Chapter 1 – Introduction

Contents

- Network-centric computing and network-centric content.
- Cloud computing.
- Delivery models and services.
- Ethical issues in cloud computing.
- Cloud vulnerabilities.

Network-centric computing

- Information processing can be done more efficiently on large farms of computing and storage systems accessible via the Internet.
  - Grid computing – initiated by the National Labs in the early 1990s; targeted primarily at scientific computing.
  - Utility computing – initiated in 2005-2006 by IT companies and targeted at enterprise computing.
- The focus of utility computing is on the business model for providing computing services; it often requires a cloud-like infrastructure.
- Cloud computing is a path to utility computing embraced by major IT companies including: Amazon, HP, IBM, Microsoft, Oracle, and others.

Network-centric content

- Content: any type or volume of media, be it static or dynamic, monolithic or modular, live or stored, produced by aggregation, or mixed.
- The “Future Internet” will be content-centric.

  The creation and consumption of audio and visual content is likely to transform the Internet to support increased quality in terms of resolution, frame rate, color depth, stereoscopic information.

Network-centric computing and content

- Data-intensive: large scale simulations in science and engineering require large volumes of data. Multimedia streaming transfers large volume of data.
- Network-intensive: transferring large volumes of data requires high bandwidth networks.
- Low-latency networks for data streaming, parallel computing, computation steering.
- The systems are accessed using *thin clients* running on systems with limited resources, e.g., wireless devices such as smart phones and tablets.
- The infrastructure should support some form of workflow management.

Evolution of concepts and technologies

- The concepts and technologies for network-centric computing and content evolved along the years.
  - The web and the semantic web - expected to support composition of services. The web is dominated by unstructured or semi-structured data, while the semantic web advocates inclusion of sematic content in web pages.
  - The Grid - initiated in the early 1990s by National Laboratories and Universities; used primarily for applications in the area of science and engineering.
  - Peer-to-peer systems.
  - Computer clouds.

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*.

- 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?

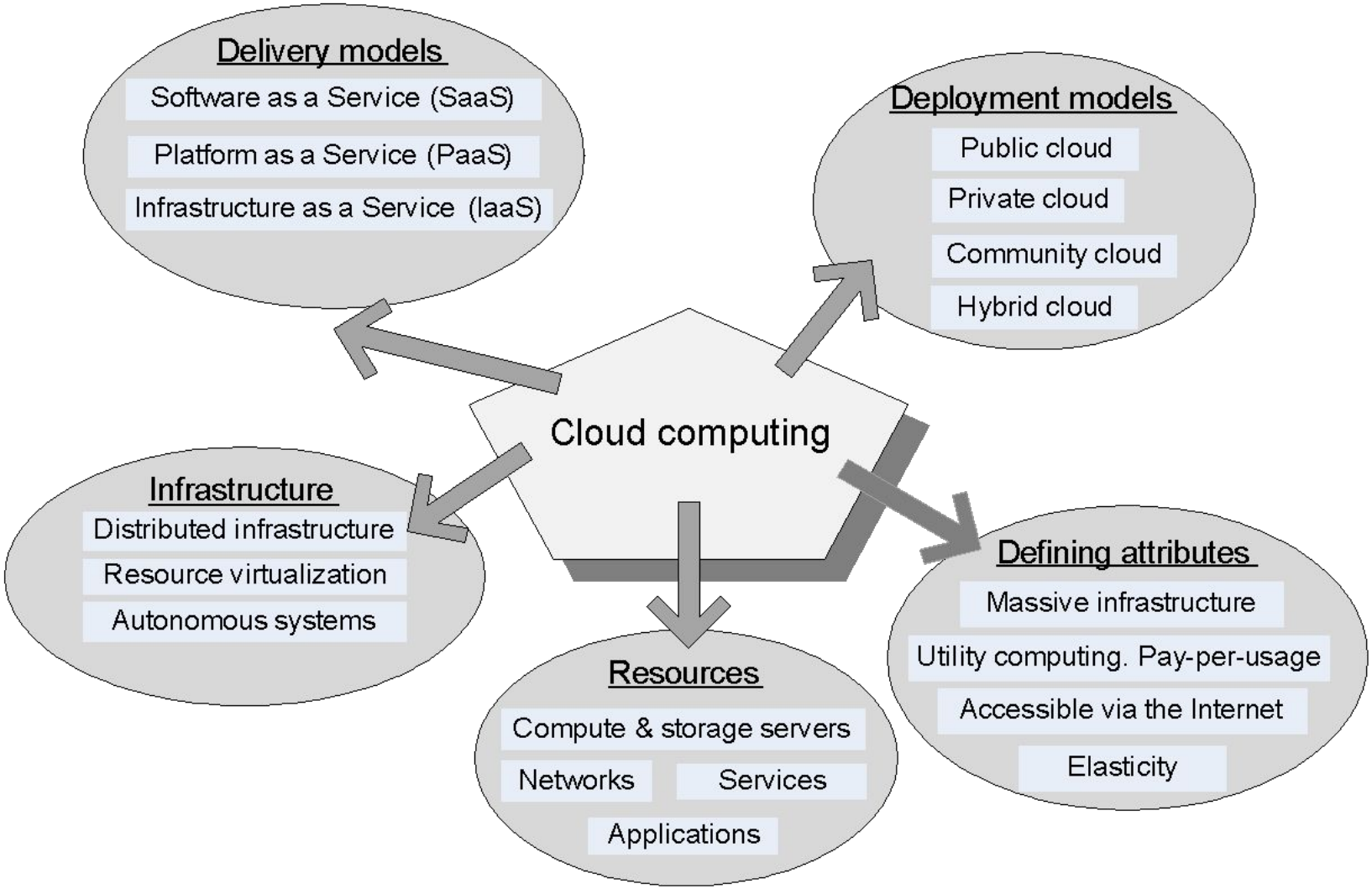
- It is in a better position to exploit recent advances in software, networking, storage, and processor technologies promoted by the same companies who provide cloud services.
- It is focused on enterprise computing; its adoption by industrial organizations, financial institutions, government, and so 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.

