Simulation of Hierarchical Storage Systems for TCO

Jakob Lüttgau and Julian Kunkel

German Climate Compute Center (DKRZ)

June 22, 2017

Overview

1. Motivation and Background
2. Modeling and Simulation Tape Storage Systems
3. Evaluation
4. Conclusion / Discussion

Motivation

Long-term storage and upcoming challenges for exascale supercomputers.

- Storing data from a supercomputer is a common bottleneck
- Deep storage hierarchies to balance cost and performance
- Tape is among the most affordable storage solutions
- RAIT and object semantics change how tapes are used
- Innovation mostly dependent on vendors

But given a simulator it would be possible to:

- Experiment with alternative configurations
- Better understand cost and deploy more informed
- Explore strategies to rollout LTO generations

This is still a work in progress.

Automated Tape Libraries

Archives; Data reduction and compression; Encryption; Self-describing tape formats;

IBM TS3500 Library Complex (IBM, 2011b)

TFinity Library Complex (Spectralogic, 2016b)

StorageTek SL8500 Library Complex (Oracle, 2015)

Scalar i6000 Library Complex (Quantum, 2015)

1. Motivation and Background

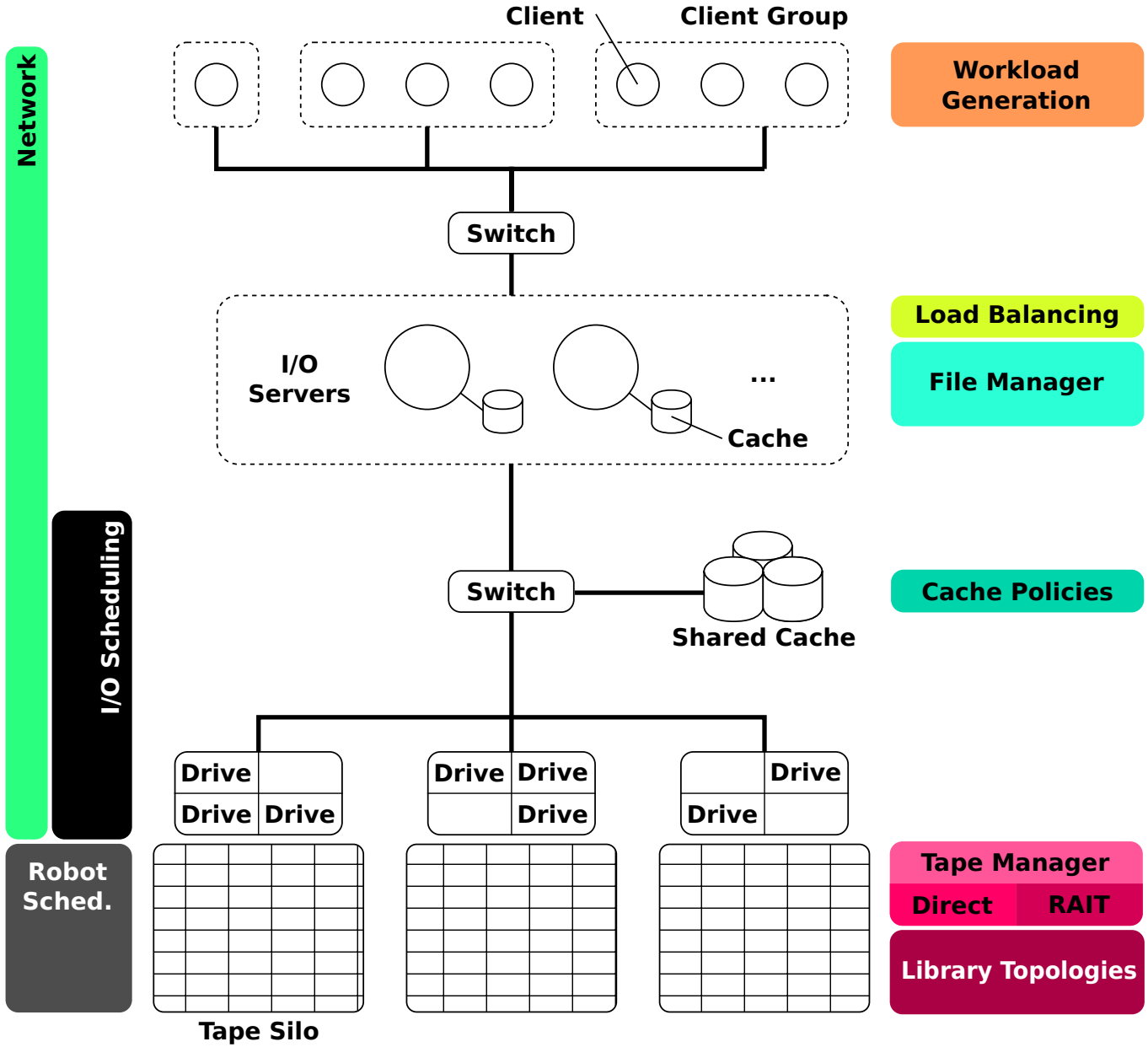
2. Modeling and Simulation Tape Storage Systems

3. Evaluation

4. Conclusion / Discussion

Model Overview

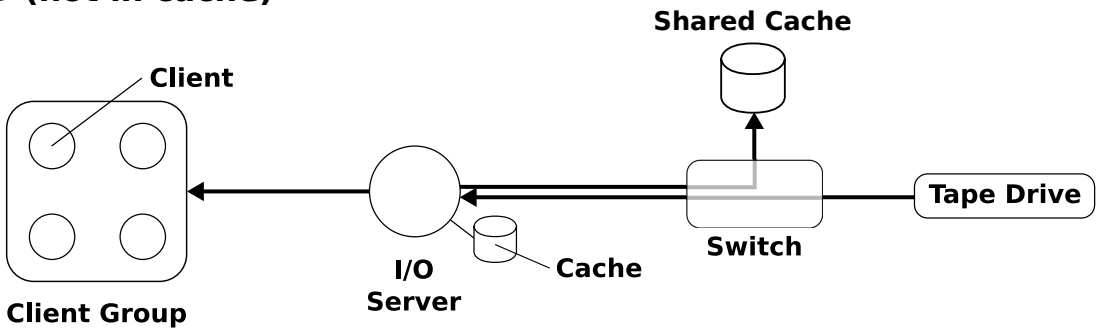
Hardware and software components in a combined overview.



