, treinamos os modelos com os eventos de janeiro, e avaliamos ao longo do tempo sem realizar atualizações. A figura 1 mostra o desempenho da classificação das técnicas selecionadas quando não são realizadas atualizações periódicas dos modelos. Os classificadores selecionados diminuem significativamente o seu desempenho de classificação após o período de treinamento. Por exemplo, o classificador RF aumenta a sua taxa FP em $82,4\%$ apenas um mês após o treino (Figura 1a, *Jan.* vs *Fev.*), ao mesmo tempo que fornece o maior erro em novembro, atingindo $46,7\%$ na taxa de FN.

O segundo experimento avalia o desempenho da classificação das técnicas selecionadas quando são realizadas atualizações periódicas dos modelos. Neste caso, atualizamos os modelos selecionados no início de cada mês com os dados que ocorreram durante os últimos 30 dias. A Figura 2 mostra as taxas de erro dos classificadores selecionados quando se realizam atualizações periódicas do modelo. É possível notar que, ao contrário da abordagem sem atualização, os classificadores selecionados foram capazes de fornecer uma baixa taxa de erro ao longo do tempo. Por exemplo, o classificador RF forneceu uma taxa média de FP de $3,40\%$, enquanto a sua contraparte sem atualização atingiu uma taxa média de FP de $5,08\%$.

Como conclusão se verifica que para manter as taxas de erro baixas, à medida que o tempo passa, é necessário efetuar atualizações periódicas do modelo. Contudo, esta não é uma tarefa trivial nos NIDS, considerando os desafios relacionados com a tarefa de

rotulação de dados, uma vez que muitas vezes requer assistência especializada, exigindo vários dias ou mesmo semanas para ser aplicada. Para permitir a implementação adequada dos NIDS baseados em AM, a literatura deve primeiro abordar a tarefa de atualização do modelo de uma forma aplicável, permitindo ao adminsitrador realizar tal tarefa sem exigir quantidades significativas de dados rotulados.

5. Atualização de Modelo baseado em Aumento de Dados e Transferência de Aprendizagem para Detecção de Intrusão

Estamos propondo um modelo de aumento de dados baseado em Redes Adversárias Generativas (*Generative Adversarial Networks*, GANs), combinado com um esquema de Transferência de Aprendizagem com o objetivo final de facilitar a tarefa de atualização de modelos. A abordagem proposta aborda as atualizações de modelos de três maneiras, como mostra a Figura 3.

Por primeiro diminuímos o número de amostras que devem ser rotuladas; as atualizações de modelos são realizadas considerando um mecanismo de janela deslizante. Neste caso, o conjunto de dados de treinamento é amostrado a partir do tráfego da rede que foi observado antes da execução da tarefa de atualização do modelo, por exemplo 10% dos eventos selecionados aleatoriamente nos últimos 7 dias. O objetivo foi permitir que a quantidade de tráfego de rede utilizada para atualizações de modelos possa ser significativamente reduzida, facilitando assim a tarefa de rotulação do administrador de rede. Posteriormente, para diminuir o impacto causado pela abordagem de amostragem da rede, o esquema proposto baseia-se em uma técnica de aumento de dados através de uma GAN. O objetivo de tal implementação é recriar o comportamento do tráfego de rede, por amostragem, por meio do mecanismo de aumento de dados. Por fim, para diminuir os custos computacionais durante as atualizações do modelo, o conhecimento aprendido no modelo desatualizado é aproveitado nas atualizações sendo feitas, usando um esquema de transferência de aprendizagem.

Nossa técnica é capaz de diminuir significativamente o número de amostras que devem ser rotuladas, devido à janela deslizante de amostragem sendo utilizada. Diminuiu também o impacto na acurácia do modelo causado por tais lacunas, devido ao aumento de dados com base na GAN. Consequentemente, diminuir os custos computacionais, devido à abordagem de transferência de aprendizagem.

As subsecções seguintes descrevem ainda a proposta e os módulos que a implementam.

5.1. Classificação

O modelo proposto assume um esquema tradicional de classificação baseada em AM para NIDS, que faz uso de um modelo de AM com transferência de aprendizagem. O modelo de AM implantado deve ser atualizado regularmente para lidar com as alterações do comportamento do tráfego de rede ao longo do tempo.

