

Detecção de Falhas em Redes de Sensores sem Fio Baseada na Medição do Tráfego e em Técnicas de Fusão de Dados

Eduardo F. Nakamura^{1,2}, Fabíola G. Nakamura¹,
Carlos M.S. Figueiredo^{1,2}, Antônio A.F. Loureiro¹

¹ Departamento de Ciência da Computação – DCC
Universidade Federal de Minas Gerais – UFMG
Av. Antônio Carlos, 6627 – 31270-010 Belo Horizonte, MG.

{nakamura, fgnaka, mauricio, loureiro}@dcc.ufmg.br

²Fundação Centro de Análise, Pesquisa e Inovação Tecnológica – FUCAPI
Av. Danilo de Matos Areosa, 381, Distrito Industrial – 69075-351 Manaus, AM.

{eduardo.nakamura, mauricio.figueiredo}@fucapi.br

Resumo. *A disseminação de dados é uma tarefa fundamental em Redes de Sensores sem Fio. Nessas redes os dados sensorizados são normalmente disseminados através de esquemas multi-saltos (redes planas), em virtude do alcance limitado dos rádios e de restrições no consumo de energia dos dispositivos. Frequentemente uma topologia em árvore é utilizada para disseminação de dados em redes planas. Entretanto, as soluções atuais não se preocupam com o momento em que esta topologia deve ser reconstruída. Este trabalho lida com este problema através do uso de mecanismos de fusão de dados, onde o tráfego é tratado como um sinal que é filtrado e traduzido em evidências que indicam a probabilidade de ocorrência de falhas críticas. Usando a teoria de Dempster-Shafer, essas evidências são combinadas para identificar a necessidade de reconstrução da topologia. A solução proposta é avaliada através de simulação e os resultados mostram que a fusão de dados é uma abordagem que pode ser utilizada para melhorar o desempenho de algoritmos de disseminação de dados para redes de sensores sem fio evitando tráfego desnecessário.*

Abstract. *One of the main tasks in a wireless sensor network is to deliver the sensed data to a monitoring node. Due to the limited radio range and energy constraints, a multihop approach is commonly adopted to disseminate the gathered data. In this multihop communication, data is frequently disseminated through a tree topology. However, to the best of our knowledge current solutions do not define the right moment when the dissemination topology needs to be reconstructed. To address this problem we handle the traffic as a signal that is filtered by a moving average filter. Then, the traffic is translated into evidences that indicate the failure probability. Finally, these evidences are combined, using the Dempster-Shafer theory, to determine when the topology needs to be reconstructed.*

1. Introdução

Redes de Sensores sem Fio (RSSFs) [Pottie and Kaiser, 2000, Akyildiz et al., 2002] constituem um tipo especial de redes *ad hoc* compostas por um grande número de nós com capacidade de sensoriamento. Essas redes possuem limitações de energia e de capacidade computacional. As RSSFs podem ser utilizadas em aplicações ambientais, médicas, industriais e militares. Este tipo de rede precisa ser capaz de adaptar-se de maneira autônoma a eventuais mudanças resultantes de intervenções externas, como mudanças topológicas, reação a um evento ou requisições de uma entidade externa.

O objetivo principal de uma RSSF é a coleta dados do ambiente e seu envio a um nó sorvedouro para posterior processamento e avaliação. Conseqüentemente, a disseminação de dados é uma tarefa fundamental que, devido à limitação do alcance dos rádios e às restrições de consumo, é tipicamente executada de forma plana em um esquema multi-saltos. De acordo com Tilak, Abu-Ghazaleh e Heinzelman [Tilak et al., 2002] a disseminação de dados pode ser executada segundo um modelo contínuo, onde a aplicação recebe continuamente os dados coletados do ambiente.

Topologias em árvore são freqüentemente usadas para disseminar dados em redes de sensores planas contínuas [Sohrabi et al., 2000, Krishnamachari et al., 2002, Zhou and Krishnamachari, 2003, Woo et al., 2003, Heidemann et al., 2003]. Neste cenário a Difusão Direcionada (Directed Diffusion) [Intanagonwiwat et al., 2000] provê uma variante chamada *One-Phase Pull Diffusion* [Heidemann et al., 2003] baseada na estrutura em árvore. Embora a topologia em árvore seja explorada em diferentes soluções [Krishnamachari et al., 2002, Zhou and Krishnamachari, 2003], nenhum dos trabalhos correntes considera o momento em que a árvore deve ser reconstruída. Estratégias como a reconstrução periódica [Zhou and Krishnamachari, 2003] ou a reconstrução solicitada pelo usuário podem resultar em reconstruções desnecessárias e/ou atrasadas.

Este trabalho propõe o uso de mecanismos de fusão de dados, o Filtro de Média Móvel (FMM) e a inferência de *Dempster-Shafer*, para prover uma solução viável que detecta de maneira autônoma quando a topologia de disseminação precisa ser reconstruída. De forma simples, a fusão de dados de sensores lida com a combinação de múltiplas fontes para melhorar a informação (mais barata, ou mais qualidade, ou mais relevância). A fusão de dados é normalmente usada em tarefas de detecção e classificação para aplicações militares e de robótica [Brooks and Iyengar, 1998]. Recentemente estes mecanismos têm sido usados em aplicações como detecção de intrusos [Bass, 2000] e de DoS (Denial of Service) [Siaterlis and Maglaris, 2004]. No domínio de RSSFs, técnicas simples de fusão de dados (e.g., métodos de agregação como *máximo* e *mínimo*) têm sido usadas para reduzir tráfego de dados e economizar energia [Intanagonwiwat et al., 2000, Krishnamachari et al., 2002]. Neste trabalho é introduzido o uso da fusão de dados para melhorar o desempenho de algoritmos de disseminação de dados em redes de sensores sem fio.

