VI Simpósio de Iniciação Científica 2023
Ciência como ferramenta de transformação da sociedade

KITSUNE - NIDS DE MONITORAMENTO DE REDE NÃO SUPERVISIONADO

Vinicius Silva Gonçalves[1] (IC), Edvard Martins de Oliveira[1] (PQ)
[1] *UNIFEI – Universidade Federal de Itajubá.*

Palavras-chave: KITSUNE - NIDS DE MONITORAMENTO DE REDE NÃO SUPERVISIONADO

Introdução

A Temática de segurança cibernética se evidencia como um tema cada vez mais importante em diversas áreas como no mundo dos negócios, saúde e segurança. De acordo com a Cybersecurity Ventures, os crimes cibernéticos devem causar um prejuízo na ordem de US$8 trilhões em 2023, tornando o cibercrime a terceira maior economia do mundo segundo [1]. Em agosto de 2018 foi retificado a Lei Geral de Proteção de Dados Pessoais(LGPDP) estipulando uma diretriz de como os dados dos cidadãos brasileiros devem ser coletados e tratados. Em agosto de 2023, hackers interromperam o funcionamento das unidades de saúde operadas pela Prospect Medical Holdings, que possuem hospitais e clínicas na Califórnia, Texas, Connecticut, Rhode Island e Pensilvânia [2]. Estes ataques podem ser extremamente comprometedores para vida,economia e ordem de um país. Outro ponto importante é o advento do 5G que devido a sua alta latência proporciona uma inovação nos dispositivos *Internet of Things* (IOT), permitindo assim uma maior conectividade entre vários dispositivos [3]. Todavia existem sistemas comumente usados para proteger redes como o sistema de detecção de intrusão de rede (NIDS).

Um NIDS é um dispositivo ou *software* que monitora todo o tráfego de rede que passa por um ponto estratégico para o surgimento de atividades maliciosas. Quando a atividade maliciosa é detectada, um alerta é gerado e enviado ao administrador, as redes neurais se tornaram solução cada vez mais eficiente para a implementação de um modelo NIDS. Entretanto uma desvantagem das redes neurais é a quantidade de recursos necessários para treiná-las, ou seja, muitos *gateways* de rede e dispositivos roteadores, que poderiam potencialmente hospedar um NIDS, simplesmente não possuem memória ou capacidade de processamento para treinar e executar os modelos. Além disso, as soluções de redes neurais existentes são treinadas de forma supervisionada, o que significa que um especialista deve rotular o tráfego da rede e atualizar o modelo manualmente de tempos em tempos, o que torna redes neurais supervisionadas uma solução não viável.

O Kitsune é um NIDS baseado em aprendizado de máquina, mas ao contrário das demais redes neurais ele é não supervisionado, com funcionamento online e possui uma baixa complexidade de implementação, se tornando uma boa opção para lidar com a detecção de anomalias em rede se tratando de um NIDS. O objetivo deste artigo é apresentar o funcionamento e como é realizada a análise dos resultados gerados pelo Kitsune.

Metodologia

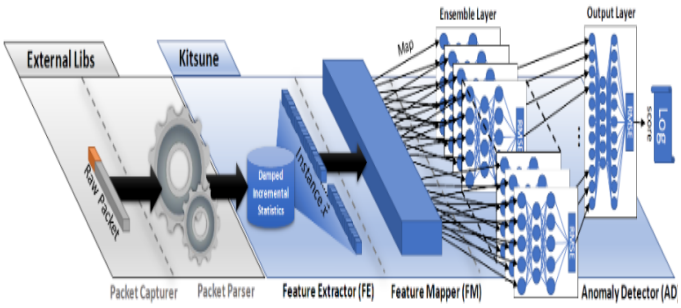
O Kitsune possui um conjunto de pequenas redes neurais (*autoencoders*), que são treinadas para imitar (reconstruir) padrões de tráfego de rede e cujo desempenho melhora gradativamente com o passar do tempo. Os *autoencoders* são um tipo especial de rede neural usada para copiar a saída para entrada, ele pega uma entrada codifica os dados em um representação latente de menor dimensão e a partir dessa representação ele codifica para a dimensão original.

O Kitsune é baseado em um sistema de aprendizagem de máquina não supervisionado, ou seja, os dados inseridos para treinamento não são rotulados por um ser humano, é a própria rede neural que entende e rotula os dados. Esses dados são agrupados através de uma técnica de *clustering*, isto significa que ela se baseia na semelhança ou diferença entre os dados agrupados. Um bom exemplo disso seria uma empresa que estuda a segmentação de mercado, aplicando um algoritmo de clusterização que atribui valores semelhantes em grupos correspondentes ao conjunto de parâmetros. Esse agrupamento pode ser baseado em localização, índice de renda ou outras variáveis.

Outra vantagem do Kitsune é que ele trabalha com processamento online, ou seja, o Kitsune processa um pacote por vez e não mantém os pacotes processados na memória, gastando assim menos hardware que outras redes neurais.Sua eficiência pode ser dimensionada com seu parâmetro de entrada $\boldsymbol{m}$: onde $\boldsymbol{m}$ é o tamanho máximo de qualquer autoencoder na camada ensemble (os *autoencoders* menores são exponencialmente mais baratos para treinar e executar), o modelo que representa o funcionamento do kitsune com as respectivas partes

que o formam está presente na Figura 1

Figura 1-Representação do modelo Kitsune, Fonte 4



O Packet Capture é o primeiro componente presente no modelo da Figura 1, ele é responsável por capturar os pacotes da rede e envia para o *packet parser*.