O tráfego de rede é coletado em um determinado ambiente monitorado pelo módulo *Coleta de Dados* (Figura 3, *Coleta de Dados*). O comportamento dos dados coletados é extraído pelo módulo de *Extração de Características* que produz um fluxo de rede correspondente em um formato vetor de características. O vetor é classificado como

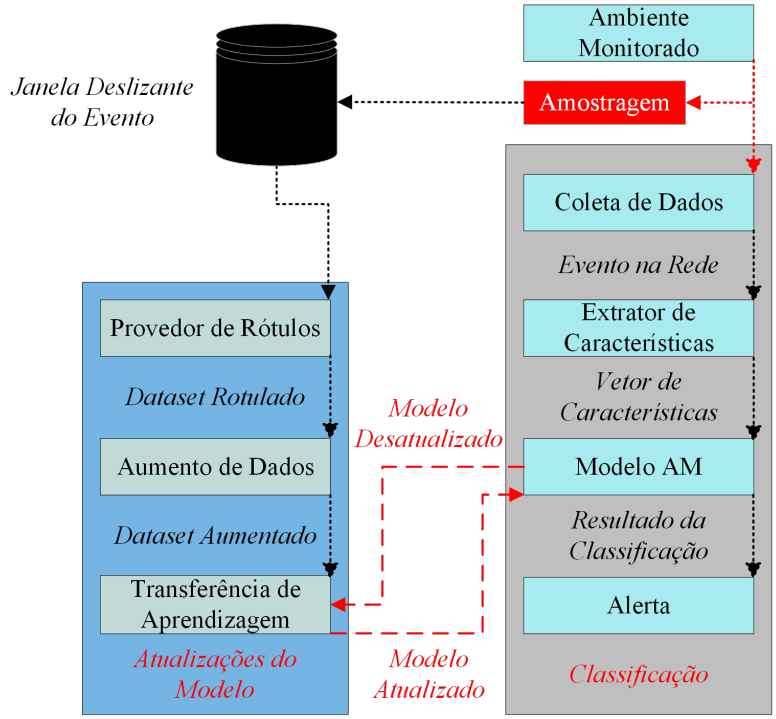


Figura 3. Mecanismo baseado em GAN para aumento de dados e transferência de aprendizagem que visa facilitar a atualização de modelos em NIDS.

normal ou *ataque* por um modelo de AM, que sinaliza condutas erradas ao módulo de *Alerta*.

5.2. Atualizações do Modelo

Como avaliado anteriormente (Seção 4), o comportamento não estático do tráfego de rede exige que sejam realizadas atualizações periódicas do modelo de AM. No entanto, as atualizações de modelos representam um grande desafio nos NIDS baseados em AM devido às dificuldades relacionadas com a tarefa de rotulação de eventos. Para enfrentar tal desafio, o modelo proposto baseia-se em três principais fases, que incluem um mecanismo de amostragem de dados, um esquema de aumento de dados com base em GAN, e uma abordagem de transferência de aprendizagem, como mostrado na Figura 3.

A amostragem de dados diminui o número de amostras que devem ser rotuladas pelo administrador da rede, diminuindo assim os custos de atualização do modelo. O aumento de dados baseado no GAN visa reconstruir a distribuição original do tráfego da rede a partir dos dados amostrados. Finalmente, a transferência de aprendizagem visa diminuir os custos computacionais de atualização do modelo, uma vez que partirá do modelo antigo (desatualizado devido à mudanças) para gerar a versão mais recente do modelo, ao invés de construir o modelo do zero. O esquema proposto é capaz de diminuir significativamente o número de amostras do tráfego de rede que devem ser rotuladas pelo administrador de rede, mantendo ao mesmo tempo a distribuição original do tráfego de rede, e também diminuindo os custos computacionais das atualizações do modelo.

