Um Sistema Inteligente para Detecção de DDoS em Ambientes Inteligentes baseado em Fog and Cloud Computing

Wanderson L Costa[1,2], Rafael L.Gomes[2] (Orientador)

[1]Instituto Federal do Piauí - (IFPI)

[2]Universidade Estadual do Piauí - (UECE)

wanderson.leonardo@ifpi.edu.br, rafa.lopes@uece.br

Abstract. *Urban spaces are engrafting Smart Environments (SE) to develop infrastructure, resources and services nowadays. SEs are composed of a huge number of heterogeneous devices (personal and IoT devices). One of the existing problems of SEs is the detection of distributed denial of service (DDoS) attacks, due to the vulnerabilities of IoT devices. It is necessary, therefore, to implement solutions that can detect DDoS in SEs, dealing with issues such as scalability, adaptability and heterogeneity. In this context, this paper presents an Intelligent System for detection of DDoS in SEs, applying machine learning approach (ML), fog computing and cloud computing. Besides that, the research presents a study on the most important traffic characteristics for the detection of DDoS in SEs, as well as a traffic segmentation approach to improve the accuracy of the system. The experiments performed, using real network traffic, suggest that the proposed system reaches 99% accuracy while reducing the volume of data exchanged and the detection time.*

Resumo. *Atualmente, os espaços urbanos estão implantando Ambientes Inteligentes (SE) para desenvolver infraestruturas, recursos e serviços. Os SEs são compostos por uma grande quantidade de dispositivos heterogêneos (dispositivos pessoais e IoT). Um dos problemas existentes dos SEs é a detecção de ataques de negação de serviço distribuída (DDoS), devido às vulnerabilidades dos dispositivos IoT. Dessa forma, é necessário implantar soluções que possam detectar DDoS em SEs, lidando com questões como escalabilidade, adaptabilidade e heterogeneidade. Nesse contexto, este artigo apresenta um Sistema Inteligente para detecção de DDoS em SEs, aplicando abordagem de aprendizado de máquina, computação em névoa e computação em nuvem. Além disso, a pesquisa apresenta um estudo sobre as características de tráfego mais importantes para a detecção de DDoS em SEs, bem como uma abordagem de segmentação de tráfego para melhorar a acurácia do sistema. Os experimentos realizados, usando tráfego de rede real, sugerem que o sistema proposto atinge* 99% *de acurácia, enquanto reduz o volume de dados trocados e o tempo de detecção.*

1. Motivação e Caracterização do problema

Em um futuro não muito distante, todos os nossos objetos do dia a dia estarão conectados à Internet e equipados com capacidades de sensoriamento, e poder de processamento suficientes para explorar todos os benefícios potenciais da chamada (*Internet of Things* - IoT). Novos ecossistemas surgem e tem sido denominado como ambiente inteligente

(*Smart Environments* - SEs), que pode ser implementado em diversos contextos: universidade inteligentes, casas inteligente, cidades inteligente, saúde inteligente e a Indústria 4.0. Esses contextos possuem serviços singulares para aprimorar a qualidade de vida dos usuários finais. Os SEs são compostos por dispositivos IoT (como sensores e atuadores) e dispositivos pessoais (como notebooks, smartphones, tablets, etc) [Li et al. 2018].

Ademais, ao observar esses equipamentos no contexto dos SEs ressaltam-se duas características indiscutíveis: grande quantidade de dispositivos e heterogeneidade. Como consequência, os SEs tendem a produzir maiores volumes de informações na rede do que redes tradicionais, devido à enorme escala de dispositivos na rede, bem como aos vários tipos de aplicativos executados nas camadas de redes mais superiores desses equipamentos. Todas essas questões trazem novos desafios relacionados à gestão e planejamento das SEs e os seus serviços neles executados [Ahmed et al. 2016].

Um desses desafios dos SEs é a detecção de ataques de negação de serviço distribuído (DDoS), que visam tornar o acesso a um ou mais alvo indisponíveis ao esgotar seus recursos por meio de múltiplas solicitações ilegítimas. Os ataques DDoS vêm de inúmeras vulnerabilidades de segurança nos dispositivos, especialmente dispositivos IoT [Diro and Chilamkurti 2018], que afetam diretamente a Qualidade do Serviço (QoS) e a Qualidade da Experiência (QoE). Como resultado, nos últimos anos, diversos ataques cibernéticos realizados na Internet ocorreram por meio da infecção de dispositivos IoT [Brun et al. 2018].

