

Proposta e Avaliação de um Mecanismo de Agregação de Dados em Redes de Sensores

Jorgito Stochero , Antonio José Gonçalves Pinto , José Ferreira de Rezende *

¹Grupo de Teleinformática e Automação (GTA)
Universidade Federal do Rio de Janeiro (UFRJ)

stochero, antonio, rezende@gta.ufrj.br

Abstract. *In wireless sensor networks, data aggregation is critical to network lifetime. It implies that data will be processed while flowing from multiple sources to an specific node named sink. However, there are trade-offs for fusing multisensor data and creating a path between sources and sink that increases the likelihood of aggregation. In this work, we propose a decentralized aggregation mechanism using a parametric technique, such as Bayesian Inference and Dempster Shafer Method, and an extension to an existing data-centric routing protocol in order to favor aggregation without incurring in the overhead related to delay, scalability, robustness, and aggregation tree building and maintenance.*

Resumo. *Em redes de sensores sem fio (RSSF), a agregação de dados é crítica para o tempo de vida da rede. Isto implica em um processamento eficiente nos nós intermediários dos dados que fluem de múltiplas fontes para o sorvedouro. Entretanto, existem prós e contras para essa fusão distribuída de informações, e para otimizar o processo é preciso estabelecer caminhos que a favoreçam. Neste trabalho, propomos um mecanismo descentralizado que identifica nós agregadores e emprega neles técnicas como Inferência Bayesiana e de Dempster-Shafer para agregar a informação, sem incorrer em retardo ou perda de escalabilidade e robustez. A nossa proposta é aplicada sobre um protocolo de roteamento já existente e avaliada por meio de simulação.*

1. Introdução

A pesquisa em RSSF tem aumentado consideravelmente face ao seu amplo potencial de aplicações, como a coleta e tratamento de dados meteorológicos, o rastreamento de alvos e o monitoramento ambiental, dentre outras [1]. Uma rede de sensores é formada por um grande número de nós, empregados sob demanda em uma área de interesse. Cada nó possui um ou mais sensores, capacidade de processamento, armazenamento e de comunicações utilizando rádios de baixa potência. O principal requisito da rede, que afeta todo o seu ciclo de vida e independe da aplicação, é o judicioso consumo de energia.

A comunicação sem fio consome muito mais energia do que o processamento e o sensoriamento realizado pelos nós da rede [2]. Essa característica sugere que o processamento dos dados seja realizado de forma distribuída e não centralizada. Uma rede

*Este trabalho foi realizado com recursos do CNPq, CAPES, COFECUB, FAPERJ e SUFRAMA.

para sensoriar um fenômeno comum poderá gerar um elevado nível de redundância de informação, o que favorece a agregação de dados enquanto estes são roteados. As técnicas empregadas podem variar desde a supressão de mensagens duplicadas ou funções básicas (tais como, *máximo*, *mínimo* e *média*) até sofisticados métodos de fusão de dados [3].

Alguns pesquisadores têm proposto protocolos de roteamento [4] cujo objetivo é a construção de uma árvore reversa com várias fontes convergindo para o sorvedouro de forma a obter uma agregação ótima. Entretanto, estes protocolos consomem muita energia devido a necessidade do conhecimento global da topologia da rede. Outras alternativas, como o roteamento baseado em agrupamentos de nós [2], favorecem a agregação em um nó previamente selecionado dentro de cada grupo. No entanto, estes protocolos apresentam um elevado nível de complexidade na seleção do nó agregador, além da necessidade de balanceamento de energia entre os nós selecionados. Outra abordagem envolve protocolos de roteamento centrados em dados, que favorecem o processamento dentro da rede. A Difusão Direcionada (*Directed Diffusion*) [5] é um exemplo de protocolo que utiliza esse paradigma, não assume nenhum conhecimento da topologia da rede e é relativamente simples. Contudo, na construção da árvore N para 1, ele favorece rotas com baixa latência em detrimento dos caminhos que permitam a agregação.

O objetivo desse artigo é explorar dois aspectos complementares em redes de sensores. Inicialmente, a construção da árvore de agregação em protocolos de roteamento centrados em dados e, em seguida, o mecanismo de agregação destes utilizando técnicas paramétricas, de forma a proporcionar economia de energia com um baixo custo. A principal motivação deve-se ao fato de que os nós de uma RSSF trabalham de forma colaborativa para responder a uma tarefa de sensoriamento de um fenômeno comum.

O artigo está organizado da seguinte forma. Na seção 2 é apresentada uma revisão dos trabalhos relacionados com roteamento centrado em dados e os principais aspectos referentes à agregação em redes de sensores. A seção 3 descreve o protocolo de roteamento e o mecanismo de agregação incorporados à Difusão Direcionada. Na seção 4 são apresentados os resultados das simulações e avaliações do uso conjunto das propostas de roteamento e agregação. Por fim, a seção 5 traz as conclusões do trabalho.

2. Roteamento Centrado em Dados e Agregação

Energia é o fator mais importante em uma rede de sensores. O seu tempo de vida depende da energia economizada em cada solução de *hardware* e *software* que compõe a sua arquitetura. De acordo com o modelo de transmissão via rádio proposto em [2], a comunicação de dados é responsável pela maior parcela de consumo de energia quando comparada com o sensoriamento e processamento. Portanto, para economizar energia é melhor aumentar o processamento dos dados em cada nó de forma a minimizar suas transmissões. Adicionalmente, é desejável a comunicação em curtas distâncias, uma vez que longas exigem uma maior potência de transmissão. Considerando ainda que na maioria dos cenários as fontes estão próximas ao fenômeno de interesse e distantes do sorvedouro, a comunicação de curta distância obriga que os pacotes de dados sejam encaminhados por meio de nós intermediários em rotas de múltiplos saltos.

Para minimizar o consumo de energia, os protocolos de roteamento podem explorar duas particularidades das redes de sensores: redundância dos dados coletados e

a associação muitos-para-um entre fontes e sorvedouros. A redundância é fruto da alta densidade de nós na rede e das fontes normalmente estarem sensoriando um fenômeno comum. Como os pacotes de dados passam por múltiplos saltos até o sorvedouro, é possível proceder algum tipo de agregação nos nós intermediários, minimizando o número de transmissões e, portanto, economizando energia. Os mecanismos envolvidos no roteamento devem preocupar-se em identificar *onde* agregar os dados. Para que esta *agregação-na-rede* se torne efetiva, um esquema apropriado para o endereçamento e encaminhamento dos dados das fontes até o sorvedouro é essencial.

Nos protocolos centrados em dados [5, 6], os nós intermediários criam tabelas de encaminhamento baseadas em interesses enviados pelo sorvedouro e nos dados coletados pelas fontes. Estas últimas são reconhecidas pelo tipo de informação que podem gerar (temperatura, presença de fumaça, etc.) e pela sua posição geográfica atual (caso disponível). Assim, ao invés da designação de um identificador global único (como um endereço IP) para cada nó, é utilizado o *endereçamento baseado em atributos*. Este tipo de endereçamento elimina a necessidade de resolução de nomes e favorece a agregação. Um exemplo é o algoritmo chamado *Data funneling* [7], no qual por meio de agregação de pacotes apenas um fluxo de dados de um grupo de fontes é enviado até o destino. Nesse trabalho, o campo de sensores é considerado heterogêneo, e seus nós consistem de alguns poucos controladores (ou atuadores), com maior capacidade de processamento, e muitos sensores. O controlador divide o campo em regiões distintas e envia interesses para cada uma, usando inundação direcional. Ao atingir a *borda* da região alvo, o interesse é modificado para que os nós internos se organizem de forma sincronizada para reportar-se ao nó da borda, de forma a permitir nesta agregação e compressão de dados, antes do seu envio ao sorvedouro. Dessa forma, a agregação é realizada próxima do fenômeno de interesse, o que é interessante, mas existe um alto custo em estabelecer, sincronizar e manter a árvore de agregação dentro de cada região alvo.

