Otimização do Ataque Grayhole ao Protocolo GOOSE

Guilherme C. Mundt[1], Carlos S. M. Zomer[2], Silvio E. Quinconzes[1], Juliano F. Kazienko[2], Rodrigo Miani[3], Vagner E. Quinconzes[4], Daniel Mossé[5]

[1] Universidade Federal do Pampa (UNIPAMPA)

[2] Universidade Federal de Santa Maria (UFSM)

[3] Universidade Federal de Uberlândia (UFU)

[4] Universidade Federal Fluminense (UFF)

[5] University of Pittsburgh

guilhermemundt.aluno@unipampa.edu.br, carlos.zomer@acad.ufsm.br

Resumo. *Subestações elétricas baseadas na norma IEC-61850 utilizam o protocolo GOOSE para comunicação crítica entre dispositivos de proteção. Ataques do tipo Grayhole — que descartam seletivamente pacotes de rede — ameaçam a confiabilidade dessas redes sem acionar alarmes imediatos. Embora modelos de Machine Learning detectem essas ameaças, sua natureza caixa-preta limita a transparência das decisões. Este artigo apresenta uma análise sistemática do ataque sob taxas de descarte de 3% a 90% e integra técnicas de Inteligência Artificial Explicável (XAI) para interpretar a lógica de detecção em redes IEC-61850. Os resultados reforçam a importância da explicabilidade na segurança de infraestruturas críticas.*

Abstract. *IEC 61850-based electrical substations rely on the GOOSE protocol for critical communication among protection devices. Grayhole attacks—which selectively drop network packets—pose a threat to the reliability of these networks without immediately triggering alarms. Although machine learning models can detect such threats, their black-box nature limits the transparency of the decision-making process. This paper presents a systematic analysis of Grayhole attacks under packet-dropping rates ranging from 3% to 90% and integrates Explainable Artificial Intelligence (XAI) techniques to interpret detection logic in IEC 61850 networks. The results highlight the importance of explainability in enhancing the security of critical infrastructures.*

1. Introdução

A norma IEC-61850 padroniza a troca de mensagens entre dispositivos eletrônicos inteligentes (IEDs) em subestações elétricas, com foco em comunicação de proteção e automação [Baigent et al. 2004] [Gonçalves et al. 2023]. No entanto, estudos anteriores identificaram lacunas na estrutura da norma no que tange à sua resiliência contra ameaças cibernéticas, especialmente no contexto de *smart grids* [Reda et al. 2021] [Lázaro et al. 2021]. Essas vulnerabilidades ressaltam a necessidade de mecanismos de segurança aprimorados para proteger os sistemas de energia modernos contra ataques cibernéticos em constante evolução [McLennan et al. 2022].

As vulnerabilidades identificadas na norma IEC-61850 possibilitam a exposição do sistema a várias ameaças cibernéticas, com ênfase em ataques de negação de serviço

(DoS). Esses ataques têm como objetivo interromper o acesso legítimo aos serviços oferecidos pelo sistema, sejam eles iniciados por usuários ou sistemas automatizados [Ashraf et al. 2021]. Embora os ataques DoS direcionados a redes de comunicação em conformidade com a IEC-61850 tenham sido amplamente modelados na literatura [Quincozes et al. 2023] [Ashraf et al. 2021], ainda há uma lacuna significativa no que tange a ataques de descarte seletivo de mensagens, como o ataque *Grayhole* [Quincozes et al. 2023] [Gonçalves et al. 2023].

Apesar disso, trabalhos [Quincozes et al. 2023] [Gonçalves et al. 2023] [Basumallik 2020] que buscam atuar neste escopo ainda não exploraram totalmente os cenários em que as condições de ataque são desproporcionalmente vantajosas para o atacante, deixando lacunas na compreensão do comportamento de IDSs nesses cenários.

Este trabalho tem como objetivo a realização de uma investigação dos possíveis cenários de ataque *Grayhole*, conforme modelado por [Gonçalves et al. 2023], para identificar o comportamento de IDSs dada a variabilidade dos parâmetros de descarte. O domínio do ataque concentra-se nos protocolos de comunicação entre IEDs, especificamente o protocolo GOOSE (*Generic Object-Oriented Substation Event*). Conjuntos de dados representativos para o protocolo são gerados usando a ferramenta ERENO [Quincozes 2022]. Além disso, este estudo emprega técnicas de Inteligência Artificial Explicável (XAI) para analisar a interpretabilidade das predições de modelos de classificação, elucidando assim os motivos subjacentes da atuação de IDSs na detecção de ataques *Grayhole*.

O presente estudo busca contribuir para a compreensão dos impactos da variância da taxa de descarte no desempenho de modelos de detecção de ataques, bem como para o entendimento do comportamento de sistemas de detecção de intrusão diante do descarte seletivo de mensagens em redes baseadas na norma IEC-61850. Além disso, pretende identificar os atributos de maior relevância para os modelos preditivos em cenários com taxas de descarte estáticas e dinâmicas, assim como evidenciar possíveis lacunas de segurança presentes na comunicação entre dispositivos dessas redes.

