

**UNIVERSIDADE ESTADUAL DE MARINGÁ**  
**ESPECIALIZAÇÃO EM DESENVOLVIMENTO DE SISTEMA PARA**  
**WEB**

**Ercal Netto de Lima**

**GERENCIAMENTO DE CACHE EM BANCO DE DADOS MÓVEIS**

Monografia submetida à Universidade Estadual de Maringá como parte dos  
requisitos para obtenção do grau de Especialista em Desenvolvimento de  
Sistema para Web

**ORIENTADORA**

**Heloise Manica**

Maringá – Pr , 27 de Julho de 2005

Dedico esta monografia àqueles que de forma simples e honesta contribuíram para sua realização. Em especial a meus pais, Maria e Expedito. A minha esposa Sirlei que sempre me incentivou a buscar novos conhecimentos.

## **AGRADECIMENTOS**

A DEUS pela graça do dom da vida, por sempre ter me orientado para que possa adquirir capacidades segundo minhas oportunidades.

A Professora Heloise Manica, minha orientadora, que me acolheu e me conduziu ao longo desta jornada.

Aos demais professores e funcionários do departamento do DIN da UEM, pela acolhida e conhecimentos a mim transmitidos.

Aos meus pais que foram os grandes mestres de minha vida que sempre me aconselharam e conduziram por caminhos corretos.

A minha esposa: Sirlei por todo o seu amor que soube compreender a minha ausência nos momentos em que me dedicava a esta tarefa.

Aos amigos e colegas que direta ou indiretamente contribuíram para realização desse trabalho.

## SUMÁRIO

<b>LISTA DE FIGURAS</b> .....	<b>vii</b>
<b>LISTA DE QUADROS</b> .....	<b>viii</b>
<b>LISTA DE ABREVIATURAS</b> .....	<b>ix</b>
<b>RESUMO</b> .....	<b>x</b>
<b>ABSTRACT</b> .....	<b>xi</b>
<b>1 INTRODUÇÃO</b> .....	<b>1</b>
1.1 Motivação .....	2
1.2 Objetivos .....	3
1.2.1 Objetivo geral .....	3
1.2.2 Objetivos específicos .....	3
1.2.3 Estrutura do documento .....	3
<b>2 CONCEITOS BÁSICOS</b> .....	<b>4</b>
2.1 Tecnologia wireless .....	4
2.1.1 Redes locais sem fio .....	4
2.1.1.1 Tecnologias WLANs .....	5
2.1.2 Redes geograficamente distribuídas sem fio.....	8
2.1.2.1 Tecnologia de telefonia móvel celular (tmc) .....	8
2.1.2.2 Tecnologia de satélites .....	9
2.2 Computação móvel .....	11
2.2.1 Arquitetura da computação móvel .....	12
2.2.2 Arquitetura sem infra-estrutura .....	12
2.2.3 Arquitetura com infra-estrutura .....	13
2.2.4 Características .....	15
2.2.4.1 Desconexão e modo doze .....	15
2.2.4.2 Handoff .....	16
2.2.4.3 Portabilidade .....	17
2.2.4.4 Mobilidade (localização de computadores móveis).....	18
2.2.4.5 Gerência de energia.....	19
2.2.4.6 Conectividade .....	20
2.2.4.7 Interface limitada .....	20
2.2.4.8 Capacidade de armazenamento.....	20
2.2.4.9 Aplicações no mundo real.....	21
2.3 Sistemas cliente/servidor .....	22
2.4 Processamento de consultas.....	24

<b>3</b>	<b>GERENCIAMENTO DE CACHE NO CLIENTE</b>	<b>26</b>
3.1	Formas de armazenamento de cache	26
3.1.1	Arquitetura de cache em tuplas	27
3.1.2	Arquitetura de cache em pagina	27
3.1.3	Arquitetura de cache semântico	28
3.2	Invalidação de cache (cache invalidation)	29
3.2.1	Métodos de invalidação	30
3.2.2	Tipos de servidor	30
3.2.3	Formas de disseminação de dados	31
3.2.4	Estratégia de invalidação temporal	31
3.2.4.1	Broadcastin timestamp (TS)	32
3.2.4.2	Adaptive invalidation reports (AIR)	32
3.2.4.3	Ammetic terminals (AT)	33
3.2.4.4	Signatures (SIG)	34
3.2.4.5	Asynchronous and stateful (AS)	34
3.2.4.6	Cache coherence schema with incremental update propagation (CCS-IUP)	34
3.2.4.7	Quasi-copies	35
3.2.4.8	Lazy pull-based	36
3.2.5	Estratégia de invalidação espacial	37
3.2.5.1	Regional endpoints (PE)	37
3.2.5.2	Aproximate cicle (AC)	38
3.2.5.3	Metdo genérico caching-efficiency-based (CEB)	38
3.2.5.4	Bit vector with compression (BVC)	39
3.2.5.5	Grouped bit vector with compression (GBVC)	39
3.2.5.6	Implicit scope information (ISI)	40
3.3	Substituição de cache	41
3.3.1	Tempo	41
3.3.2	Frequência	41
3.3.3	Área válida	42
3.3.4	Distância	42
3.3.5	Distância em áreas urbanas	43
3.3.6	Direção do movimento do cliente	44
<b>4</b>	<b>CONCLUSÃO</b>	<b>45</b>
4.1	Relevância deste trabalho	46
4.2	Trabalhos futuros	47
<b>5</b>	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b>	<b>48</b>

## LISTA DE FIGURAS

FIGURA 1 - ARQUITETURA BÁSICA DE UM SISTEMA CELULAR .....	9
FIGURA 2 - REDE WAN COM SATÉLITE .....	11
FIGURA 3 - REDES AD-HOC .....	13
FIGURA 4 - MODELO DE COMUNICAÇÃO EM REDES MÓVEIS INFRA – ESTRUTURADAS .....	14
FIGURA 5 - REDES WIRELESS COM INFRA-ESTRUTURA .....	15
FIGURA 6 - PROCESSO DE HANDOFF .....	17
FIGURA 7 - ARQUITETURA CLIENTE SERVIDOR WIRELESS .....	24
FIGURA 8 - BROADCASTING TIMESTAMP (BARBARA, IMIELINSKI 1995) ..	32
FIGURA 9 - AMNESIC TERMINALS (BARBARA, IMIELINSKI 1995) .....	33
FIGURA 10 - IUP (CHUNG, CHO, 1998) .....	35
FIGURA 11 - APROXIMATE CICLE (AC) .....	38
FIGURA 12 - METODO GENÉRICO CACHING-EFFICIENCY-BASED (CEB) .....	39
FIGURA 13 - CONSTRUÇÕES E RUAS EM ÁREA URBANA (KIM ET AL, 2002) ..	43
FIGURA 14 - POLITICA FAR (REN, DUNHAM, 2000) .....	44

## LISTA DE QUADROS

QUADRO 1 - COMPARAÇÃO ENTRE AS TECNOLOGIAS .....	7
QUADRO 2 - SISTEMA DE COMUNICAÇÃO VIA SATÉLITE (MATEUS, 1998)..	10
QUADRO 3 - EXEMPLO DE UM ÍNDICE DE CACHE SEMÂNTICO (REN, DUNHAM, 1999) .....	29
QUADRO 4 - CARACTERÍSTICA DE ESTRATÉGIAS DE INVALIDAÇÃO DE CACHE TEMPORAL .....	36
QUADRO 5 - RESUMO DAS CARACTERÍSTICAS EM ESTRATÉGIAS DE INVALIDAÇÃO DE CACHE ESPACIAL .....	40

## LISTA DE ABREVIATURAS

AMPS	- <i>Advanced Mobile Phone Systems</i>
AT	- <i>Amnesic Terminals</i>
CCC	- Centro de Comutação e Controle
CCS-IUP	- <i>Cache Coherency Schema with Incremental Update Propagation</i>
EF	- Estação Fixa
ESM	- Estação de Suporte Móvel
FDD	- <i>Frequency Division Duplexing</i>
GC	- Gerenciador de Conexão
GCCC	- Gerenciador para Coerência de Cache no Cliente
GCCS	- Gerenciador para Coerência de Cache no Servidor
GNI	- Gerenciador de Notificações de Invalidação
INTELSAT	- Internacional Telecommunications Satellite Organization
IR	- <i>Invalidation Report</i>
IU	- Interface com o usuário
LAN	- <i>Local Area Network</i>
LS	- Lista de Solicitações
MH	- <i>Mobile Hosts</i>
MSS	- <i>Mobile Support Stations</i>
NI	- Notificação de Invalidação
NMT	- <i>Nordic Mobile Telephone</i>
PDA	- <i>Personal Digital Assistants</i>
RAM	- <i>Random Access Memory</i>
TACS	- <i>Total Access Communication System</i>
TDD	- <i>Time Division Duplexing</i>
TDMA	- <i>Time Division Multiple Access</i>
TMC	- Telefonia Móvel Celular
TS	- <i>Broadcasting Timestamp</i>
UM	- Unidade Móvel
UMTS	- <i>Universal Mobile Telecommunication System</i>
Wireless	- Sem Fio
Wlan	- Redes Sem fio
LAN	- <i>Wireless Local Area Network</i>

## RESUMO

Com o avanço das tecnologias de redes sem fio e a utilização de equipamentos portáteis tais como palm, celular e notebook, surge um novo paradigma computacional denominado Computação Móvel. Neste ambiente, dispositivos portáteis podem armazenar informações de forma local, visando reduzir o uso da rede e economia de baterias. No entanto, a manutenção da consistência destas informações torna-se uma tarefa complexa, devido às desconexões e mobilidade dos clientes. Assim, o gerenciamento dos dados em ambiente de computação móvel torna-se uma peça chave. Caching é uma importante técnica que armazena dados utilizados com maior frequência nos clientes móveis. Para o gerenciamento dos dados em cache, tem sido proposta uma diversidade de técnicas nas literaturas que enfocam os problemas de invalidação e substituição de cache. Portanto neste trabalho é apresentado um estudo sobre o problema de gerenciamento de cache em ambientes de computação móvel. Nesta monografia apresentamos uma revisão bibliográfica de trabalhos na área de gerenciamento de cache em computação móvel contribuindo com uma relação de conceitos e propostas importantes e indispensáveis para estudos nesta área.

**Palavras-chaves:** Gerenciamento de Cache, Invalidação de Cache, Substituição de Cache, Redes Sem Fio, Computação Móvel.

## **ABSTRACT**

With the advances in wireless network technology and the use of portable equipments such as palm top, cellular and notebook a new nominated mobile computing arises. In this environment, portable device can store information in a local way, aiming at reducing the use of the network and saving batteries. However, the maintenance of that information consistence becomes a complex task due to the clients' disconnections and mobility. Thus, the data management in a mobile computing environment turns into a key piece. Caching is an important technique that stores the most frequently used data in the mobile clients. A lots of techniques which focus on cache invalidation and replacement have been proposed in the literature for the data management. This work presents a study about the cache management problem in mobile computing environment. A bibliographic review in the area of cache management in mobile computing is presented, contributing with a list of concepts and important and indispensable proposals for studies in this area. for this area of study.

**Key words:** Cache management, Cache invalidation, cache substitution, wireless network, mobile computing

## INTRODUÇÃO

O crescente desenvolvimento das tecnologias de redes sem fio e a utilização de equipamentos portáteis tais como palm, celular e notebook, introduz um novo paradigma computacional denominado Computação Móvel.

A mobilidade permite aos usuários deste ambiente enviar e receber informações a qualquer hora e em qualquer lugar, porém uma série de desafios impostos pela mobilidade faz com que soluções tradicionais para o gerenciamento de dados sejam revistas de modo que atendam as novas aplicações.

Entre outras, podemos citar as seguintes características deste ambiente: a largura de banda disponível em redes sem fio normalmente é menor que nas redes cabeadas e possui um alto nível de interferências; desconexões da unidade móvel (UM) nem sempre devem ser consideradas uma falha, podendo ser voluntárias; os equipamentos móveis possuem baterias limitadas (de 2 a 3 horas) e menor capacidade de processamento; alta latência.

Na área de projeto de um software, grandes mudanças ocorrem no levantamento de requisitos, linguagens, interface do usuário, etc. Na área de gerenciamento de dados, a arquitetura cliente/servidor em ambiente de computação móvel permite que um cliente se comunique com o servidor por um link sem fio requerendo serviços ou dados. Normalmente essas solicitações são processadas no servidor e devolvidas ao cliente que a solicitou por meio de conexão sem fio. As dificuldades desta arquitetura na computação móvel estão na limitação do tempo de conexão das unidades móveis (UM), pois pode haver desconexão no meio de um serviço por diferentes motivos, tais como, término de bateria do equipamento ou desconexão por perda de sinal.

Estes são alguns exemplos das dificuldades, todavia, existem benefícios importantes oriundos dessa tecnologia tais como a portabilidade e mobilidade. Estas duas características, aliadas à replicação local de dados, permitem aos usuários um deslocamento físico do dispositivo sob a área de cobertura do serviço de rede, sem que um possível processamento de informações seja interrompido (BERKENBROCK, 2004).

Estes fatos tornam interessante o estudo de métodos para gerência de cache e seus respectivos desempenhos em redes sem fio. Armazenar cópias de itens de dados mais frequentemente acessados em memória no cliente é uma solução adotada para aumentar o desempenho de consultas, a disponibilidade dos dados durante desconexões e reduzir o número de mensagens entre cliente e servidor.

Métodos eficientes de gerência de cache são indispensáveis para garantir que as aplicações em ambientes sem fio tenham uma performance satisfatória. Estas cópias,

denominadas cache, devem ser consistentes em relação aos dados originais do servidor, e quando não houver mais espaço deve-se adotar uma política para substituição dos dados.

## 1.1 Motivação

As características inerentes ao ambiente móvel e sem fio tornam interessantes o armazenamento de dados em cache no cliente para reduzir a quantidade de informações transferidas na rede, assim como a latência de acesso aos dados.

O sistema de cache desempenha um papel fundamental em diversas áreas. Para banco de dados em ambiente de computação móvel, esta tecnologia é essencial por diversas razões. Primeiramente, porque é um meio de aumentar o tempo de resposta das consultas, pois tende a reduzir a média de latência no acesso aos dados no servidor. Em segundo lugar, permite a redução no tráfego dos canais de comunicação sem fio, já que tem a vantagem de trabalhar desconectado com dados em memória, reutilizando informações permitindo um acesso mais rápido aos dados, possibilitando trabalhar com unidades móveis em regiões não cobertas por um sistema de comunicação sem fio, haja visto que podem trabalhar com dados em cache sem precisar estar conectado o tempo todo no servidor.

Além dos benefícios citados anteriormente podemos dizer que se obtém uma economia financeira, seja em equipamentos ou no custo da comunicação sem fio, pois o cliente não necessita ficar o tempo todo conectado ao servidor, dessa forma haverá uma economia significativa com o tempo de conexão.

Em um sistema de computação móvel, é importante não somente armazenar os dados em cache, mas também haver um bom gerenciamento para que a informação desejada esteja disponível. Através de métodos adequados, a tecnologia de cache garante a disponibilidade de dados com eficácia a qualquer hora e em qualquer lugar.