Este trabalho possui duas contribuições. Primeiro, a melhoria de algoritmos de disseminação reduzindo o número de construções de topologia. Segundo, a expansão da utilização de fusão de dados. Além disso, é mostrado como adequar métodos de fusão de dados a um problema específico.

O artigo está organizado como descrito a seguir. Os trabalhos relacionados são apresentados na Seção 2. A Seção 3 provê uma definição formal e uma análise do pro-

blema tratado, bem como define seu escopo. A Seção 4 detalha como foram adequados os mecanismos de fusão de dados ao problema tratado. Na Seção 5 o protótipo desenvolvido é avaliado através de um conjunto de simulações. Finalmente, na Seção 6 são apresentadas as conclusões e trabalhos futuros.

2. Trabalhos Relacionados

Entre as estratégias correntes de disseminação para redes de sensores planas [Intanagonwiwat et al., 2000, Sohrabi et al., 2000, Krishnamachari et al., 2002, Zhou and Krishnamachari, 2003, Woo et al., 2003, Heidemann et al., 2003] o uso da topologia em árvore [Sohrabi et al., 2000, Krishnamachari et al., 2002, Zhou and Krishnamachari, 2003, Woo et al., 2003] destaca-se pela simplicidade e eficiência [Woo et al., 2003].

O algoritmo SAR (Sequential Assignment Routing) [Sohrabi et al., 2000] cria múltiplas árvores. Cada árvore é construída evitando nós com baixos valores de QoS (e.g., baixa vazão/alto atraso) e baixa energia residual. No final do processo, a maioria dos nós pertencem a múltiplas árvores, permitindo que um nó sensor escolha a melhor árvore para transmitir sua informação até o sorvedouro.

Em [Krishnamachari et al., 2002], Krishnamachari, Estrin, e Wicker avaliam três estratégias diferentes para construção de uma árvore de disseminação. Na estratégia CNS (*Center at Nearest Source*), cada fonte envia seus dados diretamente ao nó fonte mais perto do sorvedouro. Na estratégia SPT (*Shortest Paths Tree*), cada fonte envia seus dados para o sorvedouro através do caminho mínimo entre os dois nós. A última estratégia, GIT (*Greedy Incremental Tree*), é um esquema seqüencial onde a árvore de disseminação consiste inicialmente do caminho mínimo entre o sorvedouro e o nó fonte mais próximo dele, a cada passo o nó fonte que está mais próximo da árvore corrente é incluído na árvore.

Em [Zhou and Krishnamachari, 2003], Zhou e Krishnamachari avaliam a topologia em árvore com quatro estratégias diferentes de seleção de pais. Na estratégia EF (*Earliest-first*), cada nó escolhe o primeiro candidato como seu pai. Na estratégia Aleatória, o nó escolhe seu pai aleatoriamente e com igual probabilidade dentre todos os possíveis candidatos. Na NF (*Nearest-first*), o nó escolhe o candidato mais próximo como pai. Finalmente, na WR (*Weighted-randomized*), cada nó atribui um peso para cada possível pai de acordo com seu número de vizinhos; a seleção do pai é aleatória com estes pesos; o nó com mais vizinhos terá o menor peso.

O Algoritmo *One-Phase Pull Diffusion* [Heidemann et al., 2003] é uma variação da Difusão Direcionada [Intanagonwiwat et al., 2000] que constrói uma árvore de disseminação. Neste algoritmo não há nenhum dado exploratório, o nó sorvedouro simplesmente dissemina seu interesse e os nós fontes enviam os dados aos vizinhos que primeiro enviaram o interesse (construindo assim uma árvore de disseminação de dados). A topologia de disseminação resultante comporta-se exatamente como a estratégia EF [Zhou and Krishnamachari, 2003].

Zhou e Krishnamachari [Zhou and Krishnamachari, 2003] sugerem reconstruções periódicas para recuperação de eventuais falhas de nós e *One-Phase Pull Diffusion* [Heidemann et al., 2003] permite disseminação de interesses periódica e requisitada pelo

usuário. Entretanto, nenhuma destas estratégias detecta de maneira autônoma quando a topologia precisa ser reconstruída.

3. Problema Tratado

Esta seção provê uma definição formal do problema tratado e define seu escopo. Além disso, é apresentada uma investigação mais detalhada do problema.

3.1. Formulação do Problema

Para fins de simplificação, este trabalho assume que uma topologia em árvore é usada para disseminar os dados até o nó sorvedouro. Assim, o problema pode ser definido e delimitado como segue.

