



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

INPE-14105-NTC/326

**REDES DE GEO-SENSORES: ROTEAMENTO, ARMAZENAMENTO
E ANÁLISE DE DADOS**

Ilka Afonso Reis

Publicação Interna - Sua reprodução ao público externo está sujeita à autorização da chefia

INPE
São José dos Campos
2005



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

INPE-14105-NTC/326

**REDES DE GEO-SENSORES: ROTEAMENTO, ARMAZENAMENTO
E ANÁLISE DE DADOS**

Ilka Afonso Reis

Publicação Interna - Sua reprodução ao público externo está sujeita à autorização da chefia

INPE
São José dos Campos
2005

“Em 2010, sensores estarão em todo lugar. (...) eles poderão sobreviver à raça humana. Descendentes dos golfinhos poderão descobri-los no gelo ártico e se maravilhar com esta tecnologia extinta.”

KRIS PISTER, Projeto *Smart Dust*.

RESUMO

Redes de geo-sensores são formadas por minúsculos dispositivos eletro-mecânicos com capacidade de sensoriamento, processamento e comunicação, que são distribuídos numa área geográfica com o objetivo de coletar dados sobre um fenômeno de interesse. Estruturas assim podem ser utilizadas para monitorar a poluição do ar, estudar o micro-clima do *habitat* de um animal, detectar vibrações anormais nas vigas de um edifício e realizar um grande número de tarefas de estudo e monitoramento. O objetivo deste trabalho é caracterizar as redes de geo-sensores, incluindo algumas alternativas de transporte, armazenamento e análise de dados, assim como a descrição de suas aplicações mais importantes.

Palavras-chave: roteamento, agregação de dados, bancos de dados de sensores, estatística espacial.

ABSTRACT

Geosensor networks are dense wireless networks of small, low-cost sensors, which collect and disseminate environmental data of a vast geographical area. These networks hold the promise of revolutionizing sensing in a wide range of application. This includes environmental monitoring and surveillance applications. The aim of this work is to examine this emerging subject, especially some alternatives for data routing, databases and data analysis in geosensors networks context.

SUMÁRIO

Pág.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE SÍMBOLOS

LISTA DE SIGLAS E ABREVIATURAS

CAPÍTULO 1 - INTRODUÇÃO	17
CAPÍTULO 2 - O QUE SÃO REDES DE SENSORES ?	21
2.1 Aplicações e Exemplos.....	24
2.2 Características das Redes de Sensores	27
2.3 Classificação das Redes de Sensores	28
2.4 Redes de Geo-sensores	29
2.5 Algumas considerações	29
CAPÍTULO 3 - ROTEAMENTO E AGREGAÇÃO DE DADOS.....	31
3.1 Propostas para roteamento de dados.....	33
3.2 Algumas considerações	36
CAPÍTULO 4 - BANCOS DE DADOS PARA REDES DE SENSORES.....	37
4.1 Armazenamento e consultas	37
4.2 Propostas	40
4.3 Algumas considerações	44
CAPÍTULO 5 - ANÁLISE ESTATÍSTICA EM REDES DE SENSORES.....	47
5.1 Análise Instantânea.....	47
5.2 Análise Retrospectiva	48
5.2.1 Análise na presença de agregação.	51
5.3 Algumas considerações	55
CAPÍTULO 6 - CONSIDERAÇÕES FINAIS	57
REFERÊNCIAS BIBLIOGRÁFICAS	59

LISTA DE FIGURAS

2.1 – Representação esquemática de uma rede de sensores	21
2.2 – Alguns exemplos de sensores	22
3.1 – Exemplos de roteamento	32
5.1 – Rede de sensores com agregação de dados intra-rede.....	52

LISTA DE TABELAS

- 4.1 – Classificação de algumas alternativas de armazenamento de dados dos sensores segundo o local (centralizado ou distribuído) e período (histórico ou não). 40

LISTA DE SIGLAS E ABREVIATURAS

ALERT	- Automated Local Evaluation in Real-Time
CAR	- Condicional Auto-Regressivo
EUA	- Estados Unidos da América
GAF	- Geographic Adaptive Fidelity
GEAR	- Geographic Energy Aware Routing
GIS	- Geographic Information System
GPS	- Global Positioning System
GPSR	- Greedy Perimeter Stateless Routing
GWR	- Geographic Weighed Regression
LEACH	- Low Energy Adaptative Clustering Hierarchy
MANET	- Mobile Ad hoc Network
MCMC	- Markov Chain Monte Carlo
NASA	- National Aeronautics and Space Administration
SIG	- Sistema de Informação Geográfica
SPIN	- Sensor Protocols for Information via Negotiation
SQL	- Structure Query Language
STWin	- Spatio-Temporal Window)

CAPÍTULO 1

INTRODUÇÃO

A compreensão do funcionamento do mundo físico é preocupação constante da espécie humana. Há séculos, o homem observa o ambiente ao seu redor, faz indagações, conclui e observa de novo. O estudo do ambiente, seja para fins de preservação, reconhecimento ou vigilância, é tarefa importante para a sobrevivência de muitos seres vivos. Como reconhecer a iminência de um terremoto senão conhecendo seus sinais? E como saber que sinais são estes sem coletar dados sobre os terremotos?

Esta necessidade de conhecimento sobre o mundo físico tem impulsionado pesquisadores a instalarem estruturas de observação capazes de coletar dados sobre diversos ambientes, abrangendo grandes áreas geográficas. No leste da Noruega, mecanismos para detecção de movimentos de geleiras foram enterrados em buracos no gelo. Medidores da qualidade do ar estão presentes em 23 pontos da cidade de São Paulo. Mini-estações meteorológicas foram instaladas no topo de grandes árvores para estudar o micro-clima da floresta de uma região da Califórnia, EUA. Além destes exemplos, há ainda muitos outros.

Apesar de ter se tornado uma atividade freqüente, monitorar fenômenos em grandes áreas geográficas ainda é uma tarefa dispendiosa e difícil. Estruturas de observação caras, que exigem manutenção constante, muitas vezes têm restrições quanto aos locais de instalação, inviabilizando a cobertura adequada da área de estudo. Além disto, a resolução temporal dos dados coletados nem sempre é suficiente para a compreensão do fenômeno. No entanto, uma nova tecnologia surge para sanar estas dificuldades e auxiliar a observação do mundo físico : as redes de geo-sensores.

Sensores são minúsculas estruturas eletro-mecânicas capazes de medir variáveis como temperatura, pressão barométrica, partículas em suspensão, salinidade e várias outras. Estes mecanismos comunicam-se entre si e, distribuídos numa área geográfica, formam uma rede de geo-sensores. O prefixo *geo* serve para enfatizar o aspecto geográfico de uma rede de sensores.

Uma característica importante da comunicação entre os sensores é que ela acontece sem a necessidade de fios (*wireless*). Esta característica, aliada ao tamanho diminuto de seus componentes, tornam as redes de sensores um valioso instrumento para a coleta de dados de forma não-invasiva e não-perturbadora, o que é especialmente importante no estudo de ecossistemas e da vida animal sensíveis à presença humana.

Os sensores numa rede funcionam segundo o princípio da colméia: sozinhos são quase nada; juntos constroem uma visão rica do ambiente sensoriado. Redes de sensores são especialmente úteis para aplicações de monitoramento: pequenos vigilantes, sempre atentos, em todo lugar, o tempo todo.

As pesquisas em redes de sensores foram motivadas por aplicações militares, como o desenvolvimento de sistemas de vigilância acústica para monitoramento oceânico e redes de detecção de alvos terrestres. Entretanto, os recentes avanços tecnológicos, diminuindo o custo de produção e aumentando a capacidade dos mecanismos, ampliaram as aplicações para as redes de sensores, que vão desde a detecção de contaminação biológica até a coleta de dados meteorológicos, passando pela organização de tráfego e as casas inteligentes.

Este trabalho tem o objetivo de descrever as características e aplicações mais importantes das redes de geo-sensores, incluindo algumas alternativas de transporte, armazenamento e análise de dados. O texto está organizado como se segue. O Capítulo 2 descreve os componentes de uma rede de sensores, apresenta alguns exemplos de redes já instaladas, enumerando outras possíveis aplicações e destaca as características

e classificações mais importantes das redes de sensores. O Capítulo 3 aborda o transporte dos dados coletados, questão que tem bastante destaque na literatura. O armazenamento dos dados de sensores é tratado no Capítulo 4. No Capítulo 5, são apresentadas algumas alternativas para a análise dos dados vindos de uma rede de sensores. Finalmente, o Capítulo 6 traz algumas considerações sobre o assunto.

CAPÍTULO 2

O QUE SÃO REDES DE SENSORES ?

Redes de Sensores consistem de nodos (sensores) que podem medir características do seu ambiente local, fazer cálculos e comunicar-se uns com os outros numa rede sem fio (Paskin et al., 2005). Em geral, estes sensores são distribuídos dentro do fenômeno a ser monitorado ou muito perto dele. A Figura 2.1 mostra a representação esquemática de uma Rede de Sensores. Os nodos sensores coletam dados sobre o ambiente sensoriado, processam estes dados e os transmitem a outros sensores ou a um nodo especial chamado sorvedouro (*sink*). Os dados são então enviados ao *gateway*, que se comunica com a estação-base através da Internet ou de um satélite. O usuário tem acesso à estação-base, que é um computador comum.

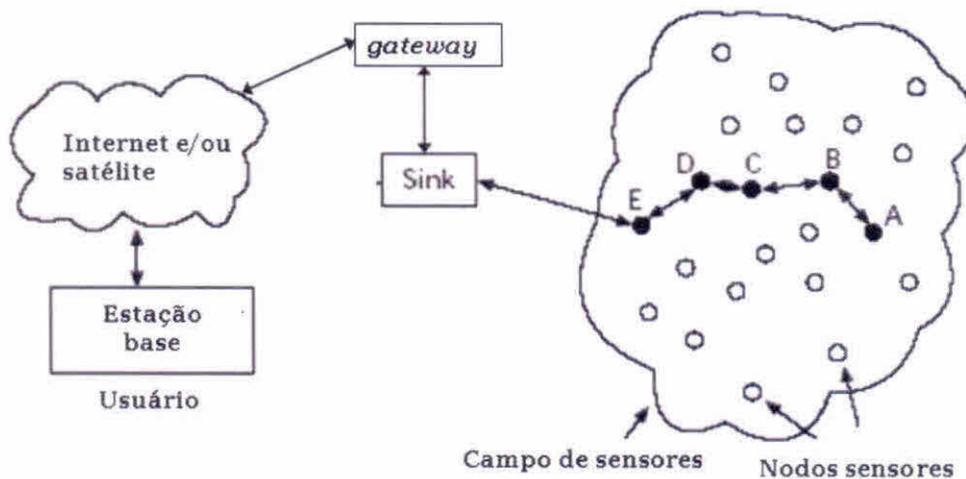


FIGURA 2.1 - Representação esquemática de uma rede de sensores (adaptada de Akyildiz et al., 2002)

Um **nodo sensor** é composto por um ou mais sensores, um processador, memória, bateria e um transceptor (componente que transmite e recebe dados). A comunicação entre os nodos sensores pode acontecer através de rádio-frequência ou infravermelho

(Ruiz et al, 2004 e Akyildiz et al., 2002). Na literatura, é comum encontrar o termo *mote* (abreviatura de *remote sensor*) para se referir ao componente sensor de uma rede de sensores. A Figura 2.2 apresenta alguns tipos de sensores. A tendência atual é a de que estes mecanismos se tornem cada vez menores. Um exemplo desta tendência são os sensores desenvolvidos no projeto *Smart Dust* da Universidade da Califórnia (Berkeley) (figuras 2.2a e 2.2b). Este projeto tem por objetivo construir mecanismos sensores tão pequenos quanto grãos de areia e tão baratos que poderão ser distribuídos às centenas.

