RT: PESC 776 08/2021

Universidade Federal do Rio de Janeiro

IPTraf: Coleta e Detecção de Anomalias em Fluxos de Rede

Autores
Felipe M. F. de Assis
Marco A. Coutinho
José B. da Silva Filho
Evandro L. C. Macedo
Luís F. M. de Moraes

Rio de Janeiro, 24 de agosto de 2021

Sumário

Resumo

Considerando o crescimento acelerado do já enorme número de dispositivos conectados à Internet, a segurança das redes de computadores torna-se cada vez mais importante. Este relatório aborda a plataforma IPTraf – uma ferramenta projetada para coletar dados de fluxos que compõem o tráfego em redes IP – e sua aplicação na identificação de anomalias. A arquitetura da plataforma em questão é apresentada juntamente com resultados obtidos a partir dos fluxos coletados nos enlaces de borda da Rede-Rio/FAPERJ, um Sistema Autônomo que compõe a rede acadêmica e de pesquisa do Estado do Rio de Janeiro. A utilidade da plataforma apresentada, bem como dos resultados obtidos com os dados coletados, é evidenciada a partir das anomalias identificadas.

Abstract

Considering the rapid growth of the already huge number of Internet-connected devices, the computer networks security become rather important. This report tackles the IPTraf platform – a tool designed to collect data of flows that constitute traffic in IP networks – and its application on anomaly identification. The architecture of such platform is presented along with the results obtained from flows collected from edge links of Rede-Rio/FAPERJ, an autonomous system that composes the academic research network in the State of Rio de Janeiro. The usefulness of the presented platform, as well as the results obtained with the collected data, is thus evidenced from the identified anomalies.

Palavras-chave: *anomalias de tráfego, fluxos em redes IP, segurança de redes.*

1 Introdução

Redes de Interconexão (*backbones*) são redes de extrema importância para o funcionamento da Internet, visto que estas proporcionam a conectividade necessária à interligação das redes mundialmente difundidas. Os *backbones* apresentam proporções de grande escala com elevado desempenho, funcionando como meio de acesso para conexão de redes de escalas menores. Um exemplo dessas redes de interconexão é a Rede-Rio/FAPERJ, um *backbone IP* situado no estado do Rio de Janeiro, que realiza atividades à serviço da ciência, tecnologia e educação. Hoje a capacidade de transmissão do núcleo da Rede-Rio já alcança 10 Gbps e mantém uma conexão com a rede acadêmica brasileira, a Rede Nacional de Ensino e Pesquisa (RNP), e com o Ponto de Troca de Tráfego, o *Internet Exchange* do Brasil (IX.br). A Figura 1 ilustra a topologia atual da Rede-Rio/FAPERJ.

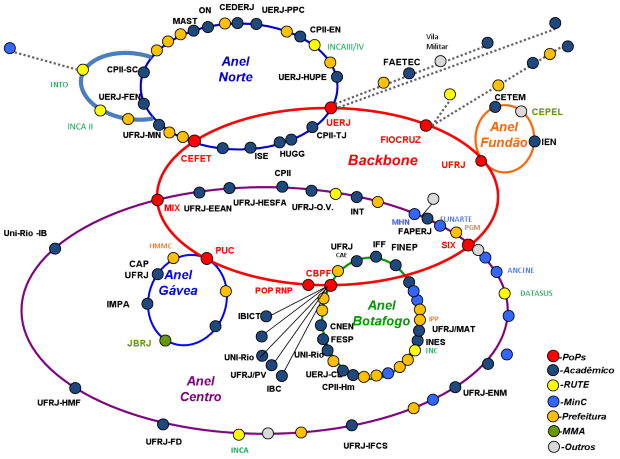


Figura 1: Topologia da Rede-Rio [Rede-Rio 2021]

Com a enorme quantidade de dispositivos conectados prevista para os próximos anos [Ericsson 2019], juntamente com o fato dos princípios básicos de segurança geralmente não serem observados com a devida preocupação, novos fatores de risco são potencialmente gerados. Além disso, o aumento da superfície de ataque para as redes em operação também se torna uma realidade. Como exemplo, ataques de DDoS tendem a crescer cada vez mais, explorando vulnerabilidades já conhecidas e causando instabilidade na rede com tráfegos de mais de 1 Tbps [Antonakakis et al. 2017]. Dessa maneira, é de fundamental importância para o gerenciamento e planejamento futuro das redes e o bom funcionamento do núcleo da rede, o desenvolvimento de uma ferramenta que, além de identificar anomalias de tráfego, também o caracterize e classifique, o que ainda permanece um desafio. Na literatura encontram-se uma série de aplicações de análise e detecção de anomalias que

são baseadas em aprendizado de máquina [Omar et al. 2013], decomposição de tensores [Streit et al. 2020], séries temporais [Wu and Shao 2005], e ainda através de telemetria de rede considerando o uso de Redes Definidas por Software (*Software-Defined Networking* – SDN) [Paolucci et al. 2018, Ribeiro et al. 2021, Giotis et al. 2014].

Este relatório apresenta a plataforma IPTraf, que é composta por ferramentas de coleta, monitoramento contínuo e ferramentas de análise de anomalias identificadas a partir dos dados coletados. São consideradas abordagens baseadas em séries temporais, baseadas em aprendizado de máquina e decomposição de tensores. Apresentamos propostas de detecção, bem como alguns resultados numéricos obtidos das técnicas que já possuem implementação utilizando os dados coletados a partir dos enlaces de borda da Rede-Rio/FAPERJ.

O restante do relatório está organizado da seguinte maneira. A Seção 2 comenta os trabalhos relacionados ao tema do relatório. A Seção 3 apresenta a plataforma IPTraf e as ferramentas que a compõe. A validação da plataforma é apresentada na Seção 4. Os resultados das detecções de anomalias utilizando outras abordagens propostas são mostrados na Seção 5. Por fim, a Seção 6 conclui o relatório e aponta os trabalhos futuros.

