Event Processing on Mobile Phones: Mobile 3.0?

Archan Misra,
Senior Scientist,
Telcordia Technologies
+1.732.699.5278
archan@research.telcordia.com

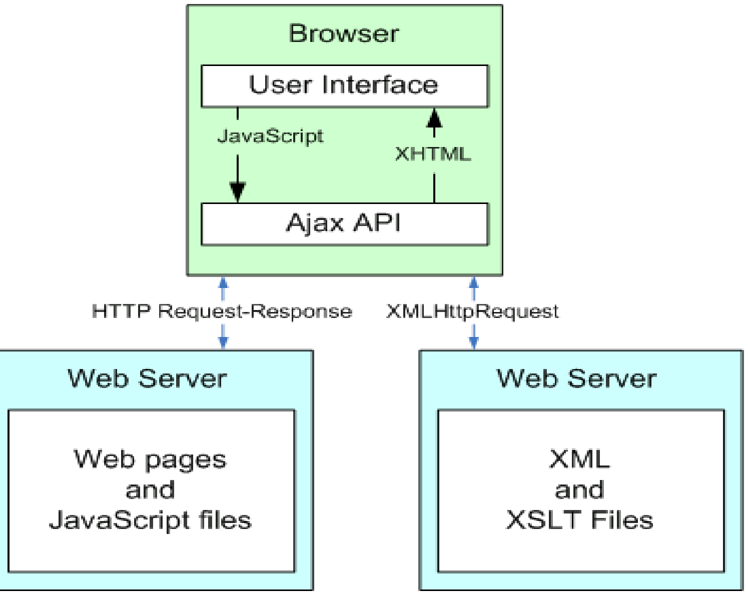
Outline

- **The Evolution of Mobile Computing and Mobile 3.0**
- My work on Efficient Stream Processing on Mobile Devices
  - Harmoni, MediAlly and LE3SE
- Ongoing Work and Thoughts on Real-Time Analytics and Mobile 3.0

Acknowledgments:
Harmoni is joint research with Iqbal Mohomed (summer intern) and Maria Ebling (IBM)
MediAlly is joint research with Ben Falchuk (Telcordia) and Atanu Roy Chowdhury (summer intern, 2009.)
LE3SE is joint research with Lipyeow Lim (Univ. of Hawaii)

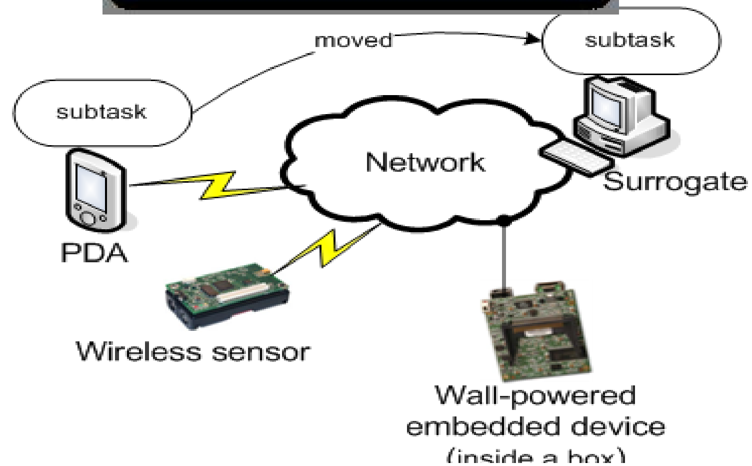
Cloud vs. Local Computing?

Desktop World



Ajax Scripting and Efficient Browsing

Mobile World



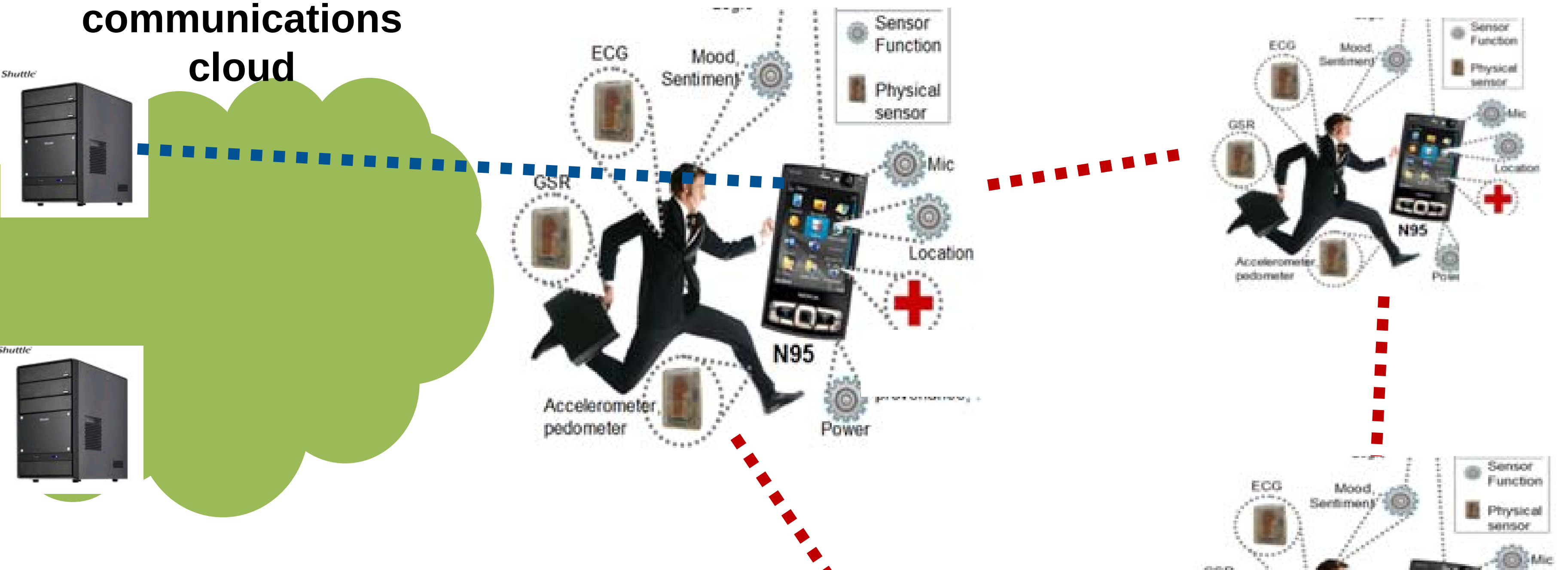
Type	Response Time	Battery
local	61.47 (1.29)s	1.5 (0.109)%
cyber foraged	2.9 (0.066)s	0.06 (0.003)%
cyber foraged (20s sleep)	24.4 (0.648)s	0.1 (0.008)%

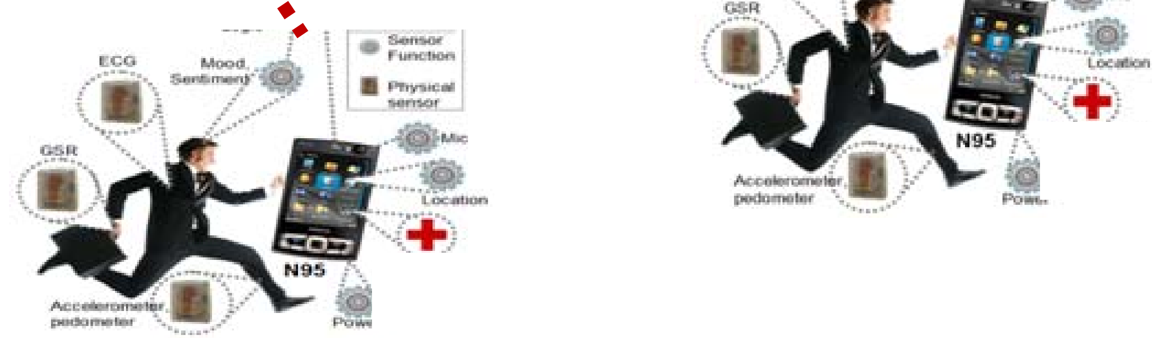
Voice Processing via Elastic Applications
(reproduced from Goyal and Carter, WMCSA 2004)

- Move compute logic from server to local client.
- **Goal: Improved Responsiveness for Interactive Applications**

- Move computing from local client to server/clone/surrogate.
- Goal: **Lower footprint on resource-constrained mobile client.**

The Brave New World: beyond Static & Mobile





Account for the dynamic
formation embedded in
- nearby mobile phones
- nearby sensors

A new compute model:

he Personal Cloud

Mobile xxx.0: The Evolution

1.0

ody defined it!!

cess to static &
amic Web content on
ile phone.

st applications were
t-based

personalization or
ext-dependence

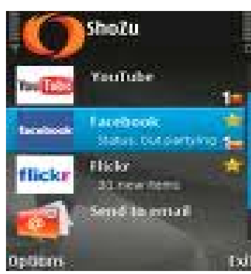
2.0

•Includes: Web 2.0..
- User-generated content & apps
- Mashups (application composition)

• Context-awareness of personal client platform.
- Location-aware search, maps, directions

Significant social network interaction and collaborative focus

•However, social interaction & application mashup occurs in the "cloud" (server-based)



3.0?

