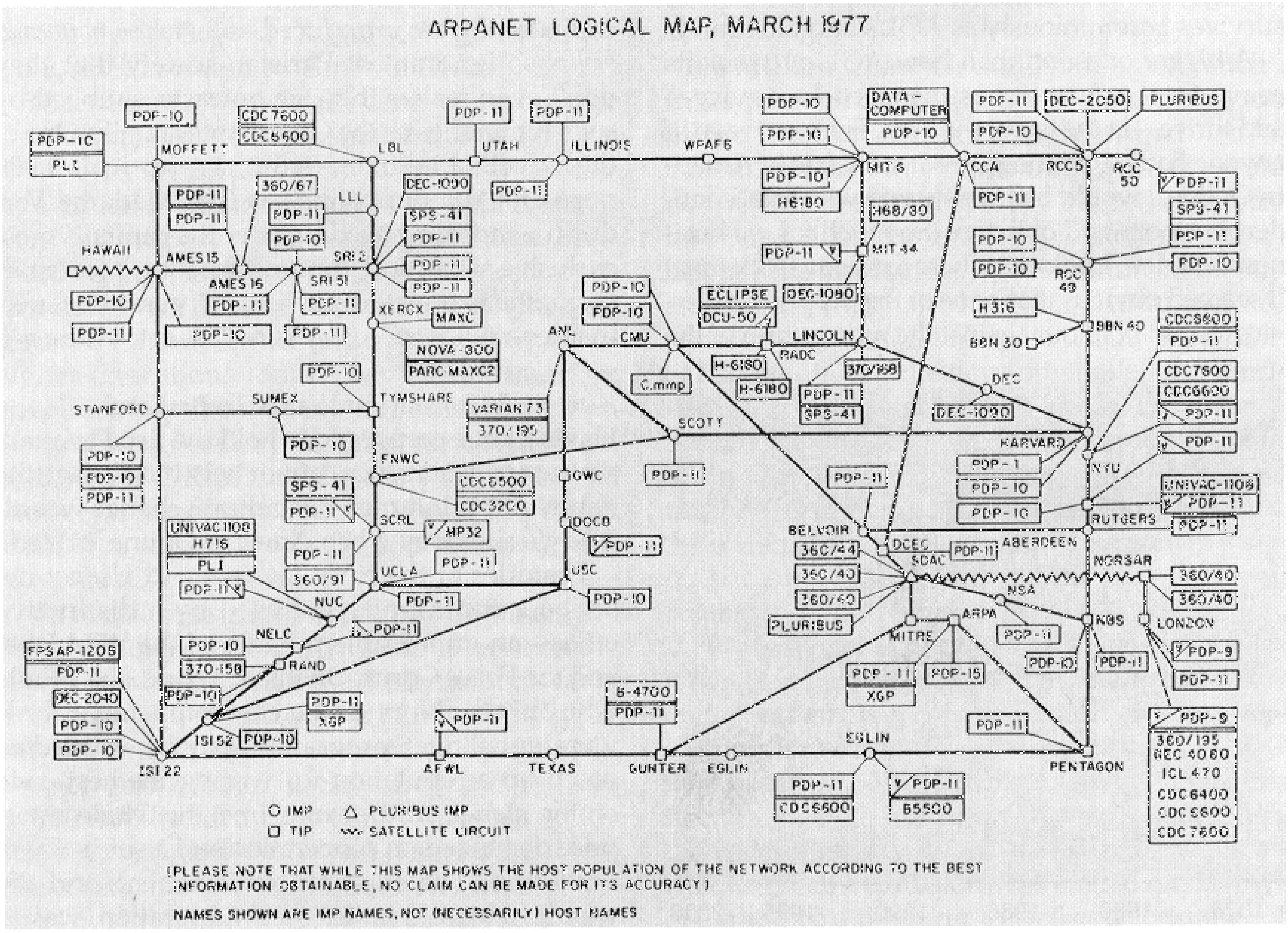
Distributed Systems + Middleware
Introduction

Gianpaolo Cugola
Dipartimento di Elettronica e Informazione
Politecnico di Milano, Italy
cugola@elet.polimi.it
http://home.dei.polimi.it/cugola
http://corsi.dei.polimi.it/distsys

Distributed systems: Yesterday

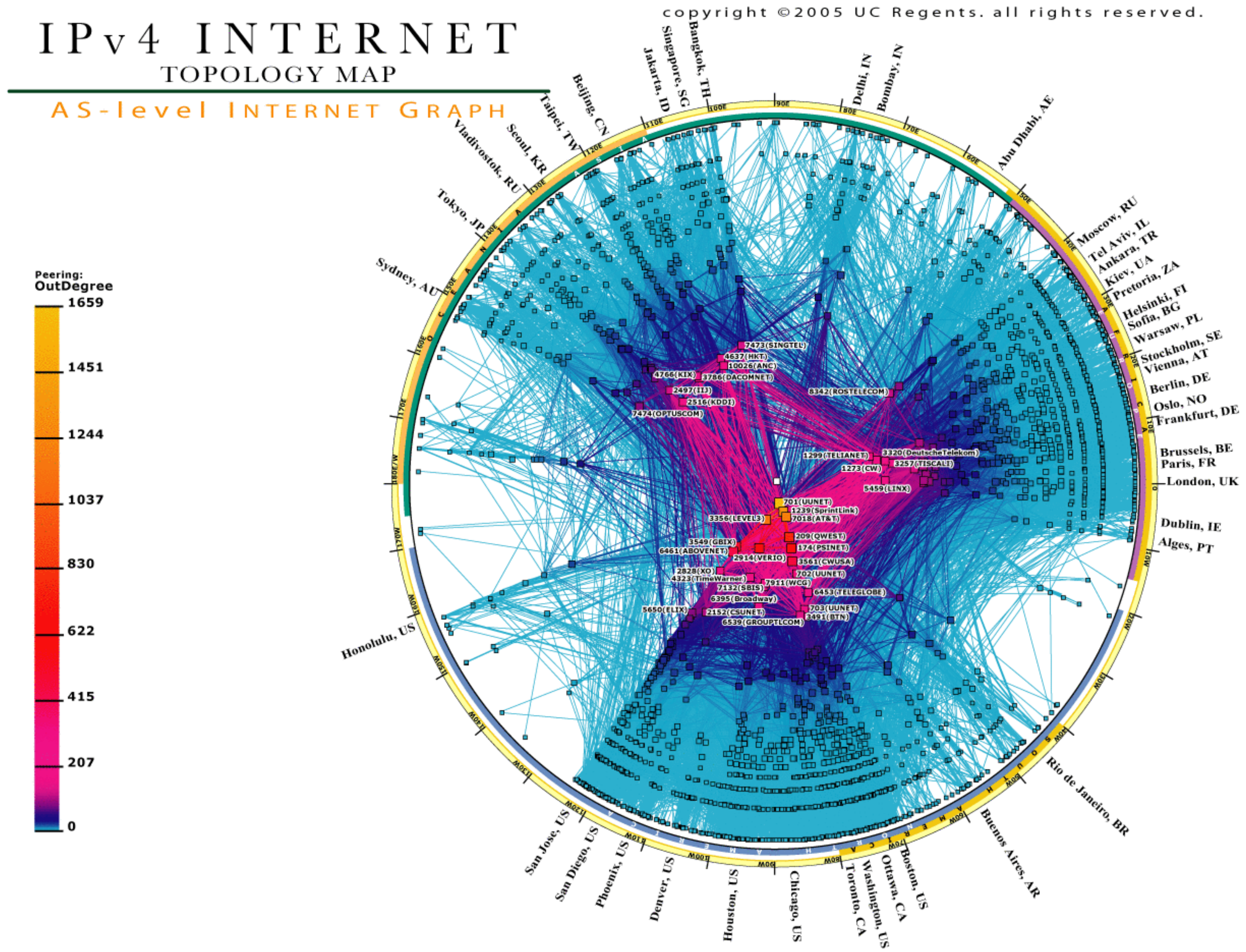

ARPANET LOGICAL MAP, MARCH 1977

(PLEASE NOTE THAT WHILE THIS MAP SHOWS THE HOST POPULATION OF THE NETWORK ACCORDING TO THE BEST INFORMATION OBTAINABLE, NO CLAIM CAN BE MADE FOR ITS ACCURACY)