Mesmo com a constante evolução da capacidade de memória das UMs e da qualidade do canal sem fio, o armazenamento em cache continua sendo uma solução interessante para contribuir com o aumento de desempenho no processamento de consultas em banco de dados móveis.

Para o gerenciamento dos dados em cache, tem sido, proposto diversas técnicas nas literaturas que enfocam os problemas de invalidação e substituição de cache. Tais soluções enfatizam um ou mais fatores impostos pela mobilidade dos clientes dificultando uma comparação entre as abordagens propostas. Diante do exposto, percebemos que esta é uma importante área de pesquisa.

## **1.2 Objetivos**

Abaixo, segue uma descrição dos objetivos gerais e específicos pretendidos neste trabalho.

### **1.2.1 Objetivo geral**

Esta monografia tem como objetivo realizar uma pesquisa bibliográfica sobre diferentes abordagens de gerenciamento de cache em um ambiente de computação móvel.

### **1.2.2 Objetivos específicos**

Além dos objetivos gerais apresentados, esta monografia tem os seguintes objetivos específicos:

- Pesquisar estratégias de gerenciamento de cache em dispositivos móveis;
- Contribuir no sentido de introduzir conceitos importantes na área de gerência de cache em banco de dados no ambiente de computação móvel;
- Apresentar as soluções já existentes bem como suas principais características.

### **1.2.3 Estrutura do documento**

Este trabalho está organizado conforme segue:

- O capítulo 2 apresenta conceitos básicos necessários para o entendimento de sistemas de cache em computação móvel. São abordados conceitos sobre redes sem fio, computação móvel, modelo cliente-servidor e processamento de consultas;
- O capítulo 3 aborda o problema da gerência de cache. Neste capítulo são apresentados os métodos para invalidação de cache, as formas do servidor gerenciar o cache nos clientes móveis, formas de disseminação dos dados e políticas de substituição de cache;
- Finalmente, no capítulo 4 são apresentadas as conclusões, problemas encontrados e propostas para trabalhos futuros.

## 2 CONCEITOS BÁSICOS

Neste capítulo são apresentados conceitos básicos sobre redes sem fio e computação móvel, destacando suas principais características, tecnologias e aplicações.

### 2.1 Tecnologia wireless

Wireless (wire=fio, less=sem) é uma tecnologia que permite a conexão entre equipamentos sem uma conexão física direta. Refere-se a todo tipo de conexão efetuada sem fios utilizando algum tipo de tecnologia tais como infravermelho rádio, telefone celular e outros.

Redes sem fio são geralmente utilizadas onde é necessária mobilidade dos pontos da rede ou quando existe dificuldade de cabeamento, por exemplo, quando a arquitetura de um prédio torna difícil (ou impossível) a passagem de cabos de rede. Neste caso, sem o cabeamento físico, a rede torna-se mais flexível, pois uma estação pode ser movida ou retirada com mais facilidade. Em muitas aplicações, redes de comunicação sem fio podem eliminar o alto custo de instalação e manutenção de sistemas cabeados (HOU, 2001).

Segundo (TANENBAUM, 1997), não há nada de novo no conceito de comunicação sem fio digital. Em 1901 o físico italiano Guglielmo Marconi demonstrou como funciona um telégrafo sem fio que transmitia informações de um navio para o litoral por meio de código Morse (afinal de contas, os pontos e traços são binários). Os modernos sistemas sem fios digitais têm um desempenho melhor, mas a idéia básica é a mesma.

O princípio de funcionamento da comunicação sem fio se baseia na transmissão de dados através da camada atmosférica utilizando a propagação das ondas eletromagnéticas como caminho entre o transmissor e o receptor. As redes sem fio utilizam as ondas eletromagnéticas como meio de propagação para fazer a comunicação de dados de uma ponta à outra sem a utilização de cabos. Operam com placas de rede que possuem adaptadores para antenas, através da qual recebem e transmitem informação de um ponto a outro sem necessitar de conexão física.

#### 2.1.1 Redes locais sem fio

As redes locais sem fio (*Wireless LAN* ou *WLAN*) tornaram-se uma alternativa às redes convencionais com fio. Elas fornecem as mesmas funcionalidades das redes cabeadas,

mas de forma flexível, de fácil configuração e com boa conectividade, (BERKENBROCK, 2004).

São geralmente utilizadas onde é necessária mobilidade dos pontos da rede ou quando existem dificuldades de cabeamento, quando, por exemplo, a arquitetura de um prédio torna difícil (ou impossível), a passagem de cabos de rede. Sem o cabeamento físico, a rede torna-se mais flexível, pois uma estação pode ser movida ou retirada com facilidade. A comunicação entre as estações e uma rede *WLAN* pode ser feita através de um *Hub* central, instalado em uma parte alta do ambiente, transmitindo e recebendo dados das antenas que foram instaladas nas estações da rede.

Em redes locais sem fio, os usuários podem acessar informações compartilhadas sem se preocupar com um lugar para se conectar à rede. Redes podem ser gerenciadas pelos administradores sem precisar instalar ou remover cabos.

#### **2.1.1.1 Tecnologias WLANs**

Existem vários tipos de tecnologias de transmissão utilizadas em redes locais sem fio. A seguir são apresentadas resumidamente algumas delas.

- **Infravermelho:** as WLANs baseadas em infravermelho utilizam a mesma tecnologia usada em produtos como telefones sem fio, controles remotos de aparelhos de TV e videocassetes, entre outros. É mais utilizado em ambientes internos como escritórios onde as distâncias entre os pontos de conexão são bem menores em relação às encontradas em ambientes externos. A maior vantagem do infravermelho é sua habilidade de carregar grande largura de banda. Uma desvantagem é devido à facilidade como que ele pode ser obstruído, pois a luz não pode atravessar objetos sólidos e opacos como paredes, sendo que também pode receber interferências de iluminação do ambiente. Quanto à segurança, apresenta fraca ou ausente criptografia nos dados.
- **Radio-microondas:** serviços de transmissão de dados radio-microondas são oferecidos pelas empresas de serviços em telecomunicações para comunicação em altas velocidades. Esse tipo de conexão é interessante quando não há facilidades terrestres na comunicação entre dois pontos. Muitas vezes é impraticável ligarmos dois pontos por fibras ópticas em regiões urbanas que não tenham facilidades. Basicamente é composta de duas antenas e não deve ter

obstáculos no meio. Caso haja, deve-se colocar uma ou mais antenas repetidoras ao longo do caminho. Quanto à segurança, para as ondas de rádio, baixo nível de força na transmissão são utilizados, restringindo o alcance das ondas, obrigando ao espião utilizar caros, sensíveis e difíceis de esconder receptores adjacentes ao prédio.

- Spread Spectrum: a spread Spectrum é atualmente uma das tecnologias de transmissão mais utilizadas em WLANs, por ser considerada menos sensível a interferências do meio que as outras tecnologias. A transmissão com Spread Spectrum atravessa obstáculos com mais facilidade, por utilizar frequências mais baixas, sendo, portanto, mais fáceis de ultrapassar barreiras como paredes.
- Bluetooth: é um protocolo que estabelece comunicações espontâneas entre equipamentos digitais a curta distância, sem auxílio de conexões físicas. Para funcionar, basta aproximar dois aparelhos digitais equipados com chips Bluetooth, e eles se conectam automaticamente, através de frequência de rádio (BRITO, 2001). A idéia é permitir a interoperabilidade desses dispositivos de forma automática e sem que o usuário necessite se preocupar com isso.
- IEEE 802.11 - Wi-Fi: Seguem os padrões técnicos internacionais estabelecidos pelo IEEE (Institute of Electrical and Electronics Engineers). No padrão IEEE 802.11, é especificada a forma de ligação física e de enlace de redes locais sem fio, existem diversas implementações/versões da rede 802.11, como 802.11a, 802.11b, 802.11g, 802.11h, etc. Porém as mais usuais são as redes 802.11b e 802.11g. A arquitetura é baseada no conceito celular, onde o sistema é composto por uma ou mais células, denominadas BSS (Basic Service Set) e são controladas por estações base, denominadas AP (Access Point). Os AP's são interligados através de uma rede física, que é denominada DS (Distribution System), normalmente sendo utilizado a ETHERNET. A tecnologia Wi-fi são destina a Redes locais internas de escritórios e residências, substituindo ou complementando redes que utilizam cabos coaxiais. O nome Wi-fi vencendo muito utilizado em redes públicas de acesso a internet. Um modelo de negócio que vem crescendo é a criação de Hot-Spots, pontos de cobertura indoor de redes 802.11. Estes pontos são instalados em locais de grande concentração de executivos em trânsito, como hotéis, aeroportos, centros de convenções, onde é permitido ao usuário acesso á rede pública INTERNET. Caso o usuário queira

realizar uma conexão com seu escritório, esta normalmente é feita através da criação de uma VPN através da INTERNET.

- Wi-Max, é uma rede sem fio de alto desempenho, com uma taxa de transferência máxima de 75 Mbps, alcance de até 48 Km e permitindo até 1.000 usuários por estação-base. Esta é uma solução para redes metropolitano sendo um ponto realmente fantástico, pois não será necessário passar cabos para atingir os usuários. O padrão WiMAX suporta as features FDD (frequency division duplexing) e TDD (time division duplexing) para permitir a inter-operabilidade com sistemas celulares e outros sistemas sem fio. FDD é um método de duplex legado, que tem sido muito implantado em telefonia móvel Opera em bandas licenciadas e não licenciadas. As bandas não licenciadas são as de 2,4 GHz e 5,8 GHz. Como o padrão WiMAX é projetado para transmissões outdoor em longas distância, a banda de 5,8 GHz é um voto de confiança para o mercado permitindo a operação em uma faixa não poluída pois sabemos que a de 2,4 GHz (também do Wi-Fi) pode sofrer uma série de interferências. As bandas licenciadas são as de 3,5 GHz e 10,5 GHz (com algumas variações) e nos EUA na faixa de 2,5 - 2,7 GHz (Nos EUA, a Nextel e a Sprint são proprietárias de quase 2/3 desta banda). No Brasil, a EMBRATEL possui a banda de 3,5 GHz e a Brasil Telecom as bandas de 3,5 GHz e 10,5 GHz. O quadro 1 mostra uma comparação entre essas tecnologias.

QUADRO 1: COMPARAÇÃO ENTRE AS TECNOLOGIAS

<b>Tecnologia</b>	<b>Frequência HTz</b>	<b>Taxa de Transmissão</b>	<b>Obstáculos</b>	<b>Alcance Máximo</b>
Infravermelho	100HTz	1 a 2 Mbps	Não ultrapassa paredes	400m
Radio-microondas	2.4 GHz	2 a 34 Mbps	Ultrapassa paredes finas	47Km
Spread pectrum	902 a 928MHz, 2.4GHz a 2.5GHz e 5.9GHz	2Mbps e 11Mbps	Ultrapassa obstáculos com maior eficiência que microondas	8km
Bluetooth	2.4 GHz	700Kbps até 2 Mbps	Ultrapassa obstáculos	10m
Wi-Fi	2.4 GHz	11Mbps	Prejudica o sinal vidros blindados e metais	100m
Wi-Max	2.4 GHz até 10.5 GHz	75 Mbps	Não prejudicado	48Km

Fonte: Elaborado pelo Pesquisador

## 2.1.2 Redes geograficamente distribuídas sem fio

Uma rede geograficamente distribuída sem fio (ou *Wireless WAN*) é um novo conceito de rede que reúne os conceitos de *WAN* e de comunicação sem fio, como a telefonia celular, por exemplo. Assim como a *WAN*, a *Wireless WAN* é uma rede que se estende por uma área geograficamente considerável, onde o número de usuários seja grande e esteja distribuída por uma extensa área geográfica, e onde a conexão não depende de um meio concreto para ser estabelecida, do mesmo modo que qualquer tipo de comunicação sem fio.

As definições de *WAN* e *Wireless WAN* se confundem bastante, mas se diferenciam no ponto que diz respeito à tecnologia. Uma *WAN* usa fibra ótica, modem de alta velocidade, fios etc. Uma *Wireless WAN* usa antenas, transmissores e receptores de rádio, e apresenta as seguintes características:

- Pode ser utilizada para a comunicação entre vários prédios;
- É utilizada por grupos de usuários bem definidos;
- Está distribuída por uma extensa área geográfica;
- Apresenta um grande número de usuários;
- As conexões não dependem de um meio concreto para serem estabelecidas.

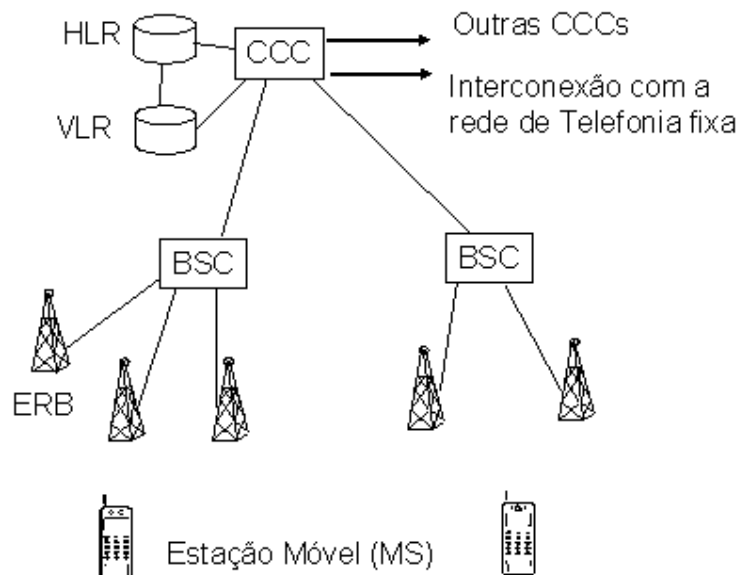
### 2.1.2.1 Tecnologia de telefonia móvel celular (TMC)

Existem várias tecnologias de transmissão empregadas em uma *Wireless WAN*. Uma delas é a tecnologia de Telefonia Móvel Celular (TMC). Uma célula de telefonia móvel contém diversos canais (frequências), onde um ou mais podem ser reservados para a comunicação de dados e compartilhados entre diversos usuários.

Para a implantação da telefonia móvel celular, subdivide-se a área geográfica em pequenas células hexagonais, cada uma dispendo de uma estação de radio-base e de conjuntos de antenas direcionais para supervisão e controle das suas radiofrequências disponíveis.

Estações base são conectadas a um Centro de Comutação e Controle (CCC) de radiofrequência e interligadas com o sistema telefônico convencional, como mostra a figura 1, este por sua vez, é conectado entre si e com a rede pública nacional e internacional de telefonia.

FIGURA – 1: Arquitetura básica de um sistema celular



FONTE: [http://www.teleco.com.br/tutoriais/tutorialcdma/pagina\\_1.asp](http://www.teleco.com.br/tutoriais/tutorialcdma/pagina_1.asp) 25/07/2005

Quando um telefone celular deseja fazer uma chamada envia uma mensagem ao rádio-base. Essa mensagem, após ser processada e aceita pelo terminal de controle, implica a conexão do telefone celular à estação telefônica celular através da concessão de uma radiofrequência disponível.