Outros protocolos buscam estabelecer árvores de agregação. A árvore que fornece o caminho ótimo para agregação em termos de número de transmissões é, em geral, um problema NP-completo [8]. Os autores em [4] analisam três esquemas quase ótimos para obtenção dessa árvore em função de dois modelos de posicionamento e de um número crescente de fontes. O roteamento baseado em agrupamentos hierárquicos (*clusters*) favorece a agregação nos *cluster heads* e é analisado em [2]. O estabelecimento de uma árvore de agregação cria uma relação de perdas e ganhos entre o retardo e a economia de energia na rede. Quando as fontes estão próximas umas das outras e afastadas do sorvedouro, os menores retardos e maiores economias são atingidos [4]. Outra questão importante é o balanço de dispêndio de energia entre os nós intermediários. Quanto maior é o número de nós agregadores na rota até o sorvedouro, maior é o retardo, mas melhor é o balanço de energia. Todas estas propostas não levam em conta o gasto em energia para estabelecer e manter árvores de agregação em topologias dinâmicas devido a falhas de nós. Na abordagem hierárquica, há ainda eleição dos *cluster heads* e sua permuta, que também envolve troca de mensagens e consome uma quantidade não desprezível de energia.

Alguns pesquisadores têm explorado os impactos da agregação de dados em RSSF [4, 9, 10], examinando os aspectos relativos ao consumo de energia e retardo. Os resultados obtidos mostram significativa economia de energia, garantindo uma maior robustez

e prolongando o sensoriamento do fenômeno de interesse. Verifica-se, também, que o retardo introduzido pelas técnicas de agregação tradicionais de armazenar e encaminhar não deve ser negligenciado, devendo ser considerado durante o projeto da rede.

A grande variedade de aplicações e a diversidade de técnicas tornam o estudo da agregação de dados complexo. Algumas dessas técnicas são baseadas em métodos matemáticos e heurísticos envolvendo diversas disciplinas como a estatística, a inteligência artificial, o processamento de sinais e a teoria de decisão [3]. A função mais trivial de agregação de dados é a supressão de mensagens duplicadas. A compressão de dados também é sugerida em [7] como uma forma de agregação. Todas essas técnicas envolvem *como* agregar dados. O aspecto temporal consiste em *quando* a agregação deve ser realizada. Na abordagem tradicional, a agregação pode ocorrer somente uma vez (*instantânea*) ou em intervalos regulares (*periódica*) [11].

As funções básicas de agregação são adequadas para agregar múltiplas informações de um mesmo tipo de sensor, enquanto a fusão de dados de sensores de tipos diferentes exige métodos mais complexos. Alguns desses métodos têm sido explorados em diversos desenvolvimentos matemáticos, tais como as chamadas técnicas paramétricas, como a Inferência Bayesiana e o Método de Dempster-Shafer. Estas técnicas compõem o modelo de fusão de dados criado pelo *Joint Directors of Laboratories Data Fusion Subpanel* (JDL DFS) do Departamento de Defesa Americano [12]. O modelo JDL DFS, ilustrado na figura 1, tem sido alvo de diversas pesquisas e desenvolvimentos envolvendo sensores tradicionais, como radares e imageadores. No entanto, a literatura de RSSF tem se limitado apenas a citações do potencial destas técnicas. Neste trabalho, aplicamos as técnicas paramétricas sobre um protocolo de roteamento centrado em dados, estabelecendo um mecanismo de agregação que busca minimizar os aspectos negativos de abordagens tradicionais.

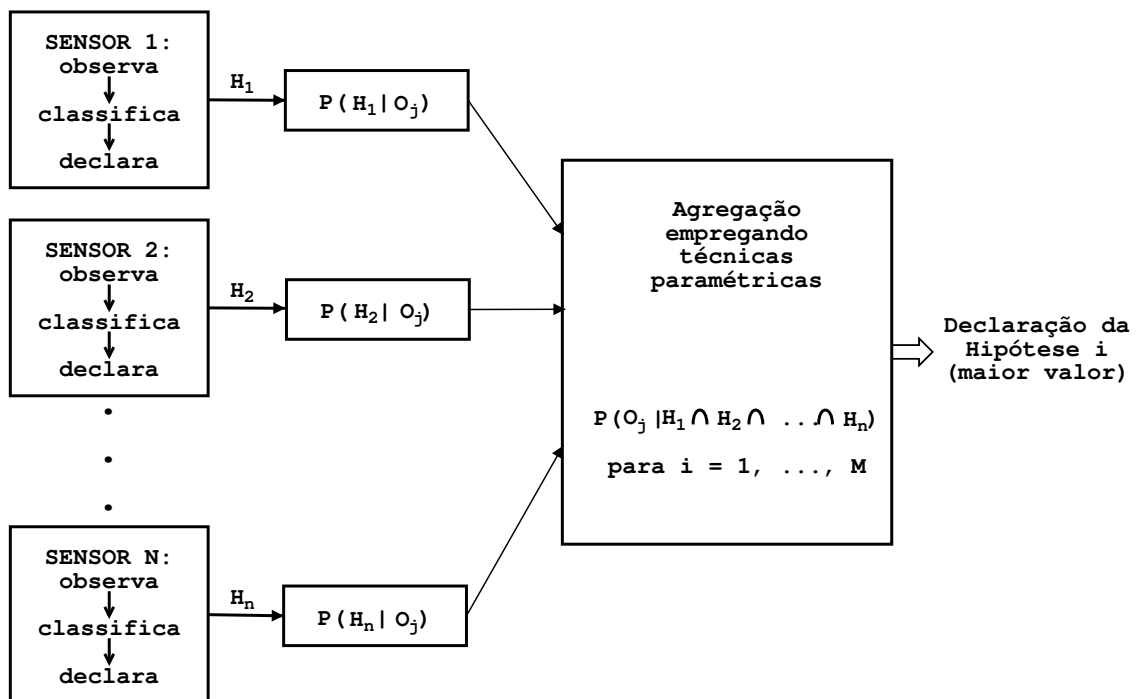


Figura 1: Modelo de fusão de dados (Adaptado de Waltz et al. [3]).

3. Roteamento e Mecanismo de Agregação

Este trabalho ocupa-se dos três diferentes aspectos que envolvem a agregação de dados em redes de sensores, que são *onde*, *como* e *quando* agregar, implicando o uso de protocolos de roteamento e mecanismos de filtragem nos nós sensores. Estes mecanismos são implementados de forma a explorar seus impactos na rede, em particular, a economia de energia. Neste artigo, propomos uma extensão a um protocolo de roteamento centrado em dados, de forma a aumentar a possibilidade de agregação na rede sem incorrer em altos custos relacionados com o estabelecimento e manutenção de uma árvore de agregação, e ainda, um mecanismo de agregação, utilizando técnicas paramétricas, com o mínimo de perdas em termos de retardo, escalabilidade e robustez.

