Cloud Computing

The concept

- Cloud computing
  - is the use of computing resources (hardware and software) that are delivered as a service over a network (typically the Internet).

- NIST Cloud Computing definition
  - “Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.”
    - http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

Cloud computing Essential Characteristics

On-demand self-service. A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider.

Broad network access. Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, tablets, laptops, and workstations).

Resource pooling. The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand.

Rapid elasticity. Capabilities can be elastically provisioned and released, in some cases automatically, to scale rapidly outward and inward commensurate with demand. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be appropriated in any quantity at any time.

Measured service. Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported, providing transparency for both the provider and consumer of the utilized service.

From NIST: http://csrc.nist.gov/publications/nistpubs/800-145/sp800-145.pdf

Cloud computing

- Uses Internet technologies to offer scalable and elastic services.

- The term “**elastic computing**” refers to the ability of *dynamically acquiring computing resources* and supporting a variable workload.
    - The resources used for these services can be metered and the *users can be charged only for the resources they used*.
        - Utility computing service provisioning model

- The maintenance and security are ensured by service providers.

- The service providers can operate more efficiently due to specialization and centralization.

Cloud computing (cont’d)

- Lower costs for the cloud service provider are past to the cloud users.
- Data is stored:
  - closer to the site where it is used.
  - in a device and in a location-independent manner.
- The data storage strategy can increase reliability, as well as security, and can lower communication costs.

Types of clouds

- Public Cloud - the infrastructure is made available to the general public or a large industry group and is owned by the organization selling cloud services.
- Private Cloud – the infrastructure is operated solely for an organization.
- Community Cloud - the infrastructure is shared by several organizations and supports a community that has shared concerns.
- Hybrid Cloud - composition of two or more clouds (public, private, or community) as unique entities but bound by standardized technology that enables data and application portability.

The “good” about cloud computing

- Resources, such as CPU cycles, storage, network bandwidth, are shared.

- When multiple applications share a system, their peak demands for resources are not synchronized thus, *multiplexing leads to a higher resource utilization*.

- Resources can be aggregated to support data-intensive applications.

- Data sharing facilitates collaborative activities. Many applications require multiple types of analysis of shared data sets and multiple decisions carried out by groups scattered around the globe.

More “good” about cloud computing

- Eliminates the initial investment costs for a private computing infrastructure and the maintenance and operation costs.

- Cost reduction: concentration of resources creates the opportunity to pay as you go for computing.

- Elasticity: the ability to accommodate workloads with very large peak-to-average ratios.

- User convenience: virtualization allows users to operate in familiar environments rather than in idiosyncratic ones.

Why cloud computing could be successful when other paradigms have failed?

- It is in a better position to exploit recent advances in software, networking, storage, and processor technologies promoted by the same companies who provide cloud services.
- It is focused on enterprise computing; its adoption by industrial organizations, financial institutions, government, and so on could have a huge impact on the economy.
- A cloud consists of a homogeneous set of hardware and software resources.
- The resources are in a single administrative domain (AD). Security, resource management, fault-tolerance, and quality of service are less challenging than in a heterogeneous environment with resources in multiple ADs.

Computing Paradigm Concepts / Distinctions

Centralized Computing

All computer resources are centralized in one physical system.

Parallel Computing

All processors are either tightly coupled with central shared memory or loosely coupled with distributed memory

Distributed Computing

Field of CS/CE that studies distributed systems. A distributed system consists of multiple autonomous computers, each with its own private memory, communicating over a network.