### 2.1.2.2 Tecnologia de satélites

Em alguns casos pode ocorrer a necessidade de uma tecnologia que prove cobertura em áreas extremamente remotas. Neste caso uma cobertura universal para estes tipos de usuários móveis só é possível através de redes sem fios baseado em satélites. Geralmente quem utiliza satélite são redes públicas, pois o custo da comunicação é bastante elevado. A transmissão do sinal é por antenas, para o satélite, na faixa de frequência conforme o sistema utilizado.

Existem vários tipos de redes satélite que proporcionam comunicação de alturas diferentes da superfície da terra. Eles geralmente podem ser caracterizados por dois modos: baixa comunicação de qualidade de voz, baixa capacidade de transferência (GEOS) tem cobertura extensa, de forma que um número grande de usuários possa ser alcançado de uma só vez, satélite de órbita abaixo da terra (LEO), e satélite de órbita de altura média (MEIOS). O quadro 2 abaixo ilustra alguns exemplos de sistemas de comunicação via satélite.

Os satélites LEO foram os primeiros a serem lançados e apresentam um complexo problema de roteamento dos sinais e rastreamento em terra. Devido às baixas altitudes é

necessário um número mais elevado de unidades para uma maior cobertura, apesar dos equipamentos serem também menores por trabalharem em baixas potências. Os atrasos nos processos de comunicação também são menores.

QUADRO 2 - SISTEMA DE COMUNICAÇÃO VIA SATÉLITE

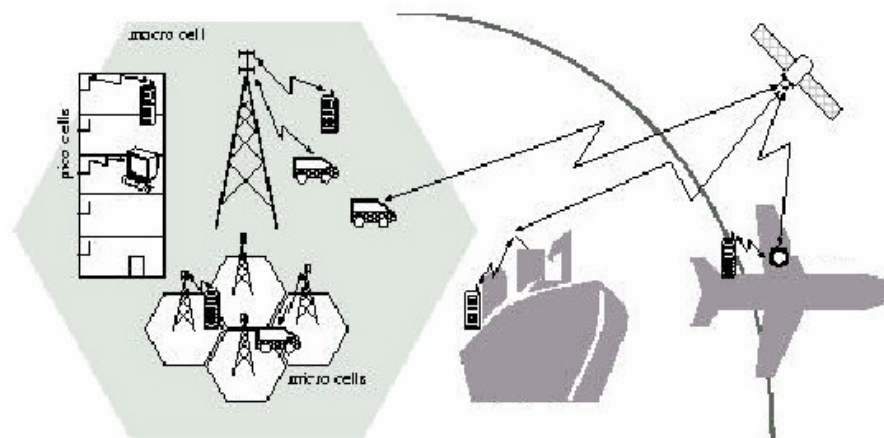
Sistema	Tipo	Altura	Serviços	Custo US\$ bi
Msat	GEO	19.000	Veicular e tel. Fixo	0.55
Globis	GEO	20.000	Tel. Fixo e TV	n.d.
Odissey	MEO	4.600	Voz, Dados	1.3
Ellipso	MEO	4.212	Voz, Dados, Fax	0.7
Archimede	MEO	n.d.	Voz, Dados, Fax	n.d.
Iridium	LEO	413	Voz, Digital, Dados	3.4
Globalstar	LEO	750	Voz, Digital, Dados	1.7
Áries	LEO	550	Voz, Digital, Dados	0.5
Teledisc	LEO	378	Tel.Fixo, ideo	9
Orbcom	LEO	424	Dados	0.5
Starsys	LEO	702	Dados	n.d.
LeoStar	LEO	432	Dados	n.d.
Ecco	LEO	1.100	Voz, Dados, Paging	1.5

FONTE: (MATEUS, 1998)

A segunda geração são os satélites GEO que movimentam sincronamente com a terra, mantendo a mesma posição em relação à linha do equador. Isto permite manter as estações terrestres em posições fixas. O primeiro satélite GEO foi lançado pela INTELSAT (Internacional Telecommunications Satellite Organization) em 1965 e, a partir daí, passaram a predominar. Com o sincronismo os problemas de roteamento e rastreamento são reduzidos.

Aumentando a altitude também se reduz o número de unidades para uma maior cobertura. Uma unidade, com antena não direcionada pode cobrir até 30% da superfície terrestre, bastando três satélites distanciados a 120 graus para uma ampla cobertura. Mas, a proximidade à linha do equador deixa algumas regiões polares sombreadas. Também se eleva à dimensão dos equipamentos pelo uso de grandes potências, reduz-se a portabilidade e dificulta atendimentos de massa. Outra característica importante são os atrasos na aproximadamente 120 ms, portanto 240 ms de ida e volta. Envolvendo mais de um satélite, esse atraso aproxima de 1s, o que inviabiliza muitos serviços. A figura 2 mostra uma rede WAN com satélite.

FIGURA – 2: Rede Wan com satélite



FONTE: (FASBENDER, HOFF e PIETSCHMANN, 1995)

## 2.2 Computação móvel

Computação móvel representa um novo paradigma computacional. Surge como uma quarta revolução na computação, antecedida pelos grandes centros de processamento de dados da década de sessenta, o surgimento dos terminais nos anos setenta, e as redes de computadores na década de oitenta. O novo paradigma permite que usuários desse ambiente tenham acesso a serviços independente de onde estão localizados, e mais importante, de mudanças de conceito de computação distribuída. Isso é possível graças à comunicação sem fio que elimina a necessidade do usuário manter-se conectado a uma infra-estrutura fixa e, em geral, estática (MATEUS, 1998).

A comunicação sem fio é um suporte para a computação móvel, que, portanto pode ser vista como uma área da comunicação sem fio. A combinação de comunicação sem fio com a mobilidade de computadores criou problemas novos na área de informática e telecomunicações, em especial redes de computadores, sistemas operacionais, otimização, sistema de informação, banco de dados dentre outras.

A questão principal na computação móvel é a mobilidade que introduz restrições inexistentes na computação tradicional formada por computadores estáticos. Logo, o objetivo principal da computação móvel é prover para os usuários um ambiente computacional com um conjunto de serviços comparáveis aos existentes num sistema distribuído de computadores estáticos que permita a mobilidade.

### 2.2.1 Arquitetura da computação móvel

Genericamente, um sistema de computação móvel pode ser definido como um conjunto de computadores móveis e computadores fixos. Os computadores móveis se comunicam uns com os outros através de uma interface de comunicação sem fio, enquanto que os computadores fixos se comunicam com os outros através de rede cabeada (BERKENBROCK, RIBEIRO, 2004).

Existem dois modelos possíveis para o ambiente móvel: as redes móveis com infraestrutura e as redes sem infra-estruturas (denominada *rede ad hoc*).

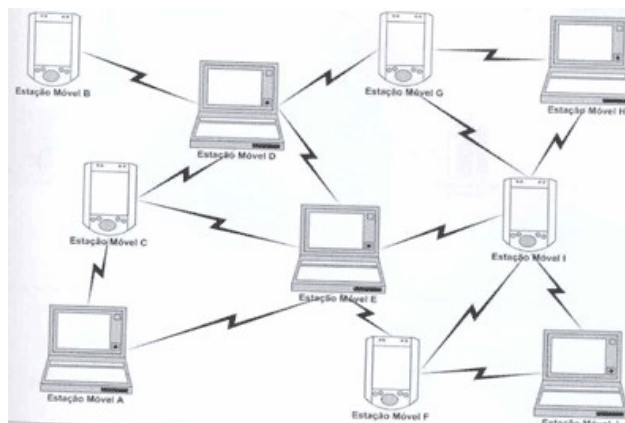
### 2.2.2 Arquitetura sem infra-estrutura

Nesta topologia não existe uma infra-estrutura fixa e todos os componentes são móveis. Uma estação móvel pode comunicar-se com a outra diretamente dentro do seu raio de atuação, ou pode usar uma outra como repetidora do sinal.

Caso exista alguma estação fixa, esta é tratada como móvel. Todos os componentes comunicam-se através das unidades móveis participantes. Se uma estação deseja comunicar-se com outras que não estão ao seu alcance, esta irá utilizar-se de outras estações móveis como roteador para que suas mensagens sejam encaminhadas, dispensando assim qualquer infra-estrutura fixa. Este tipo de rede, sem infra-estrutura é chamado também de rede *ad hoc*.

Segundo (SILVA, 2003), o grande desafio neste ambiente é encontrar uma solução ótima para o problema de roteamento. Como todas as unidades móveis conectam-se aos seus vizinho e todos são repetidores dos sinais, a unidade móvel A fala diretamente com a unidade móvel C e com a unidade móvel E, conforme a figura 3. Caso seja necessário trocar informações entre as unidades móveis A e B, será necessário usar C ou E como seu retransmissor. C por sua vez necessitará de um algoritmo eficiente para encontrar B e assim por diante.

FIGURA - 3: Redes Ad-Hoc



FONTE: (SILVA, 2003)

As redes ad hoc são indicadas para os casos onde a instalação de uma infra-estrutura é inviável ou impossível. Podem ser usadas por um exército em movimento no campo de batalha inimigo, onde uma infra-estrutura não existe, ou em casos de catástrofes onde não se tem tempo hábil para a instalação de uma infra-estrutura.

Neste ambiente as taxas de erros são elevadas, desconexões são freqüentes e a largura de banda é menor que as encontradas na rede fixa. Apesar disto esta arquitetura possui inúmeras vantagens, como:

Não dependem de infra-estrutura, entram em operação assim que as unidades móveis iniciam suas atividades, e sua conectividade é maior, desde que os elementos estejam dentro dos limites de alcance e a rede configura-se de maneira dinâmica, aumentando assim a tolerância a falhas, existindo maior mobilidade entre as unidades móveis.

### 2.2.3 Arquitetura com infra-estrutura

O funcionamento deste tipo de rede móvel é semelhante ao da telefonia celular, onde toda a comunicação deve necessariamente passar pela central, mesmo que os telefones estejam a uma distância em que poderiam, eventualmente, comunicar-se diretamente.

Redes celulares são compostas por uma rede fixa (por exemplo, um LAN), onde podem estar localizados servidores de arquivos; e por várias células (cell`s), as quais possuem estações de controle, também chamadas de estações de Suporte Móveis (*MóBILE Support*

*Stations - MSS*) e por clientes móveis – *MH Mobile Hosts* - que podem ser *notebook's*, *palmtop's*, etc..

Cada MSS coordena a comunicação entre os *MH's*. Esta comunicação pode ocorrer tanto entre *hosts* que estão contidos em uma mesma célula, como entre *hosts* em células diferentes e assim como entre um *host* móvel em uma célula e uma estação fixa (*MSS*), ou um *host* fixo.

A figura 4 descreve graficamente esta arquitetura. Uma célula é uma área geográfica que é coberta pela MSS, Isto é, qualquer comunicação entre *MH's* que estejam dentro da área de uma MSS vai ocorrer a partir da mesma. Assim, cada MSS e *MH's* dentro de uma célula formam uma “rede sem fio” (*wireless network*).

As redes sem fio possuem um tamanho de banda (*bandwidth*) restrito. Sendo que redes celulares chegam a 10bps e as LAN's sem fio chegam a 10Mbps. Como os *MH's* podem se mover, isto é, possuem uma característica de mobilidade, a localização dos mesmos pode mudar no decorrer do tempo, e isto causa uma reconfiguração dinâmica da rede.

Na figura 5 temos um outro exemplo onde as unidades móveis comunicam-se através de pontos de acessos instalados. Assim uma unidade móvel comunica-se com a outra através de uma unidade móvel intermediária. A rede fixa é responsável pelo suporte às unidades móveis, através de uma estação de suporte à mobilidade, executando as tarefas de localização da unidade móvel, de roteamento, redução de trafego, adaptabilidade, etc. Estas redes podem utilizar os padrões de comunicação da família IEEE 802.11, em conjuntos com outros padrões como o IEEE 802.3, sinais de satélite, etc.

FIGURA - 4: Modelo de comunicação em redes móveis infra-estruturadas

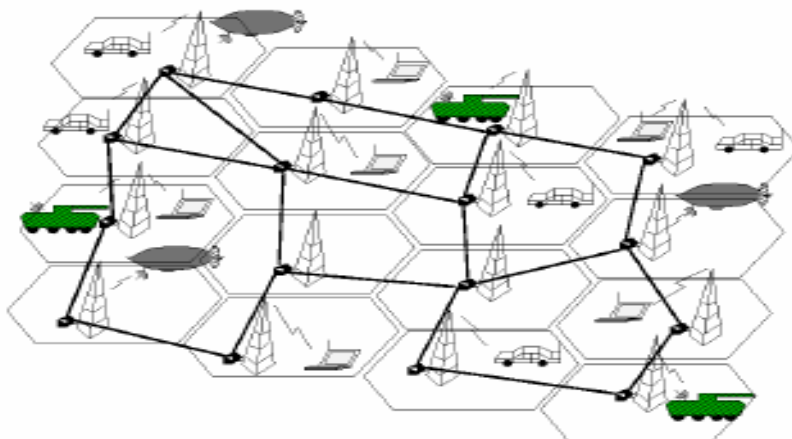
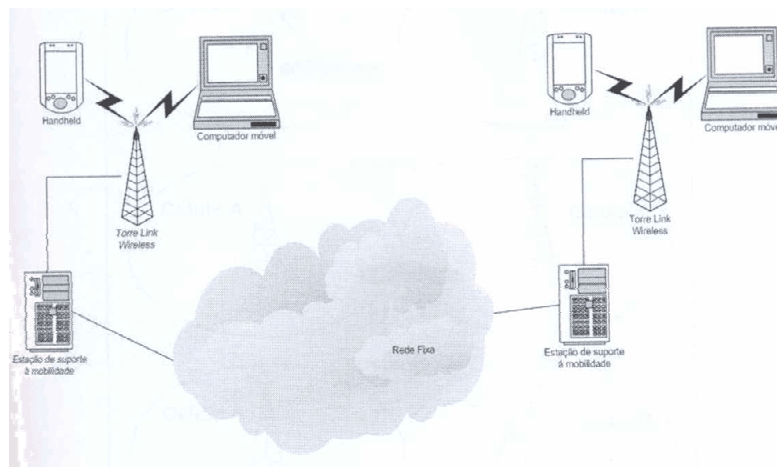


FIGURA - 5: Redes wireless com infra-estrutura



FONTE: (SILVA, 2003)

## 2.2.4 Características

Dentro da área de computação móvel existem vários tópicos a serem estudados, sendo que os mesmos caracterizam os ambientes móveis. A seguir são comentadas algumas características peculiares deste ambiente.

### 2.2.4.1 Desconexão e modo doze

Terminais móveis freqüentemente são desconectados da rede, não podendo enviar nem receber mensagens. A desconexão pode ocorrer por vários motivos como: a bateria que estava fornecendo força elétrica ao terminal necessitou ser recarregada, a conexão pode ter sido perdida quando o terminal, que estava em movimento, cruzou a fronteira de comunicação de uma MSS e não foi possível encontrar uma nova célula para continuar a comunicação, etc.

Desconsiderando um estado de desconexão, um host móvel pode estar em vários graus de conexão, que vão desde fracamente conectado (*weak connection*), através de um canal de rádio de tamanho de banda pequeno, até fortemente conectado (*strong connection*), que é obtido através de uma conexão com uma rede fixa.

