Data Management
and Ambient Intelligence

Thierry Delot
University of Valenciennes
Thierry.Delot@univ-valenciennes.fr

Sergio Ilarri
University of Zaragoza
silarri@unizar.es
General context

- **Mobile** Data Management & Query processing
- Communication infrastructure
  - Not always available
  - Direct interactions between mobile nodes
- Application to Vehicular Networks

- I don’t care! I’m driving in the opposite direction!
General context

- Recent development of:
  - Mobile devices
    - Sensors, smartphones, navigation devices, etc.
  - Wireless technologies with different ranges
    - Wi-Fi, 3G, etc.
  - Global Navigation Satellite Systems (GNSS)
    - GPS system

These mobile devices produce and/or store data!

« In 2014, the volume of mobile data sent and received every month by users around the world will exceed by a significant amount the total data traffic for all of 2008 » (ABI research)
Characteristics of mobile data

- Stamped data
  - Location-dependent, timestamped
- Personal data
- Uncertain data
- May be produced as streams
Impact on query processing

- New types of queries
  - Location-dependent queries (examples):
    - Continuous queries
    - Nearest neighbor queries
    - Range queries
    - Spatio-temporal queries
    - ...

- New processing techniques
  - Traditional techniques are no more suited
    - No placement schema as in distributed databases (example)
Impact on query processing

• Optimization objectives
  • Non classical optimization
    – Energy, financial cost, etc.
  • Local vs. global optimization

• Privacy and trust issues

• Even the notion of query result is different!
  – Open World Assumption vs. Close World Assumption
Constraints on query processing

- Different access models to consider
  - Pull vs. Push vs. Hybrid
- Different architectures to consider
  - (mobile) Client/Server
  - Hybrid P2P
  - Mobile P2P
Architectures: Mobile Client/Server
Possible issues

- Bandwidth limitations and scalability issues
- Connection not available everywhere
- Mobile telephony networks are not free
- Privacy preservation
Architectures: Alternative solutions

• Hybrid peer-to-peer architectures
  – Distinguishes mobile devices and traditional servers

• Mobile peer-to-peer architectures
  – Direct interactions between mobile devices
  – It does not require any fixed infrastructure
  – E.g., vehicular ad hoc networks (VANETs)
Mobile Query Processing
Vehicular Networks (VANETs)

Vehicle-to-vehicle (V2V) communications
Vehicle-to-infrastructure (V2I) communications

- GPS
- WIFI, UWB, IEEE 802.11p
- Storage card
- Smartphone, …
Using a push model

Where is the closest parking space?

Available parking spaces, traffic congestions, emergency brakings, ...

IEEE 802.11

Routing results towards a moving object is a (very) difficult task!
Delivery of query results

- How to route partial results towards the mobile recipient?
- Decentralized architectures with some fixed nodes:

What if no infrastructure (or only a partial one) is available?
Data dissemination

- **Objective:**
  - Push data towards (potentially interested) mobile nodes
Data dissemination

• Objective:
  – Push data towards (potentially interested) mobile nodes

• Challenges:
  – Avoid network flooding
  – Adapt the dissemination to the type of info (e.g., parking vs. accident)
The VESPA approach

• Objective: share any type of event between vehicles using vehicular ad hoc networks (unified solution)
  – Numerous events to share!!
    • Available parking spaces
    • Emergency braking
    • Obstacles on the road
    • Real-time traffic information
    • Emergency vehicles
    • Driver in state of hypovigilance / doing strange maneuvers
    • …

• The type of event considered has an incidence on its relevance (and so on its dissemination)
Representation of events

• Messages are exchanged between vehicles to describe physical events
• Different attributes. At least:
  – Identifier
  – Priority
  – Position (and reference positions)
    • GPS coordinates
  – Time
    • GPS time
  – Event type
    • e.g., available parking space, accident, etc.
  – Version
    • No invalidation message is considered!

