

iPODS: Intelligent Performance Optimisation of Virtualised Data Storage Systems

Nick Dingle, Peter Harrison, William Knottenbelt, Abigail Lebrecht and Soraya Zertal

{njd200,pgh,wjk,asl102}@doc.ic.ac.uk, zertal@prism.uvsq.fr

Department of Computing, Imperial College London; Laboratoire PRiSM, Université de
Versailles Saint-Quentin-en-Yvelines

What is iPODS?

- Three year project funded by the UK Engineering and Physical Sciences Research Council (EPSRC)
- Supported by industrial partners (Reuters and IBM)

Background to iPODS

- Importance of storage in today's world, and corresponding importance of storage system QoS
- Rise of virtualised storage infrastructures that incorporate intelligent storage fabrics
- Typically have their physical resources organised into storage tiers, each of which delivers a different cost/capacity ratio against a certain QoS
- An important challenge is to place data onto the most appropriate tier and then migrate it from one tier to another as the access profile evolves

Current State-of-the-Art

- Existing mechanisms for fabric intelligence in are relatively simple, with data migration policies that are:
 - Centred on capacity utilisation and failure recovery
 - Not sensitive to dimension of the access profile other than access frequency, type and size
- The most sophisticated tiered SAN available today offers fixed-interval block-level data migration based on access frequency
- Consequently, data-specific QoS cannot be guaranteed, and space utilisation is potentially inefficient

Objective of iPODS

“To develop more sophisticated fabric intelligence that is able to autonomously and transparently migrate data across tiers and organise data within tiers to deliver the required QoS in terms of factors such as response time, availability, reliability, resilience, storage cost and power utilisation.”

Objective of iPODS

“To develop more sophisticated fabric intelligence that is able to **autonomously and transparently migrate data** across tiers and organise data within tiers to deliver the required QoS in terms of factors such as **response time, availability, reliability, resilience, storage cost and power utilisation.**”