NAMES SHOWN ARE IMP NAMES, NOT (NECESSARILY) HOST NAMES

Politecnico di Milano

Distributed systems: Today



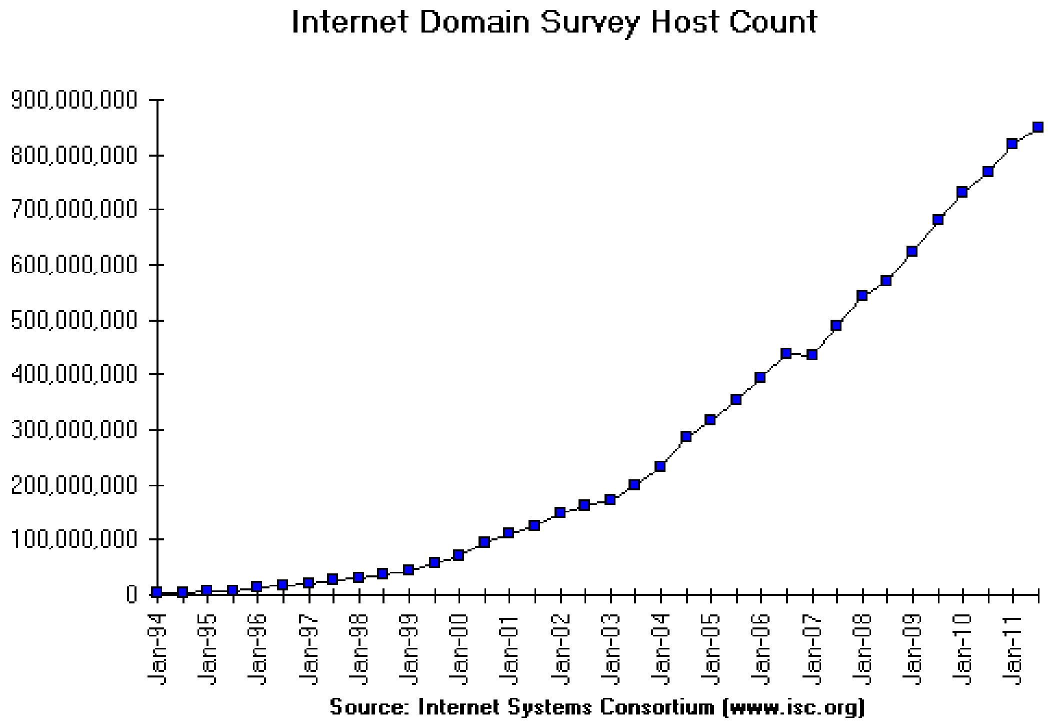
Politecnico di Milano

Internet growth: Hosts

(advertised in dns)

Date	Hosts
Dec 1979	188
July 1989	130,000
July 1999	56,218,000
July 2001	125,888,197
July 2003	171,638,297
July 2005	353,284,187
July 2006	439,286,364
July 2007	489,774,269
July 2008	570,937,778
July 2009	681,064,561
July 2010	768,913,036
Jan 2011	818,374,269
July 2011	849,869,781

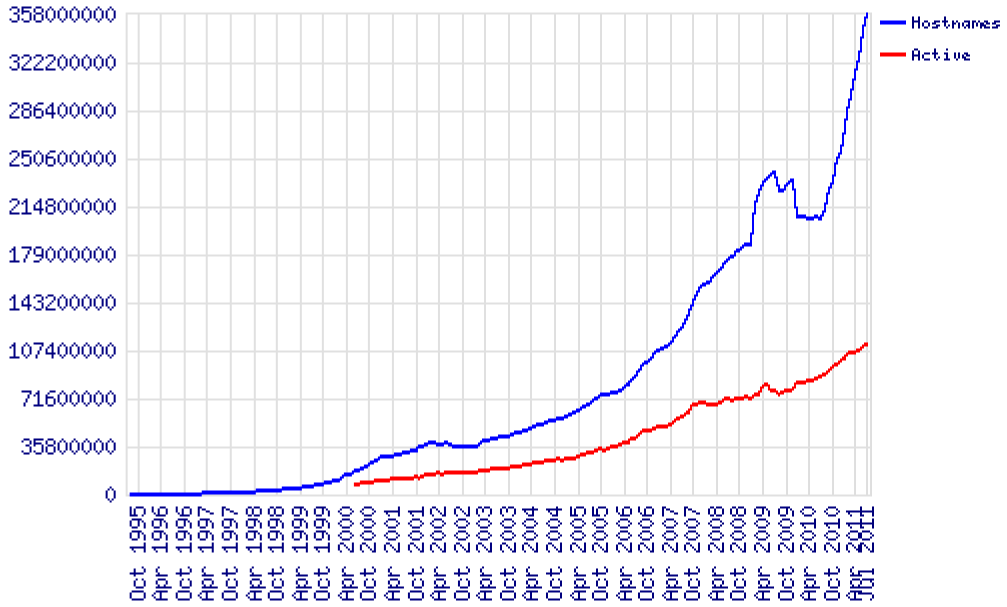
Source: Internet Systems Consortium



Politecnico di Milano

Internet growth: Web servers

Date	Hosts	Web servers	%
July 1993	1,776,000	130	0.008
July 1995	6,642,000	23,500	0.4
July 1997	19,540,000	1,203,096	6
July 1999	56,218,000	6,598,697	12
July 2001	125,888,197	31,299,592	25
July 2003	212,570,000	42,298,371	20
July 2005	353,284,187	67,571,581	19
July 2006	439,286,364	88,166,395	20
July 2007	489,774,269	125,626,329	39
July 2008	570,937,778	175,480,931	33
July 2009	681,064,561	239,611,111	35
July 2010	768,913,036	205,714,253	27
July 2011	849,869,781	357,292,065	42