O protocolo de roteamento centrado em dados utilizado é a Difusão Direcionada. Nesse protocolo, os dados são nomeados usando pares atributo-valor. O sorvedouro dissemina *interesses*, que constituem as tarefas de sensoriamento. Estes criam *gradientes* nos nós intermediários, provocando o estabelecimento de rotas para os dados que atendam ao interesse. Dessa forma, dados fluem das fontes até os originadores de interesse por múltiplos caminhos. O sorvedouro reforça um vizinho em particular, aquele do qual ele recebeu a primeira mensagem de dados. Os nós intermediários usam o mesmo critério para reforçar seus respectivos vizinhos, criando uma rota até a fonte. Esta rota privilegia caminhos de pequeno retardo, mas não necessariamente leva ao estabelecimento dos melhores caminhos para agregação.

Estes caminhos, descobertos por interações localizadas entre vizinhos, são independentes para fluxos de fontes diferentes. Propomos um novo esquema para descoberta de caminhos, reforçando nós intermediários com maior potencial para combinar dados de diversas fontes. Após a descoberta destes nós agregadores, empregamos algoritmos distribuídos de estimação para criar fluxos comuns, agregando de forma eficiente o tráfego da rede em um número reduzido de mensagens de dados. A idéia é usar as mesmas interações localizadas características da Difusão Direcionada para estabelecer uma árvore de agregação empírica na rede. Para tal, exploramos a sua arquitetura baseada em filtros, descrita em [13], que permite que um módulo de software externo, chamado *filtro*, interaja com o núcleo do protocolo e modifique suas capacidades de roteamento e agregação, influenciando, assim, como os dados são transferidos. Dois tipos de filtros são usados, um para descoberta de rotas com nós agregadores (Filtro PAR) e outro para agregação de dados (Filtro Bayes ou DS, conforme a técnica empregada).

3.1. Roteamento Privilegiando a Agregação (Filtro PAR)

Um nó tem potencial para agregação se recebe, dentro de um pequeno intervalo de tempo e de fontes diferentes, dados com atributos relacionados, mas não necessariamente os mesmos valores. A fusão dos dados vai depender da função de agregação utilizada e esta depende da aplicação da rede de sensores. Para privilegiar a agregação, buscamos escolher caminhos fontes/sorvedouro com um maior número de nós agregadores o mais próximo possível das fontes. Segundo o trabalho descrito em [4], tal abordagem proporciona os maiores ganhos em número de transmissões.

Seja o exemplo simplificado de um campo de sensores com 3 fontes, cada uma fornecendo mensagens do tipo **T**=Temperatura, **U**=Umidade, e **F**=Fumaça, contendo as evidências sobre o fenômeno a ser observado, que é a presença de fogo. A aplicação

da rede busca validar a hipótese se há ou não fogo e, para tal, aplica-se em cada nó uma técnica paramétrica de fusão de dados, discutida em 3.2, que corrobore ou refute a hipótese, com base nas evidências dos sensores. Neste caso, o número máximo de nós agregadores em determinada rota seria 3 (agregando-se em diferentes pontos as mensagens das fontes TU, TF, e UF). Ao passar por cada um destes nós, duas mensagens diferentes (T e U) seriam transformadas em uma única (TU), aumentando a confiabilidade para a decisão final sobre a hipótese fogo, e diminuindo o número de mensagens transmitidas na rede. Quanto mais próximo da fonte ocorrer esta agregação, mais mensagens serão suprimidas sem perda de informação.

Para identificar e selecionar estes caminhos, adicionamos um módulo à implementação da Difusão Direcionada disponível no simulador ns-2 [14]. Nomeamos o módulo de *Filtro PAR* (Roteamento com Privilégio para Agregação), como mostrado na Figura 2. Nosso filtro age nas mensagens de dados usadas para descoberta de rotas e configuração de caminhos, ou *exploratórias*, antes de serem processadas pelo filtro *Gradiente*, que é o módulo de roteamento do protocolo. Ao invés de imediatamente encaminhar a primeira mensagem exploratória recebida ao originador do interesse, o filtro PAR configura um temporizador associado a esta mensagem e a compara com outras mensagens até que o temporizador expire.

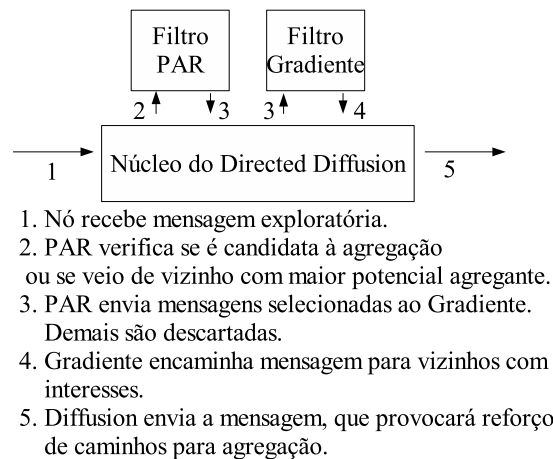


Figura 2: Filtro PAR.

Para a comparação, três atributos foram adicionados ao cabeçalho das mensagens, *distance_to_source*, *aggreg_nodes* e *aggreg_ID*. O primeiro contém o número de saltos entre a fonte e o primeiro nó capaz de agregação, indicando o quão distante estão. O segundo guarda o número de nós agregadores existentes no caminho e o *aggreg_ID* mantém uma indicação de quais fontes foram incluídas no processo de agregação. Mensagens recebidas de um vizinho em particular com os valores mais altos de *aggreg_nodes* e mais baixos de *distance_to_source* são escolhidas para encaminhamento, mesmo se recebidas depois de outras mensagens, que são descartadas. O atributo *aggreg_ID* é usado para evitar a computação duplicada da mesma fonte no caminho de agregação, o que provocaria números errados para *aggreg_nodes* e poderia prejudicar o mecanismo de seleção de rotas.

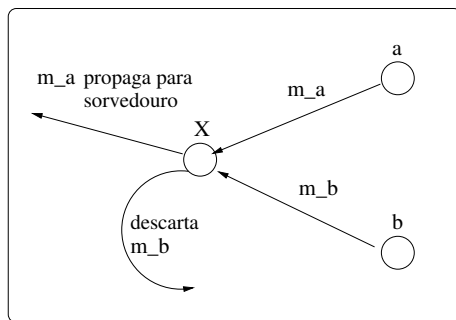
O último atributo é necessário porque um nó não tem conhecimento global da topologia. Ele transporta a informação de quais fontes foram incluídas no processo para

evitar que uma mensagem, agregada em outro ponto da rede, seja novamente marcada para agregação. Se o *aggreg_ID* da mensagem recebida casa com o das mensagens já no nó, então a mensagem não é agregada.

O mecanismo de seleção de rotas funciona em duas fases. A primeira ocorre quando a mensagem chega ao nó e é comparada com outras vindas de fontes diferentes, tendo por base *interesses* comuns e *aggreg_id* diferentes. Caso existam candidatas à agregação, os seus atributos *aggreg_nodes* são incrementados e *aggreg_ID* atualizados, incluindo as fontes da agregação em curso. O atributo *distance_to_source* passa a guardar a distância média de ambas as mensagens até as respectivas fontes. A seguir, ela é comparada com outras vindas da mesma fonte para fins de descarte. Se uma mensagem com maior *aggreg_nodes* ou menor *distance_to_source* (nesta ordem) é encontrada, a atual é descartada e nenhum temporizador é criado. Caso contrário, os temporizadores para as *piores* entradas existentes são removidos, as mensagens são apagadas e um temporizador é criado para a nova mensagem.