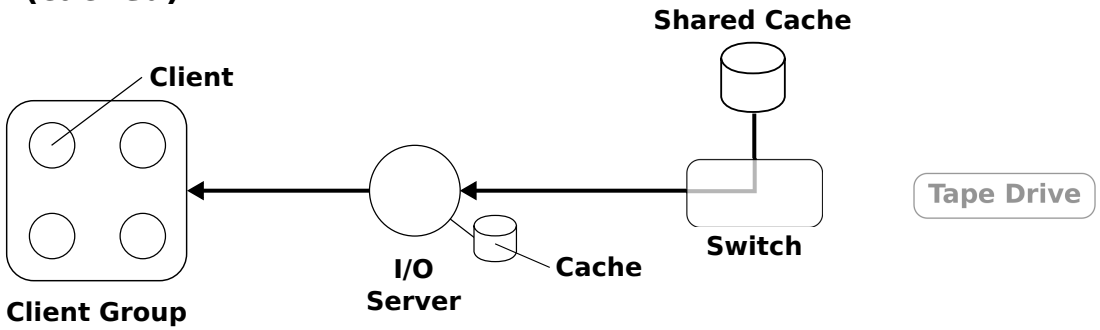
Handling `READ` Requests

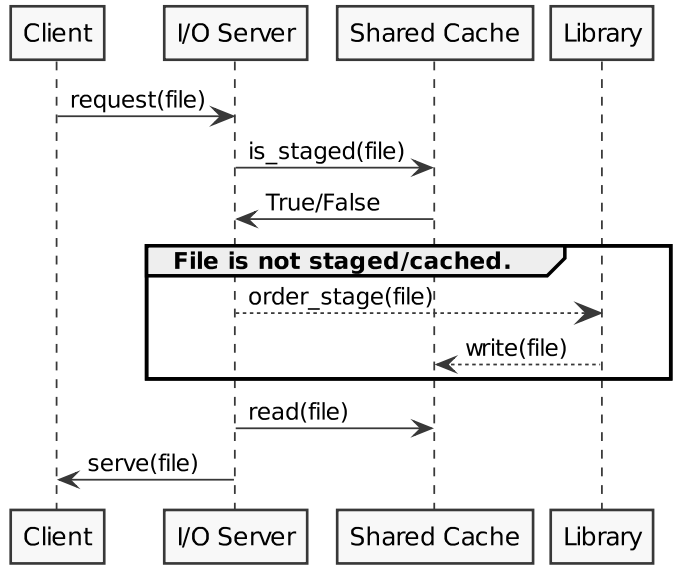
Staging of recently accessed files for reads.

READ (not in cache)



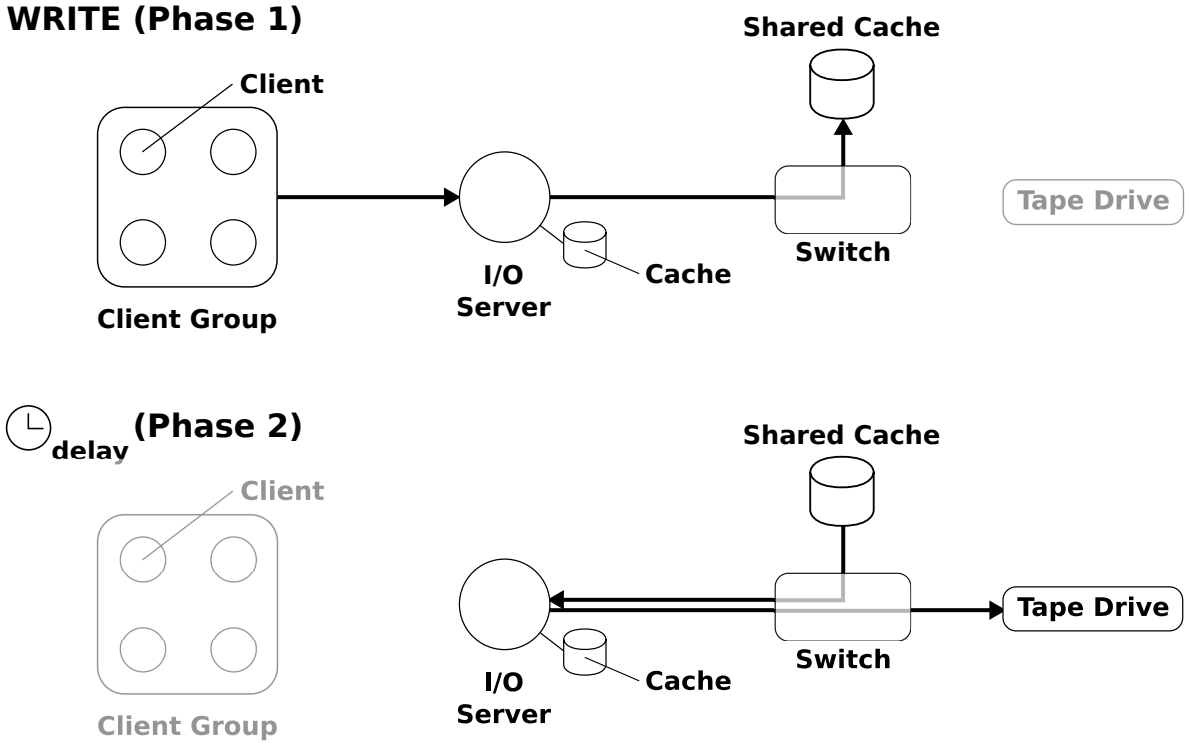
READ (cached)