2 Trabalhos Relacionados

A literatura existente afirma que o ramo de detecções de anomalias tem aplicações em diversas áreas [Chandola et al. 2009]. No contexto de redes, o principal uso se encontra em Sistemas de Detecção de Intrusões. Um Sistema de Detecção de Intrusões (IDS) pode ser dividido em diferentes bases para a detecção: por assinatura, anomalia ou especificação ([Liao et al. 2013], [Fernandes et al. 2019]). A detecção por assinatura necessita de uma base de dados com descrição do que é considerado normal e o que é um comportamento anômalo, que é usada para comparação com novos fluxos e sua consequente classificação. O maior desafio na detecção por assinatura se encontra na criação da base de assinaturas, que pode ser feita de diferentes formas, como pela abstração em eventos [Kumar and Spafford 1994] ou com técnicas de mineração de dados [Zhengbing et al. 2008]. Nos casos em que se utiliza detecção por especificação, é necessário que se especifique manualmente os parâmetros, como estudado por [Uppuluri and Sekar 2001].

A detecção com base em anomalia se mostra mais versátil no sentido de não ser dependente de ataques e padrões já conhecidos. Além de atuar como IDS, também é uma técnica para monitoramento e análise da rede. Em seu trabalho, [Ye et al. 2000] desenvolveram uma técnica de detecção de anomalia com base no modelo probabilístico de Cadeias de Markov. Já [Bereziński

et al. 2015] trabalharam em um método com base no conceito de entropia da Teoria da Informação de forma a descobrir padrões anômalos e potencialmente maliciosos em uma rede. Na pesquisa de [Barford et al. 2002], é construída uma ferramenta de detecção de anormalidades com base na análise de sinais das características de tempo e frequência do fluxo IP e dos dados SNMP. Em seu trabalho, [He et al. 2008] criam uma ferramenta para estimar largura de banda disponível entre nós da rede e, subsequentemente, um método para detectar tráfego anômalo por meio desta ferramenta. [Bartos et al. 2011], em sua pesquisa, desenvolveram uma forma de otimizar a amostragem de fluxos para minimizar seu impacto negativo na detecção de anomalias.

O uso de aprendizado de máquina para a detecção de anomalias se mostra presente no estado da arte. [Hyun Oh and Suk Lee 2003] utilizam métodos de agrupamento para reunir padrões de tráfego comum de usuários, de forma a detectar anomalia fora destes agrupamentos. Já [Pajouh et al. 2017] se utiliza da técnica do Classificador de Bayes Ingênuo para separar o tráfego normal do anormal, e complementa o processamento com um Algoritmo de K vizinhos mais próximos para refinar decisões sobre o tráfego normal.

É evidente a existência de diversas ferramentas para a detecção de comportamentos anômalos em redes. A ferramenta aqui apresentada se diferencia primeiramente por não ser de uso exclusivo como IDS. A ferramenta fornece dados para análise e previsão do comportamento da rede, assim como possibilita a distinção de diferentes níveis de anomalia. Além disso, é dividida em módulos, que atuam de forma paralela para prover maior tolerância a falhas, isto é, cada módulo tem capacidade independente de detecção de comportamento anormal. Além disso, por ser modularizável, a ferramenta é expansível de forma que se pode integrar novas funções e métodos.

3 O Projeto IPTraf

O IPTraf é uma plataforma desenvolvida para servir como fonte de consulta sobre as características dos fluxos de rede em *backbones IP*. A arquitetura da plataforma é composta por um Módulo Coletor e as ferramentas de análise associadas, conforme é ilustrado na Figura 2. Os roteadores de borda da Rede-Rio/FAPERJ exportam os fluxos coletados para o Módulo Coletor no servidor IPTraf.

Os fluxos são processados gerando metadados que são disponibilizados para as ferramentas de análise de anomalias. Por fim, o Sistema Guardião integra as ferramentas de detecção de anomalia, fornecendo alertas e uma interface única de monitoramento.

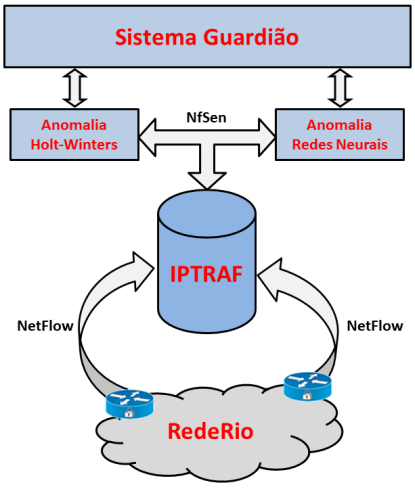


Figura 2: Arquitetura da Plataforma IPTraf

3.1 Módulo Coletor

O Módulo Coletor funciona a partir de uma infraestrutura montada para a coleta de dados efetuada na Rede-Rio/FAPERJ, através do pacote de ferramentas *NFDUMP*. Os fluxos de rede originados do roteador de borda da Rede-Rio/FAPERJ são capturados e remetidos para o Laboratório Ravel da COPPE/UFRJ, onde se encontra o Coletor.

O Módulo Coletor fornece uma interface de monitoramento através da qual é possível acompanhar e filtrar os endereços IPs dos fluxos recebidos, além de outras informações contidas em tais fluxos. Na configuração atual, são extraídos de cada fluxo: endereço de origem, endereço de destino, porta de origem, porta de destino, protocolo, flags tcp, número de pacotes, número de bytes e horário de início. Desta forma, características mais detalhadas sobre as anomalias detectadas pelas ferramentas de análise podem ser investigadas com mais precisão.

3.2 Ferramentas de Análise de Anomalias

Nesta seção são apresentadas as técnicas de detecção de anomalias em fluxos de redes desenvolvidas na plataforma IPTraf.

