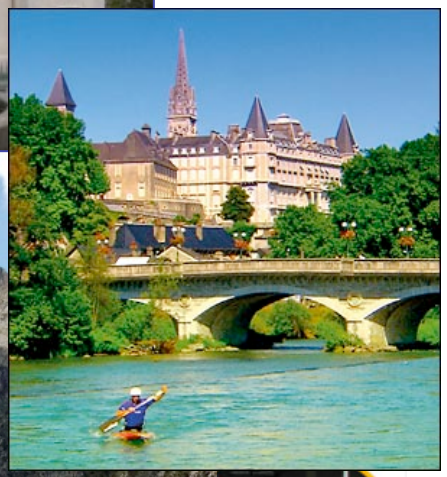
From Global Sensing in a Ubiquitous Digital Society to QoS for Cloud Computing Infrastructures

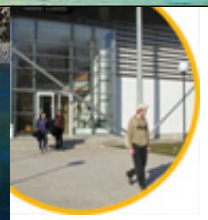
SEMINAR, GHPC group
July, 27th 2011
Bách Khoa University

Prof. Congduc Pham
http://www.univ-pau.fr/~cpham
Université de Pau, France

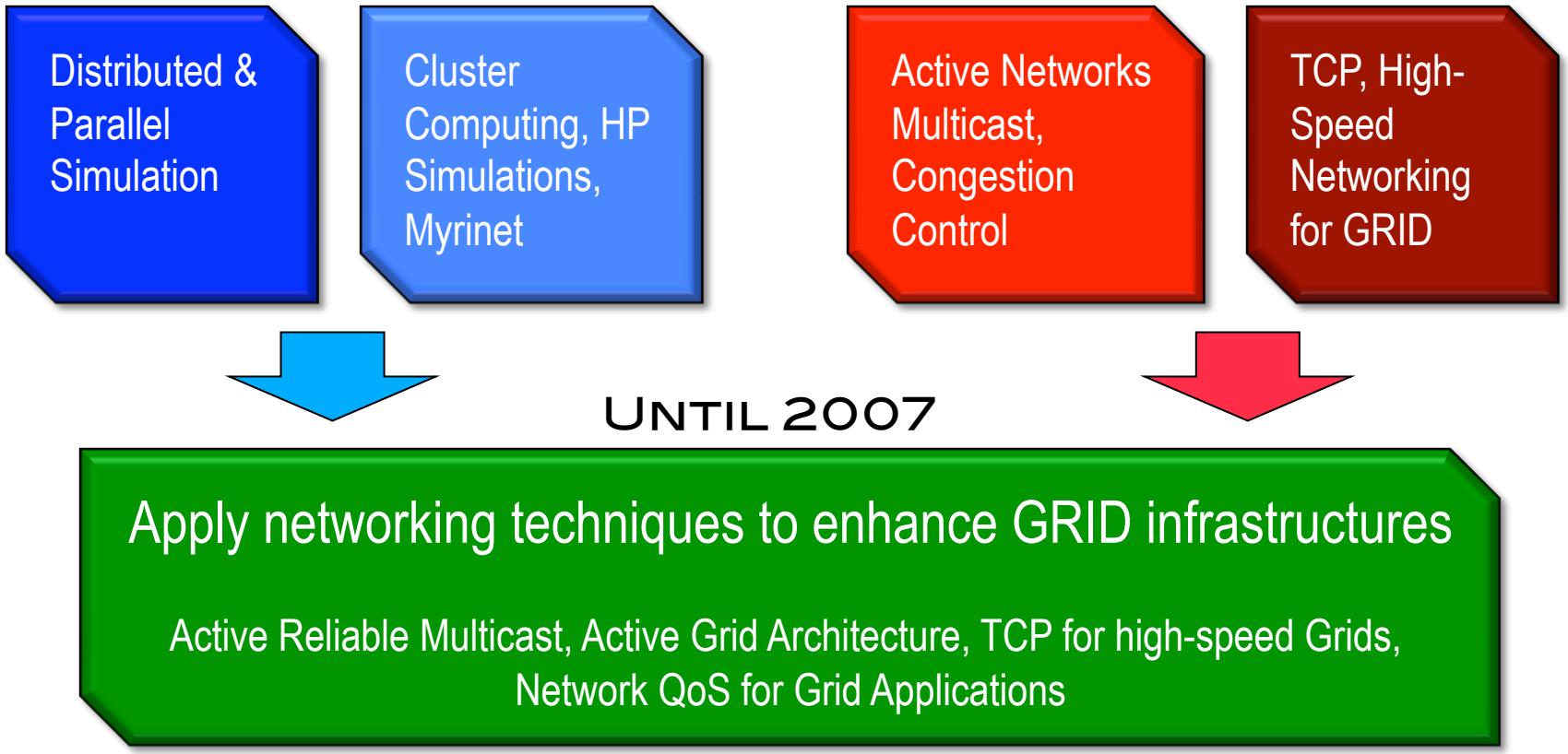
TY OF PAU

iaporama des
mpus de l'UPPA

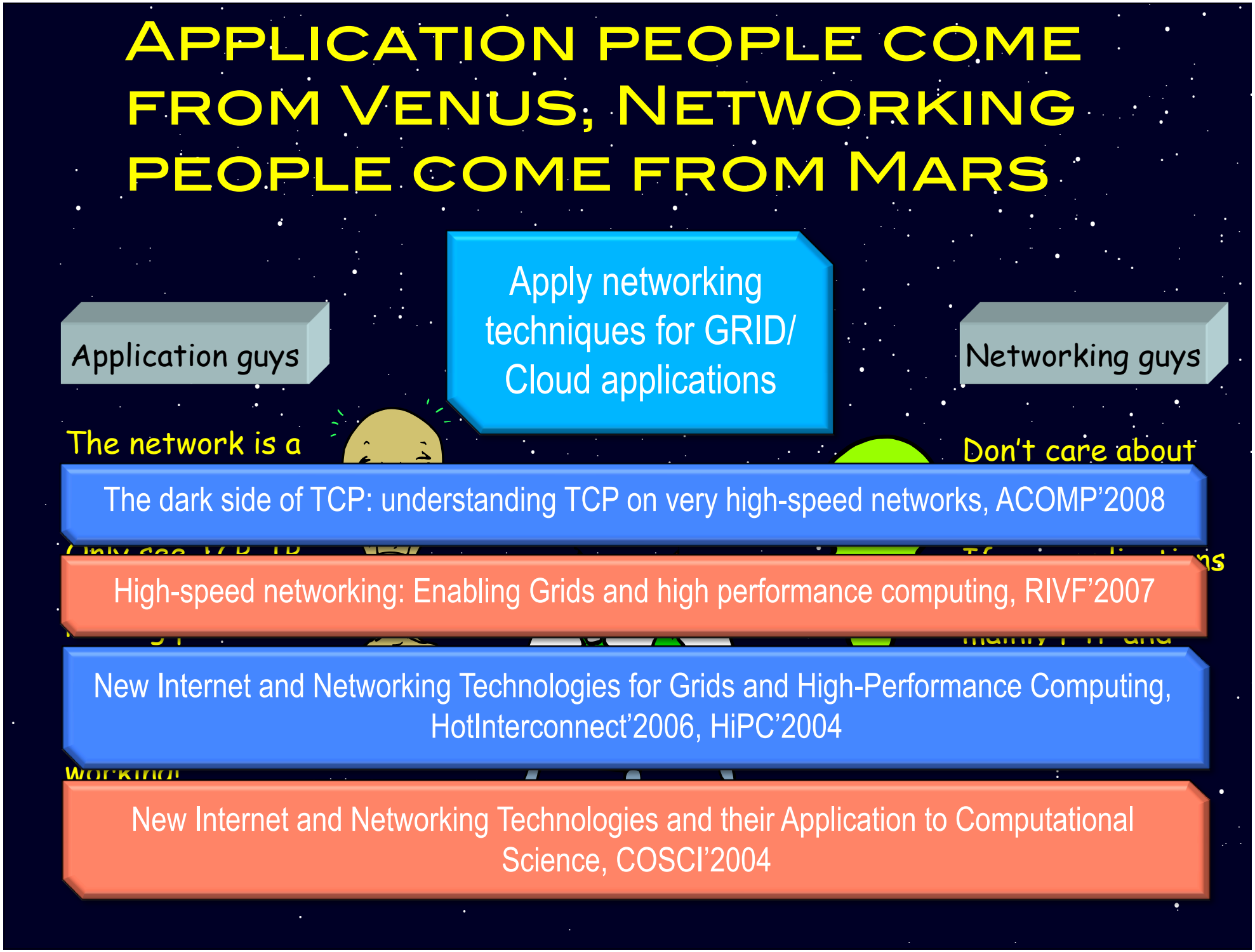




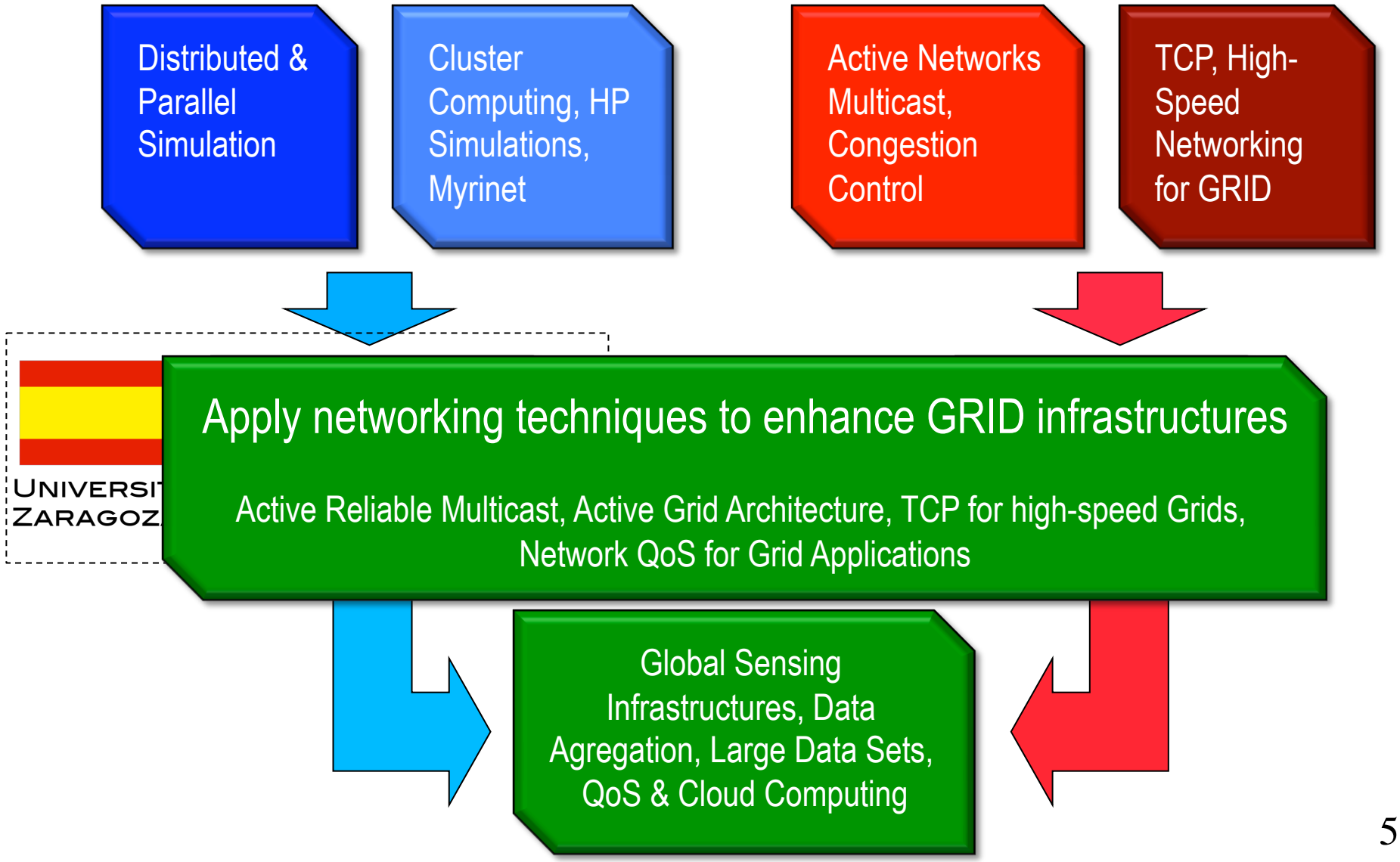
Research Interests



Application people come from Venus, Networking people come from Mars



Research Interests



GRID RELATED PROJECTS

- GRID/CLOUD FOR COMPANIES

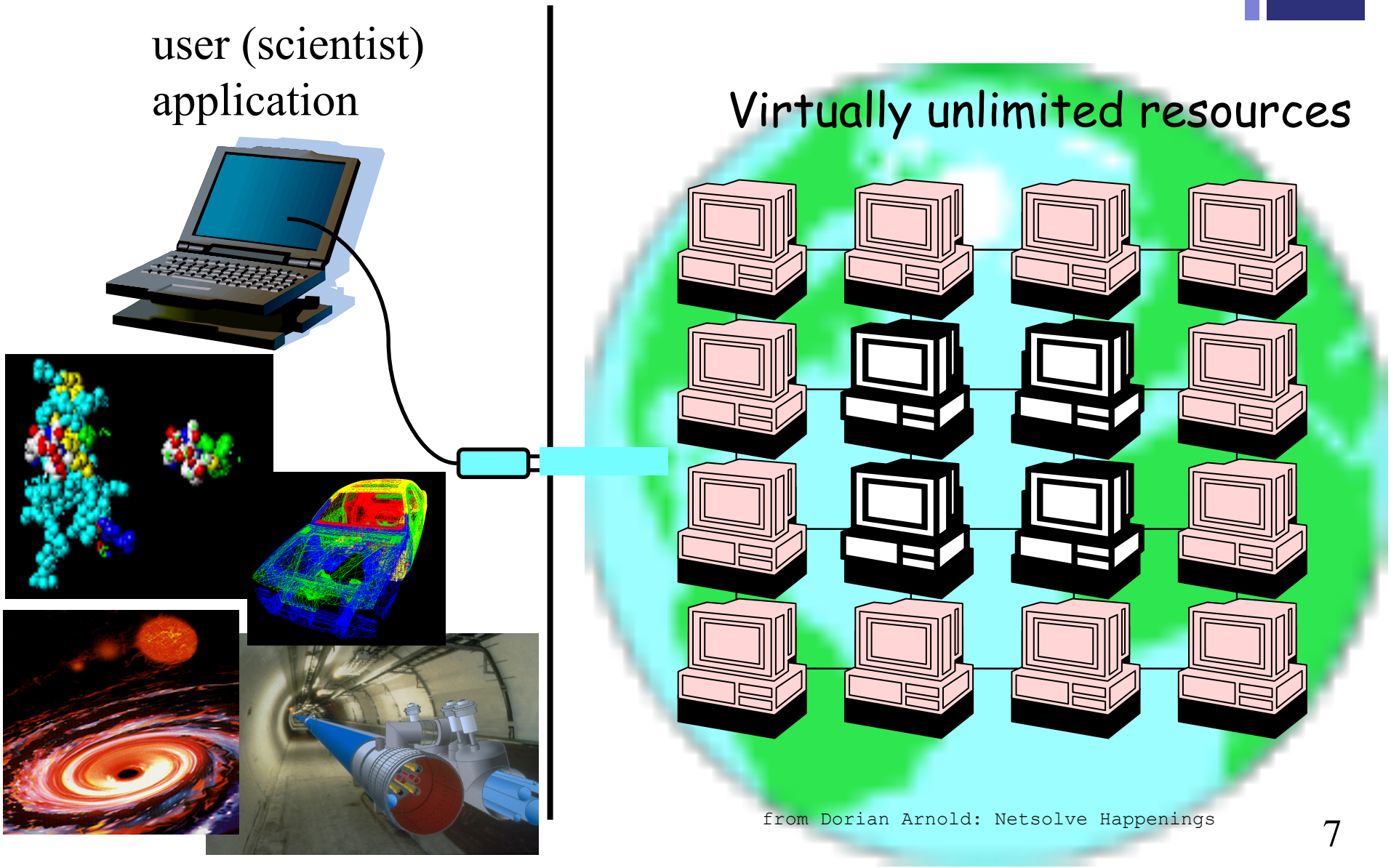
PireGrid
WWW.PIREGRID.EU

WWW.PIREGRID.EU

- LARGE DATA ISSUES (WORKFLOWS)

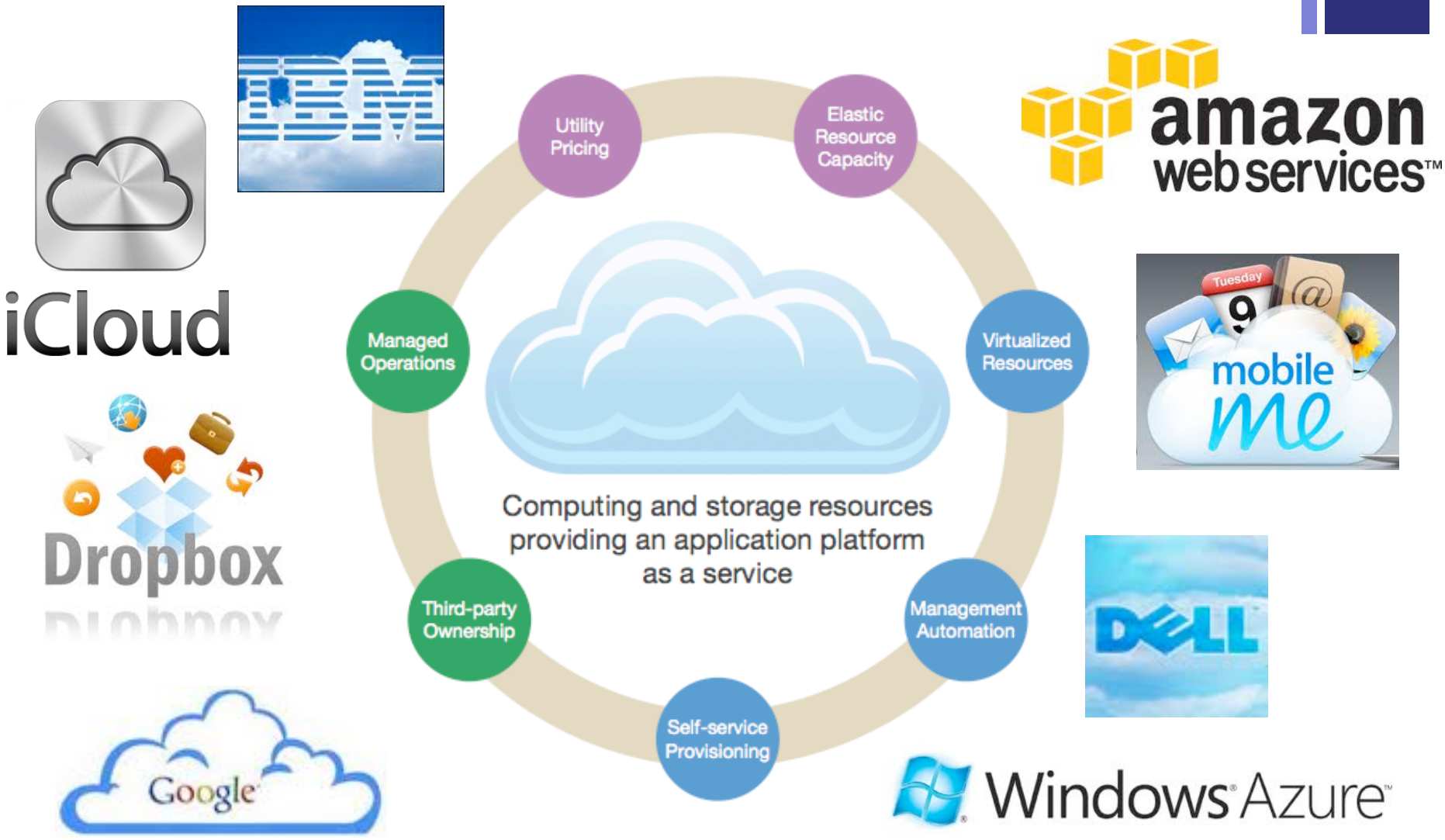
OMNIDATA

Computational grids



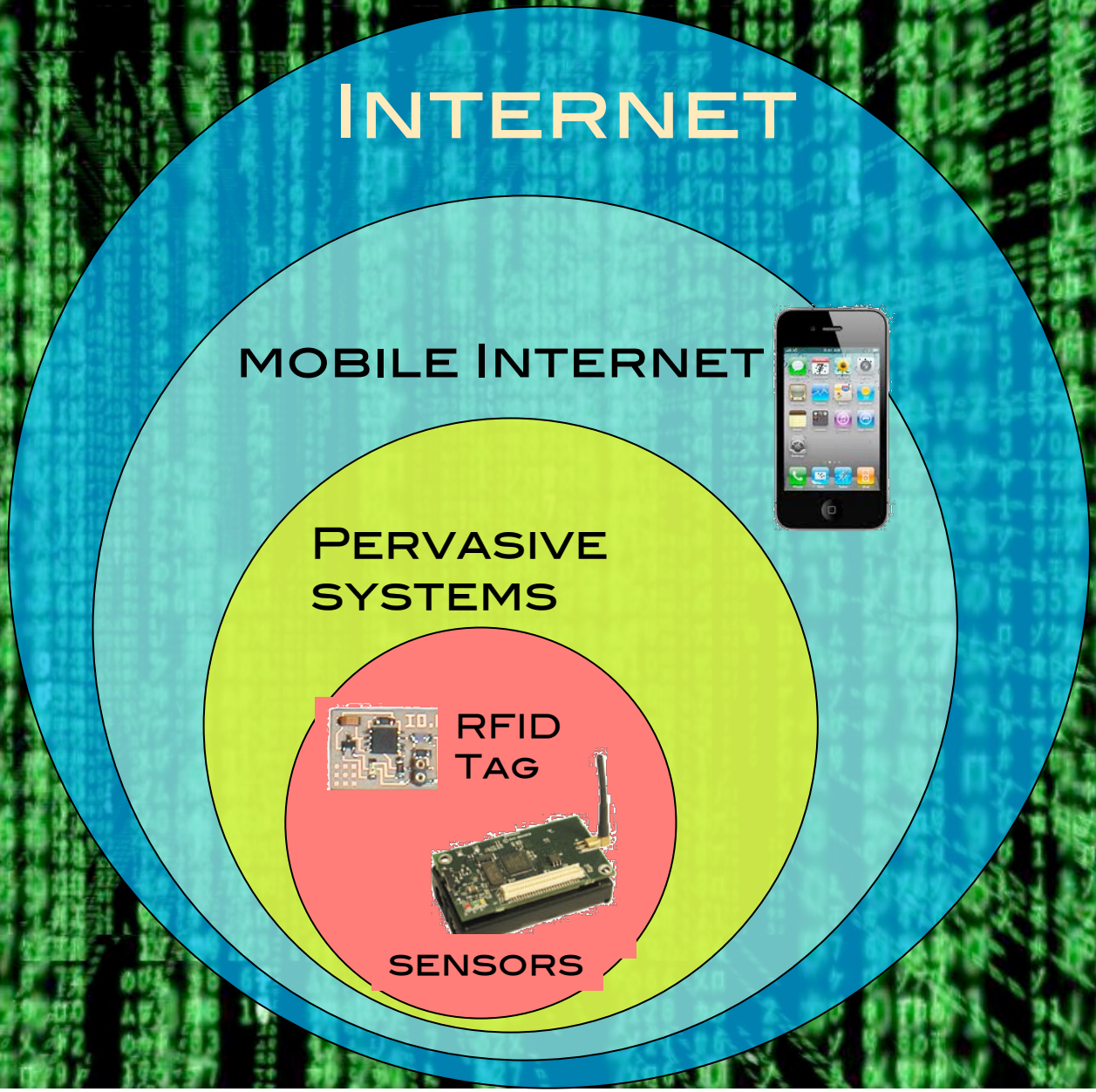
from Dorian Arnold: Netsolve Happenings

Ubiquitous Clouds

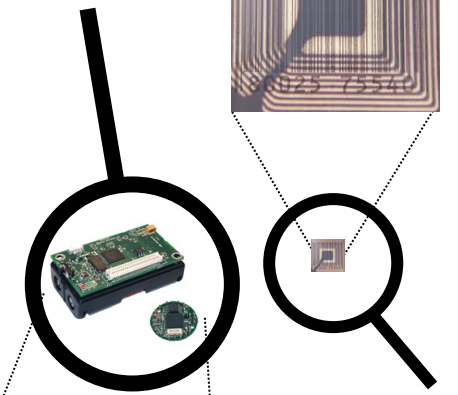


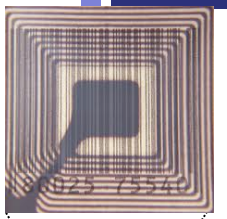
Concept: MWD Advisors, www.mwdadvisors.com

Digital Wireless World



SMALL, SMART DEVICES!





Autonomous sensors and RFID tag can be embedded in various structures or objects of our daily life to enhance localization, tracking and information collection.

From sensing...