2. Fundamentação Teórica

Esta seção estabelece os conceitos fundamentais que sustentam nossa pesquisa. Primeiro, examinamos o ataque *Grayhole*. Posteriormente, se discute o protocolo GOOSE, sob o qual o ataque ocorre, findando com a explanação acerca de XAI.

2.1. Ataque *Grayhole* em subestações

O ataque *Grayhole* consiste em uma forma de negação de serviço baseada no descarte seletivo de pacotes [Quincozes et al. 2023] [Gonçalves et al. 2023]. No contexto deste trabalho, o atacante compromete um IED da rede, geralmente um dispositivo publicador, e passa a descartar mensagens de acordo com critérios predefinidos, comprometendo a comunicação entre os dispositivos da subestação, conforme ilustrado na Figura 1.

Esse comprometimento pode ocorrer por diferentes vetores de infiltração: exploração de vulnerabilidades em interfaces de acesso remoto ou má segmentação de rede (vetor externo), abuso de credenciais legítimas por um agente com acesso físico ou lógico autorizado à subestação (ameaça interna), ou a inserção de firmware malicioso durante a cadeia de fornecimento do dispositivo (ataque de supply chain). Neste trabalho,

assume-se que o atacante já obteve controle sobre o IED publicador, independentemente do vetor utilizado, e a análise se concentra no comportamento da rede a partir desse ponto de comprometimento.

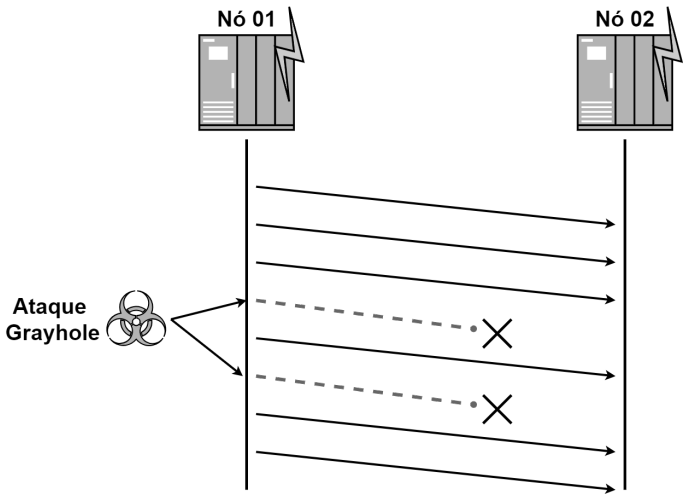


Figura 1. Ataque *Grayhole*

Como visto em [Pal et al. 2016], o objetivo do atacante é causar o máximo de dano possível, neste caso indisponibilizar o serviço, sem ser identificado pelo sistema de detecção de intrusão, sendo nesse contexto dentro das subestações elétricas. Esse tipo de ataque se torna ainda mais eficiente ao levar em conta sua capacidade de escapar da detecção por sistemas IDS, especialmente devido à alta probabilidade de falsos positivos durante situações legítimas de congestionamento de rede e às limitações computacionais dos IEDs em subestações.

2.2. Protocolo GOOSE

O protocolo GOOSE, definido na norma IEC 61850-8-1, fornece comunicação crítica em tempo real para automação de subestações, priorizando baixa latência e comportamento determinístico, sendo a entrega de mensagens dentro de janelas de tempo previsíveis e limitadas. Entretanto, a ausência de mecanismos nativos de segurança o torna suscetível a ataques. A estrutura das mensagens GOOSE tem sua comunicação baseada em um mecanismo determinístico de retransmissão por *backoff* exponencial, que busca garantir a confiabilidade da comunicação sem sobrecarregar a rede [Sidhu and Yin 2007].

O mecanismo de *backoff* exponencial do protocolo gera padrões temporais que podem ser explorados para a detecção de intrusões e ataques baseados em tempo [Quincozes 2022]. Entretanto, a implementação de medidas de segurança deve respeitar restrições rigorosas de latência, inferiores a 3 ms [Ashraf and et al. 2021]. Embora esquemas de autenticação leve tenham sido propostos [Tefek et al. 2022], eles ainda enfrentam dificuldades para atender a todos os requisitos operacionais do protocolo.

2.3. XAI: Inteligência Artificial Explicável

XAI é uma abordagem que utiliza modelos de *Machine Learning* para quantificar a importância de cada atributo de um dataset na geração das previsões de um modelo classificador. Esta abordagem torna-se especialmente necessária quando o domínio em que o

classificador está inserido necessita de alta robustez em suas saídas – como é o caso de classificadores no domínio de saúde ou segurança [Gunning et al. 2019].