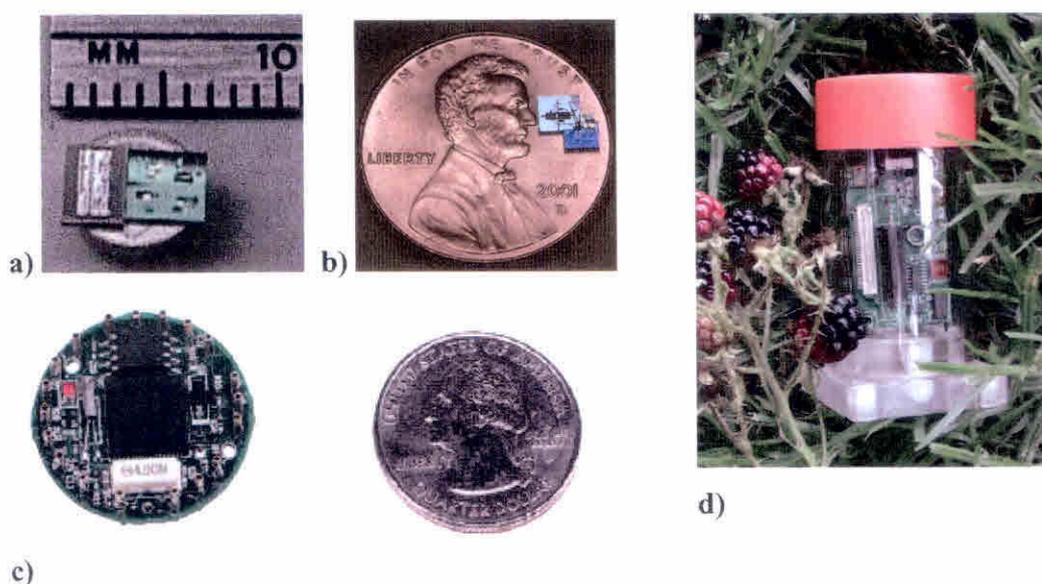


FIGURA 2.2 – Alguns exemplos de sensores: **a)** protótipo inicial do projeto *Smart Dust* (Universidade da Califórnia – Berkeley) ; **b)** *Golem Dust*, uma das gerações de sensores do projeto *Smart Dust*; **c)** MICA2DOT, sensor desenvolvido pela Universidade da Califórnia – Berkeley em parceria com a Intel© ; **d)** sensor MICA e seu invólucro de acrílico, utilizados no projeto *Great Duck Island*, Maine, USA (Universidade da Califórnia – Berkeley)

Atualmente, a instalação dos nodos sensores na região de estudo é feita manualmente. Entretanto, as previsões para o futuro mostram que eles poderão ser jogados de um avião, por exemplo, favorecendo a instalação de redes de sensores em regiões inóspitas (florestas densas, desertos, montanhas, geleiras) ou perigosas (campos de batalha, vulcões). Uma vez instalados na região de estudo, os nodos sensores devem detectar seus vizinhos, estabelecendo assim a topologia da rede.

O **sorvedouro** é um nodo especial da rede utilizado para agregar dados de outros nodos sensores. O sorvedouro desempenha papel importante no modo como os dados trafegam na rede (roteamento). Pode ajudar também na resposta a consultas feitas à rede, tornando-se responsável pelos dados dos nodos sob sua jurisdição (Loureiro et al, 2003).

Para se comunicarem com o exterior da rede, os nodos sensores utilizam um nodo especializado chamado **gateway**, que possibilita esta troca de informações através do seu contato com a **estação-base**. Esta, por sua vez, é um computador comum capaz de se comunicar com o **gateway**, sendo também o nodo da rede que mostra os resultados aos usuários. Em algumas propostas de rede de sensores, o banco de dados encontra-se na estação-base e nele são armazenados os dados de sensoriamento e as informações sobre os nodos.

Segundo Loureiro et al. (2003), dentre as tarefas típicas de uma rede de sensores, pode-se citar:

- Coletar dados sobre o fenômeno estudado e enviá-los para armazenamento na estação-base ;
- Responder a consultas, retornando os valores de uma variável em determinados locais e intervalos de tempo. Ex: quais os valores da temperatura nos nodos próximos aos ninhos das aves (nodos 1, 4, e 7) entre 18:00 e 23:00 horas do dia 5 de dezembro de 2000?
- Detectar a ocorrência de eventos de interesse e estimar valores de variáveis associadas ao evento detectado. Ex: há algum veículo no cruzamento das ruas A e B neste momento? Se sim, qual o valor de sua velocidade?
- Classificar um objeto detectado. Ex: o veículo no cruzamento das ruas A e B é um carro, um veículo de carga ou uma moto?
- Rastrear um objeto. Ex: quais as localizações do animal X foram registradas entre meia-noite e 6:00 horas do dia 18 de março de 2002?

2.1 – Aplicações e exemplos

O desenvolvimento tecnológico dos mecanismos sensores tem sido rápido, assim como o avanço das pesquisas sobre o funcionamento das redes. No entanto, até o presente momento, redes de sensores foram instaladas somente em caráter experimental (Ratnasamy et al., 2002), como é o caso dos projetos apresentados a seguir.

Exemplo 1: Projeto *Great Duck Island* (Mainwaring et. al, 2002). Em agosto de 2002, pesquisadores da Universidade da Califórnia – Berkeley, em conjunto com a *Intel Research Laboratories*, instalaram uma rede de sensores na *Great Duck Island*, no estado do Maine (EUA), para monitorar o comportamento e o *habitat* de um pássaro chamado *Storm Petrel*. Inicialmente, 32 nodos sensores MICA (Figura 2d) foram colocados nas tocas dos pássaros e também próximos a elas. Posteriormente, a rede foi aumentada, ganhando mais nodos e estações meteorológicas. Os nodos sensores podiam medir a temperatura, a umidade do ar e a pressão barométrica, entre outras variáveis. As leituras periódicas dos nodos eram transmitidas a um *gateway*, que as retransmitia até a estação-base, onde eram armazenadas. Uma réplica do banco de dados era transmitida, via satélite, para um servidor na Universidade de Berkeley a cada 15 minutos. Os usuários tinham acesso ao banco de dados via Internet (<http://www.greatduckisland.net/>).

Exemplo 2: Projeto ZebraNet (Juang et al., 2002). Colares sensores equipados com GPS (*Global Position System*) foram instalados em zebras na reserva de *Sweetwaters*, Quênia. Do ponto de vista biológico, o objetivo do projeto é monitorar o comportamento noturno dos animais e ainda responder questões acerca de migração e relacionamento inter-espécies. Os sensores foram projetados para registrarem e armazenarem a posição dos animais a cada 3 minutos e, a cada hora, dados meteorológicos, ambientais, de iluminação, de temperatura e de movimentos corporais são registrados por 3 minutos. A estação-base também é móvel e frequentemente

percorre o campo dos sensores (*habitat* das zebras) para recolher os dados. (<http://www.princeton.edu/%7Emrm/zebranet.html>)

Exemplo 3: *Automated Local Evaluation in Real-Time (ALERT)*. Instalada na década de 70 pelo *National Weather Service* (EUA), ALERT é uma rede de sensores que fornece dados sobre chuva e nível de água em tempo real, o que permite avaliações sobre o potencial de enchentes. Seus sensores podem coletar dados sobre ventos, temperatura, nível de água, entre outros. Os dados são transmitidos para uma estação-base, onde um modelo de previsão de enchentes processa estes dados e gera os alertas. Atualmente, ALERT possui sensores instalados na maior parte do oeste americano e é intensamente utilizado pelos Estados da Califórnia e Arizona. (<http://www.alertsystems.org>). Outro projeto para previsão de enchentes é o FloodNet (Envisense), no qual sensores foram instalados às margens de um rio na Inglaterra (<http://envisense.org/floodnet/floodnet.htm>).

Com a prevista diminuição do custo dos mecanismos sensores e o avanço dos estudos na área, pesquisadores vislumbram os mais diversos tipos de aplicação para as redes de sensores. Estas aplicações podem ser divididas em dois grupos segundo seu objetivo geral : monitoramento ou estudo retrospectivo. Nas **aplicações de monitoramento**, o interesse maior está em analisar as observações do fenômeno assim que elas estiverem disponíveis, como na detecção de incêndios, enchentes e outros eventos. O armazenamento das séries históricas dos dados pode ou não acontecer. Já nas aplicações cujo objetivo geral é o **estudo retrospectivo**, o interesse maior está na análise de séries históricas e o armazenamento dos dados é feito para uma análise posterior. Convém observar que uma rede de sensores pode ser projetada para cumprir os dois objetivos.

A seguir, são enumerados alguns exemplos de aplicações de rede de sensores, classificados segundo seu objetivo específico:

- Monitoramento e estudo do *habitat* e da vida selvagem de modo não-invasivo e não- perturbador. Exemplos: projetos *Great Duck Island*, *ZebraNet* e *PODS* (Biagioni e Sasaki, 2003) .
- Monitoramento ambiental : condições de temperatura e umidade no armazenamento de grãos (Armstrong, 2003) ; detecção de incêndios e desmatamento em florestas ; poluição sonora e do ar em grandes cidades ; tráfego de veículos ; movimento de geleiras (Envisense GlacsWeb) ; detecção de vazamentos de óleo e gás em rios (AmazonBots, Petrobrás).
- Monitoramento de desastres: sistemas de alarme de enchentes (Envisense FloodNet, ALERT) ; terremotos ; tornados.
- Monitoramento estrutural de construções: detecção de vibrações anormais em pontes (Basharat et al., 2005); detecção de micro fissuras nas colunas de edifícios.
- Agricultura de precisão: monitoramento das condições climáticas em vinhedos (INTEL Laboratories) ; monitoramento da umidade do solo para fins de irrigação .
- Segurança: detecção de intrusos em edificações ; detecção de incêndios em fábricas .
- Militares: detecção de alvos no campo de batalha, vigilância de oceanos.

As redes de sensores são uma tecnologia altamente orientada à aplicação. Suas características e os objetivos têm grande influência no projeto da rede: tamanho, tipo e quantidade de sensores, forma de instalação, mobilidade, envio e armazenamento de dados, entre outros. Esta dependência da aplicação será detalhada no decorrer do texto.

2.2 – Características das redes de sensores

Uma Rede de Sensores pode ser considerada um tipo especial de rede móvel *ad hoc* (MANET – *Mobile Ad Hoc Network*) (Nittel et al. 2004). MANET são redes onde os elementos computacionais (móveis ou não) trocam informação entre si com a função básica de dar suporte à comunicação entre esses elementos que, individualmente, podem executar tarefas distintas. Um exemplo de MANET é o compartilhamento de informações por um grupo de soldados em seus computadores portáteis, numa determinada distância, através de sinais de rádio-frequência (Pereira et al., 2003). Cada nodo da rede (soldado) pode realizar uma tarefa diferente e trocar informações com seus colegas. No entanto, diferentemente das MANET, redes de sensores tendem a executar uma função de modo a haver colaboração entre os nodos, que coletam dados sobre os mesmos fenômenos. Por serem um caso especial das redes *ad hoc*, as redes de sensores são naturalmente não-estruturadas. Seus nodos são capazes de se comunicar uns com os outros diretamente, não existindo estações de suporte à comunicação, como no caso das redes de telefones celulares, que são consideradas redes estruturadas.