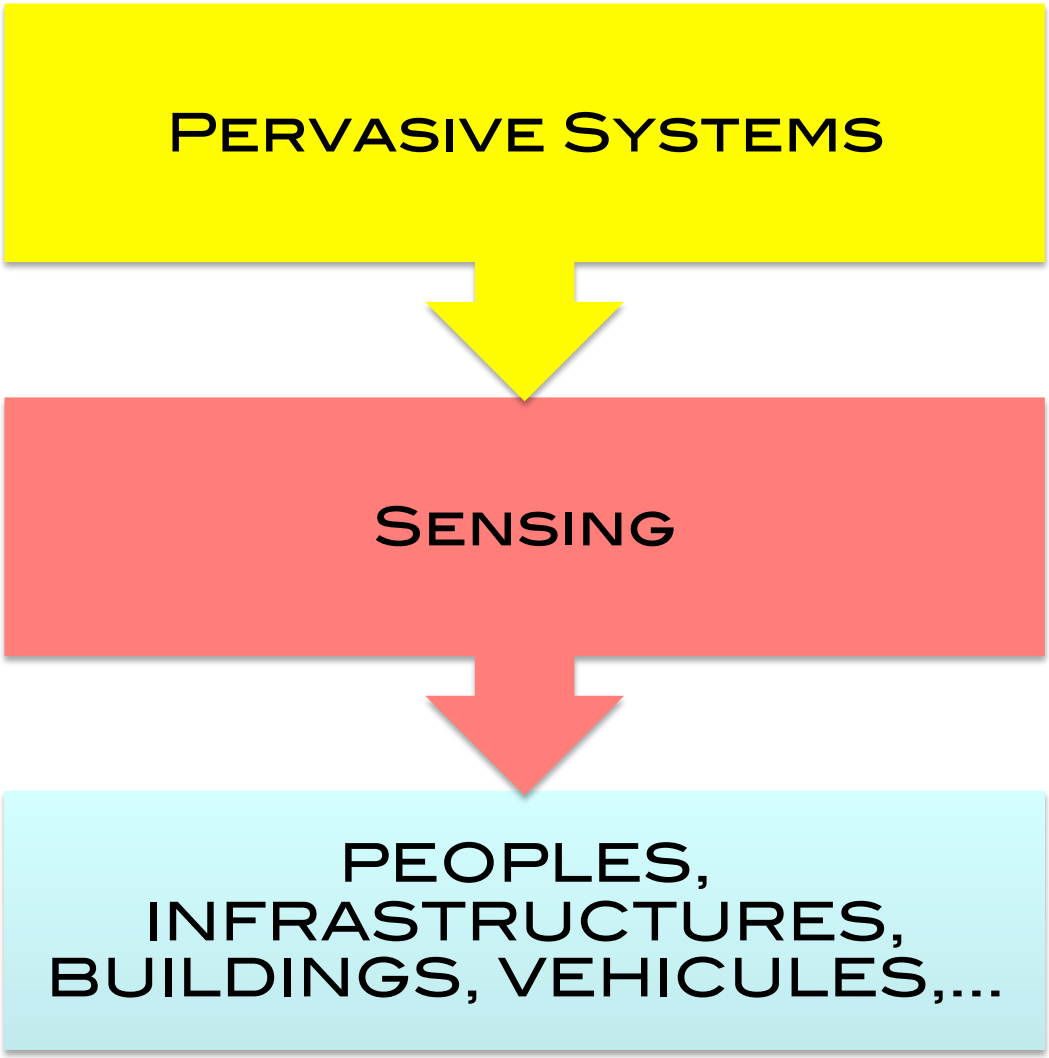
Sensing

...TO DIGITAL SOCIETY...

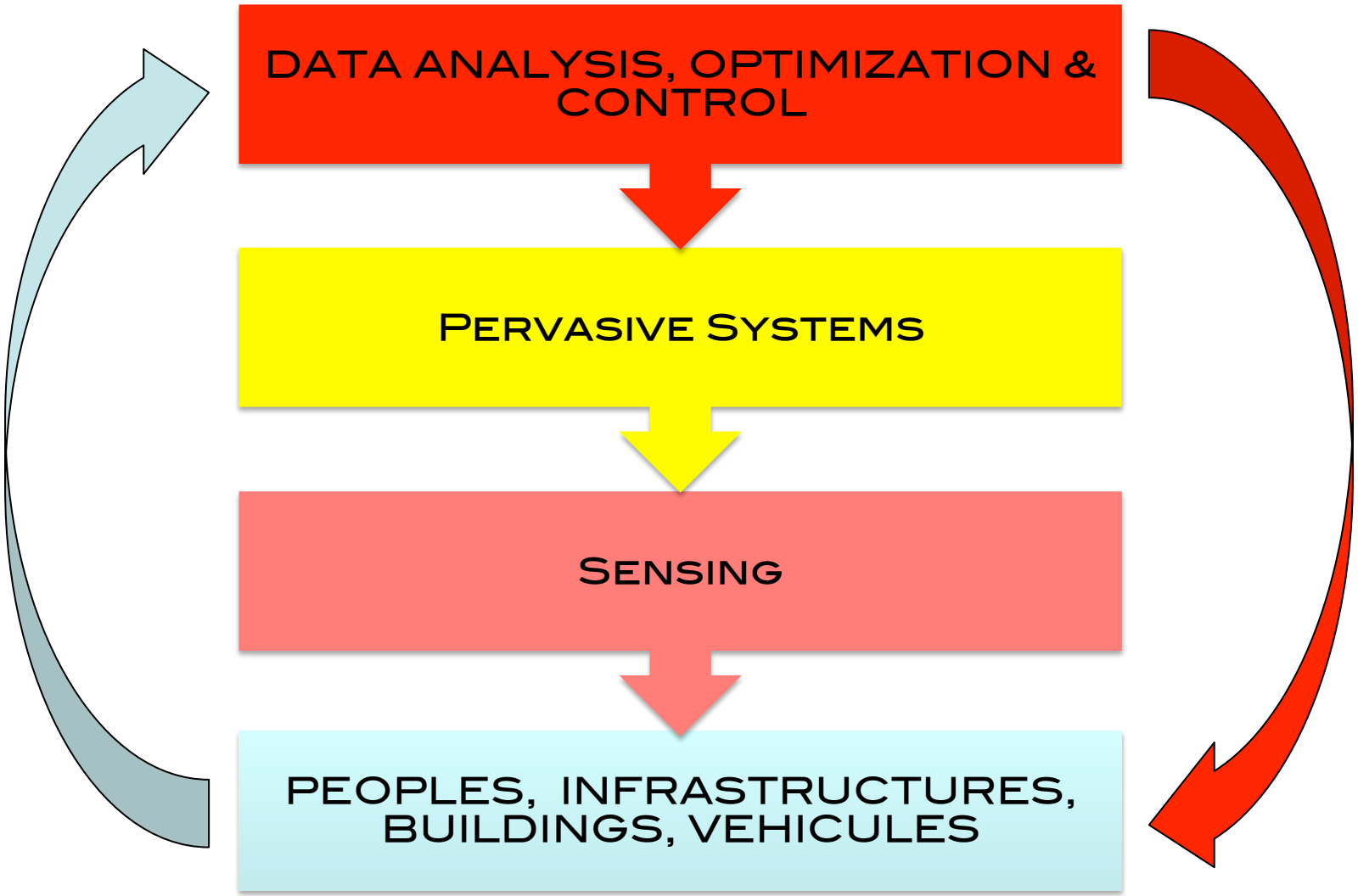
PERVASIVE SYSTEMS

SENSING

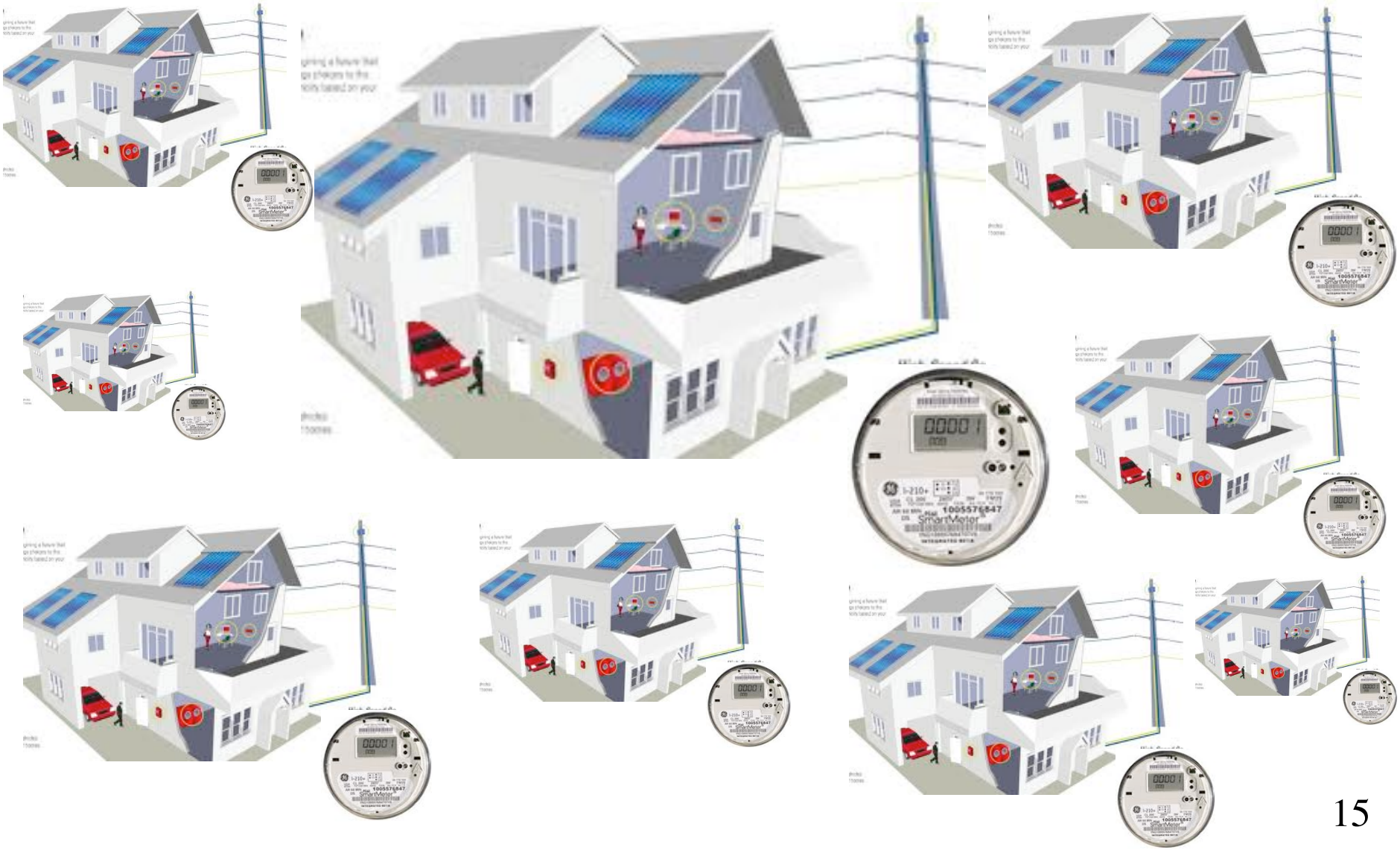
…TO UBIQUITOUS WORLD…



...TO CONTROLLED SYSTEMS.



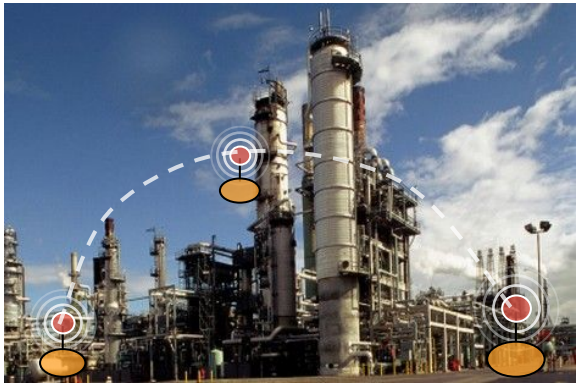
Ex: Smart electricity networks

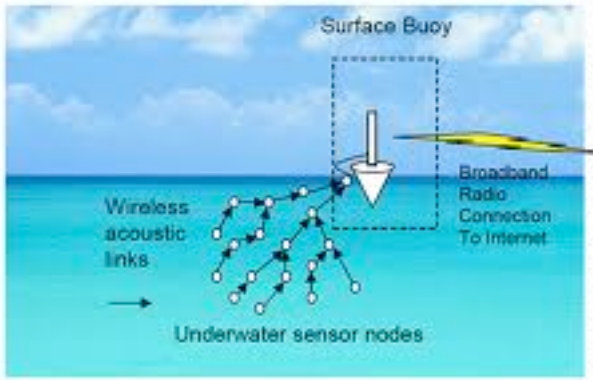


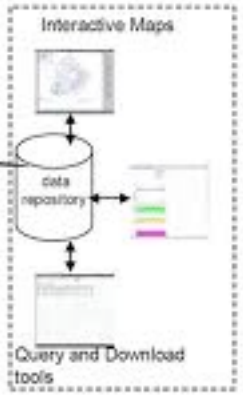
Ex: Smart electricity networks

Yogesh Simmhan, Baohua Cao, Michail Giakkoupis, and Viktor K. Prasanna. Adaptive rate stream processing for smart grid applications on clouds. In Proceedings of the 2nd ACM international workshop on Scientific cloud computing (ScienceCloud '11).

Monitoring/ Surveillance

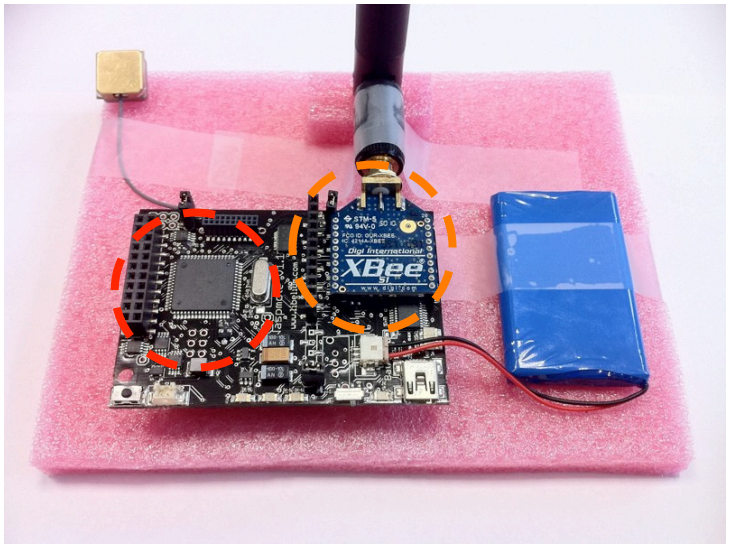


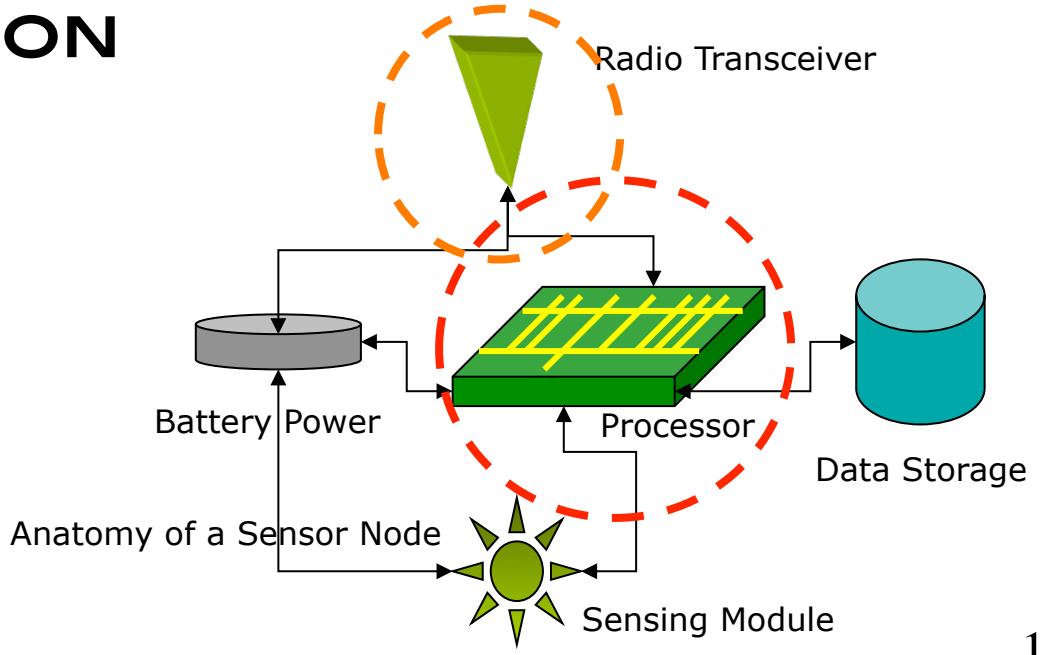




Wireless autonomous sensor

- In general: low cost, low power (the battery may not be replaceable), small size, prone to failure, possibly disposable
- Role: sensing, **data processing**, communication





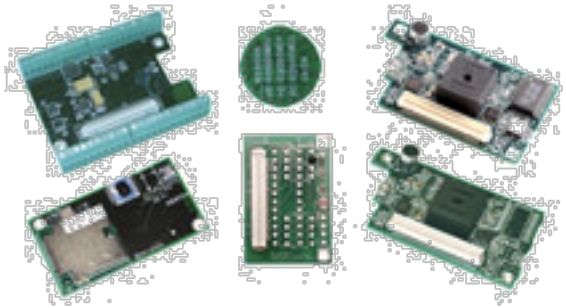
Berkeley Motes (contd.)

- Each Mote has two separate boards
  - A main CPU board with radio communication circuitry
  - A secondary board with sensing circuitry
- Decouples sensing hardware from communication hardware
- Allows for customization since application specific sensor hardware can be plugged-on to the main board

MICA2

MICAz

Imote2

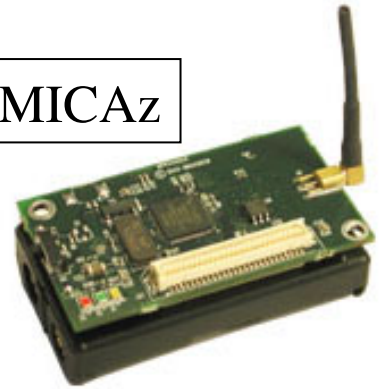


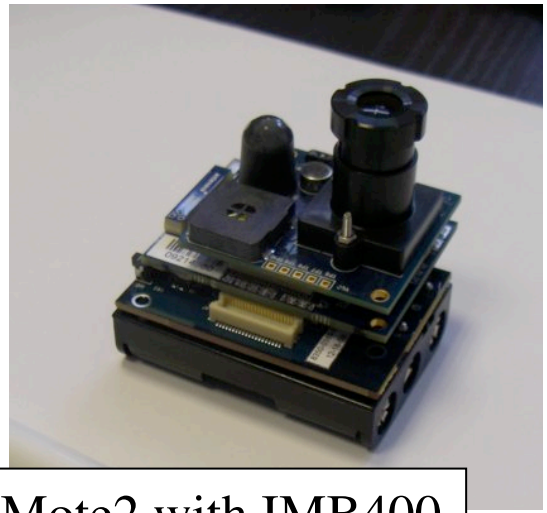
Sensing boards

Crossbow motes of our testbed

iMote2

MICAz

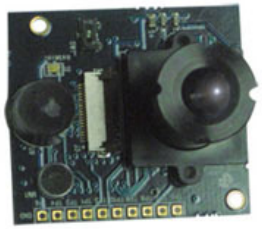




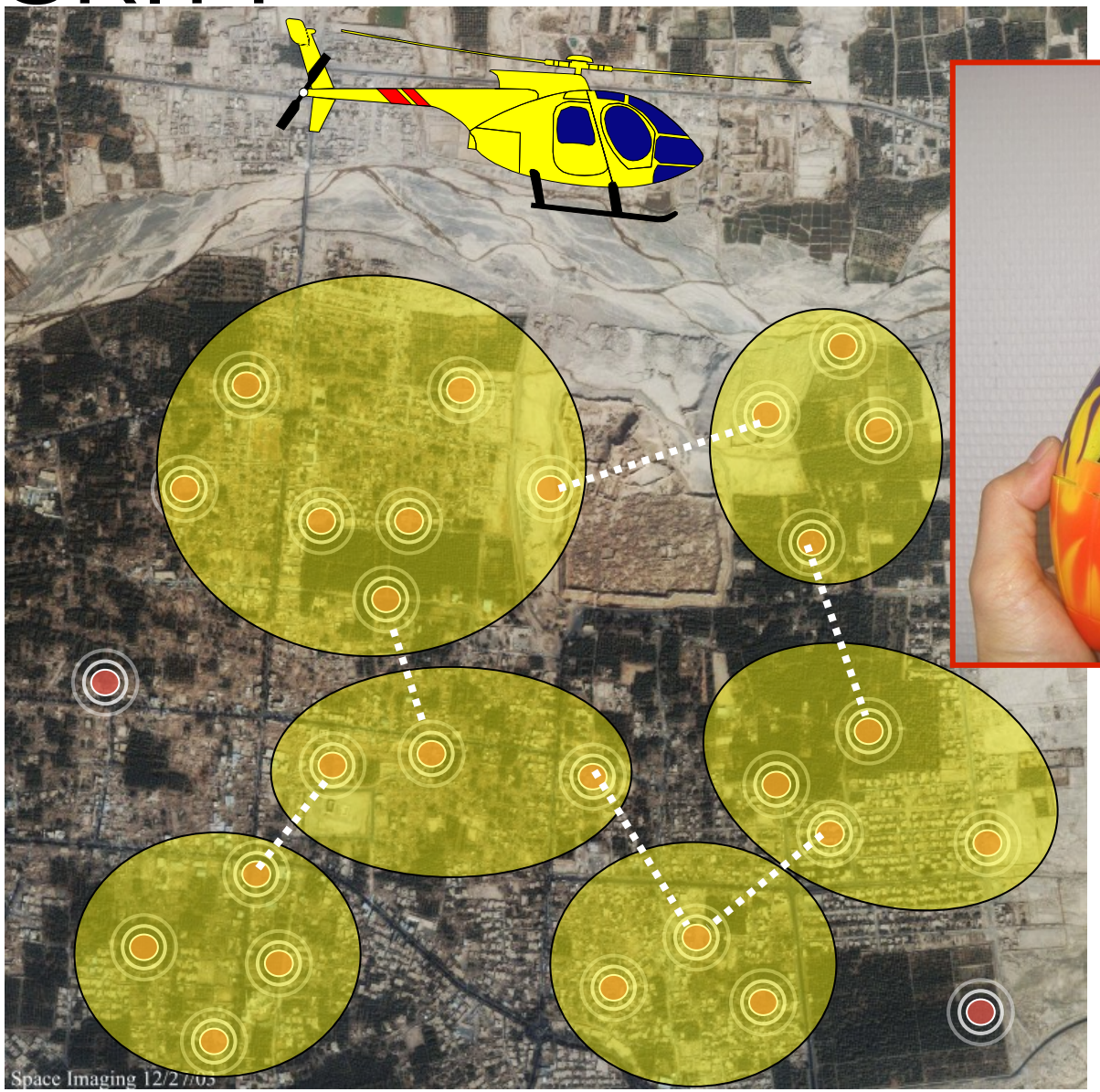
iMote2 with IMB400 multimedia board

Search&Rescue, security

Imote2

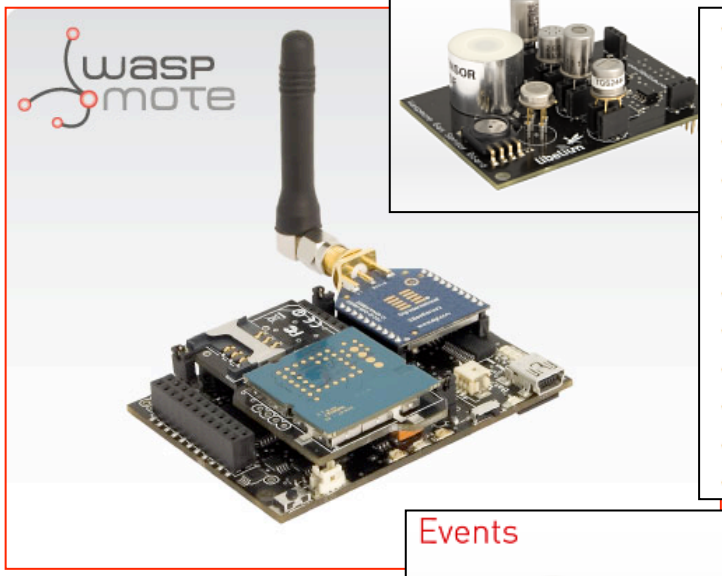


Multimedia board



libelium
comunicaciones inalámbricas distribuidas

- ATmega1281 microcontroller
- 8K RAM & 1G SD card.
- 2.4GHz IEEE 802.15.4 compatible. RF and GSM/GPRS

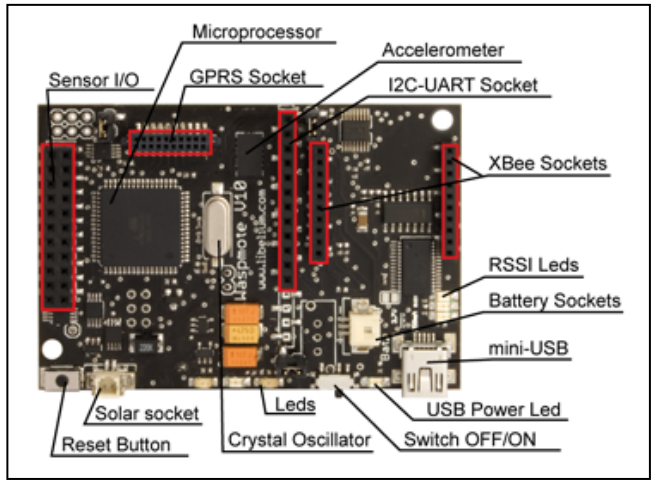


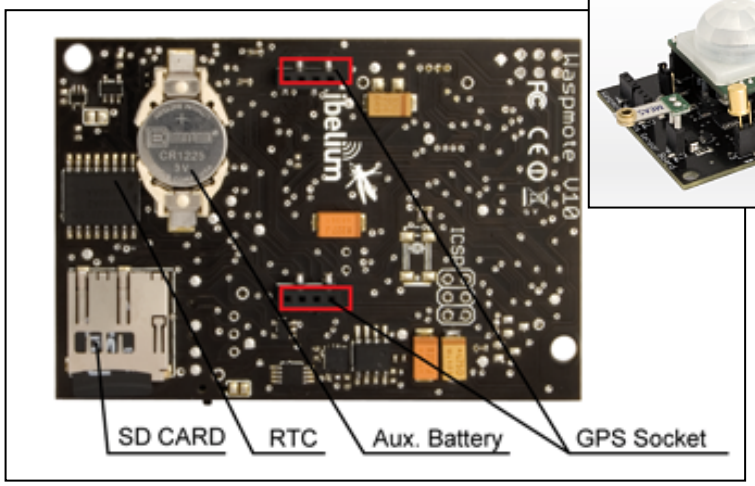
Gases

- Carbon Monoxide – CO
- Carbon Dioxide – CO2
- Oxygen – O2
- Methane – CH4
- Hydrogen – H2
- Ammonia – NH3
- Isobutane – C4H10
- Ethanol – CH3CH2OH
- Toluene – C6H5CH3
- Hydrogen Sulfide – H2S
- Nitrogen Dioxide – NO2
- Temperature
- Humidity

