CloudNet: A Platform for Optimized WAN Migration of Virtual Machines

Timothy Wood
Prashant Shenoy
University of Massachusetts Amherst

&

K.K Ramakrishnan
Jacobus Van der Merwe
AT&T Labs - Research

Presented by Assaf Israel
1. Introduction & Motivation
2. CloudNet Design Overview
3. WAN Migration in CloudNet
4. WAN Migration Optimizations
5. Performance Evaluation
6. Conclusions
It’s all about the Scale

- Single server
- Data center
- Interconnected Data centers
WAN Migration Use Cases

• Private - Public clouds transition
• Cloud burst
• Client demand
  ❖ Latency
  ❖ “Follow the Sun”
(WAN) Live Migration Concerns

• Black Box
• Maintain VLAN across WAN
• (Almost) seamless migration process
• Security
• Low bandwidth / High latency
• Storage/Data migration
CloudNet Design Overview

• Seamless, Secure Cloud Connections
• Resource Pools that Span Data Centers
• Efficient WAN Migration
Network Abstraction

• Using **Virtual Private Networks (VPN)** we can decouple the Logical network topology from the physical one

• Benefits:
  - **IP Separation:** Applications uses VPNs domain IPs, while Cloud providers are free to allocate IPs without fear of conflicts.
  - **Security:** Firewalls rules and policy are easier to manage.
  - **Dynamic & Fast VPN configuration:** Automated process by utilizing **VPN Controllers**.
Controlling VPNs within the data center

- VPC traffic is isolated using VLAN.
- A logical router (Customer Edge) connects the VPC to a VPN endpoint.
Controlling VPNs (Cont.)

- Provider Edge (PE) routers hold a Virtual Routing & Forwarding (VRF) table which holds all the routing information within each VPN.
- Tables are maintained by the VPN Controller.
- VPN Tunneling connects PEs to one another.
MANAGING IT ALL

- **Cloud Manager (Intra-data-center Mgr)**
  - Dynamically partitions the data center into Virtual Private Clouds (VPC)
  - VPNs maintenance within the cloud
  - VM placement within the cloud

- **Network Manager (Inter-data-center Mgr)**
  - Responsible for the creation and resource provisioning of VPNs.
  - Defines a Provider Edge (PE) router at each data center
Managing it all

- **Cloud Manager (Intra-data-center Mgr)**
  - Dynamically partitions the data center into Virtual Private Clouds (VPC)
  - VPNs maintenance within the cloud
  - VM placement within the cloud

- **Network Manager (Inter-data-center Mgr)**
  - Responsible for the creation and resource provisioning of VPNs.
  - Defines a Provider Edge (PE) router at each data center
Resource Management

• Individual Cloud Providers can easily map new resources to the different VPC
• Resources can be dynamically moved between Data Centers
• Application aware load-balancing
• Data center efficiency placement
WAN Migration algorithm in CloudNet

• Step 1: Establish layer-2 connectivity between data centers (Using VPN)

• Step 2: If storage is not shared, transfer all disk state

• Step 3: Transfer the memory state of the VM to a server in the new Data Center, as it continues running without interruption

• Step 4: Once the disk and memory state have been transfered, briefly pause the application for the final transition of memory and CPU state
Storage Migration

- **Asynchronous Copy**
  - New Writes initiates a Copy but do not Block
  - Order is always maintained
  - Faster writes but inconsistent replica (eventually consistent)

- **Synchronous Copy**
  - New Writes initiates a Copy and Block until a confirmation is received
  - Slower writes but consistent replica
WAN Migration Flow

- VPN Setup
- Pause VM
- Live Memory Transfer
- Asynchronous Copy
- Synchronous

Time (not to scale)
Optimizing WAN VM Migration

• So far we have a good base, but this is not enough
Smart Stop & Copy

- In Xen migration is bound by 29 iterations, or until a very small number of pages is reached.

![Graph showing iteration vs. number of pages remaining, with a maximum of 20000 and a minimum of 0.]

VM running kernel compilation
622 Mbps bandwidth
5 msec latency
Smart Stop & Copy (cont.)

- CloudNet’s solution: Track a short history of the number of pages remaining to be sent, and look for a local minimum

- Not perfect but good enough in most cases
Content Based Redundancy

- Goal: identify zero and previously sent memory blocks

Source host

Target host

Cache hit: transfers the hash entry and block address

4 Bytes ~ 4KB
Content Based Redundancy

- Goal: identify zero and previously sent memory blocks

Source host

Target host

CBR cache

Cache hit: transfers the hash entry and block address

4 Bytes ~ 4KB
Content Based Redundancy

- Goal: identify zero and previously sent memory blocks

Source host

Cache hit: transfers the hash entry and block address

Target host

4 Bytes ~ 4KB
Content Based Redundancy Statistics

![Bar chart showing redundancy statistics for Kernel Compile, TPC-W, and SPECjbb. The chart indicates the percentage of RAM occupied by zeroes and non-zero duplicates.](image)

- **Kernel Compile**:
  - Zeroes: 0%
  - Non-0 Duplicates: 10%

- **TPC-W**:
  - Zeroes: 20%
  - Non-0 Duplicates: 30%

- **SPECjbb**:
  - Zeroes: 50%
  - Non-0 Duplicates: 70%

Redundancy (% of RAM)
Page Deltas

- Idea: Only transfer what actually has changed since the last transfer
- In case of a CBR cache miss, use the page address as a secondary cache index
- Use XOR between the current page & the cached version, and length encode the result
Page Deltas

- Idea: Only transfer what actually has changed since the last transfer.
- In case of a CBR cache miss, use the page address as a secondary cache index.
- Use XOR between the current page & the cached version, and length encode the result.

![Source host CBR cache diagram]

![Target host CBR cache diagram]
Page Deltas

- Idea: Only transfer what actually has changed since the last transfer
- In case of a CBR cache miss, use the page address as a secondary cache index
- Use XOR between the current page & the cached version, and length encode the result
Performance Evaluation

- 3 data centers across the U.S - Illinois, Texas & California (Used as client site)
- VPNS Endpoint deployment took a total of 24.21 Sec
- VPN Updates took only 30 msec (propagated to all data centers)
WAN Migration Performance

- TPC-W web application ("Amazon shop" with 600 simultaneous clients actions) migrated from Texas to Illinois
- 85Mbps bandwidth
- 1.7GB of allocated RAM
- 10GB of disk

- 45% reduction in migration length
- Downtime reduced to 1 Sec (from 2.2)
- Only 6.6GB of storage transferred (Zero Blocks not transferred)
- Can be reduced 4.9 (100 MB Cache)
- Or even 3.6 (1GB Cache)
WAN Migration Performance (Cont.)

- Farther testing on other kinds of benchmarks (Kernel compilation & SPECjbb) showed 30-70% reduction in migration time, and up to 50% reduction in downtime.

- The contribution of each optimization depends on the type of application.
Impact of Network Conditions (simulated)

(a) TPC-W

(b) SpecJBB
Impact of Network Conditions (simulated)

(a) TPC-W Bandwidth Usage

(b) SPECjbb Bandwidth Usage
Impact of Network Conditions (simulated)

Increased latency has only a minor impact on the migration process, but may impact application performances due to synchronous disk replication.
Conclusions

• CloudNet dramatically reduces WAN migration & downtime lengths (~50%)
• Demonstrated migration across 1200km separated sites
• Good performance under low bandwidth & high latency scenarios
• Would be interesting to see the management (placement) efforts/issues WAN migration & VPCs arise as a result
• Clients must be part of the VPC! other IP mobility solutions?
Questions
Thank you