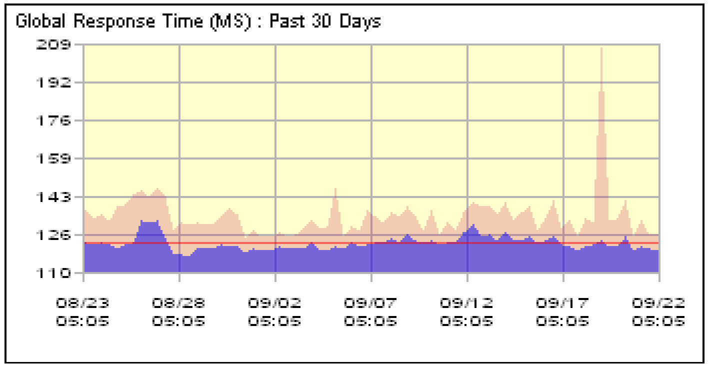


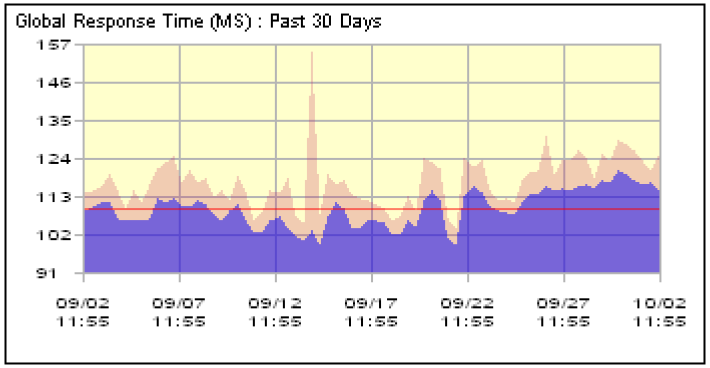
Source: Netcraft web server survey

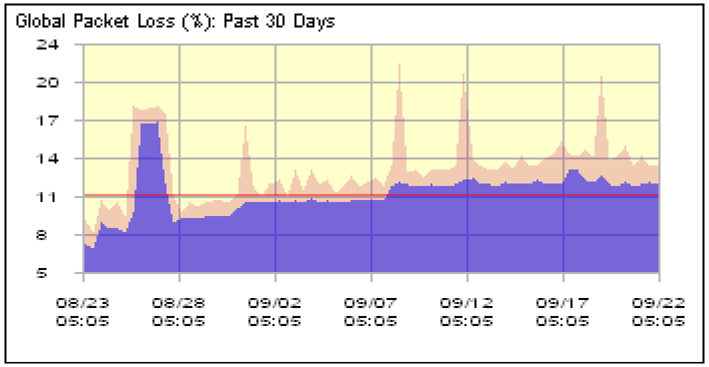
Politecnico di Milano

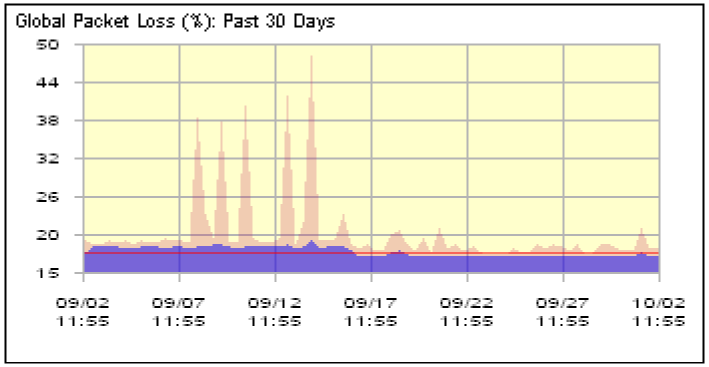
Internet performance

Last year **This year**









Legenda

red=max

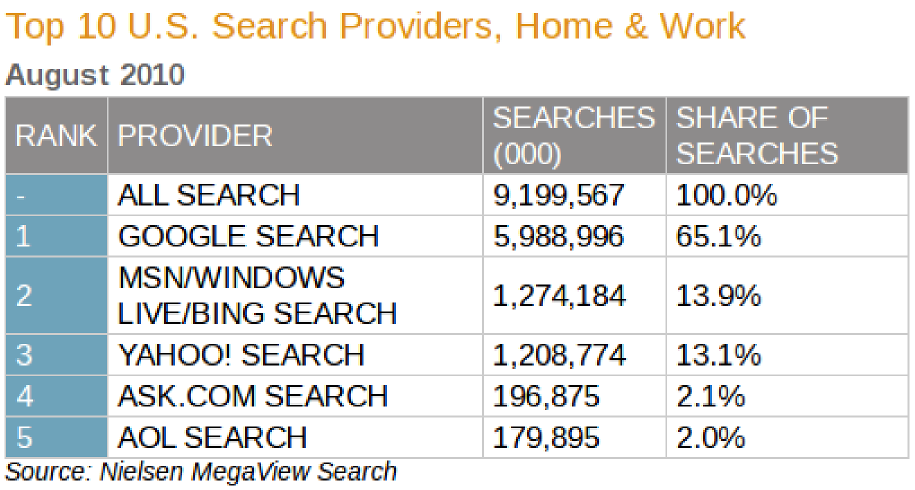
purple=avg

Source: Internet Traffic Report

Top sites and traffic

Site	% of users visiting 1/10
google.com	49.73
facebook.com	43.00
youtube.com	31.91
yahoo.com	23.43
live.com	11.55
baidu.com	10.12
wikipedia.org	13.96

Source: Alexa survey



Top 10 U.S. Search Providers, Home & Work

August 2010

RANK	PROVIDER	SEARCHES (000)	SHARE OF SEARCHES
-	ALL SEARCH	9,199,567	100.0%
1	GOOGLE SEARCH	5,988,996	65.1%
2	MSN/WINDOWS LIVE/BING SEARCH	1,274,184	13.9%
3	YAHOO! SEARCH	1,208,774	13.1%
4	ASK.COM SEARCH	196,875	2.1%
5	AOL SEARCH	179,895	2.0%

Source: *Nielsen MegaView Search*

Politecnico di Milano

However…

- Internet and the Web are distributed systems, but not all distributed systems are on (or related to) the Internet and the Web
  - Intranets (banking and corporate systems, scientific computing, server farms, grids, clouds,...)
  - Home networks (home entertainment, multimedia sharing, UPnP)
  - PANs (cellular phones, PDAs, wireless headsets, GPS radio receivers, healthcare monitoring devices)
  - Wireless Sensor Networks
  - …
- Mobility further complicates matters

Politecnico di Milano

Google – An Intranet :-)

- "Google is technologically a large supercomputer. It's a distributed supercomputer among many data centers doing all sorts of interesting things over fiber optic network that eventually are services available to end-users." - Eric Schmidt, Google CEO, 2007
- "Google runs on hundreds of thousands of servers—by one estimate, in excess of 450,000—racked up in thousands of clusters in dozens of data centers around the world. It has data centers in Dublin, Ireland; in Virginia; and in California, where it just acquired the million-square-foot headquarters it had been leasing. It recently opened a new center in Atlanta, and is currently building two football-field-sized centers in The Dalles, Ore." – 2006
- "Our view is it's better to have twice as much hardware that's not as reliable than half as much that's more reliable...You have to provide reliability on a software level. If you're running 10,000 machines, something is going to die every day." Jeff Dean, Google fellow, 2008
  - The web site monitoring service Pingdom tracked Google's worldwide network of search sites for a one-year period ending in October 2007, and found that all 32 of Google's worldwide search portals (including google.co.uk, google.in, etc.) maintained uptime of at least "four nines" – 99.99 percent. The main site at google.com was down for 31 minutes in the 12-month monitoring period. The best performer was Google Brazil (google.com.br), with 3 minutes of downtime.

