Introduction to Cloud Computing

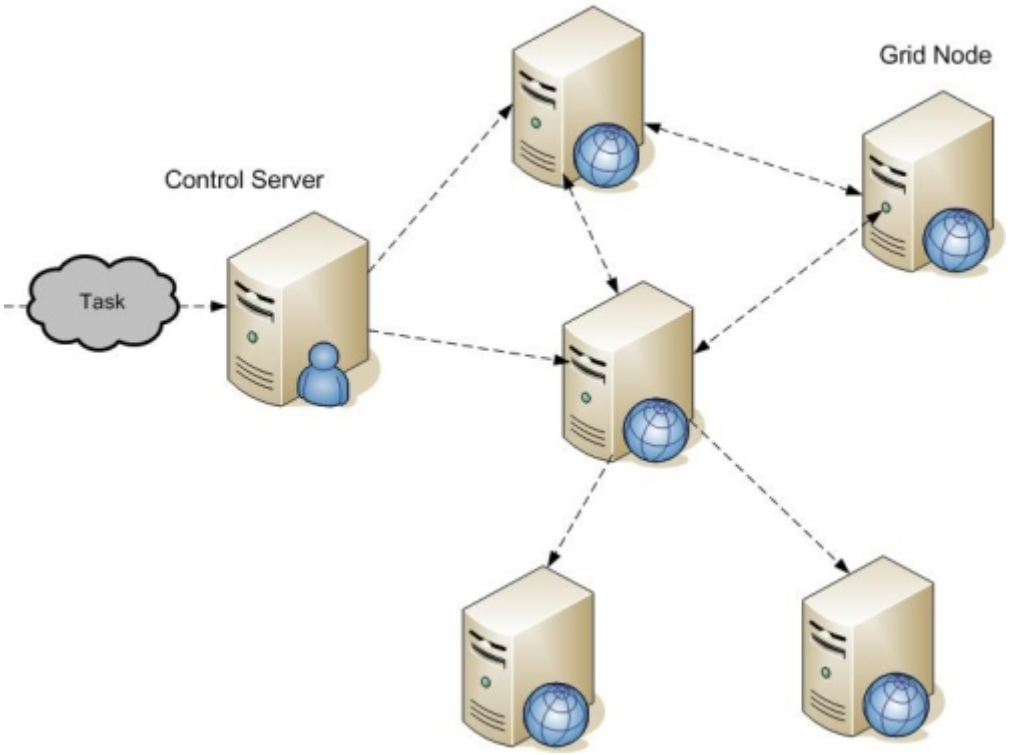
Grid Computing

Def

- *combination of computer resources from multiple administrative domains applied to a common task**

Core idea

- distributed parallel computation
  - super virtual computer



* http://en.wikipedia.org/wiki/Grid_computing

Utility Computing

Def

- *"The packaging of computing resources (computation, storage etc.) as a metered service similar to a traditional public utility"**

Observation

- not a new concept
  - *"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."* - *John McCarthy, MIT Centennial in 1961*

University of Waterloo

* http://en.wikipedia.org/wiki/Utility_computing

Cloud Computing

Is cloud computing?

- grid computing + utility computing ??
- difficult to define
  - means different things to different parties

Various definitions

- NIST – National Institute of Standards and Technology
  - "*universally*" accepted definition

Cloud Computing – NIST

Definition

- *"Cloud computing is a model for enabling 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 promotes availability and is composed of five essential characteristics, three service models, and four deployment models."**

University of Waterloo

* http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc

Cloud Computing – NIST

Definition

- *"Cloud computing is a model for enabling 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 promotes availability and is composed of five essential characteristics, three service models, and four deployment models."**

University of Waterloo

* http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc

Cloud Computing – NIST

Definition

- *"Cloud computing is a model for enabling 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 promotes availability and is composed of five essential characteristics, three service models, and four deployment models."**

University of Waterloo

* http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc

Cloud Computing – NIST

Definition

- *"Cloud computing is a model for enabling 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 promotes availability and is composed of five essential characteristics, three service models, and four deployment models."**

University of Waterloo

* http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc

Cloud Computing – NIST

Definition

- *"Cloud computing is a model for enabling 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 promotes availability and is composed of five essential characteristics, three service models, and four deployment models."**

University of Waterloo

* http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc

Cloud Computing – NIST

Definition

- *"Cloud computing is a model for enabling 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 promotes availability and is composed of five essential characteristics, three service models, and four deployment models."**

University of Waterloo

* http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc

NIST Essential Characteristics

On-demand self-service

- a consumer can *unilaterally* provision computing capabilities *without human interaction* with the service provider
- computing capabilities
  - server time, network storage, number of servers etc.