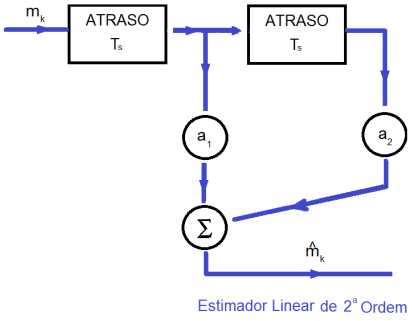
3.2.1 Detecção de Anomalias Baseada em RNA-Perceptron

Com base em [da Silva Filho 2015], um módulo de identificação de anomalias por Rede Neural Artificial é elaborado utilizando os resultados do aproximador linear definido em [Lathi 1998] e a generalização definida em

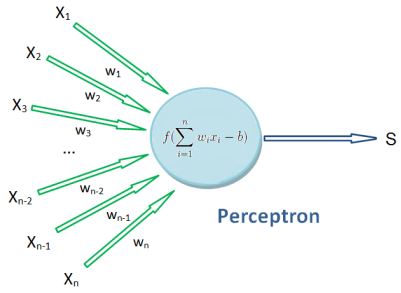
[Scalassara 2005], com base nas seguintes premissas:

- Os fluxos de redes podem ser descritos como processos estocásticos;
- O tráfego pode ser modelado como resultado do somatório de fluxos de dados;
- O tráfego e seus fluxos componentes, podem ser modelados com funções matemáticas discretas por uma transformação conveniente.

Desta forma, o tráfego (formado a partir dos respectivos fluxos), sendo considerado de natureza estocástica e estacionários em sentido amplo, são estimados pelo modelo auto-regressivo (AR), como definidos e estudados nos trabalhos de [Aguirre 2007], [Macedo 2015] e [Piltan et al. 2017]). A partir desta hipótese, admite-se estimar os fluxos de dados pelo Estimador Discreto de Segunda Ordem (Figura 3(a)), instrumentalizado como uma RNA-MLP (Figura 3(b)). A literatura na área de redes apresenta trabalhos utilizando o modelo AR para a identificação de anomalias de tráfego, como [Celenk et al. 2010].



(a) Estimador Linear [Lathi 1998]



(b) Modelo de Neurônio - Perceptron

Figura 3: Estimador e Rede Perceptron

Neste módulo, o processo de análise foi definido em três fases distintas:

Captura e Extração de Dados: O extrator de dados tem a função de totalizar a quantidade parcial de *bytes* trafegados pelas séries discretas que compõem o conjunto de características avaliadas. Para definir estas séries utilizamos a classificação de portas da RFC 6335 [Cotton et al. 2011]: tráfego (x_0), média ponderada das portas (x_1), portas baixas (x_2), portas do servidor (x_3), portas dinâmicas (x_4), compondo o vetor $\vec{X} = \begin{bmatrix} x_0 & , x_1 & , x_2 & , x_3 & , x_4 \end{bmatrix}^T$. Esta abordagem garante a modularidade e robustez do método, permitindo lidar com o problema clássico do grande esforço computacional exigido nas análises deste tipo e ainda o aumento da eficácia na identificação de anomalias na rede dentro de um intervalo tempo.

Fase de Identificação de Anomalias: Com o modelo de comportamento identificado, podemos estimar valores de qualquer momento passado e futuro. Por definição, um modelo se aproxima do seu sistema real com uma margem de erro. Para avaliar quantitativa e qualitativamente este erro, é importante estabelecer métricas, que serão utilizadas para avaliar a eficiência da identificação das anomalias, conforme descrito na fase seguinte.

Caracterização da Intensidade de Anomalias: Para estabelecer uma métrica de avaliação conjunta dos resultados, considerando o comportamento de desvio estatístico normal verificado em todas as séries discretas, definimos a intensidade de uma anomalia como a soma algébrica dos graus de anomalia observados em cada uma das séries, no mesmo instante de tempo discreto. Logo, se os desvios são classificados no intervalo [0,3], para as cinco séries avaliadas temos a intensidade de uma anomalia variando no intervalo [0,15], sendo zero o indicador de normalidade do sistema. Define-se os intervalos de intensidade: [1,4] como **Moderada**, [5,9] como **Forte** e [10,15] como **Muito Forte**.

3.2.2 Detecção de Anomalias Baseada em Holt-Winters

Este módulo utiliza uma técnica de análise baseada na construção de séries temporais e na comparação dos valores medidos e de previsões obtidas com o modelo de Holt-Winters [da Silva 2015].

Metodologia: O método consiste em um Sistema de Detecção de Intrusão com detecção baseada em anomalias que tem como principal objetivo a detecção de ataques de negação de serviço e de escaneamento de porta. O método também mantém a capacidade de identificação de anomalias mais genéricas. Pode-se entender o método como um processo de quatro etapas: a coleta, a classificação, a obtenção de métricas e a análise.

Coleta: os dados utilizados nesta ferramenta são obtidos a partir do Módulo Coletor da plataforma IPTraf. Desta forma, este módulo de detecção também se vale dos dados de fluxos de rede coletados a partir dos enlaces de borda da Rede-Rio/FAPERJ.

Classificação: a classificação tem por objetivo facilitar a detecção de anomalias através do isolamento dos eventos anômalos. A suposição feita ao desenvolver a classificação é que seja factível realizar uma partição do conjunto total de fluxos coletados de tal forma que uma determinada anomalia tenha todos os seus fluxos contidos dentro de apenas um dos conjuntos da partição.

Métricas: a cada cinco minutos o método observa todos os fluxos que foram recebidos nesse espaço de tempo e extrai as seguintes métricas de interesse: TOTAL DE BYTES, TOTAL DE PACOTES, SOCKET DE DESTINO

e PORTAS DE DESTINO.

Análise: com as séries temporais definidas, a análise destas é realizada com o objetivo de caracterizar seus respectivos comportamentos como anômalo ou normal. A análise proposta não tem por objetivo a caracterização da série como um todo, mas sim dos valores em um determinado instante de tempo, ou seja, não se busca caracterizar a série X mas sim o valor $X(t)$ para qualquer instante de tempo t, aplicando o Modelo de Previsão de Holt-Winters [Chatfield and Yar 1988].

