

Fusão Hierárquica de Dados para Sistemas de Computação em Névoa

Sandy F. da Costa Bezerra¹, Matheus Cordeiro², Atslands R. Rocha^{1 2}, Flavia C. Delicato³

¹Programa de Pós-Graduação em Ciência da Computação
Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE)

²Departamento de Engenharia de Teleinformática
Universidade Federal do Ceará (UFC)

³Departamento de Ciência da Computação - Instituto de Computação
Universidade Federal Fluminense

sandy.bezerra@ppgcc.ifce.edu.br, matheusgomescord@gmail.com

atslands@ufc.br, fdelicato@gmail.com

Abstract. *The Fog computing paradigm enables that computational resources are provided at the edge of the network so that the latency and bandwidth resulted from cloud computing can be reduced. The combination of this paradigm with data fusion and complex event processing (CEP) techniques, from raw data acquisition, has great potential for generating valuable knowledge and aiding decision making processes in IoT systems. In this context, we propose a three-level data fusion architecture (sensor node, fog, and cloud) that, combined with the use of CEP techniques, promotes timely decisions while reducing the cost of data transmission in the network. The experiments show a reduction of 77 % in the average time of sending messages in the network, and 82 % of improvement in the average flow rate when the proposed architecture is used.*

Resumo. *O paradigma da computação em névoa possibilita que os recursos computacionais sejam levados para a borda da rede, de forma que a latência e a largura de banda resultantes da computação em nuvem possam ser reduzidas. A combinação desse paradigma com técnicas de fusão de dados e processamento de eventos, a partir da aquisição de dados brutos, tem grande potencial de uso para gerar conhecimento de valor e ajudar a tomada de decisão em sistemas de IoT. Nesse contexto, foi proposta uma arquitetura de fusão de dados em três níveis (nó sensor, névoa e nuvem) que, aliada ao uso de técnicas de processamento de eventos complexos, promove a tomada de decisões com rapidez, enquanto reduz o custo de transmissão de dados na rede. Os experimentos mostram uma redução de 77% no tempo médio de envio de mensagens na rede e 82% de melhoria na vazão média com a utilização da arquitetura proposta.*

1. Introdução

A Internet das coisas (IoT, do inglês *Internet of Things*) propicia a instrumentação do mundo físico e sua conexão com o virtual a uma escala sem precedentes. O recente cres-

cimento de serviços e aplicações da Internet contribuiu para um aumento nos requisitos de armazenamento e processamento de dados. Eles são diversificados em termos de recursos exigidos por diferentes aplicações, portanto, muitas vezes requerem soluções personalizadas. A computação de borda (do inglês *Edge Computing*) veio para superar alguns das limitações do uso da computação em nuvem como plataforma de *backend* para a IoT. Tais limitações incluem a latência imprevisível e a falta de percepção quanto a localização e mobilidade do usuário e de suas aplicações [Bierzynski et al. 2017]. O conceito de computação de borda visa aproveitar as capacidades de armazenamento e processamento de um grande número de dispositivos IoT, *gateways*, estações radio-base, e de outros dispositivos de borda conectados à internet, fornecendo uma camada intermediária entre os dispositivos finais (produtores/consumidores de dados/informações) e a nuvem. Segundo [Caprolu et al. 2019], o paradigma da computação em borda inclui três tipos de dispositivos: os que geram os dados brutos (sensores); os dispositivos que oferecem a capacidade de processamento e armazenamento ou outros serviços (ex: *gateways*) e os que recebem os dados processados (ex: *smartphones*). Com a presença dos dispositivos de borda, o envio de dados e a carga de computação na nuvem são reduzidos, uma vez que algumas das solicitações são processadas localmente, não necessitando, assim, da intervenção da camada superior. Esse aspecto, por sua vez, reduz o tempo de resposta da rede e permite a manipulação de grande parte dos dados nas camadas mais próximas do local monitorado.

O paradigma de computação de borda foi aplicado a várias tecnologias emergentes e adotados por aplicações IoT que tiram proveito dos diferentes recursos oferecidos pela computação distribuída. Uma das concretizações desse paradigma é a computação em névoa (ou *Fog Computing*), que é uma infraestrutura de computação descentralizada que consiste em nós de processamento na névoa, ou seja colocados em qualquer ponto entre os dispositivos finais e a nuvem. De acordo com [Mouradian et al. 2018] a névoa é uma "nuvem mais próxima do solo". Ela estende a arquitetura tradicional de computação em nuvem até a borda da rede, permitindo o processamento e a análise de alguns dados nesse nível. Os nós na névoa são heterogêneos por natureza e, portanto, podem se basear em diferentes tipos de elementos, incluindo, mas não se limitando a, roteadores, pontos de acesso e *gateways* IoT. A heterogeneidade dos nós é oculta dos dispositivos finais, ao ser apresentada uma camada uniforme de abstração da névoa, que é composta por um conjunto de funções que fazem alocação e monitoramento de recursos, gerenciamento de dispositivos, juntamente com serviços de armazenamento e processamento. Essas funções são providas por uma camada de organização de serviço, que recebe solicitações dos usuários finais e aloca recursos de acordo com os requisitos das solicitações.