Definição 3.1 (Escopo do Problema) *Considere um rede de sensores plana contínua como um nó sorvedouro e $n - 1$ nós fontes gerando pacotes de dados a uma taxa \mathbf{R} durante um tempo de vida \mathbf{T} . A topologia da rede (árvore de disseminação) é representada como um grafo direcionado $G = (V, E)$, com as seguintes propriedades:*

- $V = \{s_0, s_1, s_2, \dots, s_{n-1}\}$, onde s_0 é o único nó sorvedouro, e s_i é um nó fonte $\forall 0 < i < n$;
- $(i, j) \in E$, se e somente se, s_j é o pai de s_i , i.e., s_i envia seu dados para o sorvedouro através de s_j ;
- Se $(i, j) \in E$ e $(i, k) \in E$, então $j = k, \forall 0 < i, j, k < n$;

Definição 3.2 (Definição do Problema) *Dado que $G = (V, E)$ é inicialmente construído no instante $t = 0$, deseja-se determinar em qual instante $t \in (0, \mathbf{T}]$ o grafo $G = (V, E)$ precisa ser reconstruído.*

3.2. Investigação do Problema

O tráfego indica qual a quantidade de dados que está sendo entregue por unidade de tempo. Esta medida pode ser expressa em diversas unidades como *bits por segundo* (bps), *bytes por segundo* (B/s) ou *pacotes por segundo* (pps). A última unidade é uma medida fraca quando o tamanho dos pacotes é diferente. Entretanto, para simplificação, este trabalho considera que os pacotes possuem o mesmo tamanho. O tráfego de dados pode ser tratado como um sinal discreto $\delta(t)$ que calcula a quantidade de pacotes recebidos durante um intervalo de tempo \mathbf{S} (taxa de amostragem). Assim, uma amostra de tráfego medida em um instante (tempo de amostragem) t_k é dada por:

$$\delta(t_k) = p_k / \mathbf{S}. \quad (1)$$

onde p_k é o número de pacotes recebidos pelo sorvedouro durante o intervalo $\mathbf{S} = t_k - t_{k-1}$, e t_{k-1} é o tempo de amostragem anterior. Para redes contínuas a taxa de dados \mathbf{R} permanece a mesma para todos os nós durante todo o tempo de vida da rede. Assim, fazendo $\mathbf{S} \equiv \mathbf{R}$ é possível obter uma boa estimativa do tráfego de dados em redes contínuas. Infelizmente, o tráfego medido é ainda está sujeito a ruídos em virtude de perda de pacotes, atraso em filas, e falta de sincronismo entre relógios.

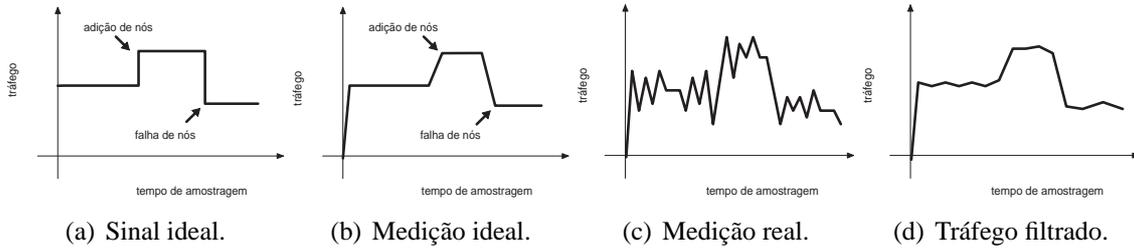


Figura 1: Comportamento do sinal de tráfego.

Idealmente, em um cenário de coleta contínua de dados, o tráfego permanece o mesmo até que novos nós sejam adicionados – levando a um aumento no tráfego – ou falhas aconteçam – levando a um decaimento do tráfego (Fig. 1(a)). Embora a medida de tráfego ideal (Fig. 1(b)) não possa ser alcançada em virtude do ruído embutido no sinal (Fig. 1(c)), a medida bruta pode ser filtrada para prover uma estimativa mais realista (Fig. 1(d)).

Assim, tratando o tráfego $\delta(t)$ como uma função temporal discreta, dado um período de amostragem S e uma amostra arbitrária $t(k)$, no cenário proposto (redes contínuas), é válido afirmar que $\forall k > 0 \in \mathbb{N}$, durante $[t_{k-1}, t_k]$:

- $t_k = t_{k-1} + S$, onde t_{k-1} é o tempo de amostragem anterior;
- Se nenhum nó é adicionado e falhas ocorrem, então $\delta(t_k) < \delta(t_{k-1})$;
- Se nenhum nó é adicionado e nenhuma falha ocorre, então $\delta(t_k) = \delta(t_{k-1})$;
- Se nós são adicionados e nenhuma falha ocorre, então $\delta(t_k) > \delta(t_{k-1})$;

O impacto das falhas depende da carga de trabalho do nó que falhou. A falha de um nó folha é chamada *falha periférica*, enquanto que a falha de um nó de retransmissão é chamada de *falha de roteamento*. Neste caso, quanto maior a sub-árvore desconectada, mais crítica é a falha. Portanto, o desafio é identificar quando uma falha crítica ocorre para que a árvore de disseminação seja reconstruída.

Uma falha de roteamento deve resultar em uma queda maior de tráfego quando comparada a uma falha periférica, conforme Figura 1(a). Na verdade, grandes quedas no tráfego significam que ocorreram poucas falhas de roteamento ou várias falhas periféricas, ambas indesejáveis. Assim, o sinal de tráfego $\delta(t)$ pode prover informação suficiente para decidir quando a topologia da rede deve ser reconstruída. Sempre que um grande decaimento de tráfego ocorrer a topologia deve ser reconstruída. Se o nível de tráfego crescer de novo significa que a rede se recuperou da falha. Caso contrário, a árvore de disseminação não se recuperou da falha e outra ação deve ser tomada, (e.g., adição de novos nós).

