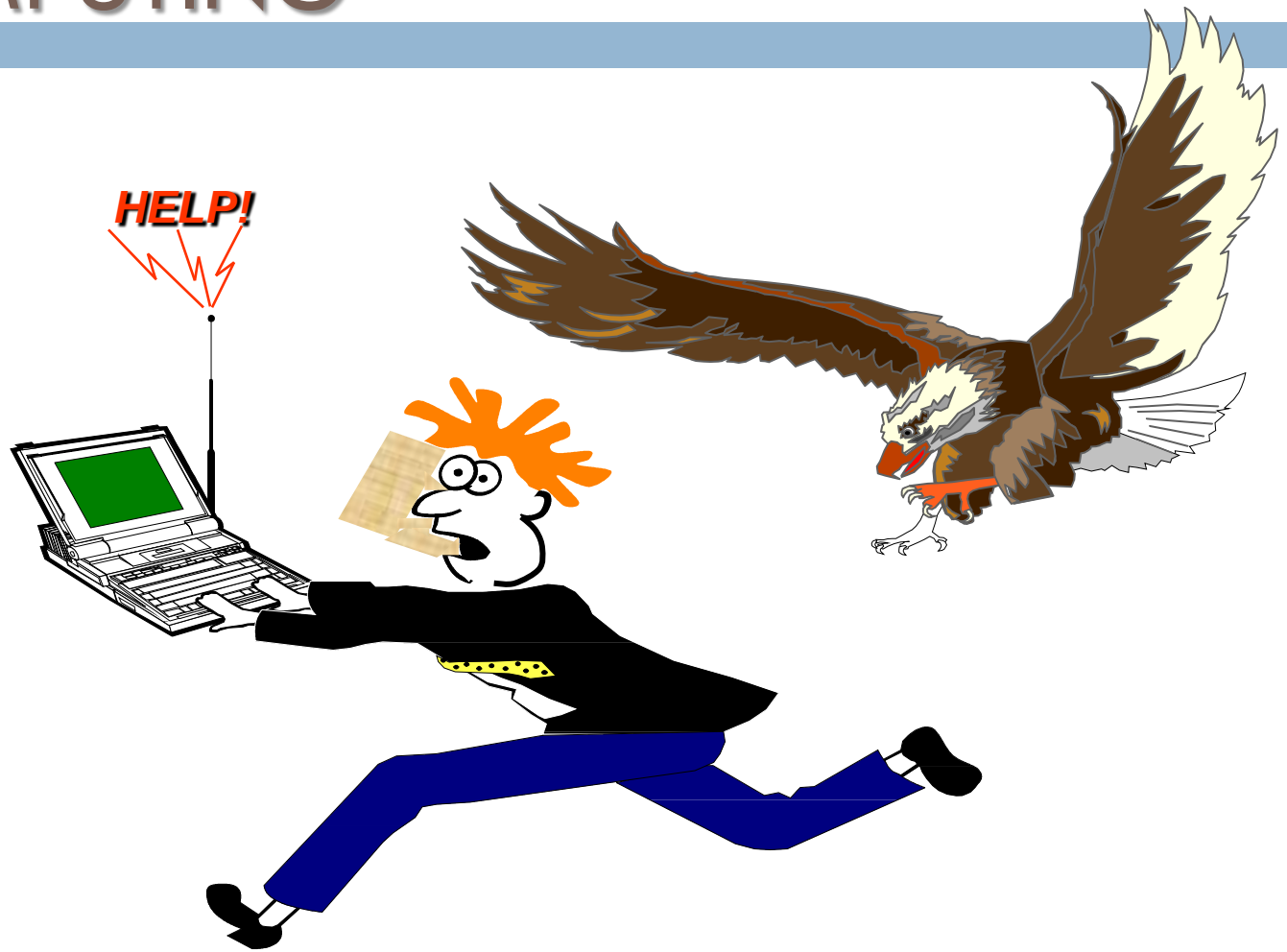
PERVASIVE DATA MANAGEMENT

DATA MANAGEMENT IN MOBILE SYSTEMS

Prof. Fabio A. Schreiber
Dipartimento di Elettronica, Informazione e Bioingegneria
Politecnico di Milano

MOBILE, CONTEXT- AWARE COMPUTING



NETWORKS FOR MOBILITY

- DIRECT RADIO LINKS
- CELLULAR NETWORKS
- INFRARED LINKS
- BLUETOOTH

DATA ACCESS **ANYWHERE** AND **ANYTIME** FOR **2.3 BILLION PEOPLE** (2011 estimate)