No ambiente da Internet das coisas, os dados brutos são frequentemente gerados por dispositivos com uma frequência muito alta. Para extrair informações significativas, precisa ser executado algum trabalho de pré-processamento nos dados brutos. Embora as técnicas de fusão de dados tradicionais também tenham capacidade de processar e extrair informações dos dados, a combinação com outras técnicas de processamento melhora ainda mais sua eficiência. A fusão de dados [Nakamura et al. 2007] pode ser entendida como um processo de agregação e combinação de dados provenientes de múltiplas fontes de modo a evitar redundância, reduzindo a quantidade de dados que circula pela rede. A fusão de dados também pode contribuir para a melhoria da acurácia dos dados, utilizando dados de várias fontes para compor uma informação mais completa. Além da fusão, técnicas de processamento de eventos complexos (CEP, do inglês *Complex Event*

Processing) também fornecem contribuições para a melhoria da informação gerada. O CEP vai além da simples consulta e transformação de dados e visa detectar padrões nos dados utilizando técnicas de filtragem, correlação e criação de novos eventos. Assim, os dados fusionados podem ser vistos como notificações de eventos que ocorrem no mundo externo utilizados para composição de eventos mais complexos.

Dentro desse contexto, neste trabalho é proposta uma arquitetura de fusão de dados aplicada em ambientes de computação em névoa, a qual atua em conjunto com o paradigma de CEP como uma solução para a alta latência e tempo de resposta em ambientes de computação em nuvem e seu uso em sistemas e aplicações de IoT. A fusão de dados é necessária para reduzir a quantidade de mensagens trocadas entre a área sensoreada e a nuvem. A ideia principal da proposta é identificar a ocorrência de um evento que desencadeia um processo de fusão de dados, visando identificar os nós que estão relacionados ao evento e extrair as características do local sensoreado. Dessa forma, informações são processadas em camadas diferentes, de acordo com a capacidade de cada dispositivo, otimizando o uso de recursos computacionais e diminuindo o tempo médio de envio de dados, tanto entre o sensor e a névoa quanto entre a névoa e a nuvem. Os dados são coletados, gerenciados e processados de acordo com os requisitos da aplicação e de forma descentralizada, onde diferentes níveis são responsáveis pela coleta, análise, tomada de decisão e suporte a sistemas IoT. Para dar suporte ao ambiente de computação de borda, neste trabalho é proposto também uma abordagem descentralizada de gerenciamento flexível para a autoconfiguração dos nós sensores em relação às névoas de acordo com parâmetros de distância e tipos de dados monitorados. Dessa forma, nós sensores podem ser adicionados e configurados rapidamente para se adaptar às alterações no ambiente.

As contribuições deste trabalho são (i) a proposta de uma abordagem hierárquica de fusão de dados visando tomadas de decisões oportunas; (ii) o uso de um mecanismo de processamento de eventos complexos (CEP), incluindo o processamento de vários eventos simples; (iii) a implementação de um mecanismo que permite a adaptação a mudanças físicas no ambiente monitorado, baseado no tipo de dado medido e na distância entre o sensor e a névoa.

Para avaliação da arquitetura proposta, foram simulados cenários na agricultura de precisão como estudo de caso. Em geral, os campos possuem sensores que monitoram as plantações e *gateways* IoT/servidores locais que se conectam a um servidor remoto ou à nuvem. Entretanto, a conectividade pode ser um fator limitante, portanto, o uso de técnicas que forneçam informações confiáveis e baixo custo de transmissão de dados é de suma importância. Os experimentos mostram uma redução de 77% no tempo médio de envio de mensagens na rede e 82% de melhoria na vazão média com a utilização da arquitetura proposta em comparação com sistemas que não usam essas técnicas na borda.

2. Trabalhos Relacionados

Em [Lan et al. 2019] é proposto um mecanismo universal CEP de monitoramento IoT utilizando computação de borda. Além disso é criado um modelo de evento complexo hierárquico formalizado, incluindo evento bruto, simples e complexo, para reduzir a complexidade da modelagem de eventos. O modelo utiliza semânticas complexas de tempo e espaço para definir eventos flexíveis por meio de algoritmos. Já os autores em [Alturki et al. 2017] utilizam uma abordagem híbrida, em que o processamento no nível

da borda e no nível da nuvem trabalham em conjunto para criar uma análise de dados IoT eficaz. Especificamente, coletam-se dados brutos localmente e os recursos são extraídos aplicando técnicas de fusão de dados nos dispositivos com maior poder de processamento e então são enviados para a nuvem.

O trabalho de [Oliveira and Kniess 2019] apresenta uma abordagem para redução de dados baseada na análise da dispersão dos dados de sensores, visando evitar o envio de detecções cujos valores sejam pouco dispersos. A abordagem proposta atingiu uma redução nas transmissões de dados, mantendo um baixo nível de erros e baixo consumo de energia. Em [Cremonezi et al. 2019] é proposto um sistema de três camadas hierárquicas (nuvem, névoa e dispositivos). O estudo foca na camada da névoa, composta por múltiplas redes capazes de estabelecer uma comunicação com a nuvem e armazenar identidades dos nós requisitadas pelos dispositivos. A camada da névoa pode possuir múltiplos níveis hierárquicos composta de diversos tipos de dispositivos com menor capacidade computacional se comparado com a nuvem (ex. pontos de acesso).

O trabalho de [Dautov et al. 2019] aborda a fusão de dados hierárquica em redes IoT que contém dispositivos de borda, unidades de rede e comunicação e plataformas em nuvem. Diferentes fontes de dados são combinadas em cada nível para produzir resultados oportunos e precisos. Em [Vora et al. 2017] é proposto um sistema de monitoramento de pacientes baseado em computação em nevoa. Os dados de movimento dos pacientes são coletados usando sensores vestíveis e são transmitidos usando os *gateways* da nevoa. Para reduzir a carga na infraestrutura de comunicação, também é apresentado um algoritmo eficiente de clusterização para transmissão de dados.

