Cloud Data Centres

from Mice to Elephants

Dr. Posco Tso
p.tso@ljmu.ac.uk

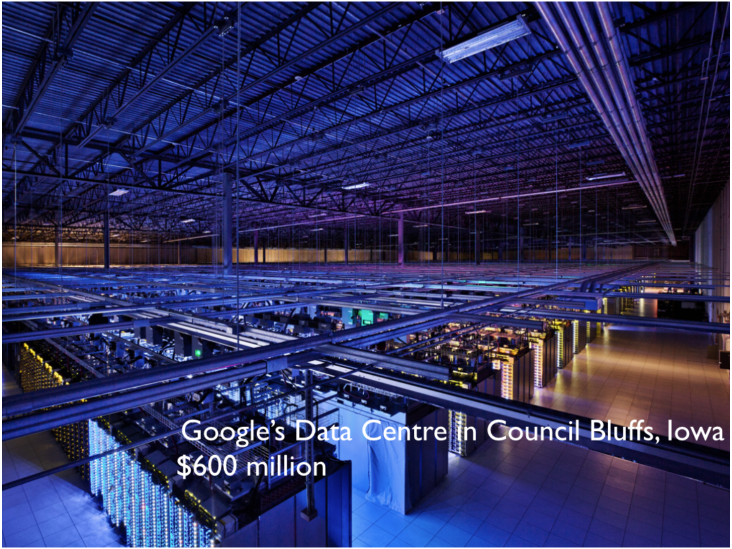
http://www.poscotso.info

Department of Computer Science
Liverpool John Moores University

where do your music/videos/pictures live?

in the cloud

in the data centres - enormous & costly



in the data centres - enormous & costly

in the data centres - enormous & costly

Facebook's Data Centre,
North Carolina
$606 million

in the data centres - enormous & costly

Apple's Data Centre,
Maiden
$1 billion

on the inside - a closer look

wait, authorised persons only!

can't get in, normally

how to find out about it?

acmqueue A Guided Tour through Data-center Networking

A good user experience depends on predictable performance within the data-center network.

Dennis Abts, Bob Felderman, Google

The magic of the cloud is that it is always on and always available from anywhere. Users have come to expect that services are there when they need them. A data center (or warehouse-scale computer) is the nexus from which all the services flow. It is often housed in a nondescript warehouse-sized building bearing no indication of what lies inside. Amidst the whirring fans and refrigerator-sized computer racks is a tapestry of electrical cables and fiber optics weaving everything together—the data-center network. This article provides a "guided tour" through the principles and central ideas surrounding the network at the heart of a data center — the modern-day loom that weaves the digital fabric of the Internet.

DATA-CENTER DEVELOPMENT

Large-scale parallel computers are grounded in HPC (high-performance computing) where kilo-processor systems were available 15 years ago. HPC systems rely on fast (low-latency) and efficient interconnection networks capable of providing both high bandwidth and efficient messaging for fine-grained (e.g., cache-line size) communication. This zealous attention to performance and low latency migrated to financial enterprise systems, where a fraction of a microsecond can make a difference in the value of a transaction.

In recent years, Ethernet networks have made significant progress toward bridging the

build a scale model

Raspberry Pi Cloud

in reality

I need £to pay for my mortgage.

more technical talk

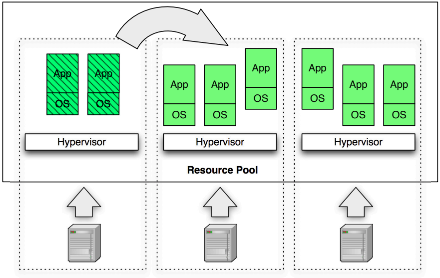
PLAN: A Policy-Aware VM Management Scheme for Cloud Data Centres

in cloud data centre

- *servers* underpin cloud computing with machine virtualisation (hence virtual machines VMs).
  Fact One: Tens of thousands of physical servers.
- *networks* have grown to support proliferated # of servers (hence apps).
  Fact Two: Networking devices are on par with servers.

server management

Rely on VMs consolidation/migration to improve server utilisation/performance/security/power efficiency.

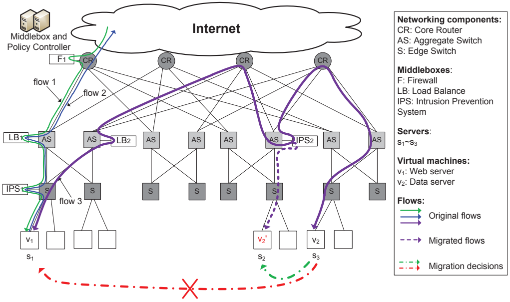


network management

Rely on *network policies* consolidation/migration to improve network utilisation/performance/security/power efficiency.

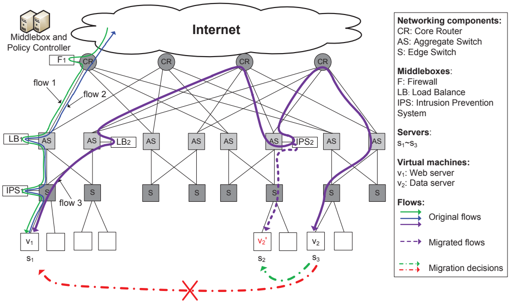
- Isolation
- Bandwidth guarantee
- block TCP:80 traffic

motivating example



- *Network polices* dictate packet forwarding rules with the help of middleboxes (firewalls, load-balancers, intrusion detection/prevention systems, etc).
- Packets has to go through pre-defined sequence of MBs governed by *Network policy*

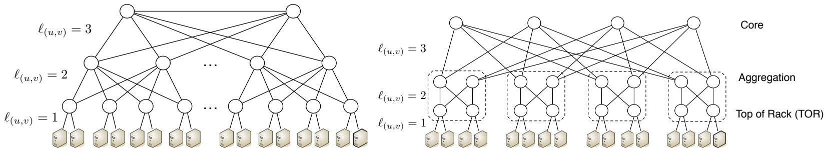
motivating example



Objectives:

- Shortest network paths amongst VMs taking into account MBs (assuming this is most efficient).
- No policy violations.