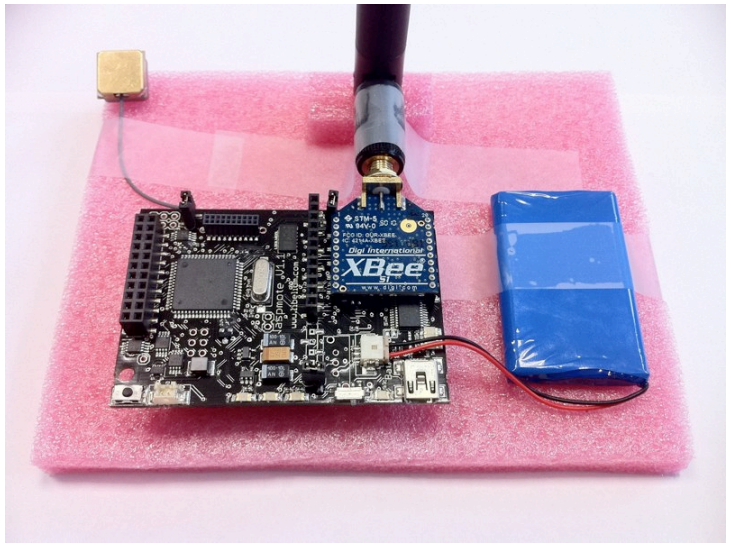
Events

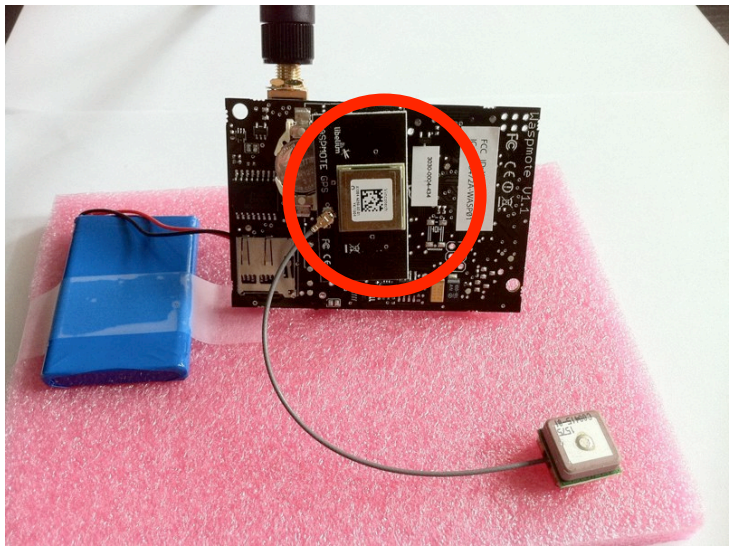
- Pressure/Weight
- Bend
- Vibration
- Impact
- Hall Effect
- Tilt
- Temperature (+/-)
- Liquid Presence
- Liquid Level
- Luminosity
- Presence (PIR)
- Stretch

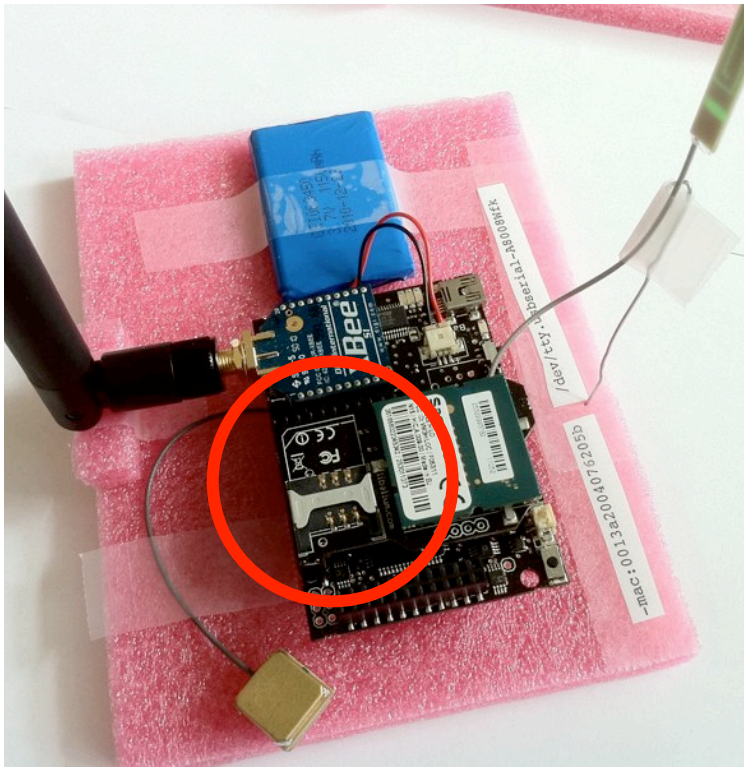




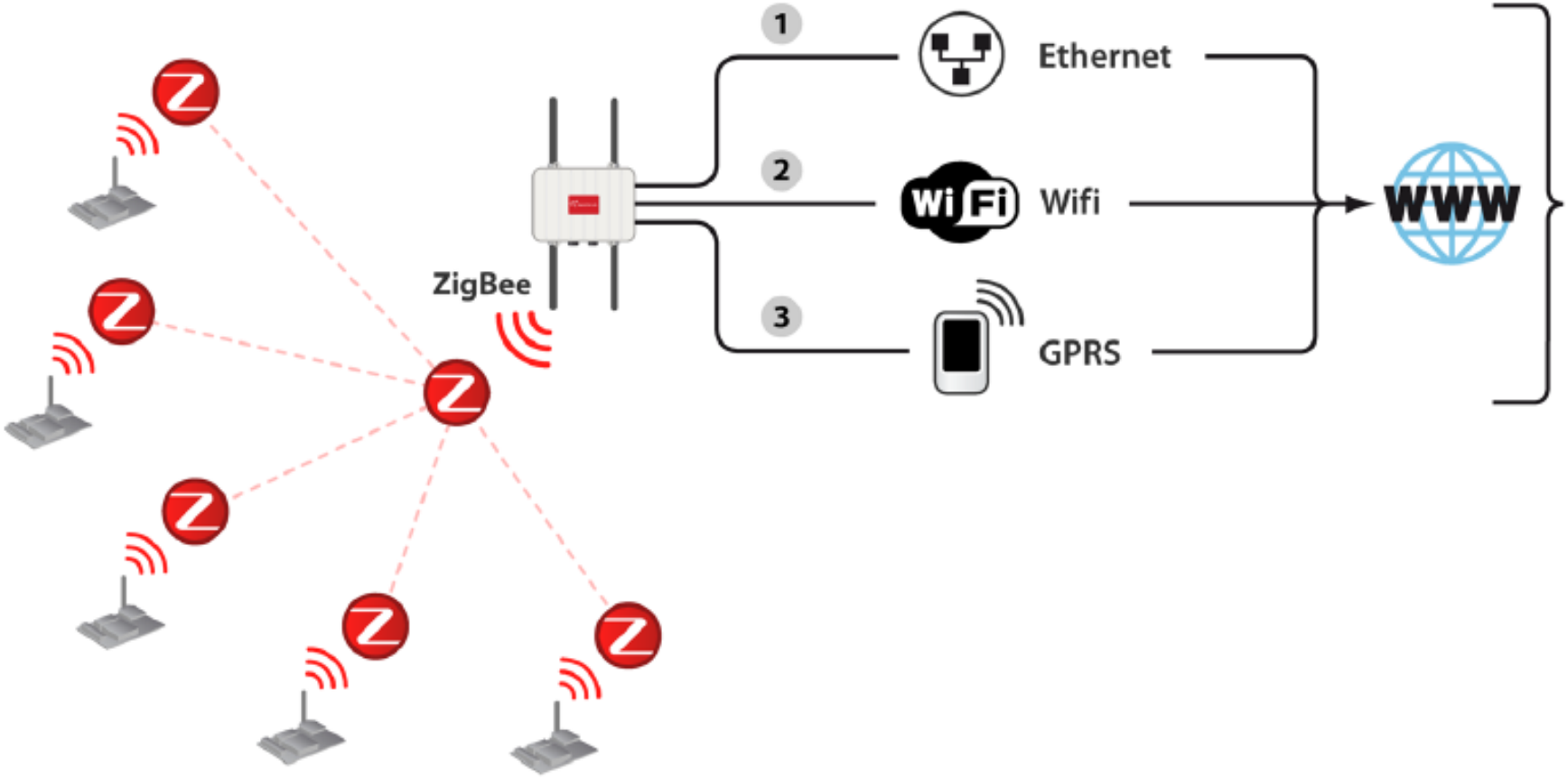
Libelium WaspMote (1)



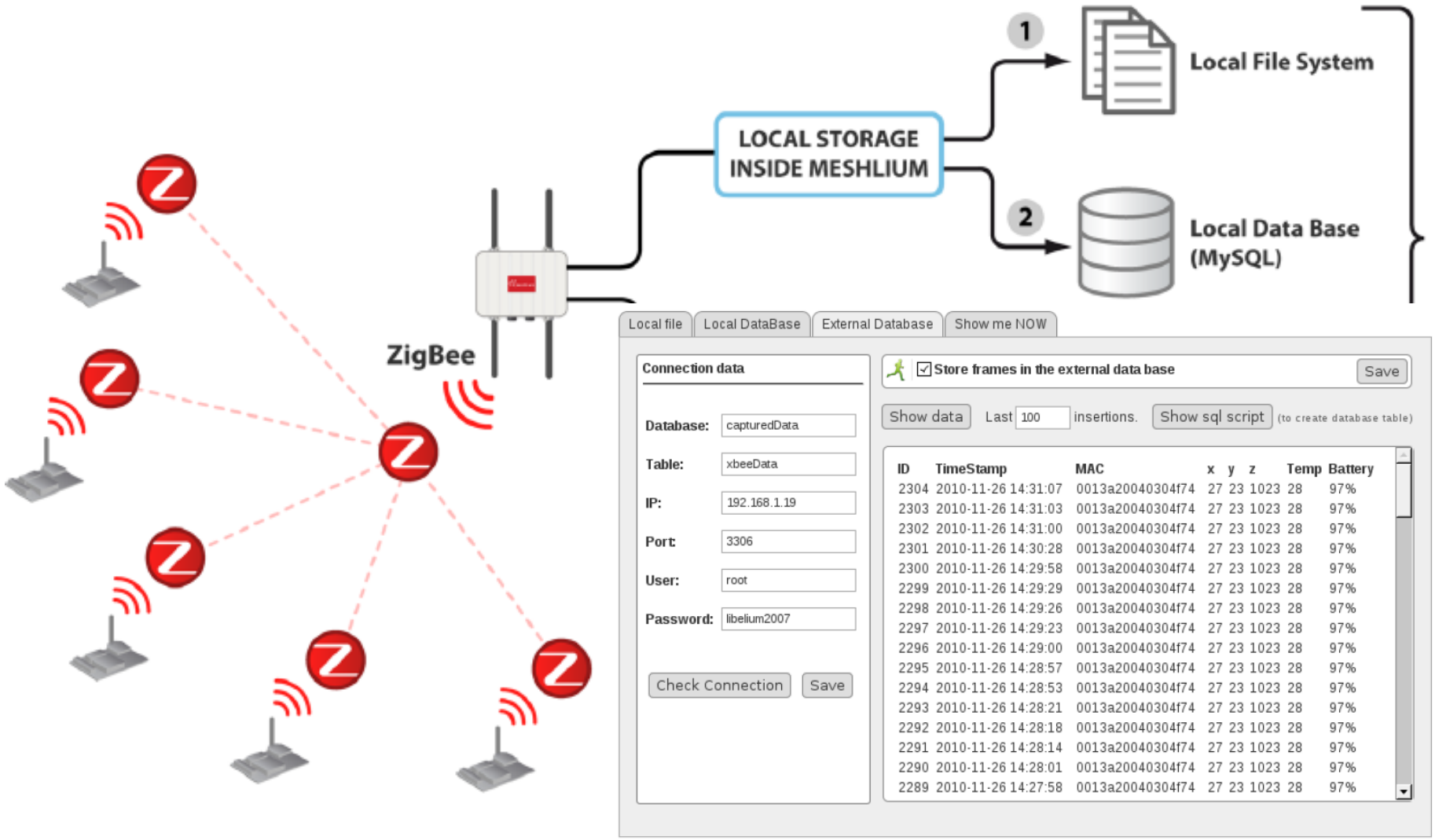




Advanced connectivity



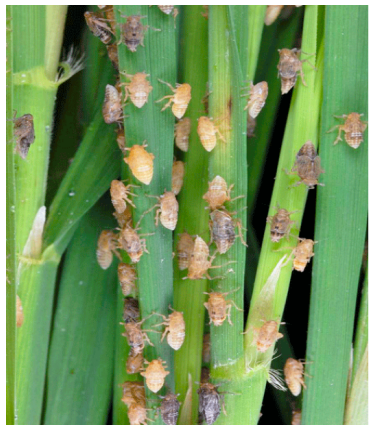
Advanced database features

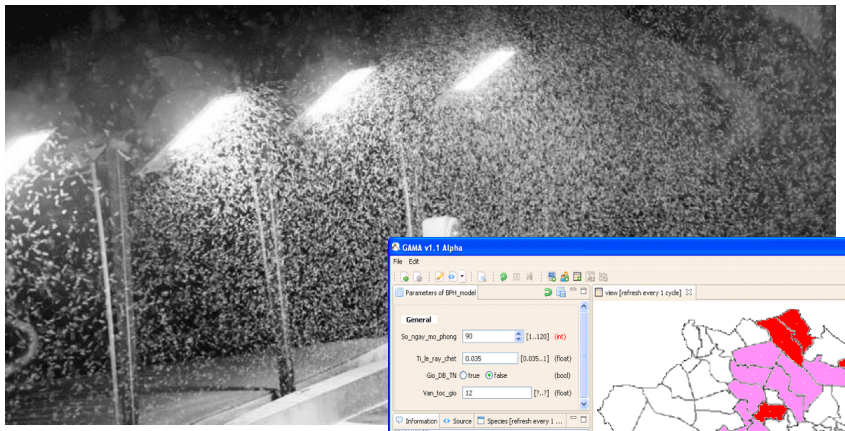


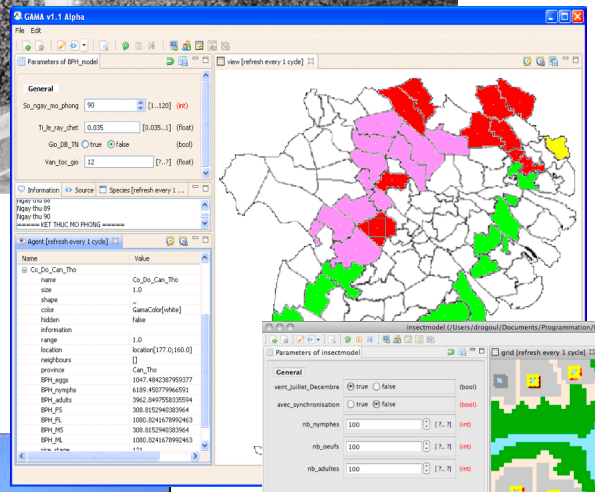
The full testbed

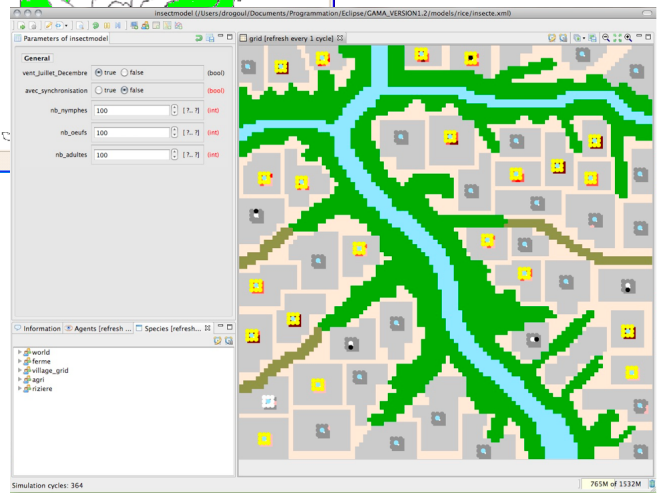
Air quality monitoring

Specific applications



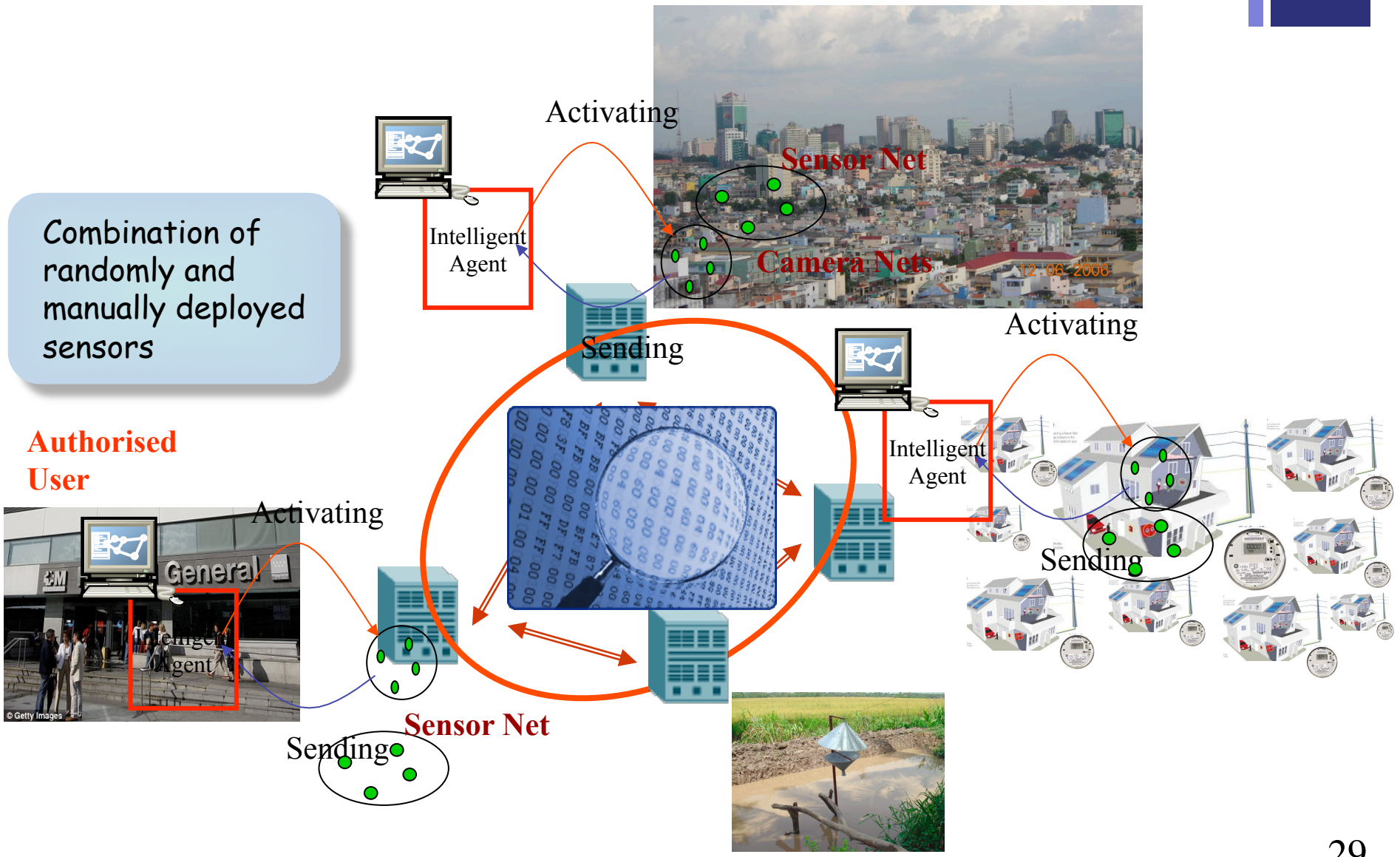




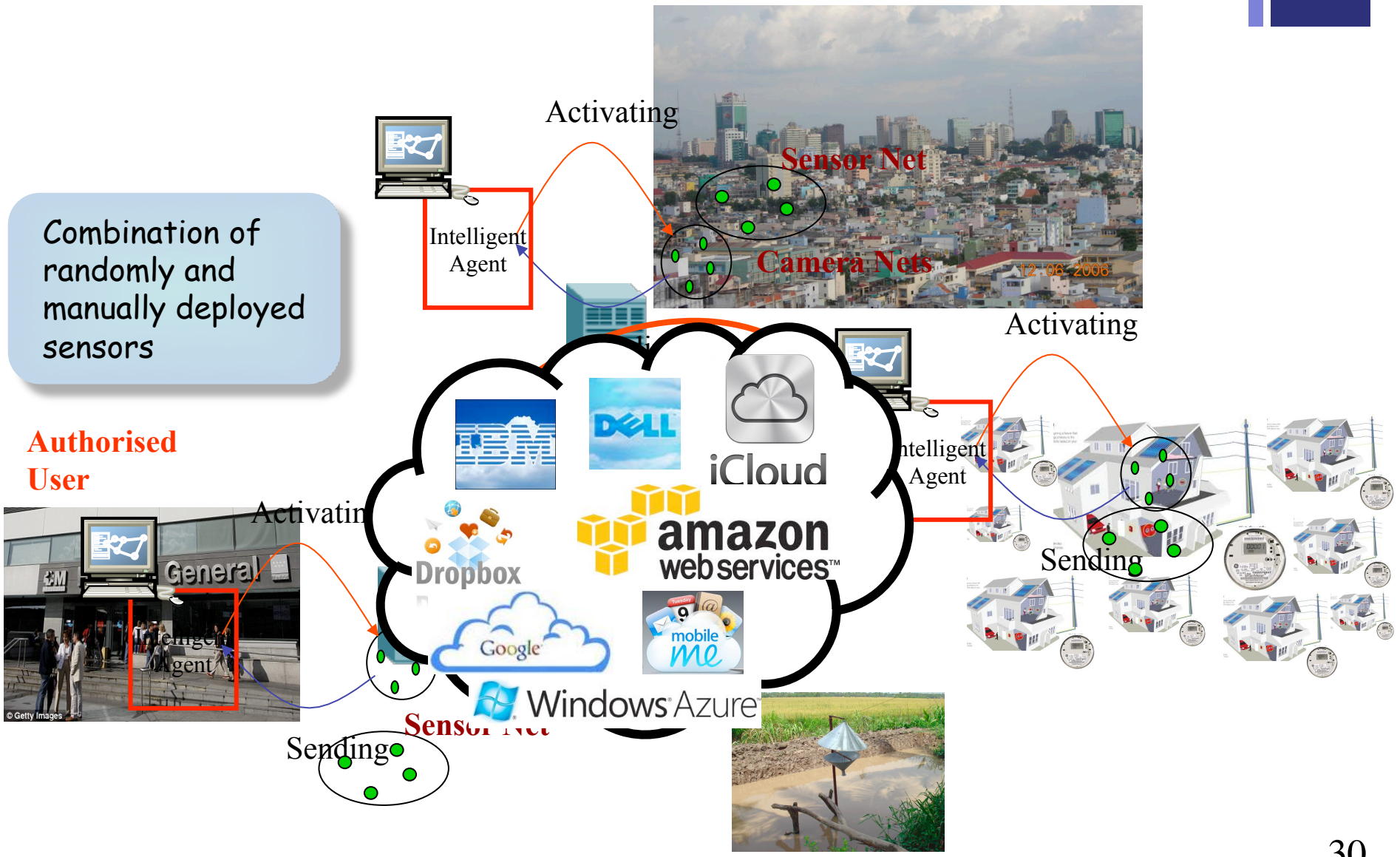


Source JEAI DREAM, U. Can Tho

Towards Global Sensing



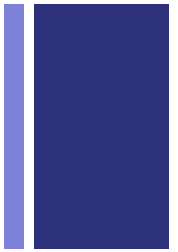
Where Clouds come in!



Cloud for every apps!

- The shared cloud assumption

What is Quality of Service?



- Quality of service is the ability to provide different priority to different applications, users, or data flows, or to guarantee a certain level of performance
- QoS criteria are numerous and is highly dependant of the app.
  - Throughput, Delay, jitter, Loss rate, Availability, Uptime, ...
- ... Or driven by the end-user
  - Image resolution, sound quality, appropriate language, ...

QoS, SLA (1)

From V.D. Cunsolo et al. « Open and Interoperable Clouds: The Cloud@Home Way », chap. 6. In « Cloud Computing: Principles, Systems and Applications », N Antonopoulos & L. Gillam (eds), Springer Computer Communications and Networks series, 2010.

« Business models – for selling Cloud computing, it is mandatory to provide QoS and SLA management for both commercial and open-volunteer Clouds (traditionally best effort) to discriminate among the applications to be run. »

QoS, SLA (2)

From Vladimir Stantchev and Christian Schröpfer,
« Negotiating and Enforcing QoS and SLAs in Grid and Cloud Computing ». In Advances in Grid and Pervasive Computing, Lecture Notes in Computer Science, 2009, Volume 5529/2009.