•Interaction with surrounding dynamic context.
- Static: RF-sensors, bar-codes, tags
- Mobile: **Other mobile phones and sensors**

• Increasing ability to distribute computing logic across client, cloud (server) and cloud-assisted "proximal peers"

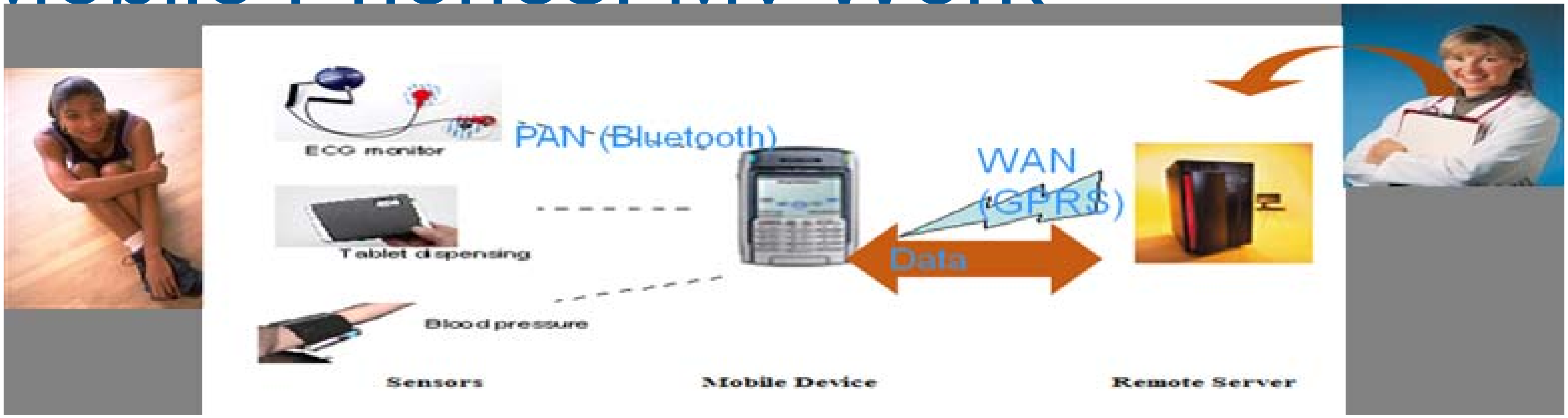
• Greater emphasis on:
- multimedia content (video, images)
- •transient interaction with unknown individuals

•Richer stream event processing on individual mobile devices & by 'groups of mobile devices'

Outline

- The Evolution of Mobile Computing and Mobile 3.0
- **My work on Efficient Stream Processing on Mobile Devices**
  - **Harmoni, MediAlly and LE3SE**
- Ongoing Work and Thoughts on Real-Time Analytics and Mobile 3.0

Stream Event Monitoring on Mobile Phones: My Work



- Carried out in the context of remote health monitoring
  - Collection of physiological and activity data from body-worn and external sensors
- Focus on chronic diseases, both physiological and mental (diabetes, CHF, PTSD)
- Three Projects:
  - **Harmoni:** Exploiting Context Awareness
  - **MediAlly:** ATDM and Contextual Provenance
  - **LE3SE:** Low-Energy Embedded Event Sensing

The Resource Challenge in "Continuous" Remote Health Monitoring

Type of Sensor Device	Bits/ sensor sample	Raw data rate (KB/day)	Operational Lifetime (hours
GPS	1408	14,850	270
SpO2	3000	94,922	42

- Mobile devices provide a promising new platform for **personalized monitoring** of medical sensor data and **ubiquitous real-time feedback**.

Need to reduce the energy overheads of high rate data collection and analytics

- Current practice addresses two extrem
  - very low frequency (e.g., glucose thrice/
  - low-duration, high-data rate data collection (e.g., triage)

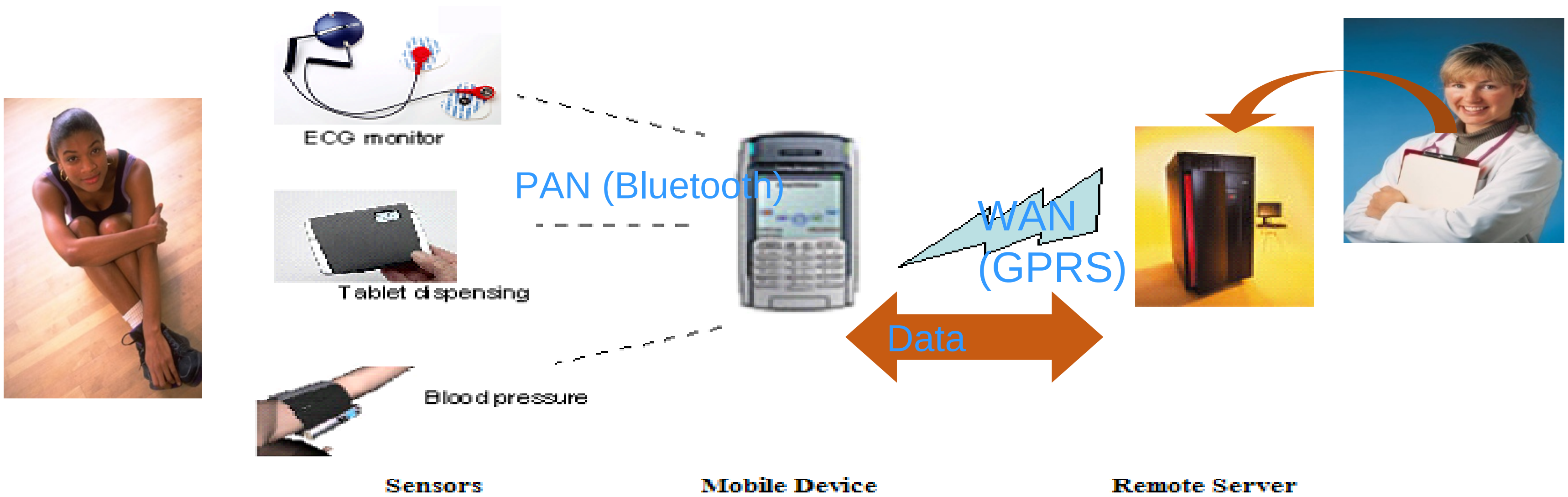
- Health and wellness comprises:
  - Physical markers (Sp02, ECG etc.)
  - Mental markers (Stress, Depression, Activity)

Future

Combination of physical and mental markers, likely to provide best medical insight and patient care.

Use of collection of phones to provide "deep ambient context".

Harmoni (**H**ealthcare **A**daptive **R**emote **Moni**toring)



1. Context-Aware Event Filtering

IF (user 'running in gym' & 90< hr<120
THEN send AVG(hr)/ min

2. Personalization of Rules

Apply Machine Learning at Server to learn
"Normal for user A: gym= 110,150)

3. Anticipation-Driven Transmission

Based on past, Prob(802.11 within 2 hrs) > 0.7 → cache data

Harmoni Novel Features

Mohomed, A. Misra, M. Ebling and W. Jerome, *Context-Aware and*

Harmoni (**H**ealthcare **A**daptive **R**emote **Moni**toring): Key Components on Mobile Device

- Lightweight Rule/Event processing engine
  - Identifies appropriate temporal patterns in data stream(s) and consequent action
    - **e.g., if {110<AVG(last 10 heart rate values) <130 && ‘user running for <10 minutes’}, then {“transmit AVG to server”}**
  - Processed events themselves act as predicates for new rules.

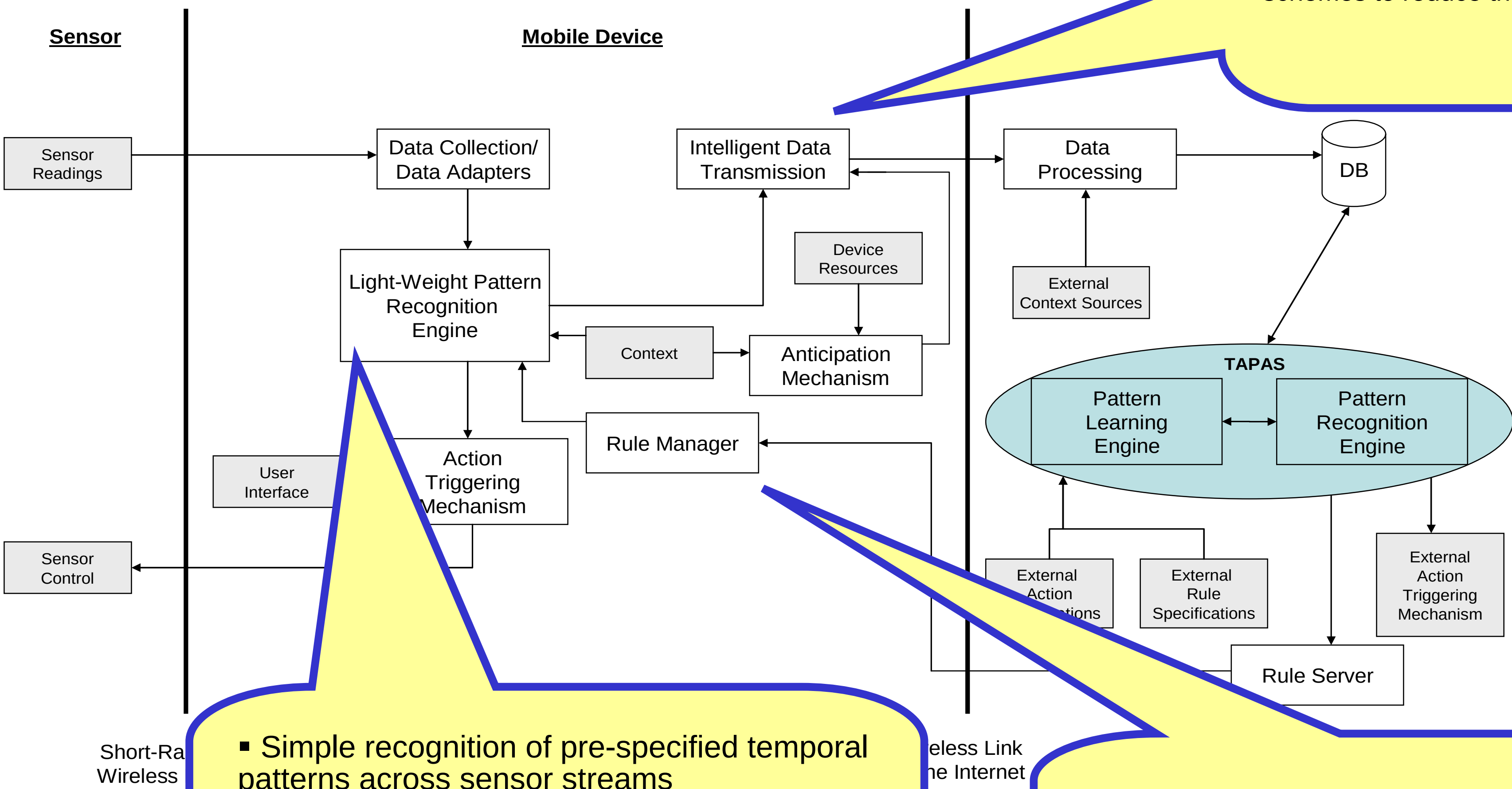
- Rule Manager
  - Coordinates with server to determine current rules and populates event engine.

- Intelligent Data Transmission
  - Compresses the filtered data (from event engine) for transmission to server.

- Anticipation Mechanism
  - Schedules the transmission of (compressed) data based on predicted availability of network connections and incoming sensor data rates.

Harmoni Functional Architec

▪ Compressed transmission ro serve
when interface is available AND
“anticipation” indicates “OK-to-Send”
▪ Use of intelligent compression/stor
schemes to reduce the volume of tra



▪ Simple recognition of pre-specified temporal patterns across sensor streams
▪ Event engine driven by Deterministic Finite Automata (DFAs)
▪ Actions associated with rules include data transmission, data transformation and local triggers (alarms, reminders)

▪ Rule Manager coordinates with backend server to download currently active or applicable rules
▪Can run, activate or deactivate multiple rul
simultaneously and dynamically.

