Um Framework Gerador de Tráfego para Detecção de Intrusões em Redes CAN

Luiz F. Junior[1], **Paulo Sérgio M. Vargas**[1],
Paulo Vitor C. Lima[1], **Silvio E. Quincozes**[1,2]

[1]Faculdade de Computação (FACOM) – Universidade Federal de Uberlândia (UFU), Brasil

[2]Campus Alegrete, Universidade Federal do Pampa (UNIPAMPA), Brasil.

{luiz.fogliato, paulo.vargas, paulo.lima1}@ufu.br

silvioquincozes@unipampa.edu.br

Abstract. *Controller Area Network (CAN) networks enable intra-vehicle communication between Electronic Control Units (ECU) and external communication via WiFi, Bluetooth, and USB, making them vulnerable to cyber attacks. This work presents a dataset generator framework to help detect intrusions in CAN networks, using GANs (Generative Adversarial Networks) and VAEs (Variational Autoencoders). GANs create datasets with similar distribution to real data, while VAEs capture variability, resulting in realistic and varied datasets. Preliminary results show that the proposed method generates data sets of adequate quality and variability and can be adapted to other environments.*

Resumo. *As redes Controller Area Network (CAN) permitem comunicação intraveicular entre as Unidades Eletrònicas de Controle (ECU) e comunicação externa via WiFi, Bluetooth e USB, tornando-as vulneráveis a ataques cibernéticos. Este trabalho apresenta um framework gerador de conjuntos de dados para ajudar na detecção de intrusões em redes CAN, utilizando GANs (Generative Adversarial Networks) e VAEs (Variational Autoencoders). GANs criam datasets com distribuição similar aos dados reais, enquanto VAEs capturam a variabilidade, resultando em conjuntos de dados realistas e variados. Resultados preliminares mostram que o método proposto gera datasets de qualidade e variabilidade adequadas, podendo ser adaptado para outros ambientes.*

1. Introdução

A *Controller Area Network* (CAN) é o protocolo predominante em redes internas de veículos, permitindo que Unidades de Controle Eletrônico (ECUs) se comuniquem e compartilhem informações cruciais para a segurança do veículo. Além disso, essas redes possibilitam a comunicação externa através de interfaces como WiFi, Bluetooth e *Universal Serial Bus* (USB), o que as torna suscetíveis a ataques cibernéticos, especialmente considerando a ausência de mecanismos como autenticação e criptografia nessas redes [Dresch et al. 2024][Aliwa et al. 2022][Avatefipour and Malik 2017].

Nesse contexto, Sistemas de Detecção de Intrusões (IDS) precisam ser considerados para proteger recursos que utilizam redes CAN [Wang et al. 2018]. Em particular, IDSs baseados em aprendizado de máquina têm demonstrado alta precisão e eficiência

[Smirti et al. 2020]. No entanto, esse tipo de mecanismo depende de conjuntos de dados para a construção de modelos preditivos.

Atualmente, na literatura existem conjuntos de dados como Survival [Han et al. 2018], Car-Hacking [Seo et al. 2018] e OTIDS [Lee et al. 2017]. No entanto, cada um desses conjuntos de dados contém dados especializados de acordo com o tipo de veículo e modelos de ataques envolvidos na coleta de suas amostras. Já geradores de dados como [Chougule et al. 2023] são limitados quanto a produção massiva de dados de forma variável e confiável. Portanto, faltam mecanismos que sejam eficientes para gerar dados com uma combinação de confiabilidade e variabilidade.