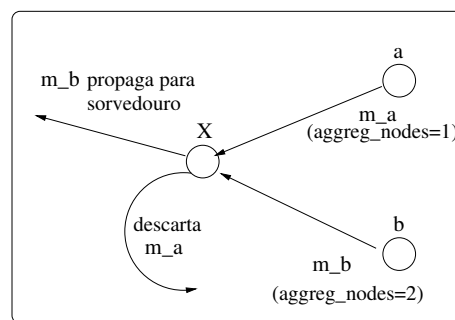
A segunda fase começa quando este temporizador expira e uma nova verificação por melhores mensagens é efetuada. Caso exista melhor mensagem, esta substitui aquela cujo temporizador expirou e é enviada em seu lugar ao filtro Gradiente, que se encarrega de encaminhá-la aos vizinhos que possuam interesses. O filtro Gradiente seguirá agindo como normalmente faz, exceto que agora ele recebe mensagens exploratórias com melhores parâmetros de agregação, e então diferentes vizinhos são reforçados.

Por exemplo, supondo que um nó tenha dois vizinhos, *a* e *b*, dos quais recebe mensagens similares, e que a mensagem de *a* chega antes, conforme ilustrado na figura 3.



1. m_a e m_b mensagens da mesma fonte que chegam em X por caminhos diferentes (por a ou b);
2. m_a chega em X antes de m_b;
3. filtro Gradient considera m_b uma duplicata de m_a e a descarta;
4. m_a encaminhada na direção do servidor.

(a) Propagação pelo Gradiente



1. m_a e m_b mensagens da mesma fonte que chegam em X por caminhos diferentes (por a ou b);
2. m_a chega em X antes de m_b;
3. filtro PAR armazena ambas e decide que m_b passou por melhor caminho (maior *aggreg_nodes*);
4. m_b encaminhada na direção do servidor.

(b) Propagação pelo PAR

Figura 3: Propagação das Mensagens Exploratórias.

Normalmente, o filtro Gradiente a encaminharia (menor latência) e descartaria a mensagem de *b*, sem maiores considerações, conforme pode ser visto em 3(a). Com o PAR, poderia ser identificado que o valor de *aggreg_nodes* de *a* é 1, significando uma possível agregação de duas fontes distintas, enquanto o valor de *aggreg_nodes* de *b* é 2, indicando agregação de três fontes. Conforme pode ser visualizado em 3(b), o filtro

PAR descartaria a mensagem vinda de a e encaminharia a mensagem de b . Para o filtro gradiente, a mensagem de a nunca existiu, então apenas a de b é encaminhada a frente. Quando chegar o momento de reforçar um caminho, o nó b será reforçado. O processo continua em cada nó intermediário até o sorvedouro ser alcançado e as rotas desejadas (agregadoras) criadas.

Desta forma, sem alterações no filtro Gradiente e tampouco no núcleo do protocolo original, rotas contendo nós agregadores são identificadas e reforçadas. Empiricamente, tais rotas possuem mais nós agregadores e estes estão mais próximos das fontes. É importante salientar que o filtro PAR não realiza nenhuma técnica de agregação de dados, apenas preparando o caminho para que esta seja acentuada. Ao agir apenas nas mensagens exploratórias, não provoca influência significativa no tráfego da rede nem no retardo, visto que estas correspondem a uma pequena fração das mensagens geradas e somente são transmitidas durante a auto-configuração da rede. A opção de não modificar o filtro Gradiente foi tomada para manter as características de robustez e escalabilidade da Difusão Direcionada, sem alterações estruturais que provocassem efeitos colaterais imprevistos. O filtro PAR deve ser usado em conjunto com outro módulo que atue nas mensagens de dados empregando uma técnica de agregação adequada. No restante deste trabalho, apresentamos dois filtros para agregação, um utilizando Inferência Bayesiana e outro o Método de Dempster-Shafer.

3.2. Mecanismo de Agregação Empregando Técnicas Paramétricas

O mecanismo proposto tem por objetivo reduzir os custos da agregação de dados em RSSF[10], e ainda, tirar proveito da formação das rotas compartilhadas. Para atingir tal objetivo, foi criado um mecanismo distribuído, no qual a agregação ocorre localmente em cada nó e sem tempo de espera. O nó agregador combina a informação disponível (isto é, as observações locais e as informações obtidas dos nós vizinhos) no momento em que recebe uma mensagem. Os dados que chegam não ficam aguardando para serem agregados com outros dados oriundos de fontes que estão mais afastadas. Nesse sentido, este mecanismo difere do procedimento tradicional de armazenar e encaminhar mensagens. O resultado é um processamento distribuído na rede, o que favorece a escalabilidade, e uma redução de transmissões sem ocorrer em aumento da latência da rede.

O mecanismo de agregação tem início após a definição da rota entre os nós fonte e o sorvedouro, daí a importância da formação de caminhos com intersecções (nós de agregação) como ilustrado na figura 4. O procedimento de agregação consiste em manter uma cópia da última evidência recebida em cada nó (do caminho fonte-sorvedouro), para ser utilizada pela função agregadora. Essa evidência pode ser usada para agregação por período T_v , como ilustrado na figura 4(a), que fixa o tempo de validade da evidência. Para evitar a agregação de dados oriundos de um mesmo sensor, o parâmetro T_v deverá ser inferior ao menor intervalo de transmissão (I_t) de mensagens dos nós fontes. Outro parâmetro que precisa ser ajustado é o tempo de descarte das mensagens T_d , ilustrado na figura 4(b). Esse parâmetro define o período em que mensagens agregadas podem ser consideradas redundantes, e portanto, devem ser descartadas. No caso do Filtro DS, existe ainda um parâmetro (C) que define a variação máxima admitida para o grau de certeza das mensagens, em um intervalo T_d .

Quando uma mensagem do tipo dado chega no nó ela é encaminhada para o filtro

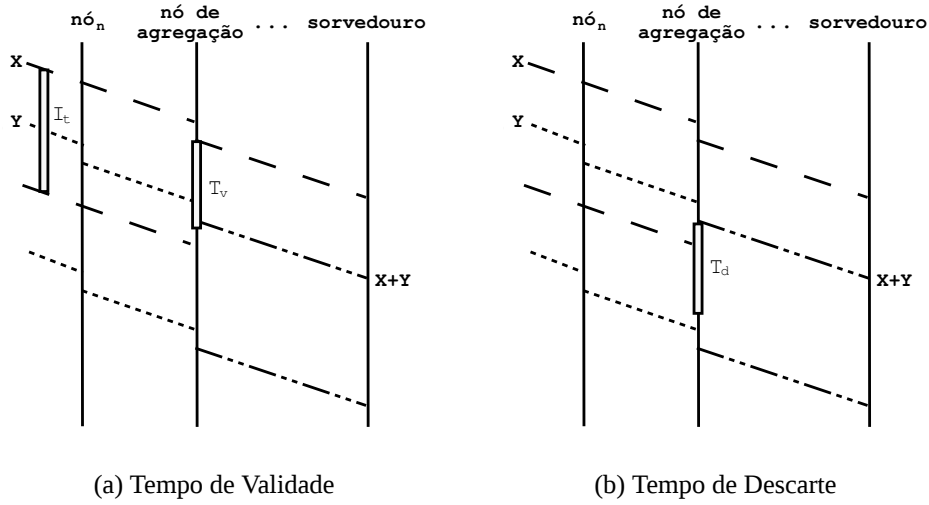


Figura 4: Parâmetros do Mecanismo de Agregação.