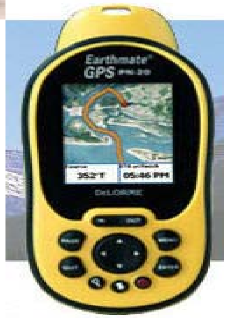
Harmoni Implementation Platform

- Nokia 770 Internet tablet → N800.
  - ARM processor, Linux-based
  - High-resolution display(800x480), touch screen with up to 65,536 colors
  - 64-128 MB RAM, 64 MB FLASH storage (expandable up to 1GB … can be used for virtual memory)
  - Built-in Bluetooth (BlueZ stack) and 802.11 interfaces
  - Relatively cheap: $350
  - http:///www.maemo.org provides open-source software and development environment.
  - Code compiled on an Intel/Debian Linux 3.1 box using cross-compiler (http://www.scratchbox.org)

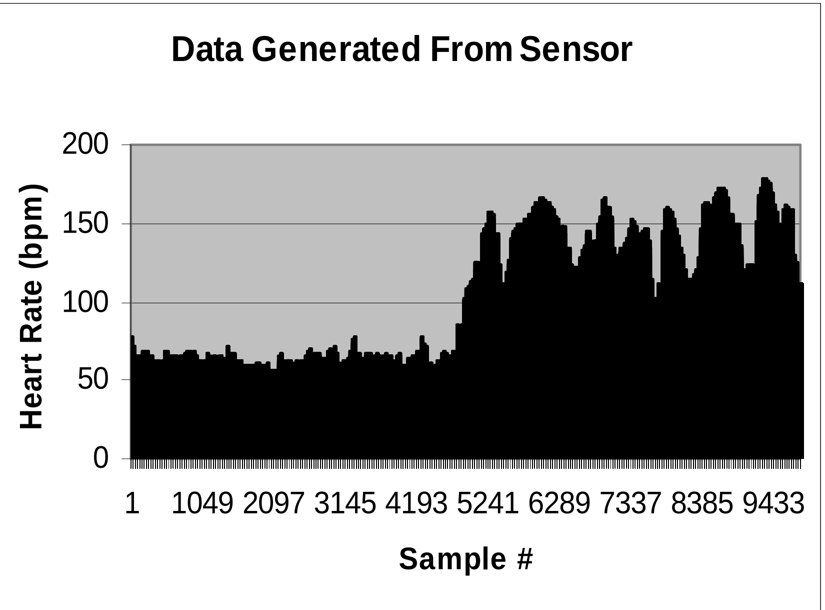
- Nonin Model 4100 Sp02/heart rate monitor
  - Provides Heart rate and Oxygen saturation
  - Supports Bluetooth Serial Port Profile (SPP)
  - 120 hours of continuous operation with 2 AA batteries
  - Three packets transmitted per second, where each packet is 375 bytes

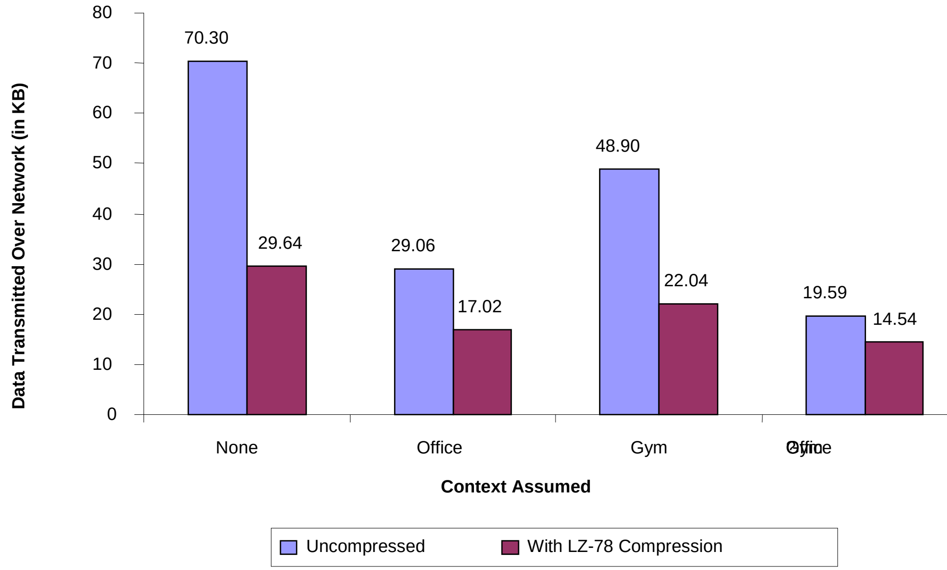
- WiTilt 3-axis Accelerometer
  - Output baud of 57.6 Kbps
  - 40 mA consumption when operating.

- Delrone Earthmate GPS



pact of Harmoni on Transmission Bandwidth: Idealized Conte



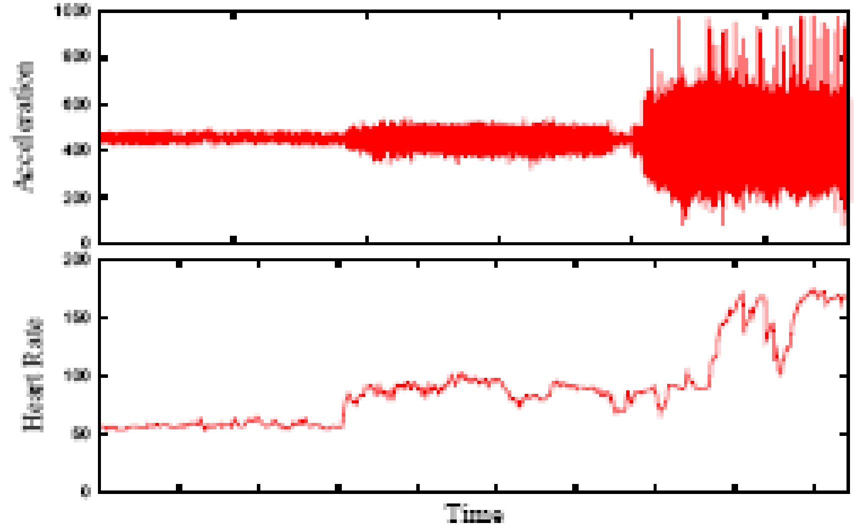


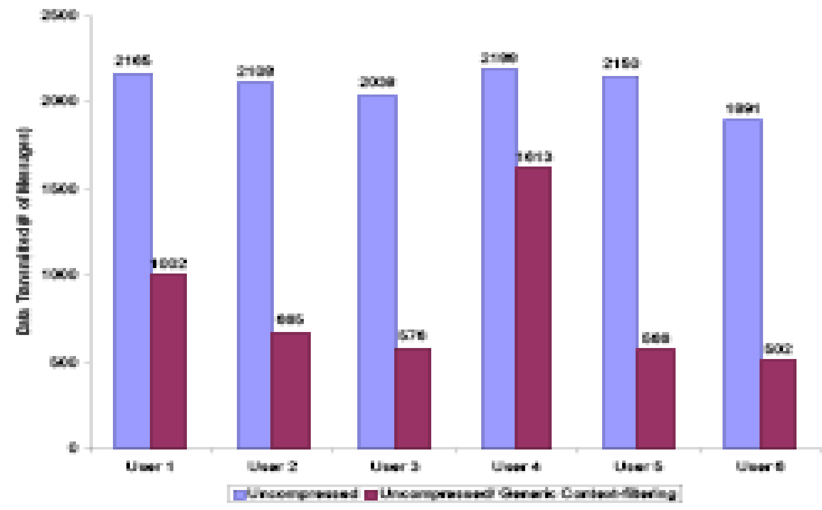
- Compression (S2), by itself, results in > 50% reduction in bandwidth consumption
- Filtering (S3) results in 85% reduction in bandwidth consumption
- Improvement from Filtering (S3) to Context-Sensitive Filtering (S4) not significant due to lack of exact pattern match of floating point values in LZ.

HARMONI in Practice: Sensor-based Context

- Accelerometer amplitude can be used to classify user into 3 different states: sitting, walking, running
  - Higher ‘normal’ threshold for 3 distinct states (0-70, 70-110, 110-170) across all users.
  - Aside: accelerometer readings used to classify ‘falls’ for elderly patients.

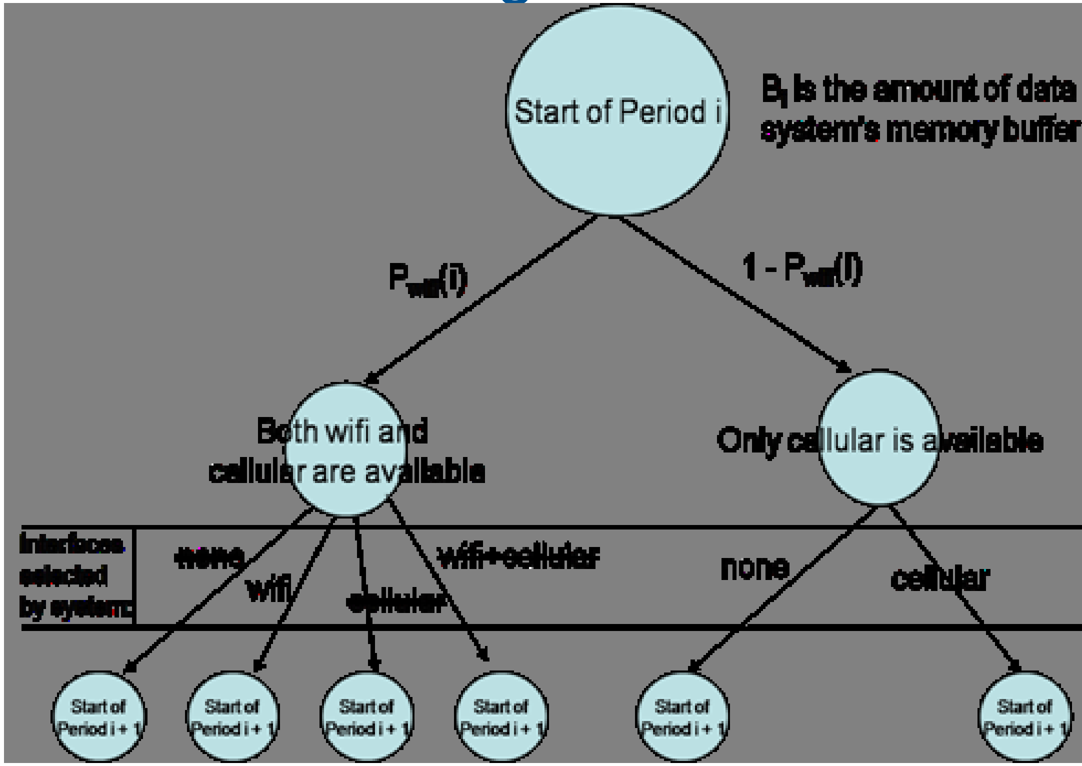
- Bandwidth savings between 26-73% for our sample population.