O procedimento de atualização do modelo proposto é executado periodicamente, por exemplo, a cada mês. Neste caso, deve ser fornecida uma janela deslizante de eventos,

construída por amostragem do comportamento do ambiente monitorado (Figura 3, *Janela Deslizante do Evento*). Por exemplo, amostragem aleatória de 10% dos eventos da rede que ocorreram durante os 7 dias que precederam a tarefa de atualização do modelo. O conjunto de dados amostrados é então fornecido ao *Provedor de Rótulos*, que associará um rotulo de modo adequado cada evento fornecido (Figura 3, *Provedor de Rótulos*). Várias técnicas podem ser utilizadas para cumprir tal tarefa, incluindo assistência manual especializada, ou mesmo a aplicação de técnicas de AM não supervisionada.

O conjunto de dados rotulado da amostra é utilizado como entrada na fase de aumento de dados com base na GAN, cujo objetivo é reconstruir a distribuição original do tráfego de rede a partir dos dados amostrados. O conjunto de dados aumentado é utilizado para o procedimento de atualização do modelo. O modelo desatualizado, implantado no ambiente de produção, é recuperado e utilizado para atualizações incrementais do modelo, em uma abordagem de transferência de aprendizagem (Figura 3, *Transferência de Aprendizagem*). Finalmente, o modelo de AM atualizado é reimplementado no ambiente de produção.

6. Avaliação

6.1. Construção do Modelo

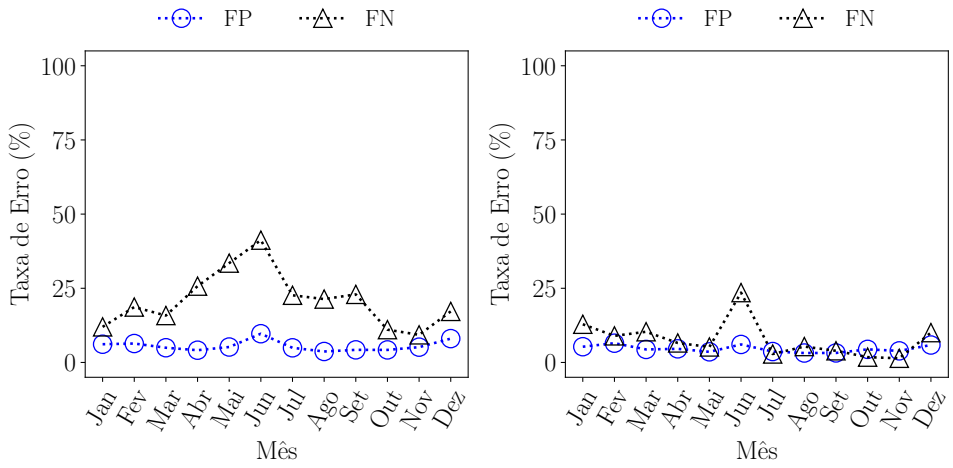
O classificador proposto (Figura 3, *Modelo AM*) foi implementado e avaliado através de uma rede neural multicamada (*Multilayer Perceptron*, MLP) para permitir a aplicação do esquema de aprendizagem de transferência durante as atualizações do modelo. O MLP foi implementado com 40 características de entrada, como fornecido pelo conjunto de características *MAWIFlow*, com 512 neurônios na camada oculta, e 1 neurônio na camada de saída. Os neurônios da camada oculta utilizam uma função de ativação *relu*, o treino faz uso de uma taxa de aprendizagem de $0,001$, otimizador *adam*, e 1.000 épocas.

O MLP foi implementado através da API *scikit-learn* $v0,24$. Para o aumento de dados baseado na GAN o esquema proposto aplicou GAN Tabular Condicional (*Conditional Tabular Generative Adverarial Networks*, CTGAN) [Xu et al. 2019]. Em cada treinamento do modelo, incluindo a atualização inicial e periódica, um novo CTGAN é treinado sobre os dados amostrados para fins de aumento de dados. Neste caso, são gerados 10% de instâncias aumentadas e adicionados ao conjunto de dados de treinamento amostrados (Figura 3, *Dataset Aumentado*).

6.2. Abordando Mudanças no Comportamento no Tráfego da Rede

O primeiro experimento avalia o modelo proposto sem a execução de atualizações periódicas. Para atingir tal objetivo, uma amostra aleatória de 10% dos eventos de tráfego da rede desde a primeira semana de janeiro foi obtida. Os dados amostrados são utilizados como entrada no modelo da GAN para fins de aumento de dados, enquanto o conjunto de dados aumentado é utilizado para realizar a construção do modelo. O modelo não é atualizado ao longo do ano. Por conseguinte, avaliamos como o aumento de dados proposto com base na GAN é capaz de melhorar o procedimento de treinamento do modelo quando são fornecidos significativamente menos dados durante a treinamento do modelo (10% de 7 dias $vs.$ 100% de 30 dias).

