



고려대학교
KOREA UNIVERSITY

KU-The Future

Cloud computing – Chapter 3

Heonchang Yu

Distributed and Cloud Computing Lab.

Concept of Cloud Computing

- **John McCarthy [1961]**

- “If computers of the kind I have advocated become the computers of the future, then computing may someday be organized as a public utility just as the telephone system is a public utility. ... The computer utility could become the basis of a new and important industry.”

- **Leonard Kleinrock [1969]**

- “As of now, computer networks are still in their infancy, but as they grow up and become sophisticated, we will probably see the spread of ‘computer utilities’ ...”.

Concept of Cloud Computing

- **Definition**

- **[Gartner report]** "...a style of computing in which scalable and elastic IT-enabled capabilities are delivered as a service to external customers using Internet technologies."
- **[Forrester Research]** "...a standardized IT capability (services, software, or infrastructure) delivered via Internet technologies in a pay-per-use, self-service way."
- **[Textbook]** "Cloud computing is a specialized form of distributed computing that introduces utilization models for remotely provisioning scalable and measured resources."

Concept of Cloud Computing

- **Definition**

- [NIST] “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. This cloud model is composed of five essential characteristics, three service models, and four deployment models.”

- **“Cloud computing” in commercial arena [2006]**

- Amazon : Elastic Compute Cloud(EC2) service
- Google Apps : browser-based enterprise applications

Basic Concepts and Terminology

- **Cloud**

- ✓ A distinct IT environment that is designed for the purpose of remotely provisioning scalable and measured IT resources
- ✓ Originated as a metaphor for the Internet which is, in essence, a network of networks providing remote access to a set of decentralized IT resources
- ❖ Whereas the **Internet** provides open access to many Web-based IT resources, a **cloud** is typically privately owned and offers access to IT resources that is metered.

Figure 3.1 The symbol used to denote the boundary of a cloud environment



Copyright © Arcitura Education

Basic Concepts and Terminology

- **Scaling**

- ✓ The ability of the IT resource to handle increased or decreased usage demands

- Horizontal Scaling – scaling out and scaling in

- Vertical Scaling – scaling up and scaling down

- **Horizontal Scaling**

- ✓ Allocating or releasing of IT resources that are of the same type

- ✓ The horizontal allocation of resources is referred to as **scaling out** and the horizontal releasing of resources is referred to as **scaling in**.