A utilização desta abordagem na literatura é limitada no escopo de ataques cibernéticos a subestações elétricas. Entretanto, percebe-se ser importante usá-la para o entendimento das consequências da falta de identificação de ataques na subestação, causando falhas na atuação de relés de proteção, podendo resultar em danos a equipamentos, interrupção do fornecimento de energia ou, em casos de maior severidade, desligamentos em cascata na rede elétrica.

3. Trabalhos Relacionados

No trabalho [Gonçalves et al. 2023], os autores modelam um ataque do tipo *Grayhole* em subestações elétricas utilizando o protocolo GOOSE como domínio de ataque. Para isso, foi implementado um algoritmo responsável por descartar 20% dos pacotes analisados, gerando um *dataset* utilizado na avaliação de diferentes algoritmos de classificação. Os resultados demonstraram que os classificadores foram capazes de identificar o ataque com relativa facilidade, destacando-se o algoritmo *J48* (árvore de decisão C4.5) em termos de desempenho e tempo de processamento. Como limitação, o estudo considera apenas uma única taxa de descarte, restringindo a obtenção de características mais genéricas sobre essa categoria de ataque.

Em [Elbez et al. 2022], se busca uma análise de ataques do tipo DoS em redes que sigam a norma IEC-61850. No desenvolvimento da pesquisa, buscou-se trabalhar ataques de *poisoning*. Os autores desenvolveram um algoritmo que se baseia em uma modelagem matemática aplicada sobre o tráfego de rede em subestações elétricas, chamando o algoritmo de EDA4GNeT. Com a finalização da simulação, percebeu-se que o método obteve ótimos resultados na deteção de ataques DoS, recebendo uma avaliação de detecção de 99.8% em média. Porém, o trabalho possui escopo limitado ao ataque *poisoning*.

Em [Quincozes et al. 2023], são avaliadas técnicas de *Machine Learning* para detecção de intrusões em WSN, considerando os ataques *Flooding*, *Grayhole* e *Blackhole* a partir do *dataset* público WSN-DS. A metodologia consiste na aplicação de diferentes algoritmos de classificação para identificação dos ataques. Os resultados indicam que abordagens supervisionadas superam as não supervisionadas em termos de precisão e tempo de processamento, alcançando um F1-Score de até 93,06% na detecção de ataques *Grayhole*. Como limitação, o estudo não detalha a modelagem dos ataques presentes no dataset, dificultando a avaliação do grau de dificuldade aos mecanismos de detecção.

Em [V and Yadav 2024], os autores analisam vulnerabilidades do protocolo GOOSE da IEC-61850 relacionadas a ataques de manipulação e *Grayhole*. Para mitigá-las, propõem mecanismos de proteção das mensagens e utilizam o classificador k-NN para detectar anomalias decorrentes do descarte de pacotes. Os resultados indicam boa capacidade de detecção com baixo custo computacional. Entretanto, o estudo não considera diferentes taxas de descarte no ataque *Grayhole*, limitando a avaliação da solução em cenários com distintos níveis de severidade.

O trabalho de [Basumallik 2020] propõe uma taxonomia de ataques cibernéticos a sistemas de energia, incluindo ataques Grayhole direcionados a unidades de medição fasorial (PMU), formalizando diferentes categorias de ataque por meio de modelos ma-

temáticos. Como limitação, o estudo tem caráter teórico e não explora cenários práticos de ataque, restringindo a avaliação dos mecanismos de defesa em condições realistas.

Para facilitar a comparação entre os trabalhos relacionados explicitados e a identificação dos pontos cruciais que este trabalho aborda, a Tabela 1 resume os pontos mais importantes nos trabalhos examinados.

Tabela 1. Comparação de trabalhos relacionados.

Referência	Grayhole	IEC-61850	Taxa de Descarte	XAI
([Gonçalves et al. 2023])	✓	✓	✓	✗
[Elbez et al. 2022]	✗	✓	✗	✗
([Quincozes et al. 2023])	✓	✗	✗	✗
[V and Yadav 2024]	✓	✓	✓	✗
[Basumallik 2020]	✓	✓	✗	✗
Este trabalho	✓	✓	✓	✓

4. Materiais e Métodos

Para o desenvolvimento do projeto e geração dos resultados, foram utilizadas diferentes ferramentas e técnicas para avaliação das defesas para ataque *Grayhole*. Nesta seção, descreve-se os materiais usados e métodos aplicados propostos.

4.1. ERENO: *Efficacious Reproducer Engine for Network Operations framework*

O ERENO é um *framework* de geração de datasets para diferentes protocolos existentes na IEC-61850 [Quincozes 2022]. Nesta ferramenta, a geração dos conjuntos de dados é acrescida da inclusão de atributos enriquecidos – valores computados a partir dos atributos originais do protocolo em questão – que auxiliam na construção de classificadores capazes de detectar o ataque sendo simulado. Os atributos enriquecidos incluem, dentre outros: *timestampDiff*, *sqDiff* e *stDiff*.