Outras características importantes das redes de sensores são sua capacidade de auto-organização, a suscetibilidade a falhas por parte dos sensores, a freqüente mudança na topologia e a limitação dos sensores em relação à energia, capacidade computacional e memória. As três primeiras características estão fortemente relacionadas entre si, pois a falha de um ou mais sensores obriga a rede a mudar a topologia dos nodos, o que somente ocorre devido à capacidade de auto-organização das redes. Esta capacidade também é essencial na fase de estabelecimento da rede. Segundo Ruiz et al. (2004), a limitação da energia é considerada o aspecto mais crítico das redes de sensores. Sendo assim, há um grande esforço por parte dos pesquisadores da área no sentido de desenvolver procedimentos de transporte e processamento de dados que levem em conta a economia de energia (Akyildiz et al., 2002)

A grande quantidade de nodos, outra característica das redes de sensores, possibilita o monitoramento de áreas extensas. Além disto, regiões densamente povoadas por sensores são um tipo de garantia contra a falha destes mecanismos, seja por defeito ou por falta de energia, principalmente em regiões de difícil acesso. Se um sensor falhar, haverá outro para fazer seu trabalho.

2.3 – Classificação das redes de sensores

As redes de sensores podem ser classificadas com relação à mobilidade dos nodos em estacionárias e não-estacionárias. Numa **rede estacionária** (ou estática), os sensores não se movem em relação ao objeto ou fenômeno sensoriado, como no caso dos sensores do projeto *Great Island Duck*. Já numa **rede não-estacionária** (ou dinâmica), existe mobilidade por parte dos sensores. Este é o caso das zebras do ZEBRANET, e de bóias lançadas ao mar, por exemplo.

O modelo de envio de dados da rede pode ser classificado em : **contínuo** (os dados são enviados continuamente a uma taxa pré-estabelecida) ; **por demanda** (o envio de dados somente acontece quando o usuário os requer numa consulta) ; **orientado a eventos** (os dados são enviados somente quando um evento de interesse ocorre) e **híbrido**, no qual dois ou mais tipos anteriores podem ocorrer (Tilak et al., 2002). Redes criadas para o monitoramento ambiental geralmente utilizam o modelo de envio de dados contínuo, enquanto as redes que têm por objetivo detectar eventos utilizam o modelo orientado a eventos. A escolha de modelo de envio de dados, além de ser dependente da aplicação para a qual foi criada a rede, tem grande impacto no gasto de energia, devido ao número de transmissões de dados requerido.

A classificação da rede tem fundamental importância na escolha do modo como os dados dos sensores serão encaminhados até o usuário, como será detalhado no próximo capítulo.

2.4 – Redes de geo-sensores

Uma rede de geo-sensores pode ser definida com uma rede de sensores que monitora fenômenos no espaço geográfico, na qual os nodos sensores capturam informação geográfica que é relevante para a aplicação em questão (Nittel et al., 2004). Os exemplos de 1 a 3 da seção 2.1 podem caracterizar redes de geo-sensores.

Alguns autores já consideram o aspecto espacial como inerente às redes de sensores (Heidemann e Bulusu, 2001 ; Al-Karaki e Kamal, 2004). Ratnasamy et al. (2002) enfatizam que o conhecimento sobre localização dos nodos é crucial para o trabalho com redes de sensores, sendo o objetivo da rede fornecer o sensoriamento de uma vasta área geográfica. A localização dos nodos pode ser conseguida através de um GPS ou, quando isto não for possível para todos os nodos, pode-se fazer uso de algoritmos como os descritos por Bharathidasan et. al (2003) e propostos por Bulusu et al. (2000).

Como esta denominação é relativamente recente, a maioria dos trabalhos publicados sobre redes de sensores não a utilizam. Deste modo, no restante do texto, os termos sensores e redes de sensores serão mantidos, tendo em mente que o aspecto espacial está subentendido nos problemas abordados¹.

2.5 – Algumas considerações

Além dos tópicos abordados neste capítulo, há ainda muitos tópicos importantes no contexto de redes de sensores. Entre eles, podemos citar os procedimentos para distribuição dos sensores na área de estudo, estabelecimento da rede, *hardware*, meios de comunicação, processamento e sincronização de sinais, segurança, cobertura da área de estudo e diminuição de custo. Em geral, estes são assuntos relacionados ao desenvolvimento tecnológico dos mecanismos sensores e ao funcionamento da rede

como um todo. Todos estes tópicos estão, de algum modo, ligados à obtenção de dados, que é o objetivo de uma rede de sensores. No entanto, este texto enfocará os aspectos mais diretamente relacionados à obtenção dos dados, como o seu transporte, agregação e armazenamento, que serão abordados nos próximos capítulos.

Boas referências sobre os vários aspectos envolvidos no desenvolvimento das redes de sensores podem ser encontradas em Al-Karaki e Kamal (2004) e Akyildiz et al. (2002).

¹ Em 2003, aconteceu o primeiro workshop sobre redes de geo-sensores, o Geo Sensor Networks (GSN2003). Nem mesmo todos os palestrantes deste encontro utilizaram o termo geo-sensores ao se referirem aos nodos da rede.

CAPÍTULO 3

ROTEAMENTO E AGREGAÇÃO DE DADOS

Uma vez coletados pelos sensores, os dados devem chegar até o usuário através da comunicação entre os nodos. O transporte destes dados pela rede (**roteamento**) pode ser uma tarefa contínua ou pode acontecer devido a uma requisição do usuário (consulta).

Um procedimento simples de transporte de dados é a **árvore de roteamento** (Figura 3.1a). O processo de construção da árvore começa a partir de um nodo raiz (o sorvedouro ou a estação-base, por exemplo) . Este nodo dissemina uma mensagem de construção a seus vizinhos. Assim que um nodo recebe esta mensagem, ele identifica o nodo que a enviou como seu *pai*. Uma vez escolhido o pai, mensagens vindas de outros nodos são descartadas. Este nodo repete o mesmo processo na procura de seus *filhos*, até que todos os nodos recebam a mensagem. Na transmissão dos dados dos sensores até o nodo raiz, filhos passam seus dados para os seus pais, que os repassam para os seus pais e, assim por diante, até chegar ao nodo raiz.

Na tarefa de comunicação, os nodos gastam mais energia do que na tarefa de processamento dos dados (Pottie e Kaiser, 2000). Assim, em alguns modos de roteamento, os dados são agregados no caminho até o usuário. Isto gera uma economia de energia, como ilustram as figuras 3b e 3c.

No roteamento do tipo tradicional (Figura 3.1b), os dados de cada nodo (A, B e C) são transmitidos ao nodo sorvedouro (S) através da rede utilizando-se um dos possíveis caminhos. Cada nodo gasta três transmissões para levar seu dado ao sorvedouro, num total de nove transmissões. Na alternativa da Figura 3.1c, o dado do nodo A é agregado ao dado do nodo B na primeira transmissão. Posteriormente, este dado agregado é transmitido ao próximo nodo e novamente agregado ao dado vindo do nodo C. A

agregação dos dados dos três nodos é então transmitida ao nodo sorvedouro. No total, foram feitas seis transmissões, 33% a menos do que no roteamento tradicional, o que certamente representou uma economia de energia.

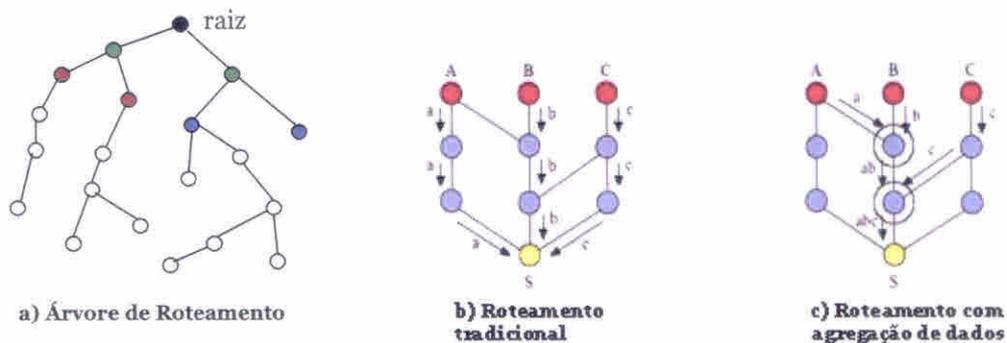


FIGURA 3.1 – Exemplos de roteamento : a) o nodo raiz (preto) é pai dos nodos verdes, que são pais dos nodos azuis e vermelhos ; b) e c) os nodos A, B e C geram os dados e o nodo S é o sorvedouro. As linhas ligando os nodos significam que eles podem se comunicar.

Para Kalpakis et al. (2003), a **agregação de dados** consiste em combinar dados oriundos de diferentes nodos sensores, eliminando transmissões redundantes e permitindo uma visão multidimensional do ambiente monitorado. A agregação é um procedimento bastante utilizado na **consulta** aos dados. Os principais tipos de operação de agregação são *média*, *máximo*, *mínimo* e *supressão* de dados redundantes. Os dados de diversos sensores são agregados em um só nodo, que então transmite um resumo destes dados pela rede. Isto gera uma **economia de energia**, como ilustrado nas Figuras 3.1b e 3.1c.

A agregação de dados também pode ser usada com a finalidade de melhorar o processo de **detecção de eventos**. Neste caso, as respostas de mais de um sensor são utilizadas para definir a ocorrência ou não de um evento. Quando a agregação de dados é utilizada com o objetivo de obter uma estimativa de maior precisão, o termo utilizado é **fusão de dados** (Steinberg et al., 1999). Esta técnica também pode ser utilizada para **diminuir o ruído** existente nos dados provenientes dos sensores. Os termos agregação e fusão

aparecem muitas vezes como sinônimos na literatura de redes de sensores. No entanto, para trazer um pouco mais de clareza à discussão, o termo agregação será utilizado neste texto para se referir ao simples processo de sintetizar dados com o objetivo de reduzir o volume de transmissões pela rede.

3.1 – Propostas para roteamento de dados

A maior preocupação de uma proposta de roteamento é transportar os dados economizando energia (Potie e Kaiser, 2000). Agregar os dados ao fazer seu transporte é uma das formas de se conseguir esta economia.

Entre as propostas de roteamento que utilizam agregação dos dados, estão o SPIN (Heinzelman et. al, 1999), o LEACH (Heinzelman et. al, 2000), o *Directed Diffusion* (Intanagonwiwat et al., 2002) e suas variações. O processo de disseminação dos dados pela rede é diferente em cada um destes tipos de roteamento. No SPIN, por exemplo, o nodo sensor dissemina para seus vizinhos uma mensagem comunicando que tem dados disponíveis e espera pelo interesse de algum outro nodo em levar estes dados ao *gateway*. O nodo interessado requisita o dado e lança também uma mensagem entre seus vizinhos avisando que tem dados disponíveis. O processo continua até que o dado chegue ao *gateway*. Já no *Directed Diffusion*, o nodo sorvedouro é quem dissemina uma mensagem com seu interesse por determinado tipo de dado e aguarda a resposta dos nodos sensores (**roteamento centrado em dados**). No entanto, é o aspecto da **agregação dos dados** o que mais interessa a este texto, pois o modo como é feita esta agregação determinará como serão armazenados e analisados os dados provenientes da rede.

No procedimento *Directed Diffusion*, o interesse em agregação é determinado localmente no sorvedouro. Já no SPIN, pode ocorrer a agregação entre o dado do nodo requisitante e o dado que foi recebido por ele e, então, uma mensagem sobre a disponibilidade destes dados agregados é disseminada pela rede. Na proposta de

roteamento do LEACH, acontece a formação de conglomerados de nodos antes de cada transmissão de dados (**roteamento hierárquico**). Nodos são escolhidos *aleatoriamente* para serem os líderes destes conglomerados (*cluster-head*), que fazem a agregação dos dados e transmitem o resumo para o *gateway*.

Algumas propostas de roteamento utilizam a localização dos nodos para transportar os dados (**roteamento geográfico**). Elas fazem uso da constatação de que geralmente as consultas aos dados incluem atributos geográficos. Entre estas propostas, estão o *Geographic Adaptive Fidelity* (GAF – Xu et al., 2001), o *Geographic and Energy Aware Routing* (GEAR - Yu et al, 2001) e o pioneiro GPSR (Karp e Kung, 2000). Segundo a classificação feita por Al-Karaki e Kamal (2004), as propostas de roteamento geográfico não utilizam agregação dos dados. A economia de energia é atingida pelo direcionamento das consultas a uma região específica da rede, diminuindo o número de mensagens a serem transmitidas.

