PERVASIVE DATA MANAGEMENT

PERVASIVE DATA MANAGEMENT: SUMMARY AND OPEN PROBLEMS

Prof. Fabio A. Schreiber
Dipartimento di Elettronica e Informazione
Politecnico di Milano
THE NEW APPLICATION ENVIRONMENT

• **A VERY LARGE NUMBER** OF LARGE DATA SOURCES

• **GENERALLY HIGHLY VARIABLE AND VOLATILE** DATA (ES. WEB)

• **HIGHLY HETEROGENEOUS** DATA SOURCES

• **DIFFERENT DATA STRUCTURING** LEVELS
  – DATABASES WITH **DIFFERENT UNDERLYING MODELS** (RELATIONAL, OBJECT ORIENTED, LEGACY…)
  – **SEMI-STRUCTURED DATA** (XML, HTML, OTHER TAGGING SYSTEMS…)
  – **NON-STRUCTURED DATA** (TEXT, IMAGE, SOUND, ETC…)

• **DIFFERENT TERMINOLOGIES AND CONTEXTS**
AUTONOMIC PERVERSIVE SYSTEMS

- CONTEXT-AWARE
- REACTIVE
- SELF-ADAPTING
PERVASIVE IS: ADAPTIVITY

• **HIGHLY DYNAMIC ENVIRONMENT**
  – RESOURCE DISCOVERY
  – DATA SEMANTICS ACROSS HETEROGENEOUS SOURCES
  – DYNAMIC OPTIMIZATION
  – CHANGE OF CONTEXT

• **CHANGING USER ROLES AND NEEDS**
  – ASSESSING INSTANTANEOUS USER NEEDS/INTERESTS
    • USER **PREFERENCES**
    • MODELING AND EXPLOITING USER CHARACTERISTICS
      • USER **PROFILE**
PERVASIVE IS: METADATA

TO EXPLOIT THE AUTONOMIC FUNCTIONALITIES, THE SYSTEM USES METADATA

– **CONTEXT** DATA
  • STATIC FEATURES AND PROPERTIES
  • INSTANTANEOUS SITUATIONS OF THE OPERATING ENVIRONMENT

– **PERSONALIZATION**
  • PERSONAL CHARACTERISTICS AND BEHAVIOUR
  • PREFERENCES OF INDIVIDUAL USERS (HIGHLY PRIVACY SENSITIVE)
CONTEXT-AWARENESS The six "W" questions

1. **What** is context?
2. **Who** might benefit from an awareness of their context; whose context is important to whom, or what?
3. **Where** can an awareness of context be exploited?
4. **When** is context-awareness useful?
5. **Why** are context-aware applications useful?

Answers to these five questions underpin the higher level, meta-question of:

6. **How** do we implement context-awareness so that we can develop context-aware applications?

CONTEXT AWARENESS

ABSTRACT SERVICES TO ASSIST THE PROGRAMMER IN IMPLEMENTING C-A APPLICATIONS

• CONTEXT SUBSCRIPTION AND DELIVERY
  – A SERVICE THAT CAN NOTIFY A COMPONENT WHEN AN EVENT OCCURS

• CONTEXT QUERY
  – A MECHANISM TO FIND A SUITABLE INFORMATION OR SERVICE

• CONTEXT TRANSFORMATION
  – CONVERSION OF LOW-LEVEL DATA INTO HIGH-LEVEL INFORMATION

• CONTEXT SYNTHESIS
  – AGGREGATION OF CONTEXT INFORMATION TO GENERATE A MORE PRECISE OR DETAILED CONTEXT

from da Costa et Al. 2008
PERVASIVE IS: DATA MANAGEMENT

HETEROGENEOUS DATA SOURCES

• A LARGE NUMBER OF DATA SOURCES
• TIME-VARIANT DATA
• MOBILE, TRANSIENT DATA SOURCES
• MOBILE USERS
• DIFFERENT LEVELS OF DATA STRUCTURE
  – Databases (relational, OO…)
  – Semistructured datasources (XML, HTML, more markups …)
  – Unstructured data (sensors, text, multimedia etc…)
• DIFFERENT TERMINOLOGIES AND DIFFERENT OPERATION CONTEXTS
CONTEXT MODELS AND USAGE

THIS TOPIC BELONGS TO PROF. TANCA’s LECTURES
The PerLa system is basically composed of three components:

• The **NODES**
  – heterogeneous devices equipped with sensors (such as RFID tags or WSN nodes such as MOTES) or more complex devices (such as palms, portable computers or ad hoc boards).