No escopo deste artigo, o *framework* tem como objetivo gerar um conjunto de datasets[1] do protocolo GOOSE com a inserção do ataque *Grayhole*, possuindo como principal parâmetro uma taxa de descarte definida pelo atacante. Tal lógica é apresentada no Algoritmo 1. Nele, observa-se a função `grayhole(p,td)`, que recebe como parâmetros um pacote GOOSE `p` e uma taxa de descarte `td`; caso o número aleatório `na` seja menor que a taxa de descarte `td` o atacante descartará o pacote `p`, senão encaminhará `p` ao nó de rede destinatário daquele pacote.

Para o prosseguimento do trabalho, foi gerado cinco datasets de aproximadamente cem mil observações cada, com taxas de descarte com os seguintes valores de `td`: 3, 15, 30, 60 e 90. Além disso, foi gerado um dataset com uma taxa de descarte dinâmica, onde a cada cem mensagens analisadas, a taxa de descarte era redefinida randomicamente.

A rotulação das amostras como ataque foi feita seguindo uma metodologia de rotulagem subsequente ao ataque. Uma vez que o registro descartado não estaria disponível para o IDS em um cenário realista, o primeiro registro após a ocorrência de um ataque é rotulada como ataque.

[1]Disponível em: https://kaggle.com/datasets/ccf1f5b7870799cbc4443a12b46459dea254647bd90cf3c849691ee7be702b74

Algoritmo 1 Ataque *Grayhole*

```
grayhole(p, td)
for p_i to p_N do
  gera um número aleatório na ∈ {1...100}
  if rn ≤ dr then
    descarta p
  else
    encaminha p
  end if
end for
```

4.2. SHAP

O SHAP é uma biblioteca em *Python* voltada à explicação de modelos de *Machine Learning*, baseada em técnicas de XAI. Por meio do cálculo dos chamados valores SHAP, a ferramenta quantifica a contribuição de cada atributo para uma determinada predição, permitindo interpretar a influência das características nas decisões do modelo.

No contexto do artigo, o SHAP será a ferramenta que nos possibilita interpretar os resultados dos classificadores e entender quais atributos auxiliam o modelo no acerto ou erro da predição. Ademais, o SHAP possibilita a geração de gráficos, auxiliando na representação para o leitor.

4.3. Classificadores de *Machine Learning*

Os classificadores de *Machine Learning*, seguindo o contexto do estudo, têm como objetivo representar uma IDS, permitindo analisar os resultados de predição e identificar possíveis dificuldades na detecção do ataque. Para fornecer maior embasamento à avaliação, foram selecionados três classificadores. O *XGBoost* utiliza a técnica de *boosting* para a classificação das amostras e apresenta bom desempenho em dados tabulares [Xu and Fan 2022]. O *Random Forest* (**RF**) baseia-se em um conjunto de árvores de decisão e também se destaca na análise de dados estruturados [Min et al. 2018]. Por fim, o *Naive Bayes* (**NB**) realiza a classificação por meio de cálculos probabilísticos, servindo como uma abordagem estatística para comparação com os demais modelos.

As diferentes abordagens de algoritmos de classificação tem como motivo buscar uma maior variedade de formatos para a efetuar o mapeamento sistemático de diferentes cenários do ataque, buscando verificar se as mudanças de abordagens entre algoritmos dos classificadores tende a modificar o resultado final.

4.4. Método K-Fold

O método K-Fold é uma técnica de validação cruzada utilizada para aumentar a validade estatística da avaliação dos modelos. Conforme definido em [Anguita et al. 2012], o dataset é dividido em K subconjuntos, sendo que, a cada iteração, um deles é utilizado para teste e os demais para treinamento. O processo é repetido até que todos os subconjuntos tenham sido empregados em ambas as etapas.

No contexto do artigo, o método K-Fold tem papel importante na validação dos resultados dos modelos, pois impossibilita a chance de um conjunto de testes em específico

auxiliar ou prejudicar na predição do classificador. Ademais, o valor de $K = 5$ foi usado, como é comum na literatura.

5. Resultados

A análise de desempenho, explicabilidade e discussão sobre os resultados estão dispostos nas subseções a seguir. Na subseção 5.1, o desempenho dos classificadores em volta dos conjuntos de dados gerados pelo ERENO. Na subseção 5.2, a explicabilidade da predição dos modelos. Na subseção 5.3, a discussão relacionada aos resultados.

5.1. Resultados dos Classificadores

A metodologia para a geração dos resultados baseou-se nos valores adquiridos do desempenho do modelo de *Machine Learning*. Este modelo buscou simular a atuação de um IDS no contexto de uma subestação elétrica. Conforme descrito na subseção 4.1, o modelo busca classificar a ocorrência de ataques, não necessariamente as amostras descartadas. As métricas recolhidas dos modelos foram acurácia, recall, precisão e F1-Score.