e *L*-step (Lookahead) Finite Horizon DP Algorithm

- Past history of user can be used to build up
  - *Future connectivity probability vector*: Connection probabilities in different time slots to 802.11/cellular (Rahmati07)
  - *Future data generation vector:* Average rate of data generated by remote monitoring and event filtering



B_{i+1} = min(0, (Bi + Di) – T_c) where T_c is the transmission capacity of the wireless interfaces selected by the system

$$\text{rg}\min_{f\in(0,1)} MEC\,(B(0),\, f\,, L, 0)$$

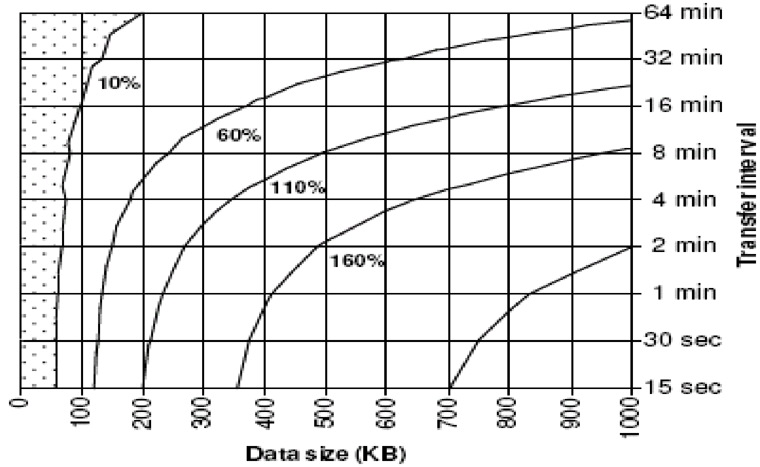
$$\text{here}\quad MEC\,(B(0),\, f\,, L, 0) = f\, *\, E_{cell} * T_{cell} * + MEC\left(B(0) + D(0) - f\, * T_{cell}\,, L - 1, 1\right)$$

$$\text{nd}\quad MEC\,(x, L, k \neq 1) = \left(1 - p^{k}_{wifi}\right) * \left[f * E_{cell} * T_{cell} + MEC\left(B(k) + D(k) - f * T_{cell}\,, L-1, k+1\right)\right]$$

$$^{k}_{wifi} * \left[E_{wifi} * T_{wifi} + MEC\left(B(k) + D(k) - f * T_{wifi}\,, L-1, k+1\right)\right]$$

Effectiveness of Finite Horizon DP

- Non-finite cost associated with loss of data (due to buffer overflow and unavailability of actual 802.11 connectivity).
    - If guarantee of no overflow, then follow slope-based transmission on cellular interface [Keshav09]
    - 64 MB cache can easily buffer 2 hours of ECG/EMG data.
- Better benefit if system willing to tolerate small probability of data loss by either
    - Associating cost with residual data B(L) and avoid paths that cause overflow at intermediate time slots
    - Bounding the total fraction of data lost due to overflow in L slots.
- Small tolerance leads to much improved energy efficiency
    - Additional constraint if cost associated with latency of delivery.



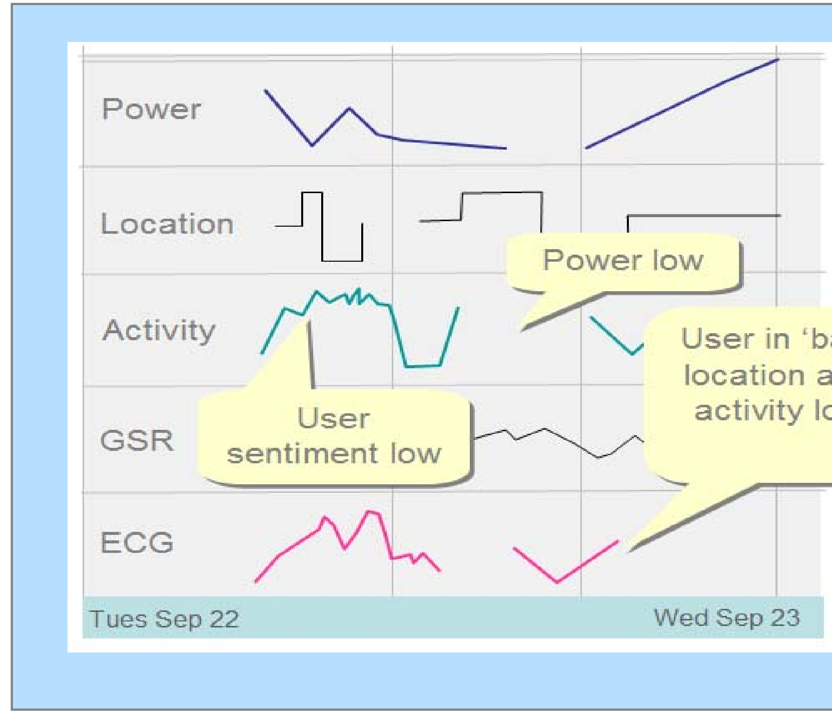
(b) 80% Wi-Fi coverage

Potential battery gain with average 80% 802.11 availability (Rahmati07): 802.11 has high startup cost, while cellular has high per bit transfer cost.

MediAlly and the ATDM Paradigm

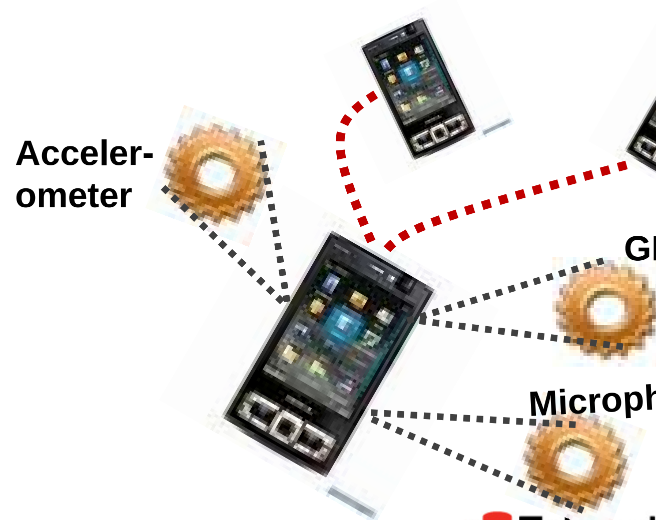
Health and wellness comprises **both**:

- Physical markers (Sp02, ECG etc.)
- Mental markers (Stress, Depression, Activity)



Use deep context as a trigger for intermittent, activity driven monitoring.

- **ATDM= Activity-Triggered Deep Monitoring**
- **Currently, no integration of local and global context for remote monitoring**.



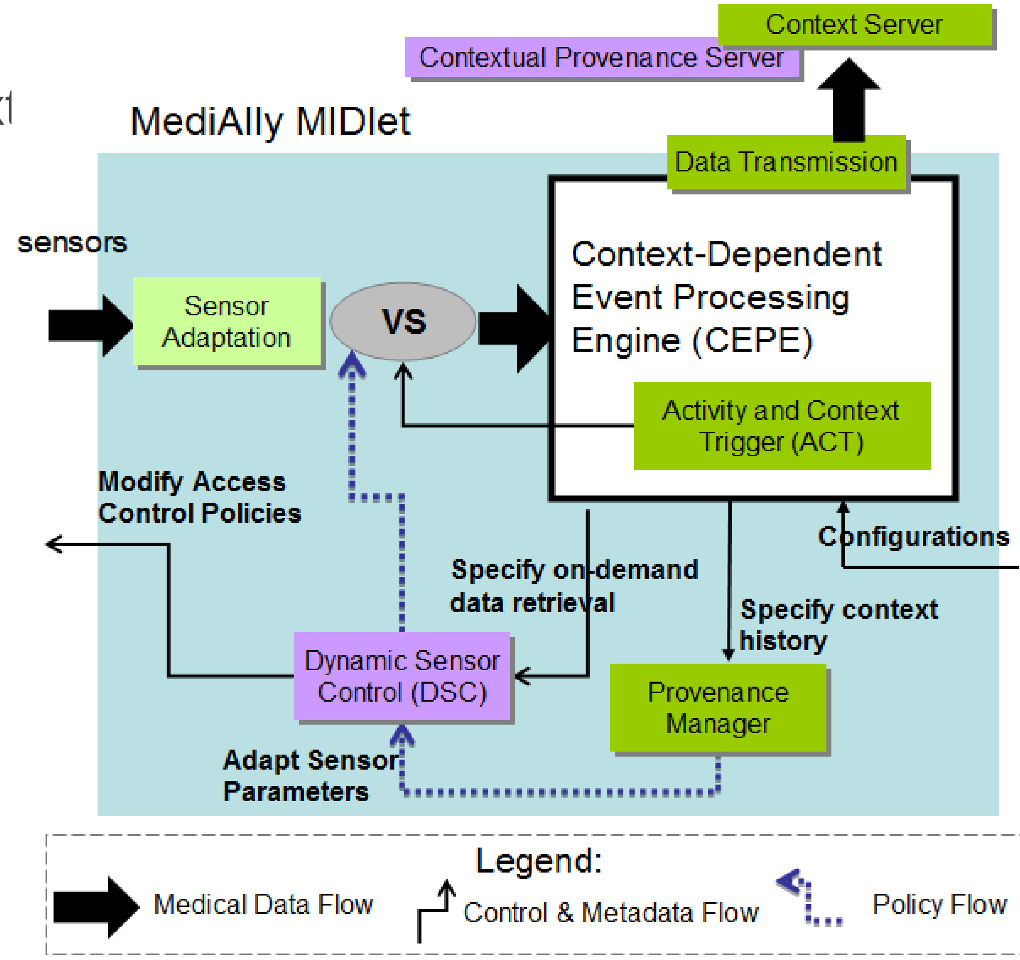
Why is context different from medical data?

- Obtained from divergent sources (mostly non-medical) of varying quality/correctness.
- Inherent uncertainty in context inference.

Key Features of MediAlly

MediAlly's infrastructure:

- Uses a combination of sensor, mobile device and global context to improve the ATDM process.
  - **Key Idea: Use context to determine "episodes" for which rich data collection from sensors is necessary.**
- Provides PHR repositories with easy access to provenance for monitored data.
  - **Key Idea: Since PHRs will only contain dis-continuous segments of data, allow observers to get answers for**
    - "why did the system collect or not collect data at certain times?"
    - "what was the user's context (both personal and non-personal" around certain medically significant episodes?"