• The **MIDDLEWARE**
  – is a stack of software layers providing a high level abstraction of each node (called LOGICAL OBJECT).
  – implements a set of functionalities to allow communications among logical objects and to manage devices that enter and leave the system (following a "Plug and Play" behaviour).

• The **LANGUAGE**
  – a full declarative high level language that allows to query a pervasive system, hiding the difficulties related to the need of handling different technologies.
  – provides a database like abstraction of the whole network in order to hide the high complexity of low level programming and allows users to retrieve data from the system in a fast and easy way.

http://perlawsn.sourceforge.net/index.php
PERVASIVE DATA MANAGEMENT

THE **MoGATU** FRAMEWORK (Perich et Al. 2004) ABSTRACTS ALL **DEVICES** IN TERMS OF

– **INFORMATION PROVIDERS**
  
  HOLD (HETEROGENEOUS) FRAGMENTS OF DATA ANNOTATED IN A SEMANTIC LANGUAGE

– **INFORMATION CONSUMERS**
  
  HUMAN OR AUTONOMOUS SW AGENTS WHO QUERY AND UPDATE DATA

– **INFORMATION MANAGERS** (ONE PER DEVICE)
  
  RESPONSIBLE OF THE NETWORK COMMUNICATIONS AND MOST DATA MANAGEMENT FUNCTIONS WHICH ARE **EQUAL SEMIAUTONOMOUS PEERS** GUIDED IN THEIR INTERACTION BY CONTEXT AND PROFILES
OTHER PERVERSIVE DATA MANAGEMENT SYSTEMS

• **AURA**

• **CoolTown**

• **DSN**

• **GAIA**

• **GSN**

• **SECAS**
HOMEWORK DEEPENING TOPICS 4

- ABOUT CONTEXT AWARENESS (Prof. Tanca)

- ABOUT DATA INTEGRATION (Prof. Tanca)

- IN-DEPTH DESCRIPTION OF A PERVERSIVE DATA MANAGEMENT SYSTEM
AUTONOMY

- AUTOMATIC CONTROL SYSTEMS

*Diagram showing a feedback control system with components: Controller, Controlled System, Feedback Transducer.*
PERVASIVE INFORMATION SYSTEM MODEL

CONTEXT MODEL (CDT)

APPLICATION VALUED DIMENSIONS

ENVIRONMENT VALUED DIMENSIONS

DATA MODEL (E-R/RELATIONAL)

DATA TAILORING

QUERY MODEL (XML)

QUERY STEREOTYPE

QUERY TAILORING

DATA CLEANING, NOISE FILTERING, DAMPING

OPERATIONAL SYSTEM

CONTINUOUS QUERIES

APPLICATION

VALUES SET PATH

ENVIRONMENT VALUES FEEDBACK PATH

QUERY/RESULT PATH
A PERVASIVE INFORMATION SYSTEM IS A DYNAMIC SYSTEM POSSIBLY COMPOSED OF HUNDREDS OR THOUSANDS OF NODES AND MULTIPLE FEEDBACK PATHS EXIST BETWEEN THE FIELD (SENSORS) AND THE CONTROL LOGIC (CONTEXT AND MIDDLEWARE)

IT IS IMPOSSIBLE TO TEST A REAL PROTOTYPE OF THE SYSTEM AT DESIGN TIME

SCALABILITY THUS BECOMES AN IMPORTANT ISSUE WHEN DESIGNING SYSTEM CONTROL ALGORITHMS

SYSTEM STABILITY ALSO IS AFFECTED BY MANY FACTORS AND MUST BE CHECKED AND CONTROLLED
SOME R&D OPEN PROBLEMS (2)

ALGORITHM COMPLEXITY (SCALABILITY)