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!!


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
Cloud computing
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

Cloud delivery models

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

Software-as-a-Service (SaaS)

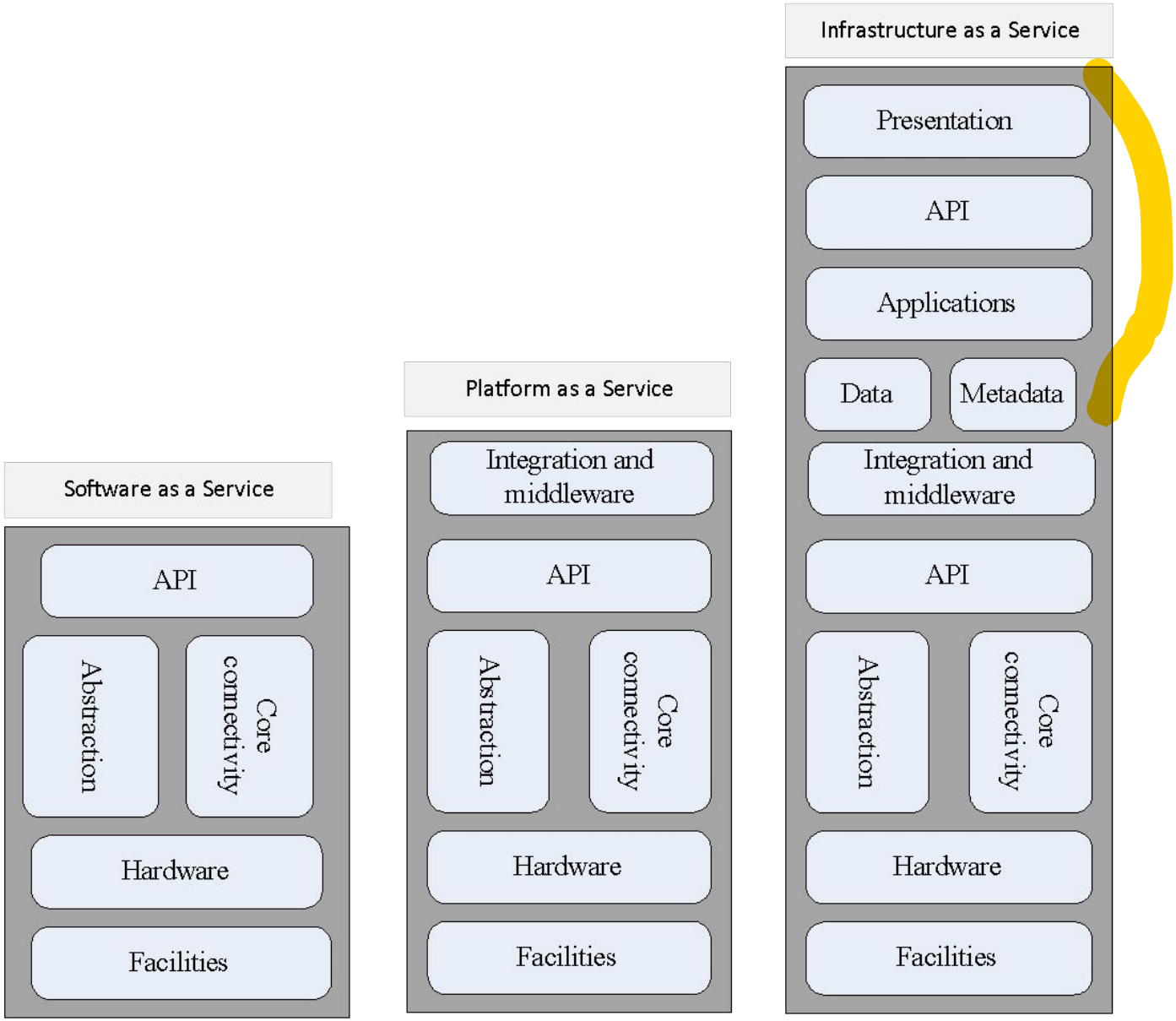
- Applications are supplied by the service provider.
- The user does not manage or control the underlying cloud infrastructure or individual application capabilities.
- Services offered include:
  - Enterprise services such as: workflow management, group-ware and collaborative, supply chain, communications, 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.