Politecnico di Milano

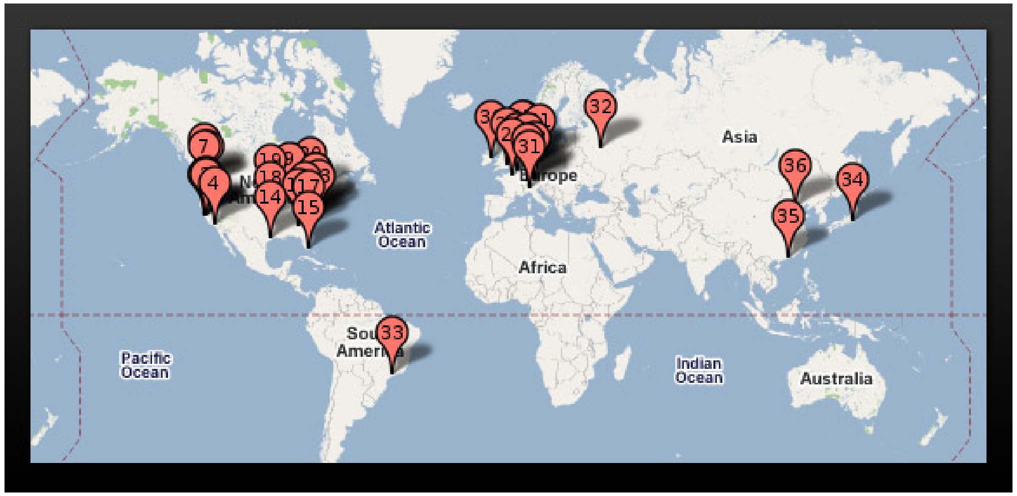
Google (continued)

- According to Google's earnings reports, they spent $1.9 billion on data centers in 2006, and $2.4 billion in 2007.
- In the Dalles, Oregon, Google's site includes three 68,680 square foot data center buildings, a 20,000 square foot administration building, a 16,000 square foot "transient employee dormitory" and an 18,000 square foot facility for cooling towers.

Politecnico di Milano

Google (continued)

- Google's criteria when selecting locations for data centers
  - Large volumes of cheap electricity.
  - Green energy. Focuses on renewable power sources.
  - Proximity to rivers and lakes. They use a large amount of water for cooling purposes.
  - Large areas of land. Allows for more privacy and security.
  - The distance to other Google data centers (for fast connections between data centers).
  - Tax incentives.



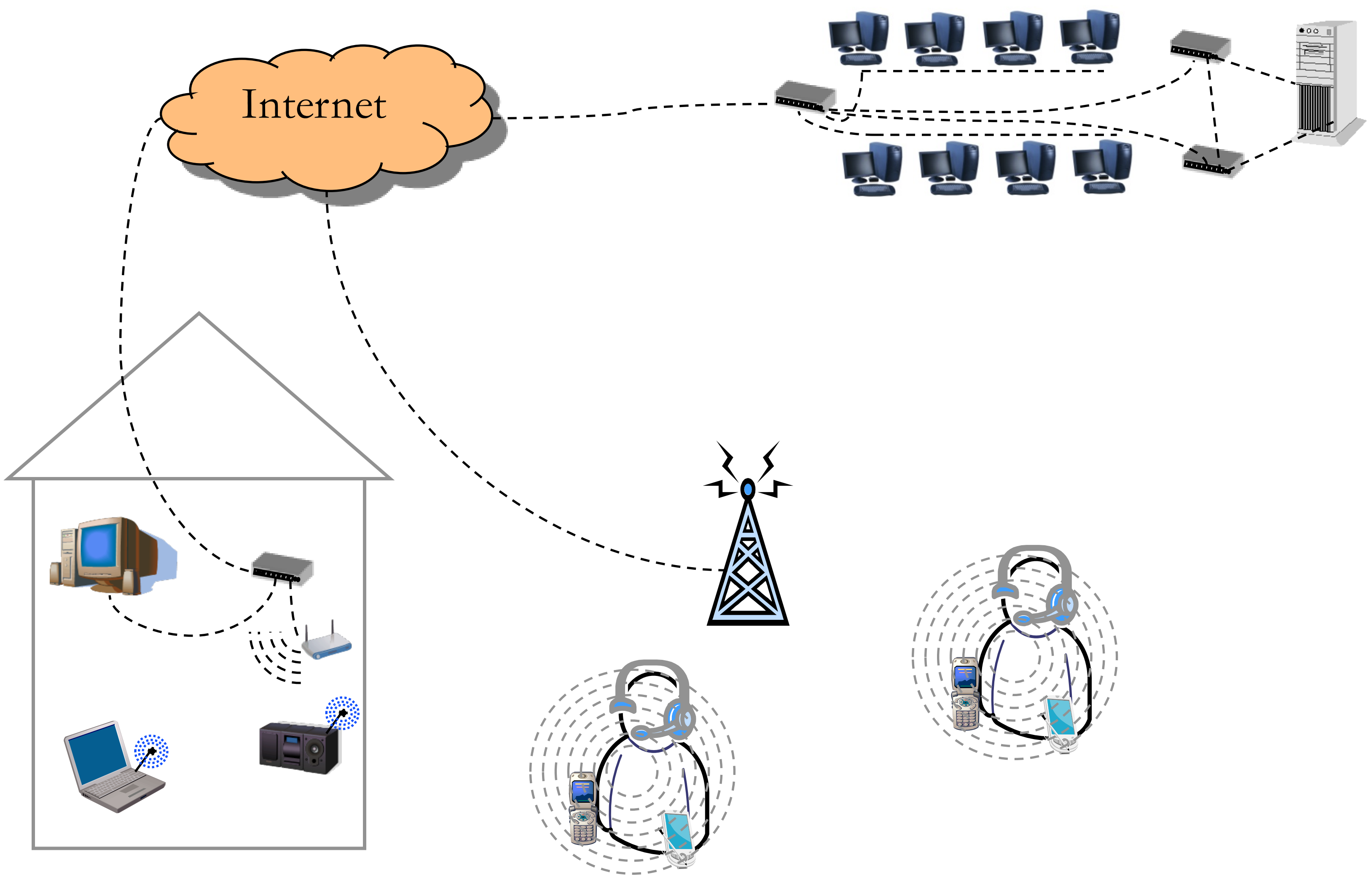
Politecnico di Milano

Google (continued)

- "A typical search will require actions from between 700 to 1,000 machines today." - Marissa Mayer, vice president of Google's search products and user experience, 2008
- "Google is fast — a typical search returns results in less than 0.2 seconds... Together with other work performed before your search even starts (such as building the search index) this amounts to 0.0003 kWh of energy per search, or 1 kJ. For comparison, the average adult needs about 8000 kJ a day of energy from food, so a Google search uses just about the same amount of energy that your body burns in ten seconds....
- ... In terms of greenhouse gases, one Google search is equivalent to about 0.2 grams of CO2.... The average car driven for one kilometer produces as many greenhouse gases as a thousand Google searches." – Urs Hölzle, Senior Vice President, Operations, googleblog, 2009