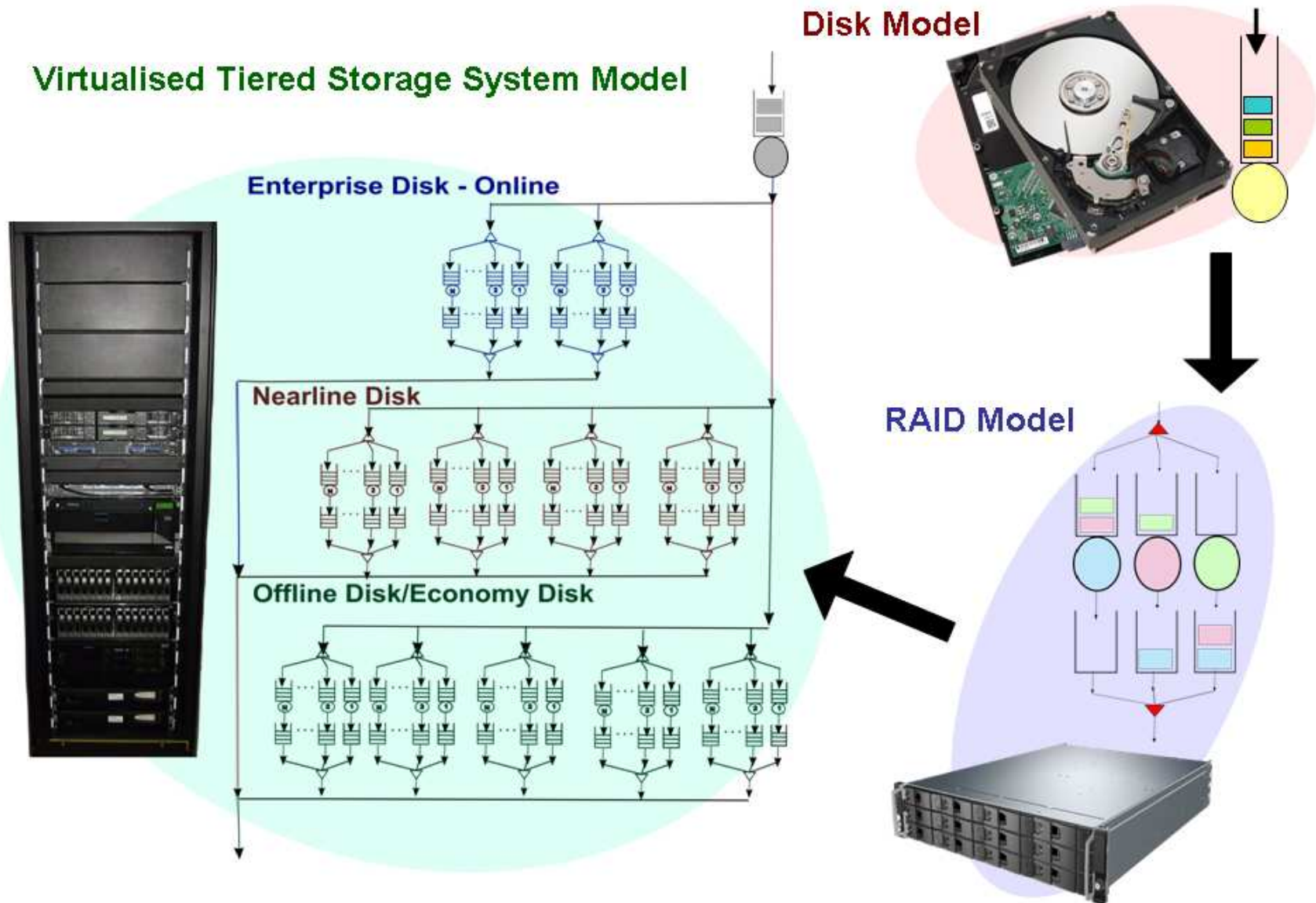
Objective of iPODS

- Two key issues:
 - How best to map data onto physical disk devices?
 - On what factors should this choice of mapping be based?
- To address these, we are seeking to develop:
 - Intelligent data placement and migration strategies
 - Performance evaluation tools to assess the benefits of such strategies quantitatively

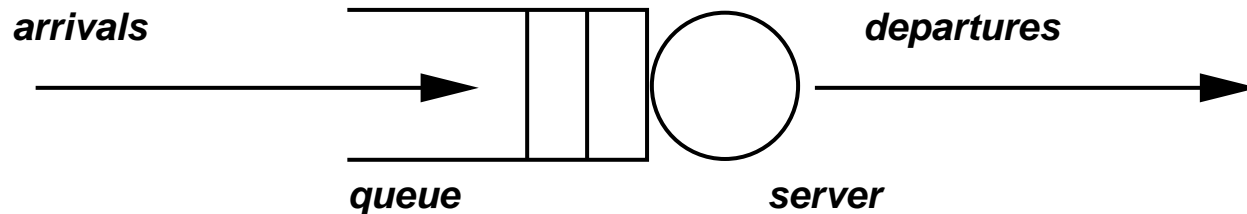
Summary of Work to Date

- Focused on creation of performance evaluation tools
- Creation of accurate analytical models of storage system performance
- General approach is to use queueing networks
- Go beyond prior work by calculating full response time distributions, rather than just averages
- Validation against measurements taken from real-life storage systems