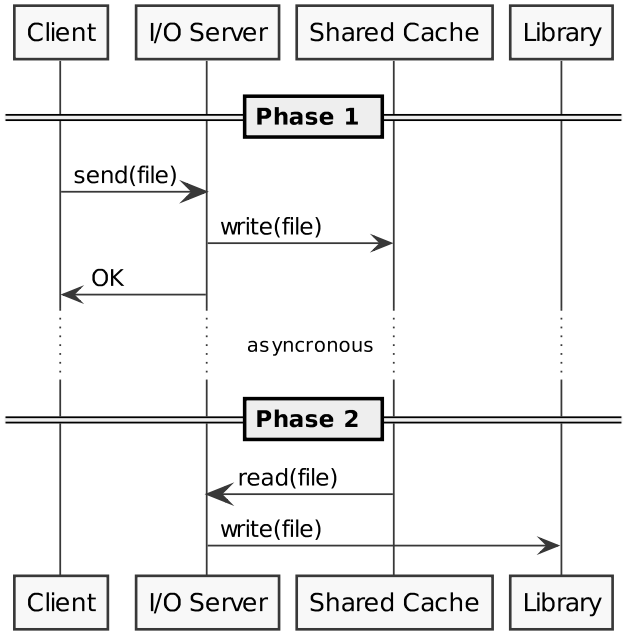




Handling `WRITE` Requests

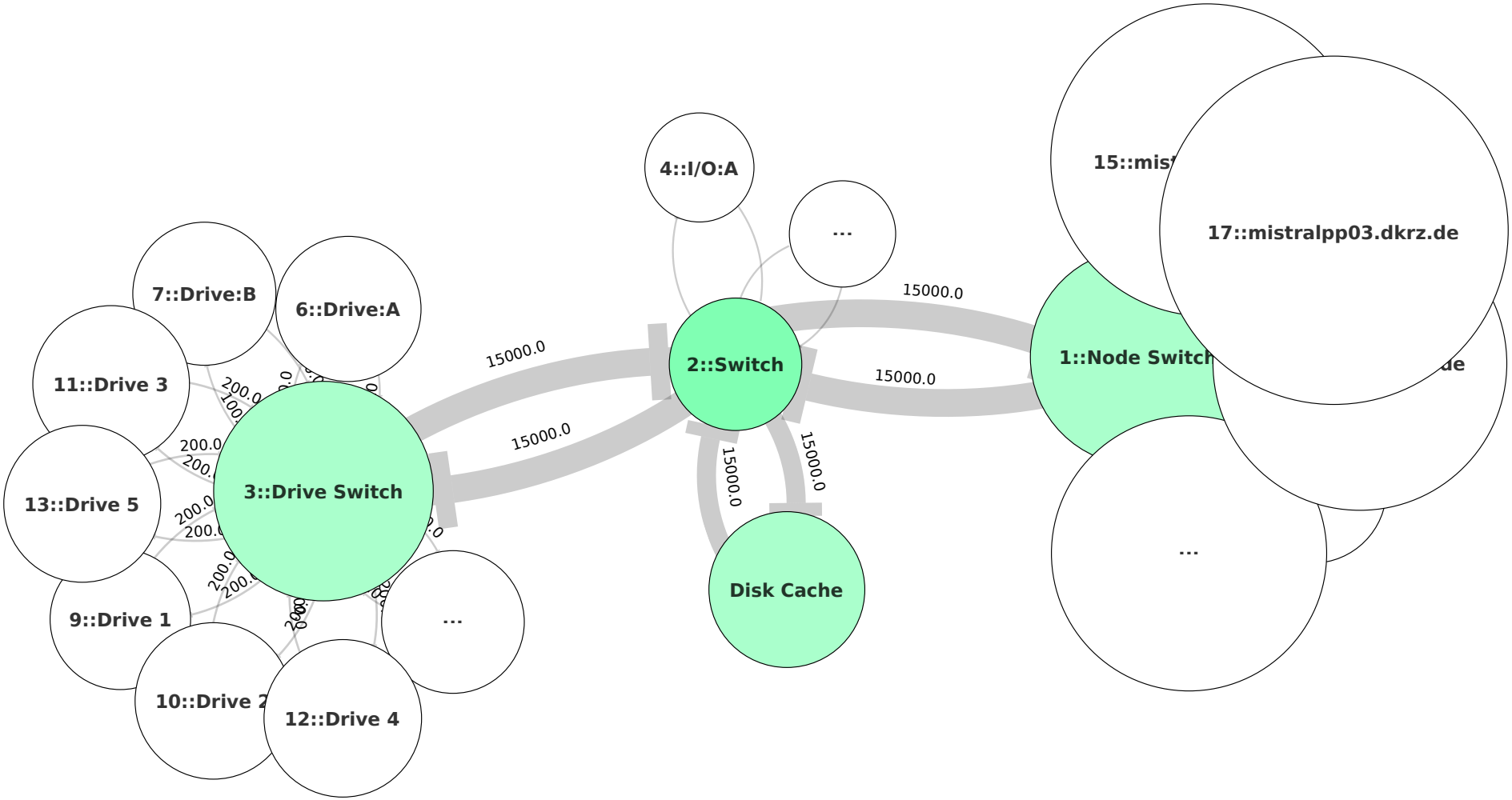
Two-Phase write with delayed persistence on tape.