Ao levar em consideração a coexistência dos problemas: grande quantidade de dados que trafegam em ambientes inteligente e o aumento significativos de ataques DDoS utilizando dispositivos IoT, faz-se necessário o desenvolvimento de uma solução que seja capaz de segmentar os dispositivos conectados em redes, ou seja, separar o tráfego dos dispositivos IoT e dispositivos pessoais em uma rede com isso reduzir a dimensionalidade dos dados, bem como, detectar ataques DDoS, ocasionados por equipamentos IoT infectados.

A proposta de dissertação de mestrado foca em desenvolver um sistema inteligente que aplique um modelo baseado em técnicas de aprendizado de máquina com o objetivo de separar o tráfego originados de dispositivos IoT e dispositivos de usuários em segmentos de rede diferentes, com a finalidade de executar uma análise mais profunda das características de tráfego, identificação dos atributos mais relevantes, dessa forma, reduzir a dimensionalidade dos dados dos dispositivos IoT. Posteriormente, em virtude do aumento significativo do número de ataques de DDoS utilizando dispositivos IoT para atacar a infraestrutura da internet, desenvolver um mecanismo que possibilite barrar esse tipo de anomalia.

1.1. Objetivos

O objetivo geral da dissertação foi desenvolver um sistema inteligente baseado em Computação em Nuvem e em Névoa para a detecção de Ataques DDoS em Ambientes Inteligentes. O sistema funciona em duas fases, onde a primeira fase o sistema segmenta o tráfego de rede entre dispositivos IoT e dispositivos pessoais.Na segunda fase, o sistema irá utilizar algoritmos de aprendizado de máquina para detectar ataques DDoS originados do tráfego de rede de dispositivos no ambiente inteligente.

A partir deste contexto, definiu-se os seguintes objetivos específicos: (1) Identi-

ficar as características mais relevantes para a classificação e segmentação de tráfego de rede IoT e de dispositivo pessoal; (2) Desenvolver um mecanismo baseado em aprendizado de máquina capaz de segmentar o tráfego de redes de dispositivos IoT e dispositivos pessoais; (3) Projetar uma arquitetura baseada em *Fog* e *Cloud* para integrar atividades do sistema de detecção desenvolvido; (4) Implementar um sistema inteligente para a detecção de fluxos que fazem parte de um ataque DDoS; e, (5) Gerar um conjunto de dados para treinamento e avaliação composto de tráfego real oriundo de dispositivos pessoais e IoT, bem como tráfego benígno e ataques DDoS.

1.2. Contribuições e Produção Científica

Essa pesquisa desenvolve um sistema de detecção de ataque de negação de serviços distribuído (DDoS) utilizando uma arquitetura Fog and Cloud. Na Fog o sistema executa o monitoramento da rede com o objetivo de categorizar e classificar os dispositivos com modelos de ML, em seguida, é utilizada técnicas de seleção de características com a finalidade de selecionar os atributos mais relevantes para a identificação de ataques DDoS na rede, e por conseguinte reduzir a dimensionalidade e volume de dados a serem processados. Ao finalizar as etapas de pré-processamento, os dados são transmitidos para Cloud para a formação do conjunto de dados de conhecimento, e treinamento do modelo. Por fim, o sistema envia da Cloud o modelo já treinado para Fog onde é executada a etapa de detecção de DDoS.

A partir das contribuições feitas, a dissertação apresentada neste artigo apresenta as seguintes contribuições: (I) Uma análise sobre os trabalhos existentes relacionados a detecção de ataques DDoS em ambientes inteligentes; ; (II) Um mecanismo de segmentação de tráfego inteligente, capaz de diferencias fluxos oriundos de dispositivos de usuários e dispositivos IoT; (III) Um modelo de detecção de ataques DDoS para ambientes inteligentes; e, (IV) Um conjunto de dados real sobre tráfego de redes de ambientes inteligentes composto por dispositivos heterogêneos (usuários e IoT). As contribuições elencadas possibilitaram a publicação dos seguintes trabalhos:

1. International Journal of Communication Networks and Information Security (IJCNIS 2021 - Qualis: A4)[1]
2. IEEE Consumer Communications Networking Conference (CCNC 2021 - Qualis: A3)[2]
3. XII Computer on the Beach 2021 (COTB 2021 - Qualis: B3)[3]
4. V Workshop de Computação Urbana (COURB 2021 - Qualis: B1)[4]
5. IEEE Latin-American Conference on Communications (LATINCOM 2020 - Qualis: B2)[5]

2. Trabalhos relacionados e Inovação em relação ao estado da arte

A partir do levantamento bibliográfico realizado, nota-se que os trabalhos existentes na literatura focam na detecção de anomalias de rede (como por exemplo as referências [Hamamoto et al. 2018], [Yamauchi et al. 2019], [Diro and Chilamkurti 2018]

[1]http://www.ijcnis.org/index.php/ijcnis/article/view/5080
[2]https://ieeexplore.ieee.org/abstract/document/9369449
[3]https://sol.sbc.org.br/index.php/sbrc$_estendido/article/view/17175$
[4]https://sol.sbc.org.br/index.php/courb/article/view/17117
[5]https://ieeexplore.ieee.org/abstract/document/9282265

e [Brun et al. 2018]), de ataques DDoS (como no caso das referências [Vinayakumar et al. 2020], [Sharafaldin et al. 2019] e [Doshi et al. 2018]) e de Intrusos (tais como as referências [HaddadPajouh et al. 2018], [Zhou and Cheng 2019] e [Meidan et al. 2018]), mas nenhum desses trabalhos focou no desenvolvimento de soluções de detecção de DDoS em ambientes inteligentes, que foi o foco da dissertação.

Assim, embora vários estudos tenham sido realizados com a finalidade de identificar categorias de tráfego de redes e outras pesquisas foram desenvolvidas para identificar algum tipo de anomalia no tráfego de rede, nenhum desses trabalhos acima abordam um cenário realístico de uma infraestrutura heterogênea em um ambiente inteligente. A seguir estão relacionados alguns dos problemas encontrados nesses trabalhos: Dispositivos IoT simulados; Cenário de ataque indefinido; Não inclui o tráfego realístico com tráfego normal e ataques simultaneamente, bem como a não inclusão de dispositivos de usuários e dispositivos IoT em um mesmo ambiente; e, a seleção de características muitas vezes é ignorada ou utilizada de forma empírica ou sem justificativa adequada.

Desta forma, o sistema inteligente proposto inova em relação ao estado da arte tarefa de detecção de ataques DDoS em ambientes inteligentes. Este integra computação em Névoa e em Nuvem, dividindo as tarefas realizadas entre esses dois ambientes computacionais para reduzir o tempo de resposta e melhorar a acurácia.

3. Sistema Inteligente Desenvolvido

O sistema desenvolvido realiza as seguintes etapas: (I) Monitoramento da Rede, que gera um conjunto de dados em um formato PCAP; (II) Extração de características de fluxo, permite a extração de até 80 características; (III) Segmentação de tráfego, separa o tráfego dos dispositivos IoT e dispositivos pessoais, permitindo o treinamento da técnica de ML utilizando os dados de acordo com a categoria dos dispositivos (melhor compreensão do comportamento); (IV) Seleção de características, seleciona as características relevantes do conjunto de dados DDoS que melhora o desempenho dos modelos, onde as diferentes estratégias aplicadas por eles (métodos de filtro, métodos de empacotamento ou métodos embutidos) levam a diferentes características selecionadas; (V) Treinamento de Modelo, envolve a entrada dos dados no conjunto de dados de conhecimento na técnica de ML para a geração do modelo; e, (VI) Detecção de DDoS, execução da técnica de ML para detectar ataques.