Existem outras propostas de roteamento que não adotam agregação de dados como estratégia para economizar energia. No contexto de redes de sensores com modelo de envio de dados orientado a eventos, Vuran et al. (2004) propuseram explorar a **correlação espacial e temporal** entre os dados dos nodos que detectam um mesmo evento. A correlação espacial entre os nodos é modelada de modo semelhante ao que se utiliza na geoestatística baseada em modelos (Diggle et. al., 1998). Somente alguns nodos são escolhidos para transmitirem seus dados, o que diminui o número de mensagens transmitidas até o sorvedouro.

Outro procedimento de roteamento que adota a idéia de escolher alguns nodos para transmitir seus dados é o **Infer** (Hartl e Li, 2005). Nesta proposta, a escolha dos nodos é aleatória e pode considerar aspectos como nível de energia do nodo ou a importância do dado a ser transmitido para o sorvedouro. Os nodos não escolhidos para transmitir ficam “dormindo”. Os dados destes nodos são estimados pelo sorvedouro utilizando os dados transmitidos por outros nodos. A estimação é feita através de **inferência Bayesiana**,

com as chamadas distribuições preditivas (Carlin e Louis, 1996). A idéia da Estatística Bayesiana é fazer inferência sobre os parâmetros de um modelo através de uma combinação da função de verossimilhança obtida com os dados da amostra e de um conhecimento *a priori* sobre estes parâmetros, expresso através de uma distribuição de probabilidades *a priori*. Esta combinação, via o Teorema de Bayes, resulta na distribuição *a posteriori* do parâmetro, que expressa como o conhecimento *a priori* se modificou à luz dos dados (verossimilhança). Para isto, os parâmetros são tratados como variáveis aleatórias e as inferências sobre eles são feitas com sua distribuição *a posteriori*. As distribuições preditivas permitem que dados faltantes sejam estimados pela distribuição *a posteriori* dos parâmetros do modelo, derivada a partir dos dados observados. No caso do Infer, a idéia é tratar os dados dos nodos dorminhocos como se fossem faltantes e estimá-los usando os dados dos outros sensores, via distribuições preditivas. Por ser tratar de um método baseado em inferência estatística, é necessário estabelecer modelos probabilísticos para as variáveis observadas, assim como para os parâmetros destes modelos (as distribuições *a priori*). Isto não deverá ser um problema quando se conhece as variáveis em estudo. Considerando que a suscetibilidade a falhas é uma característica importante das redes de sensores e deve ser levada em conta nas propostas de roteamento de dados, o Infer apresenta uma alternativa interessante para lidar com sensores falhos, estimando seus valores através dos valores dos outros sensores.

Existem várias outras propostas de roteamento de dados para redes de sensores (Al-Karaki e Kamal, 2004 ; Akkaya e Younis, 2004). A escolha do procedimento de roteamento deve considerar características da rede como sua mobilidade e seu modelo de envio de dados. O LEACH, por exemplo, é um protocolo eficiente em energia para redes de sensores estacionárias e com modelo de envio de dados contínuo, mas precisa ser adaptado para redes não-estacionárias e com modelo de envio de dados orientado a evento (Tilak et al., 2002). Já a *Directed Diffusion* é eficiente para redes não-estacionárias e pode ser aplicada a redes com modelo de envio de dados orientado a eventos e também por demanda. No entanto, não é adequada para redes com envio contínuo de dados, como as utilizadas em monitoramento ambiental, por exemplo

(Akkaya e Younis, 2004). Para este tipo de rede, a árvore de roteamento é um procedimento adequado, mas deve ser adaptado para lidar com mudanças freqüentes de topologia, seja por mobilidade ou por falhas nos sensores (Figueiredo et al., 2004). Os procedimentos de roteamento geográfico são mais adequados para redes com modelo de envio de dados por demanda, já que se aproveitam dos atributos geográficos da consulta. As propostas de Vuran et al. (2004) e a do Infer podem ser interessantes para redes densas, com grande número de nodos por unidade de área, nas quais a transmissão de todos os dados certamente esgotaria os recursos de energia da rede rapidamente. Para redes densas e com modelo de envio de dados contínuo, a proposta do Infer é particularmente interessante.

3.2 – Algumas considerações

A escolha da agregação de dados como meio de reduzir o consumo de energia deve ser feita com bastante cuidado e sempre orientada pelos objetivos da aplicação. A informação sobre a localização dos nodos que têm seus dados agregados deve ser mantida, pois é importante para a maioria das aplicações.

Em redes densas, a agregação pode auxiliar tanto no roteamento quanto na análise de dados: pequenos grupos de nodos podem ficar responsáveis por sensoriar determinada região e a agregação de seus dados pode melhorar a qualidade da informação sobre esta região (diminuição da variabilidade) e também reduzir o número de transmissões. Os procedimentos de roteamento do tipo hierárquico, em que há a formação de conglomerados antes da transmissão dos dados, podem ser utilizados neste caso. A análise de dados na presença de agregação será discutida no Capítulo 5.

O tipo de roteamento a ser utilizado numa rede de sensores e o modo como seus dados serão armazenados estão altamente relacionados. Em alguns casos, uma mesma proposta é vista como de roteamento e de banco de dados. O armazenamento de dados de sensores será discutido no próximo capítulo.

CAPÍTULO 4

BANCOS DE DADOS PARA REDES DE SENSORES

Um banco de dados de sensores é uma combinação de *dados sobre os sensores* (uma lista de sensores e seus atributos, como localização e tipo, por exemplo) e *dados dos sensores* (as leituras das variáveis que eles estão medindo). Cada nodo sensor pode ser visto com uma fonte separada que gera dados sobre seus atributos e dados sobre o que ele está sensoriando.

Utilizando o modelo de banco de dados *relacional* (Chamberlin, 1976), cada nodo sensor é considerado como uma *tupla* (linha) de uma *relação* (tabela). Seus atributos e seus dados formam os *campos* (colunas). As consultas que interessam ao usuário serão feitas sobre este banco de dados.

4.1 – Armazenamento e consultas

É importante entender como acontece o armazenamento dos dados num banco para redes de sensores: que dados são armazenados e em qual local isto ocorre. Isto determinará como serão processadas as consultas ao banco e quais tipos de análises poderão ser feitas com os dados extraídos da rede.

Segundo Ratnasamy et al. (2002), o **armazenamento dos dados** de uma rede de sensores pode ser classificado pelo local onde ele ocorre: **externo** (todos os dados são passados ao observador, que se encarrega de processá-los); **local** (os dados são armazenados localmente pelos sensores); **centrado em dados** (o dado é armazenado por nome e as consultas são feitas por aquele nome ao sensor responsável por aquele tipo de dado, que não precisa ser o sensor que o originou).

O armazenamento dos dados pode ser do tipo **histórico** (as séries históricas de dados coletados pelos sensores são armazenadas, como no caso das aplicações de **estudo retrospectivo**) ou do tipo **não-histórico** (os dados são extraídos da rede à medida que são feitas consultas ao banco, o que é mais comum nas aplicações do tipo **monitoramento**).

Ao classificar as maneiras de se processar consultas aos dados de sensores, Bonnet et al.(2001) apontaram duas alternativas para a construção de um banco de dados de sensores: alternativa centralizada e alternativa distribuída.

Na alternativa **centralizada** (*warehousing*), a rede de sensores é simplesmente utilizada como um mecanismo de coleta de dados: estes são coletados pela rede de uma forma pré-determinada (continuamente ou em intervalos regulares) e armazenados num banco de dados localizado em um servidor (a estação-base). Numa etapa posterior, a consulta é feita neste banco de dados centralizado. Esta alternativa implica em **armazenamento externo**, que também é do tipo histórico. Exemplos: bancos de dados dos projetos *Great Duck Island*, *ZebraNet*, *PODS* e *FloodNet*.

Já na alternativa **distribuída**, a construção do banco de dados é **orientada a consultas**, ou seja, as consultas determinam que dados devem ser extraídos dos sensores. Navas e Wynblatt (2001) resumiram a idéia desta alternativa afirmando que “a rede de sensores é o banco de dados”. Esta alternativa implica em **armazenamento local** ou **centrado em dados**. Entre as propostas de banco de dados distribuído, estão o *TinyDB* e o *COUGAR*, detalhados mais adiante.

A consulta a um banco de dados de sensores pode ser do tipo histórica, instantânea ou contínua. Numa **consulta histórica**, o usuário está interessado em recuperar dados de um certo período do tempo e, geralmente, numa certa região geográfica. Estas consultas podem ser chamadas também de **consultas espaço-temporais**. Um exemplo deste tipo de consulta seria : “quais as leituras de pH da água da Lagoa da Pampulha no setor

próximo à barragem entre 7:00 e 12:00 de ontem?” É importante notar que este tipo de consulta só é possível em bancos de dados com **armazenamento histórico**.

Numa **consulta instantânea**, o interesse do usuário está em leituras dos sensores no momento atual ou num futuro próximo. No monitoramento de edificações, por exemplo, uma consulta ao banco pode ser do tipo: “quais os valores de oscilação das colunas A e B (devido a fortes ventos) nos próximos 3 minutos?” Os bancos de dados com **armazenamento não-histórico** são formados a partir de consultas instantâneas.

As **consultas contínuas** são aquelas que examinam os dados assim que eles chegam até o processador, de modo contínuo, geralmente com o objetivo de monitorar algum fenômeno e detectar evento algum relacionado a ele. As consultas contínuas foram propostas por Terry et al. (1992) no contexto de filtragem de documentos, mas têm grande aplicação em redes de sensores, principalmente aquelas destinadas ao monitoramento de ambientes (Madden et al., 2002b). Uma consulta contínua pode ser vista como uma coleção de consultas instantâneas.

Um tipo de consulta também muito comum nas aplicações de monitoramento é a **consulta orientada a dados**. Neste tipo de consulta, o usuário está mais interessado em saber sobre o atributo do fenômeno do que conhecer o valor de um determinado nodo. Por exemplo, uma consulta do tipo “quais as áreas possuem temperatura maior do que 50° C” pode ser mais interessante do que saber qual o valor de temperatura de um nodo individual. A primeira consulta pode ajudar a identificar áreas com risco de incêndio ou com aquecimento além do normal.

A **linguagem de consulta** geralmente proposta é a SQL (*Structure Query Language*) ou uma variação da sua estrutura SELECT-FROM-WHERE-GROUPING BY-HAVING. Uma das propostas de linguagem, a TAG (Madden et al., 2002a), define um operador para lidar com o aspecto temporal (SAMPLE INTERVAL t , onde t é o intervalo entre

sucessivas amostras dos dados dos sensores, denominado *epoch*). Um exemplo de consulta usando a TAG poderia ser:

```
SELECT          luminosidade, temperatura , nodoID
FROM           sensores AS s
WHERE          (s.luminosidade > 10) AND (s.temperatura <
20)
GROUPING BY    s.nodoID
SAMPLE INTERVAL 2 FOR 30
```

Supondo que a unidade de tempo seja o segundo, esta consulta retornará os valores de luminosidade, temperatura e a identificação dos nodos cujo valor de luminosidade seja maior do que 10 e o valor da temperatura seja menor do que 20, agrupados pela identificação do nodo (*nodoID*). Esta consulta se repetirá de 2 em 2 segundos, durante 30 segundos. Note que um nodo pode aparecer várias vezes no resultado final desta consulta.

4.2 - Propostas para banco de dados de sensores

A seguir, são detalhadas algumas alternativas de armazenamento de dados de sensores. Antes, porém, a Tabela 4.1 apresenta uma classificação destas propostas segundo seu local e o período do armazenamento dos dados.