3.2.3 Ferramenta Guardião-Anomalia

A ferramenta Guardião-Anomalia, implementa uma metodologia automatizada e integrada, para identificação de eventos anômalos (em relação aos fluxos de rede observados) e possíveis atividades maliciosas em tempo hábil de mitigação. Desta maneira, ações de mitigação podem ser executadas dentro da janela de 5 minutos (de processamento) da ocorrência do evento. A janela de 5 minutos está relacionada às características de coleta dos roteadores de borda. Uma janela menor pode ser estabelecida, caso se faça necessário, porém é necessário considerar o *overhead* de processamento, inclusive nos equipamentos de roteamento, visto que o volume de fluxos coletados será maior.

O Sistema Guardião é composto pelos dois módulos de detecção de anomalias apresentados anteriormente, conforme ilustrado na Figura 2. Tal sistema tem por objetivo integrar os diferentes módulos de detecção de anomalias para compor um único sistema capaz de gerar alertas de anomalias. Tais alertas servem para auxiliar nas tomadas de decisões por administradores e operadores de redes, no sentido de identificar e mitigar incidentes que podem comprometer a segurança dos sistemas envolvidos, assim como dados relacionados aos seus respectivos usuários.

4 Resultados de Validação da Plataforma

Como método de avaliação da efetividade na identificação de anomalias, foi estabelecido um ensaio de laboratório comparativo de resultados de ambos os módulos do sistema Guardião-Anomalia através de uma análise de cruzamento dos dados.

Neste ensaio, consideram-se as seguintes premissas: O intervalo de análise em ambos os trabalhos coincidem, a granularidade de fluxos padrão é fixada em intervalos de 05 (cinco) minutos, foram consideradas na análise apenas as anomalias de intensidade **Forte** e **Muito Forte**, foram disseminados ataques,

de forma aleatória, criando anomalias adicionais de ciberataque na Rede-Rio, definido como um grupo de controle.

Os conjuntos das anomalias identificadas são:

- IA: técnica *RNA-Perceptron*
- ST: técnica de *Holt-Winters*;
- N_{IA}: por *Holt-Winters* e não por *RNA-Perceptron*, $\{(ST) - (IA) \cup (ST)\}$;
- N_{ST}: por *RNA-Perceptron* e não por *Holt-Winters*, $\{(IA) - (IA) \cup (ST)\}$.

Definimos a efetividade como $\mathbf{Efetividade(X)} = \frac{n(X)}{n(AtaquesSimulados)} \times 100$, com $\mathbf{n(X)}$ sendo o número total de elementos no conjunto X. Os resultados consolidados são apresentados na Tabela 1, onde se verifica que há uma identificação positiva simultânea por ambas as técnicas de 63,64 % ($n((IA) \cap (ST))$), das anomalias do período; considerando o conjunto total de 330 anomalias identificadas por ambos os métodos ($n((IA) \cup (ST))$). Destaca-se ainda a maior efetividade da técnica *RNA-Perceptron* (81,82%) quando comparada a técnica de *Holt-Winters* (72,73%) , que pode ser explicada pela abordagem de cada técnica, enquanto a primeira gera o modelo de referência a partir de um período longo (30 dias de fluxos anteriores) a segunda obtém seu modelo de uma janelas definidas a partir da última 01 hora, eventualmente podendo ser estendida até as últimas 24 horas.

Tabela 1: Análise de Anomalias

Conjunto Referenciado	**Anomalia Grave**
Efetividade(IA) %	81,82%
Efetividade(ST) %)	72,73%
Efetividade($(IA) \cap (ST)$) %	63,64%
Efetividade($(IA) \cup (ST)$) %	90,91%

Avaliando os dados obtidos para o período, encontra-se dois intervalos onde há um aumento de 20,91% na intensidade das anomalias identificadas. O primeiro intervalo é definido em 24/07/2015 e 25/07/2015 a e o segundo definido entre 31/07/2015 e 01/08/2015. Após confirmação positiva do período coincidente com os ataques do grupo de controle, os quais realmente foram disseminados nas datas em questão, identifica-se um conjunto de dez anomalias-ataques do grupo de controle, e um falso negativo.

5 Outras Abordagens de Detecção Propostas

Nesta seção são apresentadas outras abordagens de detecção de anomalias propostas para a plataforma IPTraf.

5.1 Algoritmos de Aprendizado de Máquina

Os algoritmos capazes de "aprender" novos padrões podem ser divididos em dois tipos básicos: os algoritmos de aprendizado supervisionado e os não-supervisionados [Bishop 2006]. O primeiro tipo deve inicialmente passar pela etapa de treinamento a partir de amostras reais de dados. Com este ciclo de treinamento, os parâmetros e pesos adequados são definidos para a correta execução de classificação ou previsão de um conjunto de dados. A vantagem deste tipo de algoritmo é a capacidade de utilizar o histórico para aumentar a precisão e acurácia dos dados computados. Além disso, são sistematicamente mais rápidos para processamento de grandes quantidades de dados. O segundo tipo (não-supervisionado) não depende de uma estrutura inicial de treinamento e podem obter saídas de dados somente a partir do fluxo real de dados de entrada, sem histórico. Estes algoritmos são ideais para a identificação de estruturas e relacionamento dos dados ainda não mapeados, muito comumente utilizados para o agrupamento de valores que possuam alguma característica em comum (clusterização).

Um modelo matemático típico do processo de aprendizagem é explicado por Yaser em [Abu-Mostafa et al. 2012]. Ele descreve matematicamente os elementos necessários que um processo de aprendizagem deve conter afim de obter os resultados esperados no processo de classificação e regressão de variáveis.