Em [R.Dautov and S.Distefano 2017] é proposto um mecanismo que utiliza CEP e arquitetura hierárquica como método para habilitar a fusão de uma grande quantidade de dados em vários níveis. Outro trabalho que trata de grande quantidade de dados é [Zhang et al. 2016], onde é proposto um novo paradigma de computação, projetado para processamento de *big data* em ambiente de borda colaborativa (CEE). O trabalho propõe a fusão de dados geograficamente distribuídos, criando visualizações virtuais compartilhadas dos dados que são expostos aos usuários finais por meio de interfaces predefinidas pelos proprietários dos dados.

De todos os trabalhos citados, o único que possui um modelo de evento com mecanismo de descrição hierárquico específico é o apresentado em [Lan et al. 2019]. Porém, as camadas consideradas em tal trabalho utilizam apenas CEP, sem fazer fusão de dados, o que torna sua utilização mais complexa e dessa forma, não é capaz de atender a aplicações voltadas a eventos mais abrangentes de monitoramento. Outro ponto importante que difere nosso trabalho de [Lan et al. 2019] é a possibilidade de uso de nós com baixo poder de processamento na camada mais baixa, além da capacidade de suportar uma lógica de gerenciamento flexível (autoconfiguração dos nós) que será descrita na Seção 4. As principais contribuições e diferenças entre os trabalhos estão listadas na Tabela 1.

3. Conceitos Básicos

De acordo com [Zhang et al. 2018], a fusão de dados é um pré-requisito para garantia de qualidade e mineração analítica de dados integrados. A fusão pode ser classificada levando em conta diversos aspectos. A classificação mais comum é a baseada no relacionamento entre as fontes dos dados, onde não existe dependência direta entre os sensores,

Tabela 1. Trabalhos Relacionados

Trabalho	Estrutura	CEP	Gerência dos Nós
<i>Esta proposta</i>	<i>Hierárquica Fusão + Eventos (3 camadas)</i>	<i>Sim</i>	<i>Sim</i>
[Lan et al. 2019]	Hierárquica de Eventos (3 camadas)	Sim	Não
[Alturki et al. 2017]	Fusão Híbrida (2 camadas)	Não	Não
[Oliveira and Kniess 2019]	Agregação de Dados	Não	Não
[Cremonezi et al. 2019]	Hierárquica (3 camadas)	Não	Não
[Dautov et al. 2019]	Hierárquica (3 camadas)	Não	Não
[Vora et al. 2017]	Fusão + Clusterização	Não	Não
[R.Dautov and S.Distefano 2017]	Hierárquica Multinível	Sim	Não
[Zhang et al. 2016]	Colaboração de Névoas	Não	Não

ou seja, eles funcionam de maneira descentralizada. Segundo esse critério, as técnicas de fusão podem ser classificadas em [Nakamura et al. 2007]: (i) complementar, onde os sensores fornecem informações de partes diferentes de um cenário; (ii) redundante, em que os sensores fornecem o mesmo tipo de informação de um mesmo cenário; e (iii) cooperativa, onde as informações de sensores diferentes podem ser unidas para gerar uma nova informação.

A fusão complementar busca a totalidade, compondo novas informações de diferentes partes. Um exemplo de fusão complementar consiste em fundir dados de nós sensores (por exemplo, uma amostra do sensor de umidade do solo no campo agrícola) em um mapa que descreve todo o comportamento do campo, portanto, resultando em informações mais amplas e completas. A fusão redundante pode ser usada para aumentar a confiabilidade e a precisão das informações como, por exemplo, sensores com medições de valores uniformes de umidades do solo fazendo o envio do dado apenas quando houver alteração no valor. Nas redes IoT, a fusão redundante pode fornecer informações de alta qualidade e impedir que os nós sensores transmitam informações redundantes o que resulta na economia de recursos. A fusão cooperativa também pode ser usada para reduzir o fluxo de dados em uma rede, visto que cada sensor é responsável apenas pelo envio de um tipo de informação. Um exemplo clássico de cooperação é o cálculo de um local de destino com base em informações de ângulo e distância.

Segundo [Nakamura et al. 2007], de acordo com o nível de abstração dos dados manipulados, a fusão de informações é realizada em três categorias:

- Fusão de baixo nível: os dados brutos são fornecidos como entradas e combinados em novos dados mais precisos do que as entradas individuais. Nesse nível, a fusão ocorre nos nós sensores (fontes de dados) onde os dados estão sendo coletados.
- Fusão de nível médio: é realizada a computação em um nível mais alto que o anterior, que pode ser executada com as informações pré-processadas, ou seja, os resultados do estágio de processamento anterior, no nó sensor, por exemplo, para obter estimativas de parâmetros de domínio espacial em uma determinada área ou implementar a extração de padrão ou ainda realizar cálculos redundantes.
- Fusão de alto nível: implementa o último estágio de processamento no nível mais alto de abstração em um domínio global, realizando inferência ou raciocínio com-

plexo e tomada de decisão baseada nos dados provenientes das camadas inferiores ou implementando fusão espaço-temporal.

Uma outra técnica de processamento de dados é a CEP que, ao contrário do processamento tradicional de fluxo, vai além da simples consulta e transformação de dados e visa detectar padrões de eventos complexos. Escalabilidade, eficiência, robustez e heterogeneidade são as principais características do CEP, além de detectar uma quantidade significativa de eventos com baixa latência [Akila et al. 2016a]. Além disso, os sistemas CEP são capazes de filtrar, correlacionar e agregar dados de diversas fontes para inferir eventos de alto nível e informações semântico entre os eventos.