TABELA 4.1 – Classificação de algumas alternativas de armazenamento de dados dos sensores segundo o local (centralizado ou distribuído) e período (histórico ou não).

Histórico	Armazenamento	
	Centralizado	Distribuído
SIM	Bancos de dados dos projetos <i>Great Duck Island</i> , ZEBRANET, PODS e FloodNet	STWin, DIMENSIONS
NÃO	-----	TinyDB, COUGAR, Armazenamento centrado em dados

a) **TinyDB** (Madden et. al, 2003). O TinyDB é um sistema de processamento de consultas para extrair informações de rede de sensores que usem o sistema operacional TinyOS, desenvolvido para a arquitetura dos MICA *Motes* (Figuras 2c e 2d). O TinyOS é um programa de código aberto e está disponível na Internet. O TinyDB enxerga os dados dos sensores como uma única tabela, com uma coluna por tipo de sensor, cada linha representando um nodo (**armazenamento distribuído**). Linhas (tuplas) são anexadas a esta tabela periodicamente, em intervalos de amostragem que são um dos parâmetros da consulta (**armazenamento não-histórico**). O período de tempo entre intervalos de amostragem é chamado *epoch*. Os nodos amostram seus sensores uma vez por *epoch*. Cada nodo mantém um catálogo com meta-dados que descrevem sua localização, eventos e funções pré-definidas pelo usuário. As consultas são originadas num processador central na estação-base. Dada uma consulta de interesse, o TinyDB coleta os dados dos sensores através de uma **árvore de roteamento**, filtra e agrega estes dados, transportando-os até a estação-base. A linguagem de consulta utilizada é a TAG (Hellerstein et al, 2003), uma variação da SQL. A TAG permite que *eventos* sejam pré-definidos e que determinada consulta seja feita assim que este evento ocorrer. Por exemplo, no monitoramento ambiental do projeto *Great Duck Island*, um evento poderia ser a detecção da presença do pássaro no ninho. Quando este evento ocorresse, seria feita uma consulta sobre a intensidade da luz e a temperatura dos nodos a menos de 10 metros de distância do ninho.

b) **COUGAR** (Bonnet et al., 2001 ; Yao e Gerke, 2002). Como o TinyDB, a proposta do COUGAR enxerga a rede de sensores como um enorme banco de dados distribuído, no qual os dados dos sensores são vistos como linhas (tuplas). No entanto, o COUGAR utiliza *abstract data types* (ADT)² para representar os dados dos sensores e uma forma modificada da SQL para recuperar estes dados. A estrutura SELECT-FROM-WHERE é mantida e são definidos operadores nos ADT para lidarem com o aspecto temporal.

² ADT é um tipo de abstração de dados no qual os detalhes de implementação estão escondidos (abstraídos). Os valores deste tipo só podem ser criados e inspecionados por funções de acesso definidas para este tipo. Isto permite que a implementação do tipo possa ser modificada sem a necessidade de mudar o programa no qual ele está inserido.

Assim como no TinyDB, as consultas são originadas num processador central na estação-base.

c) **STWin (Spatio-Temporal WINDOW)** (Coman et al., 2004). Nesta proposta de armazenamento, a rede de sensores é vista como um **banco de dados distribuído** e especializado que armazena dados espaço-temporais. Cada nó tem um banco onde armazena seus dados para consultas futuras (**armazenamento local**). Isto permite que sejam feitas consultas históricas e evita que todos os dados tenham que ser transmitidos para um banco fora da rede. Para isso, o STWin pressupõe que os nós têm uma capacidade de armazenamento longa. Pressupõe também que a rede é do tipo homogênea e seus nós conhecem sua localização e a de seus vizinhos, obtidas durante a fase de estabelecimento da rede. A coleta de dados pelos sensores é contínua, mas somente uma agregação temporal dos dados é armazenada (ex., média dos dados a cada 30 segundos). O tipo e o intervalo da agregação são definidos antes da instalação da rede. As **consultas podem ser originadas em qualquer nó** e chegam até o nó de interesse em duas etapas: primeiramente, o nó originador da consulta localiza um nó na área geográfica requisitada. Posteriormente, este nó assume o papel de coordenador e fica responsável por coletar as respostas de seus vizinhos e mandá-las de volta ao nó originador. Como destacado por Ratnasamy et al.(2002), um dos problemas do armazenamento local é que uma consulta aos dados gera mensagens que inundam toda a rede, implicando em gasto de energia com transmissão. Para contornar este problema, os autores do STWin recorrem a uma variação do algoritmo de inundação³, aproveitando-se do caráter espacial das consultas. Diferentemente das propostas do TinyDB e do COUGAR, o STWin prioriza o armazenamento dos dados para futuras consultas (**histórico**), não considerando consultas apenas do tipo *long-running*, na qual o usuário deseja monitorar continuamente o fenômeno sensoriado.

³ Uma mensagem é transmitida por um nó a seus vizinhos, que as retransmitem para seus vizinhos e, assim por diante, inundando toda a rede com a mensagem até que ela chegue ao nó de destino. Isto gera um número enorme de transmissões, implicando em grande gasto de energia, o que é inadmissível numa rede de sensores.

d) DIMENSIONS (Ganesan et al., 2002): na tentativa de conciliar a necessidade de **armazenamento histórico** de dados e a limitação de memória da maioria dos sensores, os autores propuseram que os dados fossem comprimidos antes de serem armazenados nos nodos sensores (**armazenamento local**). Esta compressão é feita em duas fases: na primeira fase, acontece a compressão da série temporal dos dados de cada nodo, onde ficam armazenados os dados comprimidos; na segunda fase, acontece a compressão espacial dos dados, que pode ter vários níveis, desde o mais fino até o mais grosseiro. As consultas ao banco são feitas também em etapas, examinando primeiro o resumo dos dados no nível de compressão mais grosseiro e, se houver interesse, descendo até os níveis mais finos.

e) Armazenamento centrado em dados (*Data-centric Storage*) (Ratnasamy et al., 2002). Nesta proposta, combinações de dados consideradas interessantes (**eventos**) são nomeadas e armazenadas em nodos da rede. Um evento pode ser, por exemplo, a presença de pessoas numa sala. Todos os eventos de mesmo nome são armazenados no mesmo nodo, não necessariamente aquele que gerou o evento. Consultas sobre este evento são direcionadas para o nodo responsável por armazenar suas ocorrências, evitando que mensagens sejam enviadas a todos os nodos da rede (inundação). O roteamento utilizado é uma variação do GPSR (geográfico) e, portanto, os nodos devem saber a sua localização. A proposta do armazenamento centrada em dados prevê uma rede homogênea, com grande número de nodos, razoavelmente estacionária e estável (movimentos e falhas nos sensores não são comuns). Claramente, esta é uma proposta na qual as **consultas são orientadas a eventos**, em que **os dados originais não são armazenados**, como observam os autores (“na maioria dos casos, os usuários não querem um conjunto de dados não-processados, mas estão mais interessados em eventos específicos”).

f) Banco de dados espaciais para redes de sensores (Goldin, 2003 ; Goldin e Kutlu, 2004). Os autores exploram o aspecto geográfico dos dados de sensores e propõem que o banco de dados seja do tipo espacial, um banco de dados geográficos (Câmara, 2005). Este banco estaria na estação-base (**centralizado**) e teria relações do tipo

espacial (RegiaoID, RegiaoGeometria)

sensores (SensorID, SensorLoc, Valor)

Se a área de instalação dos sensores fosse, por exemplo, uma reserva ecológica dividida espacialmente em várias regiões (lago, pântano, areal, pasto, floresta fechada, por exemplo), a tabela *espacial* conteria a identificação destas regiões (*RegiaoID*) e seu polígono (*RegiaoGeometria*). A tabela *sensores* contém a identificação dos sensores (*SensorID*), suas coordenadas geográficas (*SensorLoc*) e o valor que o sensor mede. Consultas espaciais do tipo “qual é a média de umidade nas áreas de pasto?” seriam respondidas combinando-se os dados das duas tabelas. Uma das vantagens desta combinação de tabelas é que a definição das regiões e também sua geometria podem mudar independentemente da tabela de sensores. As regiões podem ser dinâmicas, com a geometria variando constantemente. Goldin e Kutlu (2004) introduz uma forma de processar consultas espaciais, na qual as agregações espaciais são feitas **intra-rede** e somente o valor desta agregação chega até o processador central na estação-base. O transporte dos dados seria feito através de uma árvore de roteamento, onde alguns nodos seriam responsáveis pela agregação em determinada região (*rivet nodes*). Esta estrutura permite que os parâmetros da consulta sejam avaliados ainda ao nível dos *rivet nodes*, possibilitando que apenas os dados das regiões que atendem a estes parâmetros sejam transmitidas até o processador central. Quanto ao período do armazenamento dos dados, os autores não entram em detalhes suficientes para que se possa inferir se o armazenamento é do tipo histórico ou não. No entanto, por se tratar uma proposta de banco de dados **orientado a consultas**, os autores provavelmente não deviam ter em mente o armazenamento histórico destes dados.

4.3 – Algumas considerações

Devido à natureza geográfica dos dados gerados por uma rede de sensores, os Sistemas de Informação Geográfica (SIG) representam um ambiente natural para abrigar um banco de dados de sensores. No entanto, ainda são poucas as propostas de banco de dados que mencionam a integração destes dados num ambiente de SIG. Entre elas, estão

o projeto FloodNet (Envisense), que utiliza um Sistema de Informações Geográficas para armazenar seus dados, e a proposta de armazenamento de Goldin e Kutlu (2004), na qual os dados dos sensores seriam armazenados num banco de dados geográficos.

As redes de sensores são uma fonte de alimentação natural para bancos de dados espaço-temporais (Erwig e Schneider, 2002 ; Worboys, 1994). O recente desenvolvimento de modelos e linguagens de consulta espaço-temporais será de grande valia no armazenamento histórico dos dados vindos das redes de sensores.

A rápida atualização dos bancos de dados de sensores também é um desafio nesta área de pesquisa. A chegada contínua de dados ao banco tem movido pesquisadores na direção dos bancos de dados seqüenciais (*stream databases* - Abadi et al., 2004), nos quais os dados são continuamente entregues ao processador de consultas. Geralmente, consultas contínuas são utilizadas neste tipo de banco de dados e as aplicações mais discutidas envolvem o monitoramento de ambientes, como prédios, pontes e fábricas.

Analisando as propostas para banco de dados de sensores, percebe-se que a formação do banco de dados orientada a consultas é a visão mais comum entre os pesquisadores, embora a maioria dos projetos que realmente instalaram redes possui bancos de dados com armazenamento histórico. Bancos de dados orientados a consultas são mais comuns em propostas para aplicações do tipo monitoramento, nas quais a avaliação do que está acontecendo agora é mais importante do que a análise de dados históricos. A forma de armazenamento dos dados está ligada ao tipo de análise a que eles serão submetidos, assunto do próximo capítulo.

CAPÍTULO 5

ANÁLISE ESTATÍSTICA EM REDES DE SENSORES

O projeto de uma rede de sensores é altamente dependente da sua aplicação e a análise dos dados provenientes dela também está relacionada aos objetivos com os quais a rede foi instalada. De maneira geral, os dados de uma rede de sensores podem ser submetidos a dois tipos de análise: análise instantânea e análise retrospectiva.

5.1 – Análise instantânea

A **análise instantânea** envolve tomada de decisão imediata a partir dos resultados e deve acontecer no momento em que o dado está disponível. Este tipo de análise é necessário nas aplicações de **monitoramento**. Na detecção de princípios de incêndios florestais, por exemplo, a chegada contínua de dados de temperatura considerados elevados numa certa região, aliados a dados de baixa umidade e/ou dados de presença de fumaça, certamente levará à análise de que há um foco de incêndio naquela região ou próximo dela. A análise instantânea pode se beneficiar das técnicas de mineração de dados (Hand et al., 2004).