Politecnico di Milano

Home and personal networks



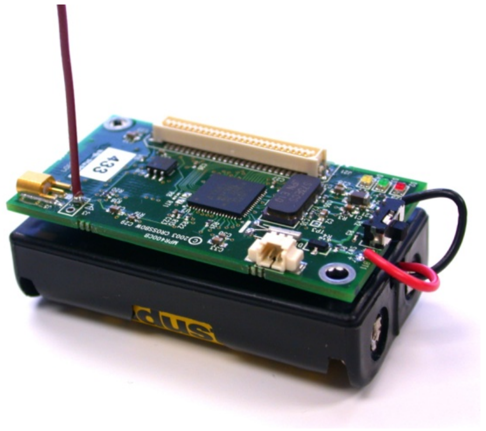
Politecnico di Milano

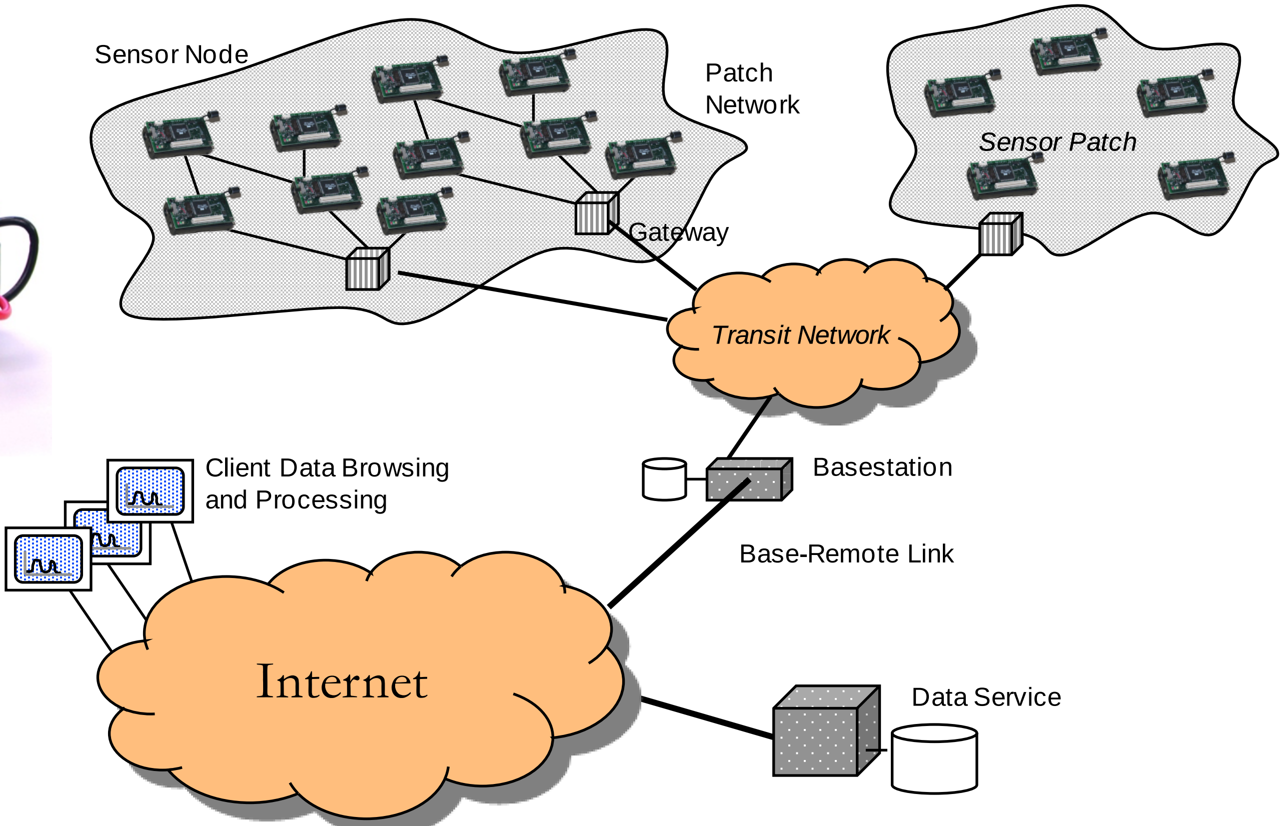
Pervasiveness: An Internet of Things

- "The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it... only when things disappear in this way are we freed to use them without thinking and so to focus beyond them on new goals." – Mark Weiser, 1991
- "... by embedding short-range mobile transceivers into a wide array of gadgets and everyday items..., a new dimension has been added to the world of information and communication technologies: from anytime, any place connectivity for anyone, we will now have connectivity for anything." – The Internet of Things, ITU Internet Reports 2005.
- "In the future, everything of value will be on the network in one form of another," - John Fowler, Software CTO of Sun Microsystems.
- "When we talk about an Internet of things, it's not just putting RFID tags on some dumb thing so we smart people know where that dumb thing is. It's about embedding intelligence so things become smarter and do more than they were proposed to do." - Nicholas Negroponte, 2005