MOBILE COMPUTING

- **NOMADIC** COMPUTING

  FIXED NETWORK INFRASTRUCTURE

- **AD-HOC** COMPUTING

  THE NETWORK IS TEMPORARILY ESTABLISHED ON DEMAND AMONG THE INVOLVED DEVICES

MOBILE APPLICATIONS

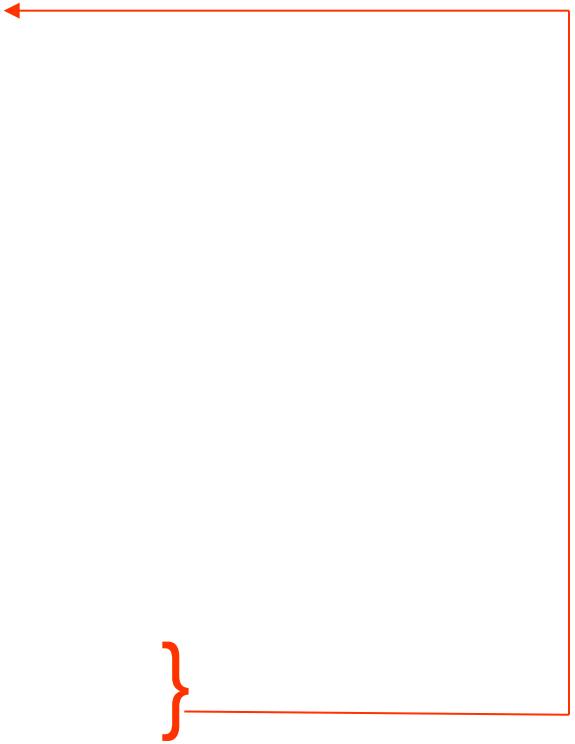
- TRADITIONAL APPLICATIONS ON MOBILE DEVICES
- NEW APPLICATIONS EXPLOITING THE DYNAMIC FEATURES OF THE MOBILE ENVIRONMENT TO PROVIDE CONTEXT-AWARE FEATURES
  - LOCATION-BASED SERVICES AND LOCATION-DEPENDENT QUERIES
    - EMERGENCY SERVICES
    - NAVIGATION AND INFORMATION SERVICES
    - LOCATION-DEPENDENT ADVERTISING
    - TRACKING SERVICES

RELATED TECHNIQUES

- CONTEXT AND ARCHITECTURE
  - MODEL AND MANAGE THE MOVING OBJECTS
    - **MOVING OBJECTS DATABASES** EXTEND THE TRADITIONAL DBMS WITH MODELS AND INDEX STRUCTURES TO EFFICIENTLY TRACK THE MOVING OBJECTS
    - **DATA STREAMS** PROCESS CONTINUOUS UPDATES OF OBJECTS POSITION
- QUERY PROCESSING TECHNIQUES
  - MOVING / STATIC RANGE QUERIES
  - NEAREST NEIGHBOUR QUERIES
  - QUERIES ON LOCATION-DEPENDENT DATA

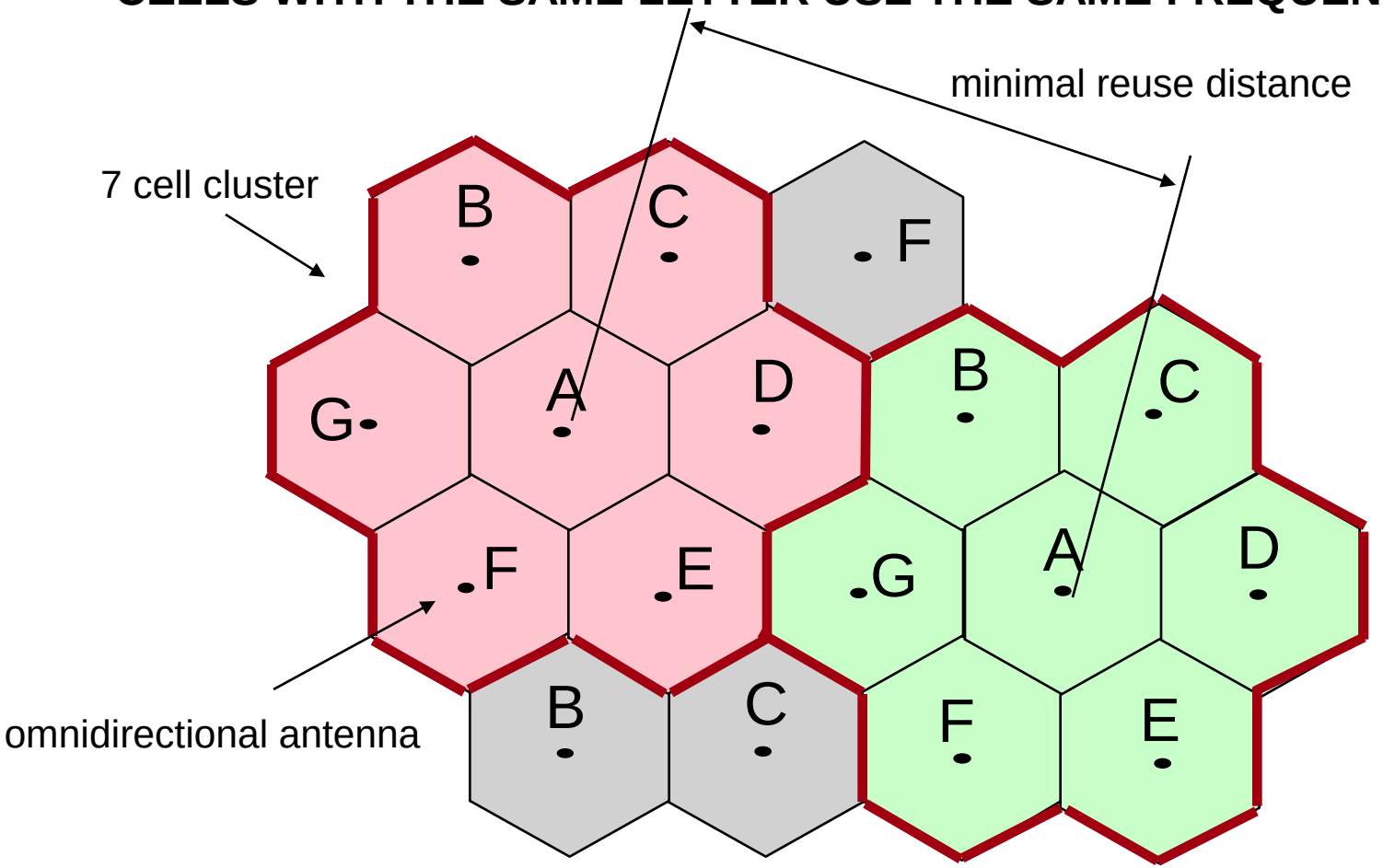
PORTABLE VSDBMS TECHNOLOGICAL CONSTRAINTS

