Cicada: Introducing Predictive Guarantees for Cloud Networks

Katrina LaCurts, Jeffrey C. Mogul, Hari Balakrishnan, Yoshio Turner

Presented by Kenneth Lui (wckl2)
24th November, 2015
Agenda

- Background
- Solution
  - Predictive guarantee
  - Cicada
- Evaluation
- Summary
Background
Motivation

- Network-bandwidth guarantee can improve predictability of application performance and cost in cloud environment
- But tenants usually do not know how much they want
  - leads to over-provisioning or under-provisioning
Challenge

- Time-varying bandwidth consumption
- Spatially inhomogeneous
Solution
Predictive guarantees

1) Predict
   a) based on traffic between VMs
   b) using an averaging interval

2) Offer
   a) offer bandwidth guarantee
   b) customers may choose to accept or reject
Advantage

- Simpler for the tenant
- Support temporal variation and spatial variation
- May support fine-grained (VM-to-VM directed path) guarantee
1. Collect + Predict + Offer
2. VM placement
Architecture

Figure 1: Cicada’s architecture.

1. customer makes initial request
2. provider places tenant VMs and sets initial rate limits
3. hypervisors continually send measurements to Cicada controller
4. Cicada predicts bandwidth for each tenant
5a. provider may offer the customer a guarantee based on Cicada’s prediction
5b. provider may update placement based on Cicada’s prediction
Prediction method

- Adapted Herbster and Warmuth’s “tracking the best expert” idea
- Linear combination of all previously observed traffic matrices
- Weights are learnt in online fashion
Placement algorithm

- Greedy
  - place clusters on the smallest subtree
  - place the most-used VM pairs on the highest bandwidth paths
Data

- Six months of traffic data from an HP Cloud Services datacenter
- Collected from top-of-rack switches
- Captures VM-to-VM traffic patterns
Evaluation

Figure 3: Inter-rack bandwidth available after placement.
Summary

- Cicada is different from existing works that it does not require the tenant to specify network demands up front.
- Able to provide temporally- and spatially-varying guarantees.
Critical Analysis

- Does it handle collocation of VMs (e.g. such that there’s no traffic detected by Cicada)?
- Does it handle data traffic with SAN?
- Ignorant about CPU usage, memory, etc.
- Figures are not clear, very vague description