4. Protótipo

Na solução proposta, as medidas de tráfego são tratadas como um sinal, que é utilizado para inferir se a topologia da rede precisa ser reconstruída. Sob condições ideais, o decaimento de tráfego ocorre em função de falhas, e grandes decaimentos são causados por falhas de roteamento. Entretanto, como citado na Seção 3.2, o tráfego medido deve ser filtrado para reduzir ruídos no sinal.

4.1. Filtrando o Tráfego

O comportamento do tráfego (Fig. 1(a)) é muito similar a uma função degrau. Conseqüentemente, o filtro de média móvel (FMM) é a melhor escolha para filtrar o sinal de tráfego, dado que o FMM é ótimo para reduzir o ruído gaussiano aleatório mantendo a convergência para o degrau a mais rápida possível [Smith, 1999]. O filtro calcula a média aritmética do número de entradas medidas para produzir cada ponto do sinal de saída:

$$output[i \geq m] = \frac{1}{m} \sum_{k=1}^m input[i - k] \quad (2)$$

onde m é a janela do filtro, que corresponde ao número de amostras de entrada que são fundidas em uma amostra de saída. O valor m também representa o número de passos que o filtro executa até detectar uma mudança no nível do sinal. Quanto menor o valor de m , mais rápido o sinal converge para o degrau; e quando maior o valor de m , mais limpo o sinal será.

O nó sorvedouro aplica o filtro de média móvel no ruído embutido no tráfego medido. O tamanho da janela m deve ser escolhido baseado nas características do ruído no tráfego, que é influenciado pelo ambiente de rede, e no tempo de resposta esperado.

4.2. Extraindo Características

Uma vez que o tráfego é filtrado, são extraídas duas características do sinal: *decaimento instantâneo* e *decaimento a longo prazo*. O decaimento instantâneo avalia como o tráfego mudou entre duas amostra consecutivas. O decaimento a longo prazo mostra como o tráfego mudou desde a última reconstrução da topologia.

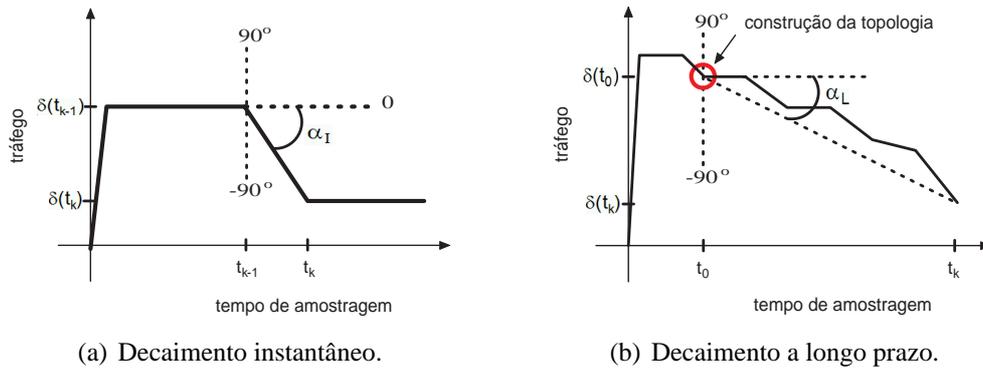


Figura 2: Características do tráfego.

4.2.1. Decaimento Instantâneo

Dadas duas amostras em uma seqüência, t_{k-1} and t_k (ver Figura 2(a)), o *ângulo de decaimento instantâneo* é definido como

$$\alpha_i = \arctan \frac{\delta(t_k) - \delta(t_{k-1})}{t_k - t_{k-1}}, \quad (3)$$

e o ângulo de decaimento máximo como

$$\alpha_{imax} = \arctan \frac{-\delta(t_{k-1})}{t_k - t_{k-1}}. \quad (4)$$

Dados α_i e α_{imax} , o decaimento instantâneo é definido como

$$\phi_i = \frac{\alpha_i}{\alpha_{imax}}. \quad (5)$$

4.2.2. Decaimento a Longo Prazo

Dada a amostra corrente, t_k (see Figure 2(b)), o ângulo de decaimento a longo prazo é definido como

$$\alpha_l = \arctan \frac{\delta(t_k) - \delta(t_0)}{t_k - t_0}, \quad (6)$$

e o ângulo de decaimento a longo prazo máximo como

$$\alpha_{lmax} = \arctan \frac{-\delta(t_0)}{t_k - t_0}. \quad (7)$$

Então, o decaimento a longo prazo pode ser definido em função de α_l and α_{lmax} como

$$\phi_l = \frac{\alpha_l}{\alpha_{lmax}}. \quad (8)$$

Ambas as características, ϕ_i e ϕ_l , são definidas em termos do ângulo de decaimento ao invés do simples decaimento de tráfego. Isto porque o ângulo também considera o tempo gasto na mudança do nível de tráfego.

4.3. Encontrando o Estado do Sistema

Para esta tarefa foi escolhida a Inferência de *Dempster-Shafer* [Shafer, 1976], porque generaliza a teoria Bayesiana e provê um formalismo que pode ser usado para representar conhecimento parcial, atualizar crenças e combinar evidências [Provan, 1992]. Além disso, comparada com o Inferência Bayesiana, a teoria de Dempster-Shafer está mais próxima da percepção e do raciocínio humanos.

