

# Um Análise da Distorção dos Dados na Coleta de Dados em Agrupamentos Formados a Partir da Similaridade Espacial

Sérgio Samuel Furlaneto <sup>1</sup>, Michel Albonico <sup>2</sup>

<sup>1</sup>Universidade Federal do Paraná (UFPR)  
Caixa Postal 81530-900 – Curitiba – PR – Brazil

sersamfur@gmail.com

<sup>2</sup>Departamento de Sistemas e Computação  
Universidade Regional de Blumenau (FURB) – Blumenau, SC – Brazil

michel@inf.ufrgs.br

## 1. Introdução

Os sensores são dispositivos eletro-mecânicos capazes de captar estímulos do ambiente e convertê-los em unidades de medidas como temperatura, pressão, velocidade e umidade, entre outros. Tais dispositivos apresentam características físicas como baixa capacidade computacional, pouca memória, certa capacidade de comunicação e uma reserva pequena e finita de energia, geralmente na forma de bateria. Uma vez que sua bateria esteja descarregada o dispositivo estará fora de operação, pois devido à natureza da maioria das aplicações de sensores, particularmente o sensoriamento remoto, torna-se inviável qualquer plano de substituição ou recarga das baterias [Akyildiz et al. 2002].

Do ponto de vista lógico os dispositivos sensores devem ser auto-configuráveis e auto-gerenciáveis, e uma vez dispersos em área devem ser capazes de formar topologias de redes complexas, definir políticas de segurança e replicar dados. Todas estas atividades devem ser realizadas pelos dispositivos sem qualquer interferência humana.

As redes de sensores sem fio (RSSF) consistem em dispositivos sensores organizados segundo um critério e encarregados do sensoriamento de determinada área dependendo do ambiente e tipo da aplicação podem ser composta de centenas ou milhares de dispositivos. Os dispositivos mais próximos à estação base comunicam-se diretamente com ela, já dispositivos em regiões mais remotas da rede comunicam-se com a estação base através da cooperação dos outros dispositivos no repasse de mensagens e com o uso de protocolos de roteamento. Logo os sensores precisam se auto-organizar, e para isso se utilizam da troca de mensagens de controle com o objetivo de estabelecer uma forma de comunicação e de prover um *backbone* para os dados sensorizados.

A coleta dos dados em redes de sensores sem fio de médio e grande porte é uma tarefa muito difícil, pois a troca de mensagens é o fator de maior consumo de energia em uma RSSF e ocupação de canal de comunicação. A inundação de mensagens de consultas para coleta das leituras de sensoriamento da rede é um dos maiores fatores de consumo de energia dos dispositivos, e conseqüentemente de diminuição da vida útil da rede e disponibilidade dos dados. Visando possibilitar a diminuição da necessidade de inundar a rede para coletar os dados são utilizadas estruturas nomeadas agrupamentos. Os agrupamentos de sensores são sub-conjuntos de sensores associados segundo critérios de proximidade [Bo et al. 2008] e valores de atributos. O critério de proximidade con-

siste em uma distância física fixa, medida em saltos, para delimitar o tamanho máximo do agrupamento. Os atributos de um sensor descrevem suas propriedades e podem ser categorizados como sendo atributos inerentes ao dispositivo e atributos de dados sensorizados. Os atributos inerentes a sensores são a carga do dispositivo, localização geográfica, altitude e mobilidade. Alguns exemplos de atributos de dados sensorizados são valores de temperatura, pressão, umidade e luminosidade.

Em alguns ambientes de sensoriamento, como o urbano, os dados lidos apresentam similaridade espacial. Nestes casos, a formação de agrupamentos baseados nesta característica representa uma maneira mais econômica e mais eficiente de coletar os dados da rede. A definição de agrupamentos baseada em similaridade espacial dos dados, considera explicitamente os atributos dos dispositivos (nós) além de sua localização geográfica. Estes agrupamentos resultam de particionamentos com três propriedades [Frery et al. 2008]:

1. **Homogeneidade interna do agrupamento:** Dispositivos pertencentes ao mesmo agrupamento tem que ser similares aos demais segundo algum atributo pré-definido.
2. **Heterogeneidade entre agrupamentos:** Dispositivos pertencentes a agrupamentos diferentes devem ser diferentes uns dos outros segundo algum atributo pré-definido.
3. **Proximidade ou contiguidade:** Dispositivos pertencentes a um mesmo agrupamento devem pertencer a uma estrutura de vizinhança pré-definida.