Basic Concepts and Terminology

- Horizontal Scaling

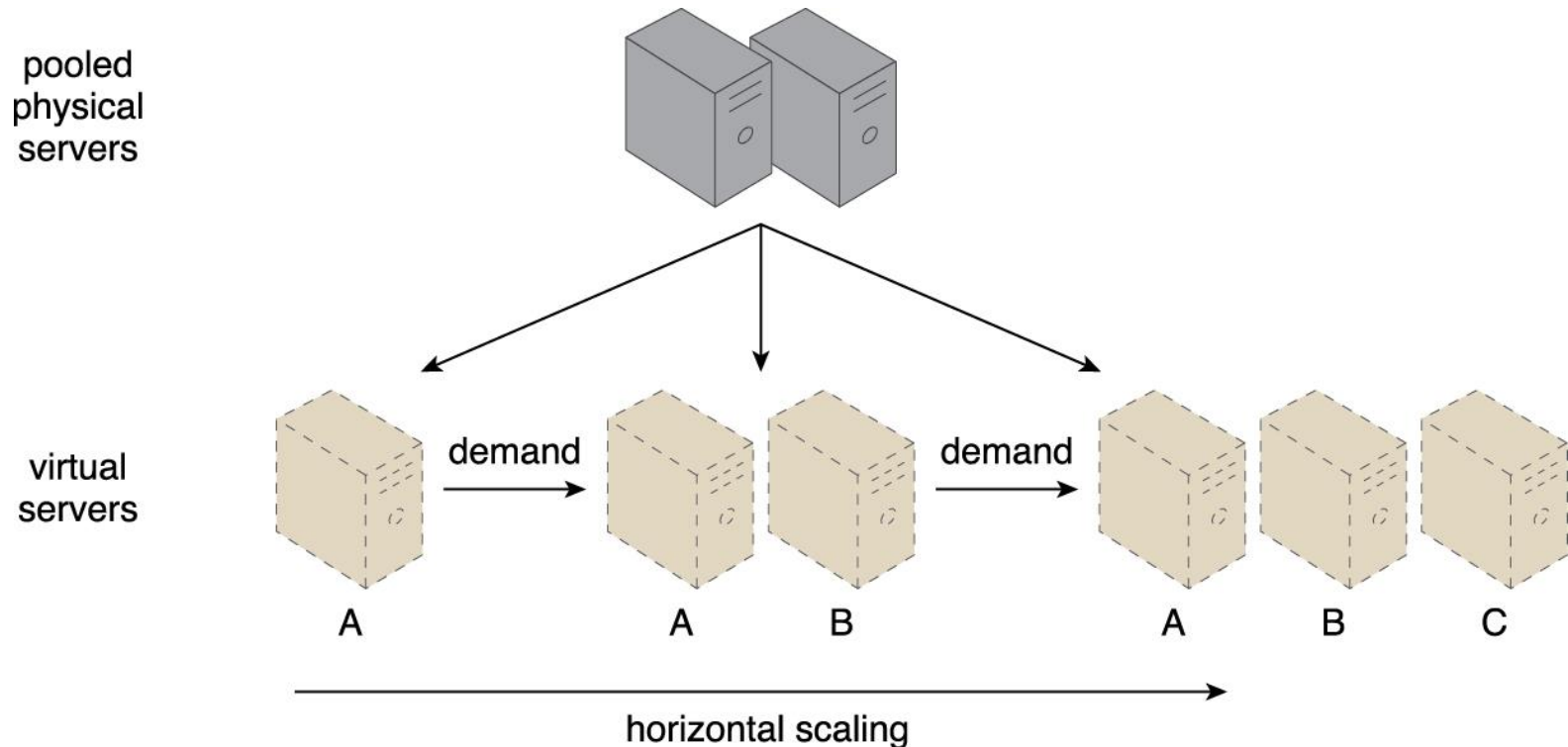
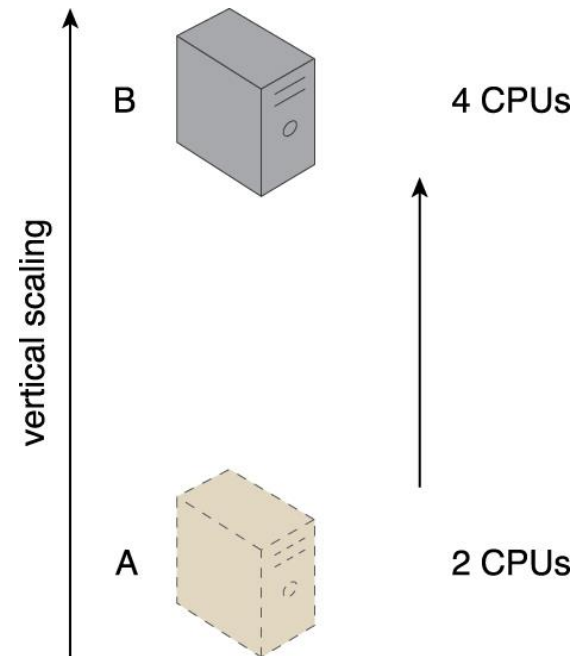


Figure 3.4 An IT resource (Virtual Server A) is scaled out by adding more of the same IT resources (Virtual Servers B and C)

Basic Concepts and Terminology

- Vertical Scaling
 - ✓ replacing by another with higher or lower capacity
 - ✓ The replacing of an IT resource with another that has a higher capacity is referred to as **scaling up** and the replacing an IT resource with another that has a lower capacity is considered **scaling down**.

