Participatory Privacy:

Enabling privacy in Participatory Sensing

DoE CRYPTODOC

21-Nov-2011

Claudio Soriente

http://lsd.ls.fi.upm.es/lsd/Members/claudio-soriente

Participatory Sensing: why?

- **Wireless Sensor Network**
  - Small-scale
  - Short-lived
  - Application-specific
  - Static
  - *Very* resource constrained
  - Wireless multi-hop
  - Deployment / maintenance costs
  - Low Real-life impact
  - People out-of-the-loop

Participatory Sensing: who?

- **Smartphones**
  - 10^9 (and counting) worldwide
  - Always -on, -carried, -connected (3/4G)
  - Multiple embedded sensors
    - GPS, thermometer, accelerometer, light sensor, etc.
    - Bluetooth, NFC to connect to other sensors
  - Powerful
    - 1.5Ghz dual-core, 1GB ram, rechargeable battery

- **People**
  - Mobile
  - Interaction w/ others
  - Interaction w/ environment

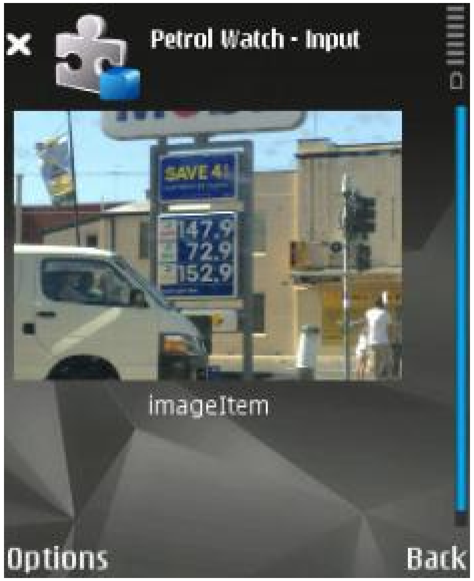
Participatory Sensing: what?

- Novel, **fast-growing** computing paradigm
- **Infrastructure-less** data collection at never-seen **scale**
- Harvest **dynamic** information about **environmental/social** trends
  - (Some) People are more interesting than motes
  - Exploit their mobility and their relationship with the environment

- That's right: **mobile phones** are "**sensors**"!

Participatory Sensing Initiatives 1

PetrolWatch
@ DCOSS'08



University of South Wales

BikeNet
@ SenSys'07

Dartmouth College

LiveCompare
@ HotMobile'09

Duke University

Participatory Sensing Initiatives 2

ParkNet
@ MobySys'10

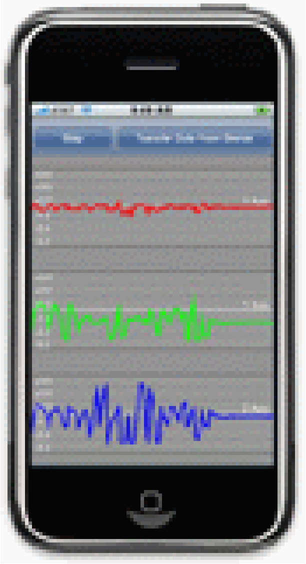
Image courtesy of Primospot

Rutgers University

SignalGuru
@ MobySys'11

Princeton University

Ishake
(tech.rep.)