Cloud Computing

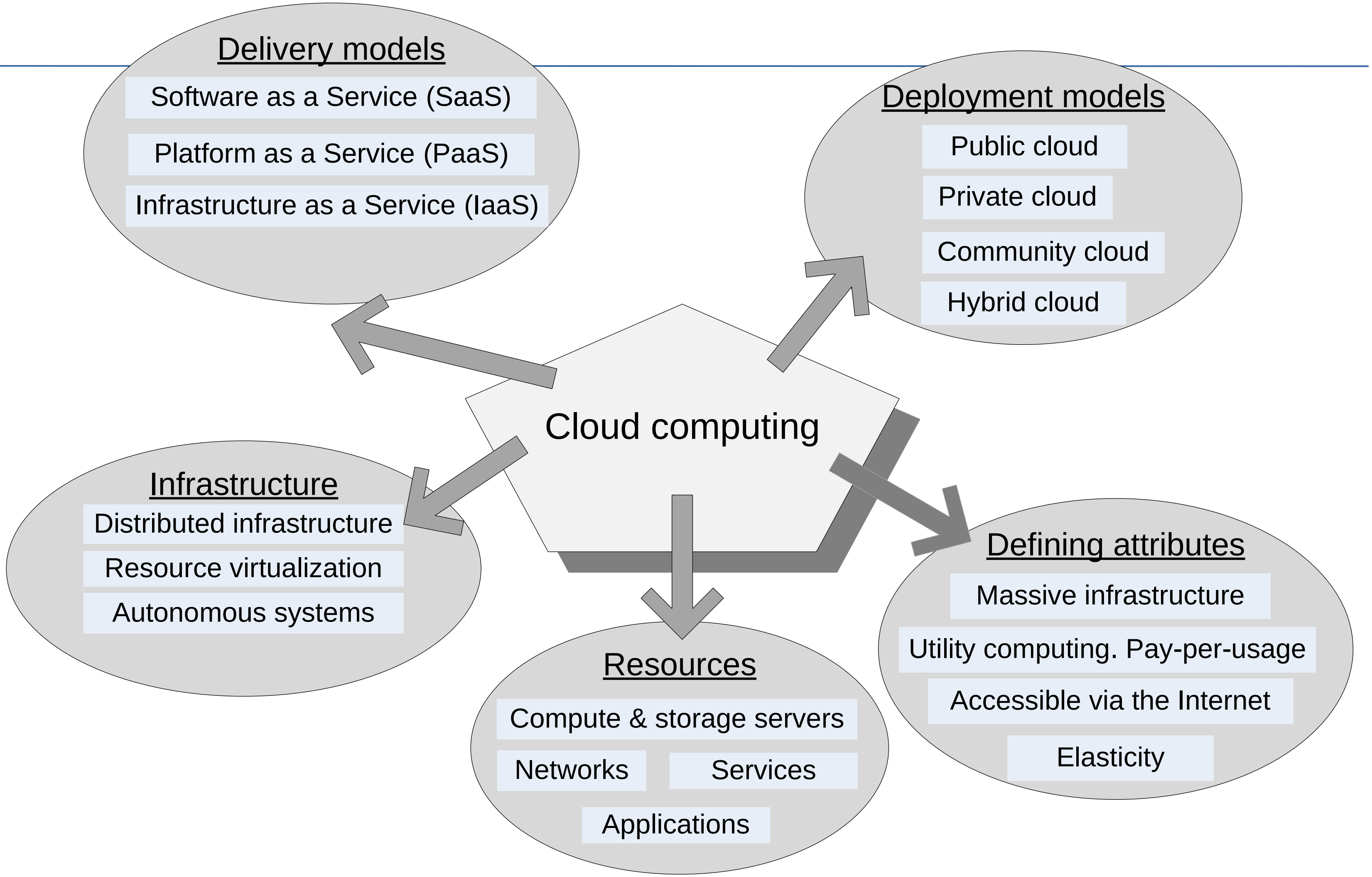
An Internet cloud of resources that may be either centralized or decentralized. The cloud can involve parallel or distributed computing or both. Clouds may be built from physical or virtualized resources.

Challenges for cloud computing

- Availability of service; what happens when the service provider cannot deliver?
- Diversity of services, data organization, user interfaces available at different service providers limit user mobility; once a customer is hooked to one provider it is hard to move to another. Standardization efforts at NIST!
- Data confidentiality and auditability, a serious problem.
- Data transfer bottleneck; many applications are data-intensive.

More challenges

- Performance unpredictability, one of the consequences of resource sharing.
  - How to use resource virtualization and performance isolation for QoS guarantees?
  - How to support elasticity, the ability to scale up and down quickly?
- Resource management; are self-organization and self-management the solution?
- Security and confidentiality; major concern.
- Addressing these challenges provides good research opportunities!!



Cloud computing

Delivery models
- Software as a Service (SaaS)
- Platform as a Service (PaaS)
- Infrastructure as a Service (IaaS)

Deployment models
- Public cloud
- Private cloud
- Community cloud
- Hybrid cloud

Infrastructure
- Distributed infrastructure
- Resource virtualization
- Autonomous systems

Resources
- Compute & storage servers
- Networks
- Services
- Applications

Defining attributes
- Massive infrastructure
- Utility computing. Pay-per-usage
- Accessible via the Internet
- Elasticity

Software-as-a-Service (SaaS)

- Applications are supplied by the service provider.
- The user does not manage or control the underlying cloud infrastructure or individual application capabilities.
- Services offered include:
  - Enterprise services such as: workflow management, group-ware and collaborative, supply chain, communications, digital signature, customer relationship management (CRM), desktop software, financial management, geo-spatial, and search.
  - Web 2.0 applications such as: metadata management, social networking, blogs, wiki services, and portal services.
- Not suitable for real-time applications or for those where data is not allowed to be hosted externally.
- Examples: Gmail, Google search engine.