NIST Essential Characteristics

Broad network access

- capabilities are
  - available over the network
  - accessed through standard mechanisms
- promote use by
  - heterogeneous thin or thick client platforms

NIST Essential Characteristics

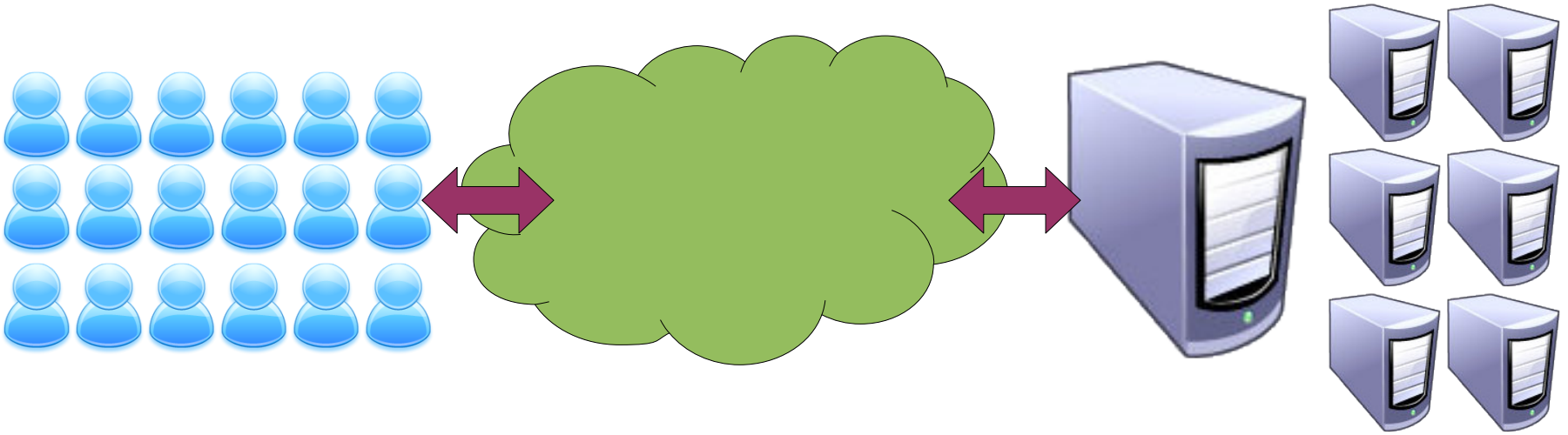
Multi-tenancy / Resource pooling

- provider's computing resources are pooled to serve multiple consumers
- computing resources
  - storage, processing, memory, network bandwidth and virtual machines
- location independence
  - no control over the exact location of the resources
- has major implications
  - performance, scalability, security

NIST Essential Characteristics

Rapid elasticity

- capabilities can be rapidly and elastically provisioned
- unlimited virtual resources
- predicting a ceiling is difficult



University of Waterloo

NIST Essential Characteristics

Measured service

- metering capability of service/resource abstractions
  - storage
  - processing
  - bandwidth
  - active user accounts
- **OK so what happened to utility computing – pay as you go model??**
  - more on this later when we discuss deployment models

University of Waterloo

Relevant Technologies

Access

- heterogeneous set of thick & thin clients
  - PCs (enterprise, home), mobile devices, hand-held devices
- high speed broadband access
  - wired & wireless
- data centres
  - large computing capacity
  - distributed
  - direct access storage devices Vs. storage area networks

Relevant Technologies

Virtualization

- decoupling from the physical computing resources

Virtualization types

- hardware
  - emulation – VM emulates/simulates complete hardware
    - QEMU
  - paravirtualization - software interface to virtual machines
    - Xen
  - full virtualization - complete simulation of the underlying hardware
    - VMWare, Parallels

Relevant Technologies

Virtualization types

- memory virtualization
  - decouples volatile random access memory (RAM) resources from individual systems
  - aggregates these resources into a virtualized memory pool available to any computer in the cluster
- storage virtualization
  - abstracting logical storage from physical storage
  - NAS - network attached storage
- data virtualization
  - data as an abstract layer, independent of underlying database systems, structures and storage

Relevant Technologies

Virtualization types

- network virtualization
  - virtualized network addressing space within or across network subnets
  - VPNs

Question?

- how do we measure virtual resources
  - Amazon ECU (elastic compute unit)
    - EC2 Compute Unit equals
      - 1.0-1.2 GHz 2007 Opteron or
      - 2007 Xeon processor

Relevant Technologies

APIs

- required for various operations and applications
  - administration
  - application development
  - resource migration
- no standards

SPI Services

SaaS (Software-as-a-Service)

- vendor/provider controlled applications accessed over the network
- characteristics
  - network based access
  - multi-tenancy
  - single software release for all

SaaS Examples

  - Salesforce.com, Google Docs

SPI Services

SaaS & Multi-tenancy

- SaaS applications are multi-tenant applications
- application data
  - Google docs

SaaS Application Design

- SaaS applications are 'net native'
- configurability, efficiency, and scalability
- SOA & SaaS

Net Native Application

Characteristics

- cloud specific design, development & deployment
- multi-tenant data
- builtin metering & management
- browser based client & client tools
- customization via configuration

SPI Services

SaaS Disadvantages

- dependency on
  - network, cloud service provider
- performance
  - limited client bandwidth
- security
  - good: better security than personal computers
  - bad: CSP is in charge of the data
  - ugly: user privacy

SPI Services

PaaS (Platform-as-a-Service)

- vendor provided development environment
  - tools & technology selected by vendor
  - control over data life-cycle

Advantages

- rapid development & deployment
- small startup cost
  - required skills set
  - money

SPI Services

PaaS – Architectural Characteristics

- multi-tenancy
  - data
- native scalability
  - load balancing & fail-over
- native integrated management
  - performance
  - resource consumption/utilization
  - load

SPI Services

PaaS Disadvantages

- inherits all from SaaS
- choice of development technology is limited to vendor provided/supported tools and services

PaaS Examples

- Google app engine
  - Google Site + Google Docs

SPI Services

IaaS (Infrastructure-as-a-Service)

- vendor provided and consumer provisioned computing resources
  - processing, storage, network, etc.
  - consumer is provided customized virtual machines
  - consumer has control over
    - OS, memory
    - storage
    - servers & deployment configurations
    - limited control over network resources

SPI Services

IaaS = utility computing??