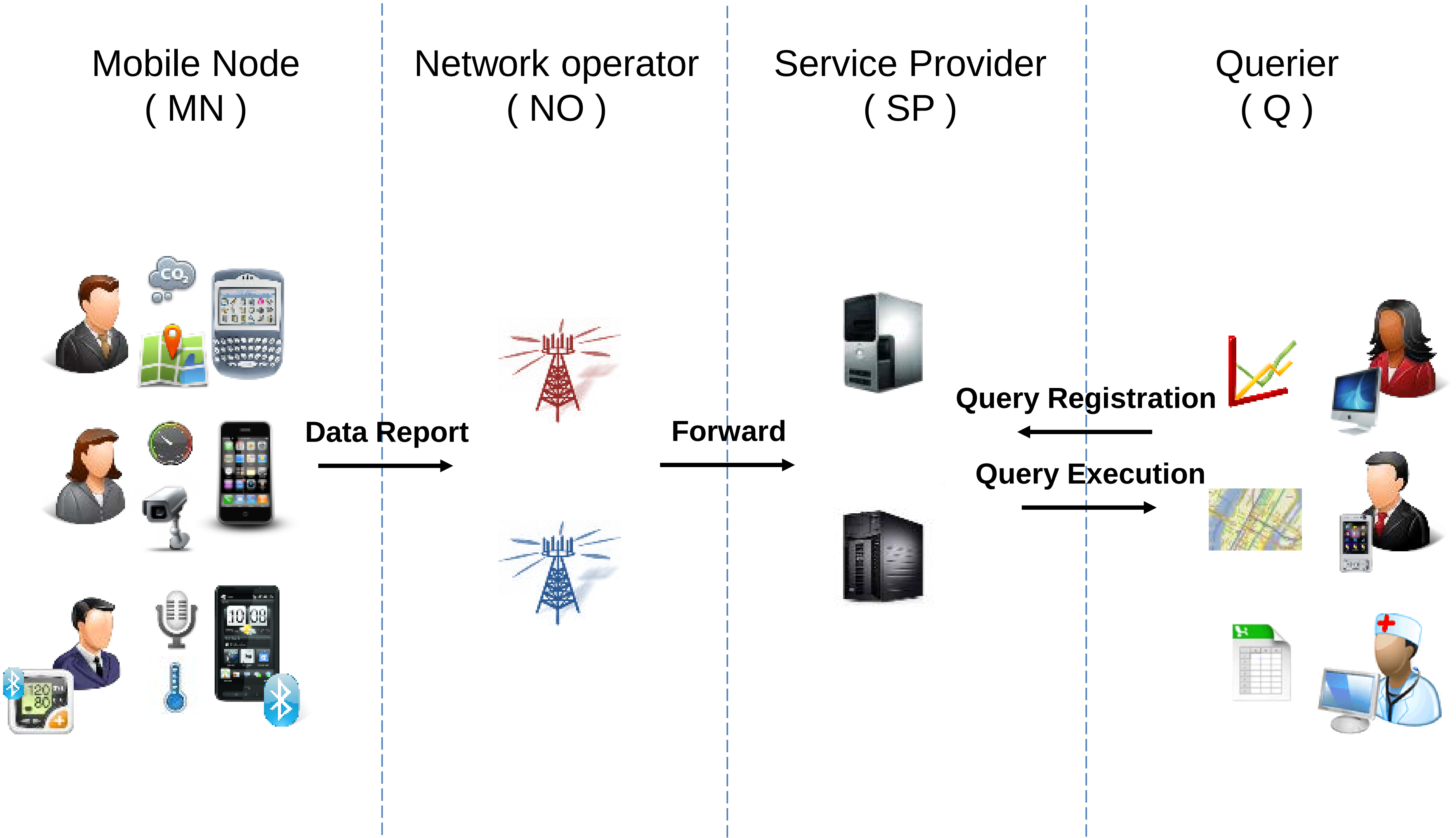


UC Berkeley

Wait… plastic surgery for WNS?

WSN	Participatory Sensing
Dull gadgets	User-carried smartphones
Poor resources	1GHz CPU
Limited battery life	Easily rechargeable
Static	Highly mobile
Network Operator owns and queries the network	Different entities co-exist and do not trust each other
Security / Dependability	Security / **Privacy**

PS (basic) architecture



Mobile Node
(MN)

Network operator
(NO)

Service Provider
(SP)

Querier
(Q)

Data Report

Forward

Query Registration

Query Execution

Parties (1)

- **Sensors**
  - Installed on smartphones
  - Emit data reports
- **Carriers**
  - People carrying their smartphone
  - Vehicles?
  - Animals?
- **Queriers**
  - Users/applications subscribing to specific information
  - *E.g., Bob interested in "Temperature in Darmstadt"*

Sensors + Carriers =
Mobile Nodes (MNs)
(*E.g., Alice's phone*)

Parties (2)

- **Network Operator (NO)**
  - Manages the network to collect and deliver reports
  - Maintains WiFi, GSM, 3G/4G, …
  - *E.g., T-Mobile*
- **Service Provider (SP)**
  - Intermediary between nodes and queriers
    - They have no mutual knowledge
  - *E.g., ps.google.com*