Mineração de dados pode ser definida como a análise de grandes conjuntos de dados com o objetivo de resumi-los e encontrar relacionamentos ocultos entre eles de uma maneira compreensível e útil para o analista. A mineração de dados utiliza técnicas estatísticas, como análise de regressão e estimação por máxima verossimilhança, teoria de probabilidades e algoritmos de aprendizagem de máquina. As técnicas de mineração de dados podem ser úteis para identificar padrões, valores discrepantes, entre outras tarefas presentes nas aplicações de monitoramento. Para exemplificar, Manganaris et al. (2000) utilizaram uma das técnicas de mineração de dados mais conhecidas, as regras de

associação, para melhorar o processo de detecção de invasores em tempo real. Sensores para detectar invasão produzem muitos alarmes falsos durante um dia. Um dos objetivos dos autores era caracterizar o comportamento normal de um sensor para ajudar no processo de distinção entre um alarme falso e um verdadeiro.

Os dados analisados instantaneamente podem ser descartados ou armazenados para uma análise retrospectiva. Uma outra opção é guardar dados sobre os eventos detectados para montar uma base de conhecimento sobre o fenômeno monitorado.

5.2 – Análise retrospectiva

A **análise retrospectiva** tem por objetivo estudar os dados observados, através de sua descrição e modelagem, permitindo a predição de futuras observações (**estudos retrospectivos**). Em geral, redes projetadas com esta finalidade devem possuir um banco de dados com **armazenamento histórico** e um modelo de envio de dados contínuo.

Considerando a natureza geográfica das redes de sensores e que cada sensor é uma fonte pontual de dados, é natural imaginar que estes dados podem ser tratados como espacialmente contínuos, uma das quatro classificações de dados dentro da Estatística Espacial (Bailey e Gatrell, 1995). Os **dados espacialmente contínuos** são o resultado de variáveis aleatórias observadas no espaço geográfico, em pontos de localização conhecida, como, por exemplo, os dados de umidade do ar coletados pelos sensores do projeto *Great Duck Island*. De uma maneira mais formal, tem-se que $Z(x_i, y_i, t)$ é uma variável aleatória observada no ponto i , de coordenadas (x_i, y_i) , e no tempo t . O conjunto de variáveis $Z(x_i, y_i, t)$, ($i = 1, 2, \dots, n$) e ($t = 1, \dots, m$), pode ser visto como um campo aleatório, uma coleção de variáveis aleatórias dependentes entre si que têm um componente espaço-temporal⁴ (Guttorp, 1995).

⁴ As variáveis de um campo aleatório podem possuir apenas o componente espacial (localização geográfica), como acontece no caso de variáveis observadas somente uma vez ($m=1$).

Numa **rede de sensores estacionária** que coleta dados em intervalos de tempo, os pontos têm coordenadas fixas e conhecidas, isto é, os pontos (x_i, y_i) são os mesmos a cada tempo t . Cada medição gera a realização de um campo aleatório bi-dimensional, correlacionados temporalmente entre si. Estas realizações podem ser vistas como uma coleção de campos aleatórios ou como um campo aleatório tri-dimensional único, no qual o tempo é a terceira dimensão. Os dados gerados pela medição são do tipo **espacialmente contínuos**.

Em **redes não-estacionárias**, os sensores devem registrar, além dos dados sobre o fenômeno e o tempo, também a sua localização geográfica. Diferentemente das redes estacionárias, os pontos (x_i, y_i) mudam a cada tempo t . Os dados coletados por uma rede não-estacionária podem ser considerados e tratados como **espacialmente contínuos**, desde que o interesse de estudo não esteja no padrão espacial dos nodos sensores, mas nos valores medidos por eles. Em outras palavras, a posição geográfica do ponto de coleta (nodo sensor) deve ser tratada como se fosse conhecida *a priori* e não como um componente aleatório. Caso contrário, os dados devem ser analisados como **processos pontuais** (Bailey e Gatrell, 1995).

As técnicas para análise de dados espacialmente contínuos são conhecidas como **Geostatística** (Isaaks e Srivastava, 1990). A mais conhecida destas técnicas é a krigagem. No modelo mais geral, a krigagem universal, o campo aleatório Z pode ser descrito como se segue

$$Z(x_i, y_i) = \mu(x_i, y_i) + U(x_i, y_i) + e(x_i, y_i) , \text{ para um tempo } t,$$

onde $\mu(x_i, y_i)$ é a média do processo e pode ser modelada com covariáveis;

$U(x_i, y_i)$ é um efeito aleatório espacialmente estruturado e

$e(x_i, y_i)$ é um ruído branco, isto é, $e(x_i, y_i)$ tem distribuição Gaussiana com média zero e variância constante e são não-correlacionados. É conhecido como *efeito pepita*.

Na krigeagem ordinária, $\mu(x_i, y_i)$ é considerada constante em todos os pontos. O objetivo da krigeagem é estimar Z num ponto $s = (x_s, y_s)$ não observado, ou seja, $Z(x_s, y_s)$.

Na abordagem de Estatística Clássica, a matriz de covariância de $U(x_i, y_i)$ é modelada e, através deste modelo, pode-se calcular as estimativas de covariância entre os pontos observados e um ponto $s = (x_s, y_s)$, para o qual deseja-se fazer a estimativa de Z . Estas estimativas, juntamente com a matriz de covariância estimada e a estimativa de $\mu(x_i, y_i)$, permitem a estimação de $Z(x_s, y_s)$.

Na abordagem Bayesiana, a matriz de covariância de $U(x_i, y_i)$ também é modelada, juntamente com os outros parâmetros, resultando na distribuição *a posteriori* destes parâmetros. A predição de Z para locais não observados é feita pelas *distribuições preditivas*, que são calculadas a partir da distribuição *a posteriori* dos parâmetros envolvidos no modelo de Z . Na verdade, a estimativa pontual do valor de Z é dada por um resumo da distribuição preditiva, por exemplo, a média.

Em 1986, Kitanidis mostrou que a krigeagem ordinária e a universal podem ser vistas como técnicas de predição espacial Bayesiana onde existe uma ignorância *a priori* sobre os parâmetros da superfície de tendência. Em outras palavras, a abordagem clássica para a krigeagem é um caso particular da abordagem Bayesiana.

Segundo Schmidt et al. (2002), a **incorporação do componente temporal** ainda é um desafio na análise de dados espacialmente contínuos. O problema está em propor uma estrutura de covariância que descreva a correlação espacial entre as observações para

cada instante de tempo. Uma proposta é que a estrutura de covariância seja a multiplicação da covariância espacial pela covariância temporal, ou seja,

$$\text{Cov}[Z(x_i, y_i, 1), Z(x_j, y_j, 2)] = \text{Cov}[Z(x_i, y_i), Z(x_j, y_j)] \times \text{Cov}[Z(1), Z(2)].$$

Entretanto, isto exige a suposição de que $\text{Cov}[Z(x_i, y_i, 1), Z(x_j, y_j, 2)] = \text{Cov}[Z(x_j, y_j, 1), Z(x_i, y_i, 2)]$, o que não parece razoável para a maioria dos fenômenos estudados. Suponha, por exemplo, que Z é a temperatura atmosférica e que os pontos $i = (x_i, y_i)$ e $j = (x_j, y_j)$ sejam tais que pertençam ao caminho normalmente percorrido pelas frentes frias vindas da Argentina. Se a frente passa por i e vai para j , é razoável pensar que um decréscimo da temperatura em i no tempo 1 estará relacionado a um decréscimo da temperatura em j no tempo 2. No entanto, isto não vai implicar que um decréscimo da temperatura em j no tempo 1 estará relacionado a um decréscimo de temperatura em i no tempo 2.

5.2.1 - Análise dos dados na presença de agregação

Em redes de sensores densamente distribuídas, o transporte dos dados de todos os nodos sensores é altamente dispendioso em termos de energia. Como visto anteriormente, a agregação destes dados é uma das estratégias utilizadas nos procedimentos de roteamento para diminuir o gasto energético. Em alguns destes procedimentos, os nodos sorvedouros (ou os líderes dos conglomerados) ficam responsáveis por agregar os dados dos nodos sob sua jurisdição e transmitir o resumo para a estação-base. Como o dado agregado está associado ao nodo sorvedouro, do ponto de vista da análise dos dados, seria como passar da configuração de pontos da Figura 5.1a (rede completa) para a configuração da Figura 5.1b (rede reduzida aos nodos sorvedouros). A redução do número de pontos seria compensada pela diminuição no nível de ruído presente nos dados, uma fonte de variabilidade que não interessa ao analista. Os dados ainda poderiam ser tratados como espacialmente contínuos.

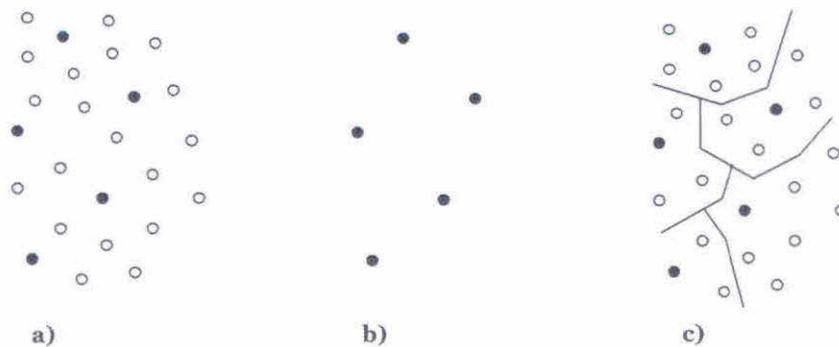


FIGURA 5.1 – Rede de sensores com agregação de dados intra-rede: a) visão da rede completa (nodos agregadores em preto) ; b) rede formada somente pelos nodos agregadores; c) campo dos sensores dividido em áreas, às quais estão associados os valores agregados pelos nodos em preto

Outra abordagem para a análise dos dados seria considerar que cada nodo sorvedouro responde por uma área, que é a união das áreas dos nodos sob sua jurisdição (Figura 5.1c). A área de um nodo pode ser obtida pela tesselação de Dirichlet, caso a informação sobre a localização dos nodos que enviaram seus dados ao sorvedouro seja preservada. Nesta abordagem, o dado agregado estaria associado a uma área e não mais a um ponto e teria que ser tratado com um **dado de área**, outra classe de dados dentro da Estatística Espacial (Bailey e Gatrell, 1995).

A **análise de dados de área** é bastante freqüente em Epidemiologia Espacial, especialmente na produção de mapas de risco de doenças (Bailey, 2001). Na modelagem espaço-temporal de dados de área, um procedimento que vem se tornando comum é a utilização da inferência Bayesiana. Nesta abordagem, é assumido um modelo probabilístico para $Z(x_i, y_i, t)$ com média μ_{it} e variância σ_{it}^2 e sua média é modelada. Um dos modelos espaço-temporais para μ_{it} pode ser:

$$f(\mu_{it}) = \alpha + \varphi_i + v_i + \delta_0 t + \delta_i t, \quad i = 1, \dots, n \text{ e } t = 1, \dots, m. \quad (5.1)$$

onde $f(\cdot)$ é uma função tal que os valores de $f(\mu_{it})$ pertence ao conjunto dos números reais;

α é a parte de μ_{it} que é comum a todas as áreas e tempos;

φ_i é parte não-espacialmente estruturada de μ_{it} ;

v_i é parte espacialmente estruturada de μ_{it} ;

δ_0 e δ_i são as partes do componente tempo não-espacialmente e espacialmente estruturadas, respectivamente.