Figure 3.5 An IT resource (a virtual server with two CPUs) is scaled up by replacing it with a more powerful IT resource with increased capacity for data storage (a physical server with four CPUs)



Basic Concepts and Terminology

Table 3.1 A comparison of horizontal and vertical scaling

Horizontal Scaling	Vertical Scaling
less expensive (through commodity hardware components)	more expensive (specialized servers)
IT resources instantly available	IT resources normally instantly available
resource replication and automated scaling	additional setup is normally needed
additional IT resources needed	no additional IT resources needed
not limited by hardware capacity	limited by maximum hardware capacity

Basic Concepts and Terminology

- **Cloud Service**

- ✓ Any IT resource that is made remotely accessible via a cloud

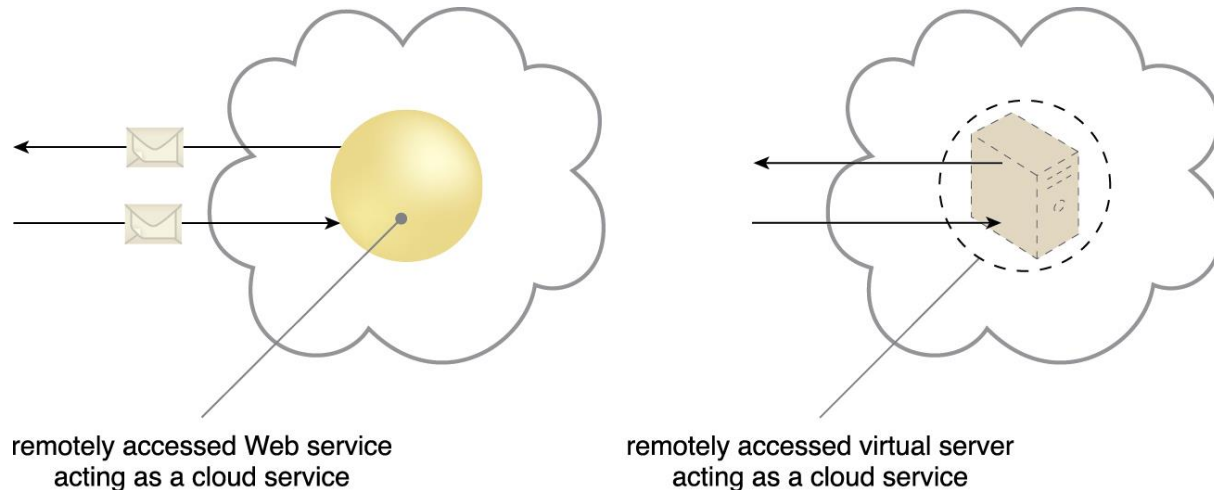


Figure 3.6 A cloud service with a published technical interface is being accessed by a consumer outside of the cloud (left). A cloud service that exists as a virtual server is also being accessed from outside of the cloud's boundary (right). The cloud service on the left is likely being invoked by a consumer program that was designed to access the cloud service's published technical interface. The cloud service on the right may be accessed by a human user that has remotely logged on to the virtual server.