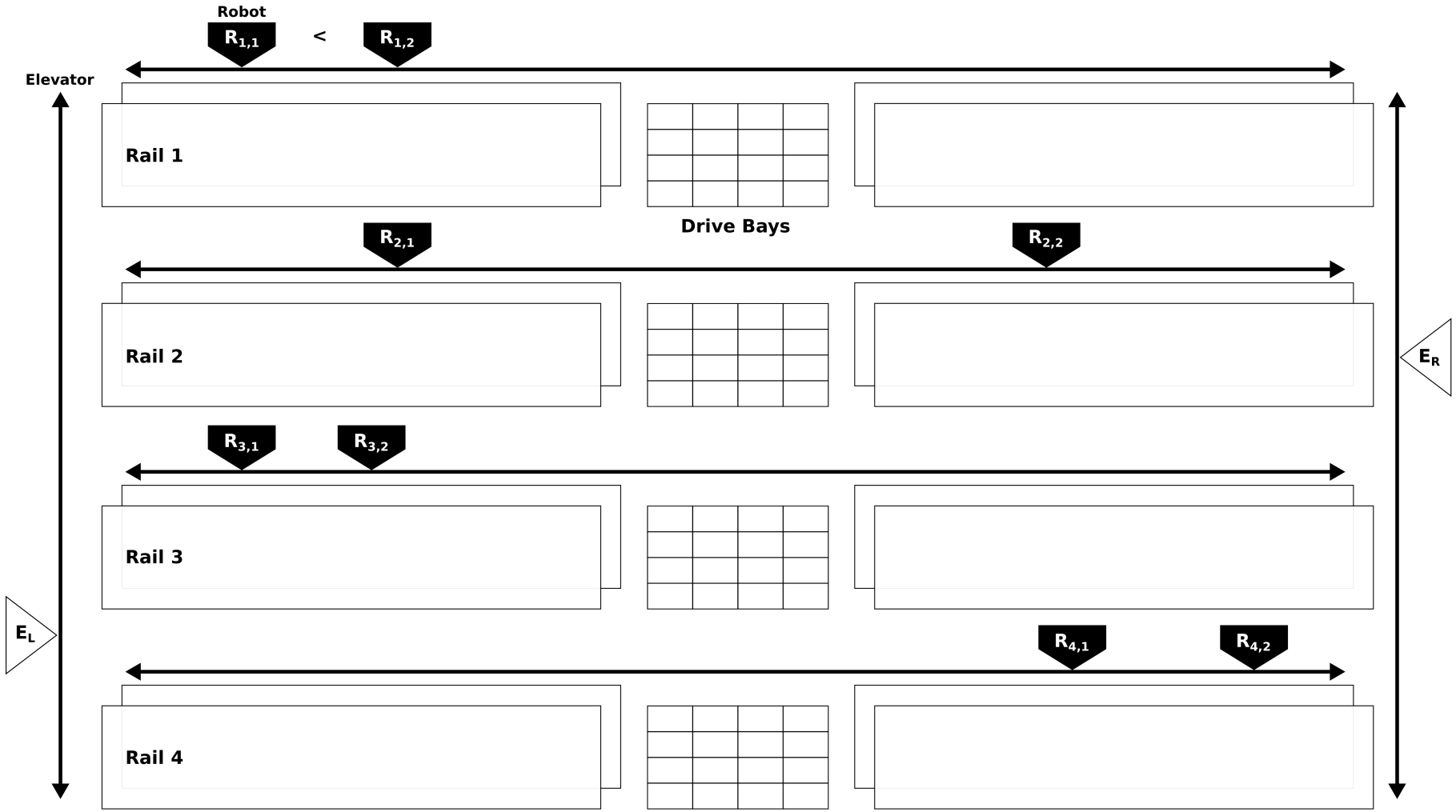
Network Topology Model

Not packet based but allocated for the duration of a transfer.



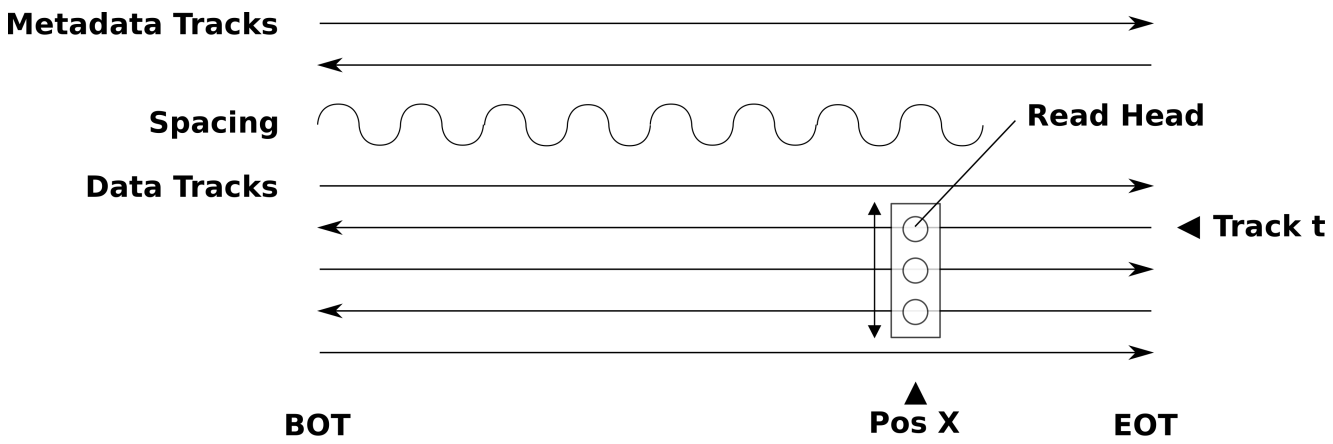
Robot Tape Library

Example: Model of the SL8500 library with robot hands and elevators.



Serpentine Tape Model

Estimating spool and seek times for tape access.



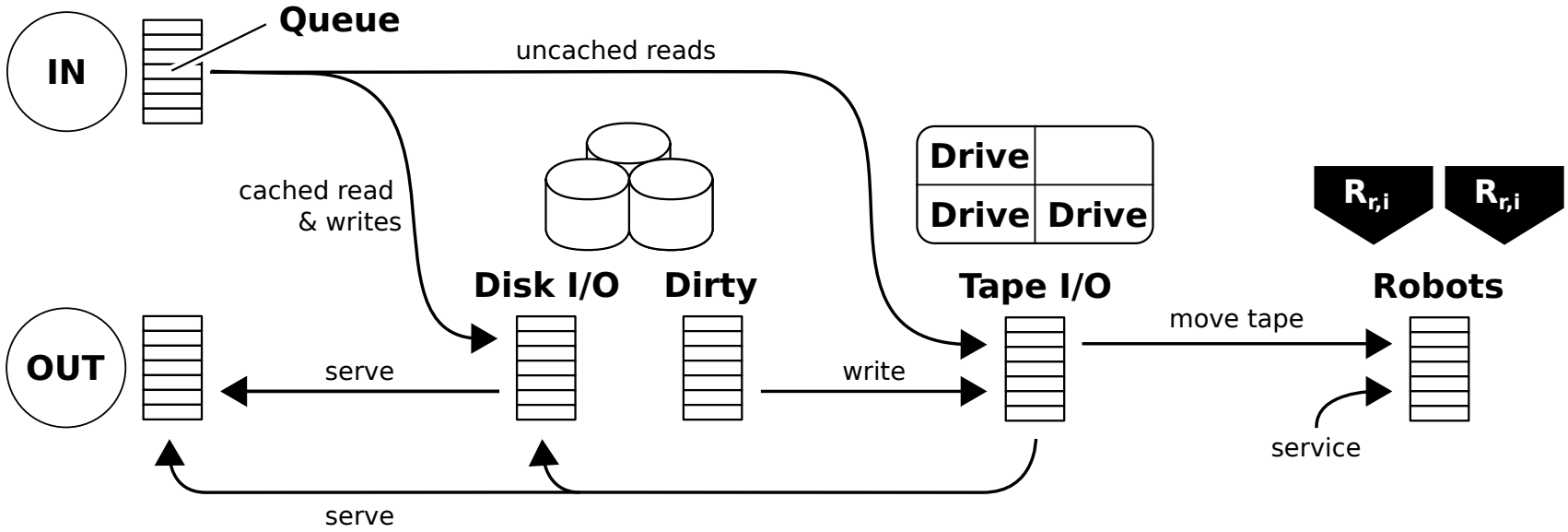
$$T_{seek}(pos_j, pos_i) = max\left(\frac{|pos_{ix} - pos_{jx}|}{v_{spool}}, \frac{|pos_{it} - pos_{jt}|}{v_{head}}\right)$$

$$T_{read/write}(bytes) = \frac{bytes}{v_{read/write}}$$

$$T_{busy} = T_{mount} + \left(\sum_{pos_i, pos_{i+1}}^{BOT,\ldots,BOT} T_{seek}(pos_i, pos_j) + T_{read/write}(bytes_i)\right) + T_{unmount}$$

Scheduling and Request Queues

Chaining specialized request queues makes resource allocation manageable.

