

1 INTRODUÇÃO

A IoT conecta globalmente variados dispositivos, incluindo dispositivos móveis, *wearables*, eletrônicos domésticos, automóveis e sensores industriais. Estes dispositivos podem, através da Internet, ser acessados, conectar-se a outros dispositivos, servidores ou aplicações, tudo com mínima intervenção ou supervisão humana [?, ?, ?, ?]. Outra característica de dispositivos IoT são os recursos computacionais dimensionados, para propósitos específicos, que limitam a capacidade de computar outras funções muito além da função original do dispositivo.

Segurança e privacidade é uma grande preocupação em IoT, especialmente em relação aos dados pessoais como localização e saúde aos quais dispositivos tem acesso [?]. Contudo, além dos dados de sensores e atuadores, esses dispositivos se subvertidos, podem gerar tráfego maligno, como o gerado pela *mirai botnet* em 2016 [?, ?]. Nesse cenário, fatores que podem favorecer a subversão dos dispositivos incluem a falta de controle sobre a origem do hardware e software embarcado nos dispositivos bem como a menor frequência de atualizações deste software. Além disso, estes dispositivos tem longa vida e, após implantação, convivem com ampla diversidade de outros dispositivos, complicando a manutenção da rede que os hospeda, aumentando sua superfície de ataque.

No contexto de segurança de redes IoT, ferramentas que facilitem a detecção e resposta à ataques são necessárias. Como a maioria dos dispositivos IoT tem recursos limitados (como bateria, processamento, memória e comunicação), técnicas de segurança tradicionais baseadas em algoritmos configuráveis não são usuais, com isso restam as técnicas de observação de rede [?]. Ferramentas como NIDS observam o comportamento da rede e de seus dispositivos e detectam possíveis ataques.

Para implementação de NIDS, técnicas de tem sido empregadas na detecção de ataques a partir de ataques conhecidos ou descobrir novos ataques o mais cedo possível [?, ?]. Apesar do uso promissor de para segurança para sistemas IoT, muitos estudos na literatura [?, ?, ?] são limitados à métodos tradicionais de . Estes métodos tradicionais utilizam modelos estáticos para descrever e prever o comportamento da rede que, por serem estáticos, não mantêm confiabilidade quando confrontados com a evolução de ataques [?, ?].

Além das complicações de confiabilidade, a grande quantidade de dispositivos em redes distantes gerando dados em volumes e velocidades consideráveis, diretamente ligados às suas funções originais ou metadados produzidos como subproduto, técnicas tradicionais que tratam grandes lotes não são aplicáveis. Considerando fluxos contínuos de dados (*Data Stream*), técnicas de mineração de fluxos de dados (*Data Stream Mining*) são amplamente necessárias. Nesses cenários, essas técnicas são aplicadas, por exemplo, em problemas de monitoramento e classificação de valores originários de sensores para tomada de decisão tanto em nível micro, como na modificação de atuadores remotos, ou macro, na otimização de processos industriais.

Dentre as técnicas de mineração de fluxo de dados, classificadores podem ser utilizados para detectar padrões conhecidos e, em conjunto com algoritmos de DNFD, detectar novos padrões. Essa capacidade de detectar novos padrões é relevante para NIDS, onde novidades na rede podem representar novas funcionalidades ou ataques por agentes maliciosos, sem assinaturas existentes em bancos de dados de ataques conhecidos. Outras características que fazem DNFD atraente para NIDS são a produção de respostas imediatas e a detecção de novidades e mudança de conceitos já conhecidos.

Análises como *Data Stream Mining* e DNFD são geralmente implementadas sobre o paradigma de computação na nuvem (*Cloud Computing*) e, recentemente, sobre paradigmas como computação em névoa (*Fog Computing*). Para *fog*, além dos recursos em *cloud*, são explorados os recursos distribuídos pela rede desde o nó remoto até a *cloud*. Processos que dependem desses recursos são distribuídos de acordo com características como sensibilidade à latência, privacidade, consumo computacional ou energético.

Como elaborado, a aplicação de DNFD para detecção de ameaças em fluxos de dados originários de redes IoT dentro de NIDS tem sido um ponto de interesse [?, ?, ?]. Este trabalho foca explora as características de implementação destas técnicas em conjunto, focando serviços localizados na borda da rede, de maneira distribuída, para uso em ambientes IoT.