De acordo com a literatura ([Liu et al. 2018], [Dasgupta et al. 2020], [Hamed et al. 2018], [Mishra et al. 2019], [Neto and Gomes 2019]), alguns algoritmos foram selecionados para testes visando identificar o melhor desempenho de classificação das anomalias. Os métodos *Random Forest*, *K-Means*, Redes Bayesianas e *Latent Dirichlet Allocation* (LDA) são descritos nas seções seguintes. Segundo Bishop [Bishop 2006], os desafios para implementação de tais algoritmos são escalabilidade, desempenho (velocidade de detecção), aprendizado automático de novas classes e precisão.

5.1.1 Random Forest

É um algoritmo de aprendizagem supervisionada que cria uma floresta de árvores de decisão de maneira aleatória. A floresta criada é treinada na maioria dos casos com o método de *bagging*. A ideia principal do método de *bagging* é

a melhoria dos resultados através da combinação de modelos de aprendizado. O algoritmo de Random Forest (RF) introduzido por [Breiman 2001] é um termo geral para métodos de combinação (ensemble) utilizando classificadores do tipo árvore. O RF constrói uma grande quantidade de árvores de decisão para fora do sub-conjunto de dados a partir de um treinamento único definido.

O algoritmo Random Forest tem se mostrado ser um classificador de uso geral eficiente e pode ser empregado com sucesso em aplicações de IDS, como é o caso da plataforma IPTraf, para aumentar a acurácia do sistema e diminuir o número de falso positivos/negativos [Almseidin et al. 2017].

5.1.2 K-Means

K-Means é um algoritmo de agrupamento [Liu et al. 2018] que pode ser descrito como uma solução de otimização, conforme

$$\mathcal{U}^* = \underset{U \in \mathcal{P}}{\arg\min} \max_{c_k \in \chi} \sum_{k=1}^{C} \sum_{i=1}^{n} U_{k,i} ||c_k - \vec{x}_i||_d^2$$

para alguma distância $|| \cdot ||_d$ pertencente a matriz U (a qual associa a cada cluster $\vec{x}$, c_k é o k^{th} centroide de cluster e $\vec{x}_j$ é o j^{th} ponto dos dados pertencentes ao cluster. Basicamente, $U_{i,k}$ é igual a 1 se $\vec{x}_i$ estiver no k^{th} cluster e será 0, caso não esteja. A equação computa a distância total de cada ponto a centroide de seu cluster.

Com técnicas de *clustering*, a eficiência geral do IDS pode ser melhorada dividindo as amostras de fluxos e classificando os diversos centróides possíveis desde "ataque"/"não ataque", até a geração de asssinaturas digitais de *malwares* específicos.

5.1.3 Redes Bayesianas

As redes bayesianas são utilizadas para a representação da probabilidade condicional entre variáveis em análise [Neapolitan 2003]. Com sucesso, pode-se estimar o grau de certeza a respeito de uma variável antes de observar o seu estado, utilizando este conhecimento *a priori* para estimar com mais precisão o modelo a ser adotado para a representação matemática desta variável.

A propagação da crença, segundo Pearl [Pearl 2009], é obtida com a atualização do modelo adotado para descrever a probabilidade *a priori* a partir dos diversos eventos observados relacionados ao estado das variáveis em estudo. Para tanto, é utilizado o algoritmo de Pearl para o cálculo da atualização de crença dado o novo estado de alguma variável na rede bayesiana.

Redes bayesianas, quando utilizadas para classificação (classificador Naïve Bayes), é uma boa opção para quando não existe um conhecimento prévio acerca da base analisada, atribuindo uma probabilidade equivalente a todas as classes. Mas, a partir dos primeiros fluxos analisados é possível treinar os modelos mais complexos com base nos algoritmos anteriormente mencionados, o que aumenta a eficiência do classificador.

5.2 Detecção por Decomposição de Tensores x Regressão Linear

Por hipótese, definimos que o modelo de representação do espaço de fluxos será obtido utilizando a ferramenta *XGBoost* para o modelo de regressão linear e a ferramenta *TensorFlow* para aproximação do modelo baseado em *RNA* [Abadi and et al 2016,Dillon et al. 2017]. Os modelos devem ser robustos o suficiente para que a técnica de identificação possa avaliar as anomalias de formato de dados (distorção de esquema) e anomalias de distribuição (distorção de valores de dados) [Ribeiro et al. 2018, Dong et al. 2020].

O treinamento dos modelos seguiu a definição da Seção 3.2.1, onde o espaço foi particionado em séries temporais, agrupadas de forma conveniente seguindo a RFC6335 da IANA [Cotton et al. 2011]. Consideramos o intervalo de 20 de abril de 2021 até 15 de junho de 2021, sendo os fluxos agrupados em intervalos discretos de 5 minutos e divididos em subconjuntos para treinamento (60%), validação (20%) e teste (20%).

Por sua característica de aprendizado não supervisionado, inicialmente utilizaremos a técnica de regressão linear sobre as séries, utilizando o Erro Médio Percentual, sendo este definido como:

$$\mathbf{E(\%)} = \frac{\sum_{i=0}^{amostras} \left(\frac{\hat{y_i} - y_i}{y_i}\right)}{amostras} \times 100$$

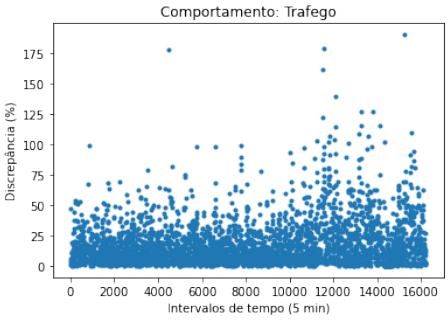
onde $\hat{y_i}$ é o valor estimado da série e y_i é o valor real da série, no intervalo discreto avaliado.

