Googling the physical world (or, my research overview)

Murat Demirbaş

SUNY Buffalo

16 October 2012
1. Wireless Sensor Networks

2. Smartphones

3. Cloud computing
Where is the nearest shade?
A sensor node (mote) has 8K RAM, 4Mhz processor, magnetism, heat, sound, vibration, infrared wireless communication up to 100 ft.

For achieving scalability distributed & local algorithms are needed.

Wireless collisions & ad hoc environments create challenges.
In OSU, we developed a surveillance service for DARPA

Detect, track, & classify trespassers as car, soldier, civilian
LiteS: 100 nodes in 2003,
ExScal: 1000 nodes in Dec 2004
Where is the nearest enemy tank?

For scalability, local ops are needed over global structures. Using geometry, we design efficient & minimal infrastructures. We achieve graceful resilience to faults via self-stabilization.

Querying structures: DQT, Glance, ... O(d) time for querying, where d is the distance to the nearest answer.

Tracking structures: MDQT, Trail, ... O(d) time for querying; O(m*\log m) for update, where m is the dist. the evader moved.
Greenhouse monitoring
Parking lot monitoring
Singlehop collaboration/coordination primitives (NSF)

Transact: Transactional framework for programming WSANs
Transact enables understanding of a system execution as a single thread of control, while permitting the actual execution over multiple threads distributed on several nodes

Pollcast, Countcast, Coordcast: Lightweight singlehop collaboration and coordination primitives for WSANs
What are the waiting times in nearby cafes?
5B cellphone users worldwide
1.13 billion phones sold in 2009
vs 0.3 billion PCs
15% = 174M were smart phones

~Pentium III, +WiFi, GSM, Bluetooth, camera, mic, GPS, compass, sensors...

Cared by user, mobile coverage, human intelligence included

Singlehop access to cloud!
DARPA’s grand challenge: Find 10 balloons in US quickly

Social networks is useful for crowdsourced sensing & collaboration
Twitter: 200M users, 200M tweets & 1.6B queries daily

Our work:
Weather app over Twitter
Location-based querying app over Twitter
Monitoring changes in location related tweets in cities
Identifying breakpoints in public opinion for a topic
Building an expert-sourced system to play Jeopardy!
LineKing: Crowdsourced line wait-time estimation

Deployed at Tim Hortons at UB
Currently used by more than 2000 people
100s readings daily, 2 min. MAE, lots of positive feedbacks
We are building a 1000 Android phone reprogrammable testbed

Geoffrey Challen, Murat Demirbas, Steve Ko, Tevfik Kosar, Chunming Qiao @ Univ at Buffalo

Dense, controlled, yet realistic environment for testing and developing next generation collaborative smartphone apps and operating systems
How do we build the backend for real-time querying?
Cloud computing provides computing as a utility. It features:

- elimination of up-front commitment by users
- illusion of infinite computing resources available on demand
- ability to pay for use of computing resources as needed
- use of SOA to provide 3rd parties APIs for services at every layer
Maestro: automated locking framework

MapReduce? Not all applications are embarrassingly parallel! Transactions are unscalable, locking is manual and error-prone

Maestro is a scalable automated locking framework. It consists of a master and several workers, which can be dynamically instantiated on demand.

Maestro examines the program actions of the workers before deployment and automatically decides which worker actions can be executed locally and which actions require synchronization through the master.
WAN filesystems for sharing big data

Scientists spend a lot of effort on solving basic data-handling issues: the physical location of data, how to access it, how to move it to visualization and compute resources for further analysis.

We aim to enable reliable, efficient, and transparent wide-area data sharing and processing for large-scale collaborative science via

- our efficient & consistent WAN virtual filesystem
- end-to-end dataflow parallelism for WAN perf. optimization
- semantic-aware wide-area data placement
Catch me!

office hours: 313 Davis Hall, Thu 12:30-14:30
web: http://www.cse.buffalo.edu/~demirbas
twitter: https://twitter.com/muratdemirbas
blog: http://muratbuffalo.blogspot.com