- maybe – NIST does not talk about $$

Advantages

- infrastructure scalability
- native integrated management
  - performance, resource consumption/utilization, load
- economical cost
  - hardware, IT support

SPI Services

IaaS Examples

- Amazon Elastic Compute Cloud – EC2

SPI Services

less flexible
more constrained
less effort

more flexible
less constrained
more effort

Google docs

salesforce.com

Windows Azure

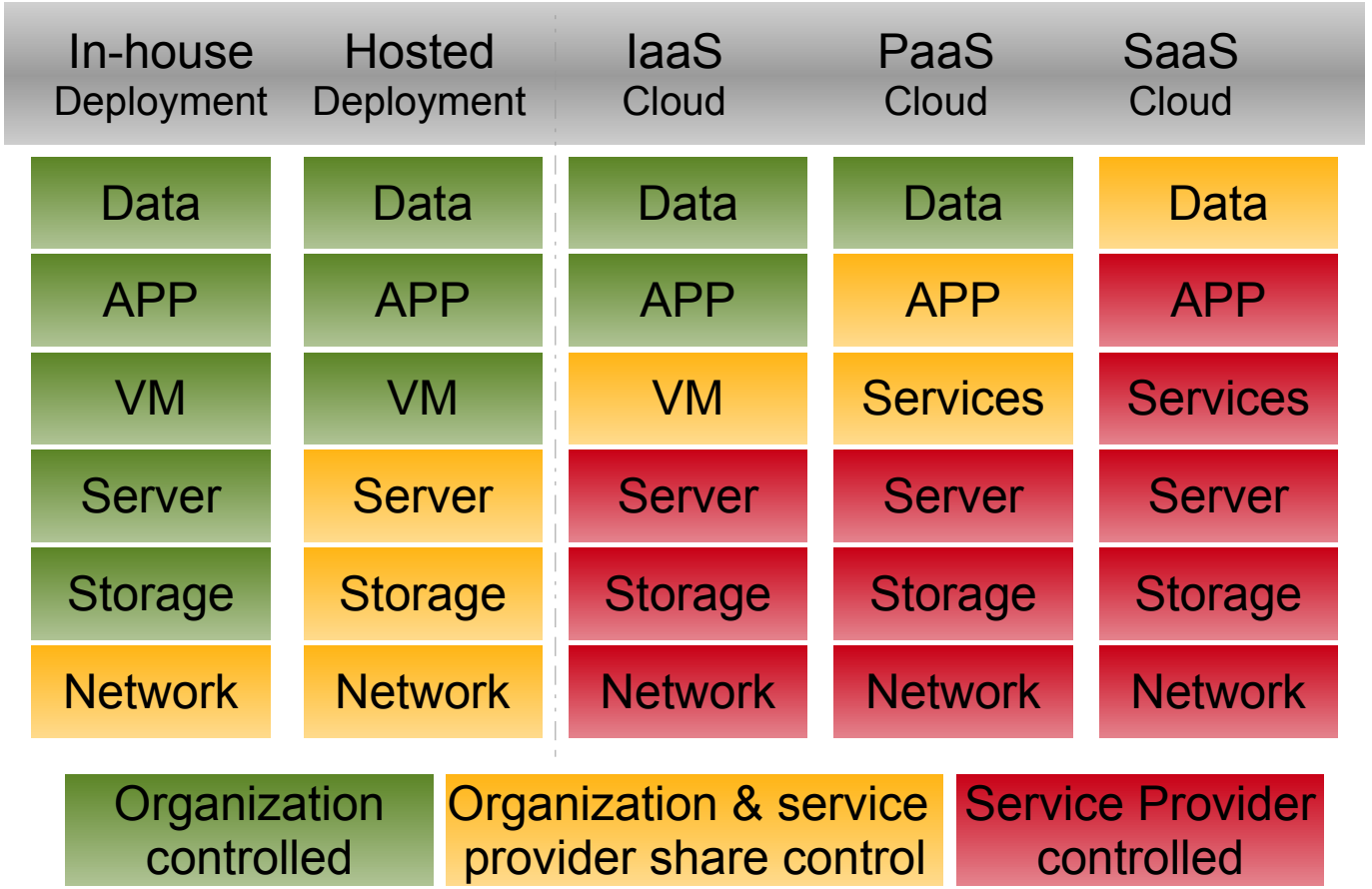
amazon web services

TurboTax

Eucalyptus

the rackspace cloud

SPI Services & Control



[1] **Visualizing the Boundaries of Control in the Cloud. Dec 2009.**
http://kscottmorrison.com/2009/12/01/visualizing-the-boundaries-of-control-in-the-cloud/