Is this enough?
Content-based Dissemination

• Objectives:
  – Support different types of events
  – Inform all the potentially interested vehicles
  – Support a high number of vehicles and events

• Challenges:
  – Avoid network flooding
    • Limit the number of vehicles relaying
      – Only the $k$-farthest vehicle will relay the info
  – Adapt the dissemination area to the type of information carried
    • A vehicle will not further broadcast a message received if this message is not relevant anymore
Encounter Probability

- Objective: estimate whether a vehicle is likely to encounter an event or not
- Not trivial because the destination of the driver cannot be assumed
Encounter Probability

• Example of computation:
  (with maps, with geographic vectors)

\[
\text{EP} = \begin{cases} 
1 & \text{if } \text{TTR} < \text{TTL} \\
0 & \text{otherwise}
\end{cases}
\]
Content-based Dissemination
Introduction of a latency in the dissemination process

\[ t_i = D \times (1 - \frac{d_i}{r}) \]
Content-based Dissemination

Messages relayed are considered as acknowledgements

Diffusion canceled!

Relays the message

Diffusion canceled!
Experimental evaluation

- Prototype
- Simulator
  - With and without maps (roads and parking lots)
- More info: [TR-C’10, IEEE ITS’11, MIS’11a]
Push-based QP: Pros and cons

+ "Easy" to provide information to the vehicle
+ "Simple" query processing techniques can then be used to deliver relevant information to the driver
- Only popular data is diffused
- The set of queries processed remains limited
Alternative: GeoVanet

• **Goal:**
  – Provide a solution to enable pull-based data gathering in vehicular ad hoc networks

• **General principle:**
  – Disseminate queries in the network
  – Consider a stationary node as a mailbox to collect the partial results obtained on the remote (mobile) nodes
  – Use both node mobility and hops in the wireless network to route the partial results towards the mailbox

• **Queries with relaxed time requirements**
Main Steps

1. Query dissemination
   - Composition of disseminated queries:
     • Request: the core of the query
       - e.g., what are the interesting sites to visit in Aussois? where are the clients looking for a taxi located?
     • Exp-date: date by which the answer is expected
     • Key: determines the location where the answer should be sent and retrieved

2. Remote processing

3. Delivery of the partial query result(s)

4. Retrieval of the query result
GeoVanet: Delivery of the query result

**Carry and forward approach**

- Every $\Delta t$ seconds, each vehicle checks whether it is driving towards the target (mailbox) or not.
- If not, it chooses the closest node (mobile node or infrastructure node) as the new carrier.
- The new carrier repeats the same algorithm until the carrier reaches the communication range of the mailbox or the expiry date is reached.

Related: use of mobile agents in VANETs
Experimental Evaluation

- Use of a simulator
  - Real road networks (TeleAtlas digital maps)
Experimental Evaluation

• Evaluation of the percentage of “interesting” nodes reached considering several strategies
  – Flooding
  – Contention-based forwarding
  – Dissemination using hotspots

• Main Results:
  – Between 60% and 70% of relevant vehicles receiving the query whatever the strategy used (considering that only 2% of the vehicles carry a query result)
  – 80% of the results are collected in the mailbox after one hour
  – 80% of the results are collected with less than 40 hops
  – More info: [MDM’11, MIS’11b]
Multi-scale mobile query processing
• We consider multi-scale query processing as any query processing that may need to access data sources of different types (e.g., local databases, remote web services, data streams, etc.) to compute the result.

• Objectives:
  – Exploit all relevant data sources, whatever their location
  – Benefit also from the information provided by Web Services
Examples

• Retrieve the list of petrol stations located in a radius of 10 Km around me where fuel prices are less than $1 (and update the result every 5 minutes)
• Retrieve the list of hotels with available rooms that I can reach in less than 30 minutes
Multi-scale query processing

Where should I refuel?

Which are the rest areas that will be near me in the next hour and offering a gas station, lodging facilities for two persons and a restaurant?