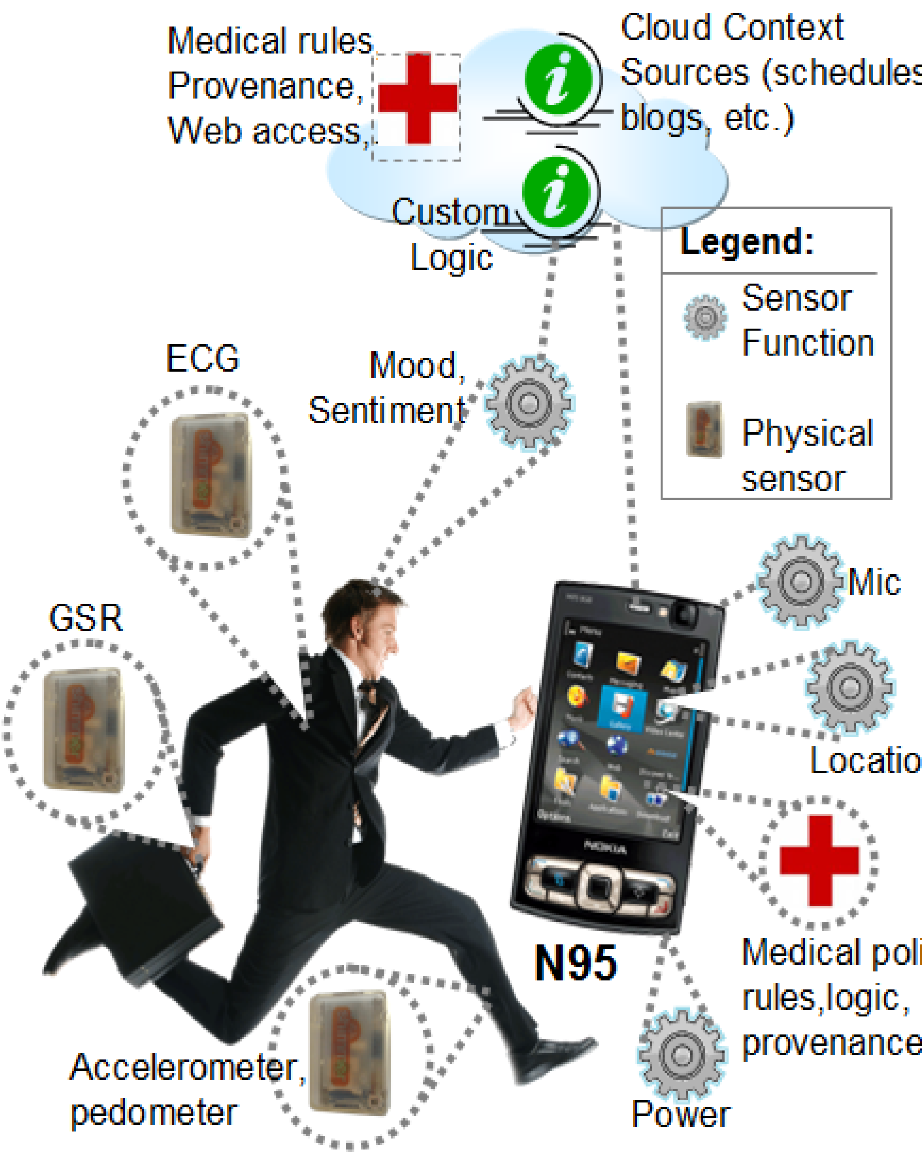
MediAlly Functional Architecture

Key Technical Innovations

Combine on-board event correlation on the mobile device along with context from cloud to determine “activity state” of patient and use as a trigger for monitoring.

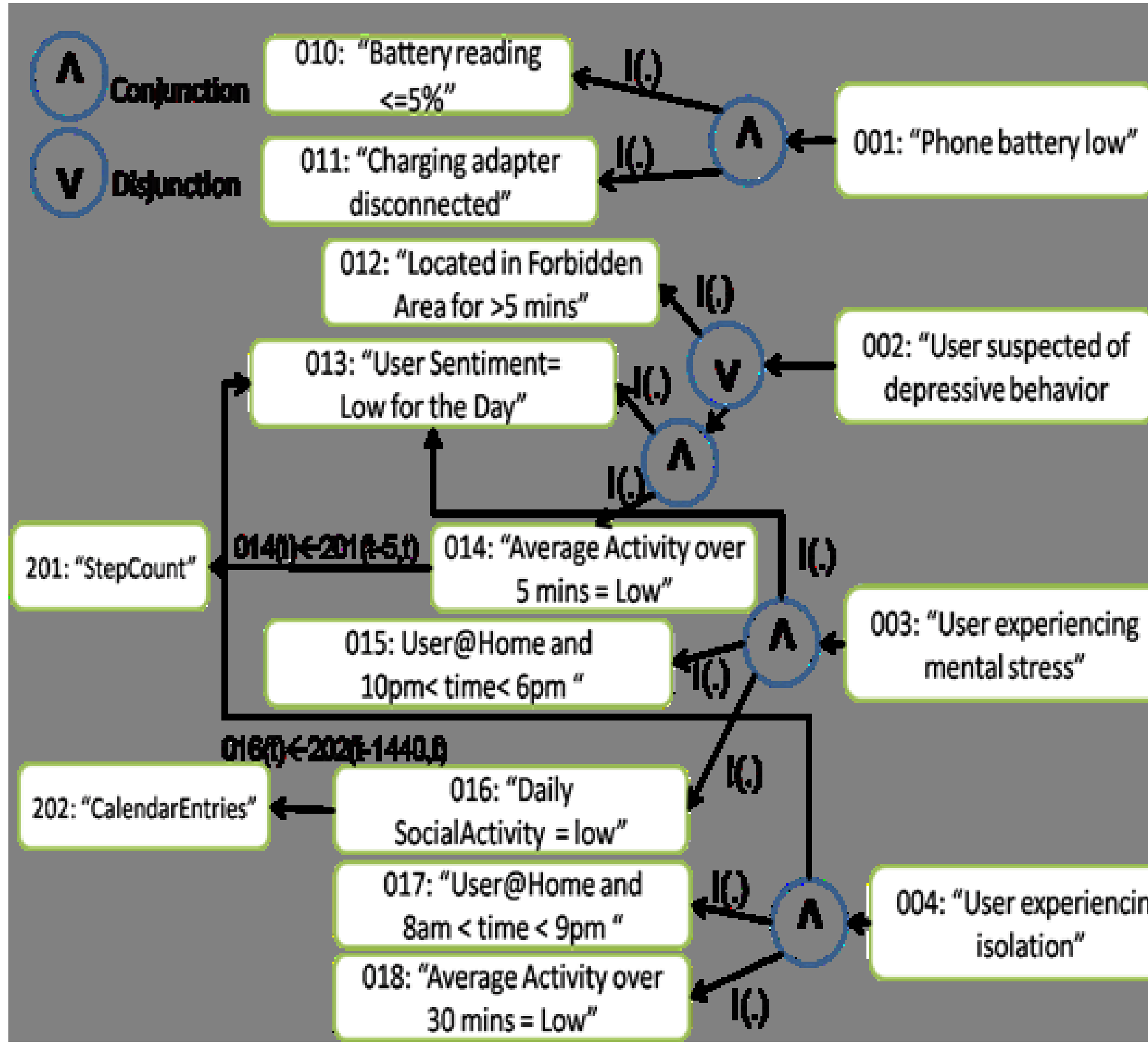
Use a combination of
- **Cloud** context (sentiment analysis, semantic location tracking)
- **Cell phone** context (e.g., GPS, noise level of phone)
- **Body-wearable** sensors (e.g., ECG, GSR, accelerometer)

Provide **low-overhead context capture** through transmission of model-based storage of processing graphs and operator state.
- Enable the storage of metadata that describes why and how the process of monitoring was affected by user’s activity, environmental conditions etc.



Low-Overhead Context Monitoring and Provenance

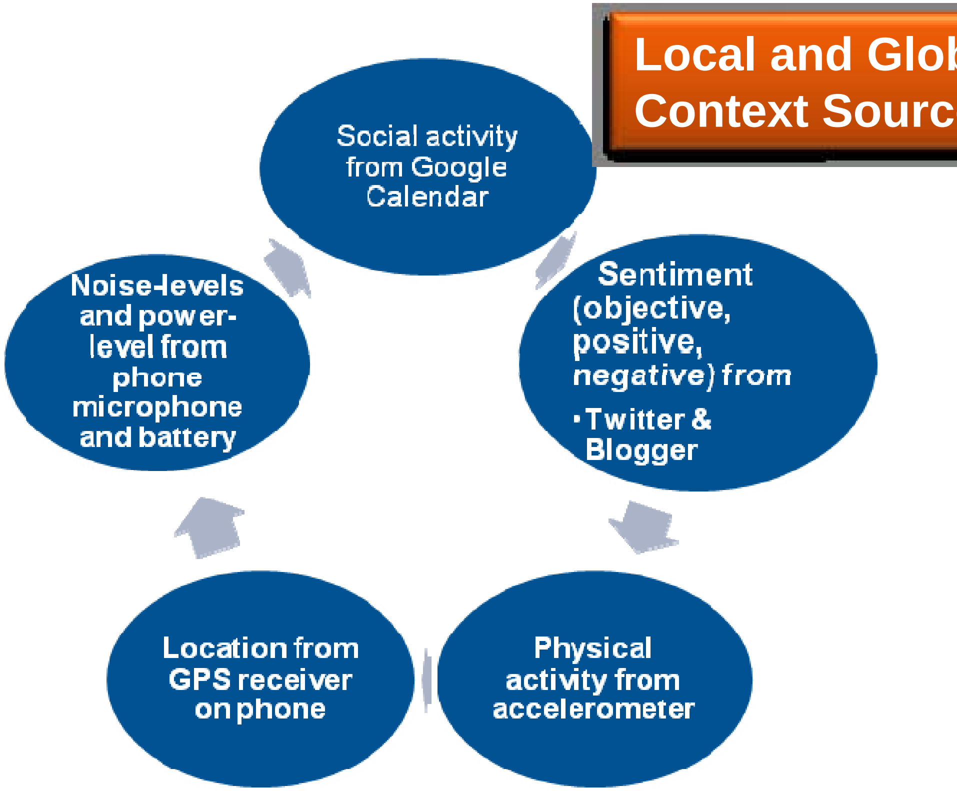
- Context composition represented as a **Context-Composition Graph (CCG)**
  - Statically configured by application
- Context History captured as evolution of CCG node states.
  - Lazy capture for reduced context overhead.