Antes de entender o processamento de eventos complexos, é necessário entender e definir o que é um evento. Segundo [Cugola and Margara 2012], um evento pode ser entendido como a ocorrência de algo, uma "notificação" como, por exemplo, um aumento da temperatura ou uma queda brusca da umidade do solo. Além disso, os eventos podem se relacionar através do tempo, agregação e causalidade. Novos eventos são formados a partir da agregação de eventos e estes são chamados de eventos complexos. Assim, do ponto de vista do CEP, as informações que chegam constantemente podem ser vistas como notificações de eventos que acontecem no mundo externo. O foco dessa perspectiva está na detecção de ocorrências de padrões específicos de eventos em uma camada inferior que, por sua vez, desencadeiam eventos na camada superior. Quando as leituras correspondentes do sensor chegam ao sistema, elas são agrupadas e transmitidas para o mecanismo CEP implementado para processamento.

Neste trabalho, é utilizado a classificação da fusão baseada na fonte dos dados (redundante, cooperativo e complementar) em conjunto com a classificação baseada em níveis de abstração (baixo, médio e alto). Além disso, a proposta inclui a integração do CEP na hierarquia dos níveis. Podem ser citadas como vantagens da proposta de fusão de dados com o uso combinado do CEP em três níveis:

- Os mecanismos CEP distribuídos na rede são capazes de equilibrar a carga de trabalho do processamento dos dados observados.
- Um novo mecanismo CEP pode ser implantado em caso de aumento do volume de dados, alterando-se apenas o nível da hierarquia em questão.
- O tráfego de rede para transferir informações entre níveis pode ser reduzido significativamente.

4. Arquitetura Hierárquica de Fusão de Dados em Névoa

Os sistemas modernos de IoT tendem a ir além da noção de monitoramento "de baixo para cima", ou seja, não somente os dados brutos são coletados pelos dispositivos de sensoriamento e transferidos pela rede, mas também implementam a comunicação de *feedback* "descendente" entre os dispositivos gerenciados e gerenciador (por ex., para tomadas de decisão baseado nos dados). Como parte dessa comunicação bidirecional, a abordagem proposta utiliza a noção de coordenação, responsável pela comunicação entre os níveis descritos. Mais especificamente, a coordenação é uma funcionalidade em duas direções. Por um lado, o dispositivo da névoa recebe dados de dispositivos de nível inferior, coletando e despachando para processadores e motores de fusão de dados, bem como para o nível superior, se necessário. Por outro lado, esse mesmo dispositivo envia as novas regras de processamento para os níveis inferiores, de acordo com as mudanças observadas nos

dados. Dessa forma, a arquitetura conceitual de fusão de dados em três níveis aqui proposta é realizada por meio da implantação e execução de instâncias da lógica de fusão de dados em dispositivos, constituindo sistemas de IoT de vários níveis. A arquitetura (Fig. 1) descreve a separação **conceitual** de tarefas em todos os três níveis. Essas tarefas possuem instâncias de mecanismos de fusão de dados (DF) dedicadas, ou seja, fusão de dados de baixo nível (LDF), de nível médio (MDF) e de alto nível (HDF) [Dautov et al. 2019]. Os dois níveis superiores incluem componentes de coordenação (coordenadores LDF e MDF), responsáveis pela comunicação bidirecional entre os nós de nível inferior e superior e o gerenciamento de solicitações recebidas dos nós de nível inferior.

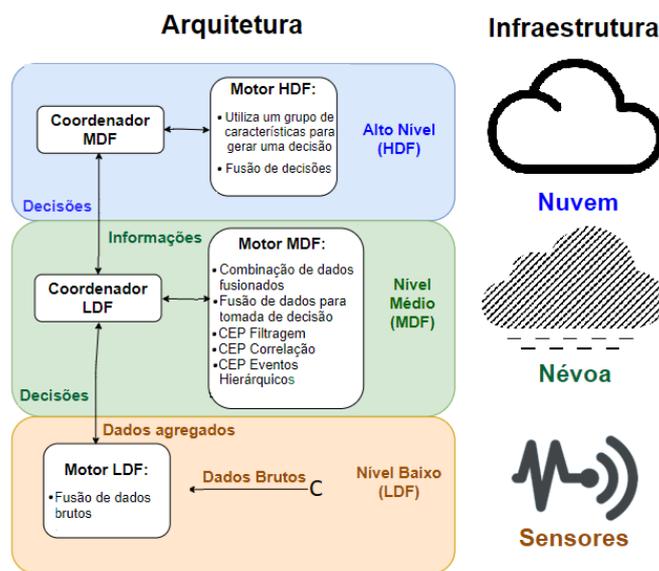


Figura 1. Arquitetura conceitual da fusão de dados hierárquicos em três níveis.