de agregação. A primeira etapa consiste em verificar as evidências válidas do cache. Isso é feito comparando a diferença entre a hora atual e o horário de chegada das evidências com T_v . Em seguida, é feita uma leitura do conteúdo da mensagem para conhecer as evidências que estão sendo informadas. Após esses procedimentos, as evidências (locais e recebidas) são combinadas, gerando uma nova mensagem. Então, o cache é atualizado com as evidências mais recentes. A nova mensagem poderá ser encaminhada para o filtro Gradiente ou simplesmente ser descartada, caso seja idêntica (ou sem modificação significativa) quando comparada à mensagem anterior, dentro do intervalo T_d .

Os nós realizam a agregação pela combinação de técnicas paramétricas de fusão de dados com o descarte de mensagens redundantes. Tais técnicas permitem um mapeamento direto entre as evidências detectadas pelos sensores e os eventos de interesse, o que é facilitado pelo esquema de endereçamento centrado em dados da Difusão Direcionada. Um exemplo desse mapeamento é ilustrado na figura 5, que modela um cenário simplificado no qual se deseja monitorar a qualidade de ar de uma região. Para esse cenário foram formuladas as hipóteses **R**=qualidade do ar regular, **I**=qualidade do ar inadequada e **C**=qualidade do ar crítica, com base nas evidências **O**=ozônio, **S**=dióxido de enxofre, **F**=fumaça e **N**=dióxido de nitrogênio. As evidências de interesse constituem as tarefas de sensoriamento, que são propagadas na rede, enquanto as hipóteses formuladas representam o resultado da agregação obtido em cada nó.

Os métodos de agregação foram implementados nos Filtros Bayes (Inferência Bayesiana) e DS (Inferência de Dempster-Shafer), que são os outros módulos de software adicionados ao protocolo e serão vistos a seguir.

3.2.1. Filtro Bayes

A Inferência Bayesiana foi implementada no Filtro Bayes. Trata-se de um método estatístico de agregação de dados baseado na Regra de Bayes [12]. Essa técnica pertence à classe de algoritmos que utiliza o conhecimento *a priori* sobre o espaço de observações

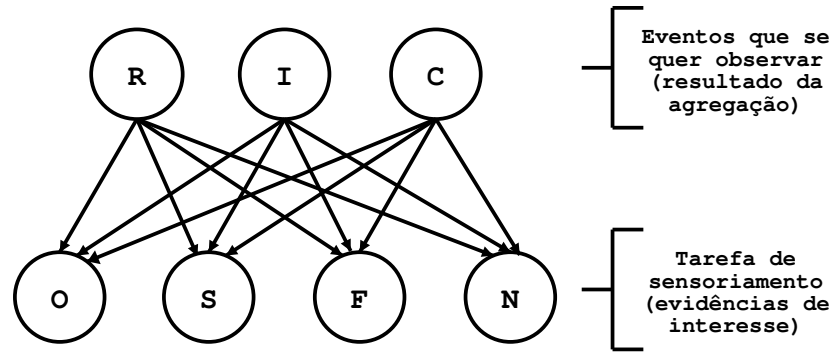


Figura 5: Mapeamento entre as evidências e os eventos de interesse.

para realizar inferências. O método consiste em atualizar a crença de uma hipótese baseada na crença estimada anteriormente e em evidências adicionais detectadas pelos múltiplos sensores. As crenças podem ser derivadas da probabilidade clássica ou, mais frequentemente, da subjetiva.

O Filtro Bayes foi implementado em um cenário genérico, de forma a atender a inúmeros problemas de agregação. Nosso modelo consiste de um conjunto de quatro hipóteses mutuamente exclusivas e exaustivas que explicam a ocorrência de um evento E (uma observação) de interesse, dentro de dois possíveis de serem identificados pelos sensores. Foram definidas as probabilidades *a priori* de cada hipótese e as probabilidades de cada hipótese ser verdadeira dado que uma evidência E ocorreu. Dessa forma, cada nó sensor pode usar a regra de Bayes, expressa em função das probabilidades condicionais (equação 1), para atualizar as crenças *a posteriori* de cada hipótese com base nos eventos detectados pelos nós fontes e propagados na rede.

$$p(H_i/E) = \frac{p(E|H_i)p(H_i)}{\sum_i [p(E|H_i)p(H_i)]} \quad (1)$$

onde

$p(H_i|E)$ = probabilidade a posteriori de que H_i é verdadeira dado E ,

$p(E|H_i)$ = probabilidade de observar E dado que H_i é verdadeira,

$p(H_i)$ = probabilidade a priori de que H_i é verdadeira,

$\sum p(H_i) = 1$, e

$\sum_i [p(E|H_i)p(H_i)]$ = probabilidade de observar E dado que H_i é verdadeira para todo i .

3.2.2. Filtro DS

O Filtro DS implementa a agregação dos dados baseado na Inferência de Dempster-Shafer [12]. Esta técnica pode ser considerada como uma generalização da Inferência Bayesiana, sendo utilizada quando os sensores contribuem com informação que não pode ser associada a cem por cento de certeza. O algoritmo permite a agregação dos dados (proposições) com qualquer que seja a crença (probabilidade de massa) dos sensores. Um nó sensor que possui múltiplas proposições, aplica a regra de Dempster (equação 2) para encontrar sua intersecção ou a conjunção com a respectiva probabilidade de massa associada.

A aplicação avaliada na seção 4 considerou a existência de dois tipos diferentes de sensores e quatro possíveis eventos. Na terminologia de Dempster-Shafer, o quadro de discernimento Θ seria constituído por $\{e_1, e_2, e_3, e_4\}$, o que dá origem a 2^4 proposições. Cada nó fonte S_i transmite uma proposição com sua probabilidade de massa m_i associada. Dessa forma, de acordo com as observações dos sensores, a probabilidade de detectar o evento e_1 , por exemplo, é indicada pelo intervalo de confiança $[Bel_i(e_1), Pl_i(e_1)]$. O limite inferior do intervalo é dado pela função de crença (Bel), que considera todas as evidências E_k que confirmam a proposição e_1 , enquanto o limite superior do intervalo é dado pela função de plausibilidade (Pl), que fornece a quantidade máxima de crença que pode ser atribuída a proposição e_1 . Assim temos $Bel_i(e_1)$ e $Pl_i(e_1)$ dado por:

$$Bel_i(e_1) = \sum_{E_k \subseteq e_1} m_i(E_k) \quad e \quad Pl_i(e_1) = 1 - \sum_{E_k \cap e_1 = \emptyset} m_i(E_k)$$

Para cada possível proposição (por exemplo e_1), a Inferência de Dempster-Shafer permite combinar as massas de probabilidades m_i e m_j das observações do sensor S_i e S_j aplicando a seguinte regra:

$$(m_i \oplus m_j)(e_1) = \frac{\sum_{E_k \cap E_l = e_1} m_i(E_k)m_j(E_l)}{1 - \sum_{E_k \cap E_l = \emptyset} m_i(E_k)m_j(E_l)} \quad (2)$$

4. Simulação e Resultados

Nesta seção, descrevemos os resultados obtidos com o nosso modelo, implementado no simulador ns-2 [14]. Escolhemos as seguintes métricas para avaliar seu desempenho: para a árvore de agregação, o número de nós agregadores nos fluxos, a distância média destes às fontes e o tempo de descoberta de rotas; para o mecanismo de agregação, o tempo de validade e de descarte das evidências, e a variação da confiança dos dados; e finalmente para o consumo de energia, o número de mensagens recebidas no sorvedouro e o número total de transmissões (saltos) na rede. A importância das métricas selecionadas para a árvore e o mecanismo de agregação é apresentar a influência desses elementos no comportamento da rede, enquanto que a combinação do número de mensagens entregues no sorvedouro e quantidade de saltos balizam o ganho do esquema para prolongar o tempo de vida da rede.