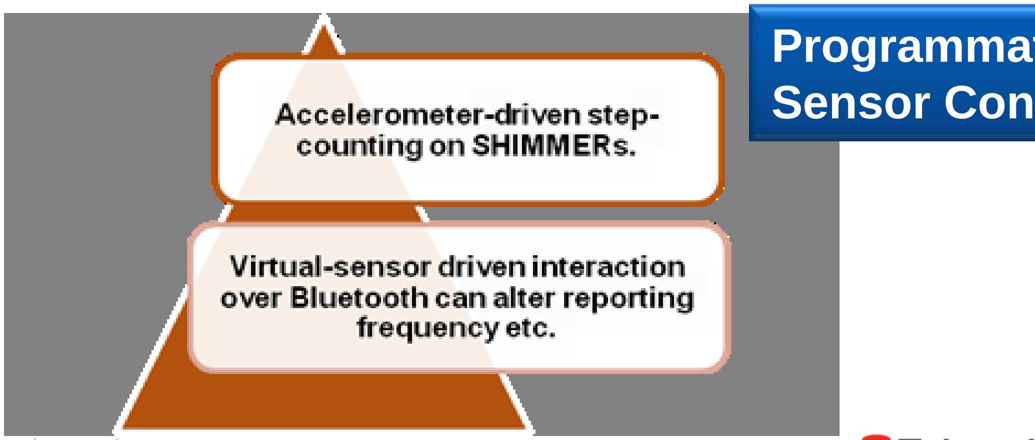


Partial CCG for our illustrative "emotional health" de

The MediAlly v0 Implementation

- Implementation of demo-level code (Nokia N95 phone and Intel SHIMMER platform) showing
  - ATDM (context-triggered data collection)
  - Context collection and replay (provenance)
  - Visualization of medical data and contextual metadata
- ~2500 lines of code
  - Midlet code on N-95 phone
  - Nes-C code on SHIMMER
  - Backend DB and Presentation tiers



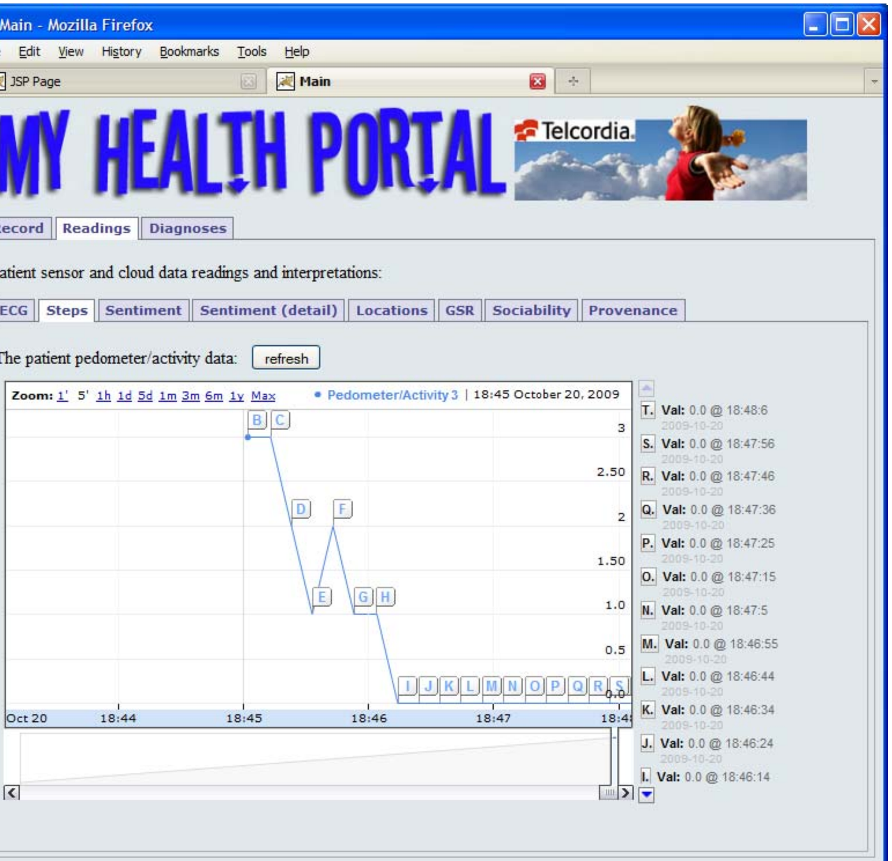


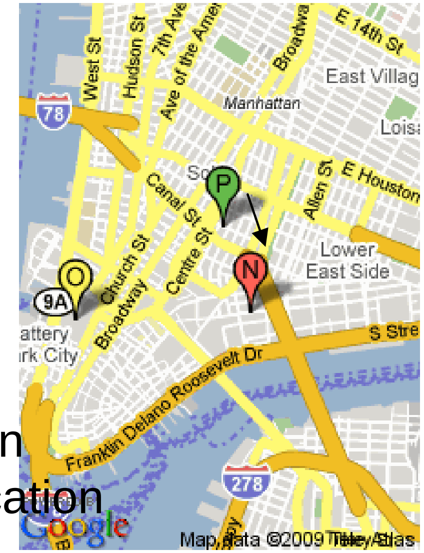
Illustrative Demo

- Uses ‘activity, location, environmental conditions’ as potential triggers for collecting data about
  - Emotional (stress)
  - Physiological (ECG)
  - Enviromental (noise level)

Medical Rationale	Context Trigger	Sensors Used for Context Inference	Sensing Action	Sensor Data Collected
Terminate monitoring due to low battery	Phone power<5% and charger disconnected	Phone battery sensor	None	None
User may be lacking composure	(In Forbidden Area >=5 min)|| (Sentiment=low and Avg. Activity (5min)=low)	GPS, Sentiment, Accelerometer	Collect noise and stress readings	GSR, Microphone
User may be feeling isolated and depressed	@Home&& Sociability=low && Avg. Activity (15min)=low	GPS, Calendar, Accelerometer	Collect noise and stress readings	GSR, Microphone

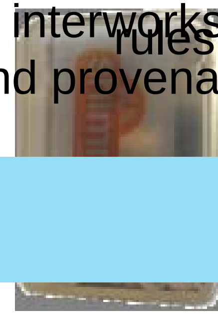
Exemplary Demo Flow





User moves to an “undesirable” location

Device in ‘steady state’, interworki efficiently with sensors; no wellnes rules are triggering..



Device interworks with sensors, begins to strea data and provenance to server(s)

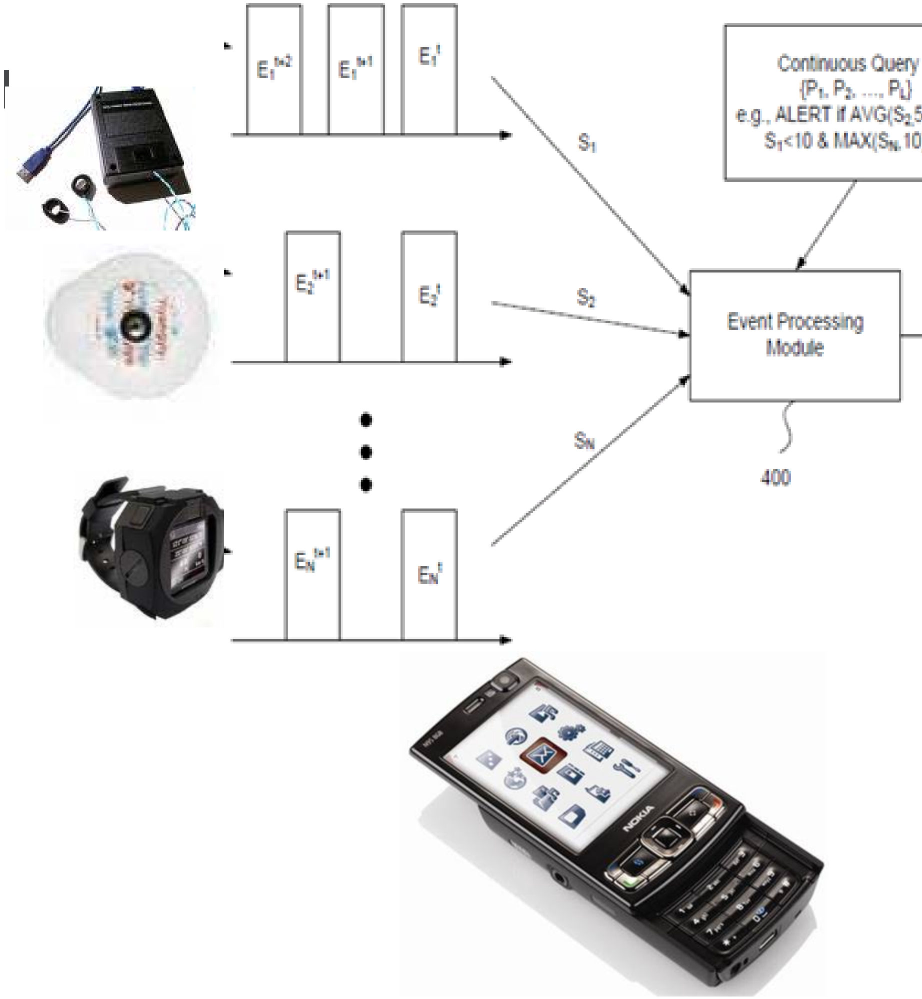
GPS Receiver

Server(s) record data and provenance for review

User begins to walk about at a rate > THRESH

LE3SE: Reducing the Energy of Event Processing

- Transforms sensor data streams from "push" model to on-demand "pull" model
- Mobile event engine retrieves data in chunks and on-demand.
  - Bulk transfer of data improves efficiency on sensors.
  - Sensors programatically instructed to cache data.
- Event Engine determines evaluation sequence dynamically based on
  - Selectivity characteristics of individual data stream: P()
  - Cost function associated with individual sensor data stream PAN transmission: C()



LE3SE: Dynamic Stream Evaluation Selection

- Dynamic retrieval and evaluation is a **continuous query optimization problem**
  - Illustrated for a single compound query
  - Borrows 'page processing' from DB & uses commn. cost as part of 'cost function'

- Assume:
  - P(S1)<10=0.05; C(S1)=0.2 nJ/sample;
  - P(S2,5)<20=0.95; C(S2)=0.02 nJ/sample
  - P(S3,10)<4=0.2; C(S3)=0.01 nJ/sample

Q1: ALERT IF AVG(S2,5)>20 && S1<10 && MAX(S3,10)< 4

Normalized Costs (evaluation stops if predicate is FALSE):
S1=0.2*1/0.5
S2= 0.02*5/0.95
S3= 0.01*10/0.8