Este trabalho visa apresentar um gerador de conjunto de dados que seja confiável o suficiente para gerar dados de maneira massiva e com a devida variação para suprir as limitações da literatura atual. Para tanto, foram combinadas duas abordagens: O uso de Redes Generativas Adversariais (*Generative Adversarial Networks* – GANs) e Codificadores Automáticos Variacionais (Variational Autoencoders – VAEs). Dessa forma, este trabalho contribui para a otimização do processo de treinamento de IDSs automotivos, minimizando ao máximo o re-treino dos algoritmos. Como prova de conceito, foi produzido um conjunto de dados, o qual está publicamente disponível[1]. Ademais, é importante observar que a ferramenta proposta, que também está publicamente disponível[2], pode ser evoluída para a geração de diversos conjuntos de dados, cobrindo uma infinidade de desafios existentes na literatura.

2. Redes CAN

As redes intraveiculares que utilizam o protocolo CAN, ou "Redes CAN" são utilizadas para permitir a comunicação entre as ECUs implementadas em veículos, conforme ilustrado na Figura 1. O CAN é amplamente adotado nas redes veiculares devido à sua alta resistência a interferências eletromagnéticas e aos seus baixos custos de implementação.

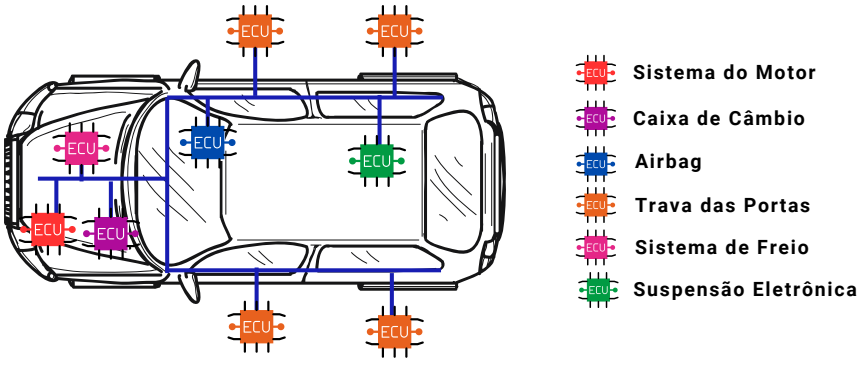

Figura 1. Rede CAN. Autoria própria. Baseado em [Khan et al. 2024]

A comunicação nas redes CAN é baseada em mensagens de *broadcast*; a transmissão no barramento CAN utiliza um método de arbitragem bit a bit para resolver a concorrência. Quando dois nós diferentes iniciam a transmissão de um quadro ao mesmo tempo, o nó com a maior prioridade continua enviando o quadro

[1]Conjunto de dados disponível em `https://www.kaggle.com/datasets/candatasetsample/can-attack-free-dataset-sample`

[2]Código-fonte disponível em: `https://anonymous.4open.science/r/dataset-generator-D20B/`

sem interrupção, enquanto o outro nó recua e tenta a transmissão novamente mais tarde [Pollicino et al. 2024].

3. Trabalhos Relacionados

O estudo de [Chougule et al. 2023] é um dos principais trabalhos relacionados, propondo uma ferramenta chamada SCAN-GAN para gerar dados sintéticos que imitam cenários reais. A arquitetura proposta pelos autores utiliza uma camada discriminadora para distinguir dados reais de dados falsos, permitindo a geração de resultados que se assemelham a situações do mundo real. A derivação de dados a partir de datasets originais seguida pela discriminação do que destoa da realidade constitui um caminho lógico e conservador. Contudo, o SCAN-GAN é limitado pelos dados de entrada, sugerindo a necessidade de métodos que gerem dados mais variados e confiáveis.

Para superar essas limitações, os V*ariational Autoencoders* (VAEs) têm se destacado na literatura por sua capacidade de reduzir ruídos [Asaoka et al. 2020], diminuir a dimensionalidade [Graving and Couzin 2020, Mahmud et al. 2020] e gerar dados realistas [Pan et al. 2019]. Em aplicações de predição de trajetória para direção autônoma, o VAE mostrou-se eficiente ao gerar dados variados que refletem diferentes terrenos e estilos de condução humana [Miguel et al. 2022].