A coleta dos dados em agrupamentos formados a partir da similaridade espacial oferece uma maior escalabilidade e disponibilidade dos dados. Porém, a escolha de apenas um valor para representar as leituras de todos os dispositivos do agrupamento pode comprometer a precisão ou a validade dos dados. Pois, é adota uma leitura a ser utilizada como base e limiar de erro que determina se duas leituras são suficientemente similares. Sempre que uma mensagem é entregue na estação base com uma variação em relação ao valor real em campo é dito que esta leitura sofreu uma distorção.

As distorções decorrem da disponibilidade finita de dispositivos para sensoriamento do campo. As distorções sofridas pelos dados podem ser classificadas como espaciais e temporais. O uso da leitura do sensor mais próximo de determinada coordenada ou o uso do valor médio das leituras coletadas em um agrupamento como resposta a uma consulta, em geral, resulta em um erro de precisão dos dados. Este tipo de erro é classificado como uma distorção espacial dos dados. A distorção temporal dos dados consiste na variação sofrida pelos dados em campo após sua coleta até sua entrega na estação base. O número de saltos exigido para a disseminação e coleta de dados para uma consulta determina a probabilidade de que os dados estejam defasados [Zhang et al. 2008] .

O trabalho [Zhang et al. 2007] avalia a distorção espaço-temporal dos dados em redes de sensores dos dados planas. A precisão dos sumários de dados coletados de *cluster-heads* é analisado em [Frery et al. 2008], porém é desconsiderado o tempo de resposta.

## 2. Objetivo

O objetivo principal deste estudo é analisar a distorção espaço-temporal dos dados em agrupamentos formados a partir da similaridade espacial dos dados. E mostrar que o uso

de agrupamentos favorece uma maior disponibilidade dos dados e o uso da similaridade espacial dos dados em sua formação atenua a distorção dos dados.

### 3. Implementação e Resultados

A avaliação será feita através do uso de métricas de latência média e precisão dos dados[Zhang et al. 2007]. A métrica de latência média representa o tempo médio de resposta de consultas. Nos ambientes em que a coleta é feita em tempo real é vital que o tempo de resposta de qualquer consulta tenda a zero, pois os dados podem não serem mais representativos. A latência média é calculada através da soma da latência de todas as consultas disseminadas e então cálculo de sua média aritmética. A precisão dos dados consiste na diferença no valor das leituras entregues à estação base e às leituras presentes no campo. O cálculo da precisão média dos dados é calculado através da soma da diferença do valor das leituras entregues e aos dados do campo.

Serão avaliadas as distorções sofridas pelos dados em redes planas, e hierárquicas com um ou mais *cluter-heads*. Estas comparações servem para determinar quais destas redes são mais eficientes relacionando o tempo de resposta e a precisão dos dados em cada uma.

Densidade	Dimensões (m)	Nº de Sensores	$Sensor/m^2$
Esparsa	400x400	1000	0,00625
Densa	200x200	1000	0,025

**Tabela 1. Parâmetros de simulação.**

Para a avaliação dos cálculos teóricos e comparação com demais trabalhos relacionados serão feitas simulação na ferramenta NS. O ambiente alvo possui dimensões de 200x200 metros para a simulação de redes densas e 400x400 metros para redes esparsas. Em cada rede serão dispersos 1000 dispositivos sensores. A Tabela 1 ilustra estes parâmetros de simulação .

### 4. Bibliografia

#### Referências

- Akyildiz, I. F., Weilian, S., Sankarasubramaniam, Y., and Cayirci, E. E. (2002). A survey on sensor networks. *IEEE Communications Magazine*, 40(8):102–114.
- Bo, W., Han-ying, H., and Wen, F. (2008). An improved leach protocol for data gathering and aggregation in wireless sensor networks. In *ICCEE*, pages 398–401, Washington, DC, USA. IEEE Computer Society.
- Frery, A. C., Ramos, H., Alencar-Neto, J., and Nakamura, E. (2008). Error estimation in wireless sensor networks. In *SAC*, pages 1923–1928, New York, NY, USA. ACM.
- Zhang, X., Wang, H., and Khokhar, A. (2008). Distortion analysis for real-time reconstruction of correlated data field in heterogeneous sensor networks. In *GLOBECOM*, pages 1–5.
- Zhang, X., Wang, H., and Khokhar, A. A. (2007). Distortion analysis for real-time data gathering of spatially-temporally correlated fields in sensor networks. In *GLOBECOM*, pages 828–832.