- COMPUTATIONAL POWER
  - SMALL FOOTPRINT DBMS
- ENERGY CONSUMPTION
  - LOW POWER BATTERIES
- FREQUENT DISCONNECTIONS
  - AREA COVERAGE
  - DEVICE SWITCH-OFF
  - ……
- PERSISTENT DATA STORAGE
  - FLASH MEMORY TECHNOLOGIES
    - BIT/BYTE ACCESS GRANULARITY
    - BLOCK EREASE/UPDATE GRANULARITY
    - UPPER LIMIT TO ERASURE NUMBER

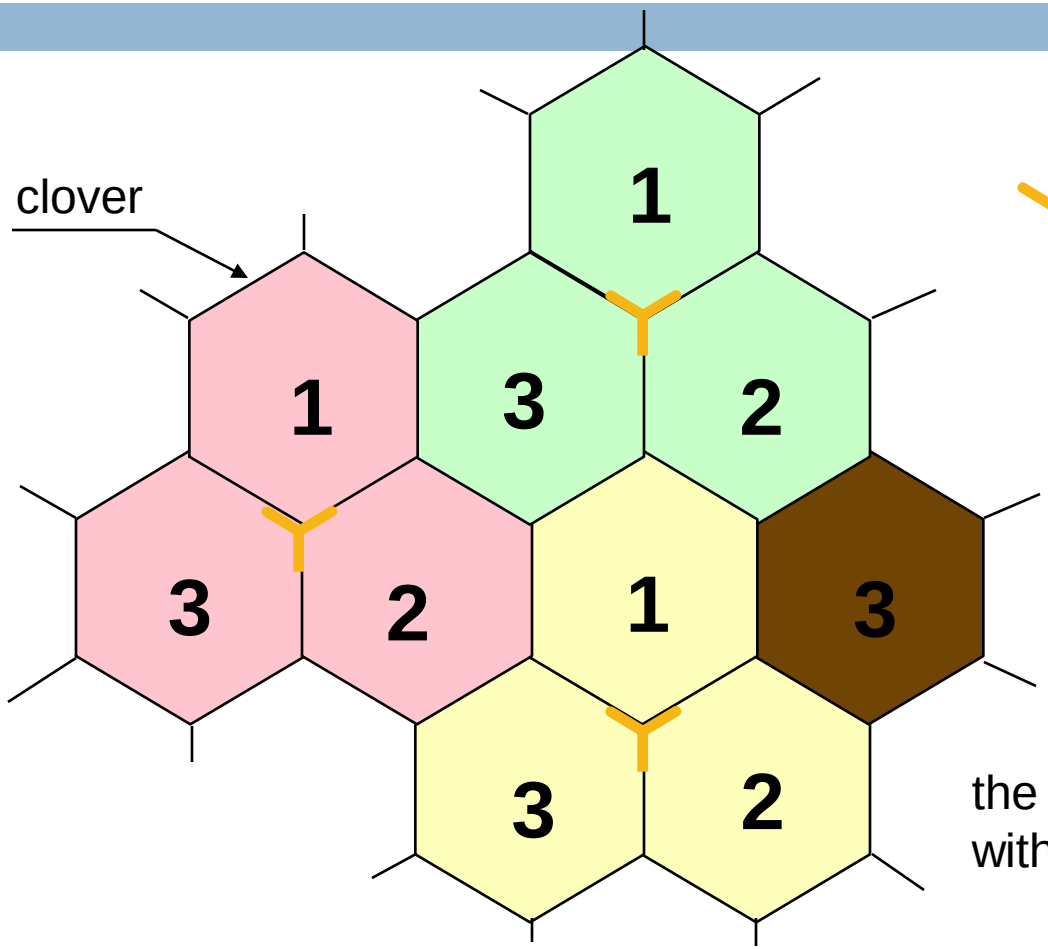


CELLULAR NETWORKS