Available parking spaces, traffic congestions, emergency brakings, ...

Ecole Masses de Données - May 2012
Challenges

• Generation of query execution plans
  – No global schema
  – How to locate relevant data sources?
  – Need to compose several services
    • e.g., to convert GPS coordinates into the name of a city or region to match with the interface of the service providing the fuel prices

• How to select the best one?
  – Solution 1: compute the list of close petrol stations locally (POIs) and obtain the prices for those stations using a Web Service
  – Solution 2: retrieve the region where I am located (first service) and then retrieve the list of petrol stations (with fuel prices) located in that region (second service)
Query Optimization

- How to select the best query execution plan?
- Trade-off between different costs
  - Time, energy, financial cost, etc.
  - Estimating the costs:

\[
C_{\text{Time}}(Q) = C_{\text{QueryDelivery}}(Q) + C_{\text{Time Processing}}(Q) + C_{\text{ResultDelivery}}(Q)
\]

\[
C_{\text{Money}}(Q) = C_{\text{QueryDelivery}}(Q) + C_{\text{Money Processing}}(Q) + C_{\text{ResultDelivery}}(Q)
\]

\[
C_{\text{Energy}}(Q) = K \times n
\]

\[
C(Q) = \sum_{i=\text{Time,Money,Energy}} w_i \times C_i
\]

- Quality expectations for the query result
  - Minimizing the above costs may lead to a poor result quality!
  - More info: [IJAIHC’11]
Prediction & query processing
What if no information is provided?

Where is the nearest parking space?
Our Approach: Aggregation

- **Objective:**
  - Process queries such as “what is the area where the probability to find a parking space is the highest?”

- **Store, aggregate and exchange summaries**
  - Do not destroy them once used to warn the driver
  - Major difference with other works on data aggregation for vehicular networks

- **Use the summaries generated to extract additional knowledge usable by drivers**
  - Estimate the probability that an event (e.g., an accident) occurs in a spatio-temporal area
Two levels space model

Spatial Model

Ecole Masses de Données - May 2012
Data Structures

Flajolet-Martin sketches
Inter-vehicle exchanges

• The quality of the information produced depends on the amount of data aggregated
• Each car/driver decides what to exchange and his/her preferences
  – Publish/subscribe process with priorities
• Duplicate detection is important
  – I might have observed the same events as my neighbor!
    → Flajolet-Martin sketches
• Need to know the vehicles with which exchanges have been performed recently
• Experimental evaluation
• More info: [RAIRO’10]
Open Issues
Just one example: context-awareness

• Many parameters frequently change in ambient environments and may impact the query processing or data dissemination
  – Connectivity
  – Autonomy
  – Location
  – etc.

• Other challenges: management of multimedia, GUIs, etc.
Example (Context-awareness)

- Context changes strongly impact the best communication solution to use:

  - I had an accident
  - Accident in 100m
  - Accident in 200m
  - ???
  - Ok I will carry and forward the information?
  - Be careful! Accident ahead!
  - Node density very low
  - I don’t care!
  - I’m driving in the opposite direction!
Conclusion

• These are some of the problems addressed within the VESPA project
• Other contributions related to resource allocation in ad hoc networks
  – Competitive environment
  – First arrived, only served...
  – More info: [ACM Mobility’09]

• From the data management point of view, a lot of (very) interesting problems to tackle!
Acknowledgements

• C. Caloca (CICESE, Mexico)
• N. Cenerario (Univ. Valenciennes, France)
• B. Defude (Telecom Institute, France)
• J.A. Garcia Macias (CICESE, Mexico)
• T. Hien (Univ. Valenciennes, France)
• S. Lecomte (Univ. Valenciennes, France)
• N. Mitton (INRIA Lille, France)
• O. Urra (Univ. Zaragoza, Spain)
• D. Zekri (Telecom Institute, France)
• ...

Ecole Masses de Données - May 2012


Thank you for your attention!

Merci!