1.1 Motivação

Um problema recente que une, em um único contexto, os métodos de computação em névoa, processamento de fluxo de dados e detecção de novidades nesses fluxos é a detecção de intrusão em redes de dispositivos IoT. Para tratar esse problema, a arquitetura IDSA-IoT, recentemente proposta por [?], aplica ao problema algoritmos atuais de detecção de novidades em fluxos, executando esses algoritmos em ambiente próximo aos dispositivos e avaliando-os quanto à detecção de intrusão.

Na arquitetura proposta, [?] avaliou os algoritmos ECSMiner [?], AnyNovel [?] e MINAS [?], sendo que o último mostrou resultados promissores.

Contudo, o algoritmo MINAS ainda não foi implementado e avaliado com paralelismo, multi-processamento ou distribuição computacional, que são necessários para tratar fluxos de dados com grandes volumes e velocidades. O tratamento de distribuição em ambiente é essencial para aplicação deste algoritmo ao problema de detecção de intrusão em redes IoT, pois esta aplicação requer tempo de resposta mínimo e mínima comunicação entre nós distantes, como aqueles na borda e na nuvem. Ainda observando o algoritmo MINAS, destaca-se a possível divisão em três partes semi-independentes, sendo elas treinamento, classificação e detecção de novidade; a classificação é o elemento central cujos resultados são utilizados para a identificação de intrusões.

Ainda no contexto de DNFD como método de detecção de intrusão, outras propostas tratam do caso de fluxos com grandes volumes e velocidades, como é o caso de [?], que apresenta o *BigFlow* no intuito de detectar intrusão em redes do tipo *10 Gigabit Ethernet*, que podem produzir um volume considerável. Essa implementação foi feita sobre uma plataforma distribuída processadora de fluxos (*Apache Flink*) executada em um cluster com até 10 nós de trabalho, cada um com 4 núcleos de processamento, totalizando 40 núcleos, para atingir taxas de até 10,72 Gbps.

Os trabalhos de [?] e [?] abordam detecção de intrusão em redes utilizando algoritmos de ND em DS, porém com perspectivas diferentes. O primeiro investiga *IoT* e processamento em *fog* e baseia-se em um algoritmo genérico de detecção de novidade. O segundo trabalho trata de *backbones* e processamento em *cloud* e implementa o próprio algoritmo de detecção de novidade. Essas diferenças deixam uma lacuna onde, de um lado, tem-se uma arquitetura mais adequada para o ambiente *fog* com um algoritmo estado da arte de detecção de novidades, porém sem paralelismo e. Do outro lado da lacuna, tem-se um sistema escalável de alto desempenho porém almejando outro ambiente (*cloud*) e com um algoritmo menos preparado para os desafios de detecção de novidades.

A proposta, aqui chamada sistema M-FOG, adapta a arquitetura IDSA-IoT [?] empregando o algoritmo de DNFD MINAS [?], tornando-o capaz de ser executado em um sistema distribuído composto de pequenos computadores com recursos limitados, alocados na borda da rede próximos dos dispositivos IoT. Utilizando a nova implementação do algoritmo MINAS, avalia-se experimentalmente como a distribuição afeta a capacidade do sistema de detectar mudanças (novidades) nos padrões de tráfego e o impacto na eficiência computacional. Por fim, algumas estratégias e políticas para configuração do sistema de detecção de novidades em fluxo de dados são discutidas.

1.2 Objetivos

Como estabelecido na Seção ??, a lacuna no estado da arte observada é a ausência de uma implementação de algoritmo de detecção de novidades no contexto de computação distribuída na borda. Neste sentido, um algoritmo que trate adequadamente os desafios de fluxo de dados contínuos (como volume e velocidade do fluxo, evolução e mudança de conceito) e considere o ambiente de computação em névoa aplicada à detecção de intrusão.

Portanto, seguindo os trabalhos do Grupo de Sistemas Distribuídos e Redes (GSDR) da Universidade Federal de São Carlos (UFSCar), propõem-se a construção de uma aplicação que implemente o algoritmo MINAS [?] de maneira escalável e distribuível para ambientes de computação em névoa, seguindo a arquitetura IDSA-IoT [?].

Além disso, propõem-se também a avaliação dessa implementação com experimentos baseados na literatura usando conjunto de dados públicos relevantes. O resultado esperado é uma implementação compatível em qualidade de classificação ao algoritmo MINAS e passível de ser distribuída em um ambiente de computação em névoa aplicado à detecção de intrusão.

Com foco no objetivo geral, alguns objetivos específicos são propostos: Implementar o algoritmo MINAS de maneira distribuída sobre uma plataforma de processamento distribuída de fluxos de dados; Arquitetar e implementar um mecanismo de execução e avaliação de qualidade e