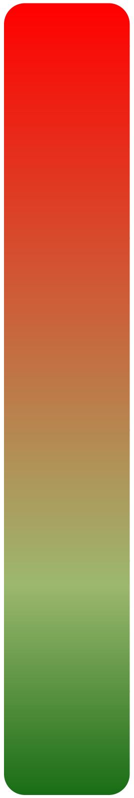
Participatory Sensing goes “live” if:

- **Users are motivated to participate**
  - Need to design appropriate business models
  - Game-theoretical models
  - Discounted data plans

- **Privacy is protected**
  - If users feel their privacy is endangered they won’t participate
  - Privacy of users reporting information
  - Privacy of users accessing/querying information

Privacy in PS

- Crypto and alike
  - Encryption, perturbation, aggregation
- Regulation
  - Who can access what, retention, etc.
- Legibility
  - Help users decide what to share and when



Pictures and Videos

- Where you are
- Who’s with you

Sound

- Personal opinions
- What you are doing

Location and Time

- GPS, WiFi AP

Biometric data

- Health condition

Acceleration

- Activity

Challenges

Shilton – Comm. ACM’09
Kapadia et al. @ COMNETS’09
Christin et al. @ ICCCN’10
Christin et al. – JSS’10

User studies

Klasnja et al @ Pervasive’09
Brush et al. @ UbiComp’10
Raij et al. @ CHI’11

Security and Privacy in PS (related work)

- Report integrity
  - Dua et al. @ HotSec’09
  - Gilbert et al. @ HotMobile’10
    - TPM-based
- Privacy-preserving aggregation
  - Dua et al. @ Securecomm’11
    - Correct behaviour of Aggregator
  - Shi et al. @ Infocom’10
    - Secret sharing based
  - Ganti et al. @ SenSys’08
    - Perturbed data w/ application-specific distribution
- Location Privacy
  - Huang et al. @ Percom’09

Anonysense (Cornelius et al. @ MobySys'08, PMC'10)

- On the plus side
  - (probably) 1st attempt to provide privacy to PS
  - AnonyTL – general purpose tasking language
  - Full implementation

- Goals
  - Carrier privacy
    - Narrow Tasking
    - Tasking de-anonymization
    - Report de-anonymization
    - Selective tasking
    - Report analysis
    - Local eavesdropping / Eavesdropping by collusion
  - Report Integrity
    - Tampering / Replay / Forgery

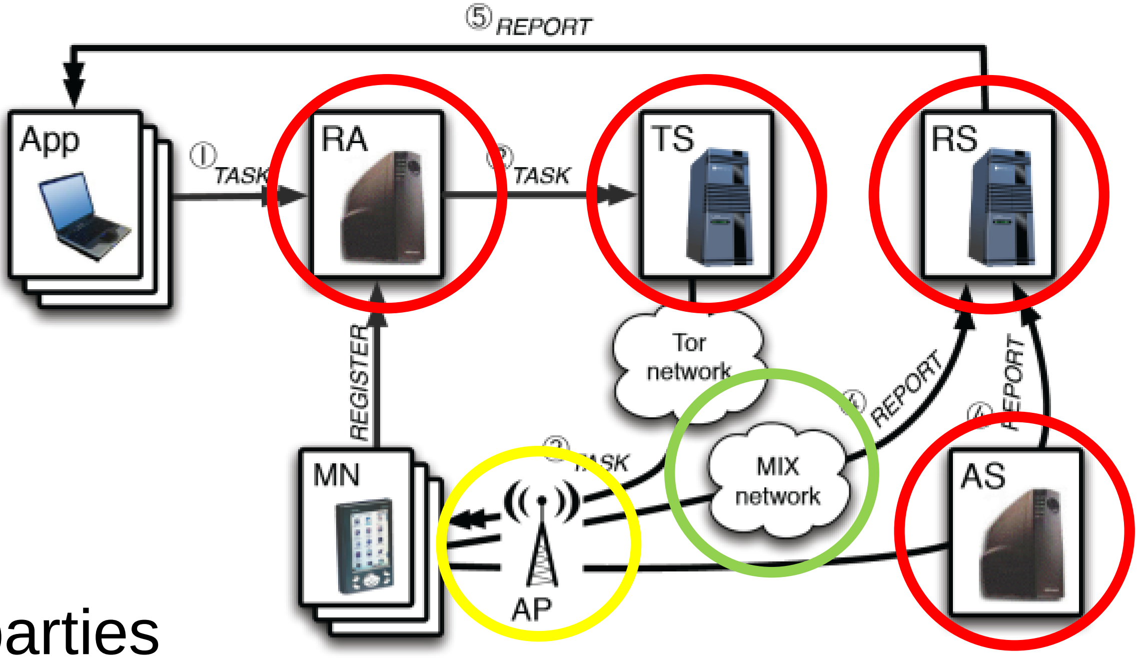
Anonysense Architecture

- Carrier privacy
  - Tor
  - MIX networks
  - AS
- Report integrity
  - Group signatures
- WiFi-based
- Many semi-trusted parties
- No provable privacy