Infrastructure as a Service
Presentation
API
Applications
Data
Metadata
Integration and middleware
API
Abstraction
Core connectivity
Hardware
Facilities
Platform as a Service
Integration and middleware
API
Abstraction
Core connectivity
Hardware
Facilities
Software as a Service
API
Abstraction
Core connectivity
Hardware
Facilities

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.

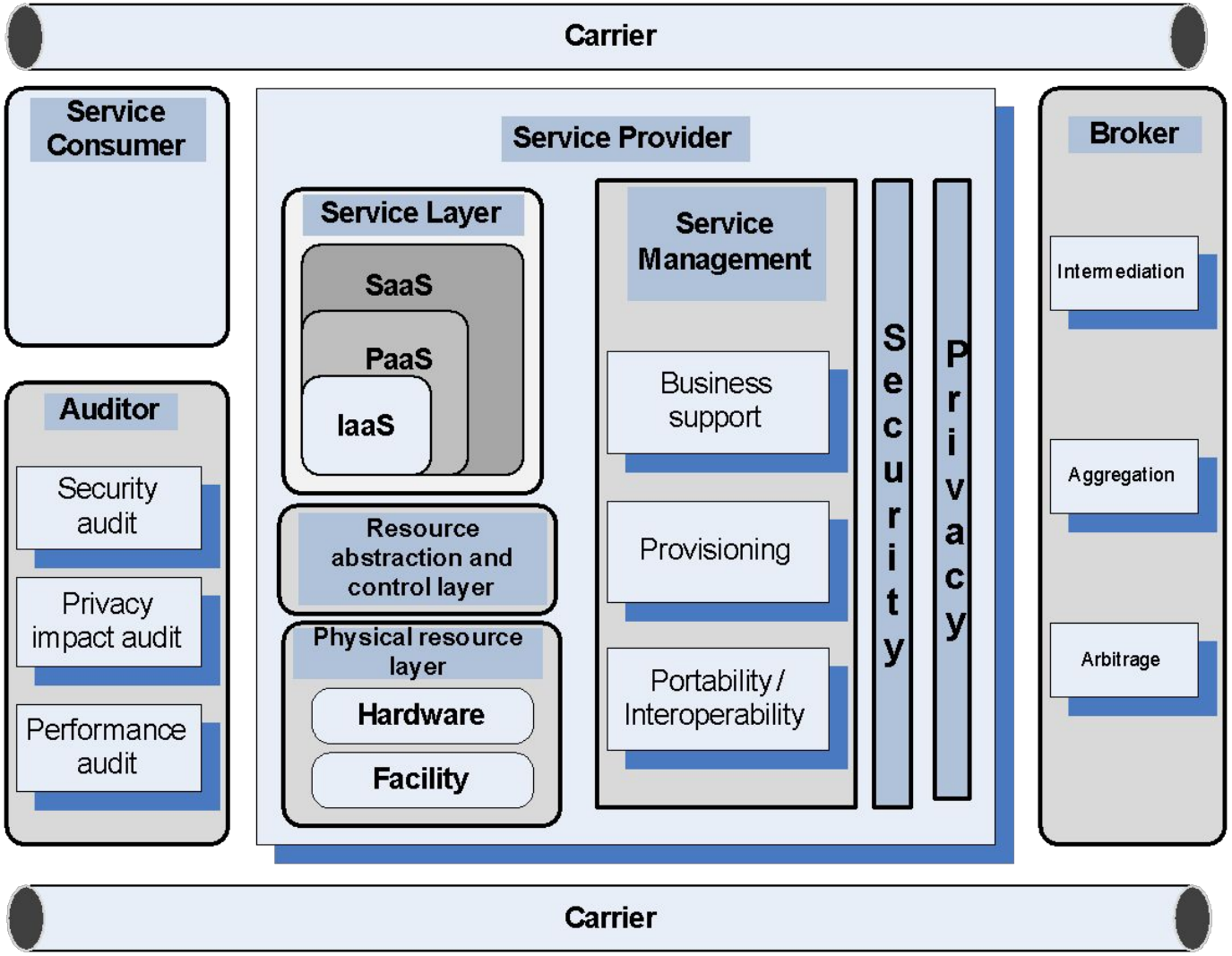
Cloud activities (cont'd)

- 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.

NIST cloud reference model



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 for several days;
  - in 2012 lightning caused a prolonged down time at Amazon.