A arquitetura da solução proposta (Fig. 1) é dividida em três níveis principais, em que níveis distintos realizam fusão de dados provenientes de várias fontes diferentes. O dispositivo da camada baixa (sensor) é normalmente incorporado em objetos físicos e coleta dados do mundo real (por exemplo, umidade e temperatura). Uma fusão redundante relativamente limitada (uma agregação simples como, por exemplo, médias aritméticas simples ou ponderadas) é realizada nesse nível e seu resultado é enviado para o nível médio (nó da névoa). Esse aspecto permite reduzir o fluxo de dados na parte da rede com menor capacidade de transmissão. O nó da névoa está localizado próximo aos nós sensores ou ao longo do caminho de comunicação para a nuvem. Ele realiza uma fusão complementar de todos os dados recebidos pelos nós sensores (utilizando técnicas de processamento de eventos), permitindo assim, o suporte a tomada de decisões localmente a respeito da área sensoreada, baseando-se nos dados coletados e com baixa latência. É importante ressaltar que um nó de névoa tem menos recursos e menor conhecimento global que um nó de nuvem. Portanto, um nó de névoa pode aplicar algoritmos de fusão de dados menos sofisticados que os de um nó da nuvem, porém melhores do que os que podem ser aplicados em um nó sensor, pois possui capacidade mais ampla de processar e fusionar informação e com maior quantidade de dados de fontes diferentes. Se for necessário, o nó da névoa também pode realizar a atuação diretamente na área sensoreada visando também diminuir a latência. Finalmente, o nó da névoa envia os dados transformados pela fusão

para o nível alto (nó nuvem), que de acordo com a aplicação, pode ter a responsabilidade de armazenar os dados de forma persistente e criar inferências globais.

Para trabalhar em conjunto com a fusão de dados, nossa proposta utiliza o modelo CEP de regras apresentado em [Akila et al. 2016b], incluindo a camada da névoa, e respeita os seguintes princípios:

- Todo evento na névoa é gerado a partir do relacionamento entre as fusões redundantes de dados simples (médias aritméticas, por exemplo) provenientes de várias fontes de dados (sensores);
- Um evento complexo é gerado pelas interações entre eventos simples na névoa.

Além das regras citadas acima, o sistema proposto utiliza um mecanismo CEP que age como um motor de inferência processando os dados fusionados de acordo com o fluxo ilustrado na Figura 2. Brevemente, os eventos processados pelos nós nos três níveis são:

1. Um evento simples é extraído a partir do resultado da fusão simples dos dados sensoreados, a agregação é feita a partir da realização de um número N de leituras realizadas pelo nó sensor;
2. na névoa existe um mecanismo de regras que descreve os eventos gerados a partir do valor recebido, ou seja, o resultado da agregação realizada no nível anterior pode desencadear um outro evento diferente baseado no valor obtido. Por exemplo, uma filtragem de eventos irrelevantes que também gera um outro evento. No contexto de agricultura, um exemplo seria uma temperatura com valor de agregação inalterada desde os últimos dados enviados, logo esse valor seria ignorado.
3. O conjunto dos eventos simples se agregam gerando eventos mais complexos, e a névoa processa o evento de acordo com o que foi descrito na regra;
4. De acordo com a complexidade do evento, caso o evento necessite de uma visão global do sistema, um novo evento envia os dados para serem processado na nuvem.

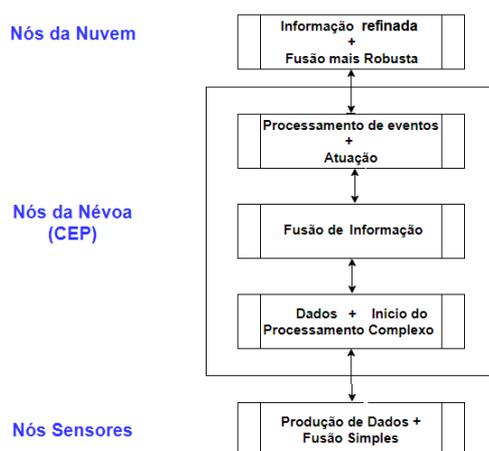


Figura 2. Fluxo do Processamento dos Eventos.

Segundo [Benarbia et al. 2018], esses 4 mecanismos de fluxo podem ser chamados, respectivamente, de receptor de dados, mecanismo de regra, executor de eventos e

encaminhador de evento. Vale ressaltar que o processamento de vários eventos simples nada mais é que o processamento complexo de eventos.

Uma das principais contribuições dessa proposta é que o modelo suporta uma lógica de gerenciamento flexível (autoconfiguração), realizada pelos nós da névoa, onde os nós podem ser adicionados e configurados rapidamente para se adaptar às alterações no ambiente monitorado. Desse modo, quando um novo nó sensor é adicionado à rede, ele consegue ser conectado a uma névoa de maneira simples. Para isso, o nó sensor envia via *broadcast* para os nós névoa uma mensagem incluindo suas coordenadas (que podem ser entendidas como latitude e longitude) e o tipo de dado coletado. Se o nó da névoa for do tipo compatível, ele calcula a distância euclidiana entre eles e envia esse dado para o nó da nuvem. Essa compatibilidade é relacionada ao tipo de dado que a névoa está apta a processar. Por exemplo, um sensor que envia dados de temperatura, por exemplo, só vai poder se conectar a um nó de névoa que processa dados de temperatura. O nó nuvem verifica qual névoa possui o menor valor de distância calculado e retorna essa informação para o nó da névoa correspondente que, por sua vez, passa a informação ao nó sensor.

5. Arquitetura da Simulação e seus Componentes

Neste artigo, foi utilizado um simulador de eventos discretos focado, mas não restrito, a ambientes de névoa chamado YAFS (*Yet Another Fog Simulator*). O YAFS foi projetado para analisar o *design* de aplicações e incorpora estratégias para posicionamento, programação e roteamento. O YAFS possui características similares as do simulador iFogSim e o utiliza como referência [Lera et al. 2019]. De acordo com [Gupta et al. 2017], o iFogSim é o simulador de nevoeiro mais utilizado para simulações de cenário utilizando computação em névoa e em nuvem, sendo ele uma extensão do conhecido simulador CloudSim [Calheiros et al. 2011]. O YAFS inclui mais funcionalidades do que os simuladores atuais para modelar cenários de IoT e apresenta uma maior facilidade de adicionar extensões.