Development of Performance Models

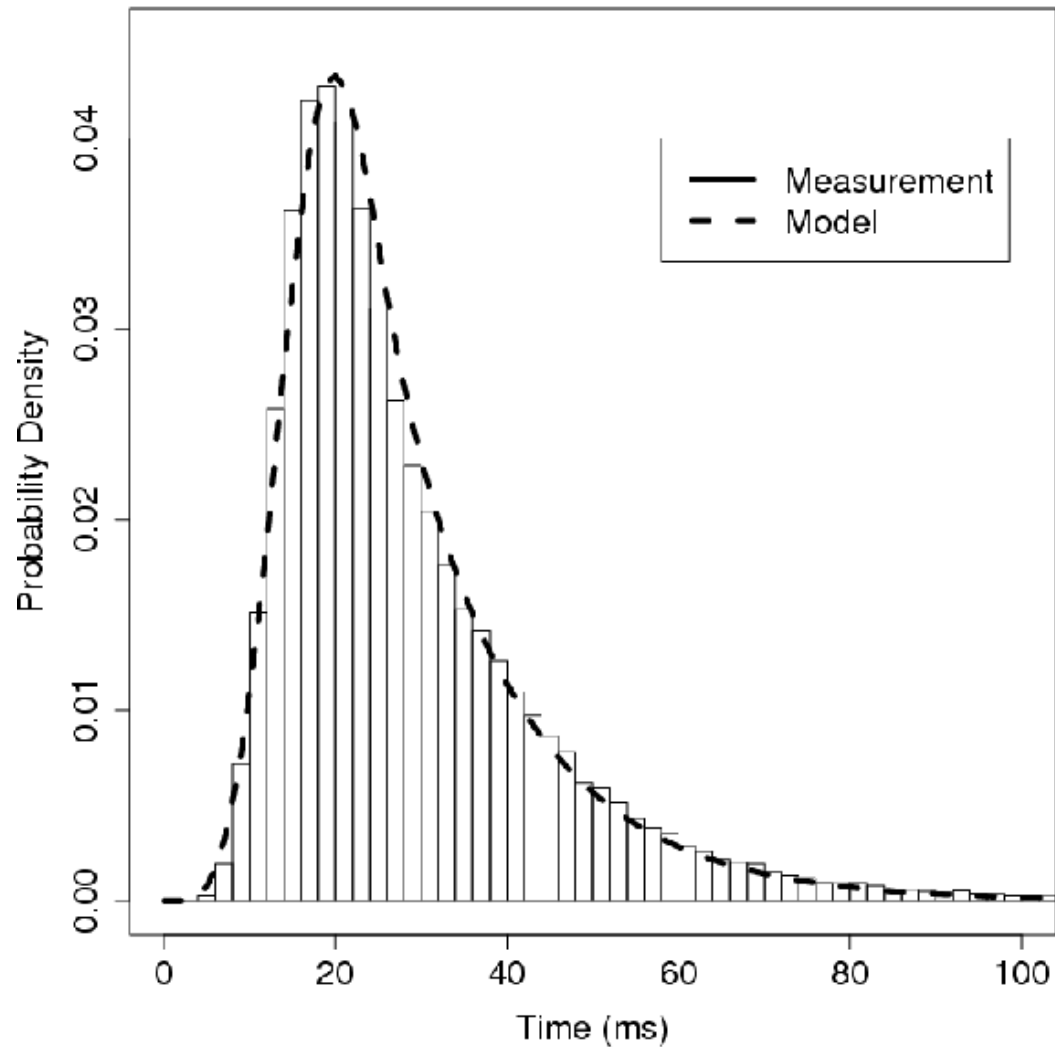


Disk Model



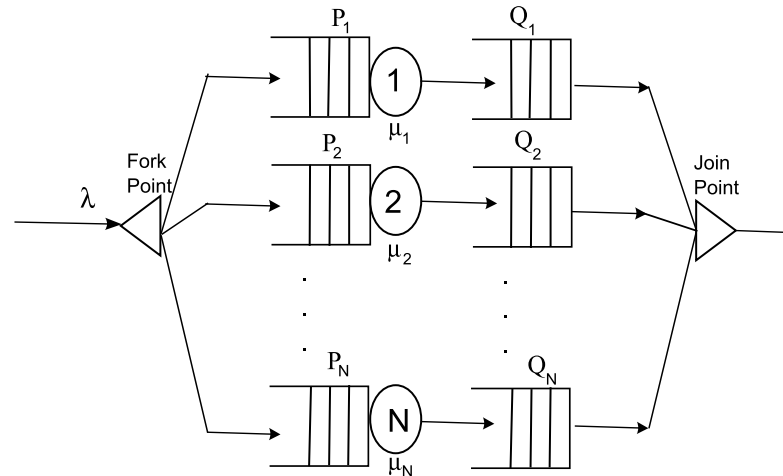
- Model a disk as a single server and associated queue, where customers represent I/O requests
- Request service time based on the rotational latency, seek time and data transfer time of a real disk
 - Taken from manufacturer's specifications or collected from our measurements
- Requests are assumed to arrive at a given random rate and can be of different types and sizes
- Experiments show excellent agreement between the disk model and its real-world counterpart