Goals and Benefits

– Reduced Investments and Proportional Costs

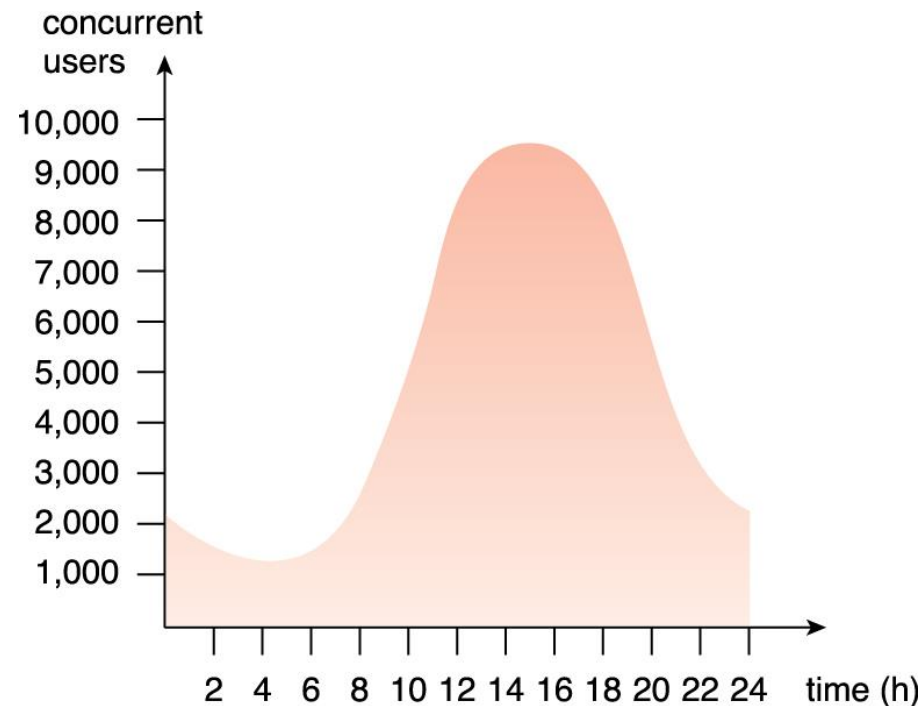
- Public cloud providers base their business model on the mass-acquisition of IT resources
- The reduction or outright elimination of up-front IT investments, namely hardware and software purchases and ownership costs
- Benefits
 - ✓ On-demand access to pay-as-you-go computing resources on a short-term basis (such as processors by the hour), and the ability to release these computing resources when they are no longer needed
 - ✓ The perception of having unlimited computing resources that are available on demand, thereby reducing the need to prepare for provisioning
 - ✓ The ability to add or remove IT resources at a fine-grained level, such as modifying available storage disk space by single gigabyte increments
 - ✓ Abstraction of the infrastructure so applications are not locked into devices or locations and can be easily moved if needed

Goals and Benefits

– Increased Scalability

- By providing pools of IT resources, clouds can instantly and dynamically allocate IT resources to cloud consumers, on-demand or via the cloud consumer's direct configuration
- The ability of IT resources to always meet and fulfill unpredictable usage demands avoids potential loss of business that can occur when usage thresholds are met

Figure 3.8 An example of an organization's changing demand for an IT resource over the course of a day



Goals and Benefits

– Increased Availability and Reliability

- An IT resource with increased availability is accessible for longer periods of time (for example, 22 hours out of a 24 hour day). Cloud providers generally offer “resilient” IT resources for which they are able to guarantee high levels of availability.
- An IT resource with increased reliability is able to better avoid and recover from exception conditions. The modular architecture of cloud environments provides extensive failover support that increases reliability.

Risks and Challenges

– Increased Security Vulnerabilities

- The remote usage of IT resources requires an expansion of trust boundaries by the cloud consumer to include the external cloud.
 - ⇒ It can be difficult to establish a security architecture that spans such a trust boundary without introducing vulnerabilities
- There can be overlapping trust boundaries from different cloud consumers due to the fact that cloud-based IT resources are commonly shared.
 - ⇒ The overlapping of trust boundaries and the increased exposure of data can provide malicious cloud consumers (human and automated) with greater opportunities to attack IT resources and steal or damage business data