Se, por exemplo, $Z(x_i, y_i, t)$ tem distribuição de probabilidade Gaussiana, $f(\cdot)$ é a função identidade, pois a média de uma distribuição Gaussiana pertence ao conjunto dos números reais. A dependência espacial das áreas entra no modelo através das distribuições *a priori* de v_i e δ_i . As opções geralmente utilizadas são: $\alpha \sim$ Uniforme $[-\infty ; +\infty]$; $\delta_0 \sim$ Uniforme $[-\infty ; +\infty]$; $\varphi_i \sim$ Gaussiana $(0 ; \sigma_\varphi^2)$.

$$v_i | v_{j \neq i} \sim \text{Gaussiana} \left(\frac{\sum_{j \neq i} w_{ij} v_j}{\sum_{j \neq i} w_{ij}}, \frac{\sigma_v^2}{\sum_{j \neq i} w_{ij}} \right) \quad , \text{ onde } w_{ij} \text{ são pesos representando a vizinhança das áreas.}$$

$$\delta_i | \delta_{j \neq i} \sim \text{Gaussiana} \left(\frac{\sum_{j \neq i} w_{ij} \delta_j}{\sum_{j \neq i} w_{ij}}, \frac{\sigma_\delta^2}{\sum_{j \neq i} w_{ij}} \right)$$

O modelo para v_i e δ_i é chamado Condicional Auto-Regressivo (CAR) Gaussiano. Ele condiciona a média de uma área nos valores de seus vizinhos, através da média ponderada deles. As inferências sobre cada parâmetro são feitas através de amostras da sua distribuição *a posteriori*. Estas amostras são conseguidas através de simulação via *Markov Chain Monte Carlo* (MCMC) (Geman, 1997). Caso v_i assuma valores

próximos a zero com grande probabilidade, isto significa que as áreas vizinhas à área i não têm muita influência sobre sua média. O mesmo vale para δ_i , indicando que as áreas vizinhas não têm influência no componente temporal da tendência μ_{it} .

O modelo em (5.1) pode incluir covariáveis, caso a análise tenha o objetivo de relacionar outras variáveis medidas pelo mesmo nodo sensor. Uma alternativa à abordagem Bayesiana no caso de inclusão de covariáveis seria utilizar a Regressão Geograficamente Ponderada (GWR – Brunsdon et al., 1998). A idéia do GWR é estimar os coeficientes de regressão para cada área, utilizando seus vizinhos como entrada de dados. A vizinhança de uma área seria definida em função de distância até os possíveis vizinhos. Levando em conta o aspecto temporal, cada intervalo de tempo seria modelado separadamente.

Do ponto de vista estatístico, tratar o dado agregado como se fosse de área é a abordagem mais adequada, já que ele não é formado apenas pelo dado do sorvedouro, mas também por uma combinação dos dados de seus vizinhos. No entanto, os resultados das análises estarão atrelados a estas áreas, que podem não ter sentido do ponto de vista prático. A proposta de armazenamento de dados de sensores feita por Goldin e Kutlu (2004) prevê que as áreas de interesse sejam definidas em um banco de dados geográficos e os dados dos sensores sejam agregados segundo estas áreas. O procedimento de roteamento LEACH-C (Heinzelman et al., 2002), uma variação do LEACH, propõe que os nodos agregadores (líderes) sejam escolhidos na estação-base através de simulação de modo a ficarem bem distribuídos em termos de energia. Este processo de escolha poderia incluir algum parâmetro geográfico de modo que o resultado fosse conglomerados bem distribuídos também em termos geográficos. Tanto na proposta de Goldin e Kutlu (2004) quanto na modificação do LEACH-C, o analista teria mais controle sobre quais seriam as áreas de trabalho.

Com a previsão de redes cada vez maiores, é quase consenso entre os pesquisadores da área que a transmissão de todos os dados até a estação-base é inviável. Redes com

modelos de envio de dados orientado a eventos ou por demanda (consulta) podem se aproveitar destas características para restringir o número de transmissões de dados através das propostas de roteamento/armazenamento centrado em dados e roteamento geográfico, respectivamente. No entanto, às redes com modelos de envio de dados contínuo, restam poucas opções de economia de energia. Uma delas é a agregação dos dados.

A agregação dos dados pode ser benéfica em termos da análise, pois ajuda a reduzir o nível de ruído, diminuindo a variabilidade, como comentado anteriormente.

5.3 – Algumas considerações

As redes de sensores abrem a possibilidade de uma visão detalhada do mundo físico. No entanto, “o tentador volume de observações produzido por estes sensores é ao mesmo tempo uma benção e uma maldição”, como bem observaram Ratnasamy et al. (2002). Mesmo sendo feita no contexto de roteamento de dados, a afirmação dos autores procede também no contexto da análise dos dados dos sensores. O que fazer com tantos dados ?

Nas aplicações de monitoramento, as análises dos dados que chegam são feitas instantaneamente, e o volume de dados pode ser tratado por técnicas de mineração de dados, por exemplo. Na análise retrospectiva, os dados devem ser pré-processados antes de serem modelados, evitando assim a preocupação com fontes de variabilidade não interessantes. Agregações espaciais e temporais podem ser consideradas como meio de melhorar a qualidade do dado.

Para Durrant-Whyte (2005), o grande desafio na área de redes de sensores é desenvolver métodos para lidar com a quantidade de dados gerados e transformá-los em informação útil. Ele propõe o uso de técnicas de fusão de dados através de modelos probabilísticos e técnicas Bayesianas (Hall, 1992).

CAPÍTULO 6

CONSIDERAÇÕES FINAIS

As redes de sensores são um objeto de pesquisa relativamente novo. As primeiras publicações com referências específicas às redes de sensores datam do final da década de 90 e os primeiros encontros organizados para a discussão do assunto foram realizados há dois ou três anos atrás⁵.

Como ferramenta de sensoriamento distribuído, as redes de sensores abrem uma nova perspectiva frente ao sensoriamento remoto tradicional, representado principalmente pelo uso de imagens ópticas e de radar. Dentre as vantagens do sensoriamento distribuído, estão a possibilidade de diminuir efeitos ambientais, colocando os sensores bem próximos ao fenômeno estudado ; o monitoramento contínuo de uma região e a localização mais precisa do fenômeno de interesse (Loureiro et al. 2003 ; Estrin et al., 2001). Outra vantagem é a possibilidade de aumentar a resolução espacial do dado coletado. As redes de sensores e o sensoriamento remoto tradicional desempenham papéis semelhantes, de maneiras diferentes e podem ser perfeitamente complementares. A integração entre estas duas ferramentas de sensoriamento é um dos objetivos do projeto SensorWeb.

Em 1997, a agência espacial americana (NASA) concebeu a SensorWeb, uma super-rede na qual um nodo pode ser uma rede de sensores ou uma plataforma de sensoriamento orbital. Como numa rede de sensores, os nodos da SensorWeb podem se comunicar uns com os outros. Para possibilitar esta comunicação, membros do *Open GIS Consortium*, incluindo a NASA, estão desenvolvendo a linguagem SensorML. A

⁵ **IPSN 2003** (International Symposium on Information Processing in Sensor Networks), **GSN 2003** (GeoSensors Networks), **SENET 2004** (International Conference on Sensor Networks), **EWSN 2004** (European Workshop on Wireless Sensor Networks), **ALGOSENSORS 2004** (First International Workshop on Algorithmic Aspects of Wireless Sensor Networks), **DMSN 2004** (International Workshop on Data Management for Sensor Networks)

SensorWeb pode ser vista como um instrumento que possibilitará o entendimento espaço-temporal de um ambiente através da coordenação de diferentes tipos de plataformas de sensoriamento, sejam elas orbitais ou terrestres, fixas ou móveis.

Uma sofisticação das redes de sensores são as redes que, além de possuírem nodos sensores, contam também com nodos atuadores. Enquanto os sensores sentem, os atuadores agem com base nos dados coletados pelos sensores. Os atuadores são mecanismos mais ricos dos que os sensores em termos de energia, capacidade de processamento e comunicação. Em geral, são móveis e instalados em um número bem menor do que o número de sensores. Um aspecto importante no desenvolvimento de redes de sensores e atuadores é coordenação entre os dois tipos de nodos (Akyildiz e Kasimoglu, 2004). Aplicações de monitoramento podem se beneficiar muito com este tipo de rede, pois os atuadores podem agir rapidamente assim que os sensores detectarem alguma anomalia no ambiente monitorado.

As previsões para as redes de sensores apontam para mecanismos cada vez menores, mais potentes e mais baratos. Isto possibilitará a instalação de mais redes e cada vez maiores. O desenvolvimento das pesquisas para viabilizar estas redes, principalmente na questão do roteamento, tem acompanhado a evolução tecnológica dos sensores.

As redes de sensores (e atuadores) prometem uma revolução na observação do mundo físico, oferecendo a possibilidade de um sensoriamento denso do ambiente. Seja na tarefa de vigilância ou de entendimento deste ambiente, as redes de sensores prometem ao homem o dom da ubiqüidade, ou quase isso.

REFERÊNCIAS BIBLIOGRÁFICAS

- Abadi, D. J. ; Lindner, W.; Madden, S. ; Schuler, J. An Integration Framework for Sensor Networks and Data Stream Management Systems. In: Proc. of the **30th Very Large DataBases (VLDB) Conference**, Toronto (CA), 2004.
- Akkaya, K. ;Younis, M. A survey on routing protocols for wireless sensor networks. **Ad Hoc Networks**, vol 3, n. 3, pp. 325—349, 2004.
- Al-Karaki, J. N. ; Kamal, A. E. Routing techniques in wireless sensor networks: a survey. **IEEE Wireless Communications**, vol. 11, n. 6, pp. 6--28, dezembro de 2004.
- Akyildiz, I. F.; Su, W. ; Sankarasubramaniam, Y.; Cayirci, E. Wireless Sensor Networks: a Survey, **Computer Networks**, 38, pp. 393–422, 2002.
- Akyildiz, I. F.; Kasimoglu, I. H. Wireless sensor and actor networks: research challenges. **Ad Hoc Networks**, vol. 2, pp. 351–367, 2004.
- Armstrong, P. Wireless Data Transmission of Networked Sensors in Grain Storage. Apresentado na **ASAE Annual International Meeting**, Las Vegas (EUA), julho de 2003.
- Bailey, T. C. Spatial statistical methods in health. **Cad. Saúde Pública**, vol.17, no.5, p.1083-1098, outubro de 2001.
- Bailey, T.; Gatrell, A. **Interactive Spatial Data Analysis**. London, Longman Scientific and Technical, 413 p, 1995.
- Basharat, A. ; Catbas, N. ; Shah, M. A Framework for Intelligent Sensor Network with Video Camera for Structural Health Monitoring of Bridges. In : Proc. of **Third IEEE International Conference on Pervasive Computing and Communications (PerCom)**, Kauai Island (Hawaii), março de 2005.
- Bharathidasan, A. ; Anand, V. ; Ponduru, S. **Sensor Networks: An Overview**. Department of Computer Science at the University of California, 2003.
Disponível em : <http://wwwcsif.cs.ucdavis.edu/~bharathi/sensor/survey.pdf>
Acesso em: 31 de julho de 2005.
- Biagioni, E. ; Sasaki, G. Wireless sensor placement for reliable and efficient data collection. In: Proc. of the **Hawaii International Conference on Systems Sciences**, janeiro de 2003.
- Bonnet, P.; Gehrke, J.; Seshadri, P. Towards sensor database systems. In Proc. of **IEEE Conference on Mobile Data Management**, pp. 3–14, 2001.

