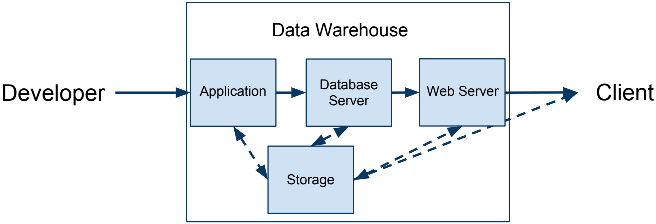
Data Security in the Cloud

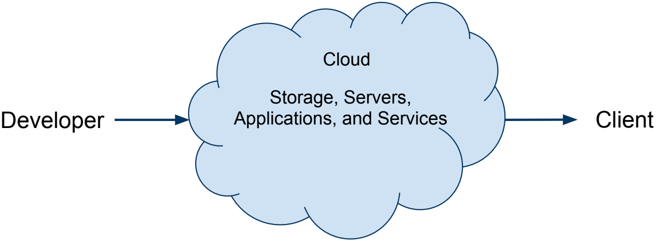
Matt Lauer

Computer Science Senior Seminar
University of Minnesota, Morris

April 30, 2011

Why Cloud Computing?





Cloud Providers

1. Introduction
   - Defining the Cloud
   - Cloud Services
2. Virtual Machine Security
   - Monitoring
   - Amazon EC2
3. Data Centric Security
   - Where it Matters
   - Secure Data Processing
4. Conclusion

Forming a Definition

Cloud computing is a buzz word

Keywords often associated with the cloud

- Virtualization
- Instant, on-demand scalability
- Pay-as-you-go service
- Parallel and distributed computing

Infrastructure as a Service

Virtual hardware available for users to run virtual machines
What is a virtual machine?

- Software implementation of a physical system
- Runs on top of existing hardware; alongside other software services
- There may be a host OS between the virtual machine manager and hardware
- Many VMs can operate simultaneously on powerful systems

Platform/Software as a Service

No hardware at this level

Platform: Tools for developers (Google App Engine)

Software: Tools for users (Google Docs, Maps)

1. Introduction
   - Defining the Cloud
   - Cloud Services
2. Virtual Machine Security
   - Monitoring
   - Amazon EC2
3. Data Centric Security
   - Where it Matters
   - Secure Data Processing
4. Conclusion

Overview of Virtualization

- Instant starting, stopping, and cloning of existing machines
- Isolation of services and applications allows for heightened security; VM only does one thing
- VMs operate in shared execution environment with other VMs
- Users setup all services and software, there are many doors potentially left open

Virtual Machine Monitoring

Cloud users do not have any access to hardware
Monitoring a VM “from the outside” is possible due to the nature of virtualization

- Cloud providers offer tools to monitor your VM
- The monitoring tools allow users to view performance of their VM and advanced tools will keep system healthy
- Advanced monitoring tools detect the guest OS on the VM and apply certain policies
- Also may detect system intrusions or anomalies

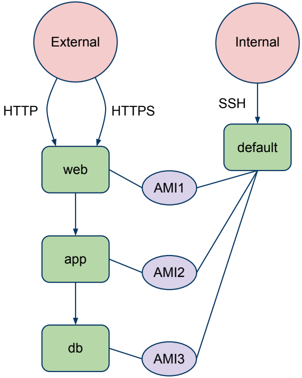
Overview

- Amazon Elastic Cloud Compute (EC2) provides VMs on Amazon infrastructure
- Amazon offers VM monitoring tools

Firewalls and Security Groups

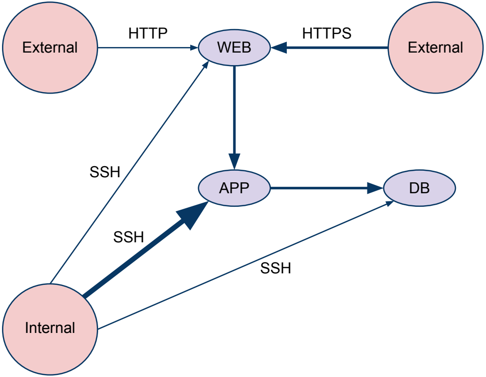
- Firewalls restrict the inbound and outbound traffic between network nodes
- Based on a set of rules that can allow or block by IP address and port
- Security Groups act as a firewall between VM and Internet to restrict undesired inbound traffic
- Security Groups do not restrict outbound traffic from Amazon VMs