As desconexões em ambientes móveis não são consideradas uma falha, pois a mesma é "eletiva" por natureza (BADRINATH, et al., 1993), sendo assim, um host móvel pode

informar quando uma desconexão irá acontecer, e deste modo, um protocolo de desconexão pode ser executado.

Outro modo de operação dos host móveis é o modo "soneca" (*Doze mode*). Neste modo de operação a velocidade de clock é reduzida e nenhuma computação de usuário é realizada, o host móvel simplesmente aguarda passivamente o recebimento de alguma mensagem enviada do restante da rede através de um canal de rádio de tamanho de banda pequeno. Se alguma mensagem é recebida, o host móvel volta ao modo de operação normal. Este modo de operação tem como um de seus objetivos a redução do consumo de energia.

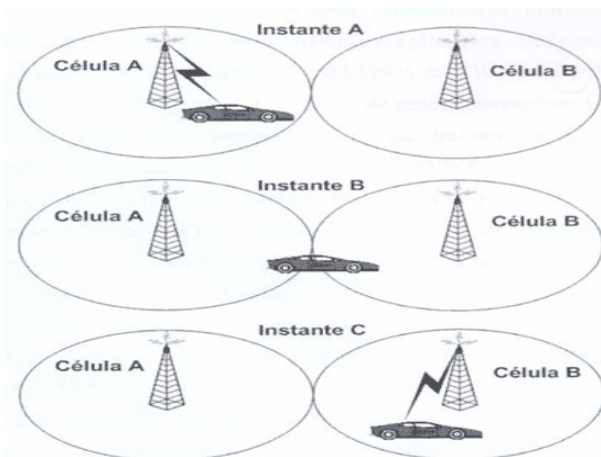
Como no caso de desconexão, esta é uma operação voluntária, mas com diferenças cruciais. Quando em desconexão, o host móvel não está alcançável dentro da rede e só pode se conectar novamente por "vontade própria", isto é, o restante da rede não tem como enviar uma mensagem para um host móvel até o mesmo estar conectado novamente, e esta é uma ação realizada única e exclusivamente pelo host móvel. Já no caso do modo "soneca", o host móvel pode continuar a receber mensagens normalmente.

#### **2.2.4.2 Handoff**

As unidades móveis em uma rede com infra-estrutura captam o sinal de uma torre. Esta torre possui uma área de abrangência denominada célula. Quando a unidade móvel se movimenta de uma célula para a outra, executando aplicações, tem-se com este movimento um processo chamado *handoff*.

*Handoff* é um processo entre duas estações bases adjacentes que tem por objetivo garantir uma transação enquanto o usuário se movimenta de uma célula para outra mantendo a conectividade de uma unidade móvel com a rede fixa. Para (PITOURA, SAMARAS, 1998), o processo de *handoff* é totalmente transparente para o usuário, conforme ilustrado na figura 6.

Uma unidade móvel localizada na célula A, encaminha-se para a célula B, no instante A. Existe um momento no instante B em que a unidade móvel está na fronteira entre as duas células, ocorrendo desconexão da célula A para após ocorrer à conexão com a célula B. No instante C, a unidade móvel já está conectada a célula B. No instante B ocorre o *handoff*.

FIGURA - 6: Processo de *handoff*

FONTE: (SILVA, 2003)

### 2.2.4.3 Portabilidade

Computadores portáteis são capazes de substituir agendas eletrônicas, calculadoras, além de poderem ser integrados a telefones celulares, (BERKENBROCK, 2004). À medida que a tecnologia avança esses dispositivos estão se tornando cada vez menores. Segundo (CUNHA, 2003), junto com essa mobilidade também se encontra algumas dificuldades de projeto no que se referencia a energia, interface e capacidade de armazenamento.

As características e limitações dos dispositivos móveis quanto à portabilidade, de acordo com (PITOURA, SAMARAS, 1998), são descritas a seguir:

- Em se comparando com dispositivos estáticos podemos observar que nos dispositivos móveis encontramos poucos recursos, interface é limitada, a tela é pequena, ou tem um buffer pequeno; tem pouca memória RAM (Random Acces Memory), processadores tem um poder de processamento menor que os dispositivos estáticos, e sua memória física é pequena; os dispositivos de entrada de dados também são pequenos e limitados (por exemplo, teclados de aparelhos celulares);
- Os elementos móveis contam com pouca capacidade em suas baterias estes elementos dependem da energia fornecida por baterias; normalmente as baterias possuem capacidades limitadas; em alguns lugares existe dificuldade para recarga da bateria;

- Elementos móveis são poucos robustos. Desta forma, podem ser danificados, perdidos ou roubados facilmente (BERKENBROCK, 2004).

#### **2.2.4.4 Mobilidade (Localização de computadores móveis)**

Para (MATEUS, LUREIRO, 1998), a localização de um elemento móvel e, conseqüentemente seu ponto de acesso a rede fixa muda à medida que esse elemento se move pela rede. Conseqüentemente com essa mobilidade existem problemas relacionados com o gerenciamento, localização, projeto de protocolo e algoritmos, heterogeneidade, segurança, risco de perda de dados e confiabilidade de conexão com a rede.

Na gerência de localização o custo de pesquisa para localização de um elemento móvel deve incluir o custo da comunicação (MATEUS, LUREIRO, 1998). Para reduzir o custo final, estruturas de dados, algoritmos e planos de consultas eficientes devem ser projetados e desenvolvidos para representar, gerenciar e consultar a localização dos dispositivos móveis, uma vez que dados de localização podem ser alterados de forma muito rápida (BERKENBROCK, 2004).

No desenvolvimento de protocolos e algoritmos distribuídos para ambientes móveis a configuração dos sistemas não é estática. Os algoritmos para processamento distribuídos tradicionais precisam ser reprojatados devido à ausência de uma topologia fixa de rede com os elementos móveis e a distribuição da carga do sistema a noção de localidade muda ao longo do tempo. De acordo com (MATEUS, LUREIRO, 1998), esses fatores não podem ser desprezados e na verdade, um dos grandes desafios da computação móvel é planejar novas aplicações e algoritmos que levem em consideração essa característica do ambiente.

Para (MATEUS, LUREIRO, 1998), a heterogeneidade é uma constante na comunicação móvel. O desempenho e a confiabilidade de um elemento móvel varia conforme sua conectividade (PITOURA, SAMARAS, 19998). Em ambientes externos (outdoors) a velocidade de comunicação em geral é mais baixa que em ambiente interno (intdoors) onde pode se oferecer uma conectividade mais confiável ao dispositivo móvel ou até mesmo permitir que seja operado através de conexão com redes fixas. O número de dispositivos móveis numa célula também pode mudar com o tempo, e conseqüentemente a carga, a estação base e a largura de banda disponível. Segundo (BERKENBROCK, 2004) os recursos disponíveis nos elementos móveis variam, como por exemplo, a sua capacidade de memória, o tamanho de sua tela e o tempo de direção.

A segurança e autenticação também é um problema introduzido pela mobilidade. As interceptações de mensagens se tornam mais fácil de ser feito na comunicação sem fio, podendo causar sérios problemas de segurança, por isso deve-se fazer o uso de técnicas de criptografia. O rastreamento do dispositivo móvel também é fácil de se fazer quando esse se comunica com uma rede fixa, o que nem sempre é bom para o usuário se o sigilo do movimento for importante.

#### **2.2.4.5 Gerência de energia**

Segundo (BERKENBROCK, 2004) devido às restrições do poder da bateria dos dispositivos móveis, estes podem estar freqüentemente desconectados. Este problema é visto como o maior empecilho no uso de computadores móveis. Logo o gerenciamento de energia é um problema importante e deve ser tratado tanto pelo Hardware quanto pelo Software (MATUES, LOUREIRO, 1998).

Estão sendo desenvolvidos vários estudos, com a finalidade de diminuir as restrições imposta pelo gerenciamento de energia. Podemos destacar estudos realizados por (FLIN, SATYANARAYANAN, 1999), que apresenta uma ferramenta para mapear o uso de energia pelas aplicações. Através desta ferramenta podemos determinar qual a fração de energia consumida por um processo em um período de tempo, com isso identificação quais componentes do sistema está consumindo mais energia. Estudos sobre o comportamento intrínseco das baterias estão sendo realizados por (CHIASSERINI, NUGGEHALI e SRINIVASAN, 2002) utilizando esse conhecimento no desenvolvimento de um modelo estocástico de bateria e em novos protocolos energy-efficient.

Projetos de Hardware para computadores móveis já incorporam algumas características nesses sistemas para diminuir o consumo de energia como desligar luz de fundo da tela, o disco quando não está sendo usado ou mesmo eliminá-lo completamente e substituí-lo por uma memória flash e projetar processador que consume menos energia no modo doze.

#### **2.2.4.6 Conectividade**

Para (BERKENBROCK, 2004), clientes móveis podem não estar conectados a redes continuamente. Um elemento móvel pode se desconectar de uma célula para se conectar à outra. Esta fraca conectividade causa várias limitações que não estão presentes quando a conectividade é forte.

O fato da comunicação entre uma unidade móvel e uma estação de suporte móvel (ESM) ser através de uma rede de comunicação sem fio, tem como principais consequências à conectividade fraca e intermitente. As redes sem fio são monetariamente mais caras e oferecem menores larguras de banda. Em adição, elas possuem uma latência maior e são menos confiáveis do que as redes cabeadas (BERKENBROCK, 2004).

Em uma rede sem fio a conectividade dos elementos móveis pode variar. Sendo assim, a qualidade da conexão pode variar em função de interferências na comunicação, da distância da unidade móvel com a ESM na qual a unidade está conectada, e ao compartilhamento da ESM por várias unidades móveis.

Segundo (YUEN, 2000), freqüentes desconexões e a mobilidade dos clientes introduzem o problema de incoerência do cache com relação aos dados do servidor.

#### **2.2.4.7 Interface limitada**

Alguns dispositivos móveis possuem visores muito pequenos como, por exemplo, PDAs (Personal Digital Assistants), que dificulta a leitura dos dados implicando em maior tempo de conexão e conseqüente mente em aumento do gasto de energia (BERKENBROCK, 2004).

De acordo com (BARBARA, 1999) o tamanho da tela em dispositivos móveis tem forçado a realização de pesquisas sobre novas interfaces para retorno de informações. Esta característica pode representar limitações na memória e poder de bateria.

#### **2.2.4.8 Capacidade de armazenamento**

Segundo os autores (HUANG, SISTLA, WOLFSON, 1994) e (CHAN, RODDICK, 1999), a capacidade de armazenamento de informações é um recurso muito importante nos dispositivos móveis. Devido a limitação de armazenamento existente em alguns tipos de

dispositivos móveis, torna-se difícil criar réplicas de banco de dados volumosos, particularmente em dispositivos como PDAs (BERKENBROCK, 2004).

A capacidade de armazenamento é um importante parâmetro de projeto em aplicações móveis. Porém, de acordo com (YE, JACOBSEN, KATZ, 1998), o tamanho do cache não possui influência significativa na sua performance.

#### **2.2.4.9 Aplicações no mundo real**

Novas aplicações são desenvolvidas para atender as diversas necessidades das pessoas. A seguir apresentam-se algumas aplicações da computação móvel:

- **Serviços de emergência:** os serviços de emergências podem fazer uso de aplicativos para prestar os primeiros atendimentos as pessoas acidentadas, coletando informações que posteriormente serão utilizadas em um hospital. Pode se ainda usar a computação móvel em casos de desastres, como terremotos ou enchentes, coletando informações.
- **Agentes de seguros:** os agentes de seguros podem elaborar uma apólice de seguro no mesmo instante da visita ao seu cliente. Pode-se ainda, fazer o termo de vistoria, imprimir o contrato. A vigência da apólice seria imediatamente, visto que todos os dados estariam nas centrais das companhias no mesmo instante.
- **Automação de vendas:** este talvez seria o apelo mais clássico. Profissionais de vendas podem fazer uso de PDAs para automatizar as suas vendas. Os pedidos seria digitados pelos vendedores de forma mais eficiente. Podem-se consultar os estoques, verificar os prazo de entrega, consultar a ficha financeira do cliente entre outro. Os aplicativos móveis podem estar diretamente ligados ao software de gestão da empresa, tendo assim uma extensão da infra-estrutura de informática;
- **Ambientes de guerra:** em campos de batalhas podem-se coletar informações para o comando geral. Estas informações serviriam para a construção de novas estratégias.
- **Transportadoras:** os funcionários de uma transportadora poderiam utilizar-se dos aplicativos em um PDAs para a emissão do conhecimento de carga no ato de uma coleta. Podem-se ainda programar as coletas das equipes de campo em tempo real. Toda chamada a central seria remetida à equipe mais próxima.

- Banco: os bancos poderão desenvolver aplicações para seus clientes. Os clientes podem ter em suas mãos todos os detalhes de sua conta e suas movimentações. O cliente terá uma extensão de um terminal do banco em seu equipamento móvel. Poderá ter diretamente da instituição financeiras extratos, saldos, serviços de pagamentos de contas, cotação de moedas, compra e venda de ações, fundos de investimentos entre outros.
- Profissional móvel: médicos, técnicos, consultores, etc. podem obter acessos a suas informações em casa ou escritório, informações sobre empresas, clientes, produtos, elaborar relatórios em tempo real, obter informações corporativas, etc.

Estas são algumas potencialidades das aplicações em dispositivos móveis. Encontrar aplicações não será algo difícil no futuro. A convergência entre os telefones celulares, Internet e dispositivos móveis implicará em um negócio com mais de 1 bilhão de usuários. Existirão mais usuários utilizando dispositivos móveis que computadores fixos (TAURION, 2002).

### **2.3 Sistemas cliente/servidor**

Segundo (OZSUS, VALDURIEZ, 2001), os SGBDs cliente/servidor entraram no cenário da computação no início da década de 1990. Este modelo apresenta uma solução simples e elegante: distinguir a funcionalidade que precisa ser fornecida e dividir essas funcionalidades em duas classes, funções do cliente e funções do servidor. Com a divisão em dois níveis torna-se mais simples o gerenciamento.

Para (MELO et al, 1997), “as aplicações com arquitetura cliente/servidor representam um caso especial de processamento distribuído, que implica a cooperação entre dois ou mais processos”. Aplicações deste tipo são caracterizadas pelo alto grau de fragmentações da aplicação. A interação é realizada através de pedidos dos clientes e pelas respostas dos servidores.

Para (OZSU, VALDURIEZ, 2001), no modelo cliente/servidor, o servidor é responsável pela gerência dos dados. Isto significa que o gerenciamento de transações, gerenciamento de armazenamento, processamento de consultas e otimização fica por conta do servidor. O cliente além da aplicação do usuário possui um módulo cliente do SGBD responsável pelo gerenciamento dos dados que são alocados no cache do cliente. Em alguns casos o módulo cliente também é responsável pelo bloqueio de transações. Pode-se ainda

colocar a verificação da consistência das consultas no lado cliente, porém exigiria a replicação do catálogo do sistema nas máquinas clientes.