Para avaliar estas métricas, simulamos um campo de sensores com uma topologia aleatória em uma área quadrada de 1000 metros de lado e outra em forma de grade com cada nó distante 200 metros dos vizinhos. Em ambos os casos o alcance de transmissão foi de 250 metros e as fontes enviam mensagens de dados a cada segundo. No modelo proposto, um sorvedouro dissemina interesses e fontes afastadas deste respondem com dados. Para avaliar o impacto na rede, variamos o número de nós no campo de sensoria-mento em ambas as topologias.

Árvore de agregação

O primeiro passo na simulação envolve a descoberta de rotas. Testamos a grade de 100 nós e três fontes com e sem o filtro PAR. As rotas estabelecidas pela Difusão

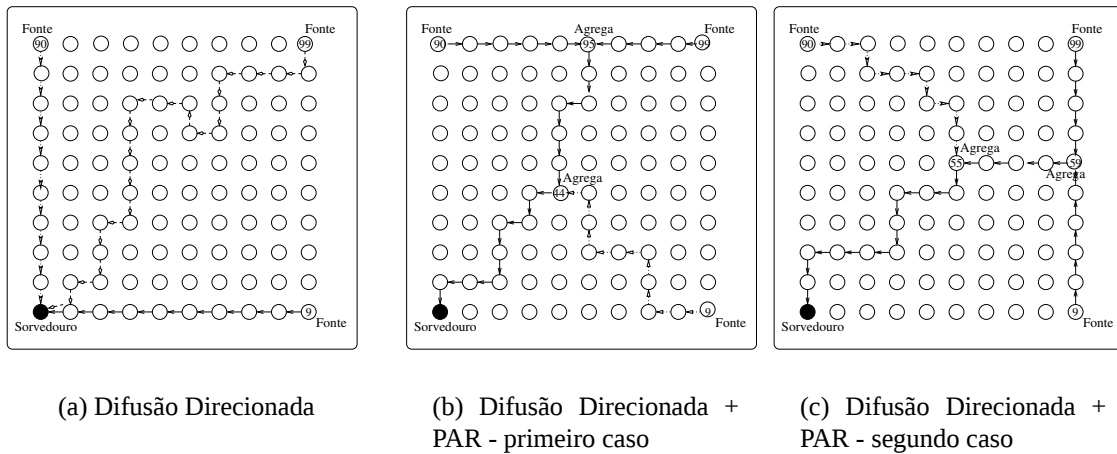


Figura 6: Rotas mais comuns.

Direcionada, sem o PAR, seguiram caminhos independentes, que variavam ao longo das simulações, mas em geral seguiram um padrão de menor retardo, como observado na figura 6(a). Esta variação freqüente das rotas evita a sobrecarga em nós com uso muito freqüente, mantendo o balanço de energia na rede. Ao adicionar o filtro PAR, rotas não mais independentes foram estabelecidas. Assim, por existirem três fontes, as rotas preferidas alternaram-se buscando nós agregadores, como pode ser observado em 6(b) (nó 95 aparece como agregador) e em 6(c) (nó 55).

A figura 7 mostra os impactos dessas novas rotas sobre os parâmetros de agregação. Em 7(a) pode ser vislumbrado o número de nós agregadores para diversos tamanhos de grade, e em 7(b) a sua distância média às fontes. Pode ser observado que o filtro PAR contribuiu para elevar o número de nós agregadores nos fluxos de dados para todas as grades ao mesmo tempo em que reduziu a distância dos mesmos às fontes.

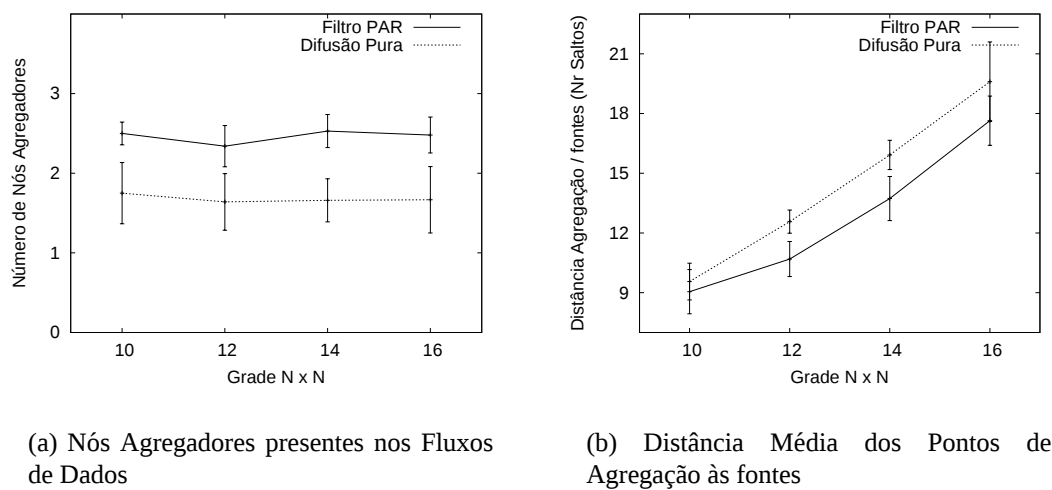


Figura 7: Parâmetros para Agregação em grade de N x N nós.

Analizamos também os efeitos do temporizador. Ele é necessário para que o PAR

tenha uma base de dados suficiente e possa decidir quais os nós com maior potencial de agregação, mas acrescenta custos no tempo necessário para descoberta das rotas. A figura 8 mostra os resultados obtidos. Nas grades maiores, os seus efeitos são acentuados, pois os fluxos passam por mais nós e o retardo em cada nó contribuiu para um tempo total inviável. Dessa forma, foi constatado que um temporizador entre 0,5 e 1 s apresenta bons resultados em termos de árvore de agregação, com menores custos em termos de latência.

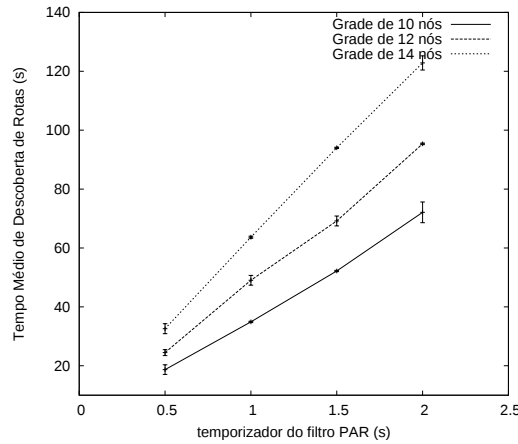


Figura 8: Tempo de Descoberta de Rotas.