O trabalho de [Razghandi et al. 2024] investiga a combinação das técnicas VAE e GAN para a geração de datasets sintéticos. Esta abordagem é particularmente relevante para a nossa pesquisa, pois a combinação dessas técnicas permite capturar a complexidade e a variabilidade dos dados reais, resultando em dados realistas utilizáveis em sistemas de detecção de intrusões. Embora aplicada em outro contexto, a combinação VAE-GAN demonstra grande potencial para melhorar a qualidade e a variabilidade dos dados gerados.

A proposta do presente trabalho utiliza o conjunto de dados descrito por [Lee et al. 2017] como *CAN-Intrusion*, que contém dados reais de intrusão em redes CAN, para criar e validar o gerador de conjunto de dados proposto na Seção 4. Esses dados fornecem uma base sólida para avaliar a eficácia das técnicas de geração de dados propostas (GAN, VAE e a combinação deles). O uso de dados reais do conjunto de dados *CAN-intrusion* como entrada para os modelos permite a criação de amostras sintéticas que refletem os padrões observados em ambientes veiculares reais.

4. Proposta

De modo a lidar com os desafios apresentados nas seções anteriores, neste trabalho é proposto um novo *framework* para a geração de conjuntos de dados a fim de viabilizar o treinamento de IDSs para serem implantados em redes CAN. Tal proposta se baseia no emprego das abordagens GAN e VAE.

O emprego de GAN permite a geração de dados através de redes neurais. Particularmente, GANs consistem em duas redes neurais competindo entre si: uma gera dados sintéticos (gerador) e a outra avalia sua autenticidade (discriminador), resultando em dados altamente realistas. Com isso, a qualidade dos dados sintéticos é maximizada a fim de que o discriminador não seja capaz de diferenciar tais dados como sintéticos ou reais.

Por outro lado, ao empregar a técnica VAE, há dois elementos principais: o *encoder* (codificador) e o *decoder* (decodificador). Estes, por sua vez, são responsáveis

respectivamente por: i) mapear os dados de entrada e realiza a distribuição probabilística deles; e ii) mapear as amostras resultantes do *encoder* novamente para os dados originais, ou seja, decodificando-os. O VAE consegue então obter uma reconstrução fiel e uma distribuição regular. Essa técnica permite a geração de dados de menor qualidade quando comparado com as GAN, porém consegue gerar uma maior diversidade com relação aos dados, podendo assim ser interessante para simular diferentes tipos de cenários.

Portanto, a presente proposta explora GANs e VAEs combinadas a fim de criar conjuntos de dados sintéticos variados e realistas, essenciais para treinar sistemas de detecção de intrusões com eficácia.

5. Experimentos

Nesta seção, serão apresentados os experimentos realizados com o objetivo de avaliar o desempenho de diferentes técnicas de geração de datasets e sua eficácia em relação aos dados reais. Três cenários distintos foram considerados: inicialmente, é utilizada a técnica GAN, seguida pela aplicação da técnica VAE, e por fim, a presente proposta GAN-VAE, que consiste na combinação de ambas as técnicas, é avaliada. Cada cenário será comparado em termos de parâmetros de similaridade com os dados reais, permitindo uma análise detalhada de suas vantagens e limitações.

5.1. Matériais e Métodos

O cenário de experimentação é ilustrado na Figura 2. O fluxo começa com a inserção do conjunto de dados reais presente em [Lee et al. 2017]. A partir desses dados, é executada a etapa de pré-processamento, a qual consiste em separar o dataset em suas features específicas e transformar o mesmo para um formato csv. Em seguida, os dados pré-processados são usados como entrada para as três abordagens comparadas: GAN, VAE e GAN-VAE (presente proposta).