Registration Authority

Task Service

Report Service



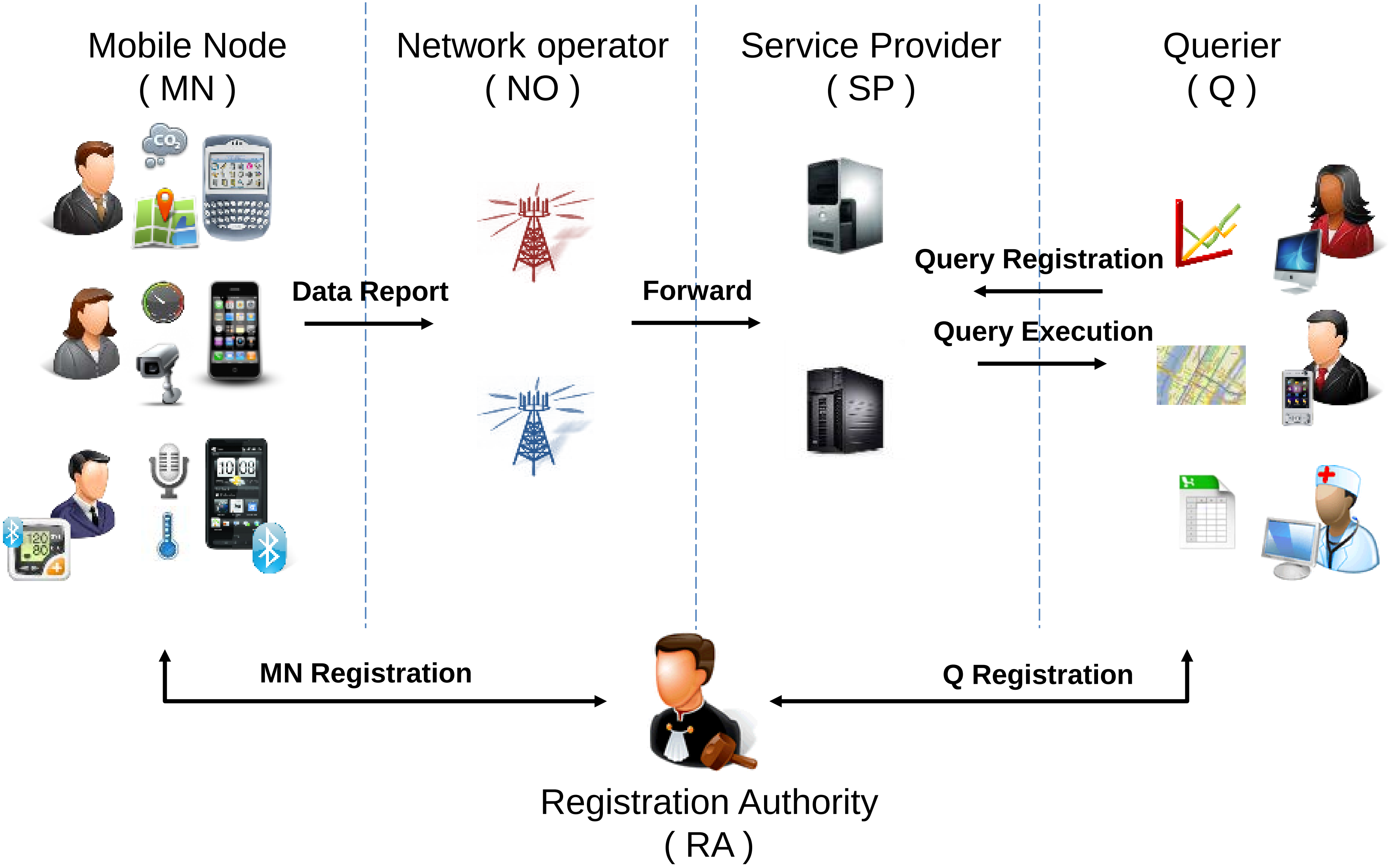
Anonymization Service

PEPSI:
Privacy-Enhanced Participatory Sensing Architecture

- Joint work with E. De Cristofaro (PARC)