(OZSU, VALDURIEZ, 2001) afirma que existem várias arquiteturas para o modelo cliente/servidor. O mais simples é a arquitetura de apenas um servidor e vários clientes denominada “vários clientes – servidor único”. Outra arquitetura, mais sofisticada, denomina-se “vários clientes - vários servidores”. Nesta abordagem o cliente gerencia a sua própria conexão com o servidor ou cada cliente conhece apenas o seu “servidor local”. A primeira abordagem, “vários clientes – servidor único”, implica em clientes pesados, isto é, os clientes possuem responsabilidades adicionais. Por outro lado, a segunda abordagem, “vários clientes - vários servidores”, implica em clientes leves, isto é, com menor número de tarefas a serem executadas.

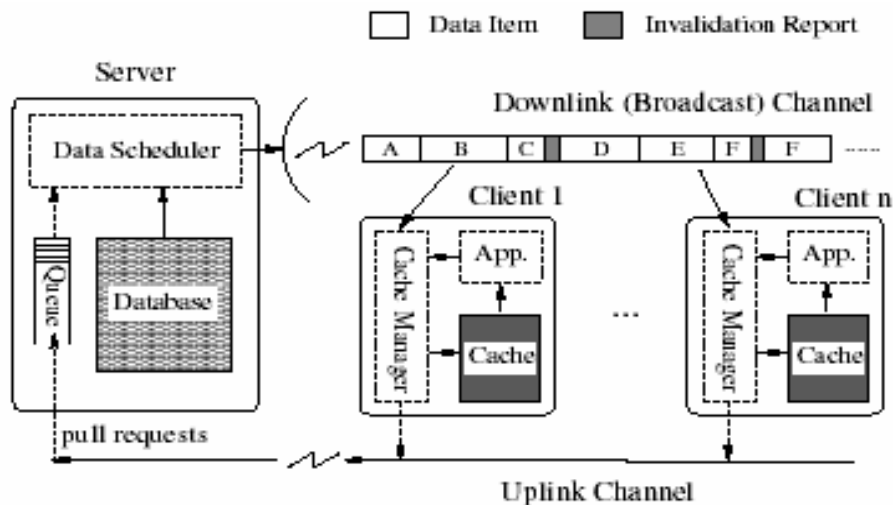
De qualquer forma, neste ambientes os usuários possuem total transparência da estrutura, aparentando assim um sistema de banco de dados logicamente único, não importando a estrutura física.

Existem diferentes formas para o envio de informações do servidor para os clientes. Para ambiente de computação móvel, a técnica de difusão de dados (*broadcast*) é a mais apropriada pelo fato de economizar energia dos computadores móveis por evitar utilizar o canal de subida (*uplink*). No método de difusão, o servidor envia os dados pelo canal de descida (*downlink*). Os clientes monitoram este canal e filtram os dados desejados.

Uma outra forma de envio é *sob demanda* na qual os clientes solicitam informações ao servidor que responde enviando os dados solicitados. Na arquitetura cliente-servidor para um ambiente de computação móvel, é comum a utilização de uma fusão dos dois métodos de difusão e sob-demanda.

A figura 7 descreve uma arquitetura genérica de transmissão de dados em uma rede sem fio. Os clientes enviam uma solicitação para o servidor através de um canal *uplink* (sentido cliente-servidor). De acordo com os autores (ACHARYA, MUTTHUKRISHMAN, 1998), (AKSOY, FRANKLIN, 1999) E (AKSOY,FRANKLIN e ZDONIK, 1998) o servidor transmite a solicitação dos itens de dados para os clientes através de um canal *downlink* (sentido servidor-cliente) de broadcast baseado em um algoritmo programado. Os clientes recuperam os itens de seu interesse monitorando o canal de broadcast.

FIGURA – 7: Arquitetura cliente servidor wireless



FONTE (XU, 2002)

## 2.4 Processamento de consultas

Em sistemas de banco de dados móveis, o processamento de consultas é significativamente diferente de sistemas fixo porque há muitos elementos envolvidos. De acordo com (ALONSO, KORTH, 1993), os otimizadores de consultas no âmbito de mobilidade devem lidar com aspectos extras não considerados em ambientes tradicionais tais como: otimizar a consulta para que a energia consumida pelos dispositivos móveis para realizar consultas seja o menor possível; e determinar os custos de comunicação entre a estação que está consultando e a estação onde os dados residem. Portanto, a consulta deve ser executada utilizando técnicas diferentes que devem considerar as características dos sistemas móveis.

Através das redes sem fio e um computador móvel é possível se consultar bancos de dados em máquinas remotas. Estes bancos de dados podem conter dados referentes a localização (ex. hotéis, restaurantes, hospitais, etc.) chamadas consultas dependentes de localização, ou dados sobre clientes de empresas, etc.. Consultas ao posto de gasolina mais próximo, hotéis, restaurantes, hospitais e outros tipos de serviços necessários enquanto se esta em movimento, são exemplos de consultas móveis.

Os novos aspectos impostos pelo ambiente de computação móvel são divididos segundo (ACHARYA, 1998) em três classes: portabilidade, mobilidade e comunicação sem

fio. Portabilidade de computadores móveis adiciona desafios especiais no desenvolvimento de componentes de hardware. Mobilidade dos usuários faz necessária a localização correta dos usuários móveis. O uso das redes sem fio resulta em novos problemas como desconexões frequentes e variações de largura de banda.

Diferentes trabalhos na área de processamento de consultas têm sido propostos. Para todos os modelos apresentados são consideradas influências externas e as soluções propostas normalmente enfatizam um ou mais dos fatores anteriormente apresentados.

Dentre os aspectos de processamento de consulta nos ambientes móveis, um tipo de consulta importante que vem se destacando é aquela baseada no contexto (localização, direção, velocidade, etc.) em que a UM (Unidade Móvel) está inserida. Desta maneira, os parâmetros para uma consulta podem expressar relativamente a atual localização do computador móvel para o servidor onde a consulta será realizada. Estes parâmetros também podem depender da direção e velocidade com que o UM está se movimentando.

Outro aspecto é que os dados dependentes de localização podem mudar durante o processamento da consulta. Estes tipos de consultas também são denominados "consultas dependentes de localização". Por exemplo, a consulta "quais os nomes e endereços dos restaurantes Chineses nesta cidade?", retorna diferentes valores dependendo em qual cidade está localizado o computador móvel.

A dependência de localização pode ser mais sofisticada, como por exemplo, a consulta "onde está o desvio para o Banco X dentro de 1 quilômetro a partir da minha localização?" ser realizada enquanto estiver no interior de um carro. O sistema tem que observar a trajetória de movimento atual do carro para que possa descobrir sua localização e responder a consulta (ALONSO, KORTH, 1993).

Após o processamento de consultas, a UM deve armazenar seus resultados. Assim, caso por algum motivo ela fique sem acesso a rede, poderá executar consultas locais aos dados disponíveis em cache.

### 3 GERENCIAMENTO DE CACHE NO CLIENTE

Conforme (ZHENG, XU e LEE, 2002), para melhorar a performance das consultas e disponibilidade dos dados durante desconexões, uma boa solução tem sido armazenar cópias de dados nos clientes.

Em um modelo de cache para banco de dados, no cliente se encontram as aplicações que solicitam dados, o conteúdo em cache que são cópias dos dados armazenadas (na memória ou em disco) em determinado momento, e o gerenciador de cache.

O componente gerenciador de cache é responsável pelo gerenciamento e substituição de cache. O gerenciador de cache deve alocar os itens que devem ser armazenados, manter um diretório sobre o, conteúdo do cache, prover acesso aos dados para o processamento de consultas, saber (juntamente com o servidor) sobre a atualização dos dados e decidir qual dado deve excluir quando não houver mais espaço e um novo dado precisa ser armazenado em cache.

Existem duas áreas de pesquisas em gerenciamento de cache:

- Invalidação de Cache (*cache invalidation*): que visa manter a consistência dos dados entre o cliente e o servidor;
- Política de Substituição de Cache (*cache replacement policy*): que determina quais dados devem ser deletados do cache quando este não possuir mais espaço para acomodar um novo item.

A seguir, primeiramente apresentamos algumas formas de armazenamento em cache e posteriormente, as técnicas propostas para gerência de cache bem como suas características. Uma taxonomia sobre técnicas de gerência de cache é proposta em (MANICA, CAMARGO, 2004).

#### 3.1 Formas de armazenamento em cache

O objetivo de um sistema de cache é armazenar dados para serem reaproveitados posteriormente. Existem várias formas de armazenamento físico dos dados em cache. As seções seguintes descrevem os principais modelos.

### 3.1.1 Arquitetura de cache em tuplas

No modelo de cache em tupla, o cache do cliente é organizado e mantido em nível de tuplas (ou objetos) individuais ao invés de páginas inteiras. Este modelo tem alta granularidade e permite a flexibilidade máxima em ajustar índices do cache às propriedades local de acesso das aplicações (WITT et al., 1990).

Este modelo pode afetar o desempenho do sistema devido ao grande número de mensagens pequenas enviadas entre cliente e o servidor. A granularidade tem grande influência no funcionamento do cache. A definição da unidade básica depende do tipo de flexibilidade que se deseja do cache.

Por exemplo, um cache com granularidade grande (por exemplo, páginas) limita a reutilização de dados, porém as informações de controle ficam reduzidas. Já em um cache com granulosidade pequena (por exemplo, tuplas) exige um maior número de informações para controle, porém existe um ganho na flexibilidade de gerenciamento que pode ser feito tupla a tupla. Segundo (Dar et al., 1996), é importante definir uma forma de se balancear o tamanho dos grupos.

### 3.1.2 Arquitetura de cache em página

Para melhorar o gerenciamento dos dados em cache, os dados podem ser agrupados para formar unidades de armazenamento maiores. No modelo de cache em páginas (*Page Caching*) os dados são armazenados em grupo de tuplas denominadas páginas. O agrupamento diminui o *overhead* causado pelas informações extras anexadas aos itens de dados que são utilizadas no gerenciamento do cache.

A unidade de transferência entre servidor e cliente é uma página, que são unidades físicas de tamanho fixos, estáticos e determinados a priori. As consultas são executadas localmente no cliente. Se uma página necessária para executar a consulta não estiver em cache, ela é solicitada ao servidor.

Em resposta para tal pedido, o servidor obterá a página de disco (se necessário) e devolve ao cliente. No lado do cliente, página em cache é mantida através de um mecanismo que é semelhante a um gerenciador de banco de dados tradicional baseado em páginas.

### 3.1.3 Arquitetura de cache semântico

O modelo de cache semântico armazena no disco do cliente as respostas das últimas consultas que foram executadas. Desta forma, novas consultas podem reaproveitar as respostas em cache.

As respostas das consultas são organizadas em segmentos semânticos, ao invés da organização tradicional dos dados em páginas de disco. O modelo semântico gerencia o *cache* do cliente como uma coleção de regiões semânticas; isto é, a informação do acesso é controlada, e a substituição do cache é executada, na unidade de regiões semânticas.

Cada região semântica possui uma fórmula que descreve seu conteúdo e um valor de substituição (*replacement value*) utilizado para decisões de substituição. Portanto, todas as tuplas de uma mesma região possuem o mesmo valor de substituição.

Para relacionar a descrição semântica com a região semântica são utilizados os segmentos semânticos que funcionam como um índice. Em um modelo básico, um segmento semântico é representado pelo conjunto ( $S_R$ ,  $S_A$ ,  $S_P$ , e  $S_C$ ), sendo que:

- $S_R$  e  $S_A$  definem a relação e os atributos envolvidos;
- $S_P$  indica a condição de seleção; e
- $S_C$  representa o resultado da consulta.

Para exemplificar o uso do cache semântico consideramos um banco de dados (REN, DUNHAM, 1999) com as duas relações: Employee (Eno, Ename, Age, Salary) e Project (Pno, Pname, Eno, Budget). Considere também que as seguintes consultas foram executadas dando origem aos seguimentos  $S_1$ ,  $S_2$ ,  $S_3$  e  $S_4$ :

$Q_1$ : *Select Ename From Employee Where 30 < Age < 40;*

$Q_2$ : *Select Pname, Budget From Project Where Budget > 10;*

$Q_3$ : *Select Ename, Salary From Employee Where Age <= 28;* e

$Q_4$ : *Select Ename, Age From Employee Where Age > 45;*

O quadro 3 ilustra um modelo de índice de cache semântico apresentado em (REN, DUNHAM, 1999).  $S_{TS}$  representa a marca de tempo que indica o momento em que o segmento foi utilizado pela última vez. Fisicamente, as tuplas que compõem um segmento podem estar armazenadas em páginas ou tuplas individuais.  $S_C$  aponta para o endereço da primeira página (ou tupla) do segmento.

S	S <sub>R</sub>	S <sub>A</sub>	S <sub>P</sub>	S <sub>C</sub>	S <sub>TS</sub>
S <sub>1</sub>	Employee	{Ename}	30 < Age < 40	2	T <sub>1</sub>
S <sub>2</sub>	Project	{Pname, Budget}	Budget > 10	4	T <sub>2</sub>
S <sub>3</sub>	Employee	{Ename, Salary}	Age <= 28	3	T <sub>3</sub>
S <sub>4</sub>	Employee	{Ename, Age}	Age > 45	8	T <sub>4</sub>

FONTE : (REN, DUNHAM, 1999)

### 3.2 Invalidação de cache (*Cache invalidation*)

A invalidação de cache visa manter a consistência dos dados entre o cliente e o servidor. Em um ambiente de computação móvel os dados do cache podem tornar-se inválidos por dois motivos: temporal e/ou espacial.

A invalidação temporal ocorre quando os dados são atualizados no servidor, tornando os dados em cache inconsistentes ao conteúdo original no servidor. A invalidação espacial ocorre quando os valores dos dados armazenados previamente em cache se tornam inválidos porque o cliente movimentou-se para uma nova área de localização, onde aquela informação já não é mais válida. Por exemplo, o usuário solicita o restaurante mais próximo da sua localização atual e armazena o resultado em cache. Após se locomover para uma nova área essa informação de restaurante mais próximo não será mais válida.

A seguir apresentamos alguns modelos para manter as cópias de dados em cache no cliente consistentes em relação ao servidor.

#### 3.2.1 Métodos de invalidação

O método de invalidação se refere ao tipo de acordo entre o cliente e o servidor para definir a forma em que o servidor irá manter o cliente atualizado das alterações.

Relatórios de invalidação é um método de invalidação temporal. Nesta abordagem, o servidor envia um relatório com as alterações a serem realizadas nos dados locais dos dispositivos móveis. O envio pode ser por broadcast (periódico ou não) ou por solicitações individuais dos clientes, isto é, sob-demanda (on-demand). As abordagens envolvendo broadcast de mensagens de invalidação para os dispositivos móveis tem sido bastante exploradas (BARBARA, IMIELINSKI, 1995), (CAI, TAN, 1999), (BARBARA, 1999), (HOU et al, 2001 ) e (LAI, TARI, BERTOK, 2003 ).

Em sistemas tradicionais normalmente são adotadas duas formas de notificação. Na primeira, o servidor envia uma mensagem de atualização sempre que um item de dado é

alterado. Se o cliente estiver desconectado no momento em que foi avisado, ele deverá contatar o servidor ao se conectar para obter uma nova versão do cache. Uma segunda forma é os clientes consultarem o servidor para verificar a validade de seu cache. Esta opção gera um alto tráfego na rede. Para ambientes móveis, nenhuns destes dois modelos são apropriados.

