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 são uma grande preocupação em IoT, especialmente em relação aos dados pessoais como localização e saúde aos quais dispositivos podem ter acesso [?]. Além dos dados de sensores e atuadores que esses dispositivos gerenciam, se esses dispositivos forem subvertidos podem gerar tráfego maligno, como o produzido pela *mirai botnet* em 2016, onde em um dos ataques de DDoS 50 000 endereços únicos, de 164 países, formaram um pico de tráfego de 280 Gbps [?, ?]. 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 têm longa vida e, após implantação, convivem com ampla diversidade de outros dispositivos, o que torna complexa 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 a ataques são necessárias. Como a maioria dos dispositivos IoT tem recursos limitados (como energia, processamento, memória e comunicação), técnicas de segurança tradicionais baseadas em algoritmos configuráveis não são usuais, restando 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 têm sido empregadas na detecção de ataques a partir de características de ataques conhecidos ou na descoberta de novos ataques o mais cedo possível [?, ?]. Apesar do uso promissor de para segurança para sistemas IoT, muitos estudos na literatura [?, ?, ?] são limitados a métodos tradicionais de . Estes métodos comumente utilizam modelos estáticos, ou com atualização manual, para descrever e prever o comportamento da rede, que não mantêm a confiabilidade frente à evolução de ataques [?, ?].

Além das complicações de confiabilidade, grande quantidade de dispositivos, redes distantes, geração de dados em volumes e velocidades elevadas, as técnicas tradicionais, que tratam grandes lotes em *datacenters*, não são aplicáveis. Para esses fluxos contínuos de dados (*Data Stream*), técnicas de mineração de fluxos de dados (*Data Stream Mining*) entre outras que tratam são promissoras [?, ?, ?]. 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 (*Novelty Detection in Data Streams*), detectar novos padrões. Um destes algoritmos de DNFD é o MINAS [?]. 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. Neste sentido, uma avaliação do algoritmo MINAS como NIDS foi feita por [?] utilizando o conjunto de dados *Kyoto 2006+*, composto de dados coletados de 348 *Honeypots* pela Universidade de Kyoto [?].

Análises como mineração de fluxos de dados e DNFD têm sido implementadas sobre o paradigma de computação na nuvem (*Cloud Computing*) e, recentemente, também sobre paradigmas como computação em névoa (*fog computing*). Para névoa, além dos recursos em nuvem, são explorados os recursos espalhados de nós remotos até a nuvem. Processos que implementam este tipo de análise em névoa fazem uso desses recursos de acordo com características como sensibilidade à latência, privacidade, consumo de recursos computacionais ou consumo energético.

De maneira geral, 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 explora as características de implementação destas técnicas em conjunto, concentrando-se em serviços localizados na borda da rede, de maneira distribuída, para uso em ambientes IoT.

Este trabalho apresenta a construção e avaliação do sistema M-FOG[1], uma implementação paralela e distribuída em névoa de dispositivos IoT do algoritmo MINAS. Esta implementação foi construída com o padrão MPI buscando escalabilidade na tarefa de processamento de fluxo de dados e economia dos recursos limitados comumente encontrados em sistemas IoT, seguindo a arquitetura IDSA-IoT [**?**].

A avaliação do sistema M-FOG é constituída de métricas de qualidade de classificação e métricas de escalabilidade, ambas extraídas experimentalmente com o conjunto de dados *Kyoto dez 2015* relevante para NIDS. As métricas de qualidade de classificação obtidas dos resultados mostraram valores equivalentes à implementação de referência do algoritmo MINAS e as métricas de escalabilidade mostraram melhora em relação à implementação de referência porém com eficiência de paralelismo abaixo do esperado. Este trabalho contribui com uma análise do algoritmo MINAS com a ótica de distribuição em névoa, mostrando benefícios e desafios deste tipo de aplicação, iluminando os detalhes do problema abordado e apontando algumas soluções para trabalhos futuros.

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 relevantes do tema 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, multiprocessamento ou distribuição computacional, que são necessários para tratar fluxos de dados em ambientes distribuídos, como em cenários IoT e névoa.

O tratamento de distribuição em ambiente névoa é essencial para aplicação deste algoritmo ao problema de detecção de intrusão em redes IoT, pois esta aplicação requer tempo de resposta mínimo e pequena comunicação entre nós distantes, como aquelas comunicações entre borda e a nuvem. Ainda observando o algoritmo MINAS, destacam-se suas três partes: treinamento, classificação e detecção de novidades. A classificação é o elemento central cujos resultados são utilizados para a identificação de intrusões, enquanto a detecção de novidades fornece atualização automática do modelo de classificação.

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 DNFD, porém com perspectivas diferentes. O primeiro investiga IoT e processamento em névoa e baseia-se em um algoritmo genérico de detecção de novidade, sem modificações que o adaptem para o ambiente de névoa (recursos limitados, distribuídos e alta velocidade). O segundo trabalho trata de *backbones* e processamento em nuvem 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 de névoa com um algoritmo estado da arte de detecção de novidades, porém sem paralelismo. Do outro lado da lacuna, tem-se um sistema escalável de alto desempenho porém limitado ao ambiente nuvem e com um algoritmo que não foi projetado para os desafios de detecção de novidades.

A proposta deste trabalho, 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] Disponível em `https://github.com/luis-puhl/minas-flink`.