Disk Model: Sample Response Time Results



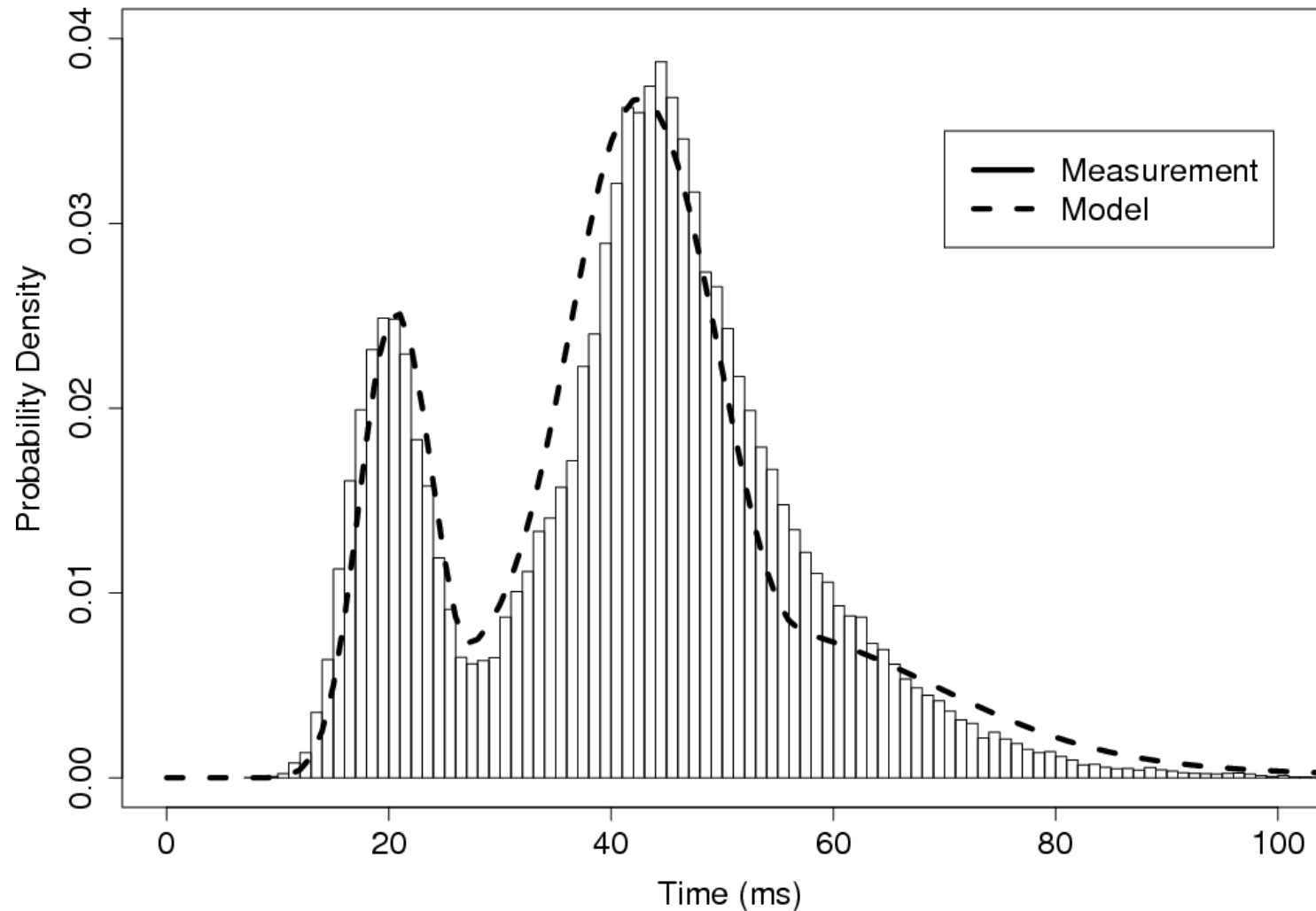
PDF for a single disk for a workload with 100% reads with mean request size of 640K and an arrival rate of 0.01 requests/ms

RAID Model



- RAID systems are collections of single disks, so we model them as queueing networks incorporating multiple instances of our single disk model
- Need to represent the intricacies of different RAID levels (e.g. RAID01 vs. RAID5)
- Again, experiments show excellent agreement between the model and its real-world counterpart

RAID Model: Sample Response Time Results



PDF for a RAID5 system for a workload of 25% reads and 75% writes with mean request size of 512K and an arrival rate 0.01 requests/ms

Current Areas of Research

- Improvements to the disk/RAID model:
 - Caching
 - RPO/NCQ
 - More realistic arrival distributions derived from real-life I/O request traces
- Modelling of solid-state drives
- Creation of a custom file system that supports specification of QoS attributes for files
- Visualisation of data layout

Future Work

- Representation of additional RAID configurations (e.g. RAID6)
- Development of analytical models of tiered storage systems (extension of our RAID model)
- Development of data placement and migration algorithms
- Evaluation of these algorithms using our models

<http://aesop.doc.ic.ac.uk/>