CELLS WITH THE SAME LETTER USE THE SAME FREQUENCY SET



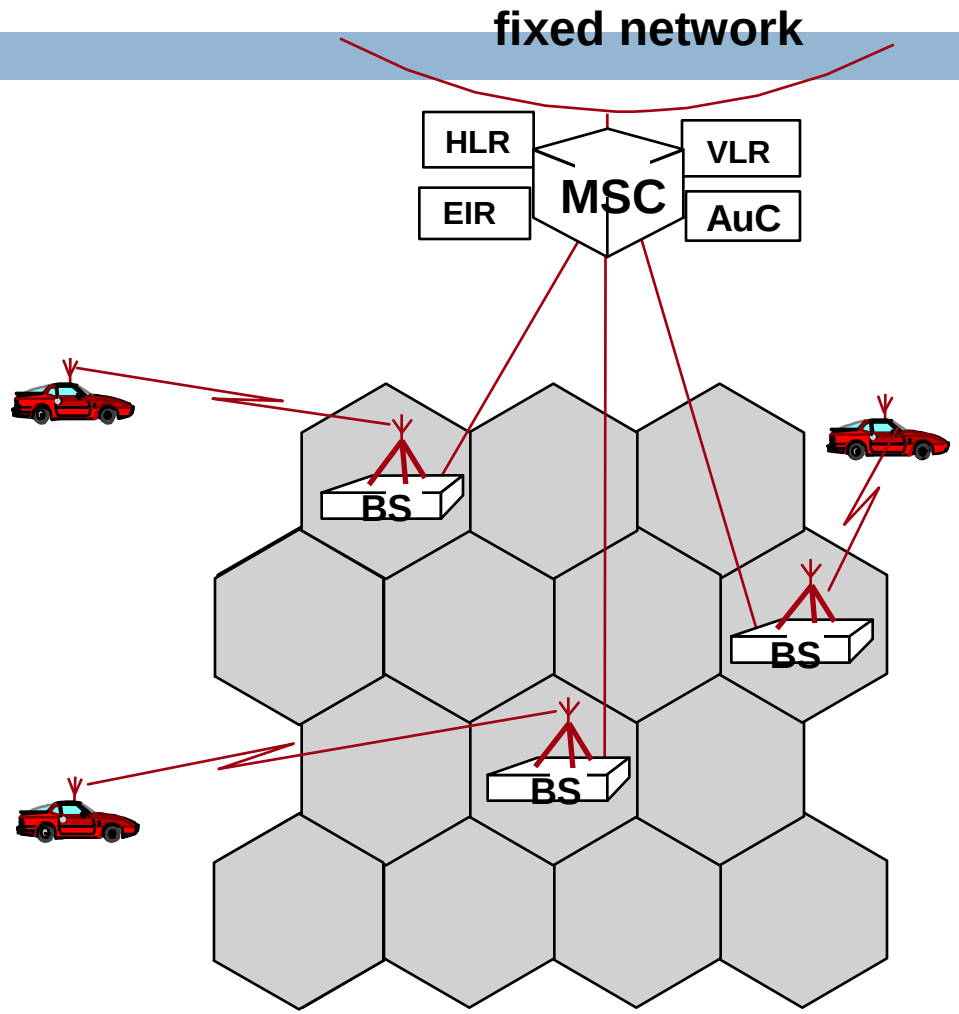
CELLULAR NETWORKS



directional antenna

the cluster includes 7 clovers
with a total of 21 cells

GSM NETWORK ARCHITECTURE



LEGENDA

Cell

Mobile Station

BS: Base Station

MSC: Mobile Switching Center

HLR: Home Location Register.

VLR: Visitor Location Register

AuC: Authentication Center.

EIR: Equipment Identification Register.