Um conceito fundamental na teoria de Dempster-Shafer é o conjunto de discernimento, que contém todos os estados possíveis de um sistema $\Theta = \{\theta_1, \theta_2, \dots, \theta_N\}$. O conjunto de discernimento deve ser exaustivo e mutuamente exclusivo, significando que o sistema só pode está em um e apenas um estado $\theta_k \in \Theta$.

Os elementos do conjunto 2^Θ são chamados hipóteses. Baseado em uma evidência E , uma probabilidade é atribuída a cada hipótese $H \in 2^\Theta$ segundo uma *atribuição básica de probabilidade* (bpa - basic probability assignment), ou função de massa, $m : 2^\Theta \rightarrow [0, 1]$ que satisfaz:

$$m(\emptyset) = 0 \quad (9)$$

$$m(H) \geq 0, \forall H \in 2^\Theta \quad (10)$$

$$\sum_{H \in 2^\Theta} m(H) = 1. \quad (11)$$

Na teoria de Dempster-Shafer, a crença total em um hipótese H é expressa por uma função de crença $bel : 2^\Theta \rightarrow [0, 1]$ sobre Θ como:

$$bel(H) = \sum_{A \subseteq H} m(A) \quad (12)$$

onde $bel(\emptyset) = 0$, e $bel(\Theta) = 1$. A plausibilidade de cada hipótese, é expressa pela função $pl : 2^\Theta \rightarrow [0, 1]$ sobre Θ como:

$$pl(H) = 1 - bel(\neg H) \quad (13)$$

Intuitivamente os estados plausíveis atestam que quanto menos duvidosa for a hipótese H , mais plausível ela é. Neste contexto, o intervalo de confiança $[bel(H), pl(H)]$ define a crença real da hipótese H .

Para combinar os efeitos de duas bpa's m_1 e m_2 , a teoria de Dempster-Shafer define a regra de combinação, $m_1 \oplus m_2$, dada por:

$$m_1 \oplus m_2(\emptyset) = 0, \quad (14)$$

$$m_1 \oplus m_2(H) = \frac{\sum_{X \cap Y = H} m_1(X)m_2(Y)}{1 - \sum_{X \cap Y = \emptyset} m_1(X)m_2(Y)}. \quad (15)$$

Os estados considerados pelo nó sorvedouro são: NORMAL e CRITICO. O estado NORMAL descreve a ausência de falhas ou a ocorrências de *falhas periféricas* na rede. O estado CRITICO descreve a ocorrência de *falhas de roteamento*. Assim, o conjunto de discernimento é $\Theta = \{\text{NORMAL}, \text{CRITICO}\}$.

Para aplicar a teoria de Dempster-Shafer as características de tráfego, ϕ_i (decaimento instantâneo) e ϕ_l (decaimento a longo prazo), são traduzidas em evidências que são fundidas pela Expressão (15).

Para o caso particular de redes de sensores sem fio contínuas, se $0 < \phi_i \leq 1$, então existe uma probabilidade não nula de uma falha de roteamento ter ocorrido e se $\phi_i = 0$ nenhuma falha ocorreu. Assim, a atribuição básica de probabilidade $m_i : 2^\Theta \rightarrow [0, 1]$ é definida como:

$$m_i(\text{CRITICO}) = \begin{cases} (\phi_i)^w, & 0 < \phi_i \leq 1; \\ 0, & \text{caso contrário} \end{cases} \quad (16)$$

$$m_i(\text{NORMAL}) = 1 - m_i(\text{CRITICO})$$

onde $w > 0$, $w \in \mathbb{R}$ é chamado peso de decaimento.

Assumindo que as mesmas observações são válidas para o decaimento a longo prazo, a atribuição básica de probabilidade de decaimento $m_l : 2^\Theta \rightarrow [0, 1]$ é definida

como:

$$\begin{aligned} m_l(\text{CRITICO}) &= \begin{cases} (\phi_l)^w, & 0 < \phi_l \leq 1; \\ 0, & \text{caso contrário} \end{cases} \\ m_l(\text{NORMAL}) &= 1 - m_l(\text{CRITICO}) \end{aligned} \quad (17)$$

Observe que ambas bpa's, m_i e m_l , atribuem probabilidades complementares às hipóteses primitivas do conjunto de discernimento $\Theta = \{\text{NORMAL}, \text{CRITICO}\}$. Isto significa que m_i e m_l são função de crença Bayesiana, onde não existe intervalo de confiança das hipóteses primitivas, isto é, a plausibilidade e a crença de cada hipótese primitiva são iguais. Portanto, o estado da rede é estimado combinando-se as probabilidades atribuídas a m_i e m_l com a Expressão (15). No segundo passo, é calculada a plausibilidade (13) de cada hipótese em função $m_i \oplus m_l$. A seguir, o estado mais plausível é escolhido como o estado atual da rede. Se ambos os estados são igualmente plausíveis é escolhido o estado NORMAL.

Sempre que o nó sorvedouro conclui que a rede está no estado CRITICO a topologia de disseminação é reconstruída para encontrar, se possível, rotas alternativas para nós que pararam de entregar dados em virtude da falha de um nó roteador. Se a rede estiver no estado NORMAL nada é feito.

5. Avaliação

A metodologia e os resultados dos experimentos realizados são apresentados e discutidos nesta seção.