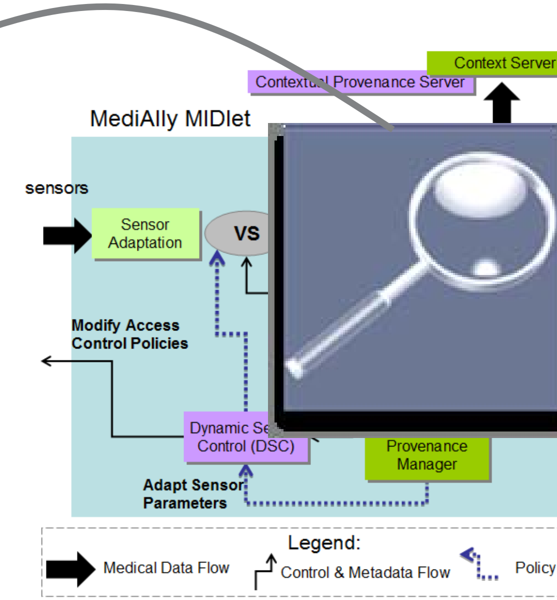
Retrieve & Evalu
in order: S2, S3,

Q1: ALERT IF AVG(S2,5)>20 || S1<10 || MAX(S3,10)< 4

Normalized Costs (evaluation stops if predicate is TRUE):
S1=0.2*1/0.5
S2= 0.02*5/0.05
S3= 0.01*10/0.2

Retrieve & Evalu
in order: S1, S3,

LE3SE Architecture



Query Logic Specification Module

- Stream-SQL based specification of query syntax

Cost Modeler

- External specification of sensor-specific trx. Cost model
- Dynamic evaluation of stream selectivity

ormalized
uery Syntax

C(.); P(.)

Dynamic Query Evaluation Optimizer

- Signals sensors to adjust push-vs-pull mode
- Determines retrieval sequence for sensor streams

Asynchronous Event Engine

- Maintains partial query evaluation state

Push/Pull, Batch commands

Dynamic Sensor Control (DSC)

Event Processing: Key Lessons Learnt

Harmoni	Use Event Procesing + Machine Learning to infer "personalized context" and reduce uplink transmission volume from phone	~70% reduction in transmission energy overhead
MediAlly	Use Event Processing over local+ global context to activate/de-activate individual sensor PAN transmissions	~80-90% reduction in PAN + WAN transmission energy overhead
LE3SE	Use dynamic stream query optimization to (a) batch transmissions for a single sensor stream and (b) eliminate unnecessary PAN transmission of	Under evaluation—initial results suggest about 60-70% reduction in PAN transmission energy overhead

All work presented here focused on event processing on single client device over "personal data streams"

Outline

- The Evolution of Mobile Computing and Mobile 3.0
- My work on Efficient Stream Processing on Mobile Devices
  - Harmoni, MediAlly and LE3SE
- **Ongoing Work and Thoughts on Real-Time Analytics and Mobile 3.0**

Acknowledgments:
Harmoni is joint research with Iqbal Mohomed (summer intern) and Maria Ebling
MediAlly is joint research with Ben Falchuk and Atanu Roy Chowdhury (summer intern, 2009.)

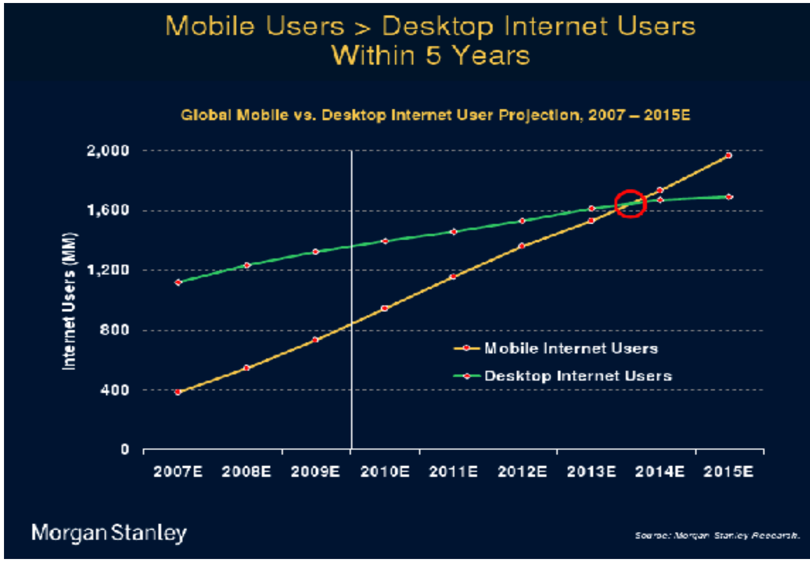
Mobile Device Security: a High-Priority Next-Gen Threat

ellphones and other hand-held computing devices e becoming the de-facto end-points of computing for oD and national communication infrastructures.

- However, pervasive devices currently have a significantly lower level of 'security assurance'→ scant acceptance of mobile-device based corporate VPNs

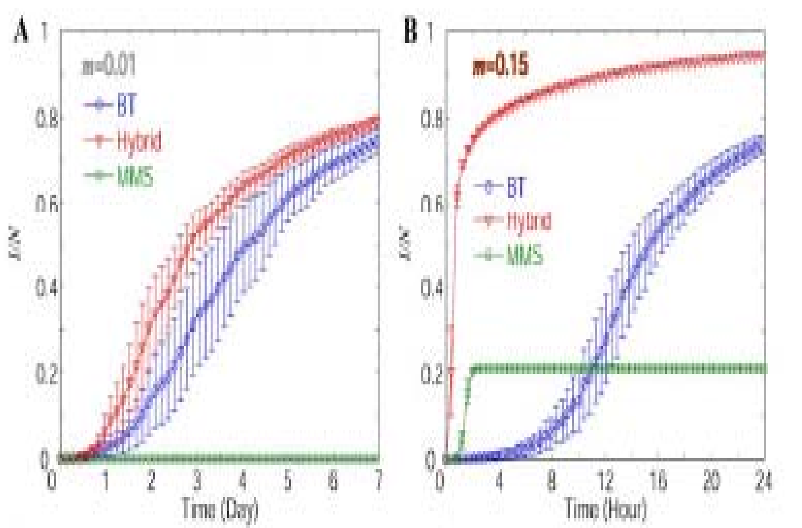
•Cellphones and mobile devices increase the infection susceptibility of malware spread. Catastrophic epidemics occurring due to a combination of:

- electronic spread (conventional)
- physical spread (proximity-driven)



nventional security solutions for computing devices not be instantiated on hand-held devices:

- Encryption, PKI, checksum, sandboxing etc. all prove too resource-intensive on such platforms and open up additional attack modes
- Mobile device capabilities typically 2-generations behind commodity fixed hardware.



The Figure shows the time evolution of infected nodes % of suscpectible nodes) for BT, MMS and hybrid viru spreads. (from: P. Wang, M. Gonzalez, C. A. Hidalgo, A Barabási , Understanding the spreading patterns of m

The Resource Challenge for Mobile Devices

Evolution of Attacks and Malware

1. Battery drainage attacks
2. Trojans for Symbian , iPhone etc. → notion of trusted applications
3. Botnets for Mobile Phones (e.g., SYMBOS_YXES.B) just beginning to appear
4. Call/Messaging attacks (Troj/Terdial-A makes expensive international calls)

Evolution of Mobile Computing Apps

1. Data-intensive communications (volume & frequency): email, SMS, MMS
2. Multimedia-enriched applications
    1. Voice- understanding
    2. Video
    3. Sensor streams (m-health)
    4. Augmented Reality (image processing)
3. Open, collaborative app models