« Emerging grid computing infrastructures such as cloud computing can only become viable alternatives for the enterprise if they can provide stable service levels for business processes and SLA-based costing. »

QoS, SLA (3)

S. Bose et al. « SLA MANAGEMENT IN CLOUD COMPUTING: A SERVICE PROVIDER'S PERSPECTIVE », chap. 16. In « Cloud Computing: principles and paradigms », R. Buya, J. Broberg, A. Goscinski (eds), Wiley.

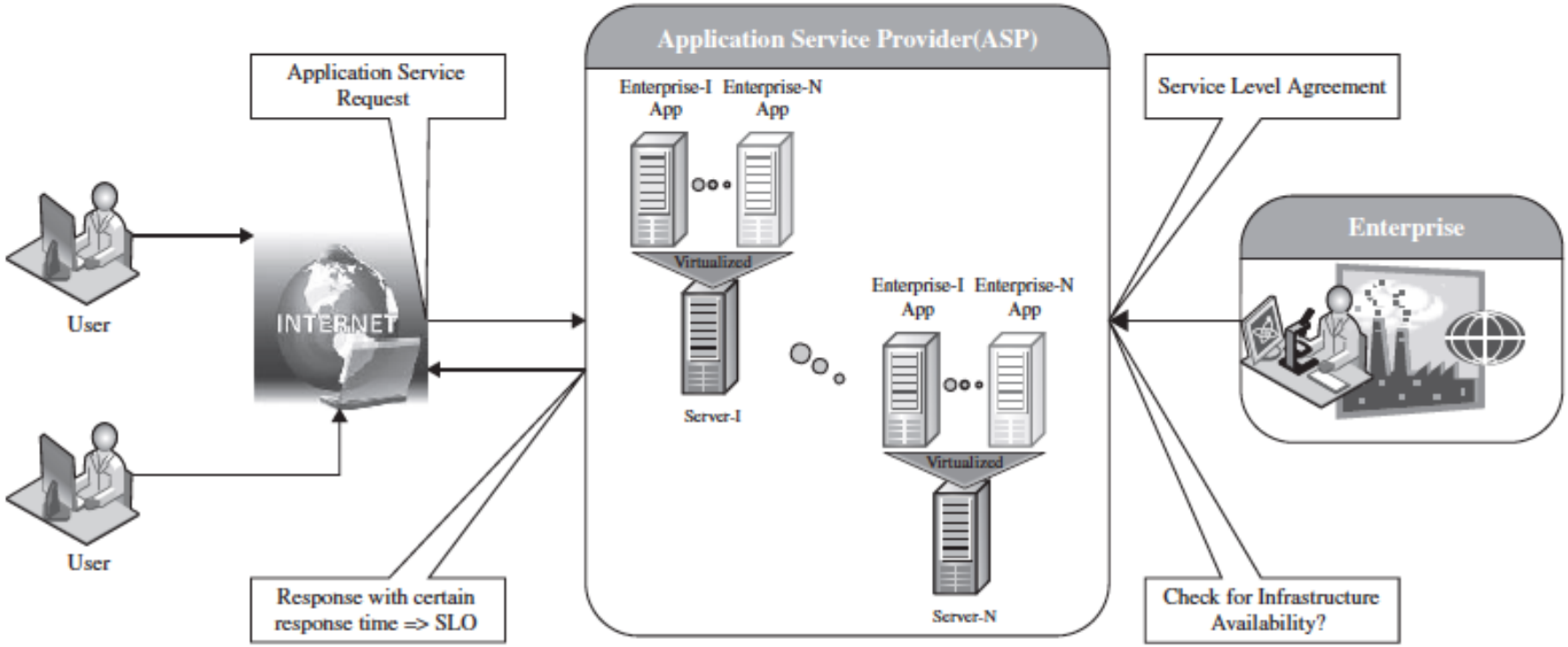
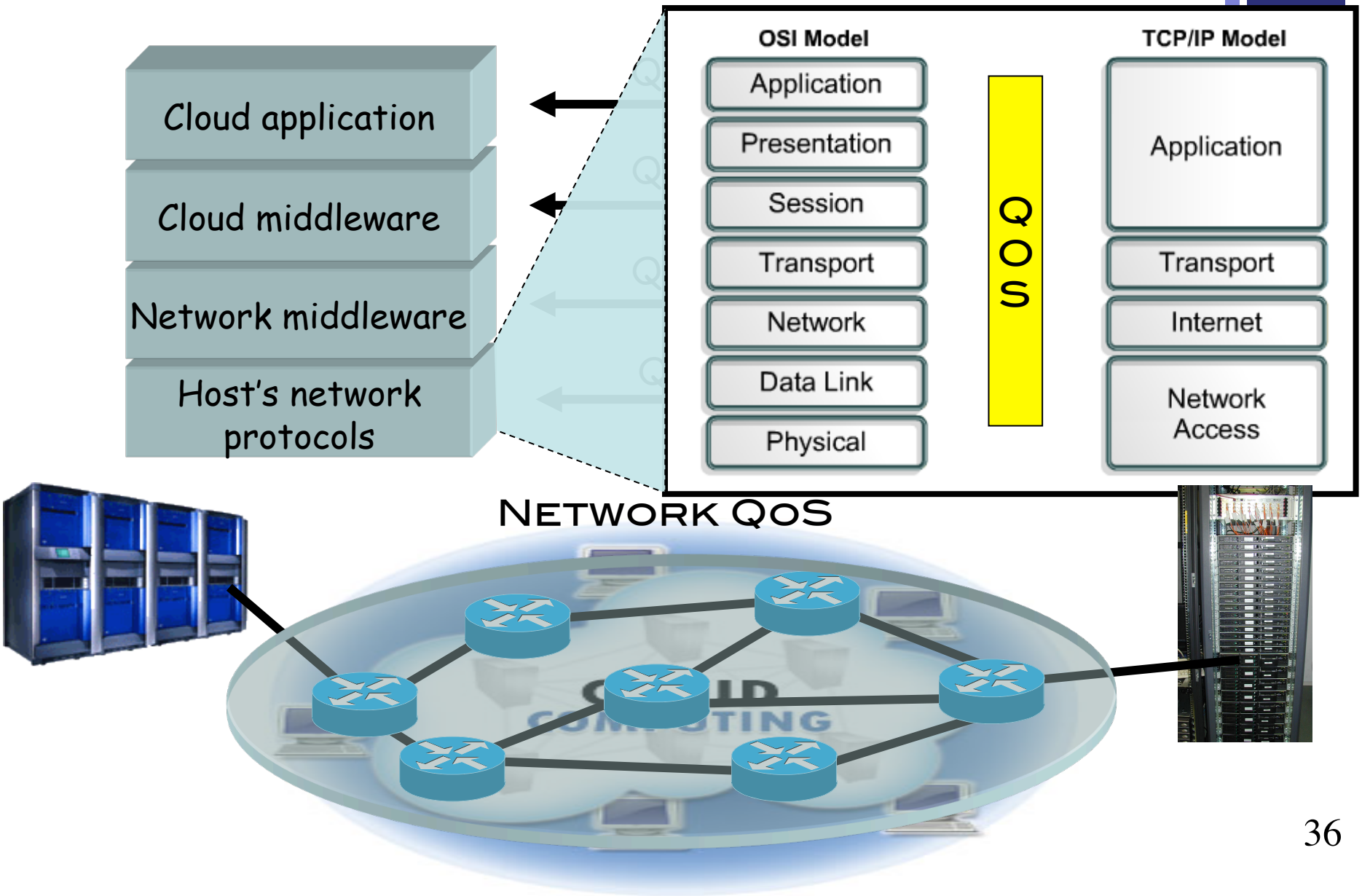
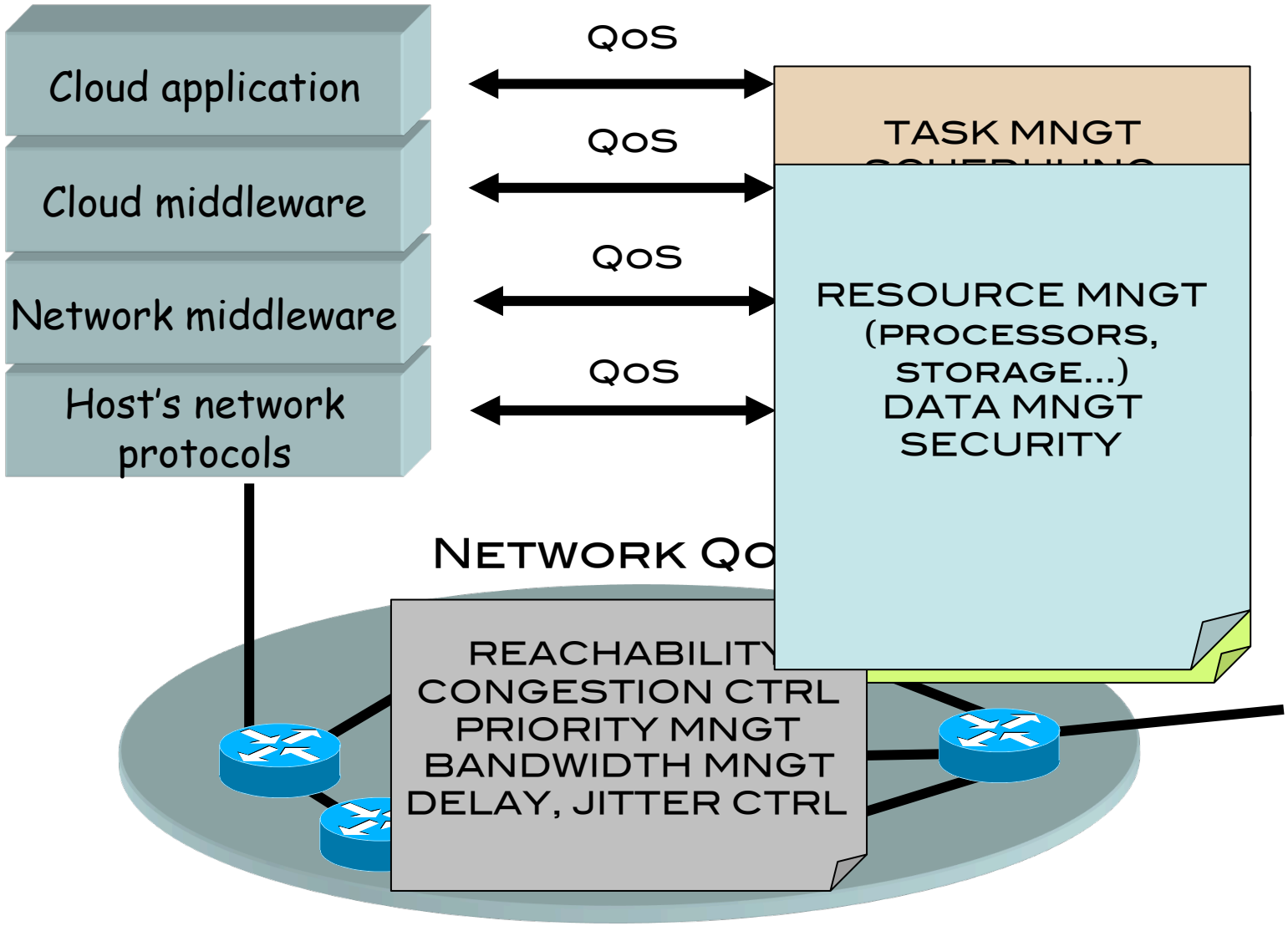


FIGURE 16.4. Shared hosting of applications on virtualized servers within ASP's data centers.

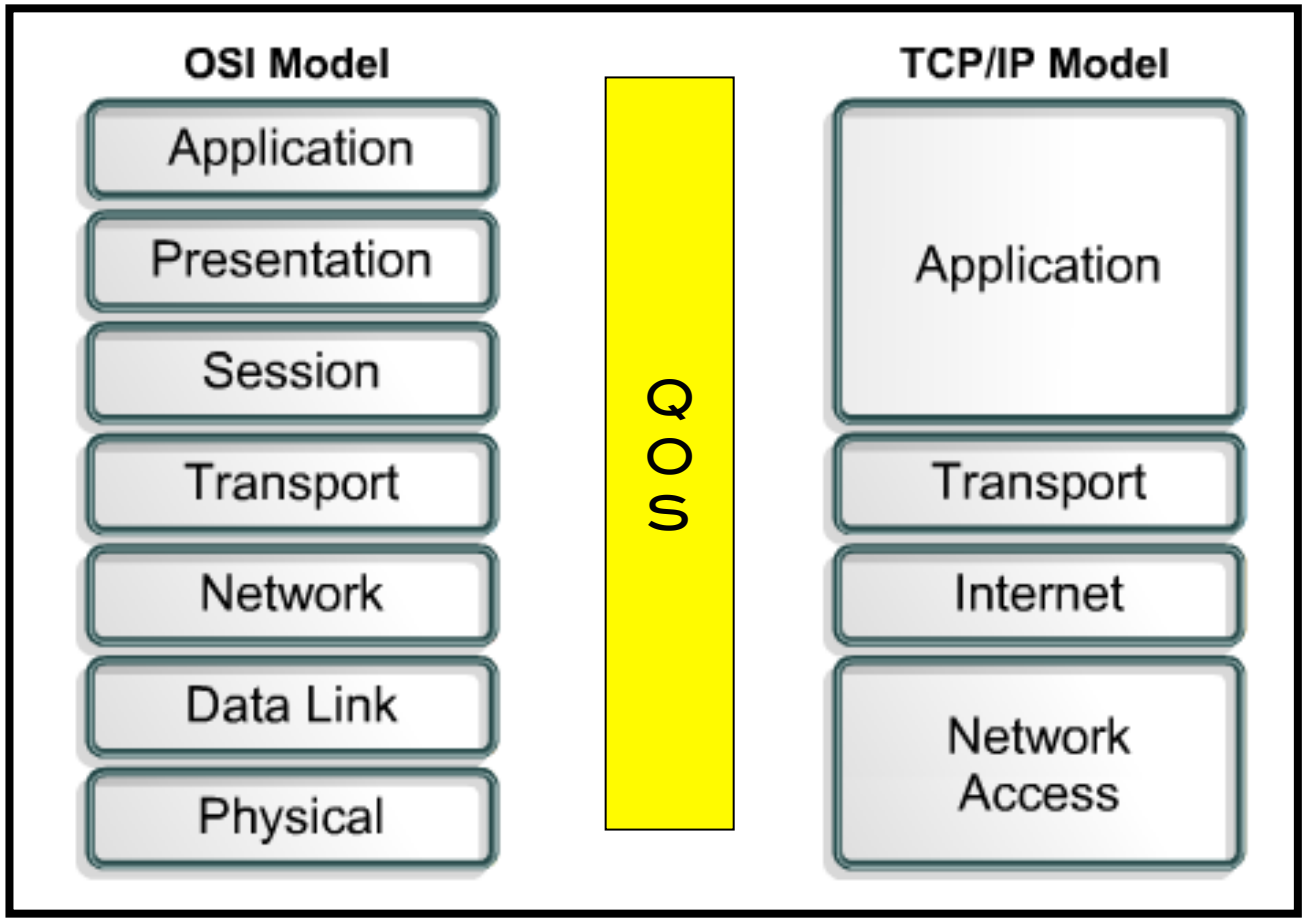
QoS in a Cloud



QoS: different meanings at different layers



30 YEARS OF INTERNET QoS...



- ...HAVE SHOWN THE POWER OF SELFISHNESS!

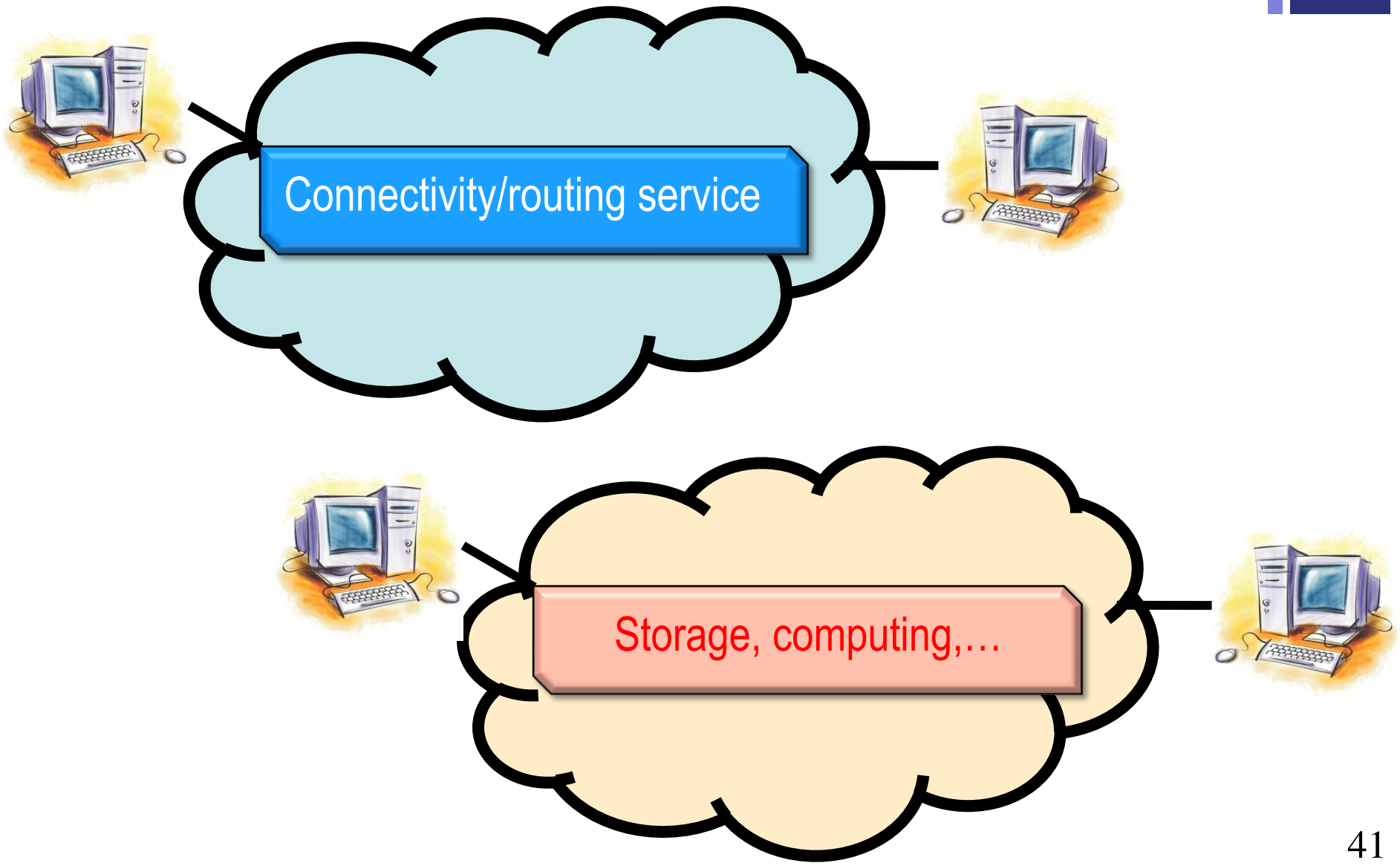
WHY SHOULD I BOTHER WITH QoS WHEN...

...OTHERS DON'T DO IT?

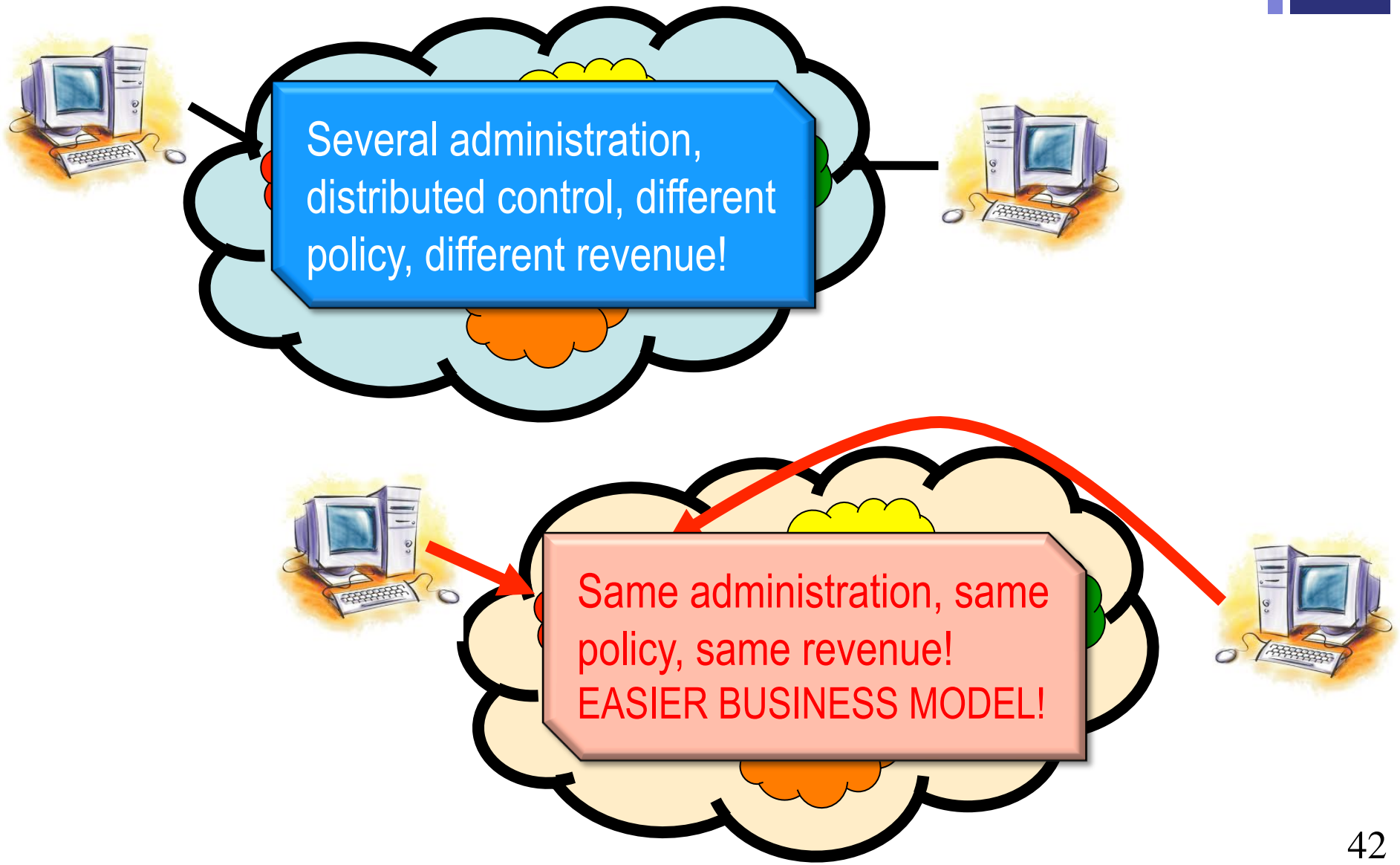
Current Internet's QoS

So why change?

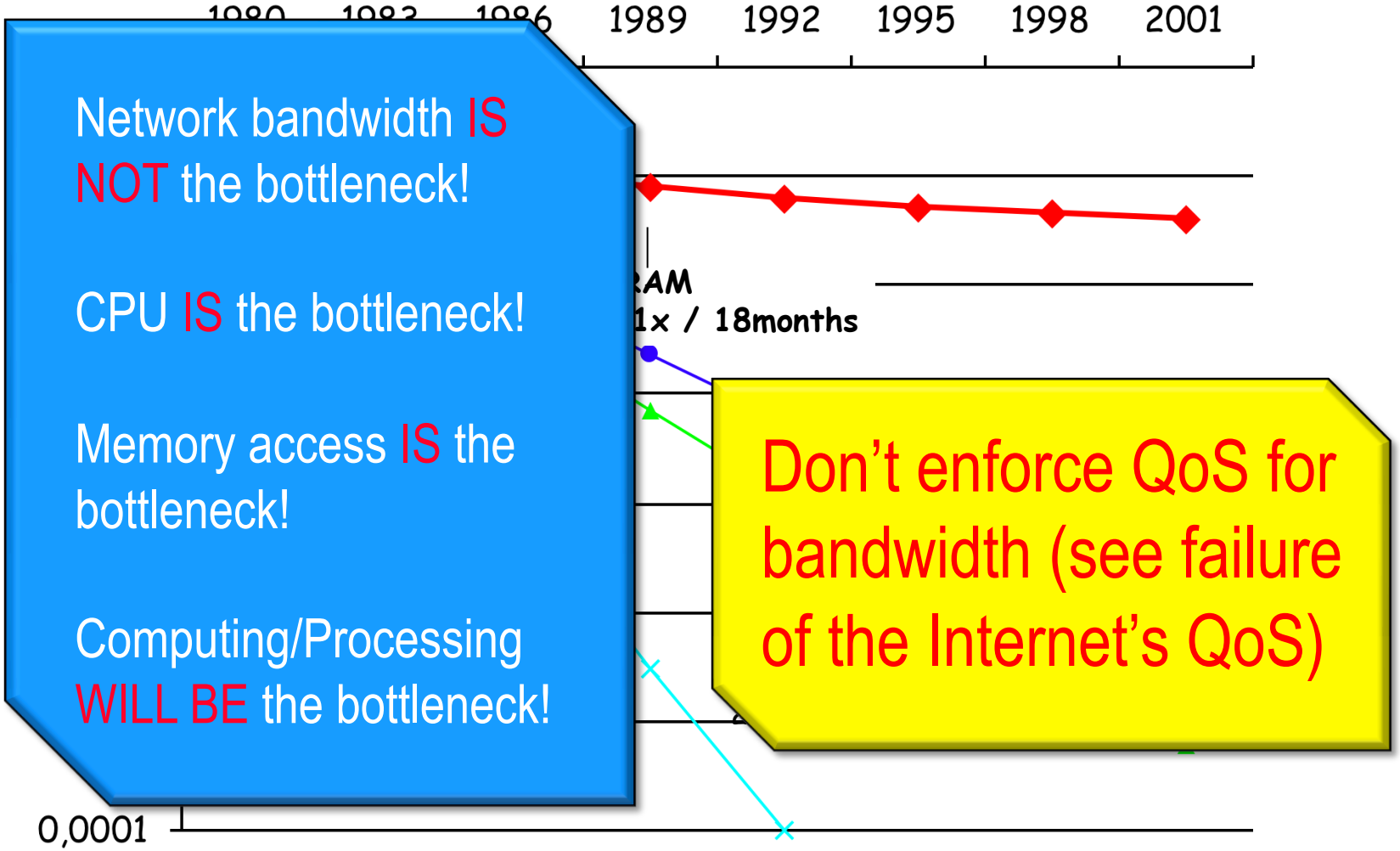
Internet vs cloud (computing)



When distributed is bad



Where to provide QoS in a Cloud?