- Brunsdon C. ; Fotheringham S. ; Charlton M. Geographically weighted regression—modelling spatial non-stationarity, **Journal of Royal Statistical Society, Series D**, vol. 47, pp. 431-444, 1998
- Bulusu, N.; Heidemann, J. ; Estrin, D. GPS-less low cost outdoor localization for very small devices, **Technical report 00-729, Computer science department**, University of Southern California, abril de 2000.
- Carlin, B.; Louis, T. A. **Bayes and Empirical Bayes Methods for Data Analysis**. Ed. Chapman & Hall, London, 399 p., 1996.
- Câmara, G. Representação computacional de dados geográficos. In : Casanova, M.; Câmara, G ; Davis, C. ; Vinhas, L.; Ribeiro, G. (org), **Bancos de Dados Geográficos**. São José dos Campos, MundoGEO, 2005.
- Chamberlin, D. D. Relational Data-Base Management Systems, **ACM Computing Surveys**, n. 8, pp. 43—66, 1976.
- Coman, A. ; Nascimento, M.A. ; Sander, J. A Framework for Spatio-Temporal Query Processing Over Wireless Sensor Networks. In: Proc. of the **First Workshop on Data Management for Sensor Networks (DMSN 2004)**, Toronto (CA), agosto de 2004.
- Diggle, P. J. ; Tawn, J. A. et al. Model-based geostatistics. **Journal of the Royal Statistical Society Series C-Applied Statistics**, vol. 47, pp. 299-326, 1998.
- Durrant-Whyte, H. Data fusion in sensor networks. In: Proc. of the **Fourth International Symposium on Information Processing in Sensor Networks (IPSN)**, pp. 2-2, 2005.
- Erwig, M.; Schneider, M. STQL: A Spatio-Temporal Query Language”, In: Ladner, R. ; Shaw, K. ; Abdelguer, M. (editors), **Mining Spatio-Temporal Information Systems**. Kluwer, 2002 .
- Estrin, D. ; Girod, L. ; Pottie, G. ; Srivastava, M. Instrumenting the world with wireless sensor networks. In: Proc. **International Conference on Acoustics, Speech, and Signal Processing (ICASSP'01)**, maio de 2001.
- Envisense GlacsWeb. Disponível em: <http://envisense.org/glacsweb.htm>. Acesso em 31 de julho de 2005.
- Envisense FloodNet. Disponível em: <http://envisense.org/floodnet/floodnet.htm>. Acesso em 31 de julho de 2005.
- Figueiredo, C. M. S., Nakamura, E. F., and Loureiro, A. A. F. Multi: A hybrid adaptive dissemination protocol for wireless sensor networks, **Lecture Notes in Computer Science**, vol. 3121 (ALGOSENSORS'04), pp. 171–186, 2004.
- Gamerman, D. **Markov Chain Monte Carlo: stochastic simulation for Bayesian inference**. Chapman and Hall, London, 245 p., 1997.

- Ganesan, D. ; Estrin, D. ; Heidemann, J. DIMENSIONS: Why do we need a new data handling architecture for sensor networks. In Proc. of **Workshop on Hot Topics in Networks**, 2002.
- Goldin, D. Spatial Queries over Sensor Networks. **University of Connecticut BECAT/CSE Technical Report**, março de 2003.
- Goldin, D.; Kutlu, A. In-Network Evaluation of Spatial Aggregation in Sensor Networks. **University of Connecticut BECAT/CSE Technical Report TR-04-3**, maio de 2004. Disponível em : <http://www.engr.uconn.edu/dqg/papers/techrep04-3.pdf> .
- Guttorp, P. **Stochastic modeling of scientific data**. Chapman & Hall, 372 p., 1995.
- Hall, D. L. **Mathematical Techniques in Multi-Sensor Data Fusion**. Artech House, 1992.
- Hand, D. ; Mannila, H.; Smyth, P. **Principles of Data Mining**. MIT Press, 546 p., 2004.
- Hartl, G. ; Li, B. Infer: A Bayesian Inference Approach towards Energy Efficient Data Collection in Sensor Networks. In: Proc. of the **25th International Conference on Distributed Computing Systems (ICDCS)**, Columbia (EUA), pp. 371-380, junho de 2005.
- Hellerstein, J. ; Hong, W. ; Madden, S. ; Stanek, K. Beyond Average: Towards Sophisticated Sensing with Queries. In: Proc. of **2nd Int. Workshop on Information Processing in Sensor Networks (IPSN '03)**, março de 2003.
- Heidemann, J. and Bulusu, N. Using Geospatial Information in Sensor Networks, In : Proc. of the **Computer Sciences and Telecommunications Board (CSTB) Workshop on the Intersection of Geospatial Information and Information Technology**, Arlington, VA. outubro de 2001.
- Heinzelman, W.R. ; Kulik, J. ; Balakrishnan, H. Adaptive protocols for information dissemination in wireless sensor networks, Proc. **ACM/MOBICOM**, Seattle, pp. 174—185, 1999.
- Heinzelman, W. B. ; Chandrakasan, A. ; Balakrishnan, H. Energy-Efficient Communication Protocol for Wireless Microsensor Networks. In: Proc. of the **33rd. Hawaii Int. Conference on System Science**, 2000.
- Heinzelman, W. B. ; Chandrakasan, A. ; Balakrishnan, H. An Application-Specific Protocol Architecture for Wireless Microsensor Networks. *IEEE Transactions On Wireless Communications*, vol. 1, no. 4, outubro de 2002.
- Intanagonwiwat, C. ; Govindan, R. ; Estrin, D. Directed Diffusion_ A Scalable and Robust Communication Paradigm for Sensor Networks. In: Proc. **ACM/MOBICOM**, pp. 56—67, 2002.

INTEL Laboratories. The Wireless Vineyard.

Disponível em URL:

<http://www.intel.com/technology/techresearch/research/rs01031.htm>

Acesso em 31 de julho de 2005.

Isaaks, E. ; Srivastava, M. **Applied Geostatistics**. New York (OUP), 555 p., 1990.

Juang, P. ; Oki, H. ; Wang, Y. ; Martonosi, M. ; Peh, L. ; Rubenstein, D. Energy-Efficient Computing for Wildlife Tracking: Design Tradeoffs and Early Experiences with ZebraNet. IN: Proc. of 10th Intl. **Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS-X)**, San Jose (CA), outubro de 2002.

Kalpakis, K., Dasgupta, K., and Namjoshi, P. Efficient algorithms for maximum lifetime data gathering and aggregation in wireless sensor networks. **The International Journal of Computer and Telecommunications Networking**, vol. 42, número 6, pp. 697—716, agosto de 2003.

Karp, B. ; Kung, H. T. GPSR: Greedy perimeter stateless routing for wireless sensor networks. In : Proc. of the **6th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom '00)**, Boston (MA), agosto de 2000.

Kitanidis, P. K. Parameter Uncertainty in Estimation of Spatial Functions - Bayesian-Analysis. **Water Resources Research**, vol. 22, n. 4, pp. 499-507, 1986.

Loureiro, A. A. ; Nogueira, J. M. S., Ruiz, L. B., Nakamura, E., Seródio, C. M., e Mini, R. Redes sensores sem fio. Em : Livro texto de mini-cursos do **XXI Simpósio Brasileiro de Redes de Computadores (SBRC)**, ISBN 85-88442-48-5, pp. 179–226, 2003

Madden, S. ; Franklin, M.J. ; Hellerstein, J.M. TAG: a tiny aggregation service for ad-hoc sensor networks. In Proc. of **Symposium on Operating Systems Design and Implementation**, pp. 131–146, 2002a.

Madden, S.; Hellerstein, J.; Shah, M. ; Raman, V. Continuously adaptive continuous queries over streams. In Proc. of **ACM SIGMOD Intl. Conf. on Management of Data**, 2002b.

Madden, S.; Franklin, M.J. ; Hellerstein, J.M. The design of an acquisitional query processor for sensor networks. In: Proc. of **SIGMOD**, pp. 491–502, 2003.

Mainwaring, A. ; Polastre, J. ; Szewczyk, R. ; Culler, D. ; Anderson, J. Wireless Sensor Networks for Habitat Monitoring, **ACM International Workshop on Wireless Sensor Networks and Applications (WSNA'02)**, Atlanta, EUA, pp. 88 – 97, 2002.

Manganaris, S. ; Christensen, M. ; Zerkle, D. ; Hermiz, K. A data mining analysis of RTID alarms. **Computer Networks**, vol. 34, pp. 571--577, 2000.

- Navas, J. C. ; Wynblatt, M. The Network is the Data Base: Data Management for Highly Distributed Systems. In: Proc. of **ACM/SIGMOD**, pp. 544—551, 2001.
- Nittel, S., Duckham, M., Kulik, L. Information dissemination in mobile ad-hoc geosensor networks. In Egenhofer, M.J., Freksa, C. and Miller, H.J. (eds) **Lecture Notes in Computer Science** 3234, Springer, pp. 206-222, 2004.
- Paskin, M. A. ; Guestrin, C. E. e McFadden, J. A Robust Architecture for Inference in Sensor Networks. In **Proceedings of the Fourth International Symposium on Information Processing in Sensor Networks (IPSN-05)**, 2005.
- Pereira, M. R. ; Castro, M. C. S ; Amorim, C. L. **Tutorial sobre Redes de Sensores**, Cadernos do IME - Série Informática (ISBN: 1413-9014), vol. 4. Universidade Federal do Rio de Janeiro, Rio de Janeiro.
Disponível em URL: <http://magnum.ime.uerj.br/cadernos/cadinf/vol14/3-clicia.pdf>. Acesso em 20/07/2005.
- Pottie, J. ; Kaiser, W. J. Embedding the internet wireless integrated network sensors, **Communications of the ACM**, vol. 43, no. 5, pp. 51– 58, maio de 2000.
- Ratnasamy, S. ; Estrin, D. ; Govindan, R. ; Karp, B. ; Shenker, S. ; Yin, L. ; Yu, F. Data-centric storage in Sensornets. In: Proc. **ACM/SIGMOD**, 2002 .
- Ruiz, L.B.; Correia, L.H.; Vieira, L. F. et al. Arquitetura de redes de sensores sem fio. In : **Livro texto de mini-cursos do XXII Simpósio Brasileiro de Redes de Computadores (SBRC)**, ISBN 85-88442-81-7. Gramado (RS), Brasil, cap. 4, pp. 167–218, 2004.
- Schmidt, A. M.; Nobre, A. A.; Ferreira, G. S. Alguns Aspectos da Modelagem de Dados Espacialmente referenciados, **Revista Brasileira de Estatística**, vol. 63, número 220, pp. 59-88, 2002.
- Steinberg, A. N.; Bowman, C. L.; White, F. E. Revisions to the JDL data fusion model. Proceedings of the **International Society for Optical Engineering (SPIE 1999)**, Orlando, FL, USA 3719, pp. 430—441, abril de 1999.
- Terry, D. ; Goldberg, D. ; Nichols, D. ; Oki, B. Continuous queues over append-only databases. In: Proc. of **ACM SIGMOD**, pp. 321–330, 1992.
- Tilak, S. ; Abu-Ghazaleh, N.B. ; Heinzelman, W. A taxonomy of wireless micro-sensor network models, In Proc.of the **ACM Workshop on Wireless Security**, ACM Press, pp 28-36, 2002.
- Vuran, M. C. ; Akan, O.B. ; Akyildiz, I. F. Spatio-temporal correlation: theory and applications for wireless sensor networks. **Computer Networks**, vol. 45, pp. 245–259, 2004.

Xu, Y. ; Heidemann, J. ; Estrin, D. Geography-informed Energy Conservation for Ad-hoc Routing," In: Proc. of the **Seventh Annual ACM/IEEE International Conference on Mobile Computing and Networking**, pp. 70-84, 2001.

Worboys, M.F. A Unified model of spatial and temporal information. **Computer Journal**, vol. 37, n.1, pp. 26-34, 1994.

Yao, Y.; Gehrke, J. The Cougar Approach to In-Network Query Processing in Sensor Networks. **SIGMOD Record**, vol. 31, n. 3, setembro de 2002.

Yu, Y. ; Estrin, D. ; Govindan, R. Geographical and Energy-Aware Routing: A Recursive Data Dissemination Protocol for Wireless Sensor Networks, **UCLA Computer Science Department Technical Report**, UCLA-CSD TR-01-0023, maio de 2001.