UNIVERSITY OF Waterloo

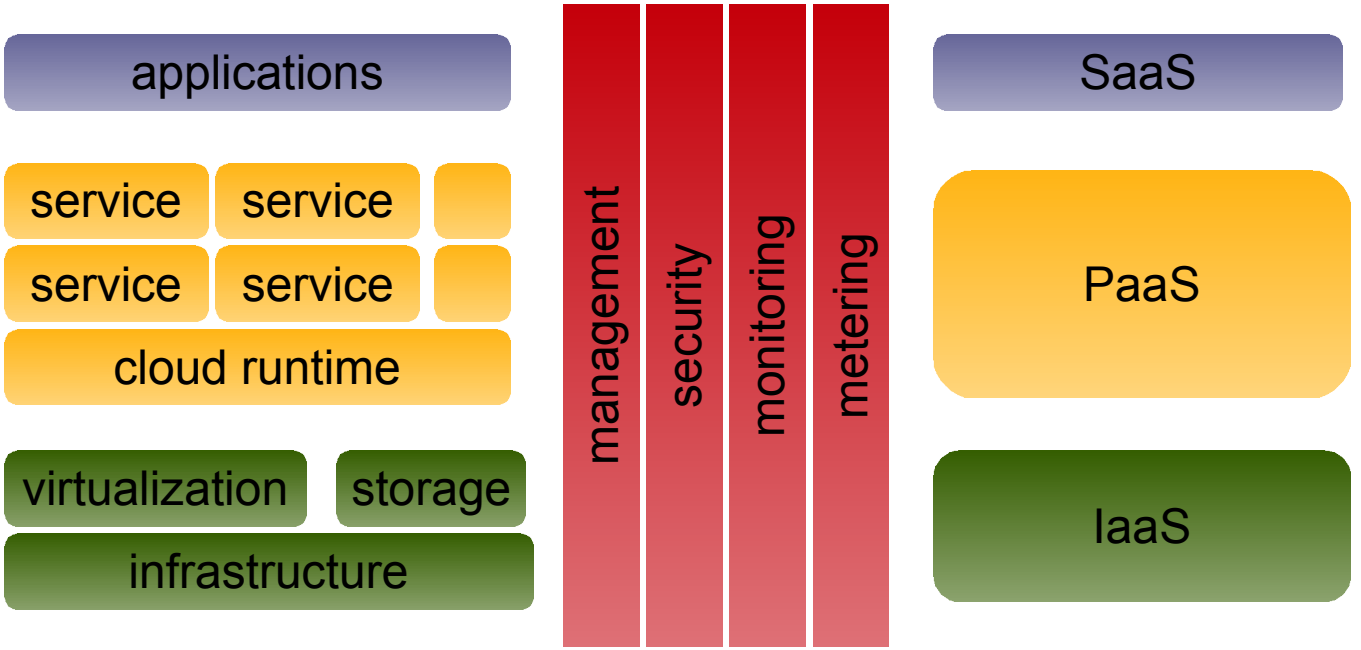
XaaS

XaaS (Everything-as-a-Service)

- composite second level services
  - Security-as-a-Service
    - McAfee*
      - McAfee SaaS Email Archiving
      - McAfee SaaS Email Inbound Filtering
      - McAfee Vulnerability Assessment SaaS (PEN Tests)
  - CaaS – Communication-as-a-Service
    - VoIP, private PBX

*http://www.mcafee.com/us/enterprise/products/hosted_security/

A Simple Reference Model



Amazon Web Services

Compute

Amazon Elastic Compute Cloud (EC2)

Amazon Elastic MapReduce

Auto Scaling

Content Delivery

Amazon CloudFront

Database

Amazon SimpleDB

Amazon Relational Database Service (RDS)

E-Commerce

Amazon Fulfillment Web Service (FWS)

Messaging

Amazon Simple Queue Service (SQS)

Amazon Simple Notification Service (SNS)

Monitoring

Amazon CloudWatch

Networking

Amazon Virtual Private Cloud (VPC)

Elastic Load Balancing

Payments & Billing

Amazon Flexible Payments Service (FPS)

Amazon DevPay

Storage

Amazon Simple Storage Service (S3)

Amazon Elastic Block Storage (EBS)