(BARBARÁ E IMIELINSKI, 1994) distinguem métodos de invalidação de cache em síncrono e assíncrono. No método assíncrono o servidor envia a mensagem de invalidação assim que o dado é alterado. Segundo (HOU et al, 2001), o maior problema na abordagem assíncrona é o imprevisível tempo de espera para as mensagens, que dependem das atividades de alteração dos dados.

No método síncrono, relatórios de invalidação são enviados periodicamente. Os servidores reúnem as alterações por um período de tempo e enviam um broadcast contendo estas alterações, em conjunto com informações sobre o tempo em que elas ocorreram. Se o cliente verificar que seus dados em cache não são válidos, ele deverá consultar o servidor e atualizar seu cache.

Conforme (CHUNG, CHO, 1998), o modo síncrono é vantajoso quando um cliente é reconectado após algum período de desconexão, pois existe uma garantia de latência devido à natureza periódica desta abordagem. (BARBARÁ E IMIELINSKI, 1994) enfatizam métodos síncronos para servidores tipo “stateless” (tipo de servidor que não sabem a posição do UM o tempo todo).

### **3.2.2 Tipo de servidor**

As estratégias de invalidação podem ser classificadas de duas formas, aquelas que utilizam servidor *stateful* e as que utilizam servidor *stateless*. Um servidor é dito *stateful* quando conhece as unidades móveis que estão em sua célula e também o estado de seus caches. Para mantê-lo atualizado, os clientes devem informar o servidor ao entrarem ou saírem de uma célula, e também quando desconectam e restabelecem a conexão. Já um servidor que desconhece informações sobre as UMs que estão em sua célula é denominado *stateless*.

A abordagem *stateful* é mais complexa (CAI, TAN, 1999). Isso acontece por vários motivos, tais como, o servidor precisa ter conhecimento de todas as UM que saíram de uma determinada célula ou se encontram desconectadas. Por existir esse tipo de dificuldade envolvida nesta abordagem *stateful*, uma grande parte dos trabalhos realizados na área de

coerência de cache são baseada na abordagem stateless (BARBARA, IMIELINSKI, 95) e (CAI, TAN, 1999).

Devido à assimetria na comunicação e a fraca capacidade de transmissão das unidades móveis é interessante que os servidores tenham um papel mais ativo na manutenção da coerência de cache (HOU et al, 2001 ).

### **3.2.3 Formas de disseminação de dados**

Basicamente existem três formas de disseminação de dados: *pull-based*, *push-based* e híbrida. Na estratégia *pull-based*, os clientes solicitam as informações desejadas enviando mensagens ao servidor, o qual devolve informações solicitadas para os clientes. O servidor entregará os dados para cada cliente de forma separada, respondendo cada mensagem de solicitação enviada pelo cliente (HARA, 2002).

Na estratégia *push-based* são enviadas informações para os clientes diretamente, não sendo necessário que os dados sejam solicitados. Uma das vantagens dos mecanismos *push-based* é o elevado *throughput* de dado (quantidade de dados transmitidos em uma unidade de tempo) em sistemas distribuídos com um grande número de clientes.

Uma terceira opção combina as duas estratégias anteriores, a estratégia híbrida (*hybrid mode*). Esta combinação pode ser proveitosa tanto para o servidor quanto para os clientes (BARBARA, 1999).

### **3.2.4 Estratégias de invalidação temporal**

Estratégias para coerência de cache garantem que os dados armazenados em cache estão sendo mantidos de forma consistente em relação ao servidor. Existem muitos algoritmos propostos para arquiteturas cliente-servidor convencionais. Entretanto, algoritmos convencionais devem ser reformulados devido às características específicas da computação móvel.

A seguir são descritas estratégias de invalidação temporal, isto é, inconsistência no cache ocasionadas pela alteração dos dados no servidor. O quadro 4 apresenta um resumo das características das estratégias apresentadas.

#### **3.2.4.1 Broadcasting timestamp (TS)**

Na estratégia TS (BARBARA, IMIELINSKI 1995), o servidor envia periodicamente por broadcast relatórios notificando sobre os itens que foram alterados nos últimos  $w$  segundos. O relatório de invalidação é composto por marcas de tempo (timestamps) das últimas mudanças destes itens.

A UM compara o valor da marca de tempo de cada item do relatório com as do cache para decidir se deve ou não atualizar o item de dado em cache. Se o item no cache tiver uma marca de tempo menor que a do relatório, então deverá ser atualizado. Para os itens não mencionados no relatório suas marcas de tempo são atualizadas com o valor da marca de tempo do relatório. A figura 8 apresenta o algoritmo desta estratégia proposto por (BARBARA, IMIELINSKI 1995).

FIGURA - 8: Broadcasting Timestamp

```

if ( $T_i - T_l > w$ ) { drop the entire cache }
else {
  for every item  $j$  in the MU cache {
    if there is a pair  $[j, t_j]$  in  $U_i$  {
      if  $t_j^c < t_j$  {
        throw  $j$  out of the cache }
      else {  $t_j^c = T_i$  }
    }
  }
}
for every item  $j \in Q_i$  {
  if  $j$  is in the cache {
    use the cache's value to answer the query }
  else { go uplink with the query }}
 $T_l := T_i$ 
}

```

FONTE: (BARBARA, IMIELINSKI 1995)

### 3.2.4.2 Adaptive Invalidation Reports (AIR)

Baseando-se na estratégia TS, (BARBARA, IMIELINSKI 1995) propõem um método adaptativo. Consideram que a frequência em que um item de dado é enviado no relatório depende do tamanho da janela  $w$ . Propõe aumentar a janela  $w$  para itens consultados frequentemente por UMs que permanecem muito tempo desconectadas.

Também acrescentam que o método TS deve ter o tamanho da janela  $w$  variável conforme o item de dado. Para isso, o servidor precisa de informações dos clientes para determinar (ou alterar) o tamanho da janela corretamente. Para recuperar essa informação, os

clientes enviam informações extras juntamente com as consultas ou é utilizada uma medida grosseira (por exemplo, se existir muitas consultas sobre um item de dado, sua janela será ampliada).

### 3.2.4.3 Ammestic terminals (AT)

Nesta estratégia o servidor informa os identificadores dos itens que mudaram desde o ultimo relatório de invalidação (BARBARA, IMIELINSKI 1995). Ao receber o relatório a UM compara os itens de seu cache com os que estão no relatório. Se um item de dado armazenado na UM for informado no relatório, a unidade móvel exclui este item do seu cache, caso contrário o cache é considerado válido. O algoritmo desta estratégia proposto por (BARBARA, IMIELINSKI 1995) é aprestando na figura 9.

FIGURA - 9: Amnesic Terminals

```

if ( $T_i - T_i > L$ ) { drop the entire cache }
else {
  for every item  $j$  in the MU cache {
    if  $j$  is in the report {
      throw  $j$  out of the cache }
  }
}
for every item  $j \in Q_i$  {
  if  $j$  is in the cache {
    use the cache's value to answer the query }
  else { go uplink with the query }}
 $T_i := T_i$ 
}

```

FONTE: (BARBARA, IMIELINSKI 1995)

#### **3.2.4.4 Signatures (SIG)**

A técnica Signatures (BARBARA, IMIELINSKI 1995) são checksums computados sobre os valores dos itens de dados. É uma técnica que compara arquivos de dados e verifica as diferenças. O problema desta comparação é que as UM não armazenam o banco de dados inteiro, portanto não podem ser comparados com o conjunto de dados completo dos servidores.

Para resolver este problema, esta técnica é adaptada da seguinte forma. As UM fazem “assinaturas” dos itens de seu interesse. Os itens que não estão no conjunto assinados são considerados iguais aos que estão no relatório que está sendo enviado por broadcast.

#### **3.2.4.5 Asynchronous and stateful (AS)**

Diferente dos modelos TS e AT, a estratégia AS, proposta em (KAHOL et al, 2001), é baseada em relatórios de invalidação assíncronos e com servidores tipo stateful. O objetivo é minimizar o overhead das unidades móveis para invalidar seu cache ao se re-conectarem na rede. Este modelo armazena mensagens de invalidação dos servidores no cache do Home Location da unidade móvel (HLC) enquanto a unidade móvel estiver desconectada. Quando a unidade móvel se reconectar, as mensagens de atualização armazenadas no HLC são enviadas a ela.

A vantagem deste modelo é que com o uso da invalidação assíncrona é minimizada a latência de acesso aos dados, e armazenando as mensagens de invalidação no Home Location da UM é minimizado o overhead gerado pela troca de mensagens entre cliente-servidor para atualizar seu cache sempre que ocorrerem reconexão.

#### **3.2.4.6 Cache coherence schema with incremental update propagation (CCS-IUP)**

Segundo (CHUNG, CHO, 1998), a estratégia CCS-IUP tenta evitar a invalidação dos dados do cache, reduzindo desta forma o número de percas de informações. Além disso, esta estratégia propaga apenas as partes relevantes das modificações que afetam o dado armazenado localmente na unidade móvel. Com isso reduz significativamente a quantidade de dados transmitidos. A figura 10 apresenta a técnica CCS-IUP.

Nesta estratégia, o banco de dados é representado por um conjunto de relações  $R_i$ . Cada relação possui um número de registros. Para cada registro na relação

$R_i$  existe um identificador único (RID). O servidor deve enviar um broadcast periódico no tempo  $T_k$  com um relatório de invalidação contendo informações sobre os registros

alterados nos últimos 4 segundos. O relatório de invalidação é definido formalmente como uma lista H cujo elemento é uma 4-tupla (Ri, RID, VAL, TS), sendo que Ri representa o identificador da relação ao qual o registro RID pertence; RID representa o identificador do registro que foi alterado no intervalo [Tk - 4, Tk]; VAL representa o novo valor do registro e TS representa o timestamp indicando quando o registro foi alterado pela última vez.

Figura -10: CCS-IUP

```

if ( $T_k - T_l > \Delta$ ) { drop entire cache } /* reconnected after long periods of disconnection */
else {
  for every element ( $R_i, RID, VAL, TS$ ) in  $\Phi$  {
    if ( $\exists \mathcal{V}_e$  such that the client caches  $\mathcal{V}_e$  and  $R_i \in PR_e$ ) {
      if ( $TS < TS(R_i)$ ) { skip  $\Phi$  } /* The log has been reflected already */
      else {
        if ( $VAL$  is null)  $D_c = D_c \cup \{(RID, TS)\}$ 
        else  $I_c = I_c \cup \{(VAL, TS)\}$ 
         $TS(R_i) = TS$  /* increments the timestamp of base relation */
      }
    }
  }
   $T_l = T_k$  /* Reset the timestamp of the client to the current time */
}
for every query  $Q_i$  {
  if (the result of  $Q_i$  is not cached) /* cache miss */
    { request  $Q_i$  to the server }
  else {
    Let  $\mathcal{V}_i$  be the cached result of  $Q_i$ 
    if ( $\exists PR_i$  such that  $D_i \neq \emptyset$  or  $I_i \neq \emptyset$ ) { /* Base relation has been updated */
      call view reconstruction
       $TS(\mathcal{V}_i) = \text{MAX}_{R \in PR_i}(TS(R))$  /* view's TS is set to maximum TS of relations */
    }
    Use  $\mathcal{V}_i$  to answer  $Q_i$ 
  }
}

```

FONTE: (CHU, 1998)

### 3.2.4.7 Quasi-copies

Na técnica *quasi-copies* (BARBARA e IMIELINSKI, 94) a consistência é relaxada para reduzir o tamanho dos relatórios de invalidação. É permitido que os valores não estejam sempre atualizados denominados “quase-copies”, isto é, quase cópias dos valores dos servidores. Um quase copie é um valor do cache que é permitido se diferencia do valor do servidor central em um modo controlado. Por exemplo, se a unidade móvel armazena preço de produtos, e é aceitável usar valores que não estão completamente atualizados, desde que esteja 0.5 % do valor do preço verdadeiro do servidor.

### 3.2.4.8 Lazy pull-based

Mecanismos de cache em um ambiente cliente-servidor convencional são normalmente baseados em páginas (page-based). De acordo com (CHAN, RODDICK, 1998), mecanismos de cache em banco de dados móveis devem ter uma granularidade menor. Consideram três níveis de granularidade: attribute caching (armazenam somente acessados) e híbrid caching (armazenam atributos que poderão ser acessados no futuro, pré-fetched).

Propõe um modelo de invalidação denominado lazy pull-based no qual cada cliente é responsável por invalidar seus itens do cache. Assim, o relatório de atualização que o servidor deverá enviar acontece sob demanda do cliente.

Baseiam-se na idéia do mecanismo de cache apresentada em (GRAY, CHERITON, 1989) que estima um tempo de provável atualização (RT) para cada objeto em cache.

#### QUADRO 4: CARACTERÍSTICAS DE ESTRATÉGIAS DE INVALIDAÇÃO DE CACHE TEMPORAL

Propriedades Invalidação Temporal	Arquitetura de Rede	Tipo de Servidor	Método de Invalidação	Hand-off	Tempo de Conexão	Tempo de atualiz./ Freq. de acesso	Consistência	Granularidade
TS, AIR e AT [3]	Celular	Stateless	Broadcast Sincrono	Exclui Cache	Considera	n.v.	n.v.	Mencionado como trabalho futuro
SIG [3]	Celular	Stateless	Broadcast Sincrono	Exclui Cache	Considera	Mencionado como trabalho futuro	n.v.	Mencionado como trabalho futuro
Quase-cópias [3]	Celular	Stateless	Broadcast Sincrono	Exclui Cache	n.v.	n.v.	Relaxa a consist.	n.v.
Lazy Pull-Based [6]	Ponto a Ponto	n.v.	Sob Demanda	n.v.	n.v.	Estima um tempo de provável atualização	n.v.	níveis de granularidade dos bits em cache
AS [18]	Celular	Statefull	Broadcast Assíncrono Call-back	Mantém Cache	Considera	n.v.	n.v.	n.v.
BS[16]	Celular	Stateless	Broadcast Periódico	n.v.	n.v.	n.v.	n.v.	Níveis de granularidade dos bits em relatórios
CCS-IUP	Celular	Stateless	broadcast periódico	Mantem cache	considera	n.v.	n.v.	n.v.

n.v.=não verificado

O valor do RT de cada objeto vai depender da sua probabilidade de atualização. Se um objeto é atualizado frequentemente seu RT será menor. O RT de um objeto indica o tempo que ele pode ficar no cache sendo considerado como válido

### 3.2.5 Estratégias de invalidação espacial

Além das alterações ocorridas nos servidores, os valores dos dados armazenados previamente em cache podem se tornar inválidos quando um cliente se move para uma nova

área de localização. A manutenção de um cache válido quando os clientes se movem é chamado de invalidação de cache dependente de localização (*location-dependent cache invalidation*).

Em um sistema dependente de localização, os valores dos dados dependem da localização geográfica. Um item de dado pode ter diferente valor conforme varia sua localização (ZHENG, XU, LEE, 2002). Estratégias tradicionais não consideram esta característica, portanto são ineficientes para sistemas dependentes de localização.