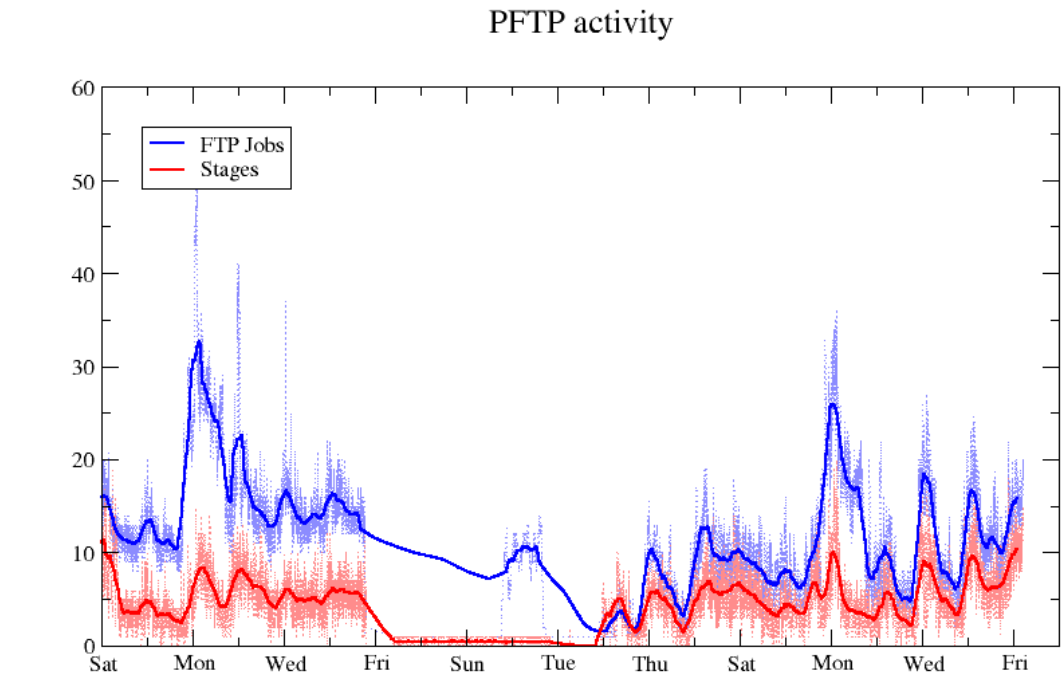


1. Motivation and Background

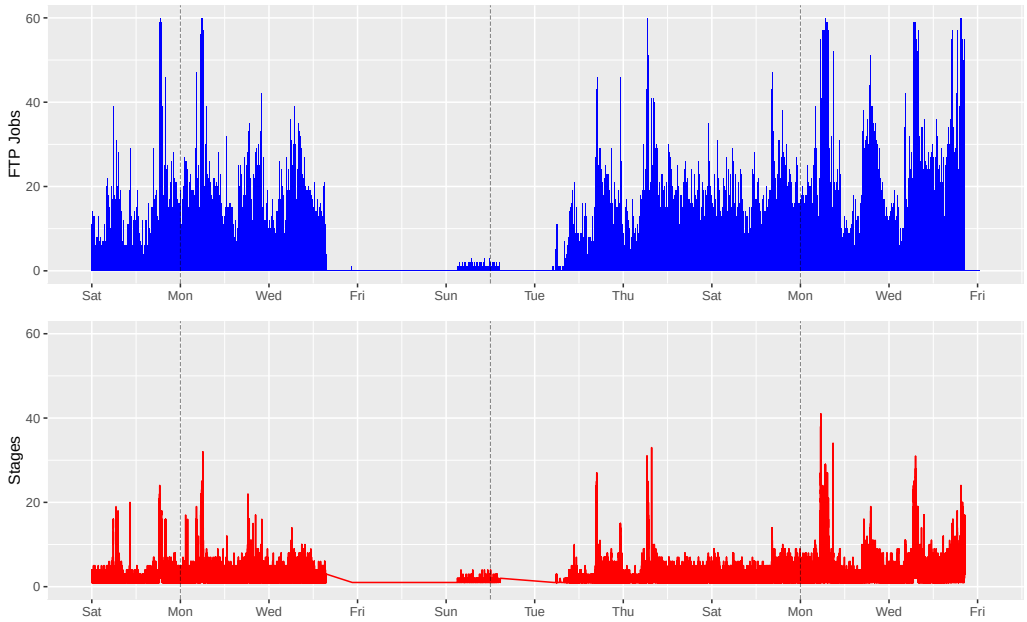
2. Modeling and Simulation Tape Storage Systems