Politecnico di Milano

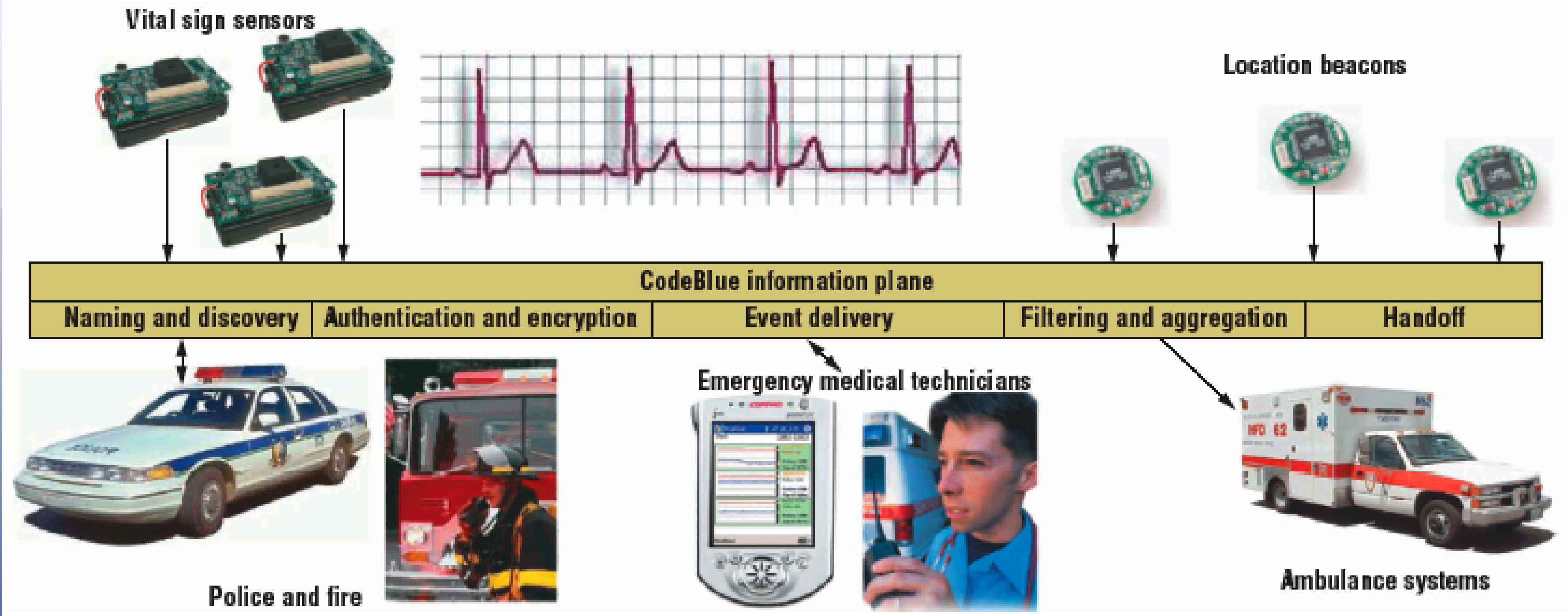
Wireless sensor networks



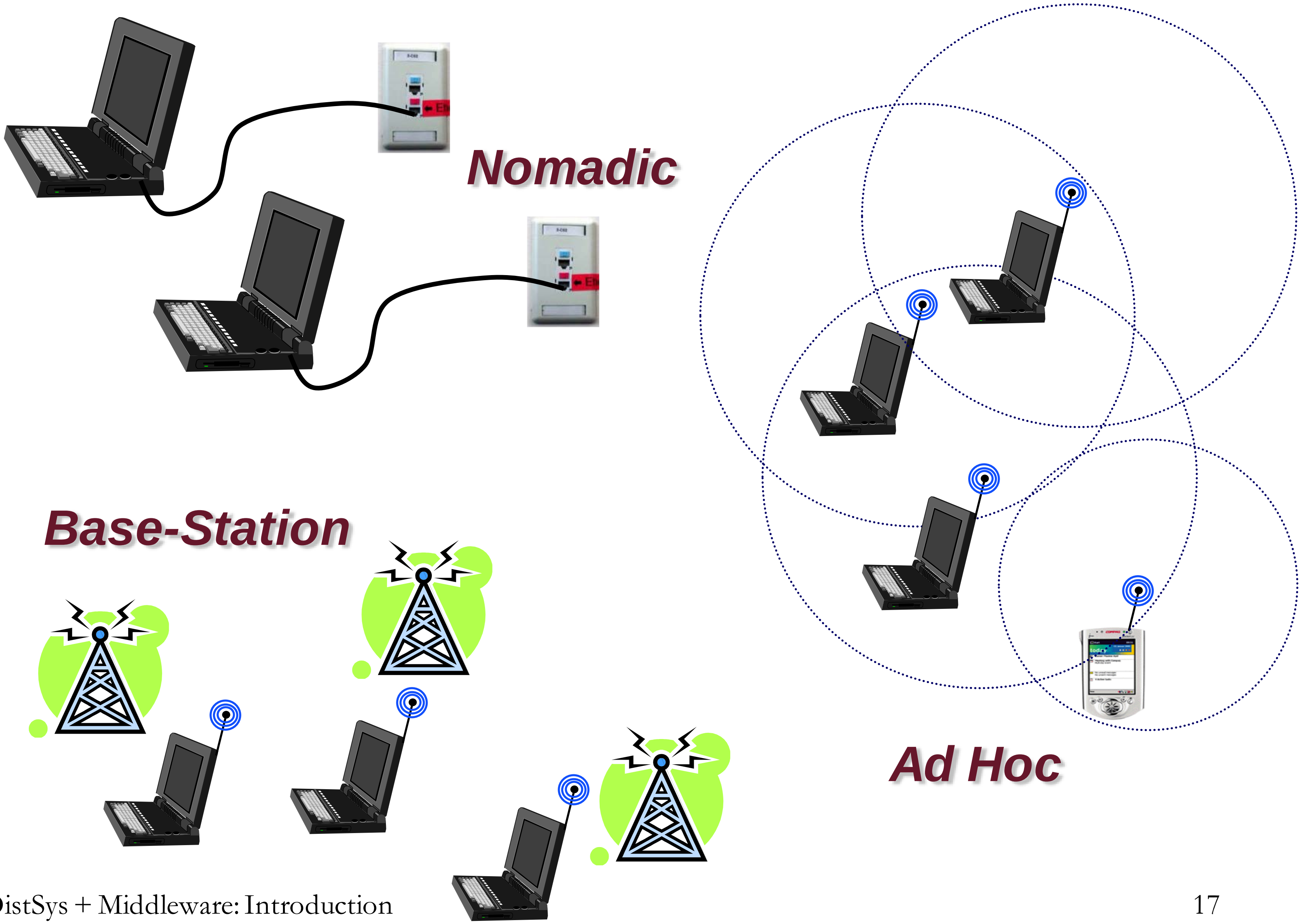


Politecnico di Milano

Example application: Emergency response


Vital sign sensors
Location beacons
CodeBlue information plane
Naming and discovery
Authentication and encryption
Event delivery
Filtering and aggregation
Handoff
Emergency medical technicians
Police and fire
Ambulance systems