A figura 4a mostra o desempenho da classificação do modelo proposto sem a execução de atualizações periódicas. É possível notar que a nossa técnica também é



(a) Sem atualizações **(b)** Com atualizações mensais

Figura 4. Comportamento das taxas de erro do modelo proposto no dataset *MAWIFlow*. Atualizações mensais no modelo são realizadas com uma amostra de 10% dos tráfegos de rede.

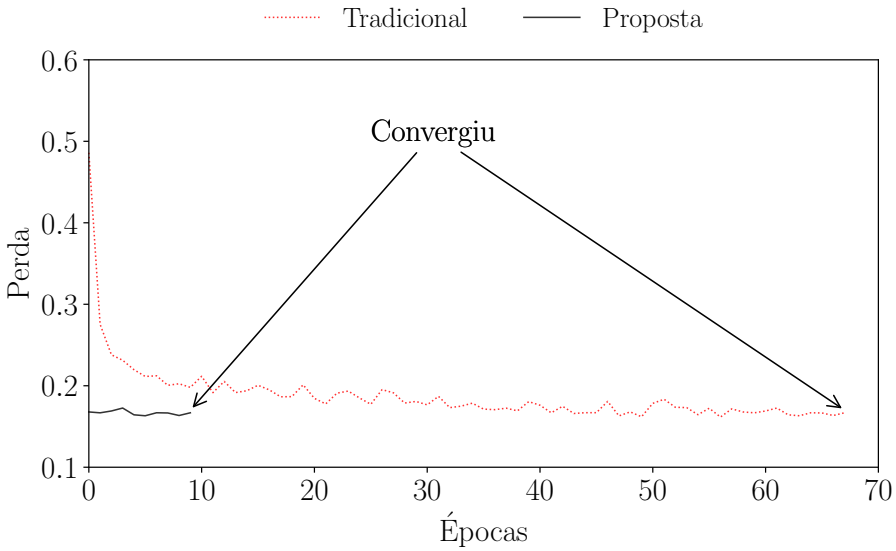
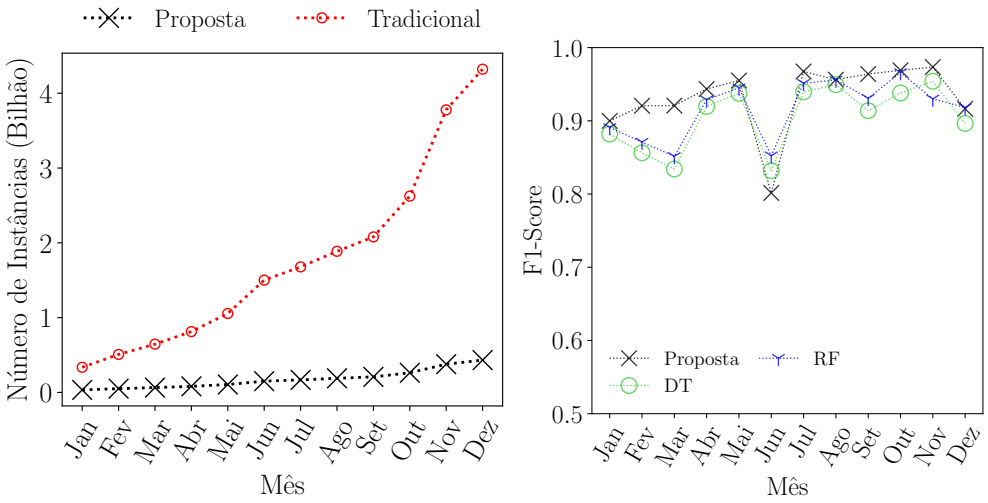


Figura 5. Convergência do modelo no mês de Fevereiro; com e sem a proposta de transferência de aprendizagem.

afetada pelo comportamento do tráfego da rede ao longo do tempo quando não são realizadas atualizações periódicas no modelo. Porém, apesar de terem sido fornecidos apenas 10% do tráfego original da rede, o esquema proposto ainda é capaz de manter as taxas médias de FP semelhantes, com uma diferença de $9,4\%$, quando comparado com a RF sem atualização (Figura 4a $vs.$ Figura 1a). A abordagem proposta de aumento de dados com base na GAN pode ser utilizada para melhorar a vida útil dos esquemas de detecção de intrusão.