3. Evaluation

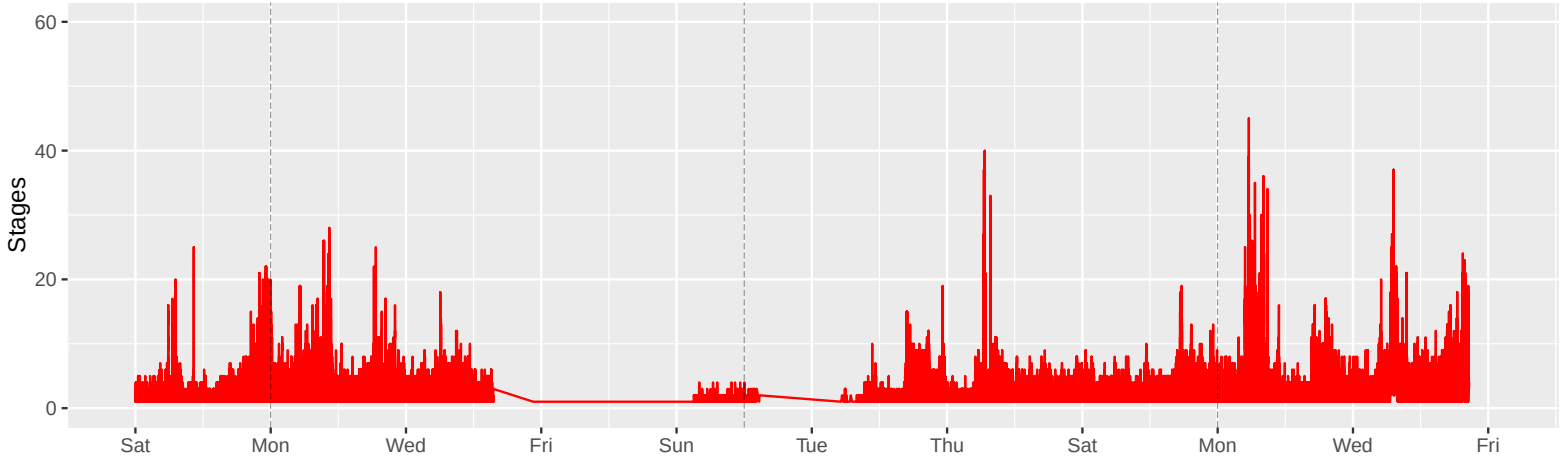
4. Conclusion / Discussion



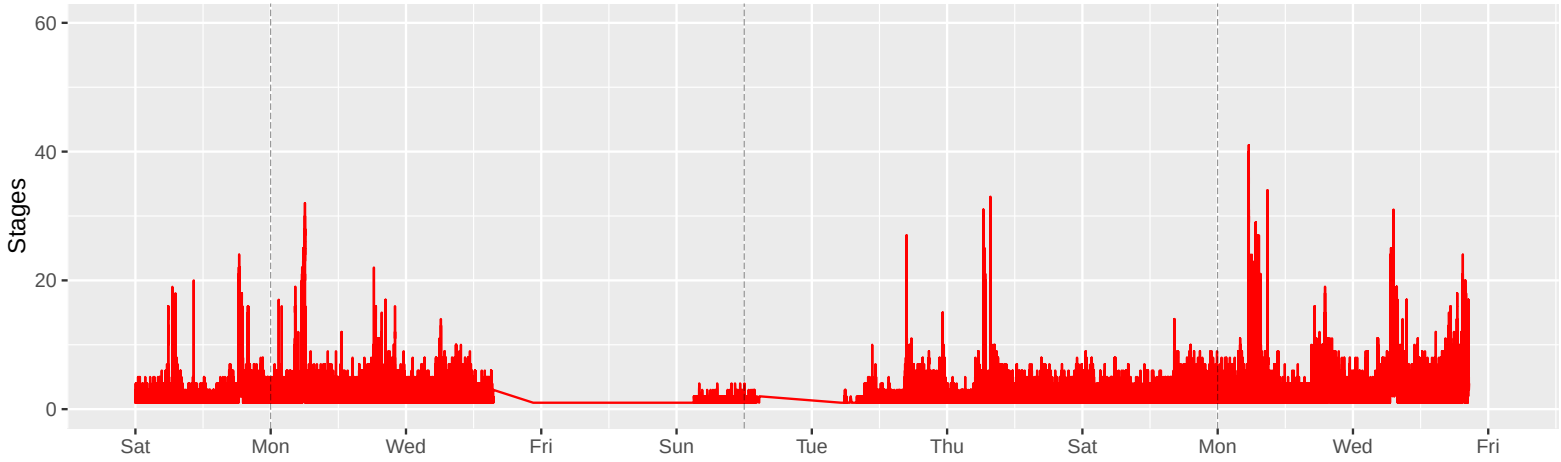
Fri Feb 5 16:40:07 2016



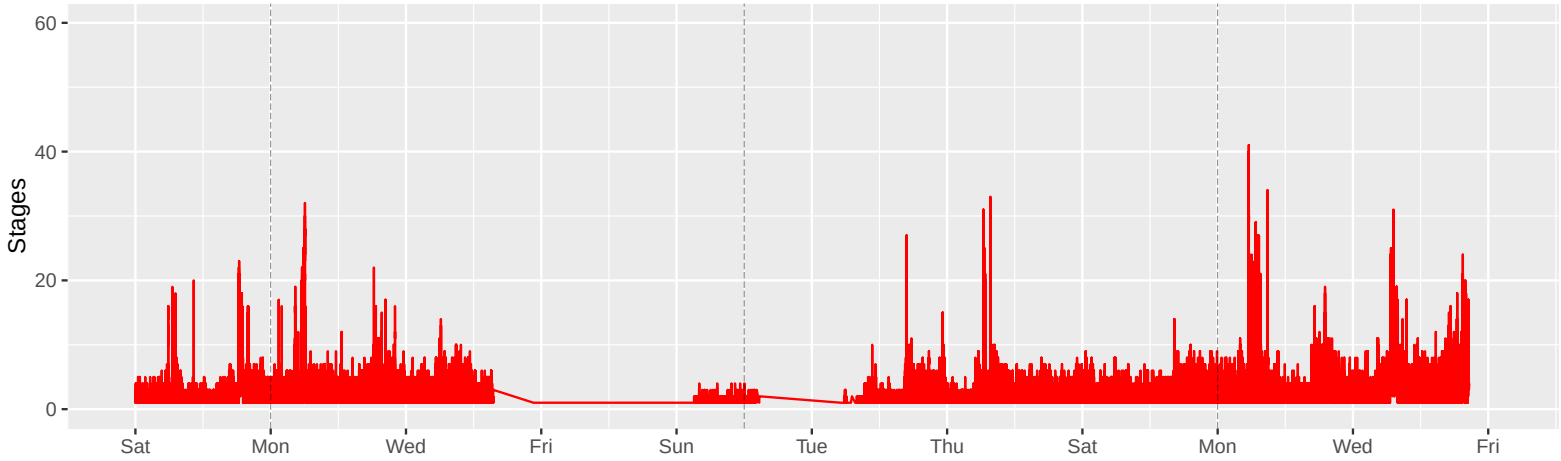
30 drives



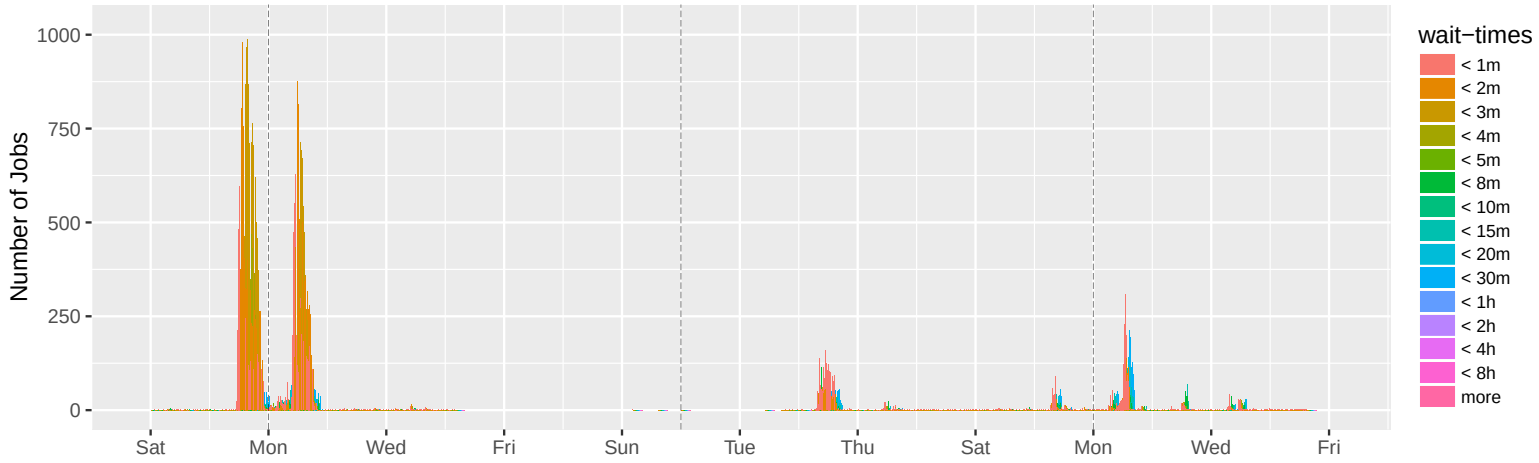
45 drives



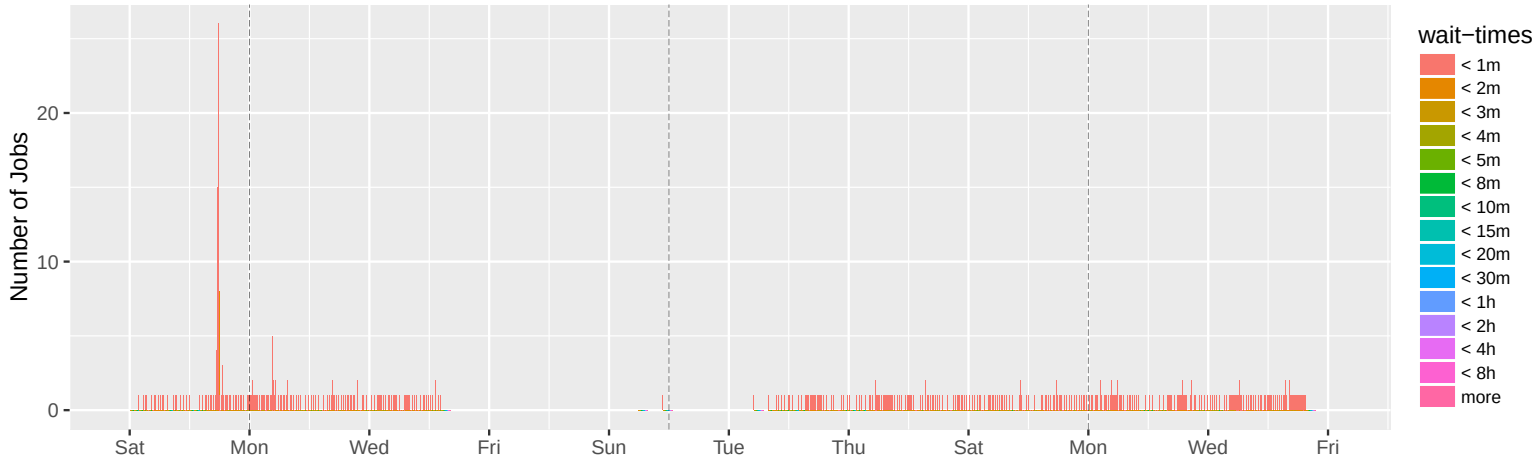
75 drives



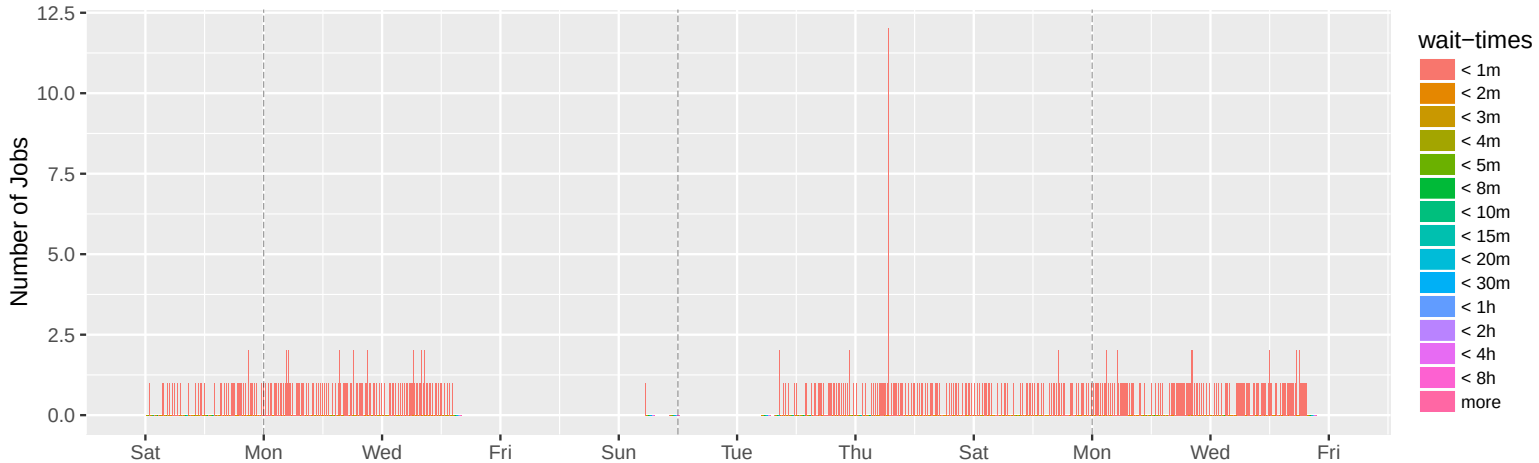
30 drives



45 drives

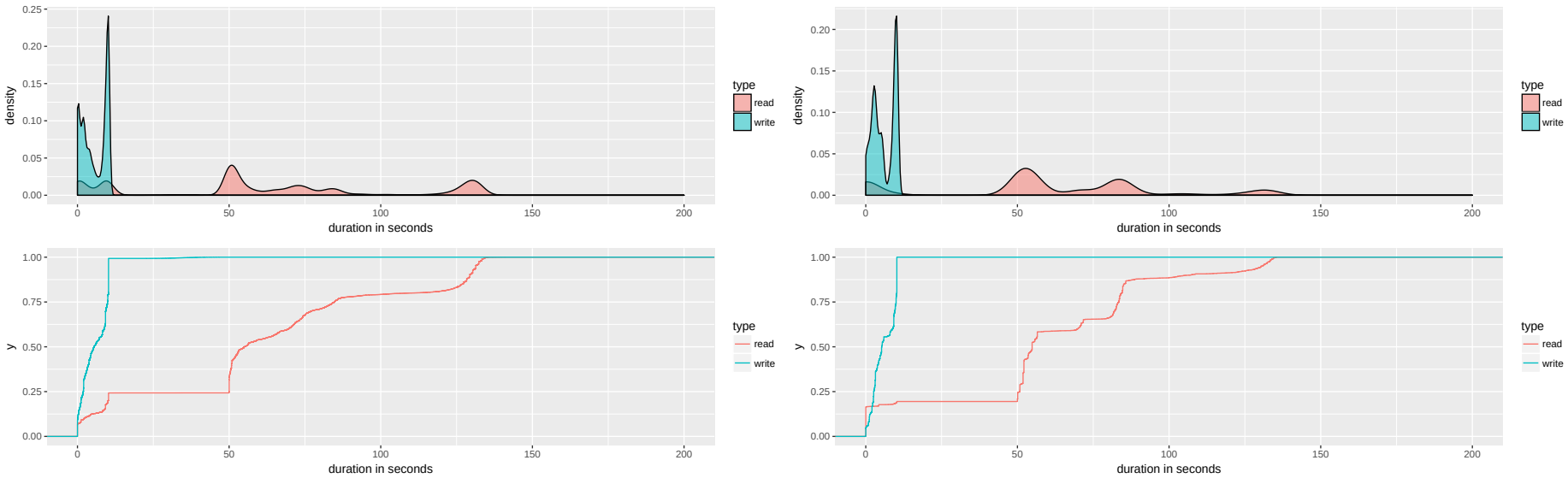


75 drives



Example: QoS for Total-Waittime

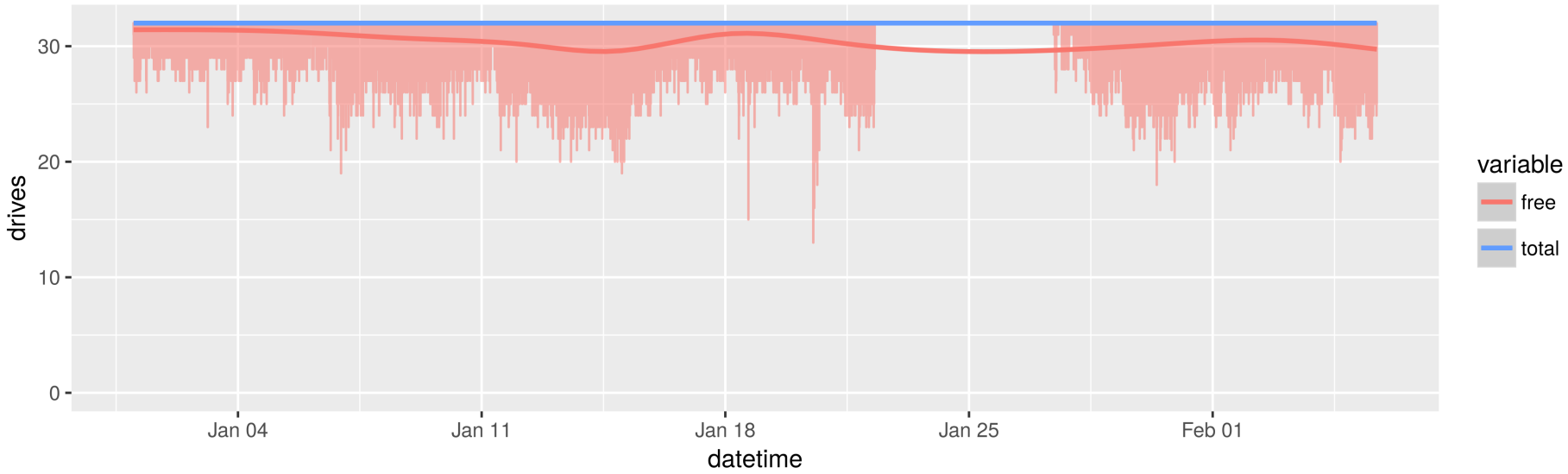
E.g.: How many drives to serve x % of requests in under y minutes.



e.g. fewer vs. more drives

Example: Power Consumption Estimates

Approximate power consumption for different configurations.



Conclusion and Discussion