Ao decorrer da execução de cada iteração do K-Fold, os valores das métricas de desempenho de cada modelo foram coletados e persistidos em uma variável global. Ao final da K-ésima iteração, foi calculada a média dos valores obtidos para cada métrica considerando todas as amostras do dataset (amostras com rótulo "normal" e amostras com rótulo "Grayhole"). Os resultados foram então sumarizados em diferentes gráficos (Figuras 2 e 3), nos quais cada barra representa o desempenho médio dos classificadores e inclui um intervalo de confiança de 95%.

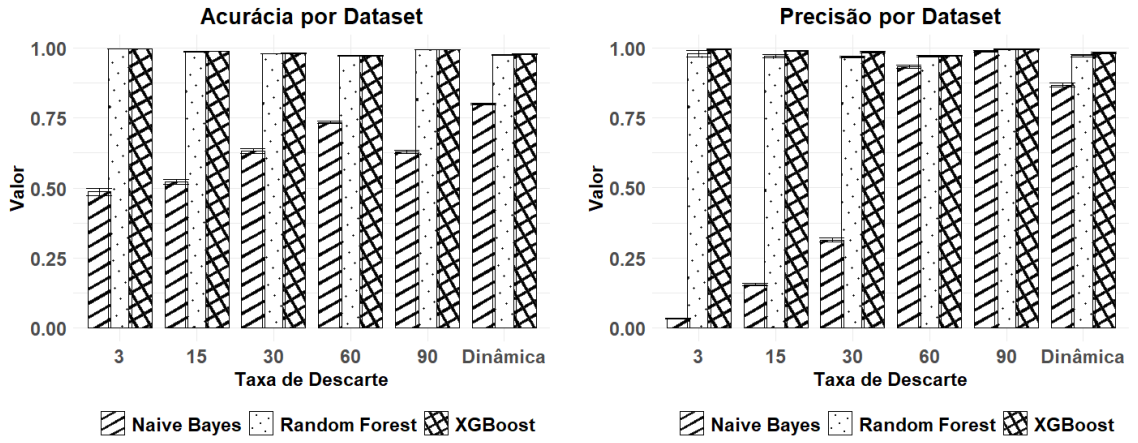


Figura 2. Acurácia e Precisão dos classificadores.

É possível observar no gráfico de F1-Score, que os modelos utilizados tiveram certa dificuldade com a detecção da ocorrência dos ataques em cenários com taxas de descarte abaixo de 30%, demonstrando que em geral taxas de descarte diminutas tem uma chance maior de escapar da detecção.

Ressalta-se a diferença de qualidade no desempenho da predição de ataques do classificador Naive Bayes em relação aos outros. Como o classificador NB trabalha com cálculos probabilísticos, enquanto XGBoost e RF trabalham com árvores de decisão, a predição do NB torna-se mais fraca para identificar padrões do ataque, obtendo melhores resultados na classificação do *Grayhole* quando o ataque é dominante no dataset.

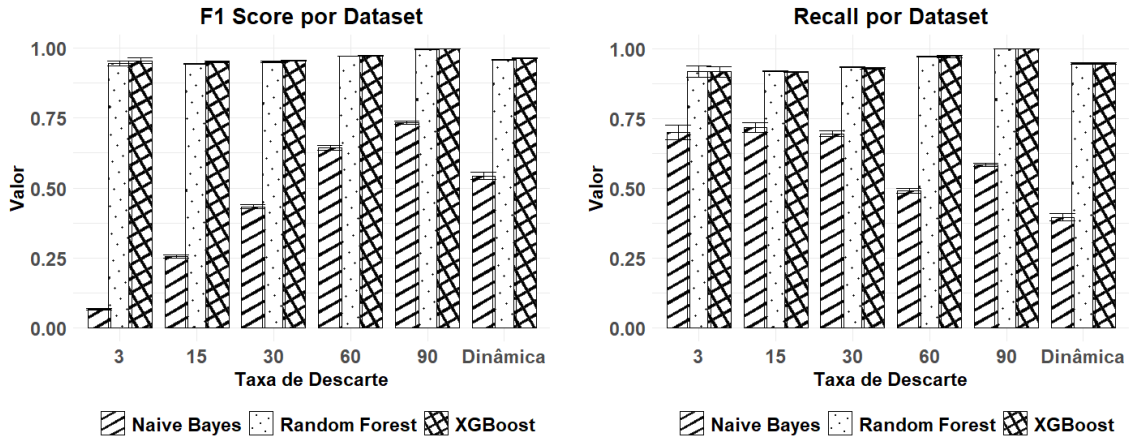


Figura 3. F1-Score e Recall dos classificadores.