5.1. Metodologia

A avaliação é executada através de simulações que comparam a solução proposta com uma árvore de reconstrução periódica, o algoritmo *One-Phase Pull Diffusion* [Heidemann et al., 2003] com disseminação de periódica interesse. É apresentado o comportamento da solução proposta para valores de $w \in \{\frac{1}{1}, \frac{1}{3}, \frac{1}{5}, \frac{1}{7}\}$, que é o peso de decaimento usado para traduzir as características em evidências (Seção 4.3). Para o filtro de média móvel, foi escolhido de forma empírica $m = 5$ como janela do filtro, porque quando $m < 5$ o ruído remanescente é muito grande e quando $m > 5$ o filtro demora um tempo excessivo para convergir.

Os experimentos foram executados no simulador ns-2 [NS-2, 2004]. Todos os experimentos foram executados com 33 diferentes sementes e foram calculados a média aritmética e o intervalo de confiança para 95% de confiança. Os gráficos desta seção incluem os valores médios com o intervalo de confiança representado por barras de erro.

5.1.1. Cenários

Os parâmetros de simulação são baseados no nó sensor Mica2 [Crossbow, 2004]: potências de transmissão, recepção e sensoriamento são 45.0 mW, 24.0 mW e 15.0 mW, respectivamente; a largura de banda é 19.2 kbps; e o raio de comunicação é fixado em 40m. A camada de acesso ao meio (MAC) usa o protocolo 802.11.

Em todos os cenários, o nó sorvedouro está localizado no canto inferior esquerdo (0, 0) da área de monitoramento. A densidade dos cenários foi fixada em 0,005 nós/m², e a área (quadrada) varia de acordo com o número de nós (posicionados aleatoriamente). Os pacotes de dados de controle possuem 20 bytes. A taxa de transmissão de dados escolhida é de um pacote a cada 20s. Para o algoritmo *One-Phase Pull Diffusion* (1PP Diffusion) os interesses são disseminados a cada 200s. O tempo de simulação para todos os cenários é de 4000s. O protótipo usa o mesmo roteamento em árvore do 1PPDiffusion, entretanto, as reconstruções são disparadas de forma reativa, i.e., somente quando o estado CRITICO é detectado.

5.1.2. Modelo de Falhas

A estratégia de inferência é avaliada com um modelo de falhas independente. Neste modelo, falhas ocorrem com um processo de Poisson, onde o tempo entre falhas sucessivas é representado por uma variável aleatória exponencial independente com uma taxa λ constante (a taxa de falhas é medida em falhas por segundo).

5.1.3. Métricas

As métricas de avaliação são: taxa de entrega, atraso por salto, número de perdas em filas e número de construção de topologia. A taxa de entrega provê uma medida de eficácia sobre a habilidade da rede para entregar os dados coletados. O atraso por salto e as perdas em filas avaliam o impacto das construções da rede no tráfego total. O número de construções, com a taxa de entrega, provê meios para avaliar quão freqüente é necessário reconstruir a topologia da rede.

5.2. Resultados

A escalabilidade é avaliada variando o tamanho da rede de 100 a 175 nós com uma taxa de falhas constante ($\lambda = 0.0025$ falhas/s), e os resultados são mostrados na Figura 3. Em relação a taxa de entrega (Figura 3(a)), usando $w = 1/5$ e $w = 1/7$, a rede entrega quase tantos pacotes quanto o *One-Phase Pull Diffusion* (1PP Diffusion), e quando o tamanho da rede é maior (175 nós), o *One-Phase Pull Diffusion* começa a sofrer uma saturação em virtude do número extra de construções de topologia.

Aumentando o tamanho da rede o tráfego total também aumenta (especialmente em nós próximos ao sorvedouro). Como resultado, o atraso (Figura 3(b)) e o número de perdas em filas (Figura 3(c)) também crescem, principalmente para os casos em que a rede é grande (175 nós). O impacto de construções de topologia extras executadas pelo algoritmo *One-Phase Pull Diffusion* é ilustrado na Figura 3(b), que mostra que os pacotes consomem mais tempo para serem encaminhados com 175 nós, e na Figura 3(c), que mostra que a rede descarta mais pacotes em virtude de estouro na fila.

A Figura 3(d) mostra as construções da topologia da rede. O número de construções tende a decrescer com o tamanho da rede quando $w = 1/5$ e $w = 1/7$. O motivo é que quando o tráfego cresce, o impacto de uma falha é reduzido, e menos construções são necessárias. Entretanto, quando a rede começa a perder mais pacotes