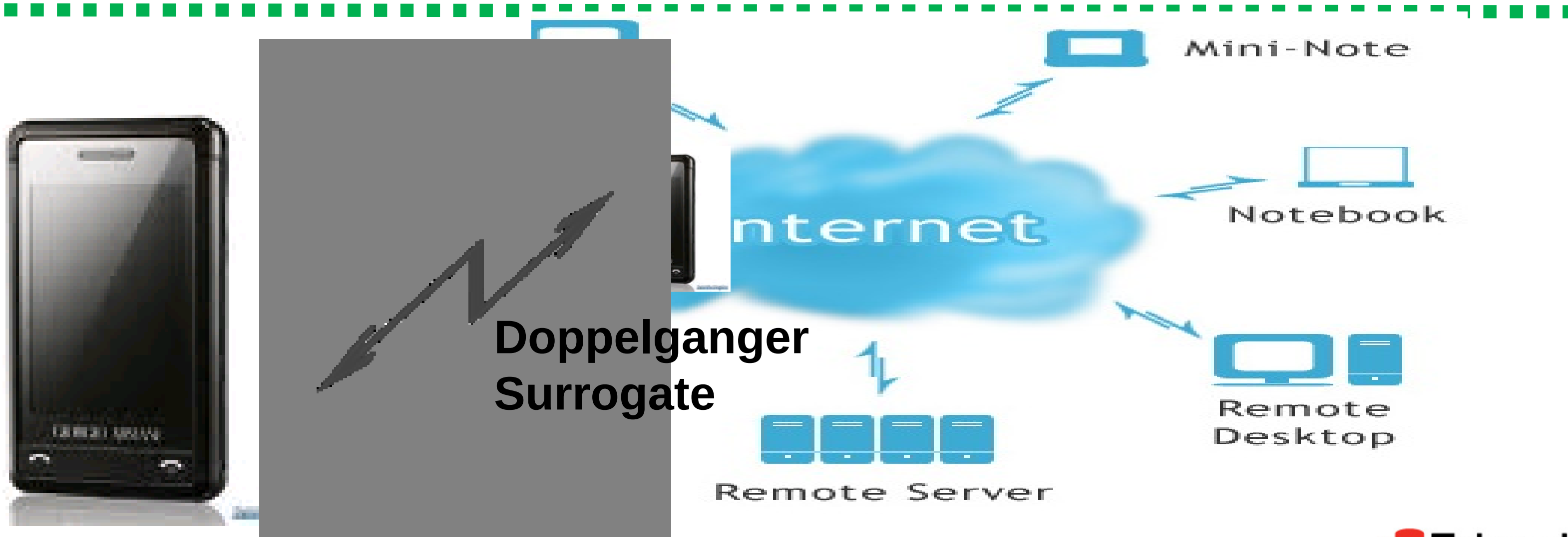
e

allenge

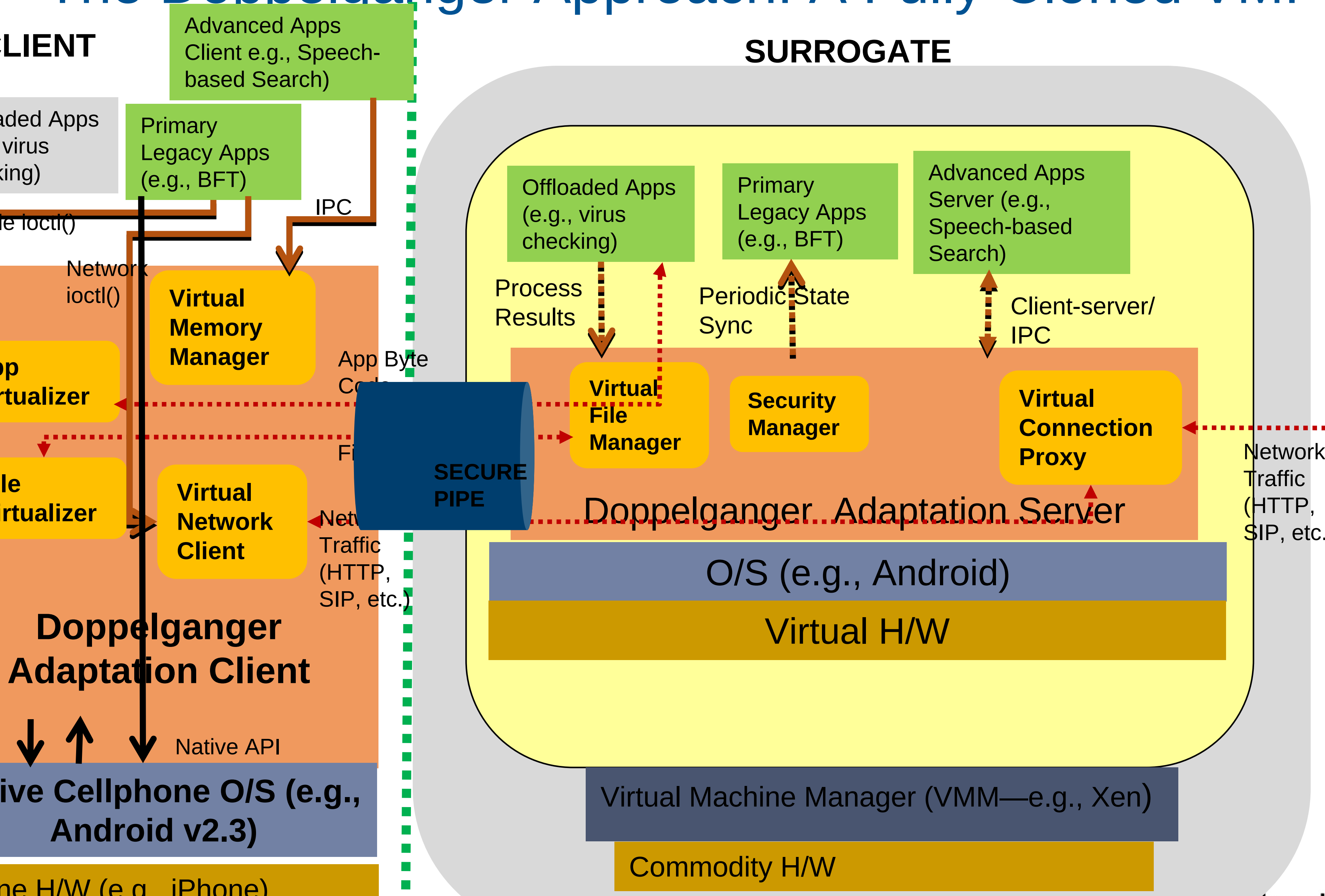
he

olution

Phone Client Device



The Doppelganger Approach: A Fully-Cloned VM!



Mobile Analytics 3.0: Examples of Cloud Dis-Intermediation

pps

An activity monitoring app that continually records and compares my jogging speed/time vs. other runners on my jogging trail
- Computes avg. speed/duration of all joggers

A 'child location' app in a sports stadium that uses the phone `flash crowd' activity to search for location/movement trail of missing child
- Source distributes child picture; recipients execute 'image matching' against recently clicked pictures in stadium.

A 'visitor' to an amusement park wants live 3-D video fe of all rides/activities in a 100 metre vicinity of her location
- Peer phones record and upload video feeds; aggregated and combined a a 'virtual feed' for the visitor ambient reality application.

nnical
e

Event Processing now distributed across ALL mobile devices in my vicinity

Requires transfer of large image files (content) and potentially expensive image matching

Requires continuous exchange of multimedia stream from a dynamically changing proximity peer se

efits

- Anonymity of Individuals/Data
- Reduced WAN load on 3G infrastructure
- Real-time "live" monitoring

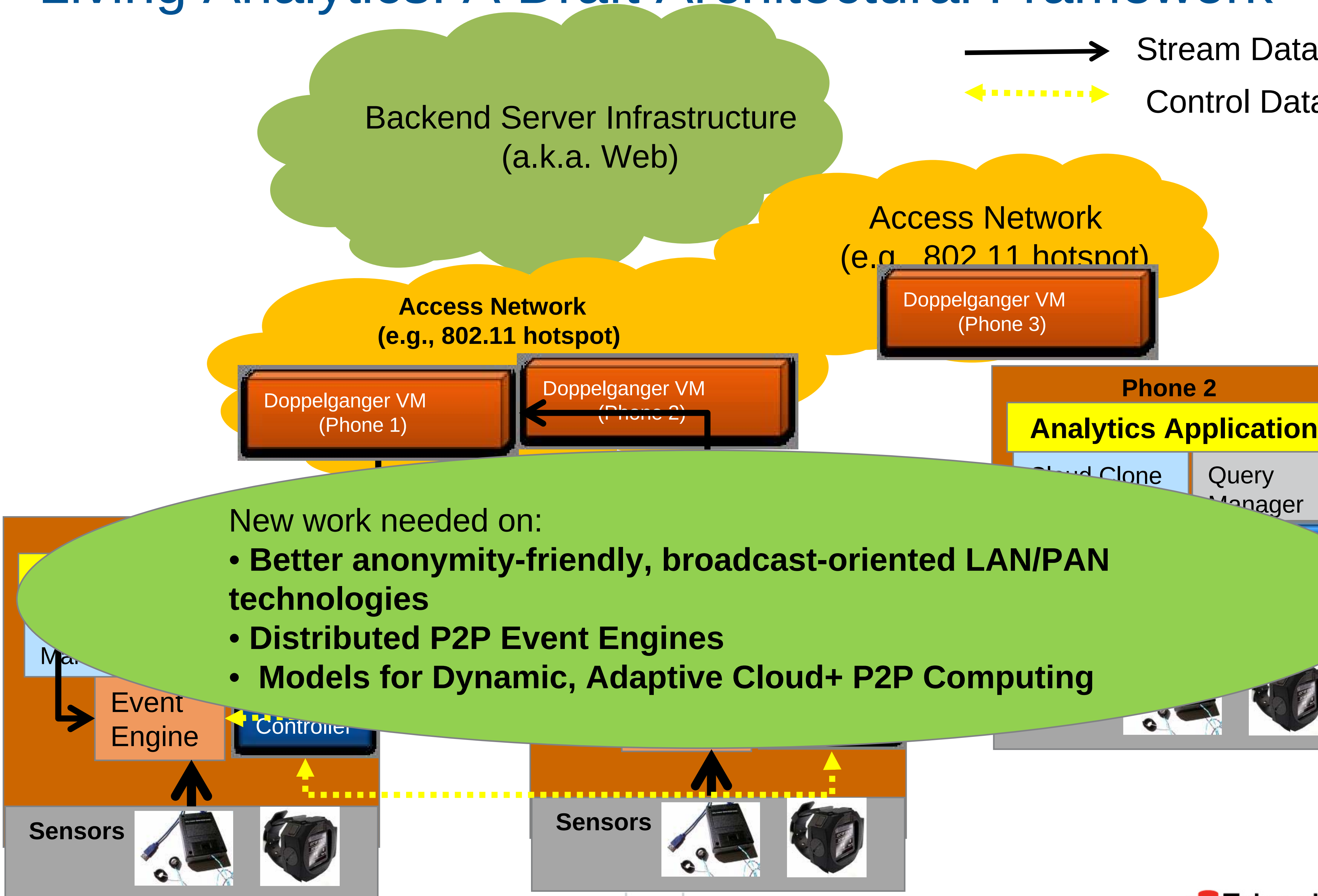
- Utilizes the full power of the collective social network.
- No dependency on any fixed infrastructure/sensors.

- Real-time awareness of surrounding "real" environment.
- Better coordination of crowd movement in public spaces.

Living Mobile Analytics: Key Challenges

Technical Challenge	Key Issue	Likely SolutionApproaches
P2P Cooperation for Analytics	Peers have no direct incentive to contribute content for my analytics	• Participatory/sharing incentives • VM-based partition of app space & resource limits on 'participatory analytics'
Privacy and Anonymity	• Query source would like to conceal its identity. • Data sources would like to maintain anonymity.	• Query source provides ID of Doppelganger VM. •Service provider services for anonymized aggregation & distribution of peer content.
Bandwidth Challenges for Continuous Stream-based Queries	•PAN/LAN interfaces not well defined in terms of bandwidth & energy for high-volume P2P communication. •3G/WAN interfaces likely to be overwhelmed by pure cloud-based computing model.	• Query distribution via PAN/LAN interface • Extensive energy-efficient onboard analytics on individual mobile device • Efficient use of WAN channels/links for high volume data dissemination
Quality of Peer Generated Data	•Need to guard against erroneous/malicious generation of data by peer mobile devices/sensors.	• Ongoing work on certified, tamper-proof sensing on mobile devices. • e.g., PM modules with private keys on embedded sensors (HotSec 09)

Living Analytics: A Draft Architectural Framework



Conclusions

- Mobile computing will evolve from Mobile 2.0 to “Mobile 3.0” characterized by exploitation of real-time multimedia data generated by ‘peer’ mobile devices and sensors
  - Importance of energy-efficient stream event processing on mobile devices
  - Opens up new class of real-time analytics applications in myriad areas—e.g., crowd coordination & navigation, situational awareness, healthcare, etc.
- Advances in event processing middleware on mobile devices
  - Context-aware event processing to reduce transmission of raw sensor data
  - Context-triggered control of sensors to avoid generation of unnecessary streams
  - On-demand adaptive acquisition and evaluation of sensor streams to reduce PAN energy overheads
- Future living/mobile analytics will require a flexible hybrid (cloud+ peer) computing paradigm
  - Leverages upon recent advances in virtualization/cloning, multimedia event processing etc.


Thank
You