Neste estudo, observamos que o Erro Médio Percentual ficou abaixo de 0,5%, em todas as séries avaliadas, o que caracteriza o modelo obtido como uma ferramenta de estimação precisa e acurada. Os resultados consolidados podem ser vistos na Tabela 2, onde todas as séries convergem apontando anomalias entre 20/04/2021 e 01/05/2021.

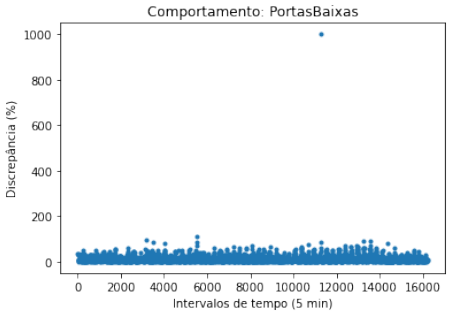
Considerando ainda a regressão linear, observamos os resultados apresentados na Figura 4, onde podemos identificar, para cada uma das séries temporais, a discrepância entre o valor observado e o valor estimado pelo modelo, na forma percentual $\Delta(\%) = (\frac{\hat{y} - y}{y}) \times 100$. Adotou-se o limiar de 30%, acima do qual é caracterizada uma anomalia na série em análise.

Série	**Erro Médio (%)**	**Anomalias Identificadas**	**Período**
Tráfego	0,19%	587	20/04/2021 03:50h até 01/05/2021 10:25h
Portas Baixas	0,14%	287	20/04/2021 05:00h até 01/05/2021 10:25h
Portas de Serviços	0,39%	1491	20/04/2021 04:10h até 01/05/2021 10:30h
Portas Dinâmicas	0,41%	1413	20/04/2021 04:25h até 01/05/2021 10:30h

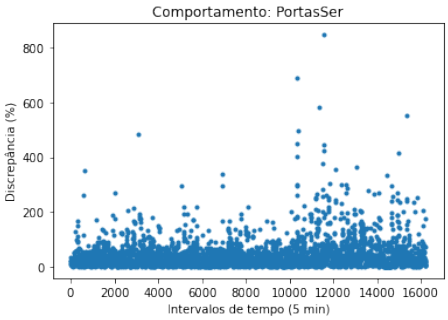
Tabela 2: Regressão linear sobre as séries de fluxos



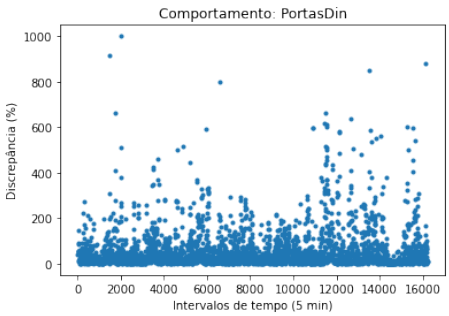
(a) Série temporal do tráfego



(b) Série temporal de portas baixas



(c) Série temporal de portas de serviços



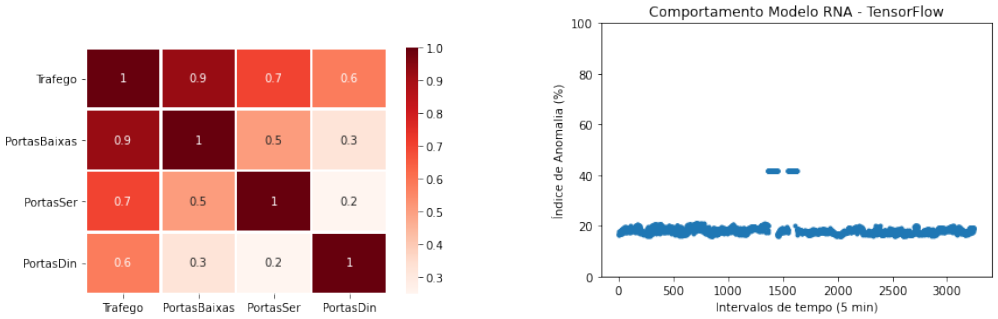
(d) Série temporal de portas dinâmicas

Figura 4: Discrepância em percentual entre o valor estimado e o observado dos fluxos para cada uma das séries temporais usando o modelo de regressão linear

Contrapondo a regressão linear, o modelo *RNA-TensorFlow* possui como característica básica o aprendizado supervisionado. Utilizamos os resultados (anomalias identificadas pela regressão) como exemplos para o treinamento da RNA. Desta forma, o modelo adotado foi definido com quatro entradas (uma por série temporal), uma camada oculta de cinco neurônios e uma saída

que representa o índice de anomalia identificado em percentual. O método de otimização utilizado foi o *Adam algorithm*, o qual convergiu com uma acurácia de 75,73%. Cabe ressaltar que estamos expandindo a técnica de *RNA-Perceptron* (ver subseção 3.2.1), propondo uma rede neural única e não uma por série. Comparamos a capacidade de identificação desta com proposta *RNA-Perceptron* e o resultado apresentado demonstra que o modelo *RNA-TensorFlow* apresentou 24,27% de falsos negativos, ou seja, anomalias que seriam identificadas pela *RNA-Perceptron* e não pela *RNA-TensorFlow*.

Na Figura 5(a), pode ser avaliado o "mapa de calor", ferramenta que permite a inspeção da correlação entre as séries. Esta inspeção permite avaliar se alguma série (ou grupo de séries) é mais significativa na predição que as demais, sugerindo ajustes no modelo. Como pode ser visto, há uma discrepância entre as séries, o que indica que será necessário utilizar ajustes ou ponderação por pesos aplicados às entradas. Desta maneira, todas as séries terão a mesma influência no treinamento e identificação.



(a) Mapa de Calor - Séries

(b) Anomalias identificadas

Figura 5: Anomalias em Modelo RNA com TensorFlow

A identificação de anomalias foi realizada com o uso da massa de testes. A Figura 5(b) apresenta todo espaço estimado, dos quais destacamos como resultado relevante um total de 164 anomalias, identificadas pela RNA, no período de 01/05/2021 às 10:35:00h até 04/06/2021 às 05:50:00h. Cabe ressaltar que, nesta avaliação, são os destacados os casos nos quais a predição (saída da rede neural artificial) apresentou mais de 30% de probabilidade de ser uma anomalia quando RNA é ativada pela massa de teste.