Ao analisar o gráfico de precisão, também se percebe a constância dos classificadores XGBoost e RF, demonstrando baixa propensão a falsos positivos independente da gravidade do ataque. Em relação ao Naive Bayes, percebe-se um aumento extremo na sua precisão a partir do dataset com 60% de taxa de descarte, trazendo embasamento para a afirmação de que ele se torna mais confiável para sinalizar descartes quando estes são majoritários.

No dataset dinâmico, os classificadores RF e XGBoost mantiveram desempenho semelhante ao observado no cenário com maior prevalência de ataques, evidenciando robustez frente a mudanças na distribuição dos dados. Em contrapartida, o Naive Bayes apresentou resultados inferiores, embora tenha alcançado desempenho razoável em métricas como F1-Score e recall, indicando capacidade limitada de detecção quando comparado aos demais modelos.

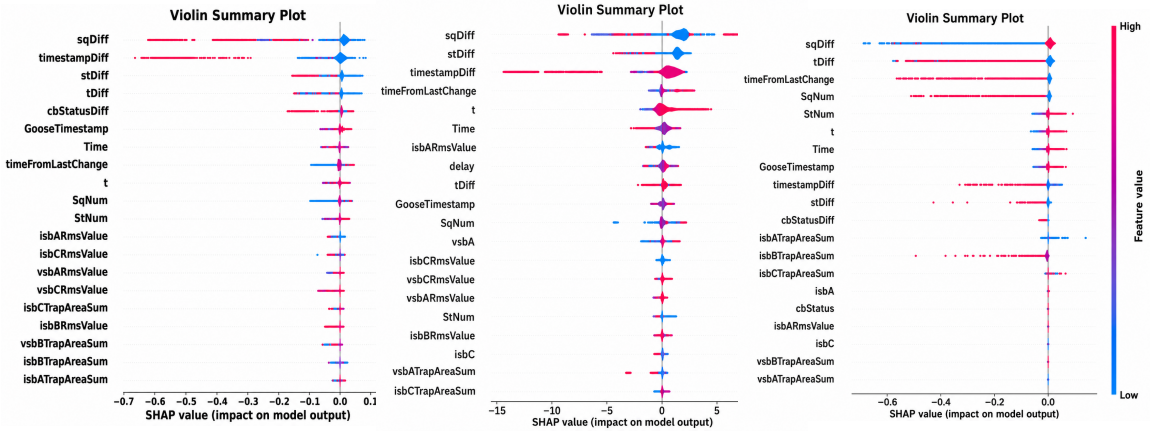
5.2. Explicabilidade dos Resultados

A explicabilidade dos modelos de *Machine Learning* foi empregada para auxiliar a análise dos resultados obtidos, conforme discutido na Seção 5.1. Ao término de cada execução utilizando um dataset, os resultados gerados pelo modelo eram transformados em um objeto *Explainer*. Em seguida, os valores SHAP eram calculados a partir desse objeto, possibilitando uma análise das predições realizadas pelo modelo.

Embora diversos tipos de gráficos SHAP tenham sido gerados ao longo deste estudo, optou-se pela utilização dos gráficos do tipo violino para representar os resultados de forma mais clara e visualmente informativa. As Figuras 4, 5 e 6 apresentam uma comparação entre os valores de explicabilidade obtidos pelos classificadores, com base nos datasets com taxas de descarte de 3% e 90%, além de analisar o comportamento dos classificadores em relação ao dataset dinâmico.

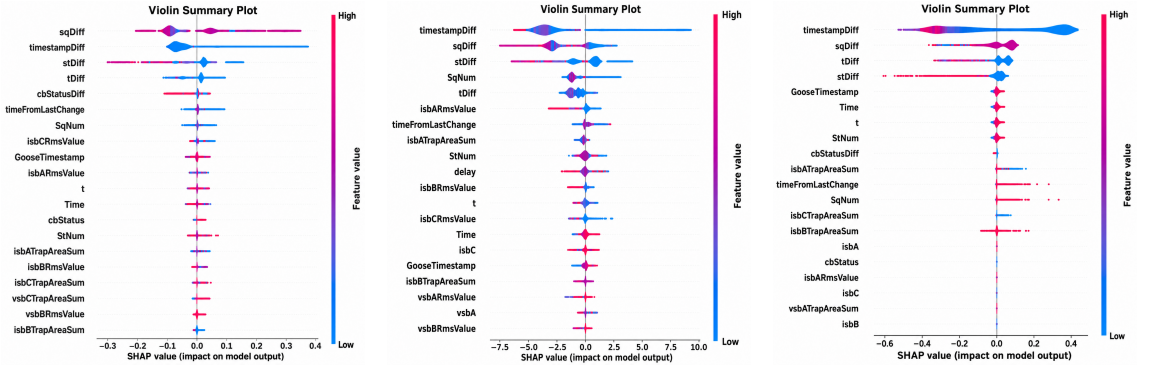
Ao comparar os datasets representados nos gráficos, observa-se uma concentração de importância nos atributos de consistência. Tal comportamento pode ser atribuído ao fato de que esses atributos derivam de cálculos realizados entre diferentes mensagens do dataset, não correspondendo a atributos originalmente definidos no protocolo GOOSE.