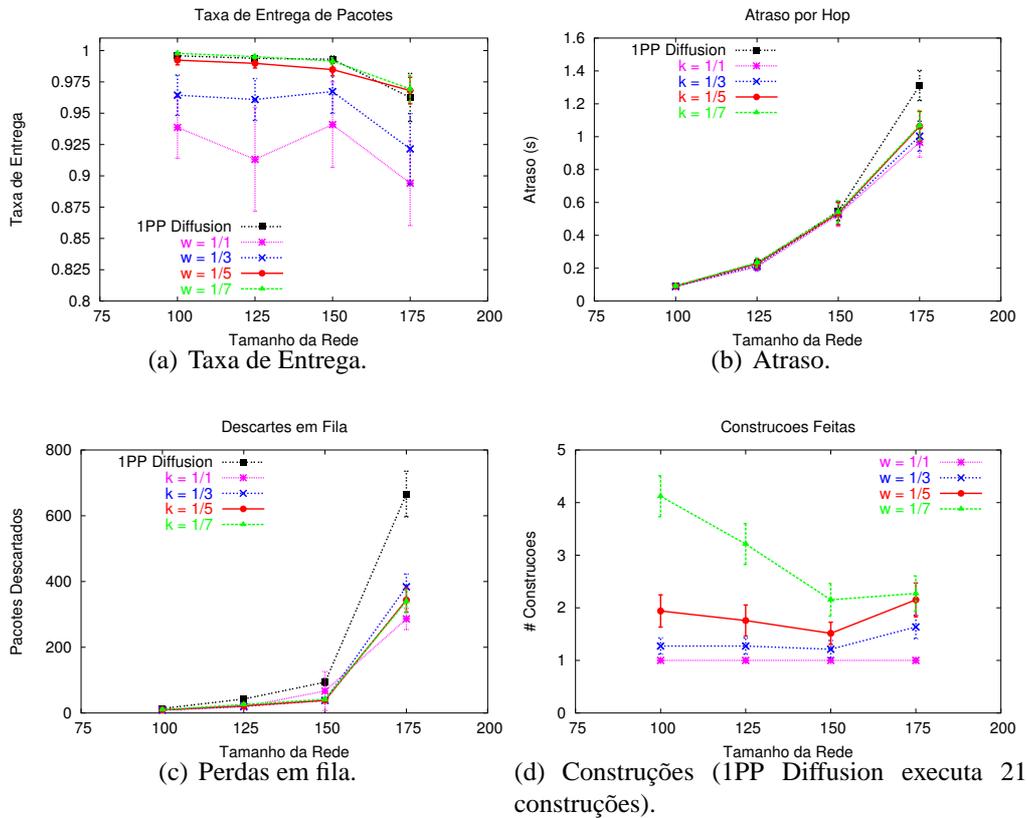


Figura 3: Escalabilidade em um cenário com falhas independentes.

em filas (175 nós), essas perdas contam como falhas e mais construções de topologia são requeridas (Figura 3(d)). Em geral, é possível concluir que para os cenários avaliados com uma construção adicional (Figura 3(d)), $w = 1/5$ e *One-Phase Pull Diffusion* (Figura 3(a)) entregam quase todos os pacotes de dados, mas $w = 1/5$ é menos afetado pelo tráfego causado por construções desnecessárias (Figuras 3(b) e 3(c)). Isto representa uma redução de aproximadamente 90% no número de construções executadas pela rede.

A robustez é avaliada com taxas de falhas (λ) de 0.0025, 0.005, 0.01 e 0.015 falhas/s, com um número fixo de 150 nós na rede. Os resultados são mostrados na Figura 4. As taxas de entrega (Figura 4(a)) de $w = 1/5$, $w = 1/7$ e *One-Phase Pull Diffusion* são praticamente as mesmas. Além disso, $w = 1/5$ e $w = 1/7$ e *One-Phase Pull Diffusion* recuperam-se com sucesso de falhas fazendo a taxa de entrega quase constante independente da taxa de falhas.

O tempo demandado para processar um pacote (Figura 4(b)) é quase o mesmo para todos os algoritmos simulados. Este tempo diminui conforme a taxa de falhas aumenta. A razão para isto é que quando a taxa de falhas cresce o tráfego total decresce reduzindo o atraso nas filas.

Em relação às perdas em filas mostradas na Figura 4(c), o *One-Phase Pull Diffusion* começa a perder menos pacotes quando o número de falhas aumenta. Este comportamento é causado também pela diminuição do tráfego. Por outro lado, $w = 1/5$ e $w = 1/7$ descartam menos pacotes conforme o número de falhas cresce porque o número de construções de topologia cresce rapidamente com o número de falhas (Figura 4(d)). Entretanto, mesmo com o número crescente de construções, a solução proposta ainda reconstrói a rede um número menor de vezes que o *One-Phase Pull Diffusion*, o que resulta também em um número menor de perdas em filas.