AWS Import/Export

Support

AWS Premium Support

Web Traffic

Alexa Web Information Service

Alexa Top Sites

Workforce

Amazon Mechanical Turk

http://aws.amazon.com/

NIST Cloud Deployment Models

4 Deployment Models

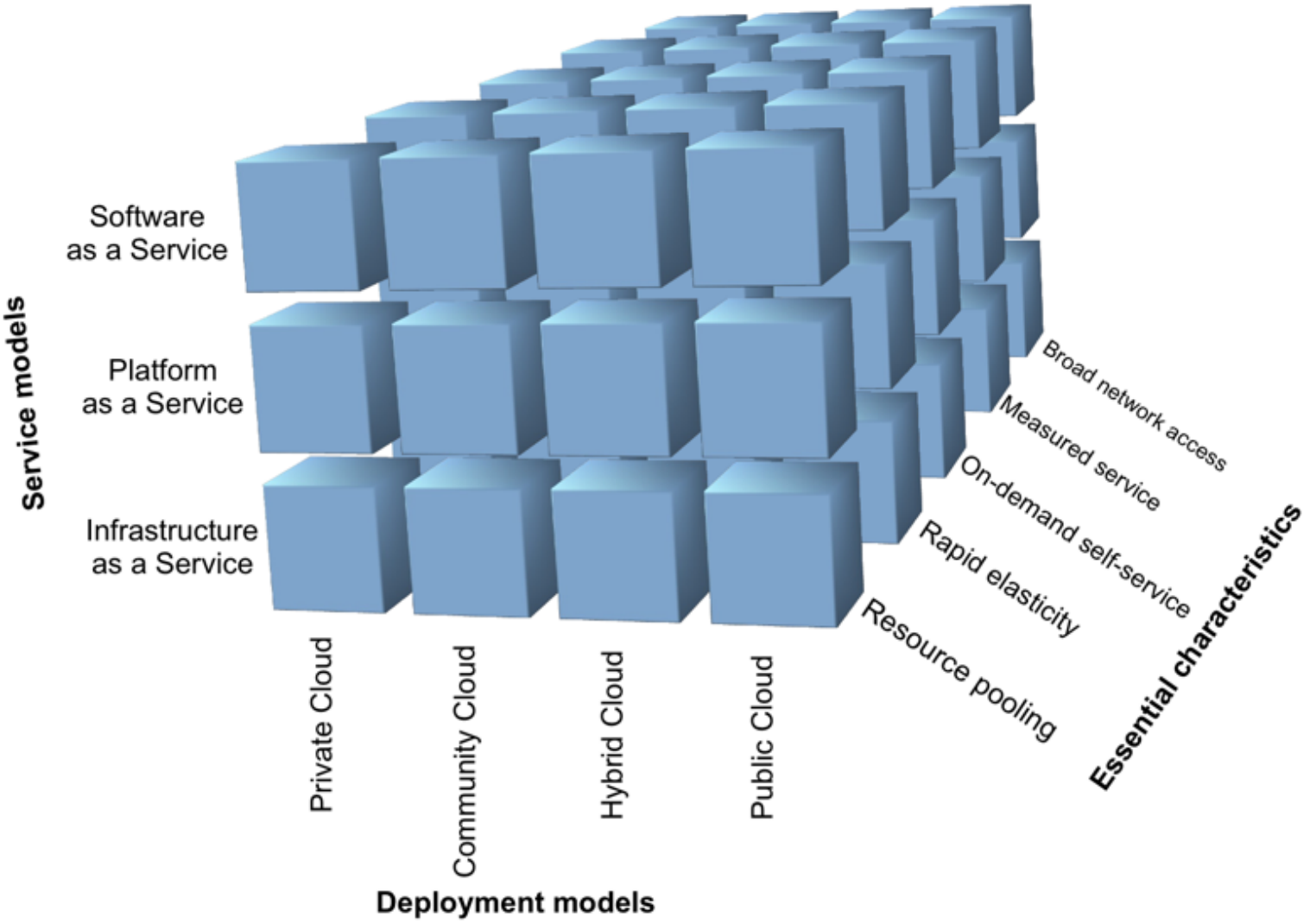
- private cloud
  - infrastructure is operated solely for an organization
  - managed by the organization or by a third party
- community cloud
  - supports a specific community
  - infrastructure is shared by several organizations

NIST Cloud Deployment Models

4 Deployment Models

- public cloud
  - infrastructure is made available to the general public
  - owned by an organization selling cloud services
- hybrid cloud
  - infrastructure is a composition of two or more clouds deployment models
  - enables data and application portability

NIST Cloud Computing

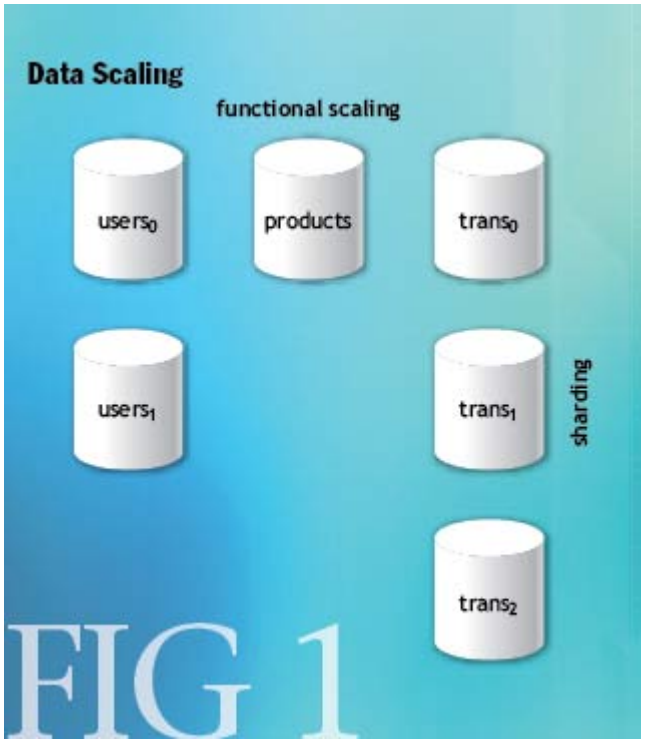


http://www.katescomment.com/images/CloudCube.png

Cloud Distributed Storage

Distributed Storage

- Two approaches to scaling
  - vertical – bigger hardware
  - horizontal – more hardware
    - functional partitioning
    - horizontal partitioning
      - sharding*