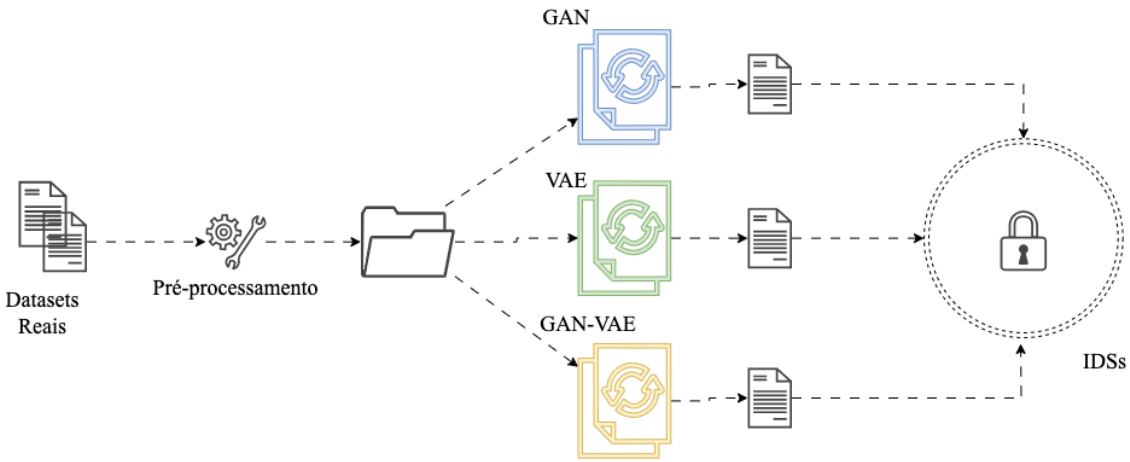

Figura 2. Distribuição dos dados reais.

Dentre as bibliotecas usadas, o *TensorFlow* funciona como uma ferramenta de código aberto para aprendizado de máquina, da qual se utilizou dos algoritmos GAN e VAE; o *keras* como uma API para trabalhar com redes neurais; o *numpy* e *pandas* para manipular os dados; o *matplotlib* para plotagem dos gráficos e imagens.

6. Resultados e conclusão

Conforme mencionado, foi realizada a execução de um gerador de datasets, utilizando três cenários diferentes. Um apenas com o modelo GAN, na sequência apenas com modelo VAE e por fim utilizando as duas em conjunto no mesmo dataset. Para realizar a execução dos dois modelos em conjunto, foi necessário a utilização de ajuste de pesos tanto para o GAN quanto para o VAE durante o treinamento. Estes pesos foram utilizados no último cenário para realizar a execução.

Conforme descrito na arquitetura, há uma etapa de pré-processamento que é importante para normalizar os dados e formatos e se tornar uma entrada padrão para os geradores. Em nossos experimentos essa etapa é realizada por um script Python que está disponível no código fonte. Assim, após esta etapa executamos a geração dos datasets utilizando os três cenários, tendo como entrada os dados reais de [Lee et al. 2017]. Podemos verificar pela Figura 3 os valores da métrica chamada de JSD (*Jensen-Shannon Divergence*) conforme [Lin 1991], onde é calculada a similaridade entre datasets. Para este cálculo realizamos a comparação entre os datasets gerados pelos métodos GAN, VAE e GAN-VAE com os dados reais. Por fim, ao consumir tais dados, recomenda-se o uso de técnicas de seleção de *features* para definir quais dados são relevantes para a detecção de intrusões [Scherer et al. 2024a, Scherer et al. 2024b].