Risks and Challenges

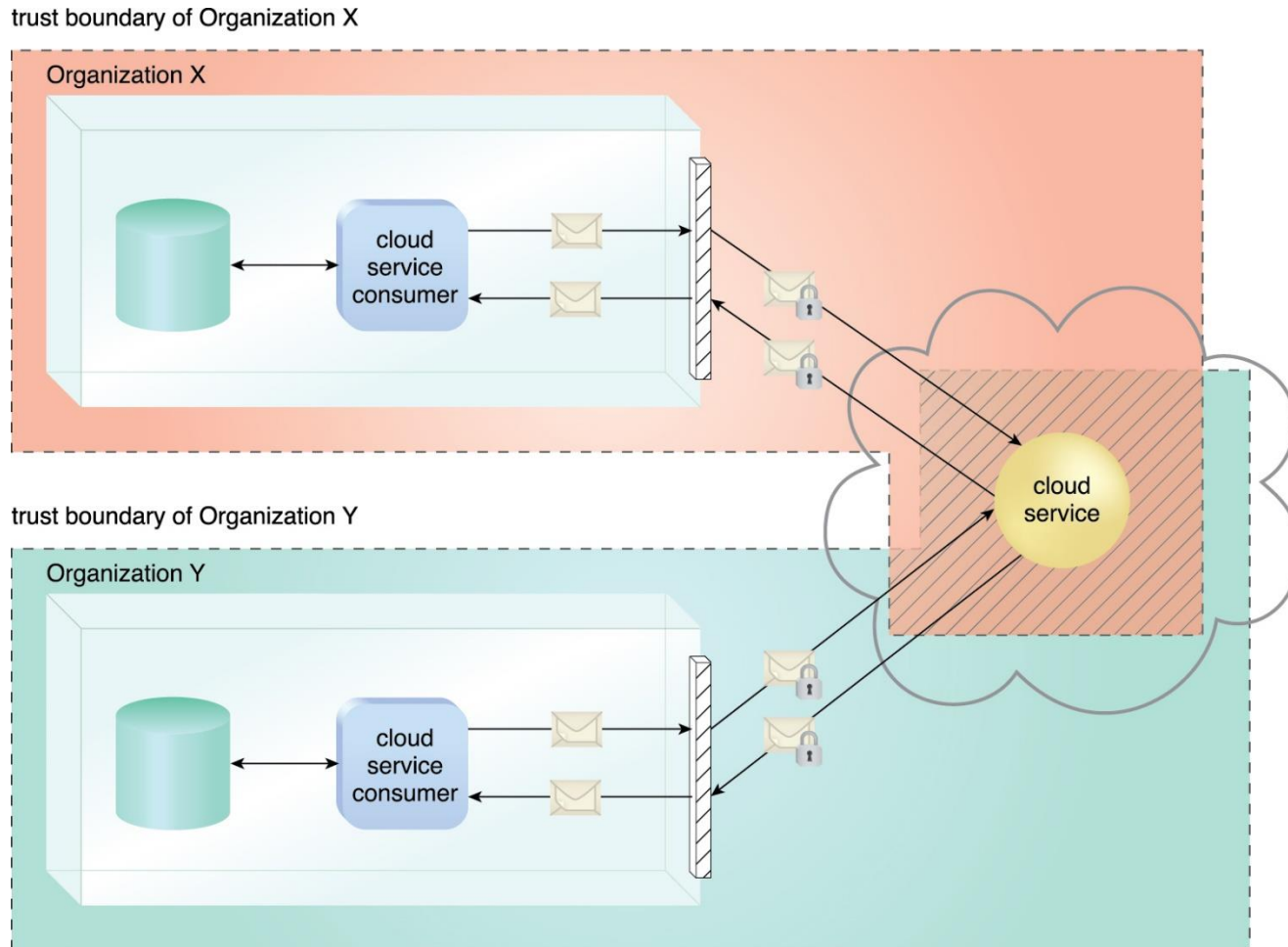


Figure 3.9 The shaded area with diagonal lines indicates the overlap of two organizations' trust boundaries

Risks and Challenges

- Reduced Operational Governance Control
 - Cloud consumers are usually allotted a level of governance control that is lower than that over on-premise IT resources.
 - Examples
 - An unreliable cloud provider may not maintain the guarantees it makes in the SLAs that were published for its cloud services. This can jeopardize the quality of the cloud consumer solutions that rely on these cloud services.
 - Longer geographic distances between the cloud consumer and cloud provider can require additional network hops that introduce fluctuating latency and potential bandwidth constraints.
 - » Figure 3.10