6 Conclusão e Trabalhos Futuros

Neste relatório apresentamos a plataforma IPTraf que é composta por ferramentas de coleta de dados de fluxos de redes, e suas aplicações de análise e detecção de anomalias em tais fluxos de rede. Diversos tipos de anomalias são

investigados, em especial as que potencialmente podem se tornar ciberataques, a exemplo do escaneamento de portas que pode ser uma anomalia preliminar a uma negação de serviço, *Low-Rate Denial of Service* (LDoS) ou sequestro de dados (*ransomware*).

A validação da ferramenta é feita utilizando dados reais de tráfego de rede obtidos a partir da infraestrutura da Rede-Rio/FAPERJ e novas abordagens vêm sendo desenvolvidas para a melhoria contínua da plataforma. Os resultados de algumas das novas abordagens se mostram promissores (Seção 5). Como trabalhos futuros, vislumbramos o desenvolvimento, a implementação e avaliação de tais módulos, permitindo realizar uma discussão comparativa das técnicas atuais (*Holt-Winters* e *RNA-Perceptron*) com as novas abordagens, em especial a *RNA-TensorFlow*.

7 Agradecimentos

Os autores agradecem o apoio financeiro oferecido pela Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) e Rede-Rio (rede de backbone acadêmico estadual) no âmbito da bolsa 150.134/2010 e também pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código Financeiro 001.

Referências

[Abadi and et al 2016] Abadi, M. and et al (2016). Tensorflow: A system for large-scale machine learning. In *12th USENIX OSDI 16*, pages 265–283, Savannah, GA. USENIX Association.

[Abu-Mostafa et al. 2012] Abu-Mostafa, Y. S., Magdon-Ismail, M., and Lin, H.-T. (2012). *Learning From Data*. AMLBook.

[Aguirre 2007] Aguirre, L. A. (2007). *Introdução à Identificação de Sistemas: Técnicas Lineares e Não-Lineares Aplicadas a Sistemas Reais*. Editora UFMG, Belo Horizonte, MG, Brasil, 3 edition.

[Almseidin et al. 2017] Almseidin, M., Alzubi, M., Kovacs, S., and Alkasassbeh, M. (2017). Evaluation of machine learning algorithms for intrusion detection system. In *IEEE SISY 2017*, pages 000277–000282.

[Antonakakis et al. 2017] Antonakakis, M., April, T., Bailey, M., Bernhard, M., Bursztein, E., Cochran, J., Durumeric, Z., Halderman, J. A., Invernizzi, L., Kallitsis, M., and et al. (2017). Understanding the mirai botnet. In *Proceedings of the USENIX SEC'17*, SEC'17, page 1093–1110, USA. USENIX Association.

[Barford et al. 2002] Barford, P., Kline, J., Plonka, D., and Ron, A. (2002). A signal analysis of network traffic anomalies. In *Proceedings of the IMW'02*, IMW '02, pages 71–82, New York, NY, USA. ACM.

[Bartos et al. 2011] Bartos, K., Rehak, M., and Krmicek, V. (2011). Optimizing flow sampling for network anomaly detection. In *IWCMC, 2011 7th International*, pages 1304–1309.

[Bereziński et al. 2015] Bereziński, P., Jasiul, B., and Szpyrka, M. (2015). An entropy-based network anomaly detection method. *Entropy*, 17(4):2367–2408.

[Bishop 2006] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg.

[Breiman 2001] Breiman, L. (2001). Random forests. *Mach. Learn.*, 45(1):5–32.

[Celenk et al. 2010] Celenk, M., Conley, T., Willis, J., and Graham, J. (2010). Predictive network anomaly detection and visualization. *Trans. Info. For. Sec.*, 5(2):288–299.

[Chandola et al. 2009] Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):1–58.

[Chatfield and Yar 1988] Chatfield, C. and Yar, M. (1988). Holt-winters forecasting: Some practical issues. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 37(2):129–140.

[Cotton et al. 2011] Cotton, M., Eggert, L., Touch, J., Westerlund, M., and Cheshire, S. (2011). Internet assigned numbers authority (iana) procedures for the management of the service name and transport protocol port number registry. `http://www.ietf.org/rfc/rfc6335.txt`. RFC 6335 (Best Current Practice), capturada em 29/06/2015.

[da Silva 2015] da Silva, V. L. P. (2015). Identificacão de anomalias em fluxos de rede utilizando o método de previsão em séries temporais de holt-winters. Dissertação de mestrado, COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.

[da Silva Filho 2015] da Silva Filho, J. B. (2015). Detecção de anomalias em fluxos de redes de computadores utilizando técnicas de redes neurais e estimadores lineares. Dissertação de mestrado, COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.

[Dasgupta et al. 2020] Dasgupta, D., Akhtar, Z., and Sen, S. (2020). Machine learning in cybersecurity: a comprehensive survey. *The Journal of Defense Modeling and Simulation*, 0(0):1548512920951275.

[Dillon et al. 2017] Dillon, J. V., Langmore, I., Tran, D., Brevdo, E., Vasudevan, S., Moore, D., Patton, B., Alemi, A., Hoffman, M., and Saurous, R. A. (2017). Tensorflow distributions.

[Dong et al. 2020] Dong, W., Huang, Y., Lehane, B., and Ma, G. (2020). Xgboost algorithm-based prediction of concrete electrical resistivity for structural health monitoring. *Automation in Construction*, 114:103155.

[Ericsson 2019] Ericsson (2019). IoT Connections Outlook - Mobility Report.