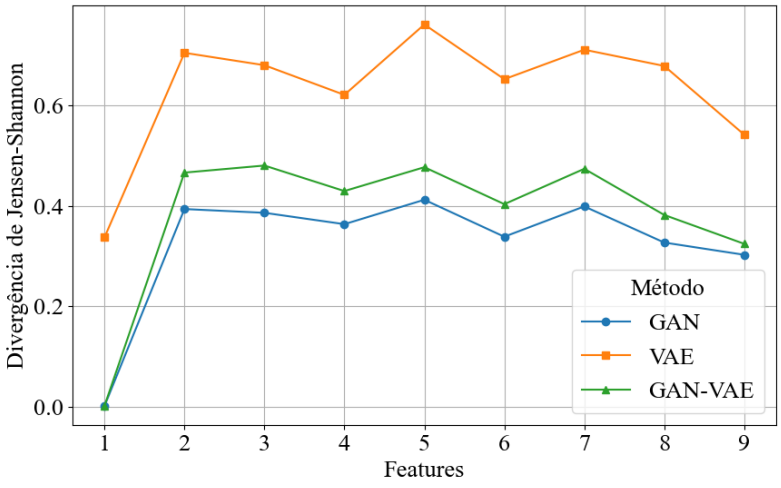

Figura 3. Divergência Jensen-Shannon para Diferentes Métodos.

Realizamos também uma análise levando em consideração a distribuição dos datasets gerados. Para isto foi utilizada a técnica FID (*Frechet Inception Distance*) que é uma métrica utilizada para avaliar a qualidade de imagens geradas sinteticamente. Esta técnica mede a distância entre as distribuições das características extraídas dos dados reais e gerados, resultando assim em um score. Mesmo que esta métrica seja bastante utilizada para imagens, podemos aplicar para o nosso cenário, pois a ideia é comparar justamente essas distribuições produzidas pelos nossos geradores com o dataset real de [Lee et al. 2017].

Este cálculo envolve a comparação das médias e variâncias dos valores originais e gerados, adaptando a fórmula de FID para trabalhar com essas estatísticas unidimensionais do nosso cenário. O score para cada gerador pode ser visto pela Tabela 1, perceba

a seguir que comprovadamente a combinação dos métodos eleva o nível de qualidade dos dados gerados em relação aos trabalhos que precedem este.

Modelo	FID
GAN	4.44404005e-05
VAE	21.6926077
Combined GAN-VAE	4.48103307e-06

Tabela 1. FID score para diferentes modelos

6.1. Considerações Finais

É perceptível que utilizando o método GAN há um dataset bastante similar em termos de distribuição quando comparado com os dados reais. Ao confirmar a distribuição pelo modelo VAE, vemos que a técnica se preocupa mais com a variabilidade realmente, não tendo uma distribuição com qualidade se comparado aos dados reais. Porém ao combinar as duas e utilizar o método GAN-VAE, é notável uma boa distribuição e também uma variabilidade maior em comparação com apenas o método GAN. Isso resulta em um gerador de dataset com uma maior qualidade, bastante similar a realidade e principalmente para entrada em sistemas de IDS, onde é preciso uma grande quantidade de dados fidedignos e de maneira variável para se ter um treinamento cada vez mais efetivo.

Em trabalhos futuros outros métodos podem ser implementados, mas, sendo especificado apenas um único gerador, avaliando este framework com diferentes tipos de datasets para verificar seu comportamento e validar diversos padrões de dados, contribuindo significativamente para a detecção de intrusões em ambientes complexos.

Referências

Aliwa, E., Rana, O., Perera, C., and Burnap, P. (2022). Cyberattacks and countermeasures for in-vehicle networks. *ACM Computing Surveys*, 54(1):1–37.

Asaoka, R., Murata, H., Matsuura, M., Fujino, Y., Yanagisawa, M., and Yamashita, T. (2020). Improving the structure–function relationship in glaucomatous visual fields by using a deep learning–based noise reduction approach. *Ophthalmology Glaucoma*, 3(3):210–217.

Avatefipour, O. and Malik, H. (2017). State-of-the-art survey on in-vehicle network communication "can-bus" security and vulnerabilities. *International Journal of Computer Science and Network*, pages 720–727.

Chougule, A., Agrawal, K., and Chamola, V. (2023). Scan-gan: Generative adversarial network based synthetic data generation technique for controller area network. *IEEE Internet of Things Magazine*, 6(3):126–130.