http://queue.acm.org/detail.cfm?id=1394128

*http://en.wikipedia.org/wiki/Shard_%28database_architecture%29

Cloud Distributed Storage

CAP Theorem*

- web services cannot ensure all three of the following properties at once
  - consistency
    - set of operations has occurred all at once
  - availability
    - an operation must terminate in an intended response
  - partition tolerance
    - operations will complete, even if individual components are unavailable

University of Waterloo

* Eric Brewer, University of California, Berkeley

Cloud Distributed Storage

Horizontal Storage Scaling

- *"any horizontal scaling strategy is based on data partitioning"**
  - forced to decide between consistency & availability

ACID

- provides strong data consistency guarantees
  - at the cost of availability
  - 2PC availability = product of availability of each

University of Waterloo

* http://queue.acm.org/detail.cfm?id=1394128

Cloud Distributed Storage

BASE – an ACID alternative

- basically available, soft state, **eventual consistency**
- characteristic
  - *"optimistic and accepts that the database consistency will be in a state of flux"**
  - supports partial failures
- scalability promise
  - *"leads to levels of scalability that cannot be obtained with ACID"**

University of Waterloo

* http://queue.acm.org/detail.cfm?id=1394128

Cloud Distributed Storage

Eventual Consistency

- consistency across functional groups is easy to relax
- we encounter this on daily basis
- some scenarios
  - update of online user profile
  - online master card payment
  - ATM cheque deposit
- **idempotent operations**
  - permit partial failures

Cloud Distributed Storage

General Characteristics

- simplified data model
- built on distributed file systems
  - GFS - Google File System
  - HDFS – Hadoop Distributed File System
- highly available
  - relaxed consistency
- fault-tolerant
  - replication

Cloud Distributed Storage

General Characteristics

- eventual consistency
  - all replicas will be updated at different times and in different order
- examples
  - Google BigTable
  - Yahoo PNUTS
  - Amazon S3

Cloud Distributed Computation

Motivation

- distributed computing
  - many thousands of computers
- large datasets
- fault-tolerant
- easy to configure & manage

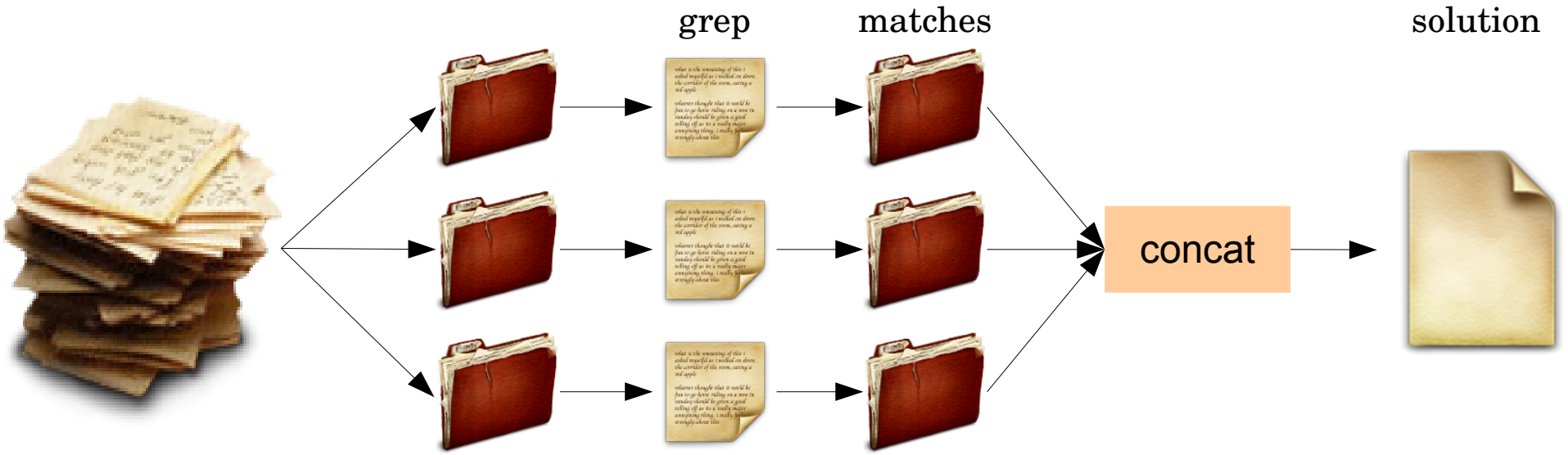
Cloud Distributed Computation

Basic Idea

- functional programming
- functional decomposition
  - large problem broken into a set of small problems
  - each small problem
    - can be solved by a functional transformation of input data
      - remember pipes & filters??
    - can be executed in complete isolation
      - parallel computing
- server (task) farm
  - to solve the big problem