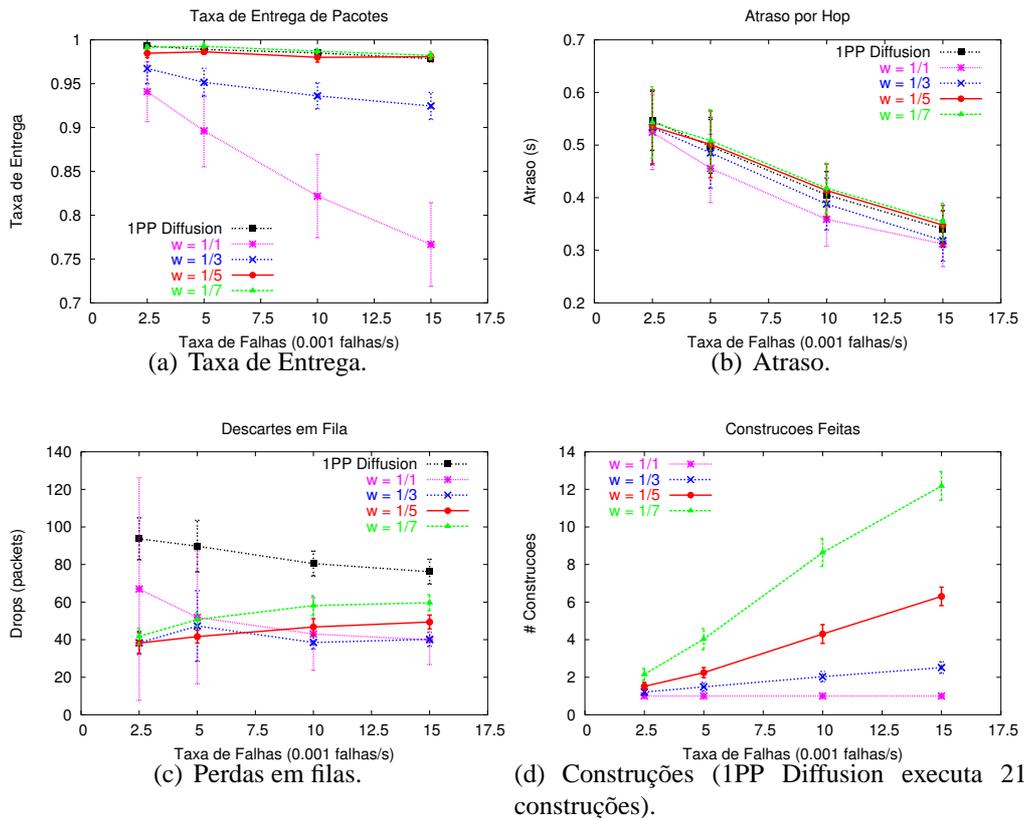


Figura 4: Robustez em um cenário com falhas independentes.

Como comentário geral, observa-se que com peso de decaimento baixo ($w = 1/1$), as barras de erro são grandes. Isto ocorre porque o algoritmo não reconstrói a topologia, tornando-se instável.

6. Considerações Finais

Este trabalho introduz o uso de mecanismos de fusão de dados (Filtro de Média Móvel e Inferência de Dempster-Shafer) para determinar quando a topologia de disseminação de dados precisa ser reconstruída em uma RSSF. Esta abordagem evita construções de topologia desnecessárias. Em alguns casos, apenas uma construção adicional é suficiente para garantir a entrega dos dados (o que representa uma redução de 90% no número de construções de topologia). Outras contribuições incluem a exemplificação de como

mecanismos de fusão de dados podem ser usados em outras aplicações e a adequação desses mecanismos aos requerimentos e limitações de um problema específico.

A avaliação atual explora redes de sensores sem fio planas com coleta e disseminação de dados contínuas. Trabalhos futuros incluem a extensão do uso desta abordagem para redes de sensores que realizam agregação de dados e redes de sensores dirigidas a eventos. A agregação de dados introduz novos desafios uma vez que o tráfego total é naturalmente reduzido. Para redes dirigidas a eventos, o tráfego supostamente aumenta quando os eventos são detectados e diminui quando os eventos deixam de ser detectados, independente da ocorrência de falhas.

Referências

- Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., and Cyirci, E. (2002). Wireless sensor networks: A survey. *Computer Networks*, 38(4):393–422.
- Bass, T. (2000). Intrusion detection systems and multisensor data fusion. *Communications of the ACM*, 43(4):99–105.
- Brooks, R. R. and Iyengar, S. S. (1998). *Multi-Sensor Fusion: Fundamentals and Applications with Software*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
- Crossbow (2004). Mica2. <http://www.xbow.com>.
- Heidemann, J., Silva, F., and Estrin, D. (2003). Matching data dissemination algorithms to application requirements. In *Proceedings of the SenSys'03*, pages 218–229, Los Angeles, CA, USA.
- Intanagonwiwat, C., Govindan, R., and Estrin, D. (2000). Directed diffusion: A scalable and robust communication paradigm for sensor networks. In *Proc. of the 6th ACM Intl. Conference on Mobile Computing and Networking*, pages 56–67, Boston, MA, USA.
- Krishnamachari, B., Estrin, D., and Wicker, S. (2002). The impact of data aggregation in wireless sensor networks. In *Proc. of the 22nd Intl. Conference on Distributed Computing Systems*, pages 575–578, Vienna, Austria.
- NS-2 (2004). The network simulator. <http://www.isi.edu/nsnam/ns/>.
- Pottie, G. J. and Kaiser, W. J. (2000). Wireless integrated network sensors. *Communications of the ACM*, 43(5):51–58.
- Provan, G. M. (1992). The validity of dempster-shafer belief functions. *International Journal of Approximate Reasoning*, 6(3):389–399.
- Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princenton University Press, Princenton, NJ, USA.
- Siaterlis, C. and Maglaris, B. (2004). Towards multisensor data fusion for DoS detection. In *Proc. of the 2004 ACM Symposium on Applied Computing*, pages 439–446, Nicosia, Cyprus.
- Smith, S. W. (1999). *The Scientist and Engineer's Guide to Digital Signal Processing*. California Technical Publishing, San Diego, CA, USA, 2nd edition.

- Sohrabi, K., Gao, J., Ailawadhi, V., and Pottie, G. J. (2000). Protocols for self-organization of a wireless sensor network. *IEEE Personal Communications*, 7(5):16–27.
- Tilak, S., Abu-Ghazaleh, N. B., and Heinzelman, W. (2002). A taxonomy of wireless micro-sensor network models. *SIGMOBILE Computing and Communications Review*, 6(2):28–36.
- Woo, A., Tong, T., and Culler, D. (2003). Taming the underlying challenges of reliable multihop routing in sensor networks. In *Proc. of the 1st International Conference on Embedded Networked Sensor Systems*, Los Angeles, California, USA.
- Zhou, C. and Krishnamachari, B. (2003). Localized topology generation mechanisms for self-configuring sensor networks. In *Proceedings of the IEEE GLOBECOM 2003*, volume 22, pages 1269–1273, San Francisco, USA.