Dresch, F. N., Scherer, F. H., Quincozes, S. E., and Kreutz, D. L. (2024). Modelos interpretáveis com inteligência artificial explicável (XAI) na detecção de intrusões em redes intra-veiculares controller area network (CAN). In *Anais do XIX Simpósio Brasileiro de Segurança da Informaçao e de Sistemas Computacionais*. SBC.

Graving, J. M. and Couzin, I. D. (2020). Vae-sne: A deep generative model for simultaneous dimensionality reduction and clustering. *BioRxiv*.

Han, M. L., Kwak, B. I., and Kim, H. K. (2018). Anomaly intrusion detection method for vehicular networks based on survival analysis. *Vehicular Communications*, 14:52–63.

Khan, M. H., Javed, A. R., Iqbal, Z., Asim, M., and Awad, A. I. (2024). DivaCAN: Detecting in-vehicle intrusion attacks on a controller area network using ensemble learning. *Computers & Security*, 139:103712.

Lee, H., Jeong, S. H., and Kim, H. K. (2017). Otids: A novel intrusion detection system for in-vehicle network by using remote frame. In *2017 15th Annual Conference on Privacy, Security and Trust (PST)*, volume 00, pages 57–5709.

Lin, J. (1991). Divergence measures based on the shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151.

Mahmud, M. S., Huang, J. Z., and Fu, X. (2020). Variational autoencoder-based dimensionality reduction for high-dimensional small-sample data classification. *International Journal of Computational Intelligence and Applications*, 19(1).

Miguel, M. M., Armignol, J. M., and Garcia, F. (2022). Vehicles trajectory prediction using recurrent vae network. *IEEE Access*, 10:32742–32749.

Pan, Z., Wang, J., Liao, W., Chen, H., Yuan, D., Zhu, W., Fang, X., and Zhu, Z. (2019). Data-driven ev load profiles generation using a variational autoencoder. *Energies*, 12(5):849.

Pollicino, F., Stabili, D., and Marchetti, M. (2024). Performance comparison of timing-based anomaly detectors for controller area network: A reproducible study. *ACM Transactions on Cyber-Physical Systems*, 8(2):1–24.

Razghandi, M., Zhou, H., Erol-Kantarci, M., and Turgut, D. (2024). Smart home energy management: Vae-gan synthetic dataset generator and q-learning. *IEEE Transactions on Smart Grid*, 15(2):1562–1573.

Scherer, F. H., Dresch, F. N., Quincozes, S. E., Kreutz, D., and Quincozes, V. E. (2024a). IWSHAP: Um método de seleção incremental de características para redes CAN baseado em Inteligência Artificial Explicável (XAI). In *Anais do XXIV Simpósio Brasileiro de Segurança da Informaçao e de Sistemas Computacionais*. SBC.

Scherer, F. H., Dresch, F. N., Quincozes, S. E., Kreutz, D., and Quincozes, V. E. (2024b). IWSHAP: Uma ferramenta para seleção incremental de características utilizando IWSS e SHAP. In *Anais Estendidos do XXIV Simpósio Brasileiro de Segurança da Informaçao e de Sistemas Computacionais*. SBC.

Seo, E., Song, H. M., and Kim, H. K. (2018). Gids: Gan based intrusion detection system for in-vehicle network. In *2018 16th Annual Conference on Privacy, Security and Trust (PST)*, pages 1–6.

Smirti, D., Medha, P., and Weiqing, S. (2020). A comparative study on contemporary intrusion detection datasets for machine learning research. *2020 IEEE International Conference on Intelligence and Security Informatics (ISI)*.

Wang, Q., Qian, Y., Lu, Z., Shoukry, Y., and Qu, G. (2018). A delay based plug-in-monitor for intrusion detection in controller area network. In *2018 Asian Hardware Oriented Security and Trust Symposium (AsianHOST)*, pages 86–91, Hong Kong. IEEE.