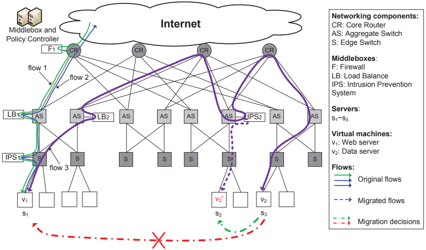
modelling - communication cost



(d) Canonical tree (e) Fat tree

Figure: The two most common DC network topologies

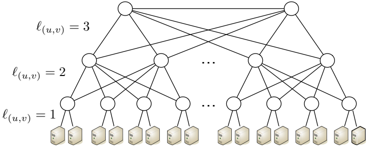
modelling - communication cost



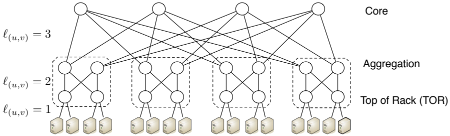
Actual routing path $L(n_i, n_j)$ is:

$$
\begin{aligned}
L_k(v_i, v_j) = \; & L(v_i, p_k^{in}) \\
& + \sum_{mb_s^k \neq p_k^{out}} L(mb_s^k, mb_{s+1}^k) \\
& + L(p_k^{out}, v_j)
\end{aligned}
\tag{1}
$$

modelling - communication cost



(a) Canonical tree



(b) Fat tree

Figure: The two most common DC network topologies

Hence, the *Communication Cost* of all traffic from VM v_i to v_j is defined as

$$\begin{aligned} C(v_i, v_j) &= \sum_{p_k \in P(v_i,v_j)} \lambda_k(v_i, v_j) \sum_{l_s \in L_k(v_i,v_j)} c_s \\ &= \sum_{p_k \in P(v_i,v_j)} \left(C_k(v_i, p_k^{in}) + C_k(p_k^{in}, p_k^{out}) \right. \\ &\qquad \left. + C_k(p_k^{out}, v_j) \right) \end{aligned} \tag{2}$$

modelling - policy awareness

$S(mb_k)$: all the servers that can reach middlebox mb_k. The *acceptable servers* that a VM v_i can migrate to are:

$$S(v_i) = \bigcap_{mb_k \in MB^{in}(v_i) \cup MB^{out}(v_i)} S(mb_k) \tag{3}$$

We denote A to be an allocation of all VMs, the feasible space of candidate servers for v_i:

$$\mathcal{S}_i = \{\hat{s} | (\sum_{v_k \in A(\hat{s})} R_k + R_i) \preceq H_j, \hat{s} \in S(v_i)\} \tag{4}$$

Let $\mathcal{C}_i(s_j)$ be the total communication cost induced by v_i between s_j and $MB^{in}(v_i) \cup MB^{out}(v_i)$, where $s_j = A(v_i)$.

$$\mathcal{C}_i(s_j) = \sum_{p_k \in P(v_i,*)} C_k(v_i, p_k^{in}) + \sum_{p_k \in P(*,v_i)} C_k(v_i, p_k^{out}) \tag{5}$$

modelling - policy awareness

The *utility* (benefit) of the migration $A(v_i) \rightarrow \hat{s}$ is defined as:

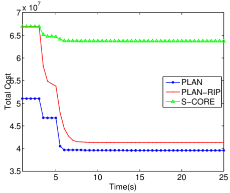
$$U(A(v_i) \rightarrow \hat{s}) = \mathcal{C}_i(A(v_i)) - \mathcal{C}_i(\hat{s}) - C_m(v_i) \tag{6}$$

Optimisation objective: find a new allocation $\hat{A}$ that maximizes the total utility:

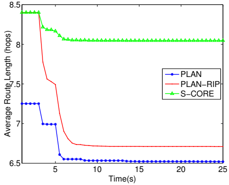
$$\begin{aligned} & \max \mathcal{U}_{A \rightarrow \hat{A}} \\ & s.t. \, \mathcal{U}_{A \rightarrow \hat{A}} > 0 \\ & \quad \hat{A}(v_i) \in S_i, \forall v_i \in \mathbb{V} \end{aligned} \tag{7}$$

The *PLAN* problem is NP-Hard - Multiple Knapsack Problem (MKP)

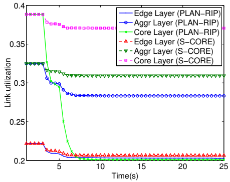
experimental results



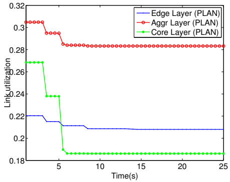
(a) Total communication cost



(b) Average route length



(c) Link utilization



(d) Link utilization

Figure: Performance comparison of *PLAN* and *S-CORE* VM migration schemes

Thank You!

slides available at:
`https://github.com/posco/cmptalk`