THE MEMORY AND TIME BEHAVIOUR OF ALGORITHMS CANNOT BE INFERRED BY TESTING SYSTEMS COMPOSED OF ONLY A BOUNCE OF NODES: CONSTANTS MATTER!

\[ O(n), O(n^2), O(2^n), O(\log n) \]
SOME R&D OPEN PROBLEMS (3)

SYSTEM STABILITY

NO GENERAL METHOD EXISTS COMPARABLE TO WHAT IS DONE IN CONTINUOUS SYSTEMS DESCRIBED BY DIFFERENTIAL EQUATIONS (root locus plot, phase plan analysis, catastrophe theory, … )

– NOISE IN THE SENSORS OUTPUT
  • CLEANING AND FILTERING

– TIME CONSTANTS IN THE FEEDBACK PATHS
  • RELATIVE COMPATIBILITY CHECKS???

– ERRORS IN THE RULES DEFINING THE CONTEXT (AND THE CONTEXT CHANGES)
  • MODEL CHECKING???
SOME R&D OPEN PROBLEMS (3)

SIMULATION Vs. TESTBEDS

- **SIMULATION** often provides **imprecise** results owing to many parameters of the system which are not accounted for by the simulation programs

- **TESTBEDS** with a very large number of components are very **difficult**, if not impossible, to organize

• Use testbeds to tune and calibrate simulation programs???
SOCIAL IMPACT OF PDM

- MASSIVE DATA COLLECTED BY SENSORS EVEN AT PEOPLE UNAWARENESS (THE «BIG BROTHER» EFFECT)
- POOR CONTROL RULES AND TECHNIQUES ON THE QUALITY OF COLLECTED DATA
- PRIVACY
  - NO OR POOR RULES ON DATA PROPERTY AND USAGE
  - ETHICAL ASPECTS
  - LEGAL ASPECTS
    - PHISHING
    - STALKING
INFORMATION PRIVACY

© Fabio A. Schreiber
INFORMATION PRIVACY

• THE CITIZEN POSSIBILITY/RIGHT TO CONTROL THE CONSISTENCY AND THE USE OF HIS/HER OWN PERSONAL DATA

• CONTROL/LIMITATION OF DATAVEILLANCE i.e. OF MONITORING OF PERSONAL
  – ACTIVITIES
  – HABITS
  – BELIEF

BY MEANS OF TRANSACTIONS AND DATA TRACES
PRIVACY DIMENSIONS

• PHYSICAL PERSON
  – OBLIGATION/DENIAL TO VACCINATE
  – BODY INSPECTIONS

• PERSONAL BEHAVIOUR
  – POLITICAL/RELIGIOUS BELIEFS
  – ALCOHOL/DRUGS ABUSE
  – SEXUAL HABITS

• PERSONAL COMMUNICATIONS
  – MAIL SECRECY
  – TELECOM LINES PROTECTION

• PERSONAL DATA
  – DATABASES IN INFORMATION SYSTEMS
PRIVACY DIMENSIONS

THE THREE LAST ELEMENTS ARE RELEVANT TO INFORMATION SYSTEMS AND CAN BE MUTUALLY INFERRED
PRIVACY RULES

• LIMITATIONS TO PERSONAL DATA COLLECTION

• LIMITATIONS TO THE LEGAL USE OF COLLECTED DATA, THE OWNER CONSENT IS REQUIRED FOR

• SAFEGUARD AGAINST NON AUTHORISED
  – ACCESS
  – DIFFUSION
  – USE
OF COLLECTED DATA
PRIVACY RULES

TRANSACTIONS ANONYMITY LEVELS

- **IDENTIFIED**: READILY RELATED TO A PARTICULAR INDIVIDUAL
  - CREDIT CARDS
- **ANONYMOUS**: TOTAL ABSENCE OF IDENTIFICATION DATA
  - PRE-PAID CARDS (TELEPHONE, …)
- **PSEUDONYMOUS**: ANONYMITY ASSURED BY ALIASES AND BY A THIRD PARTY WHO KNOWS THE PARTNERS IDENTITY
INFORMATION PRIVACY