Risks and Challenges

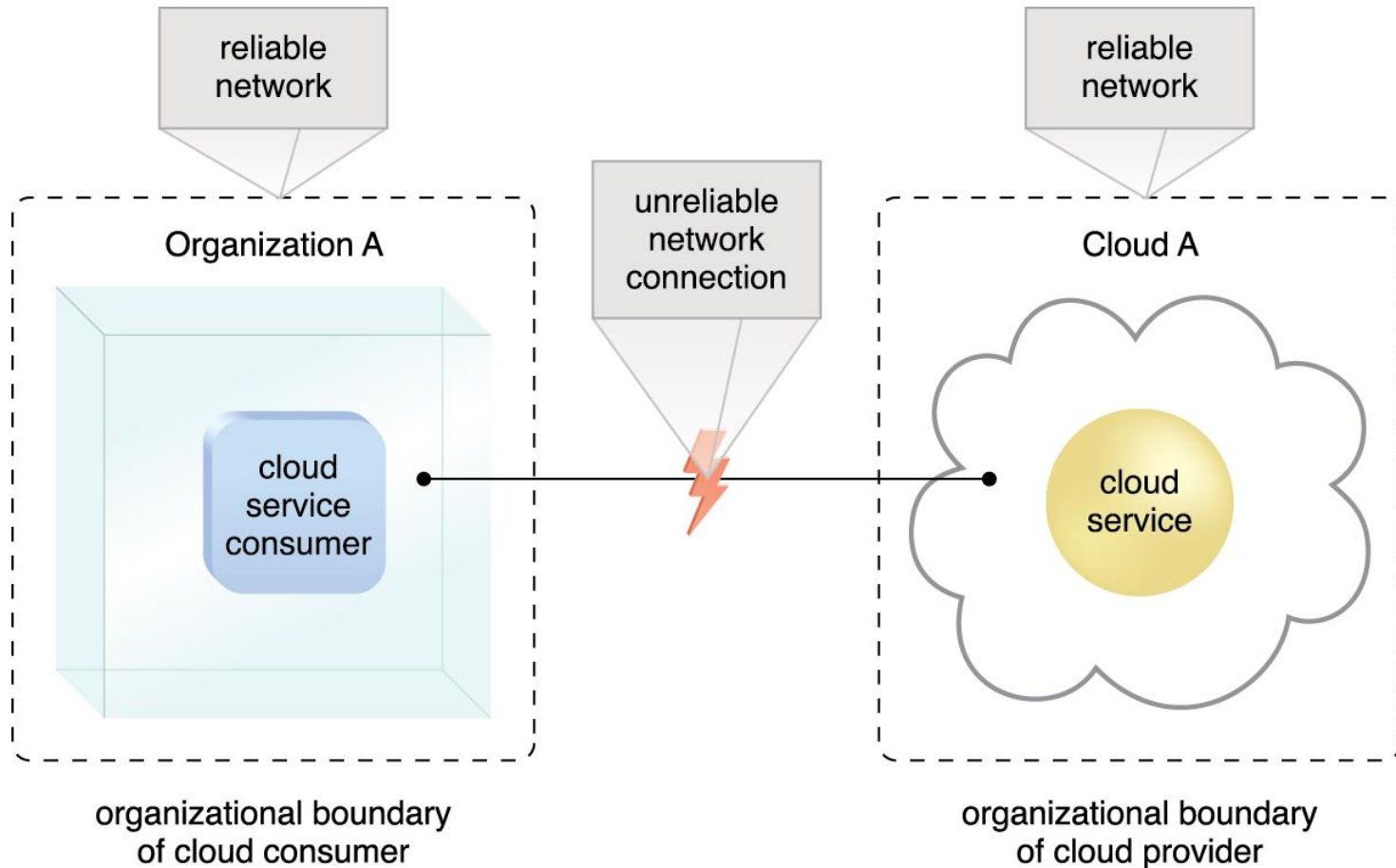


Figure 3.10 An unreliable network connection compromises the quality of communication between cloud consumer and cloud provider environments

Risks and Challenges

- Limited Portability Between Cloud Providers
 - Due to a lack of established industry standards within the cloud computing industry, public clouds are commonly proprietary to various extents
- Multi-Regional Compliance and Legal Issues
 - Third-party cloud providers will frequently establish data centers in affordable or convenient geographical locations.
 - ⇒ Serious legal concerns pertaining to industry or government regulations that specify data privacy and storage policies
 - Another potential legal issue pertains to the accessibility and disclosure of data.
 - ⇒ Countries have laws that require some types of data to be disclosed to certain government agencies or to the subject of the data.

Risks and Challenges

Figure 3.11 A cloud consumer's application has a decreased level of portability when assessing a potential migration from Cloud A to Cloud B, because the cloud provider of Cloud B does not support the same security technologies as Cloud A.