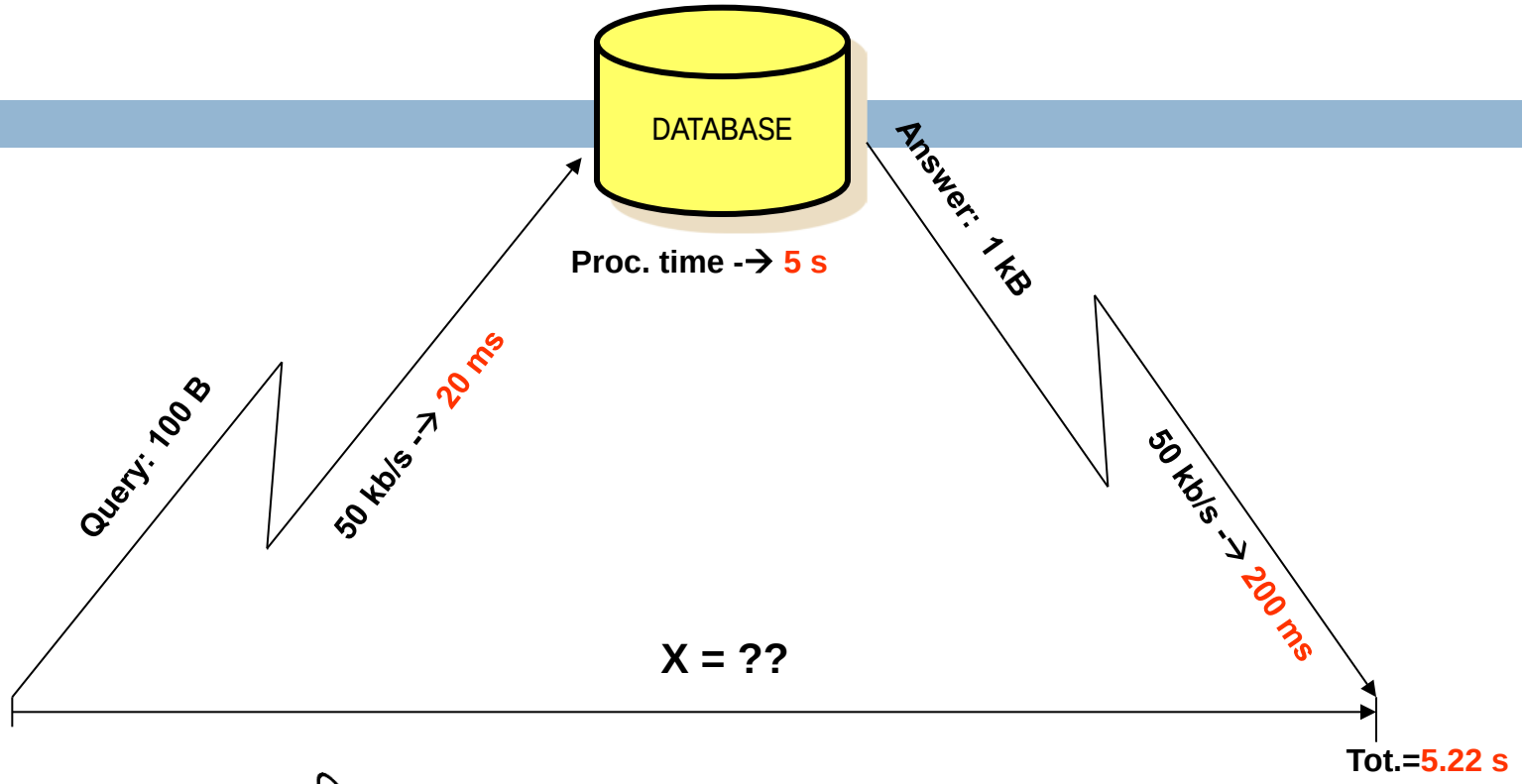
WHAT IS MOBILE

- SPATIO/TEMPORAL DATABASES FOR **MOBILE OBJECTS**
- TRADITIONAL DATABASES FOR **MOBILE DEVICES**
  - ON BOARD DBMS
  - LOCATION-DEPENDENT DATA
- BOTH

MOBILITY ISSUES IN INFO SYSTEMS

- WHICH IS THE APPLICATION **WORKING MODE**
  - WITH RESPECT TO SPACE
    - **LOCATION DEPENDENT DATA**
    - **LOCATION AWARE TRANSACTIONS**
  - WITH RESPECT TO SPACE-TIME
    - SENSIBLE OF EXTERNAL EVENTS (ACTIVE)
    - NOMADIC ROAMING (QUASI STATIC)
    - FULL-FEATURED MOBILE (REAL-TIME)
      - **QUERY ANSWERING SEMANTICS (????)**
        - **^ AS OF WHERE ISSUED**
        - **^ AS OF WHERE PROCESSED**
        - **^ AS OF WHERE RECEIVED**
- ACCURATE AND COHERENT SPATIO-TEMPORAL VIEW OF THE DISTRIBUTED STATE AND OF THE QoS

AN EXERCISE



1000 km/h → 280 m/s → 1500 m

100 km/h → 28 m/s → 150 m

5 km/h → 1.4 m/s → 7.3 m

MOBILE DEVICES

FEATURES OF MOBILE DEVICES

- TYPE AND POWER OF THE DEVICE (smart cards, cell phones, PDAs, portable PC, ...)
- OPERATING ENVIRONMENT VARIABILITY (proprietary, intranet/internet, ...)
- NETWORK CONNECTIVITY (guided or occasional disconnections, asymmetrical link bandwidth)
- ACCURATE AND COHERENT SPATIO/TEMPORAL PERCEPTION OF SERVICE STATE AND QUALITY (QoS)
- MULTICANALITY (shape the information content for the device it is aimed at)

MOBILE SYSTEMS CONSTRAINTS

DEVICE

POWER NEEDS FOR TRANSMISSION >>
POWER NEEDS FOR PROCESSING

NETWORK

LOW BANDWITH AND
UNRELIABLE LINKS

MATERIALIZE DATA ON THE MOBILE DEVICE

MOBILE OBJECTS DATABASES

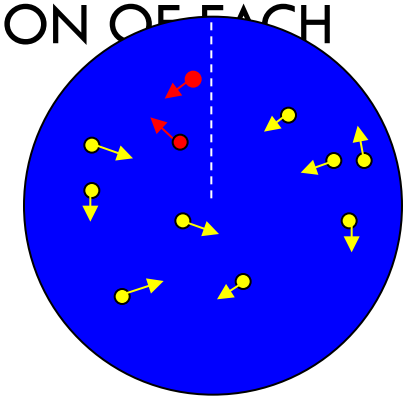
MOVING OBJECTS

- HIGH UPDATE RATE
  - ACTIVE (FAST) OBJECTS
    - STORED IN MAIN MEMORY (HARD RT)
  - INACTIVE (SLOW) OBJECTS
    - STORED ON DISK (FIRM/SOFT RT)