- Goals
  - Cryptographic “treatment” of PS
  - Protect the privacy of data producers/consumers
    - Provable guarantees
  - Realistic architectural assumptions
    - Minimize overhead

PEPSI architecture



Mobile Node
(MN)

Network operator
(NO)

Service Provider
(SP)

Querier
(Q)

Data Report

Forward

Query Registration

Query Execution

MN Registration

Q Registration

Registration Authority
(RA)

Privacy Requirements (1)

- **Soundness**
  - No false positive/false negative

- **Query Privacy**
  - Protects the query *q* subscribed by Q
  - The NO, the SP, any MN, or any other Q, learn no information about *q*
  - (Optional) Not even the RA

- **Node Privacy**
  - Protects the data report *D* contributed by MN
  - The NO, the SP, the RA, any MN, any unregistered Q, learn no information about *D*

Privacy Requirements (2)

- **Report Unlinkability**
  - No party can link two or more reports as originating from the same MN
  - Seems impossible to achieve w.r.t. the NO in cellular networks

- **Location Privacy**
  - No party can infer “who is where”
  - Again, seems impossible to achieve w.r.t. the NO in cellular networks

PEPSI intuition

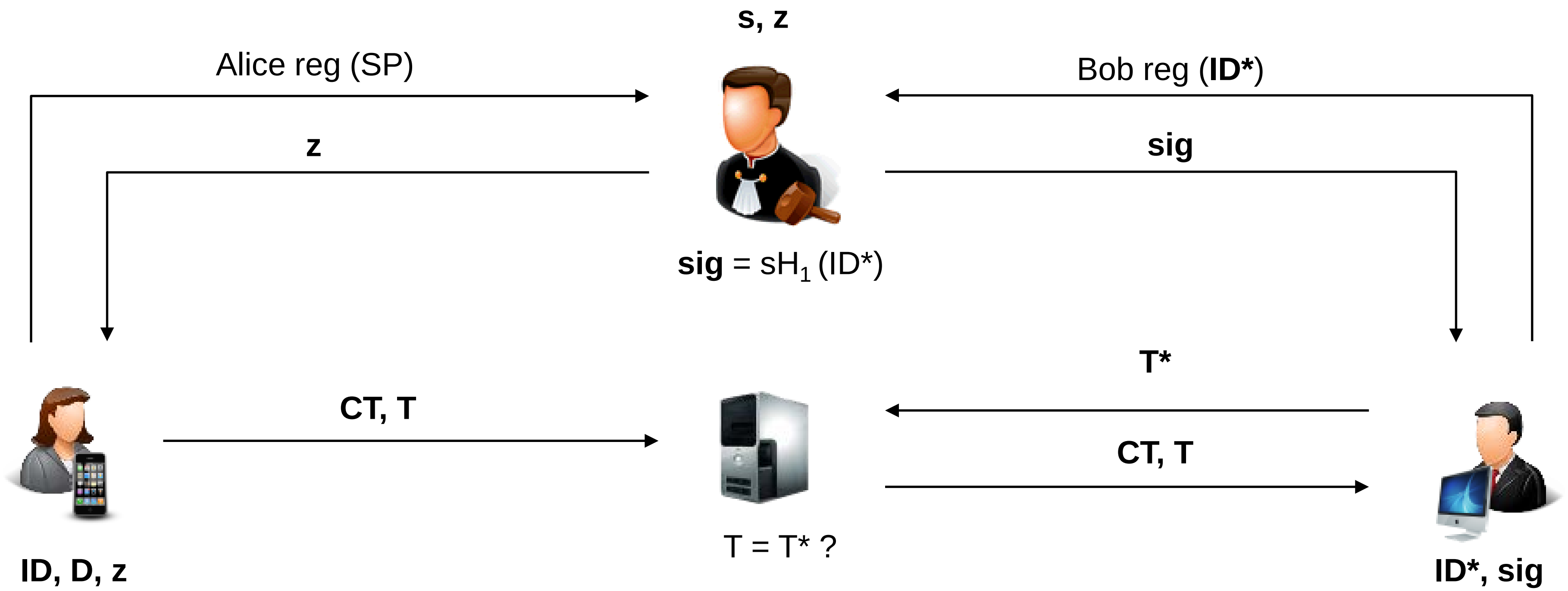
- **Hide Reports and Queries**
  - Cannot be transmitted in-the-clear, need to encrypt
  - SP needs to match queries *blindly*
- **Naïve Solutions:**
  - Queriers/Mobile Nodes share a pairwise key
  - Use public-key encryption
- **Main problem (and main intuition)**
  - Queriers and Mobile Nodes do not interact/know each other
  - We can use **Identity-based Encryption** (e.g., Boneh-Franklin):
    - *Query identifiers* are like *identities*
    - Encrypt under the identity
    - Decrypt if authorized (in possession of the corresponding secret key)

