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

Luiz F. Junior, Paulo Sergio M. Vargas,
Paulo Vitor C. Lima, **Silvio E. Quincozes**

Qual é o veículo mais seguro?

Fusca 1300 vs Tesla Model S

Hacker de 19 anos acessa o sistema de 25 carros da Tesla no mundo e culpa os proprietários

Jovem relata que conseguiu ter acesso à central multimídia, abertura das portas e janelas, e afirma que poderia até dar partida no veículo

Por Emily Nery

16/01/2022 09h27 · Atualizado há 2 anos

Introdução

- **Controller Area Network (CAN)**
  - → Comunicação entre Unidades Eletrônicas de Controle (ECU).
  - → Sem autenticação e criptografia!
  - → Em alguns cenários, são integradas à dispositivos com conectividade externa ao veículo!

Vulneráveis a ataques físicos e cibernéticos!

Motivação

Em Minneapolis (EUA), quase dois mil carros da Kia e da Hyundai foram roubados até meados de 2023.

Justificativa

- **Como detectar uma tentativa de ataque?**
  - IDSs baseados em Machine Learning = alta eficiência!
    - Mas eles precisam de **dados**!
- **Datasets disponíveis são limitados!**
  - Survival, Car-Hacking, OTIDS e X-CANIDS
    - Cada um é especializado para um tipo de veículo!

Usar dados super especializados ou confiar em um gerador de dados sintéticos?

Objetivo

- **Proposta de um gerador de conjunto de dados:**
  - **Confiável** o suficiente para gerar dados de maneira massiva!
  - Com a devida **variabilidade** para suprir as limitações da literatura!
- **Para tanto, foram combinadas duas abordagens:**
  - O uso de Redes Generativas Adversariais (GANs)
  - Codificadores Automáticos Variacionais (VAEs)

Proposta

- **GAN**
  - Permite a geração de dados através de redes neurais.
    - **Duas redes neurais competindo entre si:**
      - ✓ **uma gera dados sintéticos (gerador)**
      - ✓ **outra avalia sua autenticidade (discriminador)**
    - **Com isso, a qualidade dos dados sintéticos é maximizada:**
      - ✓ **O discriminador não deve ser capaz de diferenciar tais dados como sintéticos ou reais.**