O segundo experimento avalia o modelo proposto quando são realizadas atualizações periódicas do modelo. Para atingir tal objetivo, realizamos atualizações mensais do modelo, considerando uma janela deslizante de eventos de 7 dias e uma taxa de



(a) Requisição de rótulo **(b)** F1-Score

Figura 6. Performance das técnicas avaliadas com atualizações mensais dos modelos no dataset *MAWIFlow*.

amostragem de 10%. Lembrando que o aumento de dados com base na GAN é utilizado em cada tarefa de atualização do modelo. A Figura 5 mostra a taxa de convergência do modelo de transferência de aprendizagem proposto quando comparado com a construção do modelo a partir do zero em fevereiro.

A nossa técnica foi capaz de convergir com apenas 10 épocas de treinamento, enquanto a técnica tradicional exige 68 épocas para ser executada. Em média, a aplicação da aprendizagem de transferência durante as atualizações do modelo foi capaz de diminuir os custos computacionais em $85,3\%$. A figura 6 mostra o comportamento do erro do modelo proposto com atualizações mensais. É possível notar que a nossa técnica foi capaz de fornecer taxas de erro significativamente baixas, apresentando uma média de $4,5\%$, e $7,7\%$ das taxas FP e FN, respectivamente; uma diminuição de $18,1\%$ na taxa FP e 63% na taxa de FN quando comparada com a sua contraparte sem atualização.

A nossa técnica foi capaz de facilitar significativamente a tarefa de atualização do modelo, exigindo apenas 14% dos custos computacionais e utilizando apenas $2,3\%$ dos eventos da rede. Isto porque utiliza apenas 10% dos dados amostrados aleatoriamente dos últimos 7 dias, enquanto as técnicas tradicionais foram treinadas com 100% dos últimos 30 dias (Figura 4b vs. Figura 2).

Finalmente, investigamos os benefícios do modelo proposto quando comparado com as técnicas tradicionais. A figura 6a mostra o número cumulativo de amostras de rede que devem ser rotuladas pelas técnicas avaliadas durante as atualizações do modelo. O modelo proposto exige apenas $2,3\%$ de amostras a serem rotuladas quando comparado com as abordagens tradicionais, o que representa uma tarefa de atualização do modelo significativamente mais fácil de realizar. A figura 6b mostra a precisão de classificação do modelo proposto versus as técnicas avaliadas anteriormente (Figura 4b vs Figura 2). As abordagens foram avaliadas de acordo com o F1-Score, calculados como a média harmônica de precisão (*precision*) e revocação (*recall*).

O modelo proposto foi capaz de fornecer resultados semelhantes de F1-Score quando comparado com as abordagens tradicionais, exigindo apenas $2,33\%$ dos eventos de rede rotulados e 14% dos custos computacionais durante as atualizações do modelo.

7. Conclusão

As mudanças de comportamento do tráfego na rede são um desafio conhecido e negligenciado para os NIDS baseados em AM da literatura. Este artigo mostrou que as abordagens atuais na literatura são incapazes de manter a sua acurácia de classificação durante longos períodos, exigindo atualizações de modelos frequentes e custosas computacionalmente.

O modelo proposto foi capaz de facilitar significativamente a tarefa de atualização do modelo através de um mecanismo de janela deslizante, aliado a um mecanismo de aumento de dados baseado em GAN, e a aplicação da transferência de aprendizagem a partir do modelo desatualizado de AM, em uso antes desta atividade. Como trabalho futuro, pretendemos avaliar o esquema proposto por períodos mais longos, ao mesmo tempo que faremos uso de técnicas de aprendizagem profunda para a tarefa de classificação.

Agradecimento

Este trabalho foi parcialmente financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), processo nº 304990/2021-3.