EVEN IF MOST OF THESE ISSUES ARE A MATTER FOR JURISTS AND SOCIOLOGISTS,
AN ENGINEER MUST BE AWARE OF THE POSSIBLE CONSEQUENCES OF HIS/HER WORK
PRIVACY MANAGEMENT

• LOGICAL DESIGN OF THE DATABASE ACCESS PROFILES

• UNDERLYING PHYSICAL MECHANISMS FOR ENFORCING DATA PROTECTION
LOGICAL DESIGN OF THE DATABASE
ACCESS PROFILES

USER ROLES

• DATA OWNER
  – FULL READ/WRITE ACCESS PERMISSION TO OWNED DATA

• GUEST
  – READ ONLY ACCESS PERMISSION ON PART OF THE STORED DATA

• CARD HOLDER
  – A GUEST FOR ALL OF THE STORED DATA
  – AN OWNER OF PERSONALLY MANAGED DATA
PRIVACY MANAGEMENT

- CUSTOMISED OWNER-MULTIGUEST
  - HIGHLY SENSIBLE DATA
  - SPECIALISED USER ACCESS REQUIREMENTS

- GENERALISED OWNER-GUEST
  - YES/NO GENERIC GUEST ACCESS PERMISSION
CUSTOMISED OWNER-MULTIGUEST SCHEMA

(GROUPS OF) AUTHORITIES OR INSTITUTIONS CAN BE

– THE OWNER OF STORED DATA (R/W)
– A GUEST ACCESSING SPECIFIC DATA (R)
– UNLISTED GUESTS ACCESSING PUBLIC DATA (R)

MANAGEMENT COSTS

– VIEWS DEFINITION
– DEFINITION OF USERS GROUPS
– CUSTOMISATION OF ACCESS PERMISSIONS FOR USERS WITHIN GROUPS
CUSTOMISED OWNER-MULTIGUEST SCHEMA

• DEFINITION OF GROUPS AND OF USER ACCESS PERMISSIONS TO BE CARRIED OUT AT DATABASE DESIGN TIME

• LATE INTRODUCTION OR CHANGES TO CONSTRAINTS REQUIRE DATA REORGANISATION, A COSTLY OPERATION FOR FLASH MEMORIES

• UNLISTED GUESTS ALLOW LATE JOINING OF NEW USERS, TEMPORARILY “PARKING” THEM IN THIS CATEGORY UNTIL A CARD MAINTENANCE SESSION IS PERFORMED
CUSTOMISED OWNER-MULTIGUEST SCHEMA

AT DATABASE DESIGN TIME EACH DATA OWNER MUST SPECIFY

– THE OWNED (FULLY ACCESSIBLE) TABLES AND RELATIONS

– A SET OF VIEWS ON THEM

– FOR EACH VIEW, A LIST OF GUESTS WITH READ ONLY ACCESS PERMISSION
GENERALISED OWNER-GUEST SCHEMA

• GUESTS ARE INDISTINGUISHABLE
• AT DATABASE DESIGN TIME EACH DATA OWNER MUST SPECIFY
  – THE OWNED TABLES AND RELATIONS
  – A SET OF VIEWS ON THEM
  – GENERIC ACCESS PERMISSION FOR EACH VIEW

AS USUAL, USERS (NAMELY GUESTS) MUST UNDERGO THE AUTHENTICATION PHASE BEFORE ANY ACTION CAN BE UNDERTAKEN

© Fabio A. Schreiber

PDM - management 36
BIBLIOGRAPHICAL REFERENCES 1

• B. Babcock et Al. – Models and Issues in Data Stream Systems – Proc. ACM PODS, pp. 1-16, Madison 2002
• J. Hyde – Data in Flight – Communications of ACM, Vol. 53, n.1, Jan 2010
BIBLIOGRAPHICAL REFERENCES 2

• S. Ilarri, E. Mena, A. Illarramendi – Location-Dependent Query Processing: Where We Are and Where We Are Heading – ACM Computing Surveys, Vol. 42, n. 3, 2010


BIBLIOGRAPHICAL REFERENCES


• W. Xue, Q. Luo, L. M. Ni – *Systems Support for Pervasive Query Processing* – Proc. 25° Int. Conf. on Distributed Computing Systems (ICSCS05), IEEE, 2005