MOBILE OBJECTS DATABASES

EXAMPLE APPLICATIONS

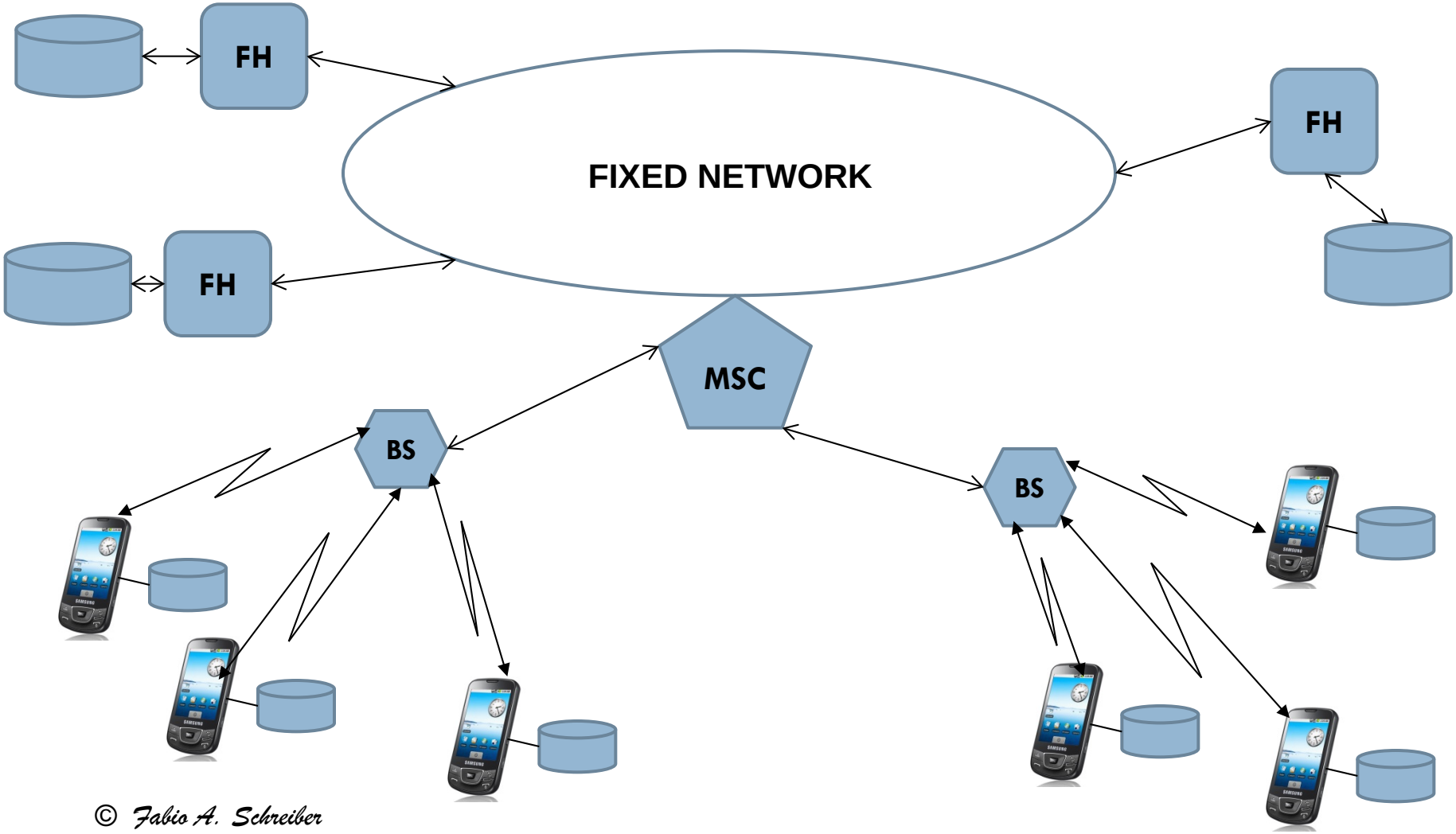
- CELLULAR PHONE NETWORKS
  - UPDATE THE USER POSITION (CELL)
  - LOCATE A USER (CELL) TO FORWARD A CALL
- AIR TRAFFIC CONTROL DATA
  - CONTINUOUSLY TRACK THE POSITION OF EACH OBJECT IN A GIVEN AREA



DATABASES FOR MOBILE DEVICES

- MOBILE USERS – FIXED HOST(S)
  - TRAVELLING EMPLOYEES → CORPORATE DB
  - MOBILE USERS → DATA SERVICE PROVIDER
  - PUBLIC BROADCASTED INFORMATION
- MOBILE HOST(S) – FIXED USERS
  - WIRELESS SENSORS → BASE STATION
- BOTH MOBILE
  - PEER-TO- PEER NETWORKS
  - PORTABLE PERSONAL FOLDERS

DATABASES FOR MOBILE DEVICES



DATABASES FOR MOBILE DEVICES

WHAT IS RESIDENT ON THE MOBILE DEVICE

- LOCAL DATA ONLY
  - **smart cards personal data**
- SMALL LOCAL PERMANENT STORAGE
  - **old cellular phones directories**
- STORAGE AND TRANSACTION MANAGEMENT CAPABILITIES
  - **smart cellular phones**
  - **palm and portable PCs**

MOBILE DB TECHNOLOGICAL ISSUES

DISCONNECTED OPERATION

- DATA SYNCHRONIZATION
  - FILE SYNCHRONIZERS (PRODUCTS)
    - MANUAL CONFLICT RESOLUTION (e.g., MS ActiveSync)
  - TRANSFORMATIONAL APPROACH (RESEARCH)
    - OPERATIONS AT EACH SITE ARE BROADCASTED TO OTHER SITES WHERE THEY ARE INTEGRATED AND TRANSFORMED FOR THE LOCAL EXECUTION CONTEXT

MOBILE DB TECHNOLOGICAL ISSUES

LONG TRANSACTIONS

- PESSIMISTIC APPROACHES
  - TOO LONG DELAYS IN COMMITTING
- OPTIMISTIC APPROACHES
  - TOO LARGE DIVERGENCES AMONG LOCAL AND SERVER COPIES

- DOUBLE COMMIT (**RESEARCH**)
  - **LOCAL**, IN DISCONNECTED MODE
  - **GLOBAL**, AT RECONNECTION TIME
    - CONTRACT AND OPERATIONAL RULES FOR GOVERNING THE REINTEGRATION OF LOCALLY MODIFIED COPIES