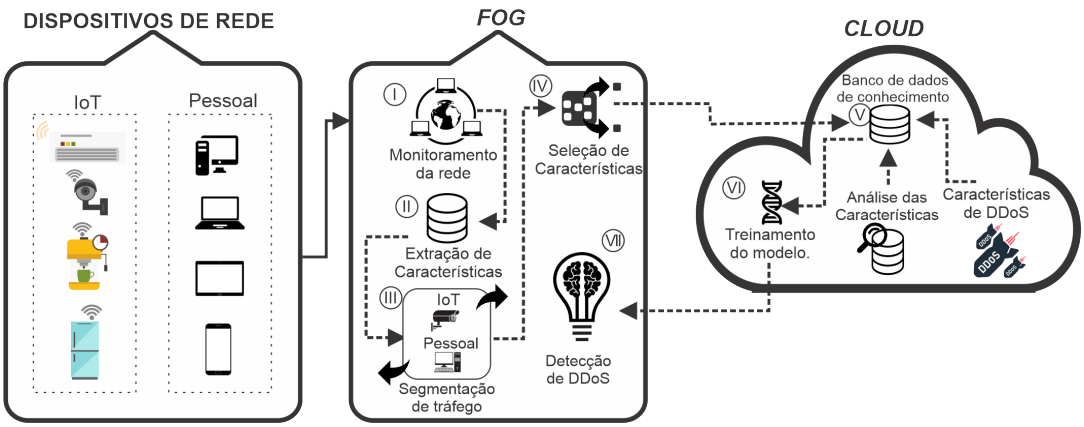


Figura 1. Visão geral do sistema desenvolvido.

Adicionalmente, o desenvolvimento do sistema proposto engloba duas etapas de pré-processamento para a construção do conhecimento básico: observação sobre DDoS, uma pesquisa para distinguir as características importantes da execução do ataque; e, análise de atributos de tráfego, um estudo para identificar as características mais importantes para melhorar a acurácia da detecção de DDoS. Essas etapas são executadas sequencialmente, trocando dados entre Névoa e Nuvem, conforme estrutura apresentada na Figura 1.

O fluxo de dados gerado pelo monitoramento da rede precisa ser processado antes da transmissão para a nuvem. Atualmente, a nuvem é o local usual para a execução de serviços. No entanto, com a escala cada vez maior dos fluxos de rede de ambientes inteligentes e, consequentemente, o volume dos fluxos de dados gerados pelos dispositivos, pode criar uma enorme sobrecarga de transmissão para a Internet.

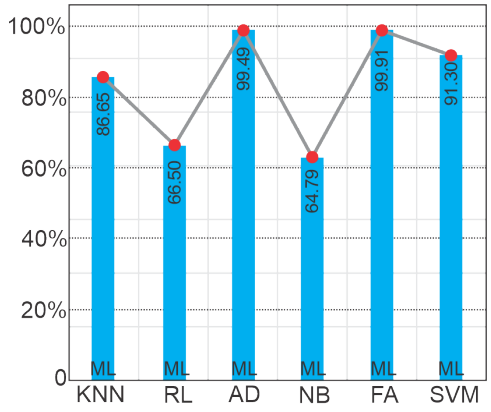
Além disso, as técnicas de ML aplicadas na detecção de DDoS demandam alto nível de recursos computacionais (processamento paralelo e capacidades de memória). Em geral, esses recursos computacionais não estão disponíveis na Névoa, exigindo vários serviços para suportar redes de acesso. Portanto, a execução de todas as funcionalidades do sistema inteligente proposto no ambiente de Névoa não é viável. Os dados brutos coletados são enviados ao ambiente de Névoa para serem processados. Após essas etapas, os dados brutos se transformam em dados processados, que serão enviados para a Nuvem. Esse processamento reduz o volume de dados, uma vez que apenas informações úteis são consideradas para serem transmitidas para a Nuvem. Após o processamento, os dados ficam disponíveis no ambiente de Nuvem e são usados como entrada das técnicas de ML para treinar o detector de DDoS. Como etapa final, o detector é implantado no ambiente de Névoa. Assim, os dados processados têm duas funções: (a) alimentar o treinamento de ML na Nuvem e (b) serem testados pelo detector de DDoS na Névoa.

A estrutura projetada do sistema inteligente proposto possibilita duas características importantes: (1) Pequeno *overhead* na infraestrutura de rede, devido ao baixo volume de dados transmitidos entre os ambientes de Névoa e de Nuvem; e, (2) Adequação de execução, uma vez que cada etapa dos módulos é executada no ambiente distinto, ou seja, as técnicas de ML são executadas na nuvem, enquanto o processamento de dados é executado na névoa. Esses dois recursos permitem que o sistema lide com os requisitos de escalabilidade, adaptabilidade e tempo de resposta dos ambientes inteligentes.