Para a prova de conceito da proposta, o ambiente de agricultura de precisão foi utilizado como estudo de caso. No setor agrícola, alcançar o máximo rendimento da colheita a um custo mínimo é uma meta da produção. Assim, a tomada de decisão é complexa, pois há vários fatores que afetam todo o processo, principalmente visando a economia de recursos. Os campos agrícolas simulados incluem nós sensores que monitoram dados de temperatura e umidade do solo de uma plantação de coqueiros. Cada nó sensor está conectado e envia seus dados para um nó névoa localizado no campo, enquanto cada nó da névoa está conectado, por sua vez, a um nó da nuvem. Especificamente, foram executadas simulações em três cenários de agricultura de precisão considerando (i) uma rede com topologia fixa e de tamanho relativamente pequeno, (ii) uma rede com uma quantidade média de nós e (iii) uma rede em que a quantidade de nós vai aumentando no decorrer do tempo, ou seja, a topologia é dinâmica. Para cada cenário, a quantidade específica de cada tipo de nó é descrita a seguir:

1. O Cenário I (Fig. 3) é composto por 25 nós sensores, 5 nós névoa e um nó nuvem. No experimento utilizando este cenário foram executadas 30 rodadas de simulação, cada rodada com uma *seed* diferente para gerar dados de tipos diferentes e com valores em intervalos diferentes.
2. O Cenário II (Fig. 3) é formado por 100 nós sensores, 10 nós de névoa e uma nuvem. No experimento utilizando este cenário também foram executadas 30

rodadas de simulação, cada rodada com uma *seed* diferente para gerar dados de tipos diferentes e com valores em intervalos diferentes.

- No Cenário III (Fig. 4), a rede é composta inicialmente por 16 nós sensores, 4 nós de névoa e 1 nó nuvem. A cada nó névoa estão conectados 4 nós sensores, e os 4 nós névoa estão conectados ao nó nuvem. A cada rodada da simulação são adicionados nós sensores e nós névoa, de acordo com a Equação 1, sem alteração na quantidade de nó nuvem, que permanece 1 (um).

$$NS = NN^2 \quad (1)$$

onde:

$$3 < NN < 33$$

NS é o número de nós sensores;

NN é o número de nós névoa.

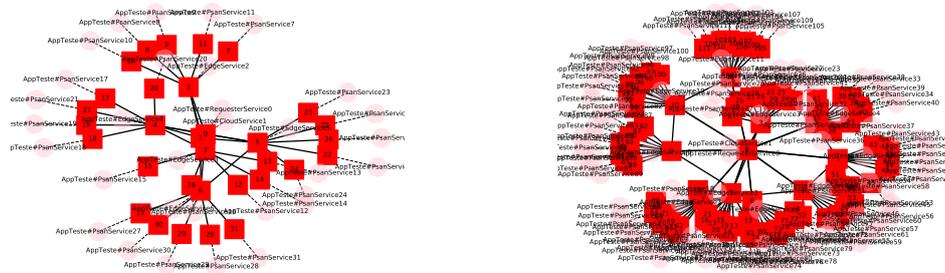


Figura 3. Topologia da Rede: Cenário 1 e 2

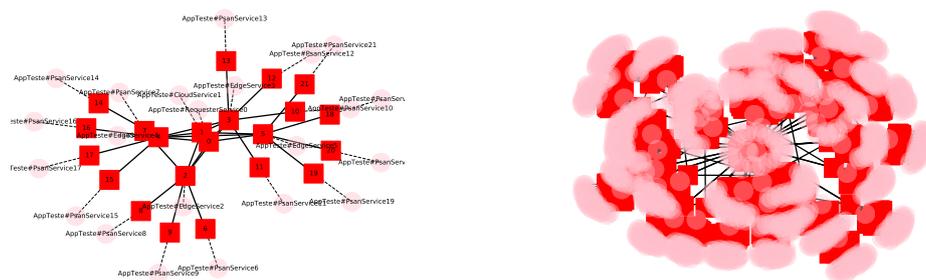


Figura 4. Topologia da Rede 3: Situação Inicial e Final da Simulação

A Figura 4 mostra a plotagem da topologia da rede no cenário 3 e suas conexões no início e fim da simulação respectivamente. A rede se autoconfigura e conecta cada nó sensor ao nó névoa baseado no dado coletado e na distancia euclidiana entre os nós (descrito na Seção 4).

Os nós sensores possuem 16 MB de RAM, enquanto os da névoa possuem 4 GB e os da nuvem possuem 8 GB. Esses valores foram escolhidos de forma a ficar mais

próximos dos valores dos dispositivos usados no mundo real. Os parâmetros configurados nos experimentos realizados são resumidos na Tabela 2.

Tabela 2. Experimentos

Experimento	Cenário	Avaliação
1	I: 25NS, 5NN	vazão e tempo de envio em densidade baixa de nós;
2	II: 100NS, 10NN	vazão e tempo de envio em densidade alta de nós;
3	III: NS = NN ²	vazão e tempo de envio em cenário adaptativo.

6. Avaliação Experimental: Resultados e Análise

Nesta seção, apresenta-se uma avaliação de desempenho dos experimentos realizados em relação as seguintes métricas: vazão e tempo de envio dos pacotes transmitidos entre os nós sensores, a névoa e a nuvem.