QUERIES IN MOBILE OBJECTS DATABASES

COORDINATE-BASED QUERIES

- SPATIO/TEMPORAL

  Find all objects within a given area during a given time interval

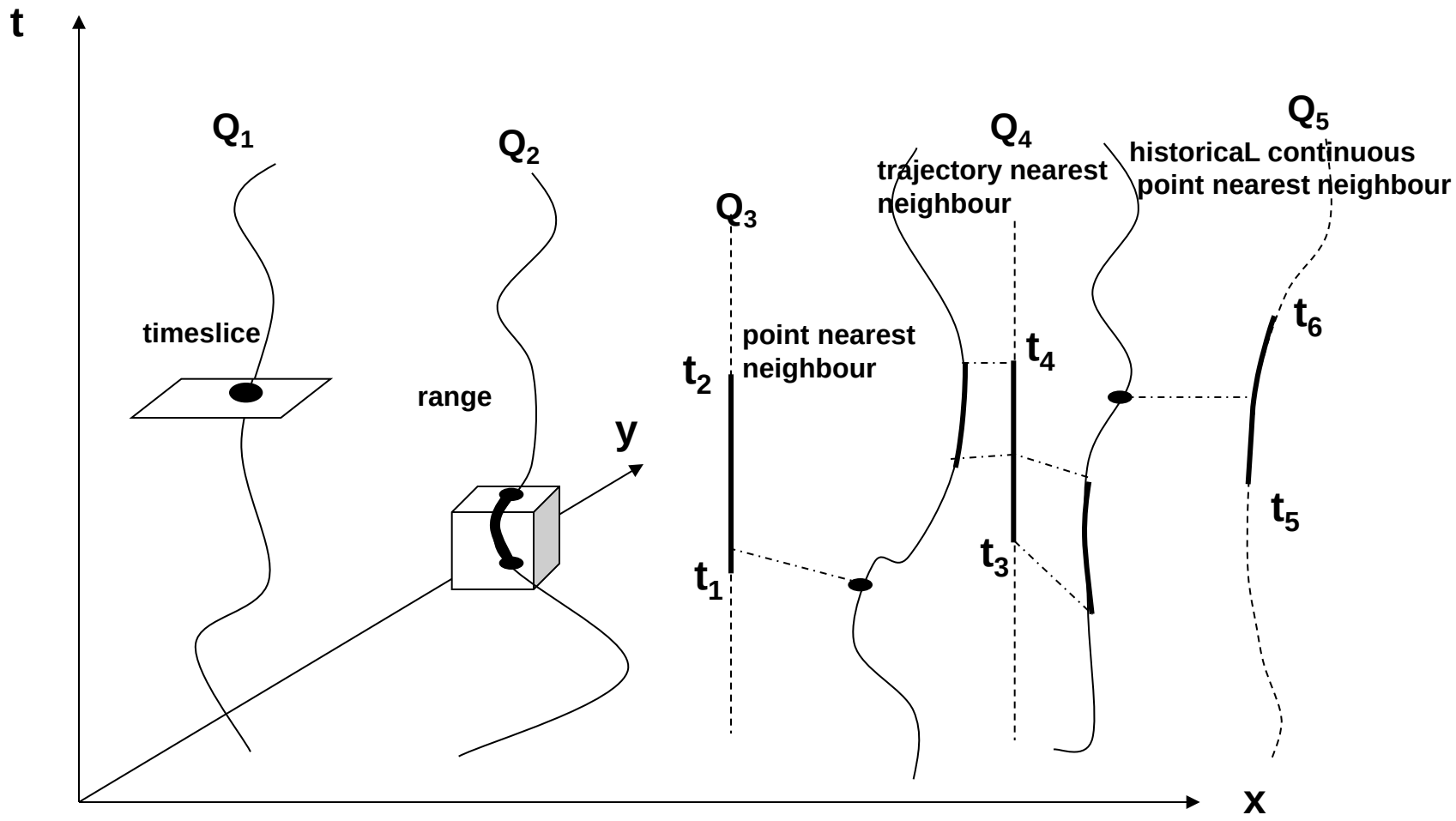
- TIMESLICE

  Find all objects' locations within a given area at a certain time instant

- NEAREST NEIGHBOUR

  Find the nearest moving object to a given object during a given time interval

QUERIES IN MOBILE OBJECTS DATABASES



QUERIES IN MOBILE OBJECTS DATABASES

TRAJECTORY-BASED QUERIES

- TOPOLOGICAL QUERIES
  - ENTER, LEAVE … A GIVEN AREA
- NAVIGATIONAL QUERIES
  - AVERAGE OBJECT SPEED, COVERED DISTANCE …

COMBINED QUERIES

- Which trajectories followed, in the next hour, the objects leaving via Ponzio 34 between 5 p.m. and 8 p.m. today
- Which is the nearest gas station at each instant during my whole trip

QUERIES IN MOBILE OBJECTS DATABASES

INDEXES

- COMPRESSION AND INDEXING TECHNIQUES FOR EFFICIENT TRACKABILITY
  - IN THE PAST (TRAJECTORIES HISTORY)
  - IN THE FUTURE (ACTUAL POSITION AND MOTION VECTOR)

A WHOLE AMAZONIA FOREST OF DIFFERENT TREE STRUCTURES HAS BEEN PROPOSED IN THE LITERATURE

NEW APPLICATION ARCHITECTURES

PEER-TO-PEER

- INFORMATION SOURCES COINCIDE WITH USER DEVICES
  - PURE P2P: FLAT ARCHITECTURE, NO DIFFERENCE AMONG THE NODES (Gnutella)
  - HYBRID AND MIXED P2P: FLAT ARCHITECTURE FOR DATA, CLIENT-SERVER ARCHITECTURE FOR META-DATA (Napster, eMule, BitTorrent)
  - USED FOR MULTIMEDIAL INFORMATION EXCHANGE
- CONTEXT-AWARENESS: ESSENTIAL IN SOLVING PARTNERS HETEROGENEITY
- ONTOLOGY: ESSENTIAL IN
  - INFORMATION SEARCHING
  - FORMULATING QUERIES AGAINST UNKNOWN PARTIES
  - RECONCILING ANSWERS FROM DIFFERENT SOURCES
- MOBILITY: NOT A NECESSARY ISSUE

EXAMPLES OF COMMERCIAL SYSTEMS (1)