Ao analisar os três cenários (Figuras 4, 5 e 6) de forma comparativa, observa-se



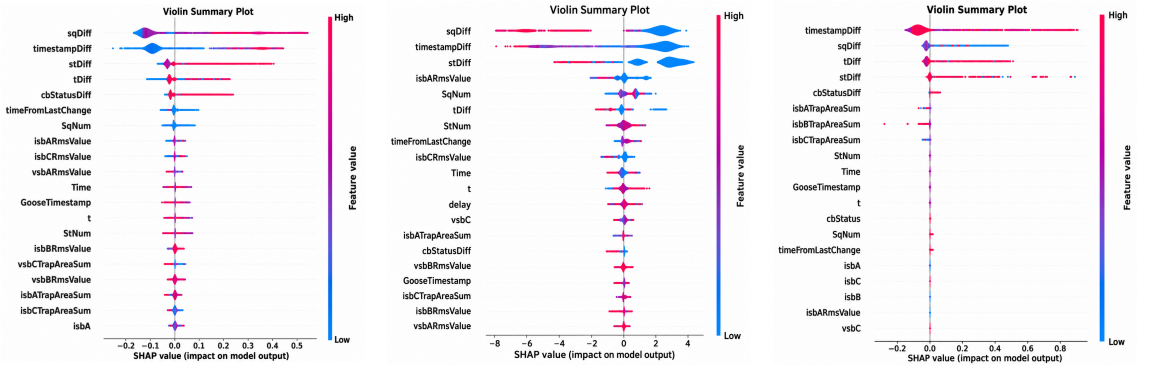
(a) SHAP (RF). (b) SHAP (XGBoost). (c) SHAP (NB).

Figura 4. Comparação entre classificadores e com a taxa de descarte mínima.



(a) SHAP (RF). (b) SHAP (XGBoost). (c) SHAP (NB).

Figura 5. Comparação entre classificadores e com a taxa de descarte máxima.



(a) SHAP (RF). (b) SHAP (XGBoost). (c) SHAP (NB).

Figura 6. Comparação entre classificadores e com a taxa de descarte dinâmica.

um padrão consistente: os classificadores baseados em árvores (RF e XGBoost) atribuem a maior importância aos atributos *timestampDiff* e *sqDiff* independentemente da taxa de descarte ou de sua natureza estática ou dinâmica, enquanto o Naive Bayes concentra sua dependência preditiva quase exclusivamente em *timestampDiff*, com uma dispersão de valores SHAP crescente à medida que a prevalência de ataques diminui. Essa recorrência

não é incidental: *timestampDiff* e *sqDiff* são atributos enriquecidos, calculados a partir da diferença entre mensagens GOOSE consecutivas, e capturam diretamente as duas consequências estruturais do descarte seletivo de pacotes, sendo a quebra da regularidade temporal de retransmissão (baseada no mecanismo de backoff exponencial descrito na Subseção 2.2) e a introdução de saltos na sequência de números esperada pelo protocolo.

Dessa forma, é esperado que qualquer classificador capaz de capturar relações não lineares entre atributos convirja para esses dois sinais como os mais discriminativos, o que explica a maior estabilidade de RF e XGBoost frente às variações de cenário. Já a maior dispersão observada no Naive Bayes decorre de sua natureza probabilística: como o classificador assume independência condicional entre atributos, ele não consegue modelar a relação conjunta entre *timestampDiff* e *sqDiff* que caracteriza o ataque, dependendo de forma mais instável de um único atributo para discriminar entre as classes.

5.3. Discussão

Considerando todos os cenários avaliados, os algoritmos baseados em árvores de decisão apresentaram desempenho superior ao Naive Bayes, devido à sua maior capacidade de capturar relações não lineares entre os atributos.

À medida que aumenta a taxa de descarte, o desempenho do Naive Bayes melhora consideravelmente, especialmente na métrica de precisão. Isso ocorre porque, com o aumento de amostras rotuladas como ataque, os atributos mais diretamente relacionados à ocorrência do ataque (como o *timestampDiff*) se tornam mais evidente e fáceis de se capturar por um modelo linear de classificação como do Naive Bayes.

A inclusão do dataset dinâmico, representando um cenário mais desafiador do ataque, revela a eficácia dos classificadores RF e XGBoost. Ambos mantêm um desempenho elevado, sugerindo sua adequação para aplicações práticas onde a taxa de ocorrência de ataques pode não ser constante.