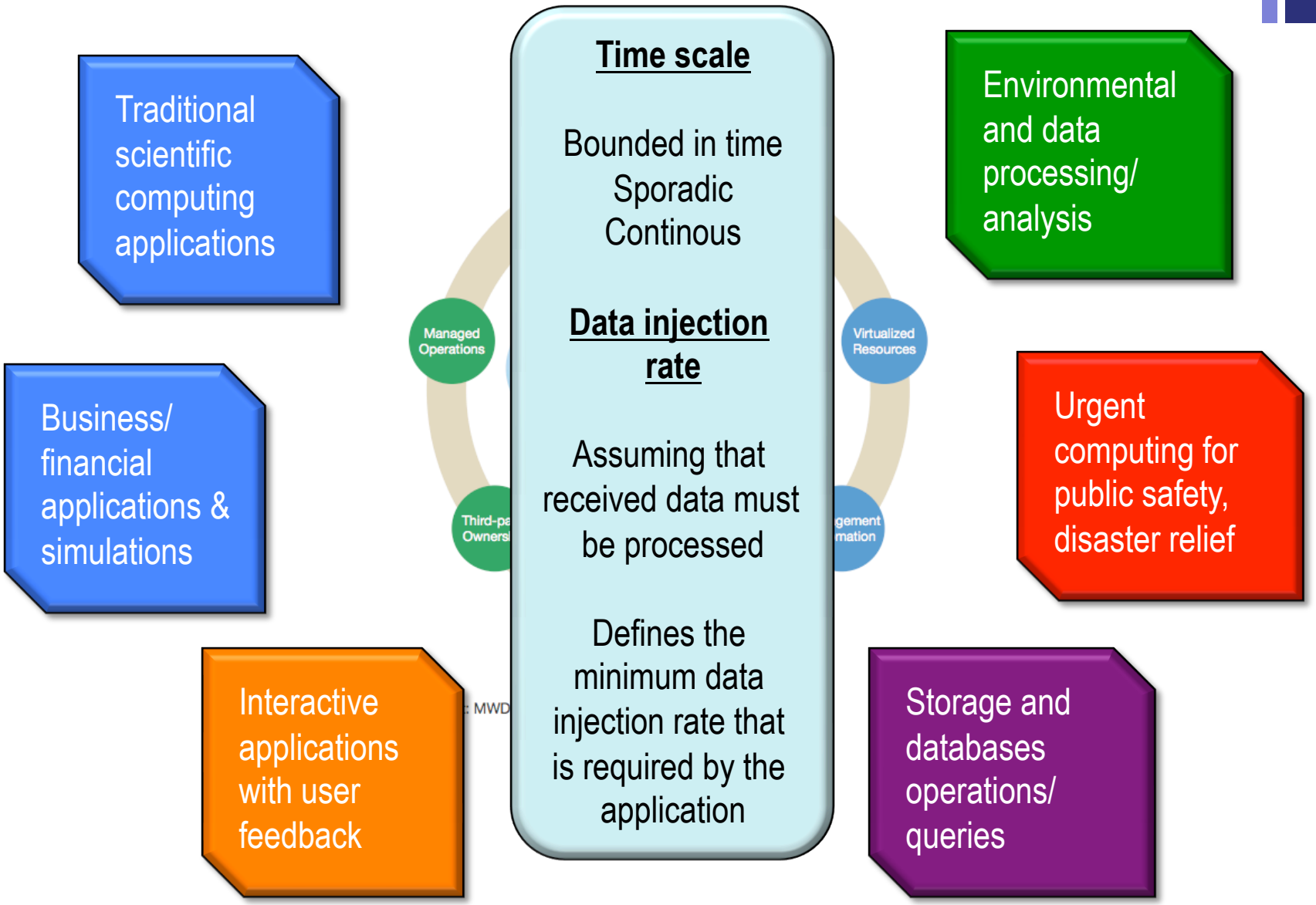
How to provide QoS in a Computing Cloud?

- No networking QoS!
- Scheduling, admission control, trafic control, dynamic resource provisioning, …

How to take into account the various application's profiles?
How to protect users from misbehaving applications?
How to handle urgent demands?

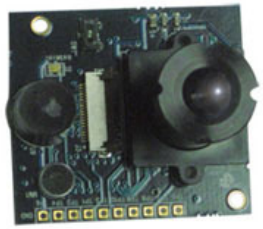
Regulate (adapt & control) the data injection rate into the computing resources

Cloud application's profile

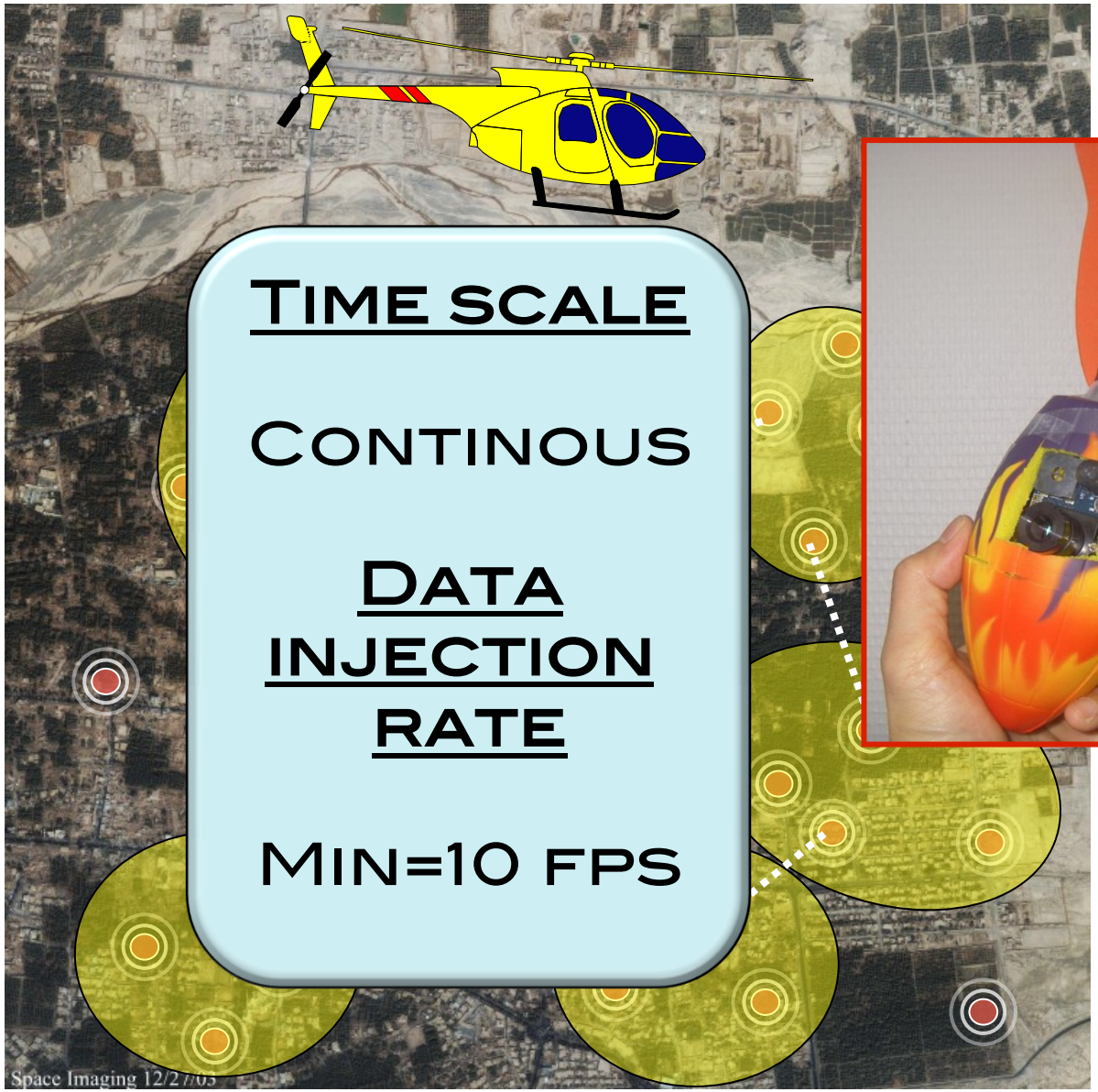


EX: URGENT DISASTER RELIEF

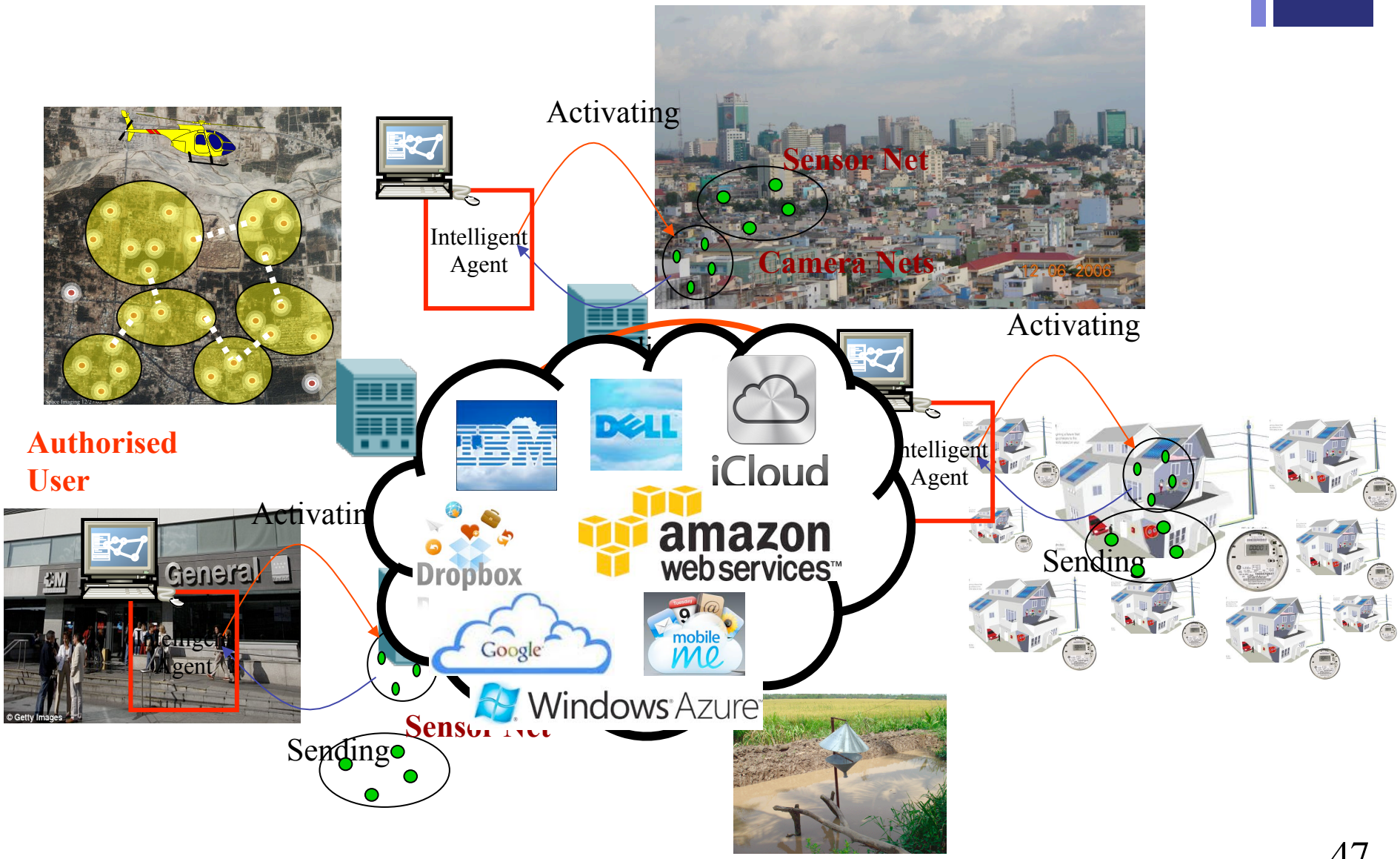
Imote2



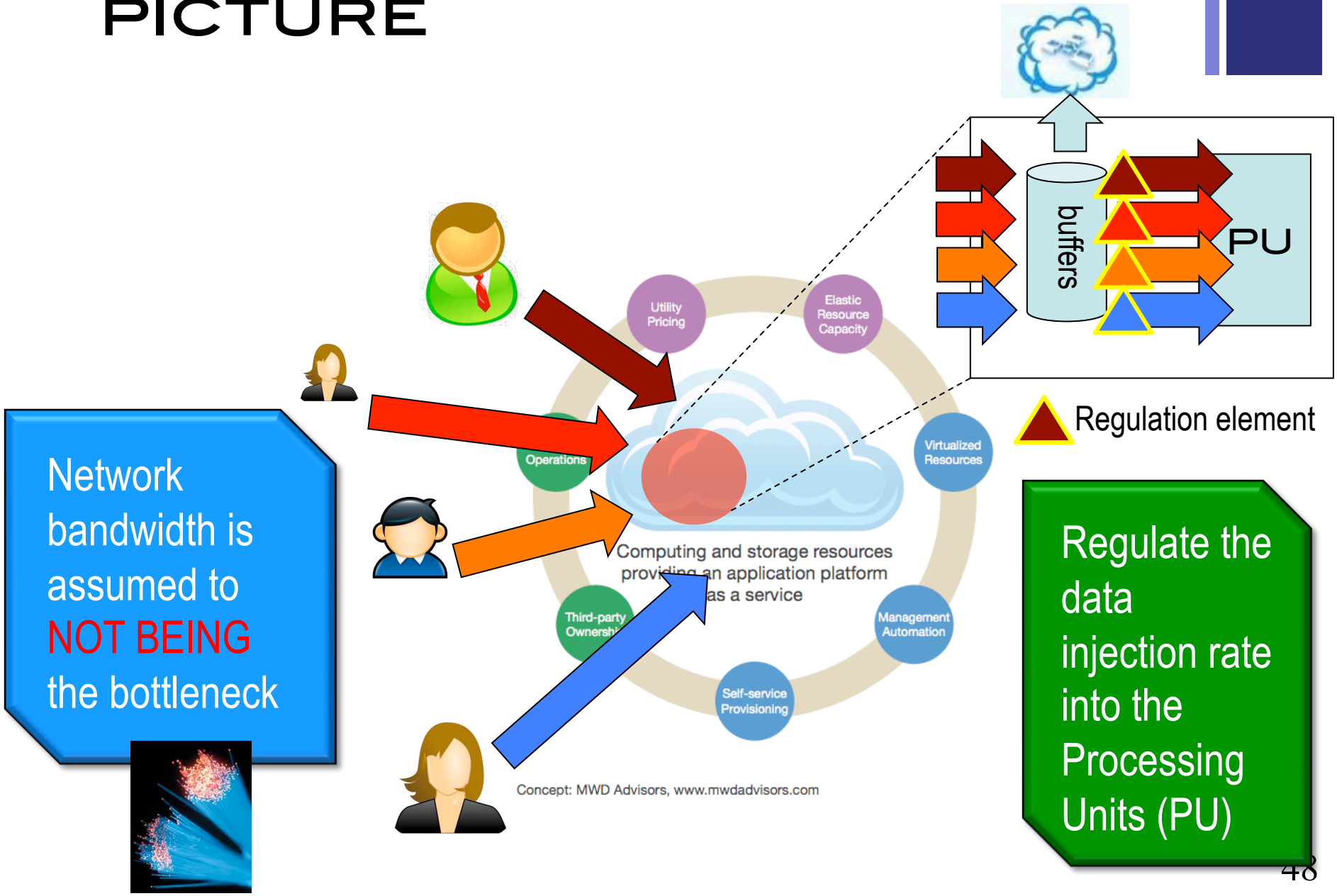
Multimedia board



Everybody's data is important!

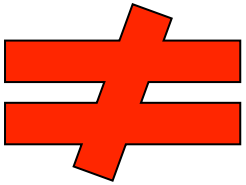


QoS on Clouds: general picture



Regulation process

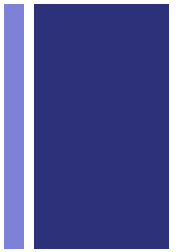
No networking QoS



Don't use networking QoS ideas

Rate regulation mechanisms are at the core of networking QoS architectures: IntServ, DiffServ, …
QoS requires traffic characterization!

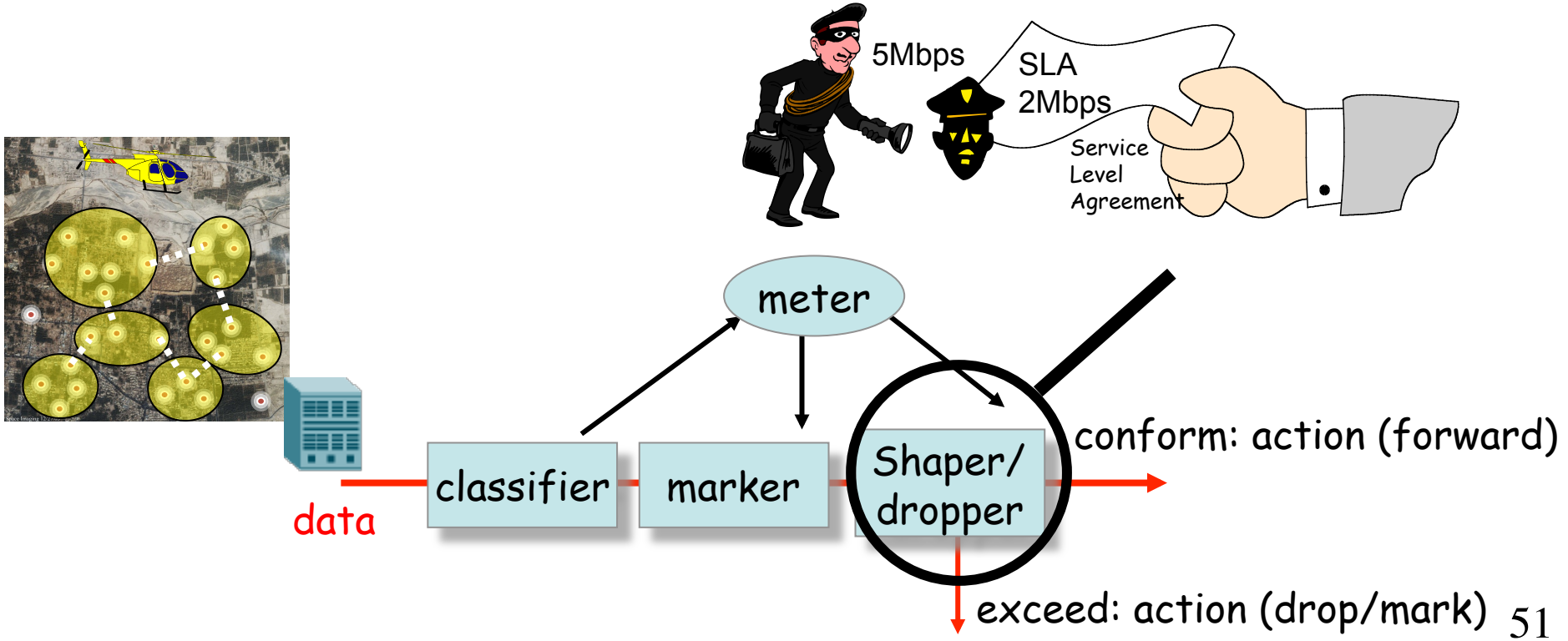
Traffic and Service Characterization



- Definitions
  - Cloud Infrastructure (CI)
  - Processing Units (PU)
- To quantify a service one has two know
  - Flow's traffic arrival
  - Service provided by the CI, i.e., resources reserved at PU
- Ideas are to
  - Bound the data injection rate to…
  - …isolate users from each others and…
  - …to provide QoS enforcement at flow level.

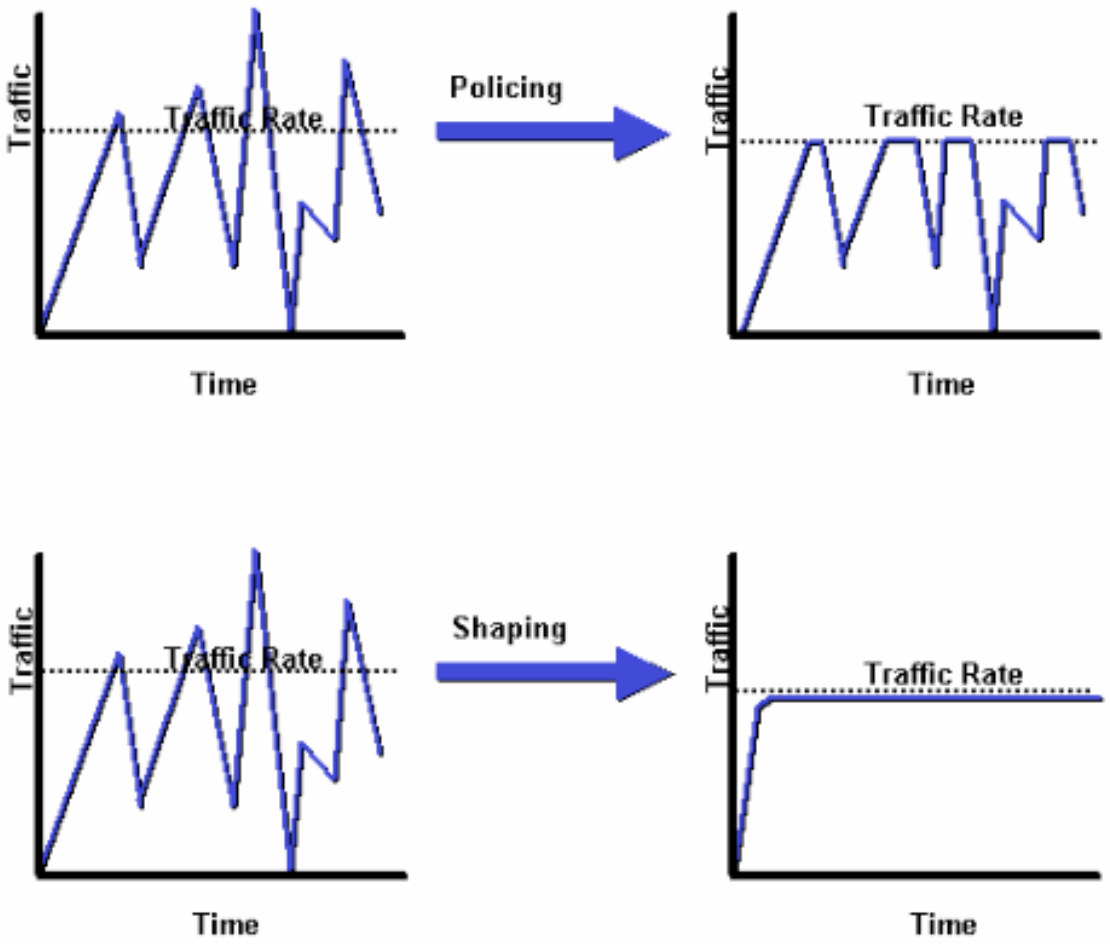
Ideas behind regulation

- User declares traffic profile (eg, data injection rate); traffic is metered and shaped if non-conforming



How to regule?

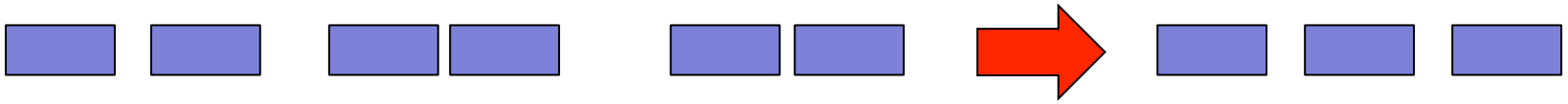
- Policing vs Shaping



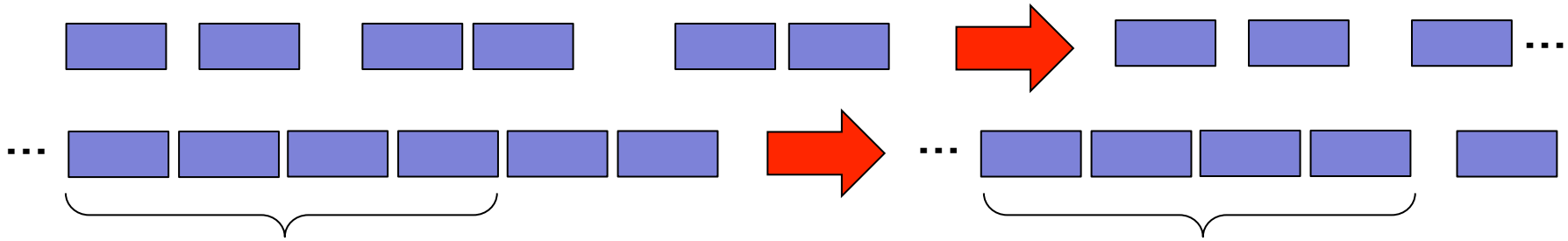
Source: Cisco

How to shape?

