Virtual-to-Physical Mapping Inference in Cloud Environments

Yang Song, Ramani Routray, and Rakesh Jain
IBM Almaden Research Center
650 Harry Road, San Jose, CA 95120
yangsong, routrayr, rakeshj@us.ibm.com

1. Motivation

Virtualization is a double-edged sword.

Pros:

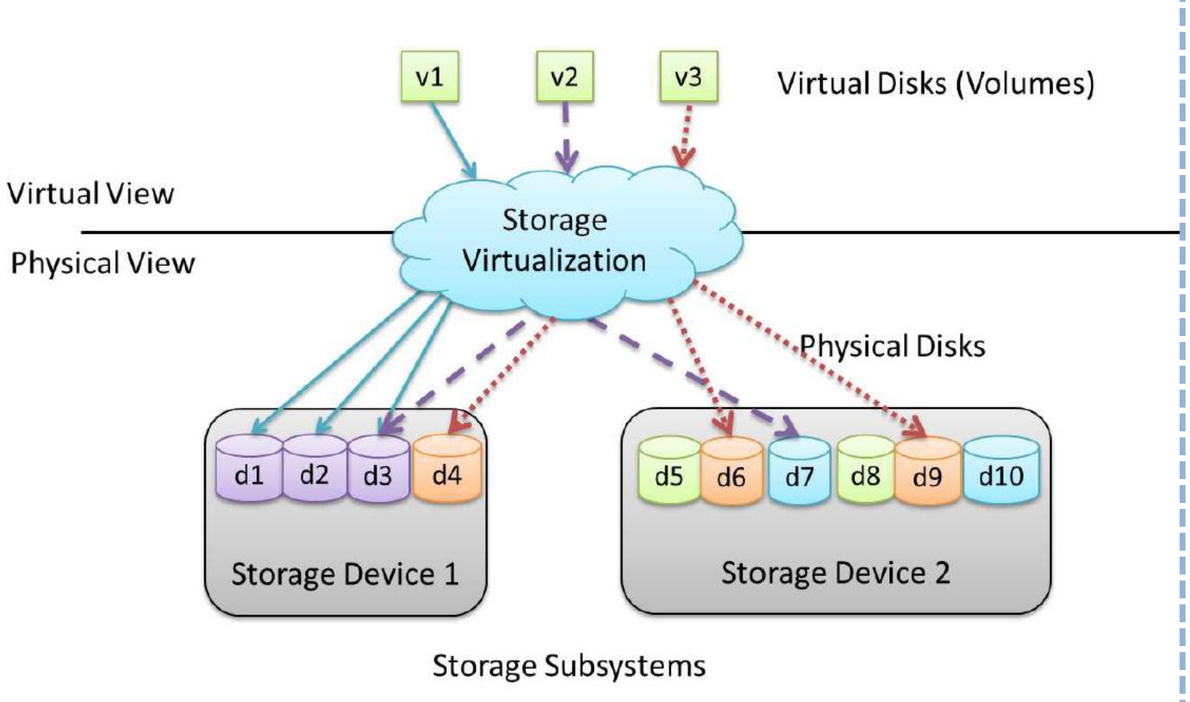
- Flexible management
- Efficient resource utilization
- Multi-tenancy

Cons:

- Additional layers of indirections
- Obscured resource mapping
- Root cause analysis is challenging

2. Objective

Can we infer the virtual-to-physical mapping relationship, by ***only*** observing the I/O counts of the inputs and the outputs?



Advantages

- Lightweight
- Non-intrusive
- Suitable for multiple layers of virtualization where end-to-end queries are prohibitive

3. Step I: Measurements

$$X = [x_{j,n}] \ \forall j = 1, \cdots, V, \text{ and } \forall n = 1, \cdots, N,$$

Number of I/O counts for volume *j* at time *n*.

$$Y = [y_{i,n}] \ \forall i = 1, \cdots, D, \text{ and } \forall n = 1, \cdots, N,$$

Number of I/O counts for disk *i* at time *n*.

4. Step II: Inference Algorithms

$$\min_H ||Y - HX||_F^2$$
$$\text{s.t. } H \geq 0$$

H is a D-by-V matrix to be inferred. Each element of H denotes the average number of I/O counts of a disk caused by a volume.

Our solutions:

(a) Nonnegative Matrix Factorization (NMF) Based Algorithm

(b) General Primal-Dual Based Algorithm

5. Numerical Example

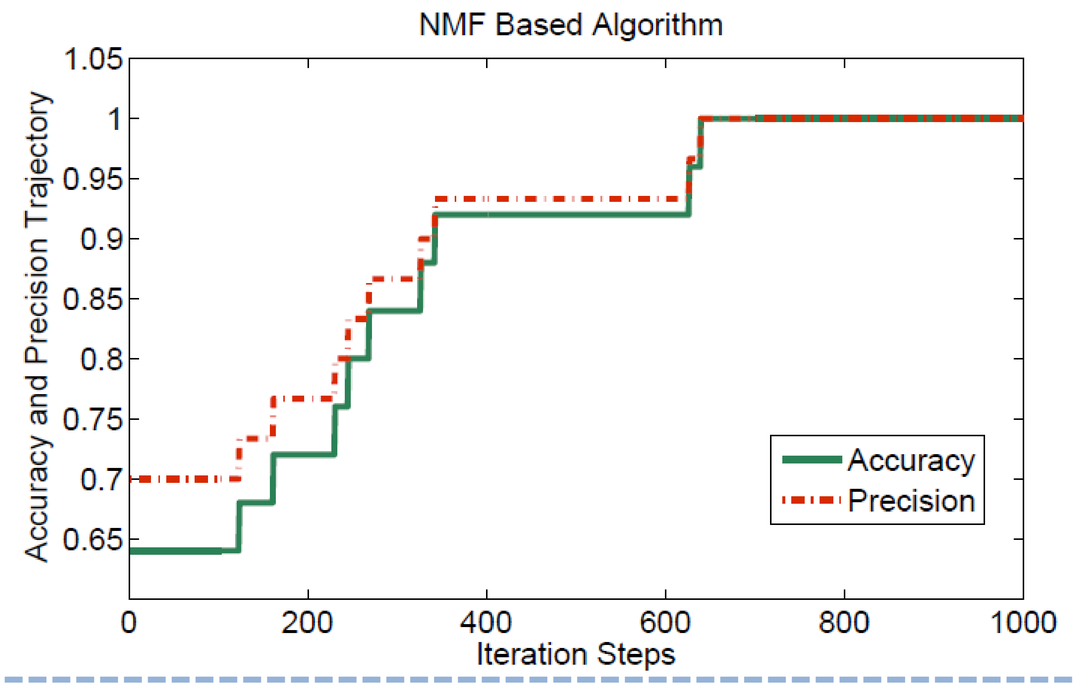
- IBM System Storage SAN Volume Controller in a virtualized storage environment.

- 10 volumes, 5 disks, I/O measurement every 5 minutes for two days.

- Evaluation criteria: *Accuracy* and *Precision*

$$\text{accuracy} = \frac{\text{no. of true positive} + \text{no. of true negative}}{\text{no. of all classifications}}$$

$$\text{precision} = \frac{\text{no. of true positive}}{\text{no. of true positive} + \text{no. of false positive}}$$



6. Takeaways

- ❑ Virtual-to-physical mapping relationship is required for many applications.
- ❑ Our work provides an inference framework without complex privileged queries.
- ❑ Lightweight, non-intrusive, easy to implement.