Referências

Abreu, V., Santin, A. O., Viegas, E. K., and Cogo, V. V. (2020). Identity and access management for IoT in smart grid. In *Advanced Information Networking and Applications*, pages 1215–1226. Springer International Publishing.

Andresini, G., Appice, A., Rose, L. D., and Malerba, D. (2021). GAN augmentation to deal with imbalance in imaging-based intrusion detection. *Future Generation Computer Systems*, 123:108–127.

Benaddi, H., Ibrahimi, K., Benslimane, A., and Qadir, J. (2020). A deep reinforcement learning based intrusion detection system (DRL-IDS) for securing wireless sensor networks and internet of things. In *Lecture Notes of the Institute for Computer Sciences*, pages 73–87.

dos Santos, R. R., Viegas, E. K., Santin, A., and Cogo, V. V. (2020). A long-lasting reinforcement learning intrusion detection model. In *Advanced Information Networking and Applications*, pages 1437–1448. Springer International Publishing.

Fontugne, R., Borgnat, P., Abry, P., and Fukuda, K. (2010). MAWILab: Combining diverse anomaly detectors for automated anomaly labeling and performance benchmarking. In *Proc. of the 6th Int. Conf. on emerging Networking EXperiments and Technologies (CoNEXT)*.

Gao, X., Shan, C., Hu, C., Niu, Z., and Liu, Z. (2019). An adaptive ensemble machine learning model for intrusion detection. *IEEE Access*, 7:82512–82521.

Gates, C. and Taylor, C. (2006). Challenging the anomaly detection paradigm: A provocative discussion. In *Proceedings of the 2006 Workshop on New Security Paradigms*, NSPW '06, page 21–29, New York, NY, USA. Association for Computing Machinery.

Horchulhack, P., Viegas, E. K., and Santin, A. O. (2022). Toward feasible machine learning model updates in network-based intrusion detection. *Computer Networks*, 202:108618.

Li, X., Hu, Z., Xu, M., Wang, Y., and Ma, J. (2021). Transfer learning based intrusion detection scheme for internet of vehicles. *Information Sciences*, 547:119–135.

Liang, J. and Ma, M. (2021). Co-maintained database based on blockchain for idss: A lifetime learning framework. *IEEE Transactions on Network and Service Management*, pages 1–1.

Martindale, N., Ismail, M., and Talbert, D. A. (2020). Ensemble-based online machine learning algorithms for network intrusion detection systems using streaming data. *Information*, 11(6):315.

MAWI (2021). MAWI Working Group Traffic Archive - Samplepoint F.

Molina-Coronado, B., Mori, U., Mendiburu, A., and Miguel-Alonso, J. (2020). Survey of network intrusion detection methods from the perspective of the knowledge discovery in databases process. *IEEE Trans. on Network and Service Management*, 17(4):2451–2479.

Otokwala, U., Petrovski, A., and Kalutarage, H. (2021). Improving intrusion detection through training data augmentation. In *International Conference on Security of Information and Networks (SIN)*. IEEE.

Ramos, F., Viegas, E., Santin, A., Horchulhack, P., dos Santos, R. R., and Espindola, A. (2021). A machine learning model for detection of docker-based APP overbooking on kubernetes. In *ICC 2021 - IEEE International Conference on Communications*. IEEE.

Sommer, R. and Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. In *2010 IEEE Symposium on Security and Privacy*, pages 305–316.

Viegas, E., Santin, A., Bessani, A., and Neves, N. (2019). BigFlow: Real-time and reliable anomaly-based intrusion detection for high-speed networks. *Future Generation Computer Systems*, 93:473–485.

Viegas, E. K., Santin, A. O., Cogo, V. V., and Abreu, V. (2020). A reliable semi-supervised intrusion detection model: One year of network traffic anomalies. In *ICC 2020 IEEE Int. Conf. on Communications (ICC)*, pages 1–6.

Xu, L., Skoularidou, M., Cuesta-Infante, A., and Veeramachaneni, K. (2019). Modeling tabular data using conditional gan. In *Advances in Neural Information Processing Systems*.

Yuan, Y., Huo, L., and Hogrefe, D. (2017). Two layers multi-class detection method for network intrusion detection system. In *2017 IEEE Symposium on Computers and Communications (ISCC)*. IEEE.