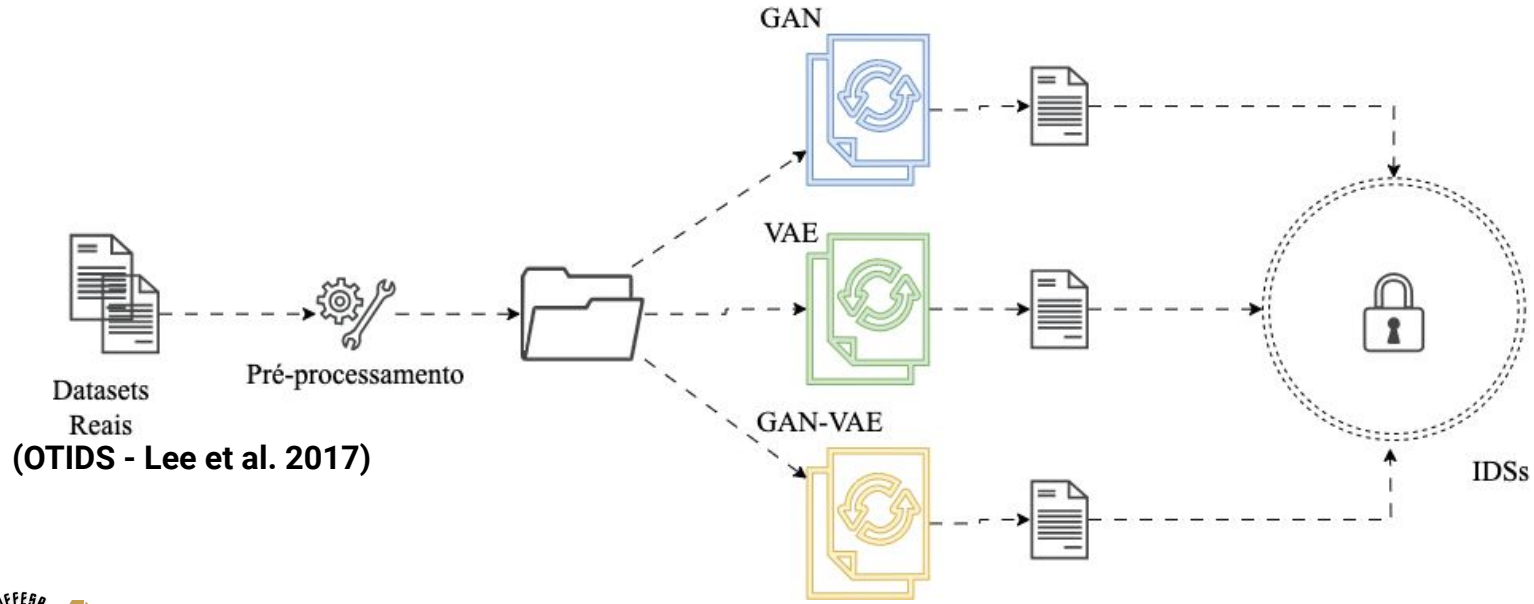
Proposta

- **VAE**
  - Permite encontrar maior diversidade de dados
    - **Através de um codificador e um decodificador:**
      - ✓ **Codificador (*encoder*) mapeia os dados de entrada e realiza a distribuição probabilística deles.**
      - ✓ **Decodificador (*decoder*): mapeia as amostras resultantes encoder novamente para os dados originais, ou seja, decodificando-os.**
    - **O VAE consegue então obter:**
      - ✓ **Uma reconstrução fiel;**
      - ✓ **Uma distribuição regular.**

Proposta

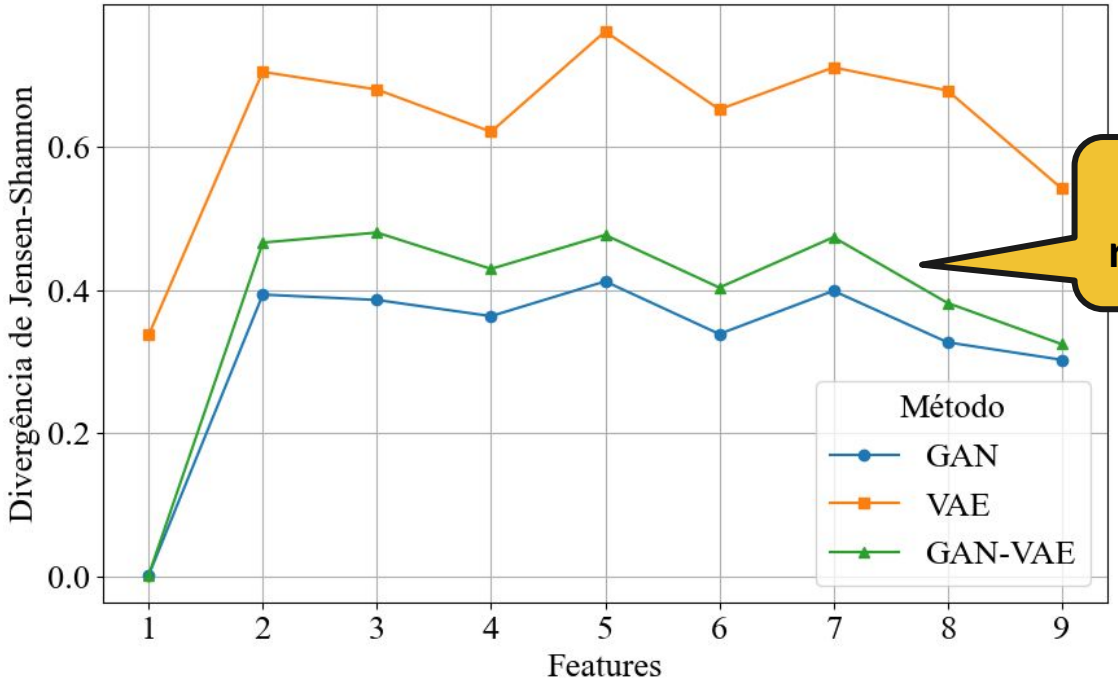
- **Combinação de GAN-VAE**
  - VAE
    - **Maior diversidade de dados!**
      - ✓ **Interessante para simular diferentes tipos de cenários!**
  - GAN
    - **Maior qualidade nos dados!**
      - ✓ **Interessante para manter os dados fidedignos!**

Experimentos



(OTIDS - Lee et al. 2017)

Resultados: Divergência de Jensen-Shannon



Variabilidade em relação aos dados reais.

Resultados: FID (Frechet Inception Distance)

- **GAN:** 4.44×10^{-5}
  - Alta **fidelidade** do modelo em reproduzir dados realistas.
- **VAE:** ~21.7
  - Dificuldade do método em reproduzir detalhes (**menor realismo**).
- **GAN-VAE:** 4.48×10^{-6}
  - **Maior fidelidade** do modelo híbrido entre os três modelos.

Considerações Finais

- **A plataforma Moodle não parece estar totalmente segura!**
  - Identificamos 894 alertas através da ferramenta OWASP ZAP;
    - **Esses alertas representam 20 tipos de vulnerabilidades.**
- **Trabalhos futuros**
  - **Explorar essas vulnerabilidades!**
  - Em 2025, uma nova lista do OWASP Top Ten estará disponível!
    - **Experimentos comparativos com versões futuras do Moodle também são importantes!**

Referências

- 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

Obrigado!

silvioquincozes@unipampa.edu.br