- Constant vs dynamic



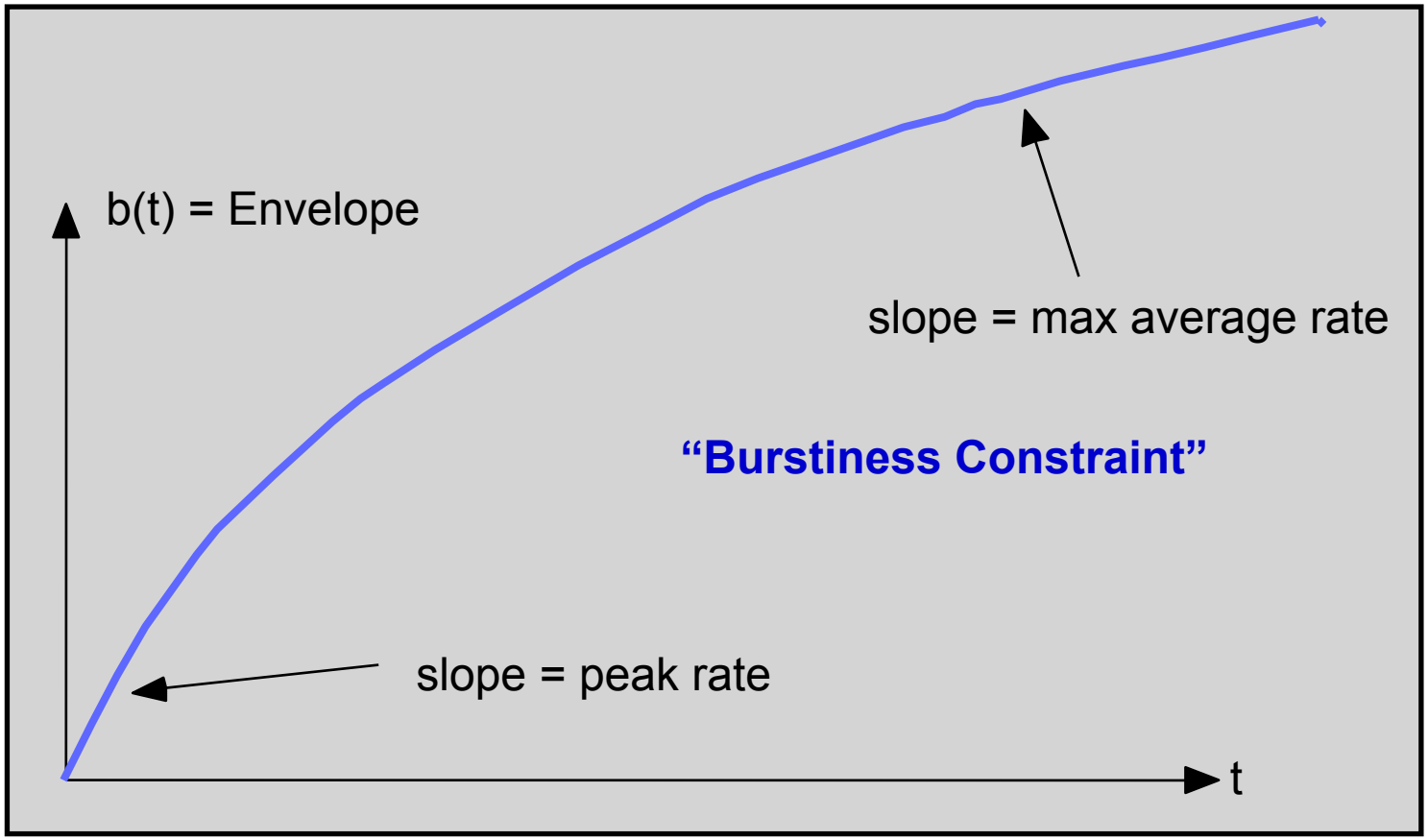
- Dynamic



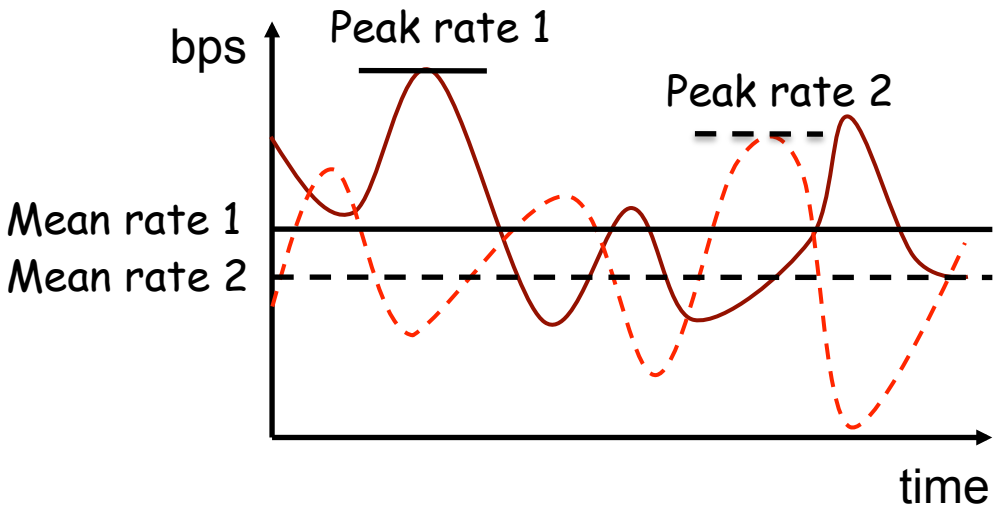
Dynamic shaping allows for variable data injection rate (variable sensing rate for instance). Need however to bound the amount of provided service per time period T.

Traffic Envelope (Arrival Curve)

- Maximum amount of service that a flow can request during an interval of time T

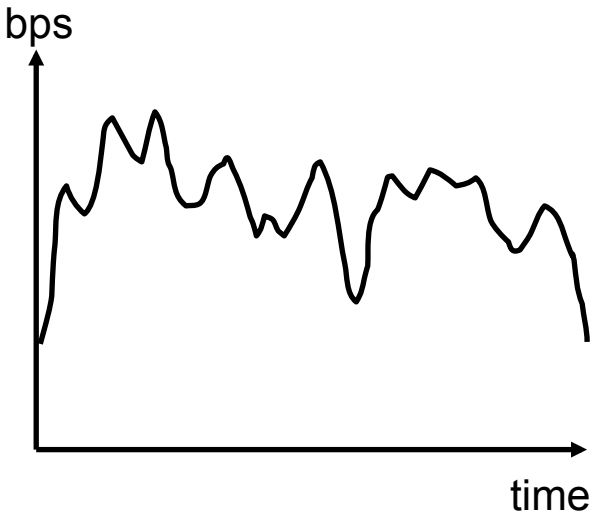


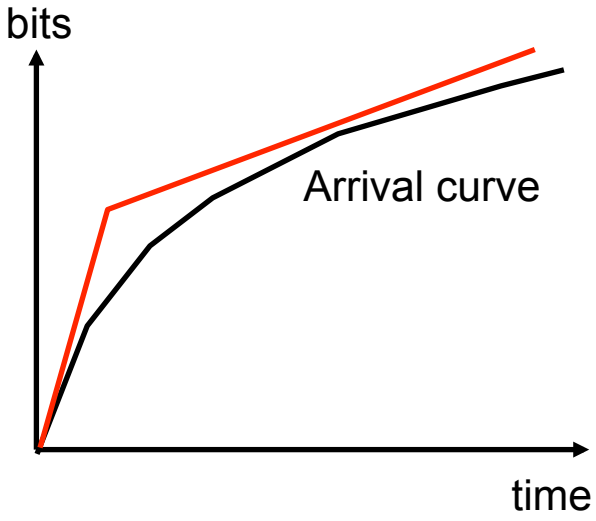
Traffic Envelope (traffic shaping)



Traffic are variable by nature, must take into account burstiness constraints

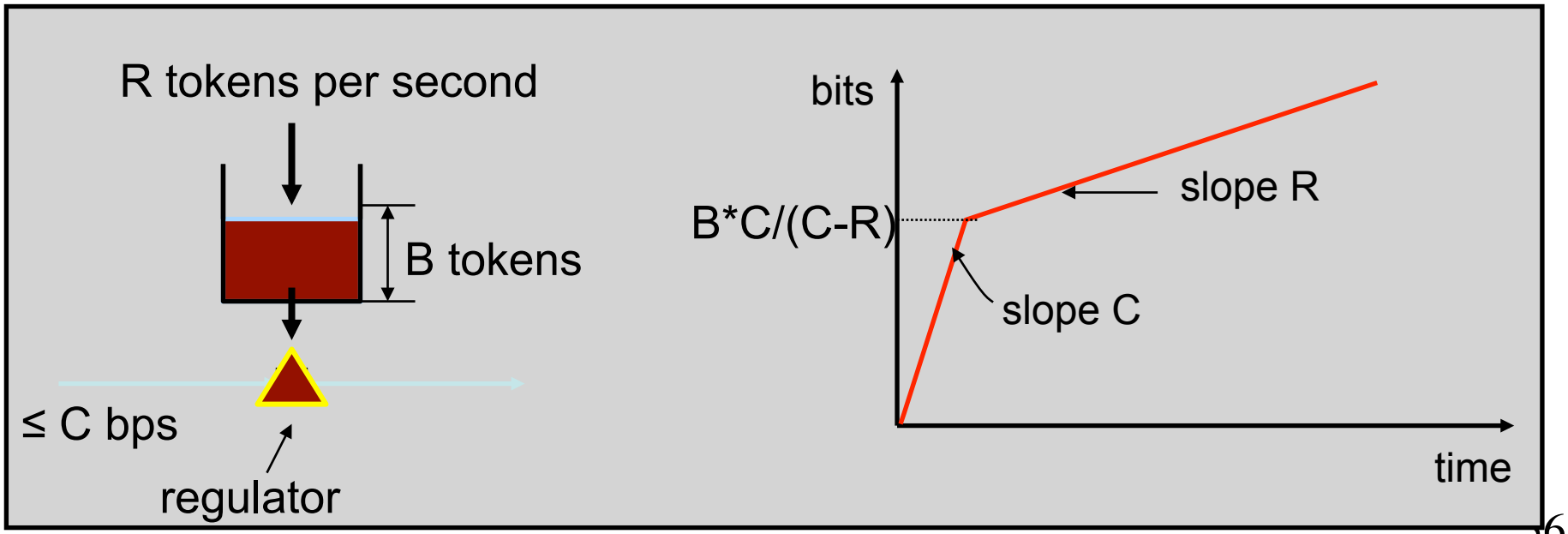
Use an envelope process to bound the data injection rate while allowing for variable, bursty traffic





Token Bucket model

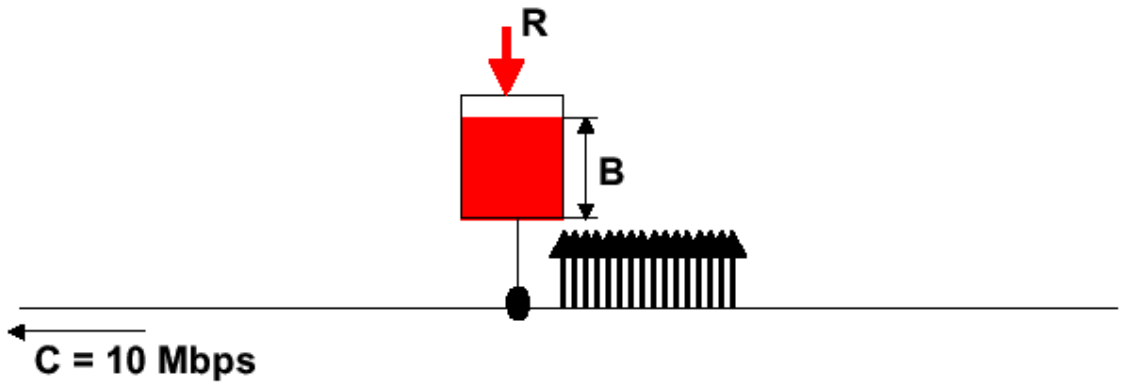
- Characterized by three parameters (B, R, C)
  - B – token depth
  - R – average arrival rate
  - C – maximum data injection rate
- A bit is transmitted only when there is an available token
  - When a bit is transmitted exactly one token is consumed



Token Bucket example (1)

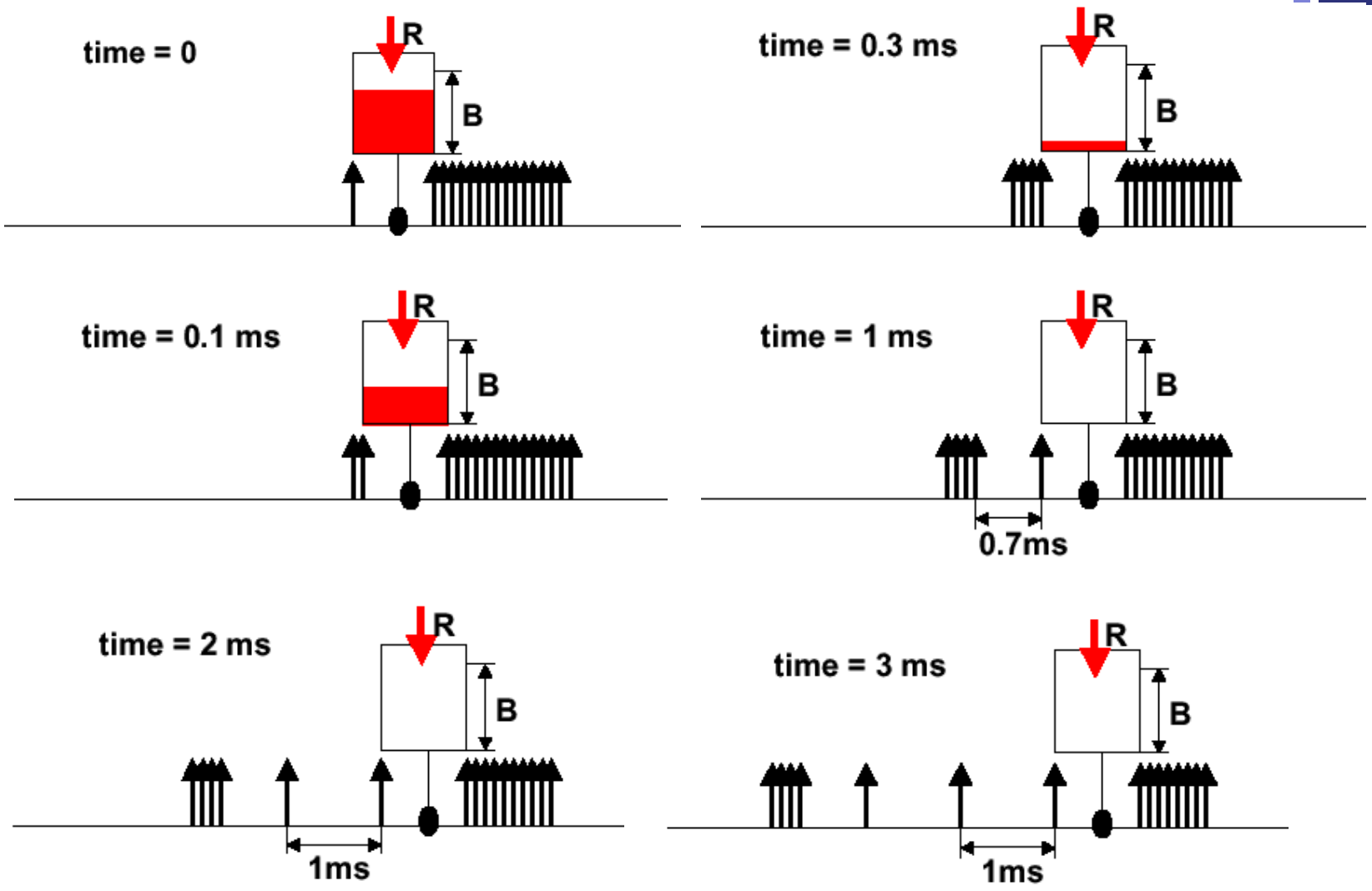
Example

- B = 4000 bits, R = 1 Mbps, C = 10 Mbps
- Packet length = 1000 bits
- Assume the bucket is initially full and a "large" burst of packets arrives



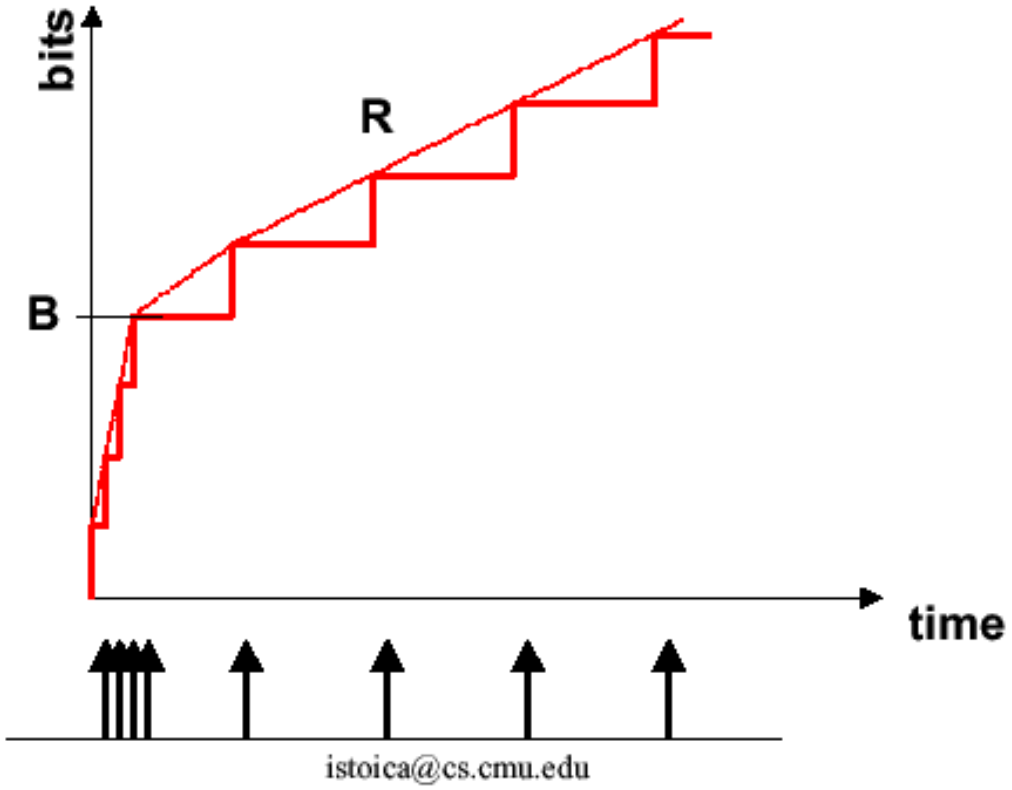
istoica@cs.cmu.edu

Token Bucket example (2)



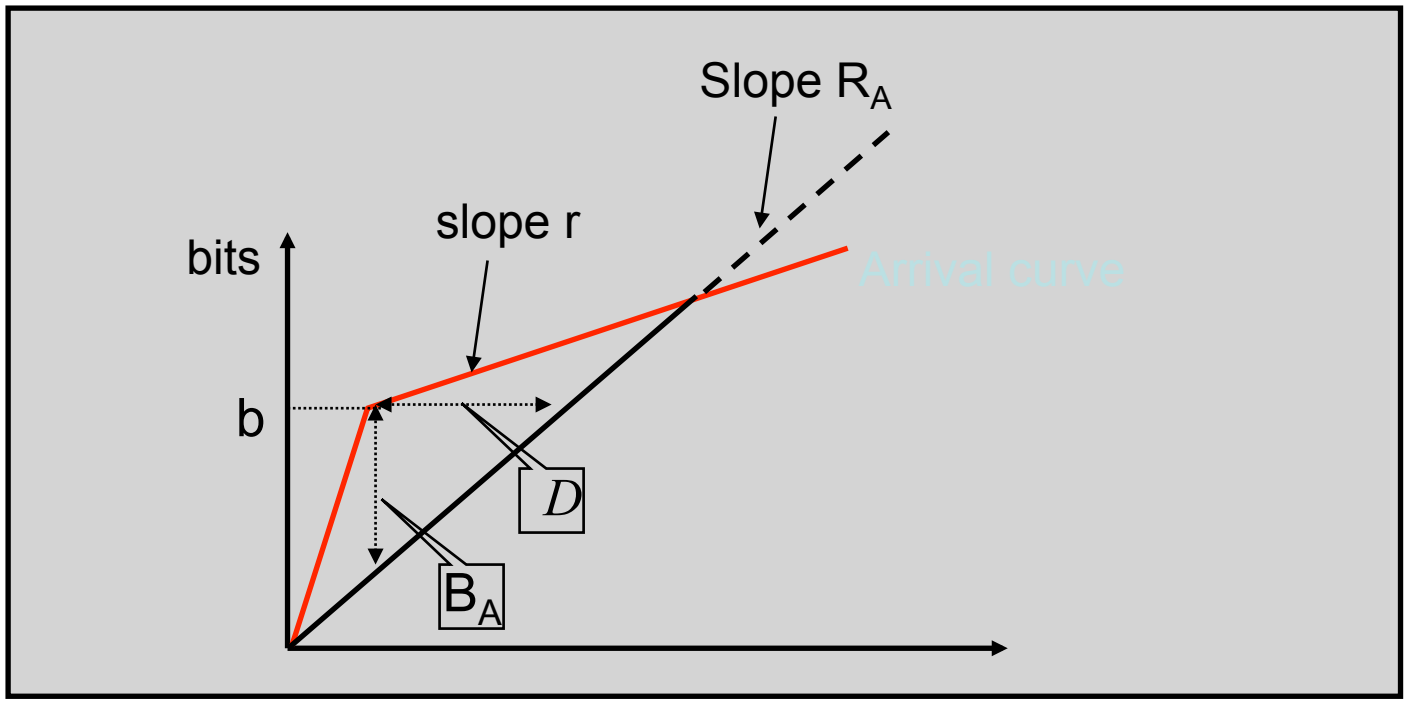
Arrival curve

A(t) – number of bits received up to time t

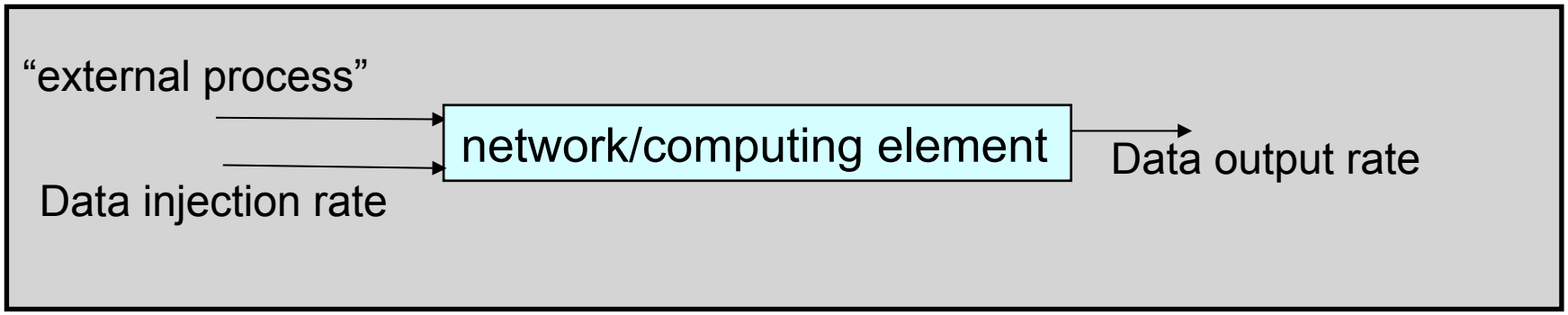


Per-hop Reservation with Token Bucket

- Given B,R,C and per-hop delay D
- Allocate resource R_A and buffer space B_A such that to guarantee D