Protocols 1

- **Setup** – executed by RA on input security parameter *λ*
  - Prime *p*
  - Groups G_1 and G_2 (of order *p*)
  - *e*: G_1xG_1-->G_2 (bilinear map)
    - $e(aU,bV) = \mathrm{e(U,V)}^{ab}$
  - *s* random in G_1 (**secret master key**)
  - *z* random in G_1 (**periodically refreshed**)
  - *P* random in G_1
  - H_1:{0,1}* --> G_1, H_2:$\{0,1\}^{G2}$ --> $\{0,1\}^{\lambda}$, H_3:$\{0,1\}^{G2}$ --> $\{0,1\}^{\lambda}$
  - Public parameters: *e*, *P*, *Q*=s*P*, *R*=z*P*, H_1, H_2, H_3

Protocols 2

Public params = P, Q=sP, R=zP, H_1, H_2, H_3



s, z

Alice reg (SP)

z

Bob reg (**ID***)

sig

sig = $sH_1(ID^*)$

CT, T

T*

CT, T

T = T* ?

ID, D, z

ID*, sig

k $= H_3[e(Q, z\,H_1(ID))]$

CT $= Enc_k(D)$

T $= H_2[e(Q, z\,H_1(ID))]$

T* $= H_2[e(R, sig)]$

k* $= H_3[e(R,sig)]$

D $= Dec_{k^*}(CT)$

$$T^* = H_2[e(R, sig)] = H_2[e(zP, sH_1(ID^*))] = H_2[e(P, H_1(ID^*))^{sz}]$$

$$T = H_2[e(Q, zH_1(ID))] = H_2[e(sP, zH_1(ID^*))] = H_2[e(P, H_1(ID^*))^{sz}]$$

Privacy

- **Node Privacy**
  - Only authorized queriers in possession of valid **sig** obtain information on (**T,CT**)
  - Reduction to CPA-security of Boneh-Franklin’s IBE

- **Query Privacy**
  - No one (except the RA) learns any information about query interests
  - Reduction to CPA-security of IBE

Privacy (2)

- **Report Unlinkability/Location Privacy**
  - Not guaranteed w.r.t. the NO: open problem
  - The NO strips off privacy-sensitive metadata (e.g., originating cell)

- **Trust Assumption**
  - RA is trusted
  - Honest-but-Curious SP
    - Does not create phantom users
    - May collude
      - But users have no incentive in colluding

Performance Evaluation

- **Focus on mobile phones**
  - Experiments on Nokia N900 (600MHz CPU, 256MB RAM)

- **Privacy-protecting layer at MNs**
  - Compute (T,CT)
  - One bilinear map pairing, one AES encryption
  - Only 93*ms*

- **Overhead at other parties**
  - No overhead for SP (only matching hashed values)
  - Negligible overhead for queriers (AES decryption)

Open Problems

- **Query privacy w.r.t. the RA**
  - Blind-IBE
- **Fine-grained authorizations**
  - Hierarchical IBE
- **Work on aggregate data queries**
  - Average Temperature
  - Sum, Mean, Variance, …
  - Predicates: e.g., “sum > 20 ?”
- **Location Privacy**
  - Possible?
- **Revocation**
  - Evict malicious MNs

Questions?

- Thank you!

- More info at http://sprout.ics.uci.edu/PEPSI

- Credits: E. De Cristofaro @ PARC, Secure Mobile Networking Lab @ TU-Darmstadt