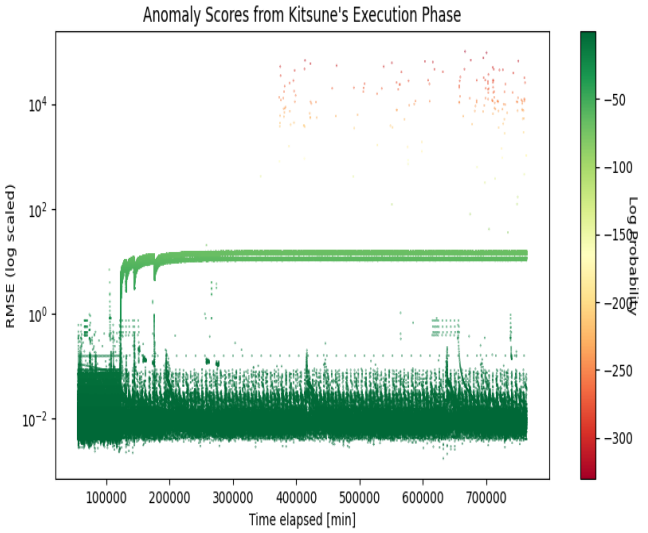
O por sua vez *Packet Parser* é responsável por extrair informações dos pacotes capturados e envia-las para o *Feature Extractor* (FE), este por sua vez recebe os dados extraídos do *Packet Parser*, armazena as estatísticas e a partir da mesma gera um vetor com 115 features que então é passado para o *Feature Mapper* (FM).

Após receber as estatísticas geradas pelo *Feature Mapper* é realizado a clusterização dos elementos de forma que eles sejam agrupados em **K** *clusters* com cada um contendo no máximo **m** *features.*

O valor de **m** é definido pelo usuário e indica o número máximo de neurônios da camada de entrada de cada autoencoder da camada composta. O *Anomaly Detector* (AD) é quem detecta os pacotes anômalos com base na representação dos **K** *clusters gerados pelo feature Mapper*.

A implementação do Kitsune é feita em Python [4], primeiramente é preciso capturar os pacotes presente em um fluxo de rede, para isso é necessário usar o s*oftware* Wireshark, mas pode ser feita através de qualquer ferramenta de captura de fluxo de pacotes em rede, desde que o arquivo gerado dessa captura seja no formato pcap ou tsv, após isso é preciso instanciar um objeto da classe Kitsune, em seguida se define o caminho para o arquivo gerado pela captura, o valor de **m** e a quantidade de pacotes usados no treino. Finalizando então na execução do método proc_next_packet() em *loop*. Após isso é possível avaliar os valores *Root Mean Squared Error* (RSME) retornado através de um gráfico gerado pelo kitsune representado na Figura 2.

Figura 2 - Gráfico gerado pelo Kitsune



Conclusão e Resultados

O gráfico presente na Figura 2 se refere ao resultado da análise de um arquivo de fluxo de rede não malicioso no formato pcap realizado pelo Kitsune, o eixo *y* a esquerda se refere ao erro quadrático médio (RMSE),quanto maior o RMSE, maior o erro na reconstrução dos dados, além disso, no eixo x temos a escala de tempo em minutos referente ao tempo total da captura dos pacotes presente no fluxo de rede. No eixo *y* a esquerda há uma barra com tons de cores que variam do vermelho até o verde, quanto mais vermelho, mais anômalo é o pacote e quanto mais verde menos anômalo. Analisando o gráfico da Figura 2 é possível notar uma diferença entre a tonalidade dos verdes, isso significa que alguns pacotes são mais anômalos que outros, mas não necessariamente se trata de um ataque malicioso, sendo assim, é possível analisar os resultados obtidos através da execução do Kitsune. A coleta do fluxo de rede testados foram todos não anômalos ou pouco anômalos, não foram testados pacotes puramente anômalos. O kitsune se mostrou um software de baixa complexidade e fácil implementação, além de possuir uma representação objetiva dos resultados gerados e alcançar os objetivos ao qual foi proposto [4]

Agradecimentos

Agradecimentos à Universidade Federal de Itajubá (UNIFEI) pela bolsa ofertada através do Programa Institucional de Bolsa de Iniciação Científica (PIBIC - UNIFEI).

Referências

J. Luz, “Qual é a real importância da cibersegurança?.”https://exame.com/bussola/qual-e-a-real-importancia-da-ciberseguranca/. Acesso em: 14 de setembro de 2023, 2023.

A. Press, “Ataque hacker interrompe funcionamento de hospitais nos estados unidos e fbi abre investigação.” https://g1.globo.com/tecnologia/noticia/2023/08/10/ataque-˜hacker-interrompe-funcionamento-de-hospitais-nos-estados-unidos-e-fbi-abreinvestigacao.ghtml. Acesso em: 13 de setembro de 2023, 2018.

P. . A. Por Di Blasi, “Iot e 5g: Avanços e perspectivas da implementação.” ˜https://diblasiparente.com.br/iot-e-5g-avancos-e-perspectivas-da-implementacao/. Acesso em: 12 de setembro de 2023, 2022.

I.Mirsky,“Kitsune.py.”https://github.com/ymirsky/Kitsune-py. Acesso em: 1 de setembro de 2023, 2020.