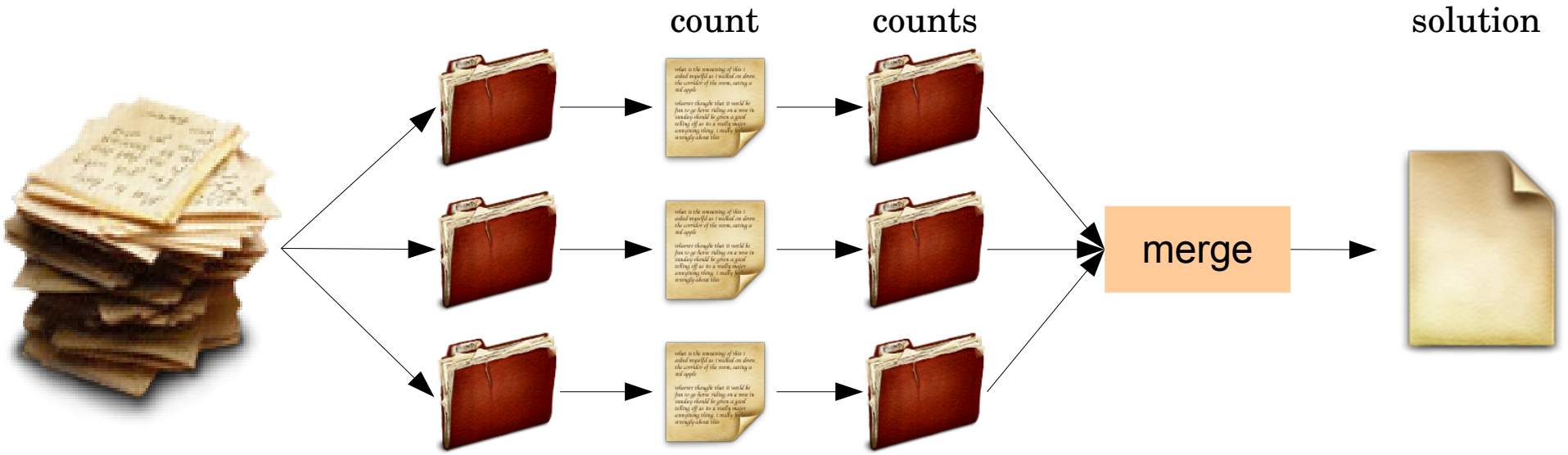
Cloud Distributed Computation

Distributed grep

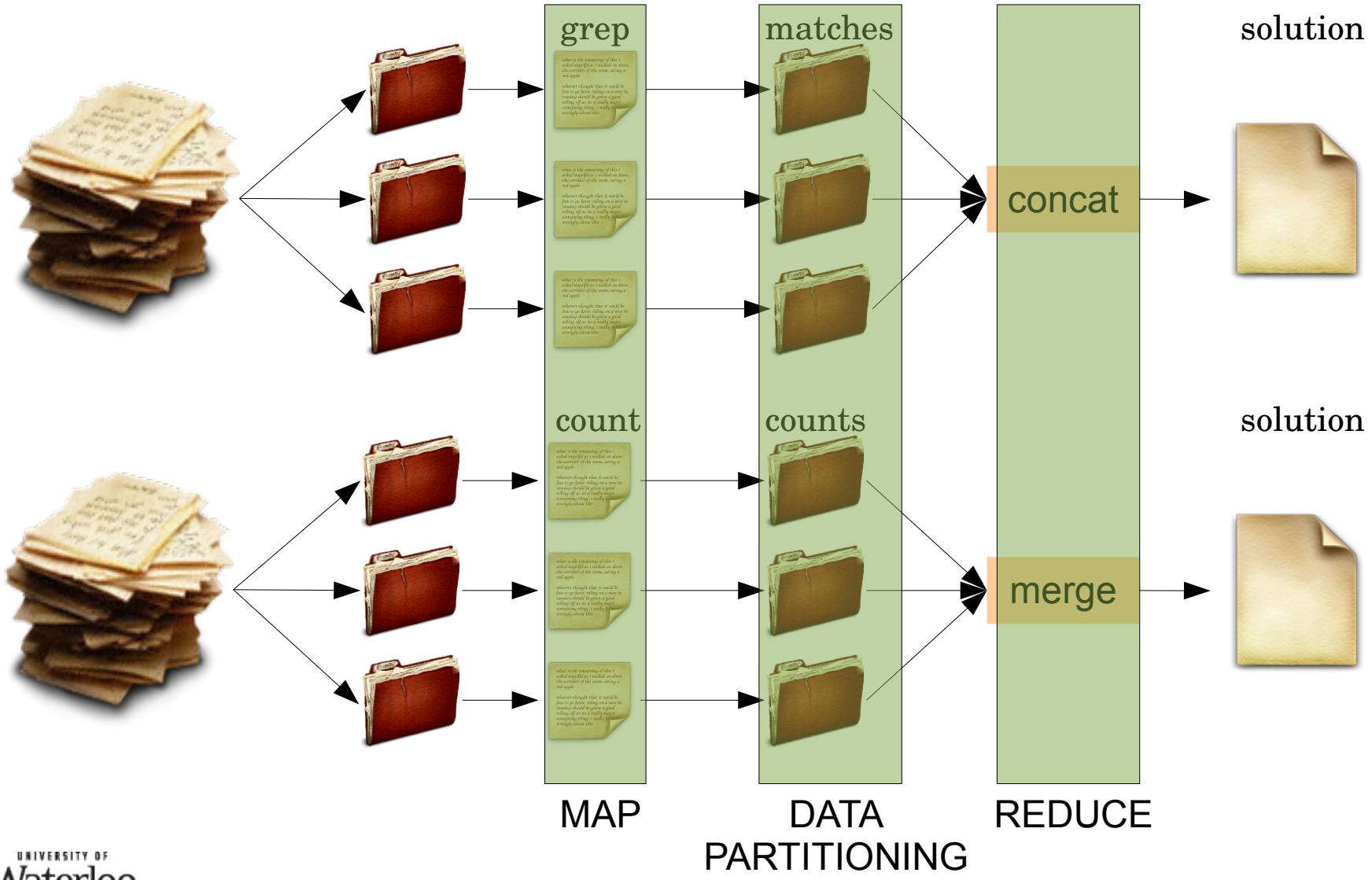


Cloud Distributed Computation

Distributed wc



MapReduce


grep
matches
concat
solution
count
counts
merge
solution
MAP
DATA PARTITIONING
REDUCE

MapReduce

Map

- input: key/value pair
- output: intermediate key/value pair

Reduce

- input: intermediate key/value pair
- output: final key/value pair

MapReduce

Examples

- distributed grep
  - map
    - if match(value,pattern) emit(value,1)
  - reduce
    - emit(key,sum(value*))
- distributed wc
  - map
    - for all w in value do emit(w,1)
  - reduce
    - emit(key,sum(value*))

Security in Cloud

Security

- Technology, provides assurance
  - confidentiality
  - integrity, authenticity

Privacy

- Right, provides control
  - anonymity
  - primary & secondary use

Information Security Concerns

Confidentiality

- safe from prying eyes
  - communication, persistence

Authenticity

- data is from a known source

Integrity

- data has not been tampered with
  - provenance (computation)
  - persistence

Information Security Concerns

Non-repudiation

- assurance against deniability

Access control

- access & modification by privileged users
  - individual vs. group access
  - multi-tenancy (PaaS, SaaS)

Information Security Concerns

Long term security

- change in authentication/authorization
- proof of possession
- confidentiality
  - crypto systems **do not** provide long term guarantees
- intersection attacks

Security Enhancing Techniques

Encryption

- symmetric encryption (*data*)
- public key cryptography (*identity, authentication*)
  - secret private key, published public key
- hash / Message Authentication Code (*integrity*)
- digital signatures (*authentication, non-repudiation*)
- TLS/SSL (*communication*)

Security Enhancing Techniques

Encryption

- homomorphic encryption*
  - allow for arbitrary computing over encrypted data
    - if E(p) = c then D(2c) = 2p (multiplication operation)
    - allows for data processing without decryption