SYSTEM NAME	PLATFORM SUPPORT	APPLICATION AREA	ARCHITECTURE / DATA MODEL	CONCURRENCY CONTROL	RECOVERY MANAGEMENT	ACCESS PATH
AT&T Bell Labs Dalì (DataBlitz) www.bell-labs.com/project/dali/	Windows 9x/NT, UNIX, Solaris	telecommunications	MM storage mngr direct access to data via shared memory relational API	multi-granularity locking s, x,, i, mode locking	•multi-level recovery •fuzzy checkpoints of dirty pages	•T-trees indexing •extensible hash tables
Centura RDM (Mbrane Touchpoint EC) www.centurasoft.com	a variety of RTOS	e-business, web+mobile solutions	RT DBMS library of C routines relational and network			Linked lists (one-to-many DBTG sets)
EMPRESS www.empress.com	Windows 9x/NT, UNIX, Solaris, Linux, Lynx, ...	RT process control network applications medical systems	distributed DBMS relational			
FastDB www.ispras.ru/~knizhnik/ fastdb.html	Windows NT, UNIX, Linux, Solaris	general applications with dominated read access pattern	MM DBMS direct data access object oriented	shadow objects	•shadow indexes •linked lists	•T-trees indexing •extensible hash tables •inverse references
Informix Cloudscape www.cloudscape.com	Windows 98/NT WinCE, Solaris, UNIX, Linux, EPOC, RTOS, ...	portable/palm mobile web solutions	distributed DBMS library of Java clas-ses object-relational	row- or table-level locking		B-trees indexing
POET FastObjects Navajo www.poet.com/fastobjects	Windows 9x/NT, UNIX, Linux	general RT and telecom applications	RT DBMS dynamically loadable Java modules object oriented			
Polyhedra www.polyhedra.com	Windows 9x/00/NT, UNIX, Linux, pSOS, VxWorks	complex RT applica-tions	MM RT Active DBMS client/server object-relational		•snapshots to disk •write-behind logging •duplicated DB (opt)	•indexes •direct pointers (?)
Sleepycat Sw Berkeley DB www.sleepycat.com	Windows 9x/00/NT, UNIX, Embedix, QNX, VxWorks		small footprint transactions library of modified UNIX modules	variable granule size locking	•write-ahead logging •checkpointing	platform's native file system

SPECIALLY SUITED FOR MOBILE APPLICATIONS

EXAMPLES OF COMMERCIAL SYSTEMS (2)

SYSTEM NAME	PLATFORM SUPPORT	APPLICATION AREA	ARCHITECTURE / DATA MODEL	CONCURRENCY CONTROL	RECOVERY MANAGEMENT	ACCESS PATH
Solid Embedded engine www.solidtech.com	Windows 9x/00/NT, UNIX, Linux, Solaris, VxWorks, ...	smart telecom networks	distributed active DBMS relational	•multiversion optimistic •row-level locking	roll forward log	•modified B-trees •bonsai-trees
Sybase SQL Anywhere UltraLite (50 KB) www.sybase.com	Windows NT/XP WinCE, UNIX, Linux EPOC, VxWorks, ...	mobile business applications	MM DBMS relational	centralized conflict resolution	checkpoints and logs	B^+-trees
Times Ten www.timesten.com	Windows 2000/NT, UNIX, Linux, Lynx	Internet application servers, mobile business applications	MM (RT) DBMS applications linkable libraries relational	row-, table-, db-level locking	•checkpoints and logs •data replication	•T-trees indexing •hash tables
VTT Inform. Technology RapidBase www.vtt.fi/tte/projects/rapid	Windows 95/98/NT, UNIX, Linux	Storage/retrieval of time series data for industrial process measurement	MM, active DBMS client/server C^{++}, Java libraries relational		checkpoints and logs	
IBM DB2 Everyplace www-106.ibm.com/developerworks /library/wi-everyplace	Windows 32 /CE, Palm OS, EPOC 5, Linux, QNX Neutrino	Mobile and embedded devices, special purpose applications	Relational database (reduced functions, 137k footprint)	Synchronization server among different platforms		Advanced indexing
Oracle9i Lite www.oracle.com/ip/deploy/ias/mobile/	Windows 32, Windows CE, Palm OS, EPOC 5	General purpose mobile and location-aware applications	Lightweight relational dabase	Two-way synchronization between the server and the mobile devices		

SPECIALLY SUITED FOR MOBILE APPLICATIONS

MOST RECENT ISSUES: *BIG DATA*

- THE **CAP** THEOREM

  ANY NETWORKED SHARED-DATA SYSTEM CAN HAVE AT MOST TWO OUT OF THREE DESIRABLE PROPERTIES

  - CONSISTENCY (C)
  - HIGH AVAILABILITY (A)
  - TOLERANCE TO NEWORK PARTITIONS (P)
    - **PARTITION ≡ TIME BOUND ON COMMUNICATION LATENCY**
- **NoSQL** DATABASES
  - SUPPORT FLEXIBLE SCHEMA (e.g.: Key – Value tables)
  - SCALE HORIZONTALLY
  - DO NOT SUPPORT ACID PROPERTIES
    - UPDATES PERFORMED ASYNCHRONOUSLY
    - POTENTIAL DATA INCONSISTENCY RESOLVED BY READERS

HOMEWORK DEEPENING TOPICS 3

- INDEXING OBJECTS IN MOBILE DATABASES
- TRANSACTION MANAGEMENT IN MOBILE DATABASES (COMMIT, UPDATE, SYNCHRONIZE)
- QUERY PROCESSING IN MOBILE DATABASES
- P2P DATA MANAGEMENT IN MOBILE DATABASES