De acordo com (ZHENG, XU, LEE, 2002) as propriedades dependentes de localização poderiam gerar novos desafios e oportunidades de pesquisas. A informação em cache resultante de uma consulta (o restaurante mais próximo, por exemplo) pode se tornar inválida quando o cliente se move de um local para outro. A manutenção de um cache válido quando os clientes se movem é chamado de invalidação de cache dependente de localização (*location-dependent cache invalidation*).

A seguir são descritas algumas estratégias propostas na literatura para relacionar o valor do dado com a sua área válida. O quadro 5 (MANICA, CAMARGO, 2004) resume as principais características de cada estratégia apresentada.

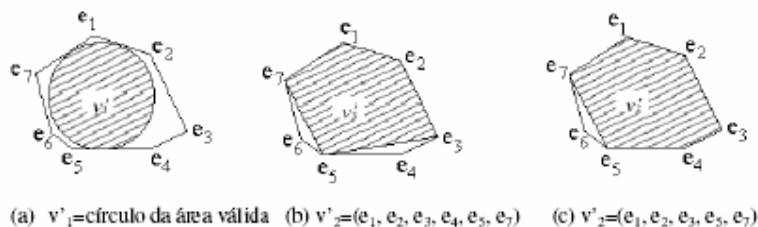
#### **3.2.5.1 Regional endpoints (PE)**

O esquema PE (ZHENG, XU, LEE, 2002) armazena todos os pontos da aresta do polígono que representa a área válida para um determinado valor de dado. Quando o número de pontos for grande irá consumir recursos de rede ao serem transmitidos e também muito do espaço reservado para cache no cliente, reduzindo o espaço para armazenar o próprio dado. Portanto, este esquema possui a desvantagem de piorar a performance, porém a vantagem de prover conhecimento completo da área válida dos valores.

#### **3.2.5.2 Approximate circle (AC)**

Também proposta em (ZHENG, XU, LEE, 2002), AC é uma alternativa para não armazenar todos os pontos das arestas do polígono que representa a área válida de um determinado valor de dado é usar um círculo aproximado ao polígono figura 11. Assim a área válida será a área aproximada definida pelo centro e raio do círculo e o overhead deste esquema é minimizado. O problema, é que quando o formato do polígono for magro e comprido, o erro de aproximação é maior. Poderá acontecer do cache tratar incorretamente dados válidos como inválidos se a localização da consulta estiver fora do círculo.

FIGURA - 11: Approximate Cycle (AC)



FONTE: (ZHENG, XU, LEE, 2002)

### 3.2.5.3 Método genérico caching-efficiency-based (CEB)

De acordo com (ZHENG, XU, LEE, 2002), um método ideal é aquele que balanceia o overhead e a precisão de áreas válidas para os valores de dados. Neste modelo é considerado um novo critério de desempenho: a eficiência do cache (cache efficiency).

Considerando que a área válida e um valor de dado é  $v$ , e  $v_i'$  é uma sub-região contida em  $v$  como mostra a figura 12 a eficiência de cache de um valor de dado com respeito a  $v_i'$  é definido por:

FIGURA 12: Método Genérico Caching-Efficiency-Based (CEB)

$$E(v'_i) = \frac{A(v'_i)/A(v)}{(D + O(v'_i))/D} = \frac{A(v'_i)D}{A(v)(D + O(v'_i))}$$

FONTE: (ZHENG, XU, LEE, 2002)

Sendo,  $D$  o tamanho do dado,  $A(v'_i)$  a área de  $v'_i$  e  $O(v'_i)$  o overhead necessário para armazenar  $v'_i$ . calculando-se a eficiência temos:  $E(v) = 0.696$ ,  $E(v'1)=0.721$ ,  $(v'2) = 0.706$  e  $E(v'3)=0.694$ . Como a eficiência em  $v'1$  apresentou maior valor, seria este que seria anexado ao valor de dado.

#### 3.2.5.4 Bit vector with compression (BVC)

O método BVC, proposto em (XU et al., 1999), considera que cada célula possui uma identificação – ID que diferencia uma da outra. Ele utiliza um vetor denominado *bit vector* para guardar informações do escopo do sistema. O tamanho do *bit vector* é igual ao número de células do sistema e todo dado em cache é associado a um bit vector.

Assim, o valor 1 (um) significa que o item de dado é válido na célula e 0 (zero) inválido. Se o sistema for muito grande, pode ocorrer um overhead e é sugerido executar uma compressão do bit vector. Por exemplo, se um sistema apresenta 10 áreas de serviços (células), cada item de dado em cache irá possuir um *bit vector* com 10 bits. Se o valor do *bit vector* de um determinado item de dado for 0000001110, significa que este item de dados é válido na sétima, oitava e nona célula do sistema.

#### 3.2.5.5 Grouped bit vector with compression (GBVC)

Em grandes sistemas, para evitar o *overhead* no modelo BVC, é proposta em (XU et al., 1999) a construção de um vetor correspondente com apenas parte das células do sistema. Considerando que as unidades móveis podem se mover em grandes distâncias, o modelo GBVC propõe armazenar no *bit vector* dos itens de dados em *cache* somente informações sobre células adjacentes e próximas a sua localização corrente.

O modelo propõe a divisão da área geográfica total do sistema em grupos e subgrupos. O ID da célula consiste em duas partes: grupo ID e intragrupo ID. Considerando o mesmo exemplo do método anterior, as áreas de serviço seriam divididas em: grupo 0 (células 1 a 5) e

grupo 1 (células 6 a 10). Com o método GBVC, um bit é usado para representar o ID do grupo e cinco bits para representar as células do grupo. O valor do *bit vector* de um determinado item de dado poderia ser (1, 11100) significa que é válido nas três primeiras células do grupo 1, ou seja, é válido na célula 6, 7 e 8.

### 3.2.5.6 Implicit scope information (ISI)

Esta estratégia também proposta em (XU et al., 1999), divide o banco de dados em múltiplas seções lógicas. Itens de dados com a mesma área válida são colocados na mesma seção.

O item de dado em *cache* terá o seguinte formato {Di, SDNi e SNi} onde Di é o valor do item de dado, SDNi o número da seção e SNi o número do dado dentro da seção (*scope number*) para diferentes áreas válidas..

QUADRO 5: RESUMO DAS CARACTERÍSTICAS EM ESTRATÉGIAS DE INVALIDAÇÃO DE CACHE ESPACIAL

Estratégias (Invalidação Espacial)	Características	Forma de Armazenar Área Válida
Regional Endpoints (PE)	Prioriza a precisão. Ocupa muito espaço prejudicando o desempenho do sistema.	Pontos (todos os pontos)
Aproximate Circle (AC)	Prioriza o overhead. Prove menor precisão, porém melhora o desempenho do sistema.	Pontos (aproximado)
Método Genérico Caching-Efficiency-Based (CEB)	Prioriza o overhead e a precisão. Considera o fator eficiência de <i>Cache</i> .	Pontos
Bit Vector with Compression (BVC)	Adequado para sistemas pequenos. Sugere o uso de métodos para compressão para sistemas maiores.	Vetores
Grouped Bit Vector with Compression (GBVC)	Minimiza o overhead Adequado para grandes sistemas	Vetores
Implicit Scope Information (ISI)	Armazenamento dos dados conforme seções lógicas para diferentes áreas válidas.	Seções

FONTE: (MANICA, CAMARGO, 2004)

### 3.3 Substituição de cachê

Quando o cache no cliente móvel esta cheio, e novos dados devem ser admitidos, uma política de substituição deve ser utilizada para identificar e deletar os dados que tiverem a menor possibilidade de serem usados novamente.

Em seus trabalhos (ACHARYA, 1998), (ALONSO, KORTH, 1993) e (FRANKLIN, FRANKLIN, JONSSON e KOSSMANN, 1996), tratar políticas de substituição de *cache* em ambiente sem fio. Os autores propuseram uma política de substituição de *cache* denominada *PIX* onde era substituído o item de dado com o menor valor de  $p/x$ , sendo  $p$  a probabilidade de acesso do item e  $x$  a frequência em que era enviado em *broadcast*. Esta abordagem também é tratada pelos autores (KHANNA, LIBERATORE, 2000) e (TASSIULAS e SU, 1997), que apresentam propostas para substituição de *cache*.

Em serviços dependentes de localização, para definir quais dados serão substituídos no *cache*, além da probabilidade de acesso existem outros fatores que devem ser considerados tais como: movimento, direção e velocidade da UM.

A seguir apresentamos algumas propostas organizadas conforme o fator principal utilizado para substituição dos dados em cache.

#### 3.3.1 Tempo

De acordo com (ARAÚJO, FERREIRA, 2000) este modelo está baseado na agregação de dados temporais, às unidades de dados do cache. Deste modo, saber quando um dado foi acessado é um forte indicador para possíveis novos acessos.

- LRU – Less Recently Used – Menos recentemente usado – O LRU considera como região “fria” do cache os dados com marcas de acesso mais antigas, considerando-os candidatos a substituição.

#### 3.3.2 Frequência

Em seu trabalho (PARKE, GRYZ, 1998), sugere que contabilize informações de frequência de acesso ao dado. Um dado pouco acessado tem baixa probabilidade de ser solicitado novamente e deve ser substituído.

- LFU – Less Frequently Used – Menos frequentemente usado – Define a região fria do cache com o conjunto de dados com a menor frequência de acesso (ARAÚJO, FERREIRA, 2000).

### 3.3.3 Área válida

Na abordagem proposta em (ZHENG, XU, LEE, 2002) o principal fator considerado além da probabilidade de acesso do item de dado é a área válida. A área válida refere-se à área geométrica em que um valor de dado é válido.

Na política denominada *Probability Área* (PA) o custo é calculado multiplicando a probabilidade de acesso do item de dado com a área válida. A função de custo de um valor  $j$  de um item  $i$  é representada por:

$$C_{ij} = P_i \cdot A(v'_{ij})$$

Sendo  $P_i$  a probabilidade de acesso de um item  $i$  e  $A(v'_{ij})$  a área válida para um valor  $j$  de um item de dado  $i$ .

Ao ser executada a substituição de dados, esta política seleciona como vítima os dados com menor custo  $C_{ij}$ .

### 3.3.4 Distância

Esta abordagem é interessante para sistemas dependentes de localização, nos quais o servidor responde uma consulta de acordo com a localização corrente do cliente. Esta política considera além da probabilidade de acesso do item de dado e sua área válida, o fator distância.

Quando uma área válida para um valor de dado está longe da localização atual do cliente, este dado terá menor chance de ser utilizado já que levará algum tempo até que o cliente entre na área válida.

(ZHENG, XU, LEE, 2002) propuseram a política denominada *Probability Área Inverse Distance* (PAID) onde a função de custo de um valor  $j$  de um item  $i$  é dada por:

$$C_{ij} = P_i \cdot A(V'_{ij}) / D(V'_{ij})$$

Sendo  $P_i$  a probabilidade de acesso de um item  $i$  e  $A(V'_{ij})$  a área válida para um valor  $j$  de um item de dado  $i$ .  $D(V'_{ij})$  a distância entre a localização corrente e a área válida  $V'_{ij}$ .

Da mesma forma que no modelo PA, ao ser executada a substituição de dados, esta política seleciona como vítima os dados com menor custo.

### 3.3.5 Distância em áreas urbanas

A localização do cliente e a área válida de um dado pode mudar quando considera-se características específicas de um determinado tipo de área. Em uma zona rural por exemplo, a distância é um fator que poderia ser calculado através de fórmula euclideana  $\sqrt{(x1-x2)^2 + (y1-y2)^2}$ .

Já em uma área urbana, devido ao grande número de construções, este cálculo não seria muito correto pelo fato de que os usuários se movem somente através das ruas. Assim, em uma área urbana o cálculo da distância deve ser diferente que nas zonas rurais.

(KIM et al., 2002) definem distância através do cálculo “Manhattan distance”, que é calculado por  $(x1-x2) + (y1-y2)$ . O valor dos dados que estão em cache varia conforme a UM muda sua posição.

O algoritmo para atualização de cache proposto calcula o cache replacement score baseando-se no seguinte cálculo:

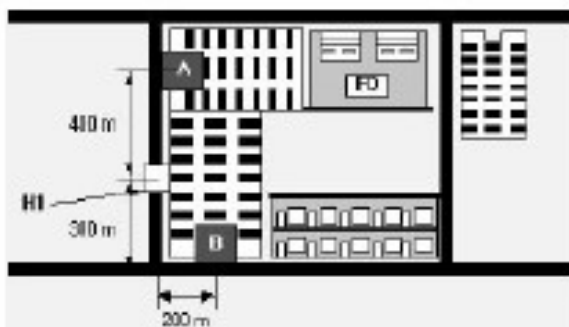
$CRS\_MD (MH_i, D_i) = (X_{Hi} - X_i) + (Y_{Hi} - Y_i)$ , sendo  $(MH_i, X_{Hi})$  a localização do host móvel e  $(X_i, Y_i)$  a localização do dado  $D_i$ .

Para atualizar o cache, o algoritmo proposto escolhe as vítimas de acordo com o CRS calculado através da localização corrente do MH. Desta forma, os dados que estiverem mais longes serão substituídos primeiro.

Quanto a frequência de cálculo do CRS, os autores consideram que em pequenos movimentos não é interessante recalculá-lo, devido a sobrecarga de trabalho no host móvel. Propõem que o cálculo deve acontecer sempre que a UM entrar em uma nova grid.

Na figura 13, pela fórmula euclideana, a distância de H1 e B seria 360 metros ( $\sqrt{300^2 + 200^2}$ ), já pelo cálculo Manhattan distance seria 500 metros, o valor real que H1 deve percorrer para chegar em B

FIGURA – 13: Construções e ruas em área urbana



FONTE: (KIM et al., 2002)

### 3.3.6 Direção do movimento do cliente



## 4 CONCLUSÕES E TRABALHOS FUTUROS

Muitos desafios impostos pela mobilidade no ambiente de computação móvel nos fazem perceber a necessidade de rever algumas soluções utilizadas no gerenciamento de dados tradicionais. Entre outras dificuldades desta nova arquitetura esta a limitação do tempo de conexão das unidades moveis, pois pode haver desconexão no meio de um serviço por diferentes motivos, tais como termino da bateria do dispositivo ou desconexão por perda de sinal.

Uma solução para esta problemática discutida neste trabalho foi o armazenamento no cache do cliente cópias de itens de dados mais freqüentemente acessados. Para que se garanta a consistência dessas cópias em relação aos dados do servidor, e que haja uma substituição de dados quando existir espaço disponível no cliente é necessário um método eficiente de gerenciamento de cache no cliente.

Para se entender o gerenciamento dos dados em cache, é preciso estudar diversas técnicas propostas na literatura que enfocam técnicas de invalidação e substituição de cache. Tais soluções enfocam um ou mais fatores impostos pela mobilidade dos clientes.