[Fernandes et al. 2019] Fernandes, G., Rodrigues, J. J., Carvalho, L. F., Al-Muhtadi, J. F., and Proença, M. L. (2019). A comprehensive survey on network anomaly detection. *Telecommunication Systems*, 70(3):447–489.

[Giotis et al. 2014] Giotis, K., Argyropoulos, C., Androulidakis, G., Kalogeras, D., and Maglaris, V. (2014). Combining OpenFlow and sFlow for an effective and scalable anomaly detection and mitigation mechanism on SDN environments. *Computer Networks*, 62:122–136.

[Hamed et al. 2018] Hamed, T., Ernst, J. B., and Kremer, S. C. (2018). A survey and taxonomy on data and pre-processing techniques of intrusion detection systems. *Computer and network security essentials*, pages 113–134.

[He et al. 2008] He, L., Yu, S., and Li, M. (2008). Anomaly detection based on available bandwidth estimation. In *IFIP*, pages 176–183.

[Hyun Oh and Suk Lee 2003] Hyun Oh, S. and Suk Lee, W. (2003). An anomaly intrusion detection method by clustering normal user behavior. *Computers & Security*, 22(7):596–612.

[Kumar and Spafford 1994] Kumar, S. and Spafford, E. H. (1994). A pattern matching model for misuse intrusion detection. In *In Proceedings of the 17th National Computer Security Conference*, pages 11–21.

[Lathi 1998] Lathi, B. P. (1998). *Modern Digital and Analog Communication Systems*. Oxford University Press, Inc., New York, NY, USA, 3 edition.

[Liao et al. 2013] Liao, H.-J., Richard Lin, C.-H., Lin, Y.-C., and Tung, K.-Y. (2013). Intrusion detection system: A comprehensive review. *Journal of Network and Computer Applications*, 36(1):16–24.

[Liu et al. 2018] Liu, Q., Li, P., Zhao, W., Cai, W., Yu, S., and Leung, V. C. M. (2018). A survey on security threats and defensive techniques of machine learning: A data driven view. *IEEE Access*, 6:12103–12117.

[Macedo 2015] Macedo, E. L. C. (2015). Previsão de tráfego em enlaces de redes utilizando séries temporais. Dissertação de mestrado, COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.

[Mishra et al. 2019] Mishra, P., Varadharajan, V., Tupakula, U., and Pilli, E. (2019). A detailed investigation and analysis of using machine learning techniques for intrusion detection. *IEEE Communications Surveys & Tutorials*, 21:686–728.

[Neapolitan 2003] Neapolitan, R. E. (2003). *Learning Bayesian Networks*. Prentice-Hall, Inc., USA.

[Neto and Gomes 2019] Neto, M. S. and Gomes, D. G. (2019). Network intrusion detection systems design: A machine learning approach. In *Anais do XXXVII SBRC*, pages 932–945, Porto Alegre, RS, Brasil. SBC.

[Omar et al. 2013] Omar, S., Ngadi, A., and Jebur, H. H. (2013). Machine learning techniques for anomaly detection: an overview. *International Journal of Computer Applications*, 79(2).

[Pajouh et al. 2017] Pajouh, H. H., Dastghaibyfard, G., and Hashemi, S. (2017). Two-tier network anomaly detection model: a machine learning approach. *Journal of Intelligent Information Systems*, 48(1):61–74.

[Paolucci et al. 2018] Paolucci, F., Sgambelluri, A., Cugini, F., and Castoldi, P. (2018). Network Telemetry Streaming Services in SDN-Based Disaggregated Optical Networks. *Journal of Lightwave Technology*, 36(15):3142–3149.

[Pearl 2009] Pearl, J. (2009). *Causality: Models, Reasoning and Inference*. Cambridge University Press, USA, 2nd edition.

[Piltan et al. 2017] Piltan, F., TayebiHaghighi, S., and Sulaiman, N. B. (2017). Comparative study between arx and armax system identification. *International Journal of Intelligent Systems and Applications (IJISA)*, 9(2):25–34.

[Rede-Rio 2021] Rede-Rio (2021). Rede-Rio/FAPERJ. http://www.rederio.br.

[Ribeiro et al. 2018] Ribeiro, A. C. F., Frazão, R., and Oliveira e Sá, J. (2018). Quebra-cabeças machine learning: Como selecionar use cases, algoritmos e tecnologias?

[Ribeiro et al. 2021] Ribeiro, A. d. R. L., Santos, R. Y. C., and Nascimento, A. C. A. (2021). Anomaly Detection Technique for Intrusion Detection in SDN Environment using Continuous Data Stream Machine Learning Algorithms. In *2021 IEEE SysCon*, pages 1–7.

[Scalassara 2005] Scalassara, P. R. (2005). Análise de sinais de ultra-som usando decomposição autorregressiva e rastreamento de polos. Dissertação de mestrado, UEL, Londrina, PR, Brasil.

[Streit et al. 2020] Streit, A., Santos, G., Leão, R., de Souza e Silva, E., Menasché, D., and Towsley, D. (2020). Network Anomaly Detection based on Tensor Decomposition. In *2020 MedComNet*, pages 1–8.

[Uppuluri and Sekar 2001] Uppuluri, P. and Sekar, R. (2001). Experiences with specification-based intrusion detection. In *International Workshop on Recent Advances in Intrusion Detection*, pages 172–189. Springer.

[Wu and Shao 2005] Wu, Q. and Shao, Z. (2005). Network anomaly detection using time series analysis. In *ICAS-ISNS 2005*, pages 42–42. IEEE.

[Ye et al. 2000] Ye, N. et al. (2000). A markov chain model of temporal behavior for anomaly detection. In *Proceedings of the 2000 IEEE Systems, Man, and Cybernetics Information Assurance and Security Workshop*, volume 166, page 169. West Point, NY.

[Zhengbing et al. 2008] Zhengbing, H., Zhitang, L., and Junqi, W. (2008). A novel network intrusion detection system (nids) based on signatures search of data mining. In *WKDD 2008*, pages 10–16.