Mobility

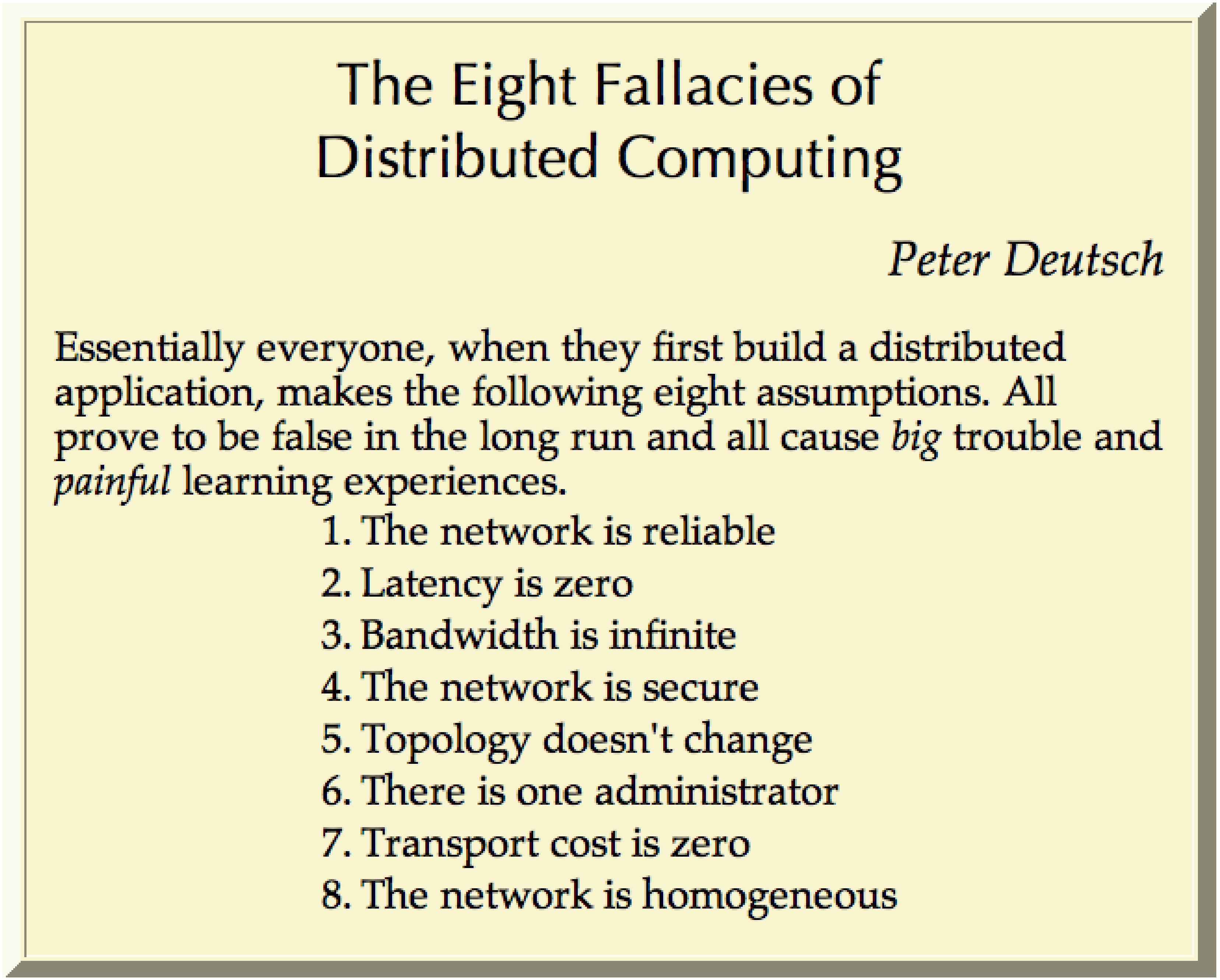


Politecnico di Milano

Why to distribute?

- Economics
  - Originally mainframes consolidated processing
  - Price/performance ratio in favor of distribution
- Incremental growth and load balancing
  - Easier to evolve the system and use its resources uniformly
- Inherent distribution
  - e.g., banks, reservation services, distributed games, CSCW, mobile apps
- Reliability
  - Failure does not bring down the whole system (important for mission critical applications)
- However, distributed systems are considerably more difficult to get right!!!

Politecnico di Milano



The Eight Fallacies of Distributed Computing

Peter Deutsch

Essentially everyone, when they first build a distributed application, makes the following eight assumptions. All prove to be false in the long run and all cause *big* trouble and *painful* learning experiences.

1. The network is reliable
2. Latency is zero
3. Bandwidth is infinite
4. The network is secure
5. Topology doesn't change
6. There is one administrator
7. Transport cost is zero
8. The network is homogeneous

Politecnico di Milano

Course outline

- Introduction (this slide set)
- Concurrent programming
  - In C and Java
- Modelling distributed systems
- Communication models
- Naming
- Synchronization
- Consistency and Replication
- Fault tolerance
- Security
- Peer-to-peer computing + JXTA (cenni)
- Simulating distributed systems with OmNet++
- Examples
  - Distributed object-based systems
  - Distributed file systems
  - Distributed coordination-based systems

- Introduction (this slide set)
- Concurrent programming
  - In C and Java
- Basic communication facilities
  - Sockets (uni & multi), RPC, Java RMI
- MOM (JMS)
- CORBA
- Sun JavaSpaces
- Middleware for service oriented computing
- Distributed programming in Erlang
- Middleware for WSNs

What is a distributed system?

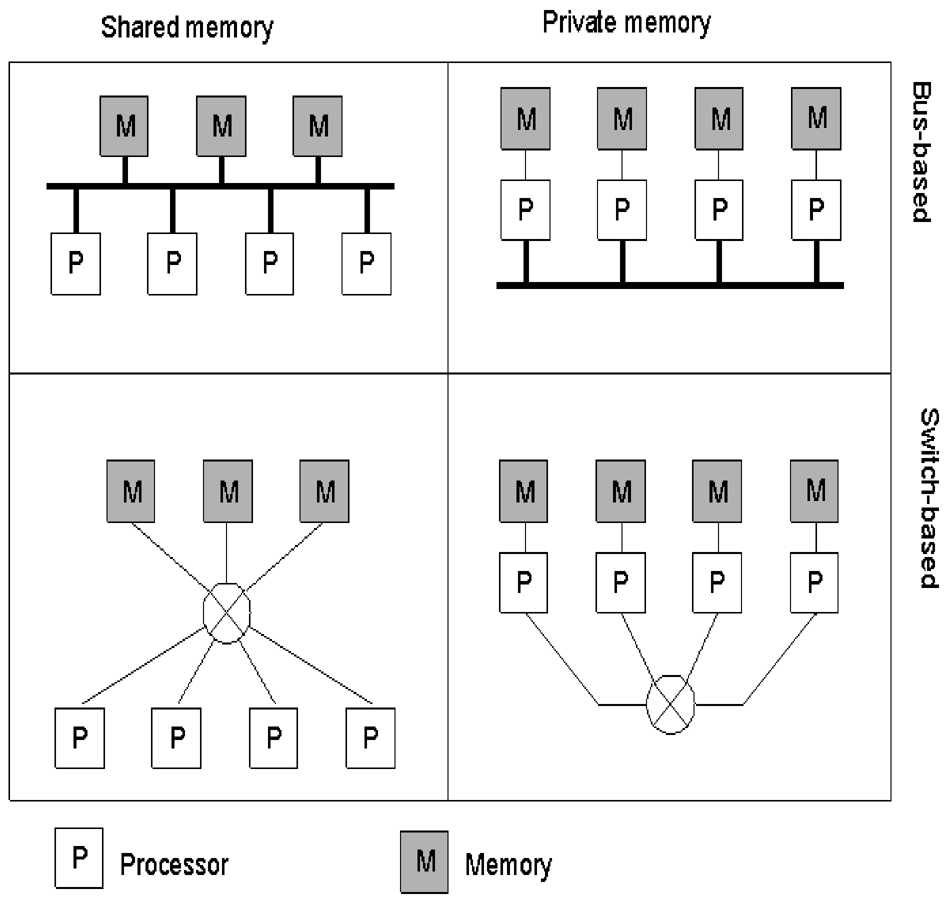
"A collection of independent computers that appears to its users as a single coherent system"
-- A.S. Tanenbaum, M. van Steen

"One in which hardware or software components located at networked computers communicate and coordinate their actions only by passing messages"
-- G. Coulouris, J. Dollimore, T. Kindberg

"One on which I cannot get any work done because some machine I have never heard of has crashed"
-- L. Lamport