  - promising but *not practical* so far**
- key management challenges
  - increase as the access control granularity increases

* Gentry, C. 2009. Fully homomorphic encryption using ideal lattices. In Proceedings of the 41st Annual ACM Symposium on theory of Computing (Bethesda, MD, USA, May 31 - June 02, 2009). STOC '09. ACM, New York, NY, 169-178.

** Bruce Schneier. Schneier on Security. http://www.schneier.com/blog/archives/2009/07/homomorphic_enc.html

Security Enhancing Techniques

Secure query & search

- PIR/SPIR (Private Information Retrieval)
  - *"allows a user to retrieve an item from the server without revealing the item to the database"**
  - under research
    - more effort required to be adopted by mainstream

* Chor, B., Kushilevitz, E., Goldreich, O., and Sudan, M. 1998. Private information retrieval. J. ACM 45, 6 (Nov. 1998), 965-981.

Security Enhancing Techniques

Secure query & search

- encrypted data search
  - matching with encrypted keywords
    - meta-data driven
    - single party query
  - secure anonymous database search (SADS)*
    - multi party queries
  - **not easy**, may require trusted third parties

* Raykova, M., Vo, B., Bellovin, S. M., and Malkin, T. 2009. Secure anonymous database search. In Proceedings of the 2009 ACM Workshop on Cloud Computing Security (Chicago, Illinois, USA, November 13 - 13, 2009). CCSW '09. ACM, New York, NY, 115-126.

Security Enhancing Techniques

Remote data checking

- client side preprocessing
  - data in chunks along with MAC for each chunk
  - server stores data chunk + MAC combinations
  - forward error correction
    - long term recoverability

Security Enhancing Techniques

Data Remanence

- "Residual representation of data after purge"
- How to purge data in cloud?
  - risk at all levels (SaaS, PaaS, and IaaS)
- Secure deletion
  - encrypt the data in the cloud
  - data deletion = key destruction

Security in Cloud

CSA (Cloud Service Alliance)

- http://www.cloudsecurityalliance.org/
- various introductory publications
  - CSA Guide ver 2.0
  - inline with NIST