Summary

- Potential helper for procurement of HSM and tape systems
- Enabler for open research otherwise only performed by vendors
- Depth of stack makes development of DES very complicated

Future Work

- Fine-tuning of various simulation parameters
- Conducting experiments related drive and tape placement, LTO generation, disk cache capacities and RAIT deployments
- Comparison with different deployments at other sites

Bibliography I

Fontana, R. E., Decad, G. M., and Hetzler, S. R. (2013). The Impact of Areal Density and Millions of Square Inches (MSI) of Produced Memory on Petabyte Shipments of TAPE , NAND Flash , and HDD Storage Class Memories. *Proceedings of the 29th IEEE Symposium on Massive Storage Systems and Technologies*.

IBM (2011a). High Performance Storage System. Technical report.

IBM (2011b). IBM System Storage TS3500 Tape Library Connector and TS1140 Tape Drive support for the IBM TS3500 Tape Library. pages 1–15.

Oracle (2015). StorageTek SL8500 Modular Library System User's Guide.

Quantum (2015). Quantum Scalar i6000 Datasheet.

Spectralogic (2016a). LTO Roadmap. `https://www.spectralogic.com/features/lto-7/`. [Online; accessed 2016-01-24].

Spectralogic (2016b). Spectralogic TFinity - Enterprise Performance. `https://www.spectralogic.com/products/spectra-tfinity/tfinity-features-enterprise-performance/`. [Online; accessed 2016-02-12].

Sun (2006). StorageTek StreamLine SL8500 - User Guide. (96154).