Mecanismo de agregação

Avaliada a árvore de agregação formada, o próximo passo é a verificação da influência dos parâmetros do mecanismo de agregação vistos na seção 3.2, no desempenho da rede. Os gráficos das figuras 9 e 10 mostram a taxa de agregação no sorvedouro em relação a Difusão Direcionada pura, com a variação dos parâmetros do mecanismo de agregação. Foram usadas dois tipos diferentes de fontes, posicionadas no canto inferior direito da grade, e o sorvedouro no canto inferior esquerdo, com o Filtro Bayes realizando a agregação privilegiada. Os valores de V e D correspondem, respectivamente, a relação T_v/I_t e T_d/I_t .

O gráfico da figura 9(a) mostra que os valores de V próximos de D proporcionam uma maior taxa de agregação no sorvedouro. No entanto, um tempo de validade das evidências muito grande pode deixar a rede menos reativa a mudanças. Na figura 9(b), observa-se que nas curvas de quatro e seis fontes, um tempo de descarte baixo provoca uma elevada taxa de agregação, o que evidencia uma concentração de mensagens redundantes com pequeno atraso entre elas. Em seguida, na medida que D cresce a taxa de agregação sofre pequenas variações. Esse intervalo indica a melhor faixa de operação do mecanismo. A partir do ponto onde D se aproxima de um, a taxa volta a crescer, pois o mecanismo passa a considerar dados de um mesmo sensor como redundantes, o que pode alterar a taxa de chegada de mensagens no sorvedouro.

As curvas da figura 10 foram geradas com os Filtros PAR e DS. Ela mostra o efeito da manipulação da variação máxima admitida para o grau de certeza (C) das proposições, com $V = 0.8$ e $D = 0.8$. Pelo gráfico, pode-se verificar que quando C se aproxima de 0.10, a taxa de agregação tende para quarenta por cento (quatro fontes) e cinquenta por

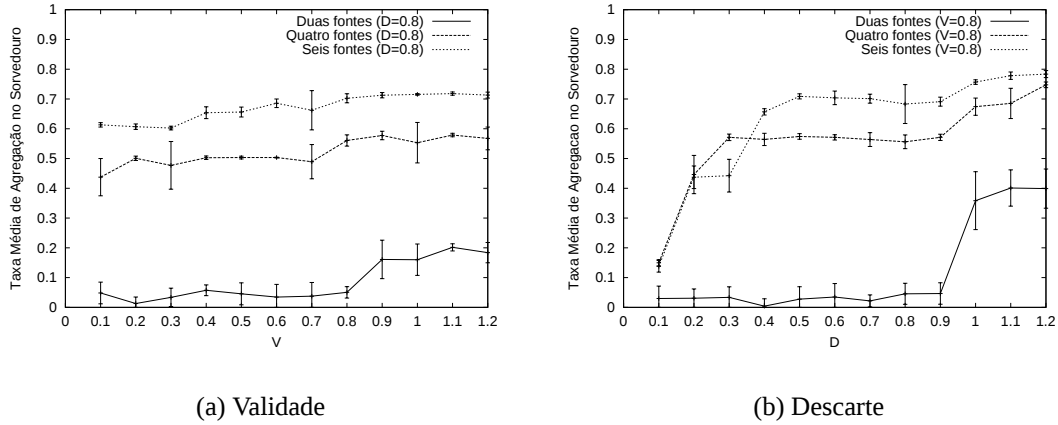


Figura 9: Mecanismo de Agregação: Tempo de validade e de descarte.

cento (seis fontes), o que garante uma redução significativa de mensagens no sorvedouro a um custo relativamente baixo para a confiança nas mensagens dos sensores.

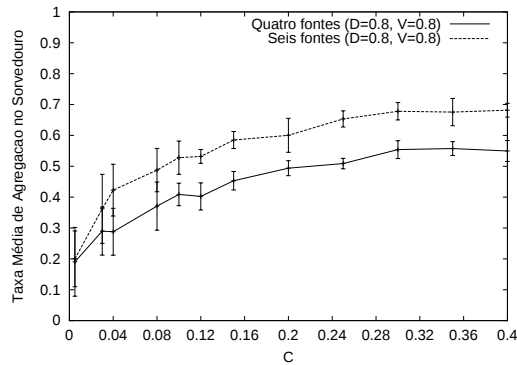
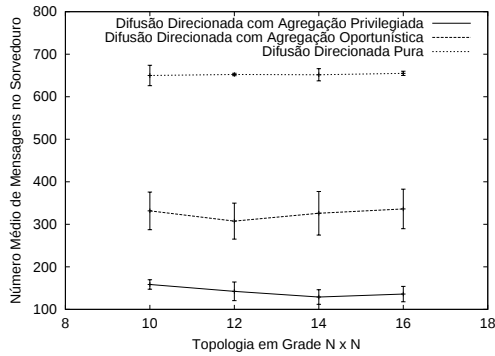


Figura 10: Variação do grau de certeza das proposições.

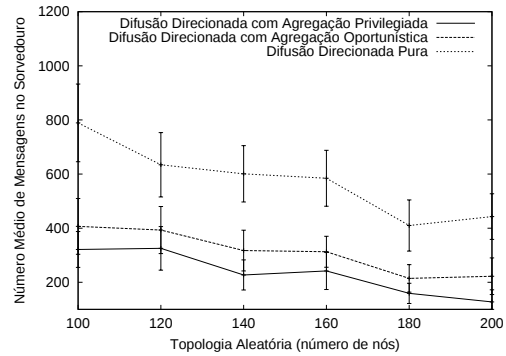
No que se refere ao número de transmissões, o aumento dos parâmetros V , D e C tendem a favorecer o desempenho da métrica (ou seja, reduzir o número médio de saltos) como observado nas simulações realizadas (estes gráficos não são apresentados neste artigo). Para a análise da economia de energia do modelo proposto, os filtros de agregação foram configurados com $V = 0,8$, $D = 0,8$ e $C = 0,1$.

Economia de energia

Finalmente, a economia de energia é analisada com o emprego das aplicações do Filtro Bayes e DS de forma distribuída na rede. A avaliação do desempenho foi feita de duas formas: com o Filtro Bayes em uma topologia regular, variando o tamanho da grade; e com o Filtro DS em uma topologia aleatória, variando o número de nós. Em ambos os casos, a quantidade de fontes é fixa. Duas métricas foram usadas na avaliação, comparando os resultados da Difusão Direcionada pura, do Bayes ou DS fazendo agregação oportunista e destes com o PAR. A primeira métrica diz respeito ao número de mensagens que chegam ao sorvedouro, mostrada na figura 11. Em 11(a) observamos que o aumento