4. Resultados

Durante os experimentos, foram avaliadas as seguintes técnicas de seleção: Extra-Arvore, SVC, Lasso, Baixa Variância e Máxima Relevância e Mínima Redundância (mRMR) (casos de 5, 10, 20, 30 e 40 características). Essas técnicas selecionaram as características a serem utilizadas nas técnicas de ML treinadas: *K-Nearest Neighbors* (KNN), Regressão Logística (LR), *Naive Bayes* (NB), Floresta Aleatória (FA), Árvore de Decisão (AD) e *Support Vector Machines* (SVM). Portanto, avaliamos todas as combinações prováveis de técnicas de seleção e ML, permitindo uma análise completa sobre os desempenhos possíveis, como pode ser visto na Figura 2.

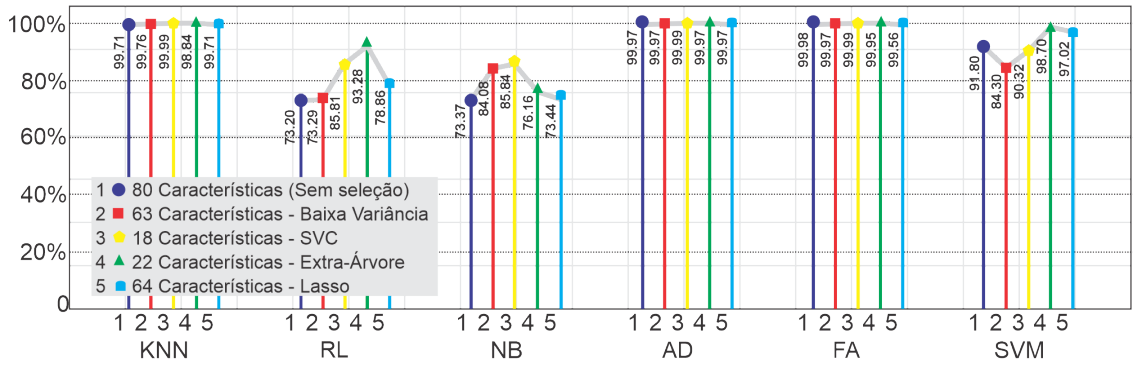
Os experimentos foram baseados em dois conjuntos de dados, que foram mesclados para representar um SE composto por dispositivos IoT e pessoais hete-



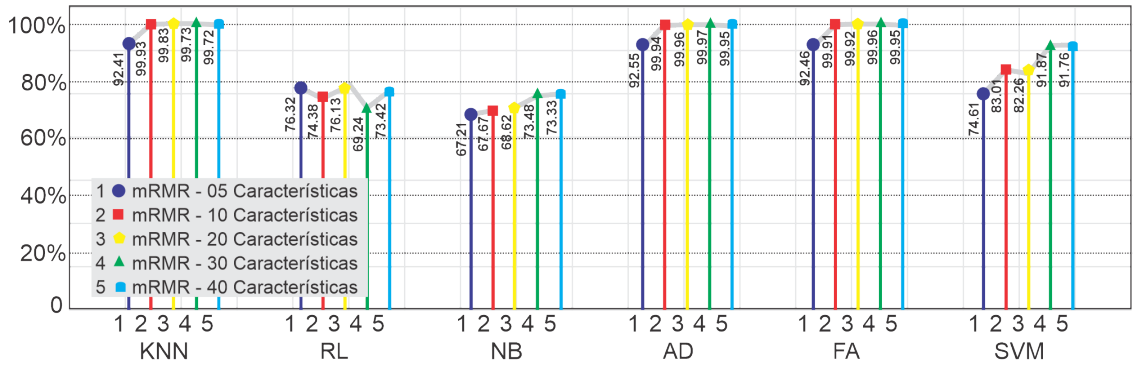
(a) Acurácia para Segmentação de Tráfego.

Técnica	Volume de dados
Dados bruto	15.16GB
Extração (80 características)	4.17GB
BV (63 características)	88MB
SVC (18 características)	21MB
EA (22 características)	22MB
Lasso (64 características)	84MB
mRMR (5 características)	5MB
mRMR (10 características)	9MB
mRMR (20 características)	22MB
mRMR (30 características)	37MB
mRMR (40 características)	50MB

(b) Volume de dados brutos (PCAP) e processados.



(c) Acurácia para detecção DDoS.



(d) Acurácia do mRMR para detecção DDoS.