Multi-tier Web System



Vulnerability of this System

Attack graph constructed from analysis of standard EC2 VM configuration (Bleikertz et al. [1])



1. Introduction
   - Defining the Cloud
   - Cloud Services
2. Virtual Machine Security
   - Monitoring
   - Amazon EC2
3. Data Centric Security
   - Where it Matters
   - Secure Data Processing
4. Conclusion

What is data security?

- Securing the flow of data between interdependent cloud services
- Encrypting sensitive data
- Digital signatures may be used to verify the authenticity of source data

Potentially dishonest infrastructure and content providers have access to large amounts of private data (e.g., Amazon employees)

Online Stores

- Online marketplaces (such as Amazon) manage many transactions between merchants
- Product information & inventory as well as monetary transactions are common exchanges of information
- The communication layer must be secured and data must be encrypted to ensure data is not tampered with.

Stock Market Analysis/Prediction

- Banks have large amount of data to process in timely manner
- Algorithms used and results produced must be kept secure
- Cloud offers necessary resources and storage for this task, but can the workloads remain private?

MapReduce

MapReduce is a framework developed by Google to utilize parallel and distributed resources

MapReduce has two steps: map and reduce

- The map step divides the workload into smaller chunks
- The reduce step aggregates the results from subworkers. Typically, new workers are created on parallel and distributed resources that other users may have access to

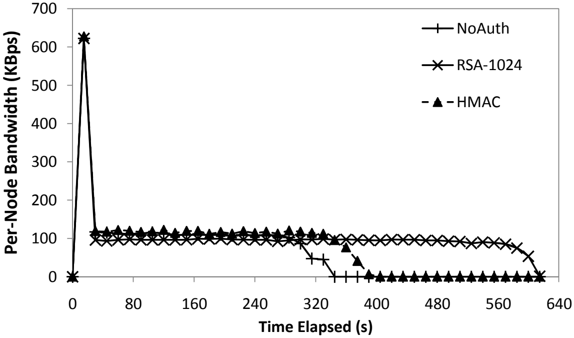
MapReduce WordCount Implementation

Zhou et al. developed a WordCount implementation using RSA-1024 and SHA-1 HMAC as encryption methods to secure the workload as it is passed around distributed resources [2].

- RSA-1024 encrypts entire message and digital signatures are often added
- SHA-1 HMAC provides only digital signatures attached to message

Results

Relationship of node-to-node bandwidth as a proxy for completion time



1. Introduction
   - Defining the Cloud
   - Cloud Services
2. Virtual Machine Security
   - Monitoring
   - Amazon EC2
3. Data Centric Security
   - Where it Matters
   - Secure Data Processing
4. Conclusion

Wrap Up

- Cloud computing offers economical and performance gains for developers and users
- While hardware infrastructure is entirely outsourced, the applications and services still must be configured
- To ensure sensitive data remains private, virtual machines must be locked down and monitored and external communications must be encrypted.

Questions?

Questions?

References

S. Bleikertz, M. Schunter, C. W. Probst, D. Pendarakis, and K. Eriksson.
Security audits of multi-tier virtual infrastructures in public infrastructure clouds.
In *Proceedings of the 2010 ACM workshop on Cloud computing security workshop*, CCSW '10, pages 93–102, New York, NY, USA, 2010. ACM.

W. Zhou, M. Sherr, W. R. Marczak, Z. Zhang, T. Tao, B. T. Loo, and I. Lee.
Towards a data-centric view of cloud security.
In *Proceedings of the second international workshop on Cloud data management*, CloudDB '10, pages 25–32, New York, NY, USA, 2010. ACM.