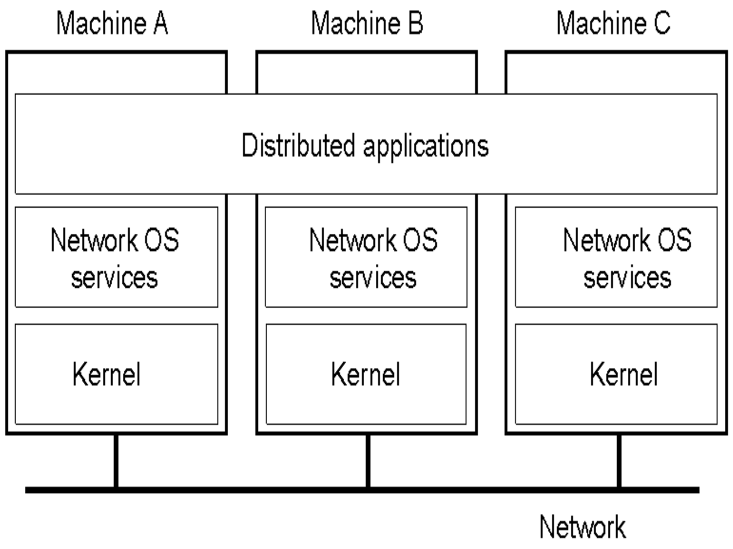
Politecnico di Milano

Parallel computing vs. distributed computing



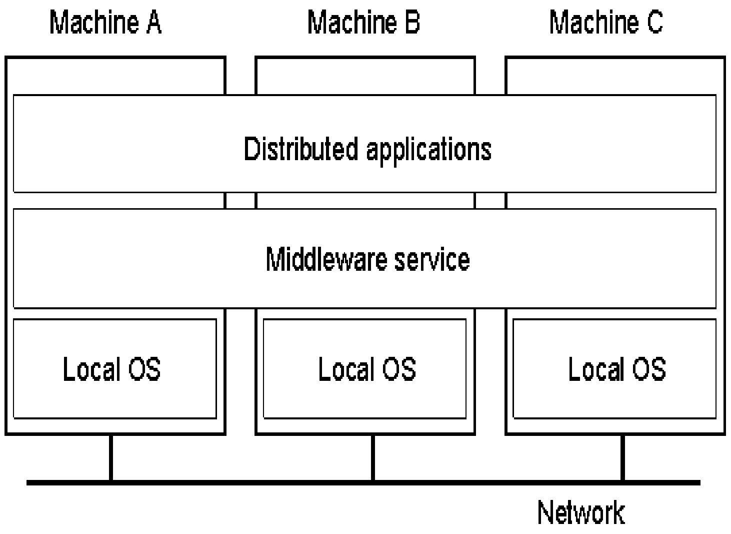
Politecnico di Milano

Where middleware fits



Network/OS based distributed system

Middleware based distributed system



- Middleware provides horizontal services to help building distributed applications
- It also masks platform differences

Politecnico di Milano

Some defining features

- ***Concurrency***
  In centralized systems, concurrency is a design choice. In distributed systems, it is a fact of life to be dealt with: computers are always (co)-operating at the same time
- ***Absence of a global clock***
  In centralized systems, the computer's physical clock can be used for the purpose of synchronization. In distributed systems, clocks are many and not necessarily synched
- ***Independent (and partial) failures***
  Centralized systems typically fail completely. Distributed systems fail only partially, and often due to communication. Each component can fail independently leaving the others still running. Hard/impossible to detect failures. Moreover, recovery is made more complicated by the fact that the application state is itself distributed

Politecnico di Milano

Some challenges

- *Heterogeneity*
  Of hosts, platforms, networks, protocols, languages, …
- *Openness*
  A distributed system should foster interoperability through standard access rules. Protocols and interfaces play a key role
- *Security*
  The ease of attaching a node to the system can be exploited maliciously
- *Scalability*
  Use decentralized algorithms to allow the system to grow with the lowest possible impact on performance
- *Failure handling*
  Hosts can fail, links are unreliable, the two are usually undistinguishable, global state complicates matters
  Detecting, masking, tolerating, recovery from failures
- *Concurrency*
  Synchronization without a physical clock and without shared memory
- *Transparency*
  Hide the most to simplify the life of programmers/users

Heterogeneity

- Networks
  - Even on the Internet, where TCP/IP seems to hide most of the networking differences, the differences at the OSI layers 1 and 2 must be carefully considered in building distributed applications
    - E.g., A wireless ad-hoc scenario is totally different from a wired, fixed scenario
  - Things become more complex when the differences span to higher OSI layers
- Computer hardware
  - Data representation vary with the hw platform and the same holds for several other aspects (e.g., performance)
- Operating systems
  - The API provided to access the basic networking protocols (including TCP/IP) may vary. Other aspects change also with the platform (e.g., the semantics of thread interleaving in Java)
- Programming languages
  - Differen languages represent data structures differently. These differences must be addressed if programs written in different languages must communicate

Politecnico di Milano

Openness

- Determines whether a system can be extended and re-implemented in various ways
- To be achieved requires
  - Publishing all the key aspects of the system
    - Protocols
    - Interfaces to services
    - …
  - Adopting standards as much as possible
  - Take design decisions that could ease interoperability and portability
    - Always choosing the simplest solution
- The Internet case: RFCs and an open standardization body (the IETF)

Politecnico di Milano

Security

- Security for the information resources made available and maintained in distributed systems has three components:
  - Confidentiality
    - Protection against disclosure to unauthorized individuals
  - Integrity
    - Protection against alteration or corruption
  - Availability
    - Protection against interfearence with the means to access the resources (DOS attacks)
- Encription is a powerfull mechanisms but several issues still to be solved
  - DOS attacks
  - Security in mobile code applications
  - Secure content-based routing
  - ...

Politecnico di Milano

Scalability

- Ability to increase the size of the system in terms of users/resources, geographic span, administrative span
  - Challenge is to reconcile scale with performance
- Typical points of centralization:
  - Services
    - E.g., a single server for all users
  - Data
    - E.g., a single on-line telephone book
  - Algorithms
    - E.g., doing routing based on complete information
- Decentralized algorithms:
  - No machine has complete information about the system state
  - Machines make decisions based only on local information
  - Failure of one machine does not ruin the algorithm
  - There is no implicit assumption that a global clock exists
- Several techniques available
  - Asynchronous communication, caching and replication, epidemic dissemination, hierarchical structuring, …

Failure handling

- Distributed systems fails only partially… This is BAD
- Issues
  - Detecting failures
    - Some failures are easy to detect (e.g., corrupted messages through checksums), while other are impossible (e.g., crash of a remote server)
    - The challenge is to manage the rpesence of failures that cannot be detected for sure but may be suspected
  - Masking failures
    - Some failures that have been detected can be easily masked (e.g., messages can be retransmitted when they fail to arrive)
  - Tolerating failures
    - Components may be designed to tolerate failures of other components, degrading the services they offer, if necessary
  - Recovering from failures
    - By periodically saving the state of a component on a permanent storage it is possible to recover from failures of that component
  - Redundancy
    - It is a common technique to mask/tolerate failures (e.g., using replicated servers or replicated routes)
    - Keeping replicas in a distributed setting is not easy

Politecnico di Milano

Concurrency

- Concurrency in distributed systems is a matter of fact
- Access to shared resources (information or services) must be carefully synchronized
- OS and language run-time systems provide several mechanisms to cope with this issue

Politecnico di Milano

Transparency

Transparency	Description
Access	Hide differences in data representation and how a resource is accessed
Location	Hide where a resource is located
Migration	Hide that a resource may move to another location
Relocation	Hide that a resource may be moved to another location while in use
Replication	Hide that a resource may be shared by several competitive users
Concurrency	Hide that a resource may be shared by several competitive users
Failure	Hide the failure and recovery of a resource
Persistence	Hide whether a (software) resource is in memory or on disk

- Some things cannot be made transparent
  - Timezones, communication delays
- Hiding too much may have a strong, negative performance impact
  - E.g., accessing repeatedly a remote object without knowing