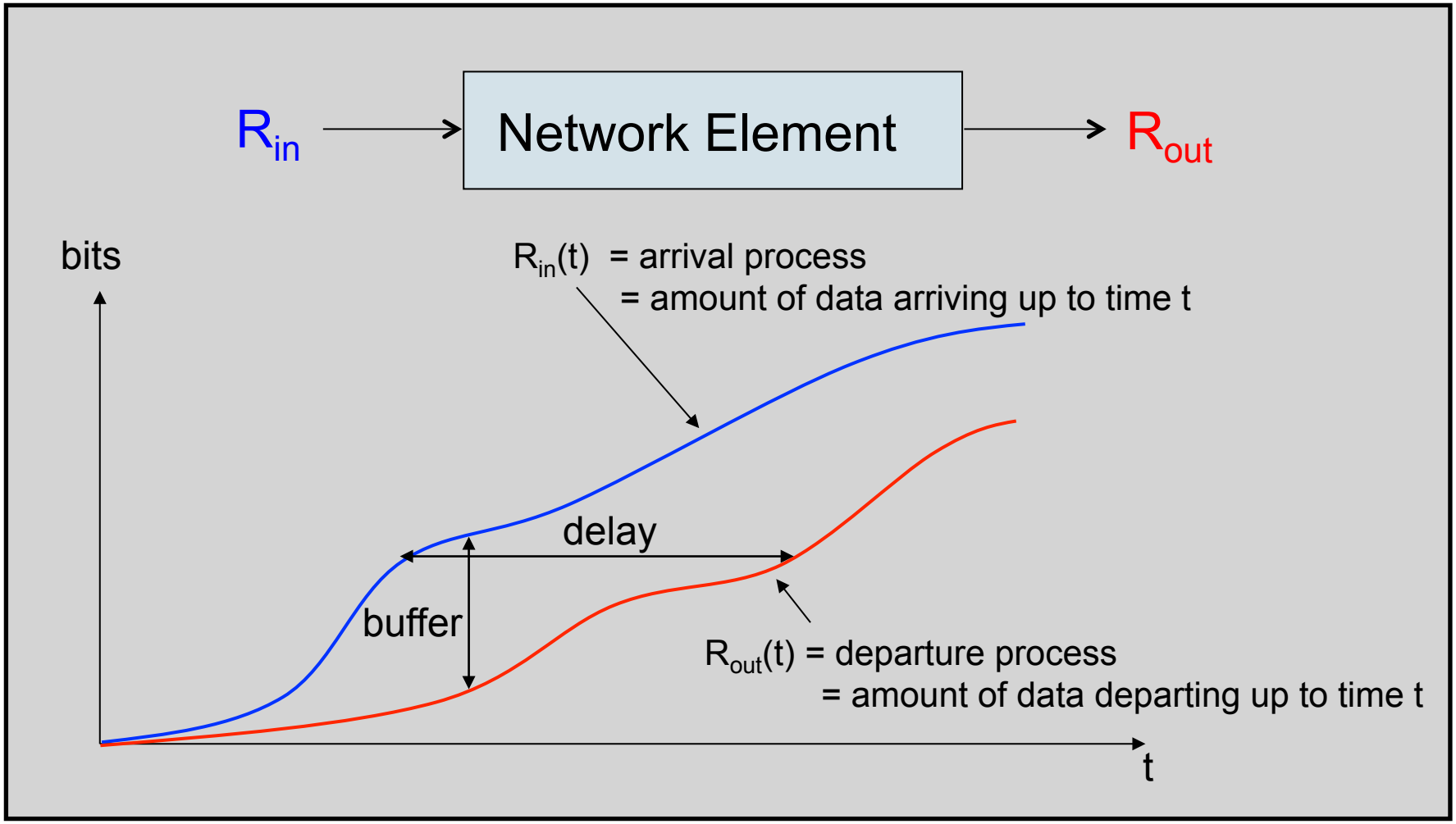


Service Model

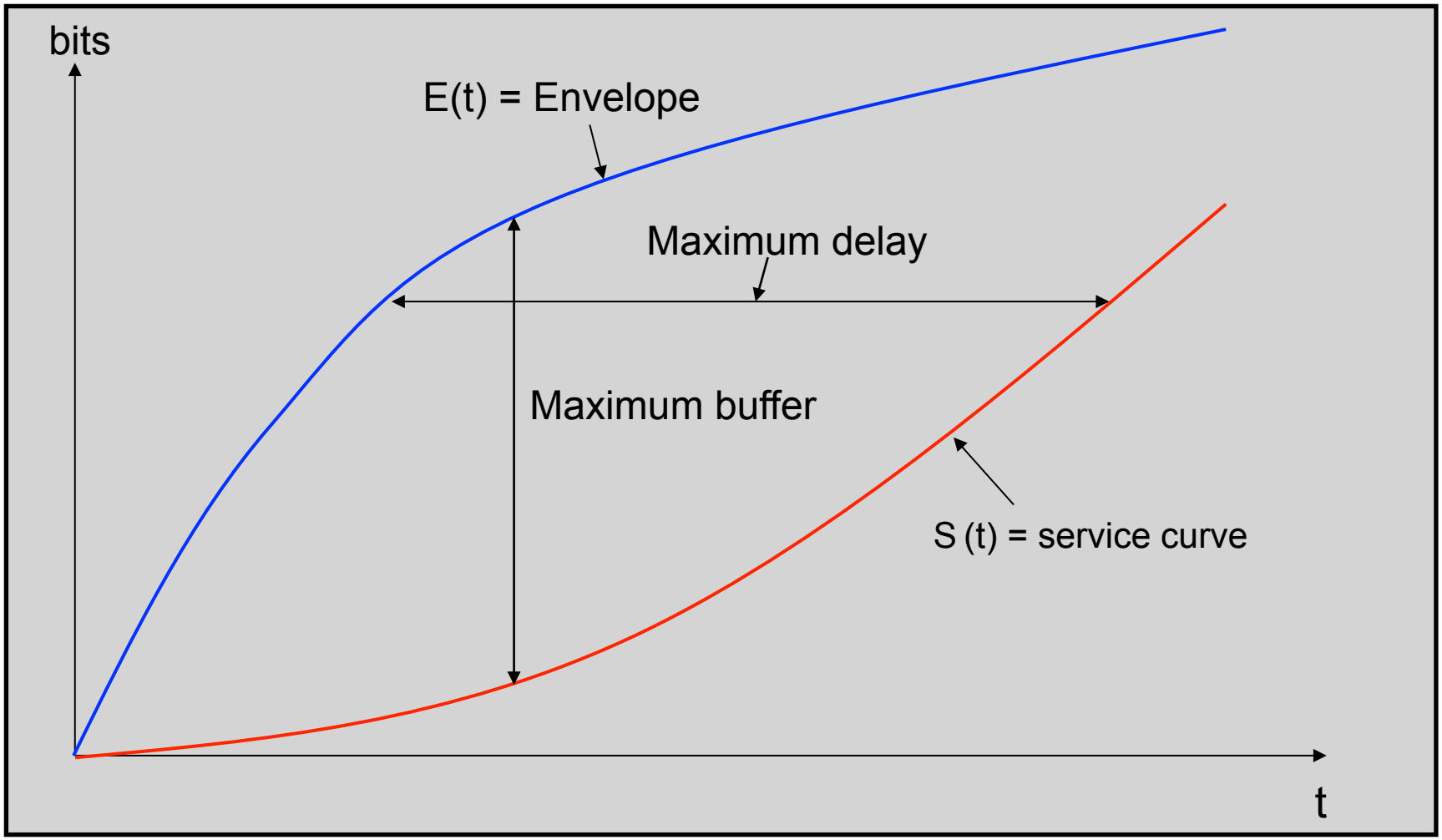


- The QoS measures (delay,throughput, loss, cost, ...) depend on offered traffic, and possibly other external processes.
- A service model attempts to characterize the relationship between inputs, outputs, and possibly other external processes.

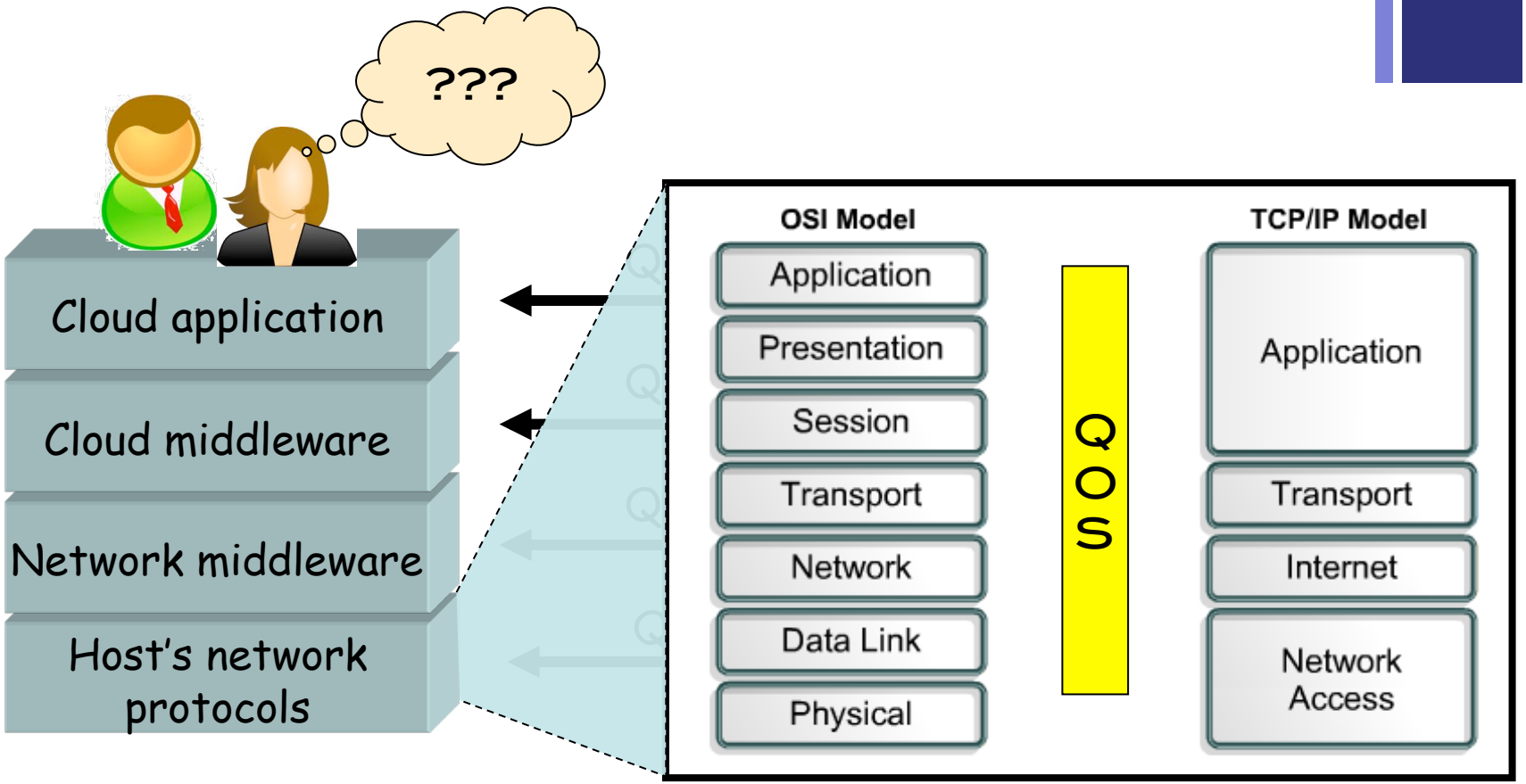
Arrival and Departure Process



Delay and Buffer Bounds



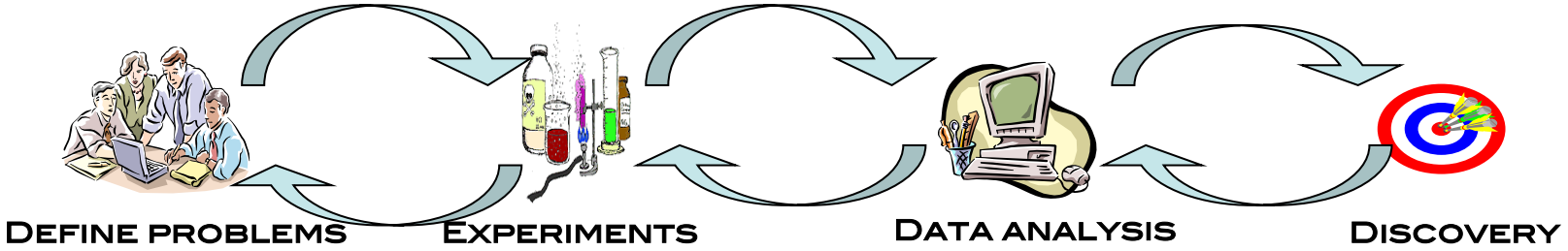
Where to put QoS?



Need to better understand the purpose of data processing from various user communities: scientific, business, industrial, ...

Scientific experiment

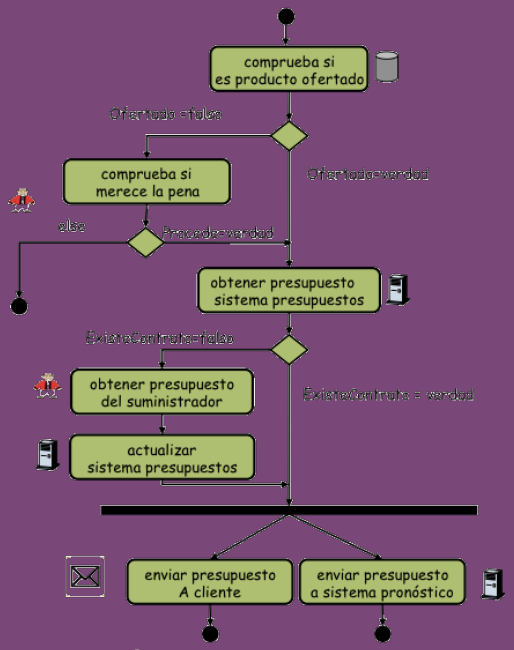
- Consists of collections of data through observation and experimentation, and the formulation and testing of hypotheses.



Activities are iterative, dynamic, and human steered !!

Slide taken from J. Banares, Univ. Zaragoza

SLIDE TAKEN FROM J. BANARES, UNIV. ZARAGOZA



Workflow

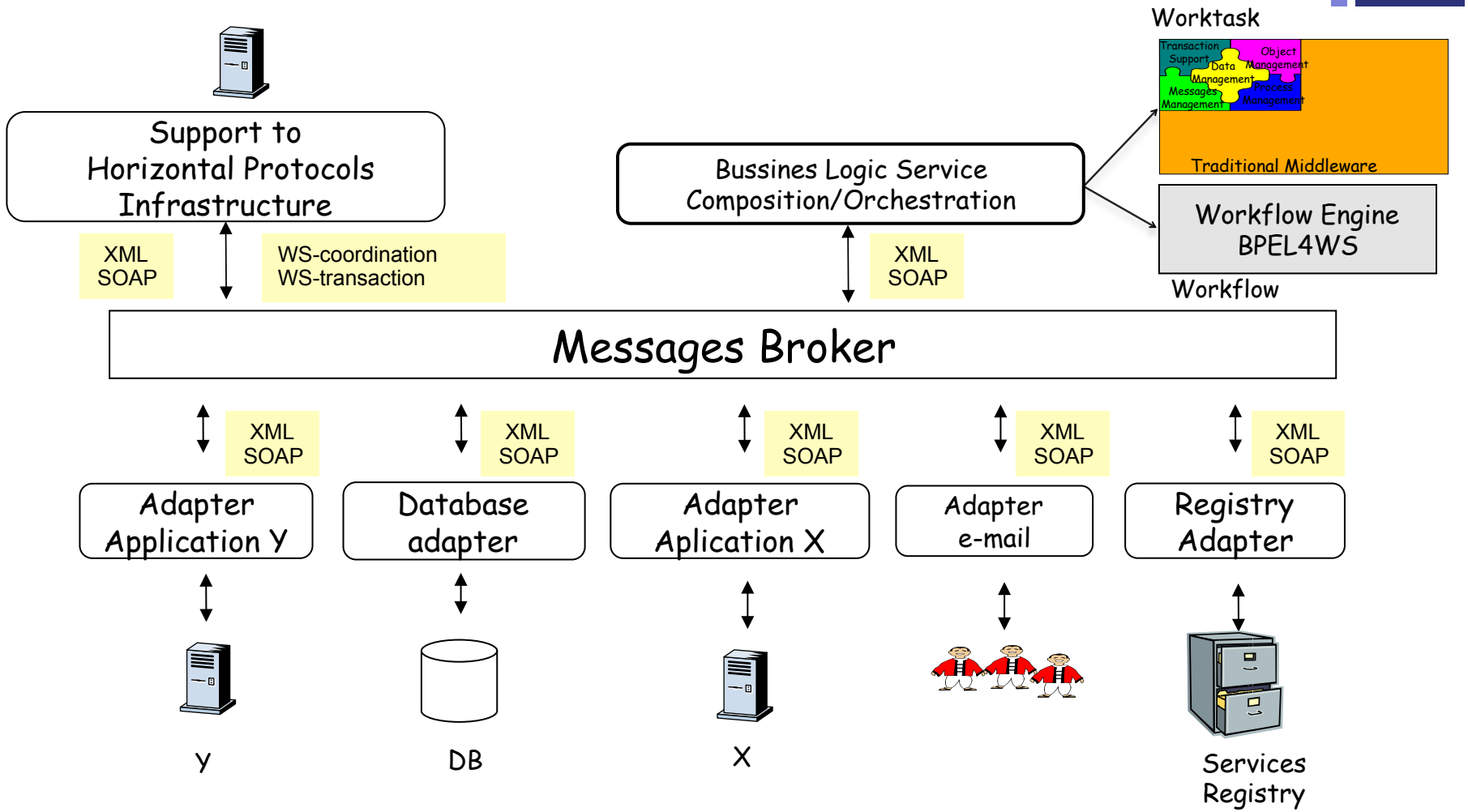
*"The **automation of a business process**, in whole or part, **where documents, information or tasks are passed from one participant to another** to be processed, according to a set of procedural rules "*

Workflow Management Coalition (WfMC)

Business processes

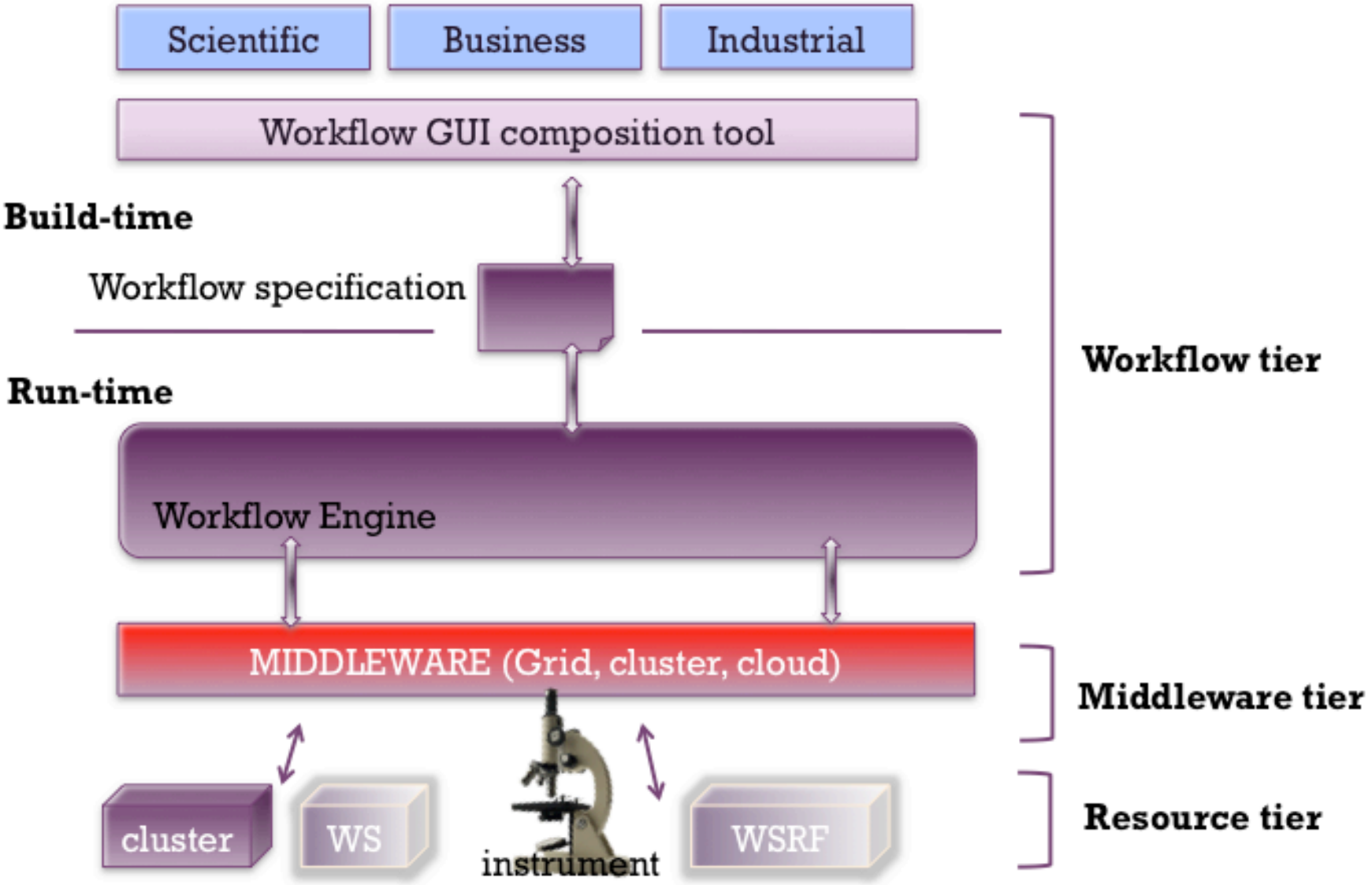
- The workflow concept has existed for decades
  - *Representation, Specification, and Automation of **Office** Procedures*
- They are very popular to manage **business processes** as **document interchanges**
- Workflow Management System (WFMS) *is a system that defines, creates. and manages the execution of workflows*
  - *Support for **large scale programming** (different time scale)*
  - ***Fault tolerance** support*
  - ***Interact** with participants*
  - ***Invoke** Enterprise applications*
- *Component of EAI tools (Enterprise Application Integration)*

Worflows & Service Oriented Architectures



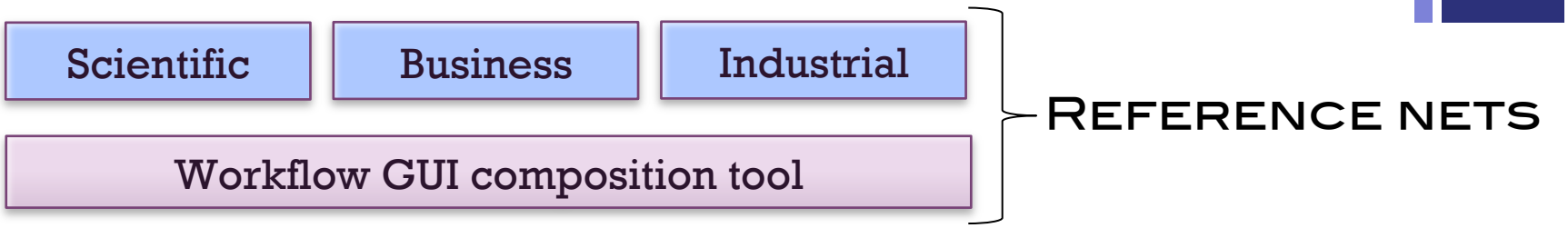
SLIDE TAKEN FROM J. BANARES, UNIV. ZARAGOZA

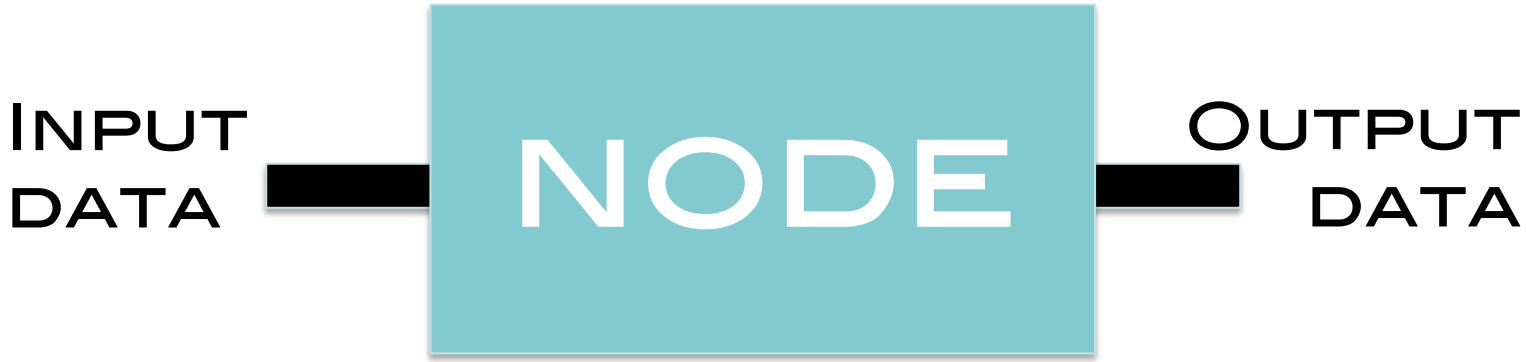
General purpose workflows

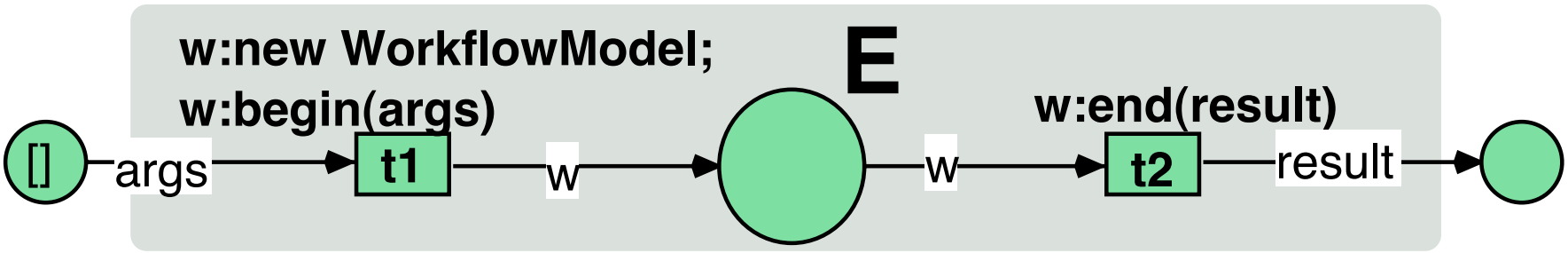


Slide taken from J. Banares, Univ. Zaragoza

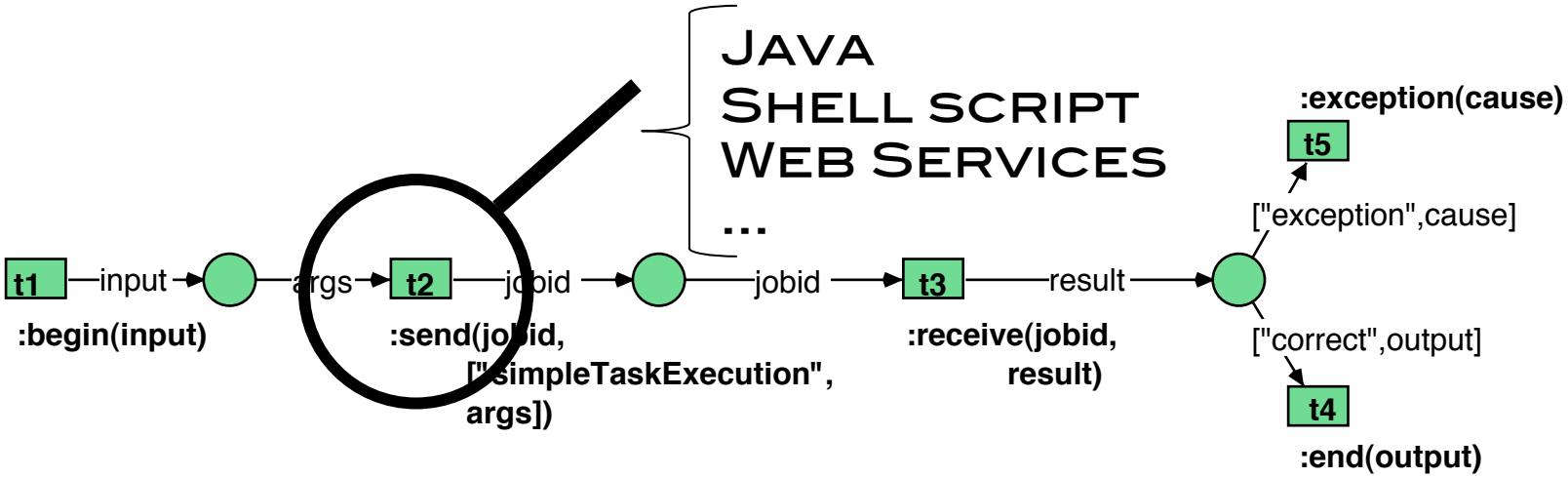
Reference Nets (RN)