Técnicas	KNN	RL	NB	AD	FA	SVM
80 Caract.	13.29	2.34	0.53	1.96	22.99	1625.04
BV	11.74	2.14	0.41	1.88	10.10	273.34
SVC	36.30	1.27	0.21	0.28	5.27	1226.96
EA	18.55	2.96	0.19	0.68	13.33	351.01
Lasso	11.91	1.20	0.14	0.74	6.73	304.38
mRMR 05	21.78	2.91	0.13	0.16	4.11	1055.70
mRMR 10	16.01	3.43	0.14	0.38	5.45	708.88
mRMR 20	10.67	3.42	0.30	0.81	10.12	1177.78
mRMR 30	7.91	0.93	0.29	0.49	10.68	385.93
mRMR 40	9.16	1.23	0.28	1.19	15.53	467.02

(e) Tempo de treinamento (em segundos).

KNN	RL	NB	AD	FA	SVM
9.02	0.02	0.53	0.02	0.53	162.57
7.98	0.03	0.03	0.01	0.34	15.24
15.62	0.02	0.02	0.01	0.35	144.13
12.95	0.01	0.06	0.02	0.44	21.15
8.15	0.01	0.14	0.01	0.45	142.08
16.29	0.02	0.01	0.01	0.48	146.50
12.43	0.02	0.01	0.01	0.53	142.84
7.72	0.01	0.02	0.01	0.58	198.56
5.46	0.02	0.03	0.01	0.49	86.29
6.26	0.05	0.04	0.01	0.52	102.73

(f) Tempo de detecção (em segundos).

Figura 2. Resultados.

rogêneos. O primeiro é o conjunto de dados "BoT-IoT"[6] desenvolvido por Meidan et al. [Meidan et al. 2018, Koroniotis et al. 2018], que contém o tráfego normal (benigno) e o tráfego relacionado aos últimos ataques DDoS. O último é o "UNSW-IoT"criado por Sivanathan et al. [Sivanathan et al. 2018], que tem tráfego normal (benigno) de IoT e dispositivos pessoais. Ambos os conjuntos de dados são formatados em dados de monitoramento do mundo real (PCAPs).

Em relação ao hardware utilizado nos experimentos, a Névoa executa em uma máquina local com Linux, CPU Intel i7-8700k 4.7GHz e 8GB de memória RAM DDR4, enquanto a Nuvem é uma máquina virtual Azure F48s-V2 com 48 vCPUs de 3.4GHz e 96GB de Memória RAM. Assim, realizamos os experimentos em ambientes de nevoa e nuvem adequados para cenários realistas. O desempenho do sistema inteligente proposto (incluindo a combinação das técnicas de seleção e ML) é apresentado na Figura 2.

Quando a seleção de características ocorre no *Fog*, a quantidade de dados chega a reduzir de 15,16 GB (brutos) para aproximadamente 25MB (processados) em média, representando menos de $0,2\%$ do volume de informações. Assim, a abordagem de integração *Fog and Cloud* aumenta a escalabilidade da rede, enquanto causa um impacto muito baixo na disponibilidade dos recursos da rede.

Os resultados dos experimentos destacam a importância da seleção de características para a acurácia, tempo de execução e volume de dados. Por exemplo, usando a técnica de seleção mais apropriada, o desempenho dos classificadores KNN e SVM aumenta em 8% e 7%, respectivamente. Além disso, a técnica RL usando os 80 atributos extraídos (sem seleção) tem uma acurácia inaceitável, enquanto usando a técnica Extra-Arvore, ela atinge mais de 93% de acurácia.

Considerando-se o tempo de treinamento, aumenta a sua importância em contextos em que é necessário um treino recorrente para atualizar o modelo de ML em virtude da alta dinâmica dos SEs, como cidades inteligentes. Assim, o modelo de ML será treinado em um período muito curto de tempo para manter a detecção de ataques DDoS de forma eficaz. O mesmo raciocínio pode ser aplicado ao tempo de detecção. Nesse contexto, o AD e FA com mRMR-10, mRMR-20, Lasso ou SVC são as combinações adequadas, pois são rápidos, têm alta acurácia e geram pequeno volume de dados. Por outro lado, se a periodicidade do treinamento for maior, devido ao comportamento estático do SE (como uma indústria inteligente), outras abordagens são viáveis.