Platform-as-a-Service (PaaS)

- Allows a cloud user to deploy consumer-created or acquired applications using programming languages and tools supported by the service provider.
- The user:
  - Has control over the deployed applications and, possibly, application hosting environment configurations.
  - Does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage.
- Not particularly useful when:
  - The application must be portable.
  - Proprietary programming languages are used.
  - The hardware and software must be customized to improve the performance of the application.

Infrastructure-as-a-Service (IaaS)

- The user is able to deploy and run arbitrary software, which can include operating systems and applications.

- The user does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of some networking components, e.g., host firewalls.

- Services offered by this delivery model include: server hosting, Web servers, storage, computing hardware, operating systems, virtual instances, load balancing, Internet access, and bandwidth provisioning.

Cloud activities

- Service management and provisioning including:
  - Virtualization.
  - Service provisioning.
  - Call center.
  - Operations management.
  - Systems management.
  - QoS management.
  - Billing and accounting, asset management.
  - SLA management.
  - Technical support and backups.

Cloud activities (cont'd)

- Security management including:
  - ID and authentication.
  - Certification and accreditation.
  - Intrusion prevention.
  - Intrusion detection.
  - Virus protection.
  - Cryptography.
  - Physical security, incident response.
  - Access control, audit and trails, and firewalls.

Cloud activities (cont'd)

- Customer services such as:
  - Customer assistance and on-line help.
  - Subscriptions.
  - Business intelligence.
  - Reporting.
  - Customer preferences.
  - Personalization.
- Integration services including:
  - Data management.
  - Development.

Ethical issues

- Paradigm shift with implications on computing ethics:
  - The control is relinquished to third party services.
  - The data is stored on multiple sites administered by several organizations.
  - Multiple services interoperate across the network.
- Implications
  - Unauthorized access.
  - Data corruption.
  - Infrastructure failure, and service unavailability.

De-perimeterisation

- Systems can span the boundaries of multiple organizations and cross the security borders.

- The complex structure of cloud services can make it difficult to determine who is responsible in case something undesirable happens.

- Identity fraud and theft are made possible by the unauthorized access to personal data in circulation and by new forms of dissemination through social networks and they could also pose a danger to cloud computing.

Privacy issues

- Cloud service providers have already collected petabytes of sensitive personal information stored in data centers around the world.
  - The acceptance of cloud computing therefore will be determined by privacy issues addressed by these companies and the countries where the data centers are located.

- Privacy is affected by cultural differences; some cultures favor privacy, others emphasize community.
  - This leads to an ambivalent attitude towards privacy in the Internet which is a global system.

Cloud vulnerabilities

- Clouds are affected by malicious attacks and failures of the infrastructure, e.g., power failures.

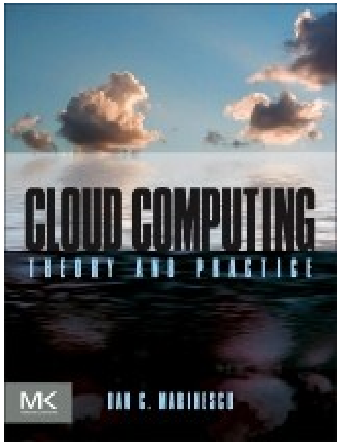
- Such events can affect the Internet domain name servers and prevent access to a cloud or can directly affect the clouds:
  - in 2004 an attack at Akamai caused a domain name outage and a major blackout that affected Google, Yahoo, and other sites.
  - in 2009, Google was the target of a denial of service attack which took down Google News and Gmail;
  - in 2012 lightning caused a prolonged down time at Amazon.

Credits, references and reading material

- *Cloud Computing: Theory and Practice*

  Dan C. Marinescu, 2013

  Chapters 1, 2



- *Cloud Computing: Principles and Paradigms*

  R. Buyya, J. Broberg, and A. Goscinski

  Wiley Press, 2011

- *Distributed and Cloud Computing*

  K. Hwang, G. Fox and J. Dongarra

  Morgan Kaufmann, 2012