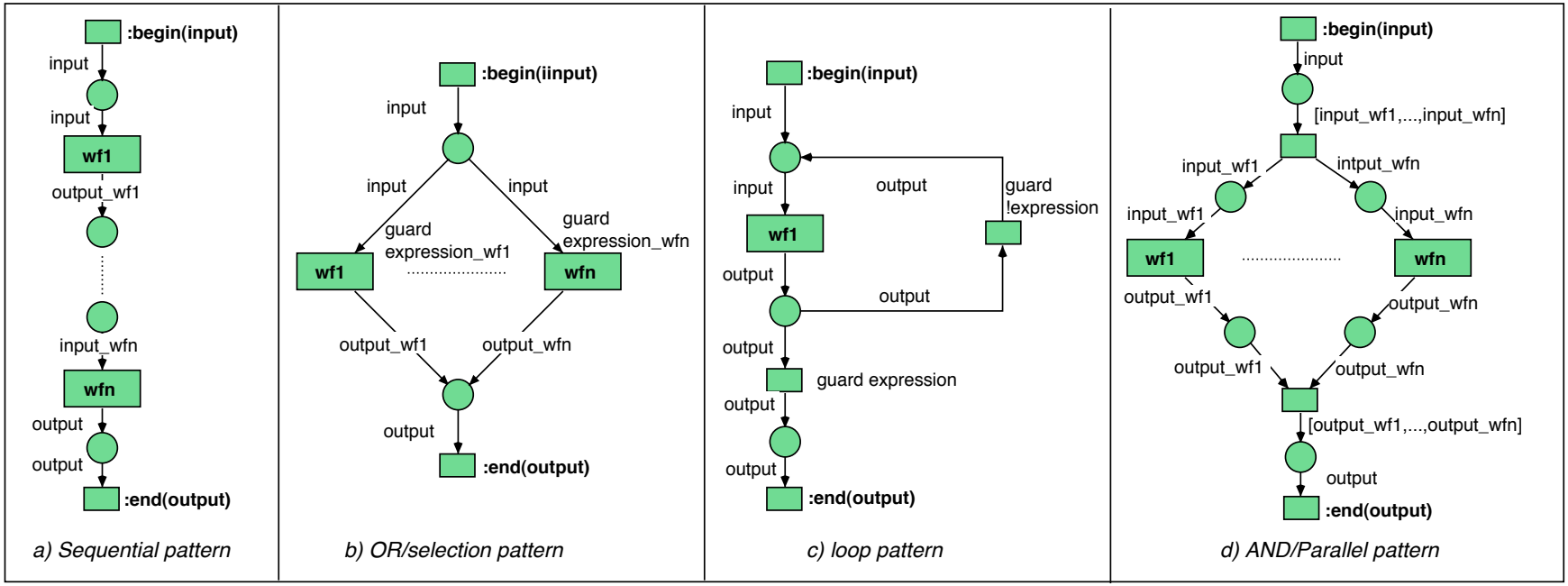




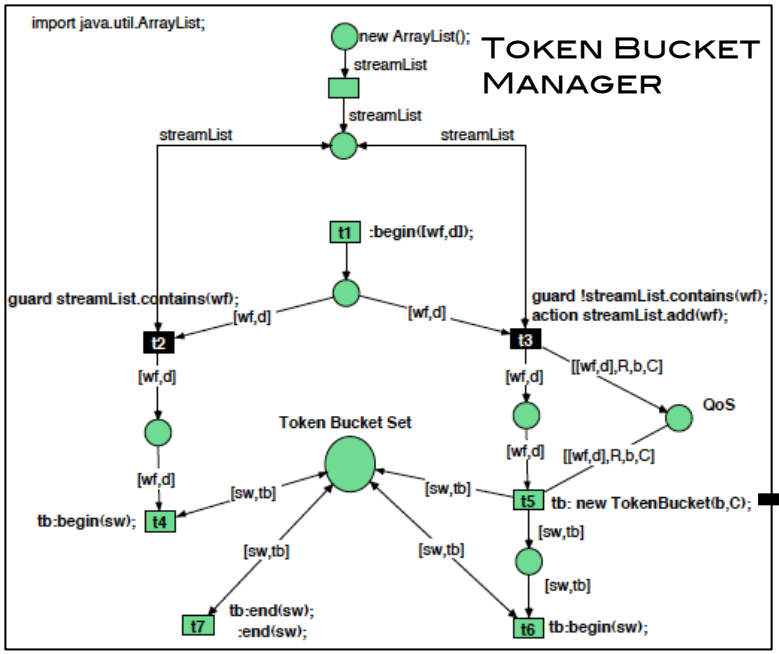


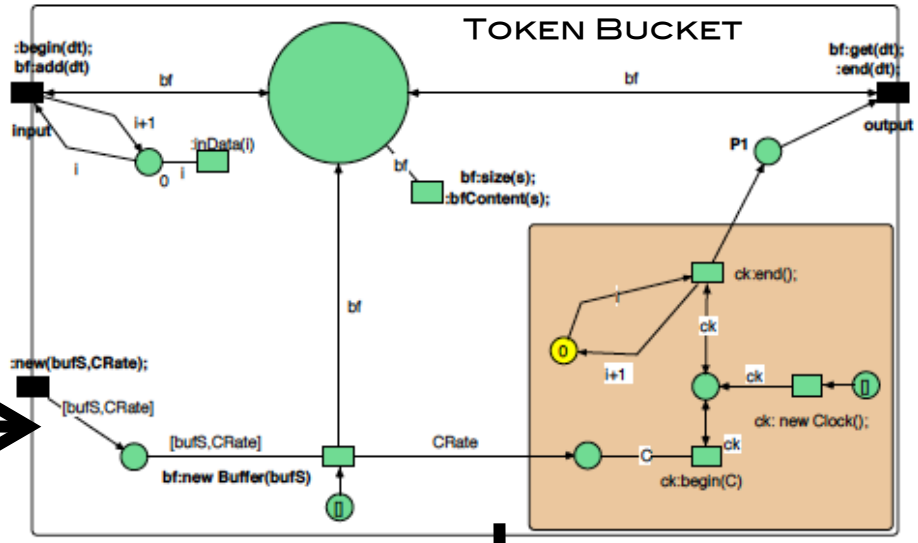
Reference Nets & Cloud



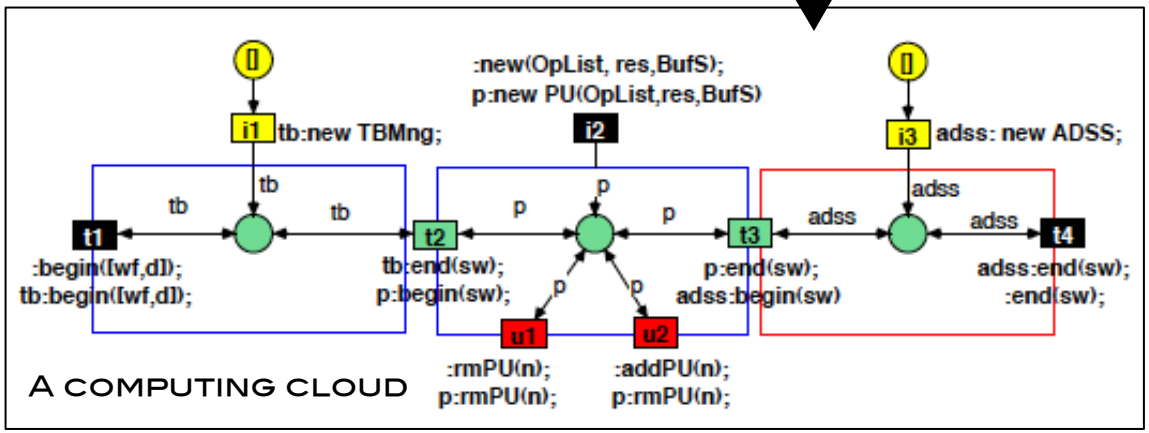


Token Bucket support in RNs workflows





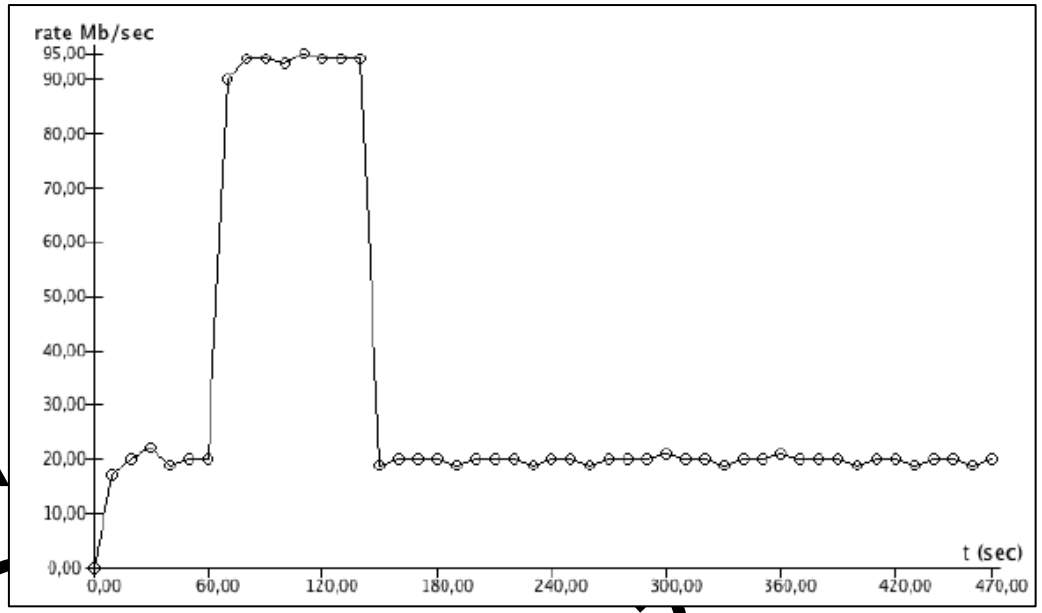
TB QoS is introduced seamlessly into workflow specifications with the Renew tools

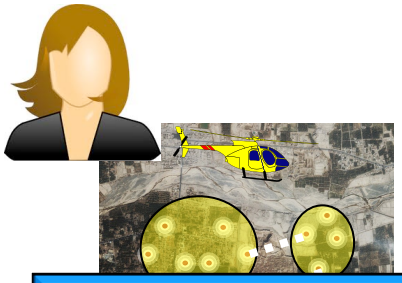


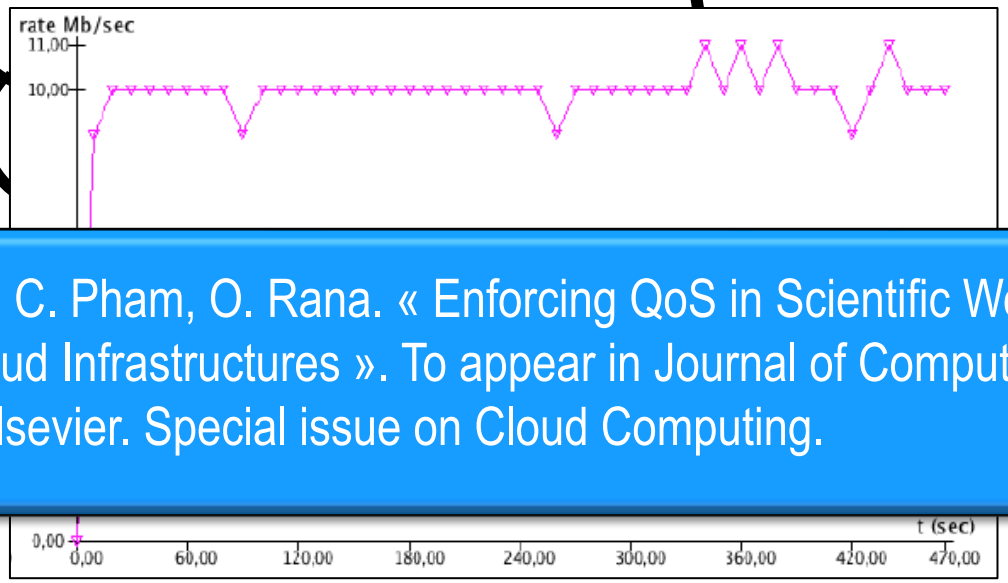
Work done with R. Tolosana, J. Banares and O. Rana

Evaluation scenario

20 MB/s

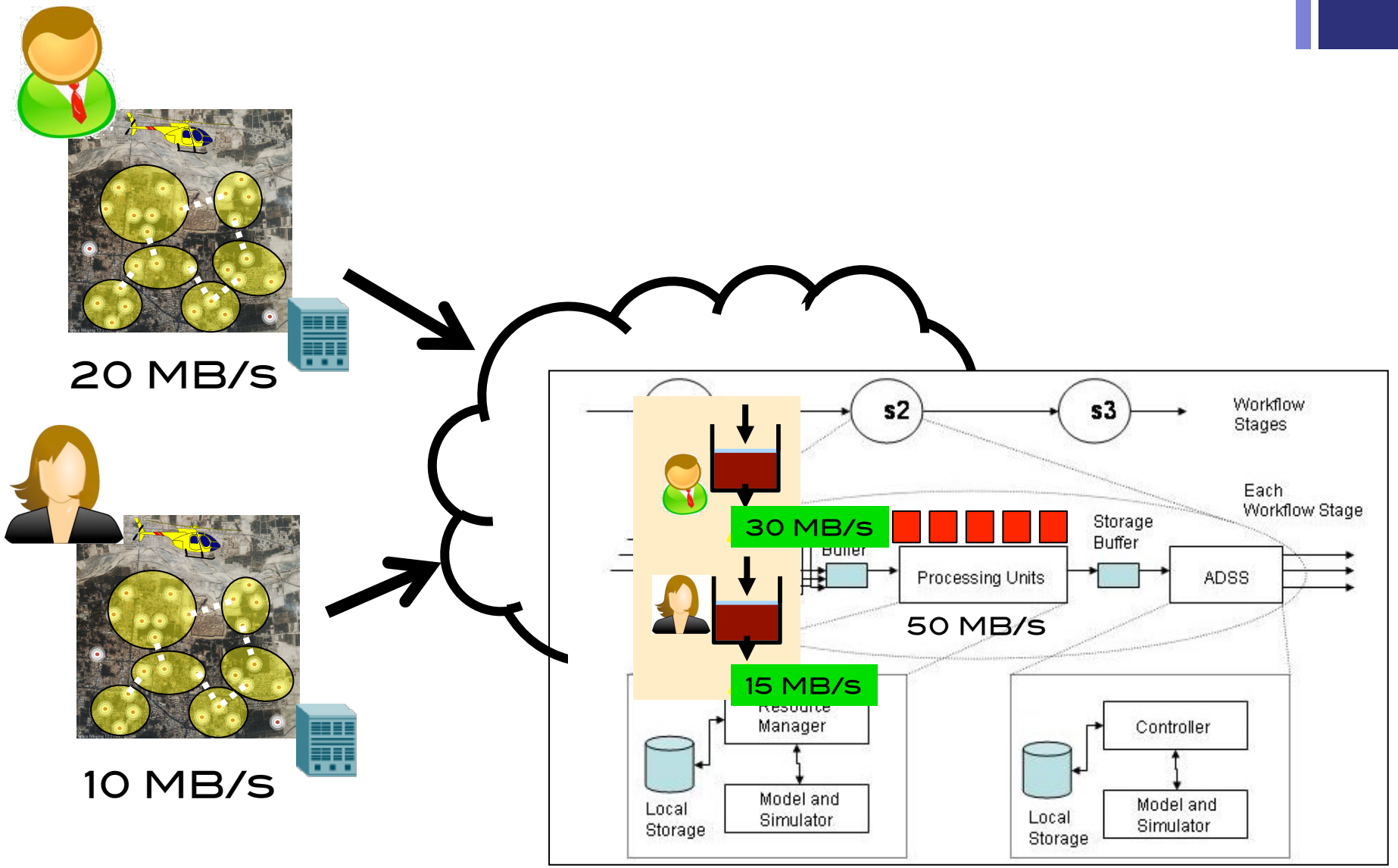






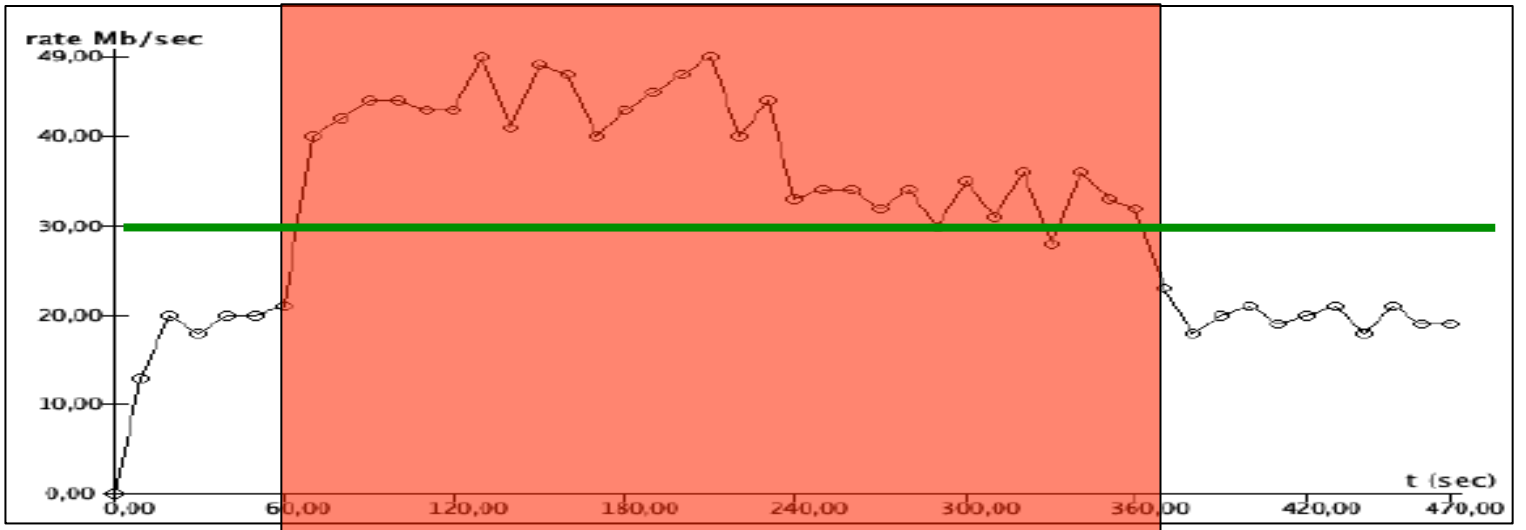
R. Tolosana, J. A. Banares, C. Pham, O. Rana. « Enforcing QoS in Scientific Workflow Systems Enacted Over Cloud Infrastructures ». To appear in Journal of Computer and System Science (JCSS), Elsevier. Special issue on Cloud Computing.

Evaluation scenario

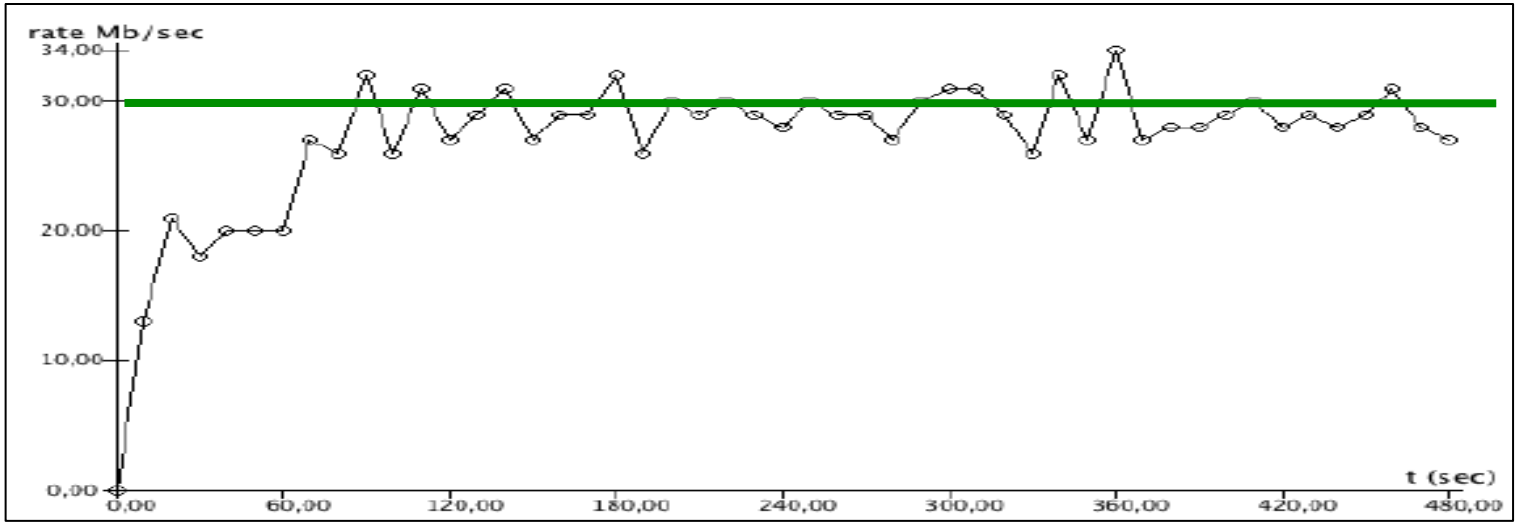


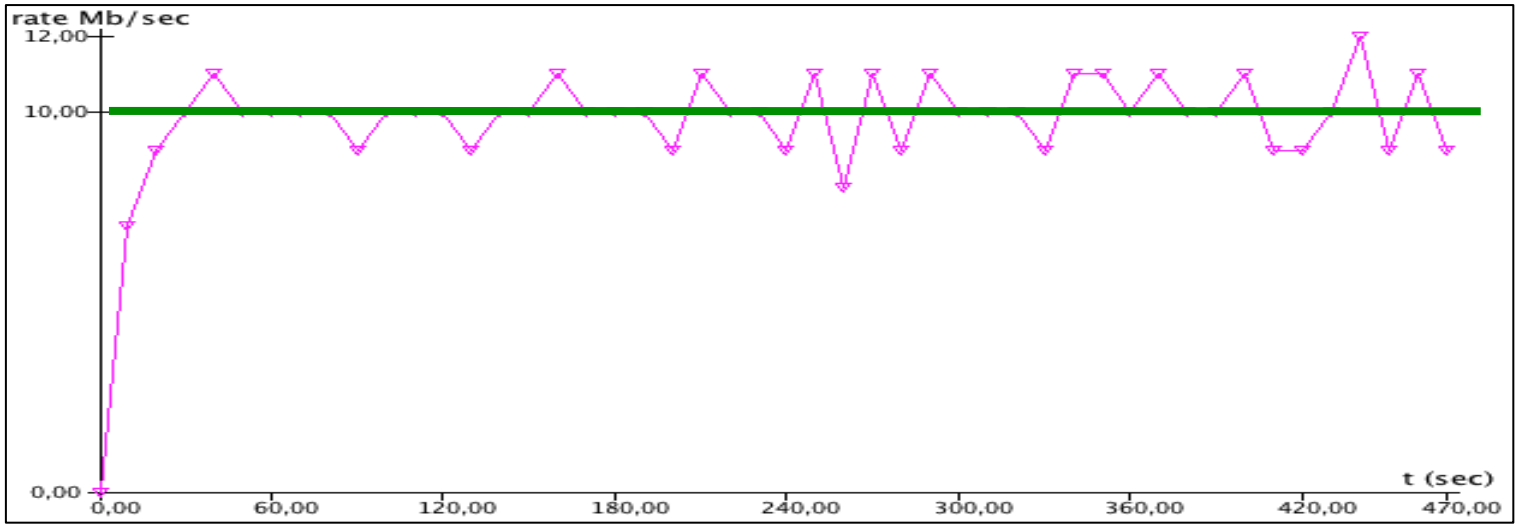
Performance results

without Token Bucket



Performance results
with Token Bucket





Conclusions (1)

- Clouds will be shared clouds driven by economical constraints
- For some applications, availability of resources and isolation are of prime importance (urgent computing)
- QoS for clouds is already a necessary and hot topic in research community

Perspectives (1)

- Add more parameters to the TB model
  - Excess burst size
  - Multiple Token Bucket models
  - Advanced mark vs. drop policy
- Dynamic configuration of TB parameters at each stage of the processing path
- Take into account data inflation behaviors
- Generalized usage of envelope processes, comparison,…

Perspectives (2)

- Tests with real computing cloud testbeds/middleware
- Demonstrator including real sensing hardware & gateways + computing clouds testbeds

Conclusions (2)

- Define & specify needs for south-asia pacific countries in terms of global sensing applications
- GRID/Cloud technology, coupled with global sensing…
  - Environmental
  - Energy
  - Disaster relief
  - …
- …opens up a lot more perspectives!