A formas de armazenamento do cache das propostas apresentadas no capítulo 2 são tuplas, pagina e semântico:

- no modelo de cache em tuplas pode-se observar que este apresenta maior flexibilidade no gerenciamento, porém pode-se ocorrer uma perda no desempenho;
- no modelo de cache em páginas ocorre o oposto, as informações de controle ficam reduzidas, porém limitadas. Sendo assim, podemos dizer que os dois métodos têm suas vantagens e desvantagens, e a estratégia a ser usada depende da aplicação;
- já no modelo de cache semântico pode-se observar através das propostas analisadas uma diminuição no overhead dos dados de controle e tem uma flexibilidade no gerenciamento da região semântica, tornando essa abordagem indicada para o ambiente de computação móvel. Porém esta técnica ainda demanda muitos estudos devido a suas restrições.

Inicialmente apresentamos conceitos fundamentais para o entendimento de gerência de cache em ambiente de computação móvel. Em seguida no capítulo 3 enfatizamos diferentes técnicas para invalidação e substituição de cache propostas na literatura.

Para invalidação de cache, muitos algoritmos têm sido propostos para arquitetura cliente-servidor convencionais, contudo, devido às características próprias de um ambiente

móvel os algoritmos convencional não são diretamente aplicável à computação móvel. Neste trabalho foram apresentadas, algumas propostas de invalidação para este ambiente.

Podemos considerar duas possibilidades de gerenciamento de cache, com o servidor stateful ou com o servidor stateless. A abordagem stateful é considerada complexa e por isso a maior parte dos trabalhos realizados na área de coerência de cache são baseados na abordagem stateless.

Para se determinar quais dados serão substituídos no cache, adota-se uma política de substituição, baseada nos seguintes fatores de classificação:

- Considerando a área válida para valores de item de dados;
- Considerando a distância;
- Considerando a distância em áreas urbanas;
- Considerando a direção do movimento do cliente;

#### **4.1 Relevância deste trabalho**

A contribuição deste trabalho, para a solução dos problemas de gerenciamento de cache em computação móvel, foi um estudo realizado sobre os problemas e soluções existentes sob a visão de diferentes autores.

Apresentamos uma revisão bibliográfica de trabalhos na área de gerenciamento de cache em computação móvel, contribuindo com uma relação de conceitos e propostas importantes e indispensáveis para trabalhos futuros nesta área.

A partir dos problemas e soluções apresentadas, poderão surgir novas propostas tratando as diversas restrições impostas pelo ambiente móvel.

## 4.2 Trabalhos futuros

A seguir algumas perspectivas de trabalhos futuros:

- Aprofundar o estudo sobre o modelo de armazenamento de dados em cache semântico.
- Implementar um ambiente real para que se possa testar estratégias com o objetivo de definir a mais indicada conforme o tipo de dispositivo móvel a ser usado, baseando-se em análise de cada dispositivo, analisando o tempo de conexão de cada um, a duração do tempo de bateria e outros aspectos.
- Após um estudo das estratégias existentes, propor um novo mecanismo de substituição de cache baseando-se na fusão de propostas já existentes, por exemplo, LFU e FAR.

## 5 REFERÊNCIAS BIBLIOGRÁFICAS

ACHARYA , S.; MUTTHUKRISHMAN, S. **Scheduling on-demand broadcasts: New metrics and algorithms.** In proceedings of the 4<sup>th</sup> Annual ACM/IEEE International Conference On Mobile Computing and networking (MobiCom' 98), pages 43-54, Dallas, TX, USA, October 1998

ACHARYA, S. **Broadcast Disks: Dissemination-based Data Management for Asymmetric Communication Environments.** In PhD Dissertation at Brown University, May 1998.

ACHARYA, S. et at. **Broadcast Disks: Data Management for Asymmetric Communications Environments.** In Proceedings of ACM SIGMOD Conference on Management of Data, pages 199-210, San Jose, CA, USA, May 1995.

AKSOY , D.; FRANKLIN, M. J.; ZDONIK, S.; **Data staging for on-demand broadcast.** Proceedings of the 27<sup>th</sup> VLDB conference (VLDB'01), Roma, Italy, September 2001.

AKSOY, D.; FRANKLIN, M. J.; **R x W A scheduling approach for large-scale on-demand data broadcast.** ACM/IEEE Transactions on networking, 1999

ALONSO, R.; KORTH, H. F. **Databases System Issues in Nomadic Computing.** Matsushita Information Technology Laboratory: Princeton, NJ, 1993.

ARAÚJO, V. L.; FERREIRA, E. J. **Cache semântico para computação sem fio baseado na Abstração de composição dos dados.** Universidade de São Paulo: Instituto de Matemática e Estatística (IME), 2000

BADRINATH, B.R. et al. **Impact of Mobility on Distributed Computations.** Rutgers University - Department of Computer Science: New Brunswick, NJ, 1993.

BARBARA D.; IMELINSKI, T. **Sleepers and workaholics: Caching strategies for mobile environments.** In Proceedings of ACM SIGMOD Conference on Management of data, paginas 1-12, Minneapolis, MN, USA, May 1994 (An extended Versio Appered in VLDB Journal 4(4): 567-602, 1995.

BARBARA, D. **Mobile computing and databases-a survey.** IEEE Transactions on Knowledge and Data Engineering, [S.l.], v.11, n.1, p.108-117, 1999.

BARBARA, D. **Mobile computing and databases-a survey.** IEEE Transactions on Knowledge and Data Engineering, [S.l.], v.11, n.1, p.108-117, 1999.

BERKENBROCK, C. D. M.; **Ivestigação e implementação de estratégias de notificação de invalidação para coerencia de cache em ambientes de computação móvel.** Universidade Federal de Santa Catarina, 2004. Dissertação de Mestrado.

BERKENBROCK, C. D. M.; RIBEIRO, M. A. **Análise comparativa entre estratégias de coerência de cache em um ambiente de computação móvel.** 3rd International Information and Telecommunication Technologies Symposium, [S.l.], 2004.

BRITO, M. **Bem- vindo à Era Bluetooth.** Extraído de TIMaster, janeiro de 2001.

CAI, J.; TAN, K. L. Energy-efficient selective cache invalidation. *Wirel. Netw.*, [S.l.], v.5, n.6, p.489–502, 1999.

CAI, J.; TAN, K.-L. **Energy-efficient selective cache invalidation.** *Wirel. Netw.*, [S.l.], v.5, n.6, p.489–502, 1999.

CAO, G. On **improving the performance of cache invalidation in mobile environments.** *Mob. Netw. Appl.*, [S.l.], v.7, n.4, p.291–303, 2002.

CHAN, D.; RODDICK, J. F. **Context-sensitive mobile database summarisation.** In: PROCEEDINGS OF THE TWENTY-SIXTH AUSTRALASIAN COMPUTER SCIENCE CONFERENCE ON CONFERENCE IN RESEARCH AND PRACTICE IN INFORMATION TECHNOLOGY, 2003. Australian Computer Society, Inc. p.139–149.

CHAN, D.; RODDICK, J. F. **Context-sensitive mobile database summarisation.** In: PROCEEDINGS OF THE TWENTY-SIXTH AUSTRALASIAN COMPUTER SCIENCE CONFERENCE ON CONFERENCE IN RESEARCH AND PRACTICE IN INFORMATION TECHNOLOGY, 2003. Australian Computer Society, Inc. p.139–149.

CHIASSERINI, C. F.; NUGGEHALLI, P.; SRINIVASAN, V. **Energy-efficient communication protocols.** In: PROCEEDINGS OF THE 39TH CONFERENCE ON DESIGN AUTOMATION, 2002. ACM Press. p.824–830.

CHUNG, H.; CHO, H. **Data caching with incremental update propagation in móbile computing environments.** In: AUSTRALIAN COMPUTER JOURNAL, 1998. [s.n.].

CUNHA, D. P. **Um Estudo das Estratégias de Replicação e Reconciliação de Banco de Dados Móveis em um Ambiente Wireless.** Universidade Federal de Santa Catarina, 2003. Dissertação de Mestrado.

DANTAS, M. **Tecnologias de redes de comunicação e computadores.** Editora Axcel Books, 2002.

DAR, S. et al. **Semantic Data Caching and Replacement**. Proc. VLDB Conf., pp. 330-341, 1996.

FLINN, J.; SATYANARAYANAN, M. **Powerscope: a tool for profiling the energy usage of mobile applications**. In: PROCEEDINGS OF THE SECOND IEEE WORKSHOP ON MOBILE COMPUTING SYSTEMS AND APPLICATIONS, 1999. IEEE. p.2–10.

FRANKLIN, M.; JONSSON, B.; KOSSMANN, D. **Performance tradeoffs for client-server query processing**. Proc. ACM SIGMOD Conf., 1996

GRAY, C. G.; D. CHERITON R. **Leases: An Efficient Fault-Tolerant Mechanism for Distributed File Cache Consistency**. In Proceedings of SOSp, pages 202-210, 1989.

HARA, T. **Cooperative caching by mobile clients in push-based information systems**. In: PROCEEDINGS OF THE ELEVENTH INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, 2002. ACM Press. p.186–193.

HKFASBENDER, A., HOFF, S.; PIETSCHMANN, M.; “**Mobility Management in Third Generation Mobile Networks**”. Proceedings of the IFIP TC6 International Workshop on Personal Wireless Communications (Wireless Local Access). Prague, April. 1995

HOU, W. C. et al. **An optimal construction of invalidation reports for mobile databases**. In: PROCEEDINGS OF THE TENTH INTERNATIONAL CONFERENCE ON INFORMATION AND KNOWLEDGE MANAGEMENT, 2001. ACM Press. p.458–465.

HUANG, Y.; SISTLA, P.; WOLFSON, O. **Data replication for mobile computers**. In: PROCEEDINGS OF THE 1994 ACM SIGMOD INTERNATIONAL CONFERENCE ON MANAGEMENT OF DATA, 1994. ACM Press. p.13–24.

JING, J. et al. **Bit-sequences: A new cache invalidation method in mobile environments**. ACM/ MONET. 2(2): 115-127, 1997.

KAHOL, A. et al. **A strategy to manage cache consistency in a distributed mobile wireless environment**. IEEE Transactions on Parallel and Distributed Systems (TPDS), 2001.

KHANNA, S.; LIBERATORE, V. **On broadcast disk paging**. In: SIAM Journal on Computing, 29(5): 1683-1702, 2000

KIM, K. et al. **Broadcasting and Caching Policies for Location-Dependent Queries in Urban Areas**. WMC'2002. Atlanta, Georgia, USA. September, 2002.

LAI, K. Y.; TARI, Z.; BERTOK, P. **Cost efficient broadcast based cache invalidation for mobile environments**. In: PROCEEDINGS OF THE 2003 ACM SYMPOSIUM ON APPLIED COMPUTING, 2003. ACM Press. p.871–877.

LAUZAC, S. W.; CHRYSANTHIS, P. K. **Personalizing information gathering for mobile database clients**. In: PROCEEDINGS OF THE 2002 ACM SYMPOSIUM ON APPLIED COMPUTING, 2002. ACM Press. p.49–56.

LEONG, H. V.; SI, A. **On adaptive caching in mobile databases**. In: PROCEEDINGS OF THE 1997 ACM SYMPOSIUM ON APPLIED COMPUTING, 1997. ACM Press. p.302–309.

MANICA, H.; CAMARGO, M. S. **Caching Estrategies for Mobile Computing Systems**. **Proceedings** of the 6th International Conference on Enterprise Information Systems. Universidade Portucalense, Porto – Portugal. April, 2004.

MATEUS, G. R.; LOUREIRO A. A. **Introdução à Computação Móvel**. DCC/IM, COPPE/SISTEMAS, NCE/UFRJ, 11ª Escola de Computação, 1998.

MURTHY, V. K. **Mobile computing deploying agents**. In: TWELFTH INTERNATIONAL CONFERENCE ON INFORMATION NETWORKING, 1998. IEEE. p.127–130.

OOI; J. C. K.-L. T. B. C. **On incremental cache coherency schemes in mobile computing environments**. In: PROCEEDINGS. 13TH INTERNATIONAL CONFERENCE ON DATA ENGINEERING, 1997. [s.n.].

ÖZSU, M. T.; VALDURIEZ, P. **Principles of Distributed Database Systems**. Prentice-Hall, Inc. Upper Saddle River, New Jersey-USA: 1999. pp. 596-603.

PITOURA, E.; CHRYSANTHIS, P. K. **Exploiting versions for handling updates in broadcast disks**. In: THE VLDB JOURNAL, 1999. [s.n.]. p.114–125.

PITOURA, E.; SAMARAS, G. **Data management for mobile computing**. In: KLUWER ACADEMIC PUBLISHERS, 1998. [s.n.]. v.10.

REN, Q.; DUNHAM M. H.. **Using Clustering for Effective Management of a Semantic Cache in Mobile Computing**. MobiDE – Seattle – USA. ACM 1999.

ROCHA, R. C. A. **Uma Arquitetura para Simulação Flexível de Protocolos para Computação Móvel**. São Paulo: Instituto de Matemática e Estatística da Universidade de São Paulo. 2001.

SILVA, C. E. **Um Estudo dos Principais Modelos de Transações em Banco de Dados Móveis e uma Proposta Diferenciada do Modelo PRO-MOTION**. Florianópolis: Instituto de Informática e Estatística da UFSC, 2003.

SOUZA, L. B. **Redes de Computadores dados, Voz e Imagem**, Editora Érica, 1999.

TANENBAUM, A. S. **Redes de computadores**. Rio de Janeiro. Editora Campus – 1997.

TASSIULAS, L.; SU, C. J. **Optimal memory management strategies for a mobile user in broadcast data delivery system**. IEEE Journal on Selected Areas in Communications (JSAC), 15 (7): 1226-1238, September 1997.

TAURION, C. **Internet Móvel – Tecnologias, Aplicações e Modelos**, Rio de Janeiro: Campus, 2002.

WEBINSIDER. **Wi-Fi, bluetooth... introdução às redes sem fio**. Disponível em <<http://www.linuxit.com.br/article3403.html>> Acesso em 28 jun. 2005.

WITT, D. de et. al. **A study of three alternative workstation- server architectures for object-oriented database systems**, Proc. VLDB conf., 1990

XU, J et al. **Cache Coherency in Location-Dependent Information Services for Mobile Environment**, Proc. the 1st Int. Conf. on Mobile Data Access (MDA'99), Hong Kong, Springer- Verlag LNCS, vol. 1748, pp. 182-193, December 1999.

XU, J et al. **Performance evaluation of an optimal cache replacement policy of Wireless Data Dissemination**. University of Science and Technology, Hong kong, 2002

YE, T.; JACOBSEN, H.-A.; KATZ, R. **Mobile awareness in a wide area wireless network of info-stations**. In: PROCEEDINGS OF THE 4TH ANNUAL ACM/IEEE INTERNATIONAL CONFERENCE ON MOBILE COMPUTING AND NETWORKING, 1998. ACM Press. p.109–120.

YUEN, J. C.-H. et al. **Cache invalidation scheme for mobile computing systems with real-time data**. SIGMOD Rec., [S.l.], v.29, n.4, p.34–39, 2000.

ZHENG, B; XU, B; LEE, D. L.. **Cache Invalidation and Replacement Strategies for Location-Dependent Data in Mobile Environments**. IEEE Trans. on Computers, Special Issue on Database Management and Mobile Computing, 51(10): 1141-1153, October 2002.