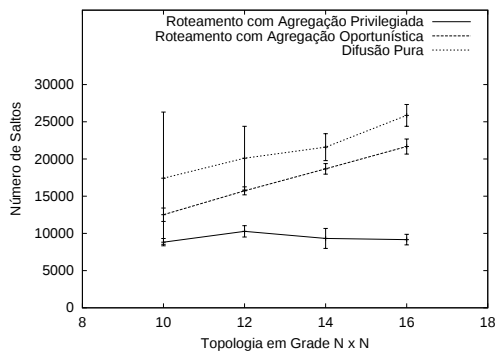
(a) Agregação com o Filtro Bayes



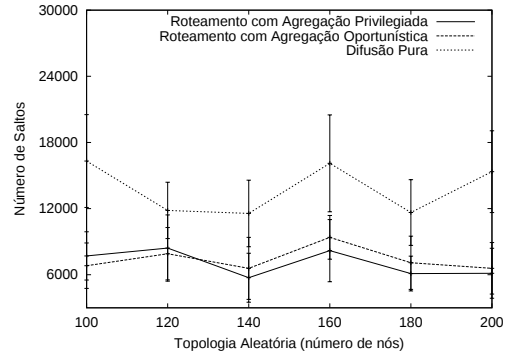
(b) Agregação com o Filtro DS

Figura 11: Mensagens no Sorvedouro.

da quantidade de nós na rede não teve influência significativa no desempenho de nenhuma das três situações, mas fica evidente a redução do número de mensagens no sorvedouro com a utilização do Bayes, para qualquer das grades utilizadas. Ao adicionar o PAR, com a conseqüente escolha dos melhores caminhos, a redução ocorre de forma drástica, da ordem de 80%. Em 11(b), com a topologia aleatória, verifica-se que o aumento da densidade de nós no campo tende favorecer o mecanismo de agregação, com vantagem para a agregação privilegiada.



(a) Agregação com o Filtro Bayes



(b) Agregação com o Filtro DS

Figura 12: Números de Saltos.

A segunda métrica diz respeito ao número total de transmissões (saltos) na rede. Isto é importante tendo em vista que o maior consumo de energia ocorre quando um nó transmite dados. Quanto mais longa é uma rota, maior é o número de saltos, mesmo que se mantenha o mesmo número de mensagens chegando no sorvedouro. Os fluxos obtidos com o PAR são naturalmente maiores que os da difusão pura, pois desviam-se da rota de menor retardo para buscar os nós agregadores, logo é importante que a árvore de agregação seja suficientemente eficiente para reduzir também o número de saltos. Os resultados estão na figura 12, e mostram que houve um ganho considerável.

5. Conclusões

Neste artigo, propomos um mecanismo que identifica nós agregadores e a seguir emprega neles técnicas paramétricas de agregação de dados de múltiplos sensores com o objetivo de aumentar o tempo de vida da rede. Neste esquema, dois filtros foram adicionados ao protocolo da Difusão Direcionada. O primeiro é chamado de PAR (Roteamento com Privilégios a Agregação) e atua na descoberta de rotas com maior potencial de possuírem nós agregadores. Estes nós são incluídos no caminho fontes-sorvedouro de forma a criar caminhos comuns entre fontes distintas e favorecer a agregação, com a contrapartida em um retardo no estabelecimento das rotas. O segundo filtro é o responsável pelo mecanismo de agregação. Exploramos aqui duas técnicas, a Inferência Bayesiana (Filtro Bayes) e o de Dempster-Shafer (Filtro DS). Ambas foram implementadas de forma a minimizar os custos da agregação em redes de sensores, uma vez que a agregação ocorre localmente em cada nó em função da informação disponível. Nossa abordagem também minimiza os impactos relacionados com a construção e manutenção de uma árvore de agregação. Mostramos por meio de simulação que nossa proposta reduz drasticamente o número de mensagens recebidas no sorvedouro, bem como o número total de transmissões (saltos) necessárias para que a tarefa da rede seja executada, além de ser flexível o suficiente para atender várias aplicações de redes de sensores.

Como trabalhos futuros, visualizamos o emprego do mecanismo de agregação com outros protocolos de roteamento centrado em dados, avaliando a adequação de outras técnicas paramétricas, assim como, a comparação entre elas. Outra abordagem a ser explorada é o emprego dos mecanismos propostos no presente trabalho em topologias aleatórias e dinâmicas, para verificação da sua eficiência em ambientes mais inóspitos. Além disso, a construção da árvore de agregação buscou a minimização do número de mensagens transmitidas, sem uma preocupação com a energia remanescente nas rotas escolhidas. Dessa forma, esta energia remanescente nos nós poderia ser também utilizada como parâmetro na escolha das rotas, no sentido de poupar nós agregadores muito freqüentes e, desta forma, equalizar o consumo de energia na rede, contribuindo para aumentar o seu tempo de vida.

Referências

- [1] AKYILDIZ, I. F., SU, W., SANKARASUBRAMANIAM, Y., AND CAYIRICI, E. A survey on sensor networks. *Communications of the ACM* (agosto de 2002), 102–114.
- [2] HEINZELMAN, W. R., CHANDRAKASAN, A., AND BALAKRISHNAN, H. Energy-efficient communication protocol for wireless microsensor networks. *Proceedings of the 33rd IEEE Hawaii International Conference on System Sciences (HICSS)* (janeiro de 2000), 1–10.
- [3] WALTZ, E., AND LLINAS, J. *Multisensor Data Fusion*. Artech House, 1990.
- [4] KRISHNAMACHARI, B., ESTRIN, D., AND WICKER, S. Modeling data-centric routing in wireless sensor networks. *IEEE INFOCOM* (junho de 2002).
- [5] INTANAGONWIWAT, C., GOVINDAN, R., AND ESTRIN, D. Directed diffusion: A scalable and robust communication paradigm for sensor networks. *Proceedings of ACM MobiCom'00, Boston, MA* (agosto de 2000), 56–67.

- [6] KULIK, J., HEINZELMAN, W. R., AND BALAKRISHMAN, H. Negotiation-based protocols for disseminating information in wireless sensor networks. *Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom'99)* (1999).
- [7] PETROVIC, D., SHAH, R. C., RAMCHANDRAN, K., AND RABAEY, J. Data funneling: Routing with aggregation and compression for wireless sensor networks. *IEEE Sensor Network Protocols and Applications (SNPA)* (maio de 2003).
- [8] GAREY, M. R., AND JOHNSON, D. S. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Freeman, San Francisco, 1979.
- [9] INTANAGONWIWAT, C., ESTRIN, D., GOVINDAN, R., AND HEIDEMANN, J. Impact of network density on data aggregation in wireless sensor networks. *Proceedings of the 22nd International Conference on Distributed Computing Systems* (julho de 2002).
- [10] KRISHNAMACHARI, B., ESTRIN, D., AND WICKER, S. The impact of data aggregation in wireless sensor networks. *International Workshop on Distributed Event-Based Systems, (DEBS '02)* (julho de 2002).
- [11] BOULIS, A., GANERIWAL, S., AND SRIVASTAVA, M. Aggregation in sensor networks: An energy-accuracy trade-off. *IEEE Sensor Network Protocols and Applications (SNPA)* (maio de 2003).
- [12] HALL, D. L. *Mathematical Techniques in Multi-Sensor Data Fusion*. Artech House, 1992.
- [13] HEIDEMANN, J., SILVA, F., YU, Y., ESTRIN, D., AND HALDAR, P. *Diffusion Filters as a Flexible Architecture for Event Notification in Wireless Sensor Networks*. USC/ISI Technical Report 2002-556, 2002.
- [14] [HTTP://WWW.ISI.EDU/NSNAM/NS](http://www.isi.edu/nsnam/ns). *The Network Simulator NS-2*.