Quanto às implicações dos resultados, a superioridade do RF e do XGBoost, aliada à capacidade de explicar os atributos mais importantes para a detecção, tornam-nos promissores para trabalhar como IDS baseados em *Machine Learning* para o protocolo GOOSE. Ademais, a identificação dos atributos cruciais através da explicabilidade pode auxiliar na otimização da coleta de dados e na compreensão do comportamento de anomalias no contexto de ataque *Grayhole*.

6. Conclusão

O presente trabalho partiu do objetivo de facilitar a compreensão e explicabilidade de diferentes cenários do ataque DoS do tipo *Grayhole* ao protocolo GOOSE em redes IEC-61850. Enquanto estudos anteriores focaram em cenários com taxas de descarte fixa ou com diferentes ataques do tipo negação de serviços, o atual estudo oferece um mapeamento sistemático do impacto diante de diferentes taxas de descarte do ataque *Grayhole*, além de oferecer uma maior explicabilidade referente aos resultados das predições.

A análise de explicabilidade evidenciou que os atributos próprios do protocolo GOOSE não oferecem alta detectabilidade do ataque, sendo os atributos enriquecidos *timestampDiff* e *sqDiff* — que capturam as irregularidades temporais e de sequência — os mais determinantes, com o XGBoost concentrando importância nesses poucos sinais e o RF distribuindo-a entre mais atributos.

Referências

Anguita, D., Ghelardoni, L., Ghio, A., Oneto, L., Ridella, S., et al. (2012). The'k'in k-fold cross validation. In *ESANN*, volume 102, pages 441–446.

Ashraf, S. and et al. (2021). A comprehensive review of cybersecurity in iec 61850 based smart grids. *IEEE Access*, 9:88569–88588.

Ashraf, S., Shawon, M. H., Khalid, H. M., and Muyeen, S. (2021). Denial-of-service attack on iec 61850-based substation automation system: A crucial cyber threat towards smart substation pathways. *Sensors*, 21(19):6415.

Baigent, D., Adamiak, M., Mackiewicz, R., and Sisco, G. (2004). Iec 61850 communication networks and systems in substations: An overview for users. *SISCO Systems*.

Basumallik, S. (2020). A taxonomy of data attacks in power systems. *arXiv preprint arXiv:2002.11011*.

Elbez, G., Nahrstedt, K., and Hagenmeyer, V. (2022). Early detection of goose denial of service (dos) attacks in iec 61850 substations. In *2022 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pages 367–373.

Gonçalves, J. C. et al. (2023). Modelagem do ataque grayhole ao protocolo de comunicação goose usando o framework ereno.

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., and Yang, G.-Z. (2019). Xai—explainable artificial intelligence. *Science robotics*, 4(37):eaay7120.

Lázaro, J., Astarloa, A., Rodríguez, M., Bidarte, U., and Jiménez, J. (2021). A survey on vulnerabilities and countermeasures in the communications of the smart grid. *Electronics*, 10(16):1881.

McLennan, M., Group, S., and Group, Z. I. (2022). The global risks report 2022 17th edition. Disponível em: `https://www.weforum.org/reports/global-risks-report-2022/`.

Min, E., Long, J., Liu, Q., Cui, J., and Chen, W. (2018). Tr-ids: Anomaly-based intrusion detection through text-convolutional neural network and random forest. *Security and Communication Networks*, 2018(1):4943509.

Pal, S., Sikdar, B., and Chow, J. H. (2016). An online mechanism for detection of grayhole attacks on pmu data. *IEEE Transactions on Smart Grid*, 9(4):2498–2507.

Quincozes, S. (2022). *ERENO: An Extensible Tool for Generating Realistic IEC–61850 Intrusion Detection Datasets*. PhD thesis, Fluminense Federal University.

Quincozes, S. E., Kazienko, J. F., and Quincozes, V. E. (2023). An extended evaluation on machine learning techniques for denial-of-service detection in wireless sensor networks. *Internet of Things*, 22:100684.

Reda, H. T., Ray, B., Peidaee, P., Anwar, A., Mahmood, A., Kalam, A., and Islam, N. (2021). Vulnerability and impact analysis of the iec 61850 goose protocol in the smart grid. *Sensors*, 21(4):1554.

Sidhu, T. S. and Yin, Y. (2007). Modelling and simulation for performance evaluation of iec61850-based substation communication systems. *IEEE Transactions on Power Delivery*, 22(3):1482–1489.

Tefek, U., Esiner, E., Mashima, D., and Hu, Y.-C. (2022). Analysis of message authentication solutions for iec 61850 in substation automation systems. In *2022 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, pages 224–230. IEEE.

V, G. and Yadav, R. (2024). Enhancing the security of iec-61850 goose messages during transmission. In *2024 IEEE International Conference on Power and Energy (PECon)*, pages 82–85.

Xu, W. and Fan, Y. (2022). Intrusion detection systems based on logarithmic autoencoder and xgboost. *Security and Communication Networks*, 2022(1):9068724.