Outro ponto importante a ser destacado é a redução do volume de dados a ser transferido da Névoa para a Nuvem. Em se tratando do volume total de dados gerado pela monitoramento da rede torna-se inviável para o envio de aproximadamente 15GB da fonte dos dados para a nuvem, todavia, ao utilizar a técnica de seleção de característica adequada, por exemplo, a SVC, é possível chegar a uma redução em até 99% do volume real a ser analisado na rede, apresentada a importância do pré-processamento (seleção de características) próximo a fonte de dados (Névoa). Logo a arquitetura baseada em Névoa e Nuvem torna-se indispensável para a implementação da proposta.

Referências

Ahmed, E., Yaqoob, I., Gani, A., Imran, M., and Guizani, M. (2016). Internet-of-things-

[6]https://www.unsw.adfa.edu.au/unsw-canberra-cyber/cybersecurity/ADFA-NB15-Datasets/bot_iot.php

based smart environments: state of the art, taxonomy, and open research challenges. *IEEE Wireless Communications*, 23(5):10–16.

Brun, O., Yin, Y., Augusto-Gonzalez, J., Ramos, M., and Gelenbe, E. (2018). Iot attack detection with deep learning. In *ISCIS Security Workshop*.

Diro, A. A. and Chilamkurti, N. (2018). Distributed attack detection scheme using deep learning approach for internet of things. *Future Generation Computer Systems*, 82:761–768.

Doshi, R., Apthorpe, N., and Feamster, N. (2018). Machine learning ddos detection for consumer internet of things devices. In *2018 IEEE Security and Privacy Workshops (SPW)*, pages 29–35. IEEE.

HaddadPajouh, H., Dehghantanha, A., Khayami, R., and Choo, K.-K. R. (2018). A deep recurrent neural network based approach for internet of things malware threat hunting. *Future Generation Computer Systems*, 85:88 – 96.

Hamamoto, A. H., Carvalho, L. F., Sampaio, L. D. H., Abrão, T., and Proença Jr, M. L. (2018). Network anomaly detection system using genetic algorithm and fuzzy logic. *Expert Systems with Applications*, 92:390–402.

Koroniotis, N., Moustafa, N., Sitnikova, E., and Turnbull, B. (2018). Towards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-iot dataset. *CoRR*, abs/1811.00701.

Li, H., Ota, K., and Dong, M. (2018). Learning iot in edge: deep learning for the internet of things with edge computing. *IEEE Network*, 32(1):96–101.

Meidan, Y., Bohadana, M., Mathov, Y., Mirsky, Y., Shabtai, A., Breitenbacher, D., and Elovici, Y. (2018). N-baiot—network-based detection of iot botnet attacks using deep autoencoders. *IEEE Pervasive Computing*, 17(3):12–22.

Meidan, Y., Bohadana, M., Mathov, Y., Mirsky, Y., Shabtai, A., Breitenbacher, D., and Elovici, Y. (2018). N-baiot—network-based detection of iot botnet attacks using deep autoencoders. *IEEE Pervasive Computing*, 17(3):12–22.

Sharafaldin, I., Lashkari, A. H., Hakak, S., and Ghorbani, A. A. (2019). Developing realistic distributed denial of service (ddos) attack dataset and taxonomy. In *2019 International Carnahan Conference on Security Technology (ICCST)*, pages 1–8. IEEE.

Sivanathan, A., Gharakheili, H. H., Loi, F., Radford, A., Wijenayake, C., Vishwanath, A., and Sivaraman, V. (2018). Classifying iot devices in smart environments using network traffic characteristics. *IEEE Transactions on Mobile Computing*, 18(8):1745–1759.

Vinayakumar, R., Alazab, M., Srinivasan, S., Pham, Q.-V., Padannayil, S. K., and Simran, K. (2020). A visualized botnet detection system based deep learning for the internet of things networks of smart cities. *IEEE Transactions on Industry Applications*.

Yamauchi, M., Ohsita, Y., Murata, M., Ueda, K., and Kato, Y. (2019). Anomaly detection for smart home based on user behavior. In *2019 IEEE International Conference on Consumer Electronics (ICCE)*, pages 1–6. IEEE.

Zhou, Y. and Cheng, G. (2019). An efficient network intrusion detection system based on feature selection and ensemble classifier. *CoRR*, abs/1904.01352.