Nos cenários 1 e 2 (Fig. 3), a topologia da rede é fixa, pois estes cenários foram gerados com o intuito de ilustrar o funcionamento da proposta em diferentes densidades. O cálculo da vazão (Fig. 5) foi realizado por meio de uma média aritmética da quantidade de mensagens enviadas em cada rodada de simulação. Além disso, o gráfico apresenta o comportamento dos cenários 1 (C1) e 2 (C2) retirando-se a proposta da camada da névoa para fins de comparação. Os resultados apresentados na Figura 5 mostram a diferença da quantidade de mensagens enviadas com o processamento de dados. Devido a fusão de dados ser executada nos três níveis da rede, existe uma diminuição na quantidade de dados enviados entre os níveis, mesmo com esse aumento considerável na densidade.

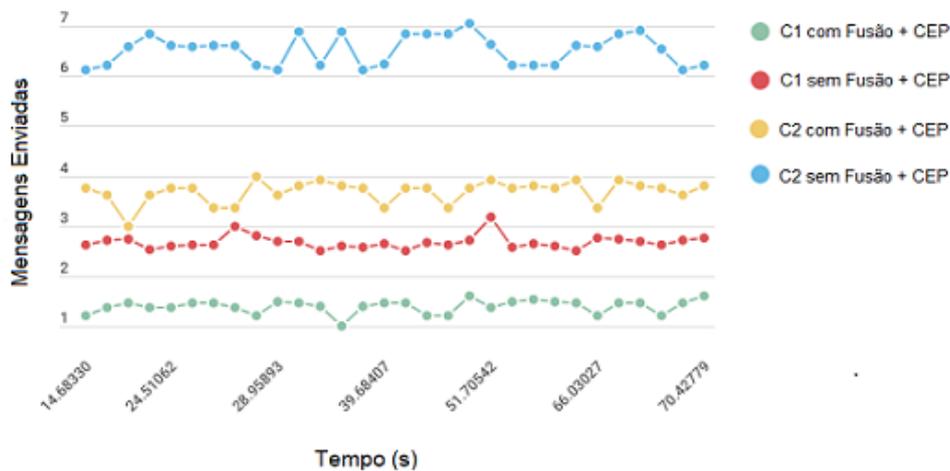


Figura 5. Vazão dos Cenários 1 e 2

No cenário 3 a topologia da rede foi sendo alterada com o passar do tempo. Logo, além do tráfego de dados que ocorreu nos outros cenários, a rede também precisou trocar informação para alocar os novos dispositivos que estavam sendo adicionados à topologia. Em suma, o aumento gradual da vazão de dados (Fig. 6) se dá tanto pelo aumento dos dados gerados na rede quanto pela troca de mensagens necessárias para sua autoconfiguração. Observadas as vazões nos cenários simulados, é possível verificar que

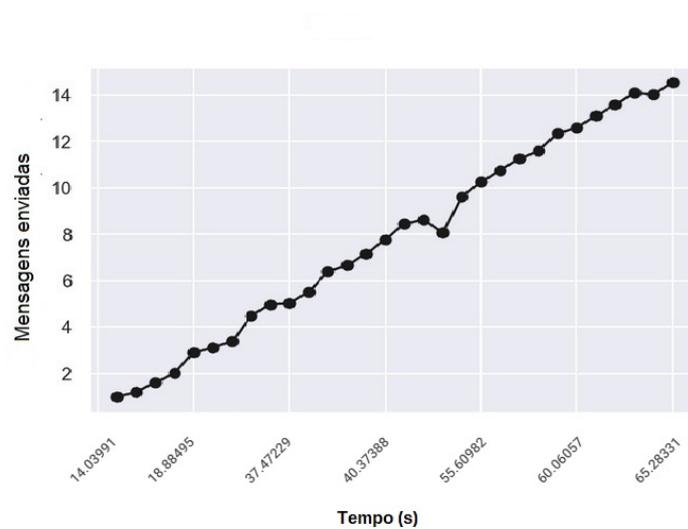


Figura 6. Vazão do Cenário 3

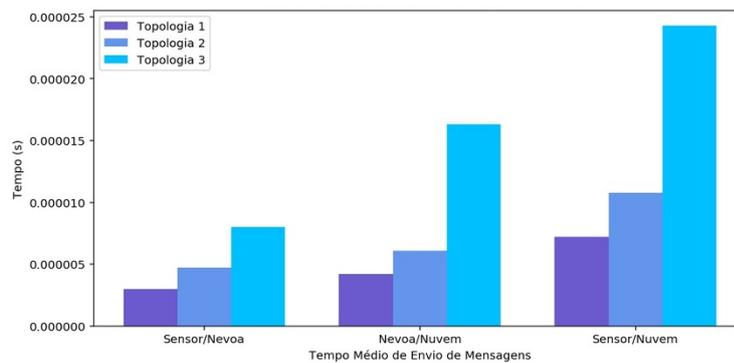


Figura 7. Tempo Médio de Envio de Mensagens nos 3 Cenários

houve redução média de 82% quando comparada a vazão da rede sem a utilização das técnicas.

Outra métrica a ser analisada é o tempo de envio médio de uma mensagem até o destino final. Através do tratamento de dados feitos pela fusão e o processamento de eventos, alguns dados que inicialmente seriam enviados à nuvem, são tratados e resolvidos nas camadas anteriores. Com base nos resultados da figura 7, pode-se verificar que o tempo médio de envio de mensagens é 77% menor quando o processamento é feito na camada da névoa, sem necessidade de envio de todos os dados para a nuvem.

