An Adaptive Technique to Model Virtual Machine Behavior for Scalable Cloud Monitoring

C. Canali
R. Lancellotti

University of Modena and Reggio Emilia
Department of Engineering “Enzo Ferrari”
Challenge: monitoring

- Large data centers (> $10^5$ VMs) → huge amount of data
- Point of view: IaaS provider → monitoring supporting infrastructure management
- VM can be anything → treat VM as black boxes
- → Scalability issues
Challenge: monitoring

- **Current approach: reduce amount of data in a uniform way**
  - Reduce sampling frequency
  - Reduce number of metrics considered
- **→ Reduced monitoring effectiveness**
  - Less information available for management
- **Solution: Exploit VM similarity**
Improving monitoring scalability

- Group similar VMs together
- Detailed monitoring of cluster representatives
- Reduced monitoring of other VMs

\[ \rightarrow \text{Data collection reduced by one order of magnitude} \]
Challenge: fast identification

- VM behavior model built starting with time series of resource usage on VMs
- Long time series to characterize VM behavior
  - → Highly accurate clustering
- Clustering accuracy decreased by shorter time series
  - → problems coping with Cloud dynamic behavior
- Need to combine fast and accurate identification of VM behavior
Reference scenario

- **IaaS, medium-long term commitment**
  - Amazon Reserved instances, private cloud

- **Reactive VM relocation**
  - Local manager

- **Periodic global consolidation**
  - Global optimization
Our proposal: adaptive approach

• Observation:
  – Some VMs are easily identified as belonging to a cluster even after short observation
  – Other VMs require more detail to build a reliable behavior model

• Proposal:
  – Cluster as fast as possible VMs clearly belonging to a cluster
  – Postpone clustering of VMs when not sure

• Adoption of fuzzy logic perspective
  – Introduce decree of belonging of VM to clusters to rete reliability of clustering result
  – Gray area of uncertain clustering
Adaptive algorithm

- Adaptive identification of time series length
  - When clustering is not ambiguous (white area)
    - VM behavior model is OK
    - No update required
  - When clustering is ambiguous (gray area)
    - Need to improve VM behavior model
    - Further observation required
Definition of Gray Area

• Feature space: k-dimensional space
  – Each VM described by a feature vector (point in feature space)
  – Each cluster has a centroid described as a point in the feature space

• For each VM n:
  – Vector of distances from the cluster centroids
    \[ D_n = \{d_{1n}^n, d_{2n}^n, \ldots, d_{Cn}^n\} \]

• Definition of gray area
  – A VM is in gray area if there exists a couple of clusters i, j such that
    \[ 1 - \varepsilon < \frac{d_{in}^n}{d_{jn}^n} < \frac{1}{1 - \varepsilon}, 0 < \varepsilon < 1 \]
Definition of Gray Area

- Higher epsilon $\rightarrow$ wider gray area
- Problem: definition of right value of epsilon
  - Open problem, still working on that...
  - Experimental results suggest $\varepsilon = 0.33$ as a rule of thumb
Case study

- **Datacenter supporting a e-health Web application**
  - Web server and DBMS
  - 110 VMs
  - 11 metrics for each VM,
  - Sampling frequency: 5 min
- **Goal: separate Web servers and DBMS**
  - Clustering accuracy
  - % of VM in gray area
- **2 VM behavior model approaches**
  - PCA-based
  - Bhattacharyya distance-based
Experimental results

Time series length: 1 day, PCA-based clustering
Experimental results

- Validating the choice of epsilon
  - For $\varepsilon \geq 0.33$ the accuracy is 100% (absence of mis-classified VMs)
  - The size of the gray area depends on the clustering algorithms

![Graphs showing clustering accuracy and percentage of gray area vs. epsilon for PCA-based and Bhattacharyya distance-based methods.](image-url)
Experimental results
Conclusion

- **Experimental results are encouraging**
  - Can achieve high clustering purity
  - Can provide accurate clustering even with very short time series
  - Works with different clustering algorithms
  - Adaptive approach to select the time series length

- **This is not a crystal ball**
  - But may be a useful tool to improve monitoring and management of cloud data centers
On-going works

- Adaptive selection of the $\varepsilon$ parameter
- Evaluation with time-series $< 24$ h
- Comparison with other fuzzy clustering algorithms
- Additional experiments with different workloads (*help appreciated*)
An Adaptive Technique to Model Virtual Machine Behavior for Scalable Cloud Monitoring

C. Canali
R. Lancellotti

University of Modena and Reggio Emilia
Department of Engineering “Enzo Ferrari”