7. Conclusão

Neste artigo foi apresentada uma metodologia de fusão hierárquica de três níveis em conjunto com técnicas de CEP em ambientes de computação em névoa. O sistema é capaz de se adaptar a mudanças físicas no ambiente monitorado, sendo possível os nós sensores adicionados se conectarem a um nó névoa baseando-se no tipo de dado e na distância euclidiana entre eles (autoconfiguração). Para avaliar a proposta foram executadas simulações em cenários distintos de agricultura de precisão, com diferentes densidades de rede e configurações. A avaliação teve como parâmetro de comparação uma

rede que usa a camada da névoa apenas como intermediária entre a área monitorada e a nuvem, sem aplicação das técnicas de fusão e CEP. Os experimentos mostraram que a rede obteve resultados bastante satisfatórios, como a redução de 77% no tempo médio de envio de mensagens e 82% de melhoria na vazão média com a aplicação da arquitetura proposta, que pode ser usada para tomada de decisões sobre a área monitorada ou para simples aquisição de conhecimento.

8. Agradecimentos

Esse trabalho foi parcialmente financiado pelo CNPq (projeto SmartGreen 432820/2016-7) e pela FAPESP (2015/24144-7).

Referências

- Akila, V., Govindasamy, V., and Sandosh, S. (2016a). Complex event processing over uncertain events: Techniques, challenges, and future directions. In *2016 International Conference on Computation of Power, Energy Information and Commuincation (ICCPEIC)*, pages 204–221.
- Akila, V., Govindasamy, V., and Sandosh, S. (2016b). Complex event processing over uncertain events: Techniques, challenges, and future directions. In *2016 International Conference on Computation of Power, Energy Information and Commuincation (ICCPEIC)*, pages 204–221.
- Alturki, B., Reiff-Marganiec, S., and Perera, C. (2017). A hybrid approach for data analytics for internet of things. In *Proceedings of the Seventh International Conference on the Internet of Things, IoT '17*, pages 7:1–7:8, New York, NY, USA. ACM.
- Benarbia, S., Alaoui, N., and Bennani, S. (2018). Complex event processing distributed architecture for massive open online courses. In *2018 International Conference on Smart Communications in Network Technologies (SaCoNeT)*, pages 273–276.
- Bierzynski, K., Escobar, A., and Eberl, M. (2017). Cloud, fog and edge: Cooperation for the future? In *2017 Second International Conference on Fog and Mobile Edge Computing (FMEC)*, pages 62–67.
- Calheiros, R. N., Ranjan, R., Beloglazov, A., De Rose, C. A. F., and Buyya, R. (2011). Cloudsim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. *Software: Practice and Experience*, 41(1):23–50.
- Caprolu, M., Di Pietro, R., Lombardi, F., and Raponi, S. (2019). Edge computing perspectives: Architectures, technologies, and open security issues. In *2019 IEEE International Conference on Edge Computing (EDGE)*, pages 116–123.
- Cremonesi, B., Nogueira, M., dos Santos, A., Vieira, A. B., and Nacif, J. A. (2019). Um sistema multinível de distribuição de identidades em névoas computacionais. In *Anais do XXXVII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 543–555, Porto Alegre, RS, Brasil. SBC.
- Cugola, G. and Margara, A. (2012). Processing flows of information: From data stream to complex event processing. *ACM Computing Surveys*, 44.

- Dautov, R., Distefano, S., and Buyya, R. (2019). Hierarchical data fusion for smart healthcare. *Journal of Big Data*, 6(1).
- Gupta, H., Vahid Dastjerdi, A., Ghosh, S. K., and Buyya, R. (2017). ifogsim: A toolkit for modeling and simulation of resource management techniques in the internet of things, edge and fog computing environments. *Software: Practice and Experience*, 47(9):1275–1296.
- Lan, L., Shi, R., Wang, B., Zhang, L., and Jiang, N. (2019). A universal complex event processing mechanism based on edge computing for internet of things real-time monitoring. *IEEE Access*, 7:101865–101878.
- Lera, I., Guerrero, C., and Juiz, C. (2019). Yafs: A simulator for iot scenarios in fog computing. *IEEE Access*, 7:91745–91758.
- Mouradian, C., Naboulsi, D., Yangui, S., Glitho, R. H., Morrow, M. J., and Polakos, P. A. (2018). A comprehensive survey on fog computing: State-of-the-art and research challenges. *IEEE Communications Surveys Tutorials*, 20(1):416–464.
- Nakamura, E., Loureiro, A., and Frery, A. (2007). Information fusion for wireless sensor networks: Methods, models, and classifications. *ACM Computing Surveys (CSUR)*, 39:9.
- Oliveira, S. and Kniess, J. (2019). Redução de dados em redes de sensores sem fio baseada em análise de dispersão. In *Anais do XXXVII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 1–14.
- R.Dautov and S.Distefano (2017). Distributed data fusion for the internet of things. In *PaCT*, pages 427–432.
- Vora, J., Tanwar, S., Tyagi, S., Kumar, N., and Rodrigues, J. J. P. C. (2017). Faal: Fog computing-based patient monitoring system for ambient assisted living. In *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*, pages 1–6.
- Zhang, L., Xie, Y., Xidao, L., and Zhang, X. (2018). Multi-source heterogeneous data fusion. In *2018 International Conference on Artificial Intelligence and Big Data (ICAIBD)*, pages 47–51.
- Zhang, Q., Zhang, X., Zhang, Q., Shi, W., and Zhong, H. (2016). Firework: Big data sharing and processing in collaborative edge environment. In *2016 Fourth IEEE Workshop on Hot Topics in Web Systems and Technologies (HotWeb)*, pages 20–25.