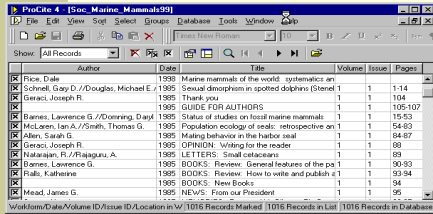


Jump Distance Based Synthetic Disk Access Patterns

Zachary Kurmas and Jeremy Zito
Grand Valley State University

New storage systems must be evaluated with respect to many different workloads

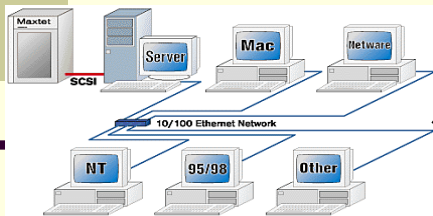


Author	Date	Title	Volume	Issue	Pages
Ross, Dale	1995	Maine mammals of the world. systematic art	1	1	1-14
Schreff, Gary D./Douglas, Michael E.	1995	Sexual dimorphism in spotted dolphins (Stenella)	1	1	104
Genaco, Joseph R.	1995	Thank you	1	1	109-107
Genaco, Joseph R.	1995	GUIDE FOR AUTHORS	1	1	15-63
Barnes, Lawrence G./Donning, David	1995	Status of studies on fossil marine mammals	1	1	54-63
McLaren, Ian A./Smith, Thomas G.	1995	Population ecology of seals: retrospective on	1	1	84-87
Allen, Sarah G.	1995	Mating behavior in the harbor seal	1	1	88
Genaco, Joseph R.	1995	OPINION: Writing for the reader	1	1	89
Husaregan, R./Ragaku, A.	1995	LETTERS: Small cetaceans	1	1	90-93
Barnes, Lawrence G.	1995	BOOKS: Review: General features of the pa	1	1	92-94
Rubb, Katherine	1995	BOOKS: Review: How to write and publish a	1	1	94
Mead, James G.	1995	NEWS: From our President	1	1	95

Database workload



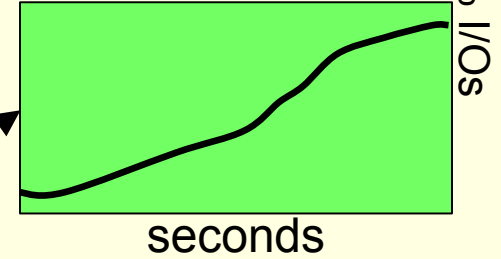
Email server workload



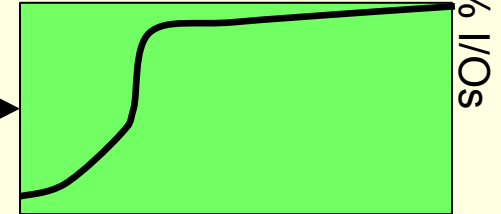
File server workload



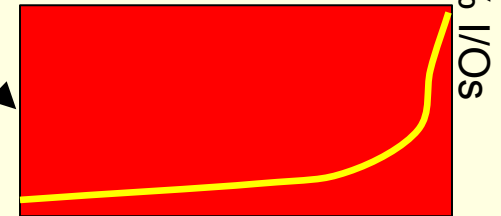
Disk array



seconds



seconds



seconds

Example Workloads

Performance (CDF of latency)

An design may be **good** for some workloads and **bad** for others.

Sources of Evaluation Workloads

Trace of real workloads

- List of I/O requests made by production workload

- **Accurate**

- **Large**

- **Inflexible**

- **Difficult to obtain
(due to security concerns)**



Synthetic workloads

- Randomly generated model of some workload

- **Usually Inaccurate**

- **Compact representation**

- **Easily modified**

- **Compact representation
contains no specific data**

My goal: Accurate synthetic workload

Synthetic Workload

```
(R, 1024, 120932, 124)
(W, 8192, 120834, 126)
(W, 8192, 120844, 127)
(R, 2048, 334321, 131)
(R, 1024, 120932, 124)
(W, 8192, 120834, 126)
(W, 8192, 120844, 127)
(R, 2048, 334321, 131)
(R, 1024, 120932, 124)
(W, 8192, 120834, 126)
(W, 8192, 120844, 127)
(R, 2048, 334321, 131)
(R, 1024, 120932, 124)
(W, 8192, 120834, 126)
(W, 8192, 120844, 127)
(R, 2048, 334321, 131)
...
```

Workload Trace

```
(R, 1024, 120932, 124)
(W, 8192, 120834, 126)
(W, 8192, 120844, 127)
(R, 2048, 334321, 131)
(R, 1024, 120932, 124)
(W, 8192, 120834, 126)
(W, 8192, 120844, 127)
(R, 2048, 334321, 131)
(R, 1024, 120932, 124)
(W, 8192, 120834, 126)
(W, 8192, 120844, 127)
(R, 2048, 334321, 131)
(R, 1024, 120932, 124)
(W, 8192, 120834, 126)
(W, 8192, 120844, 127)
(R, 2048, 334321, 131)
...
```

Shared properties

“Random” or “Fake” Workload

Mean Request Size: 8192

Mean interarrival Time: 04ms

Read Percentage: 78%

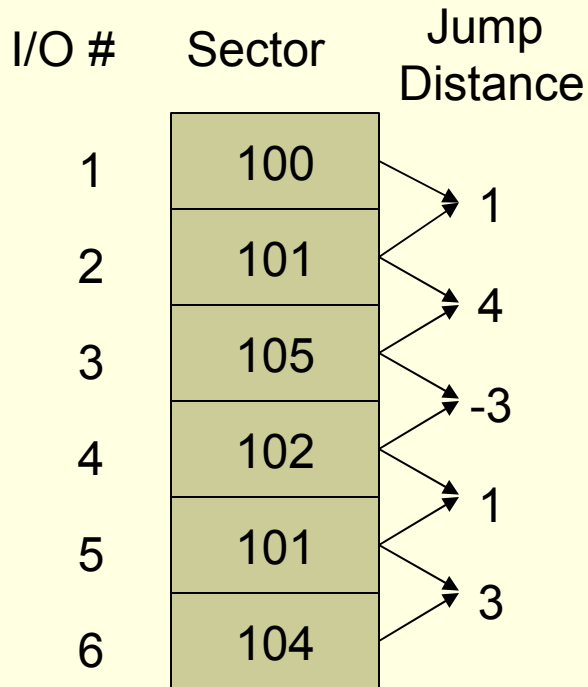
Local distribution:

(.01, .02, .09, .14, .03, .12, ...)

“Real” workload

\$64,000 question:
What goes in here?

Jump Distance



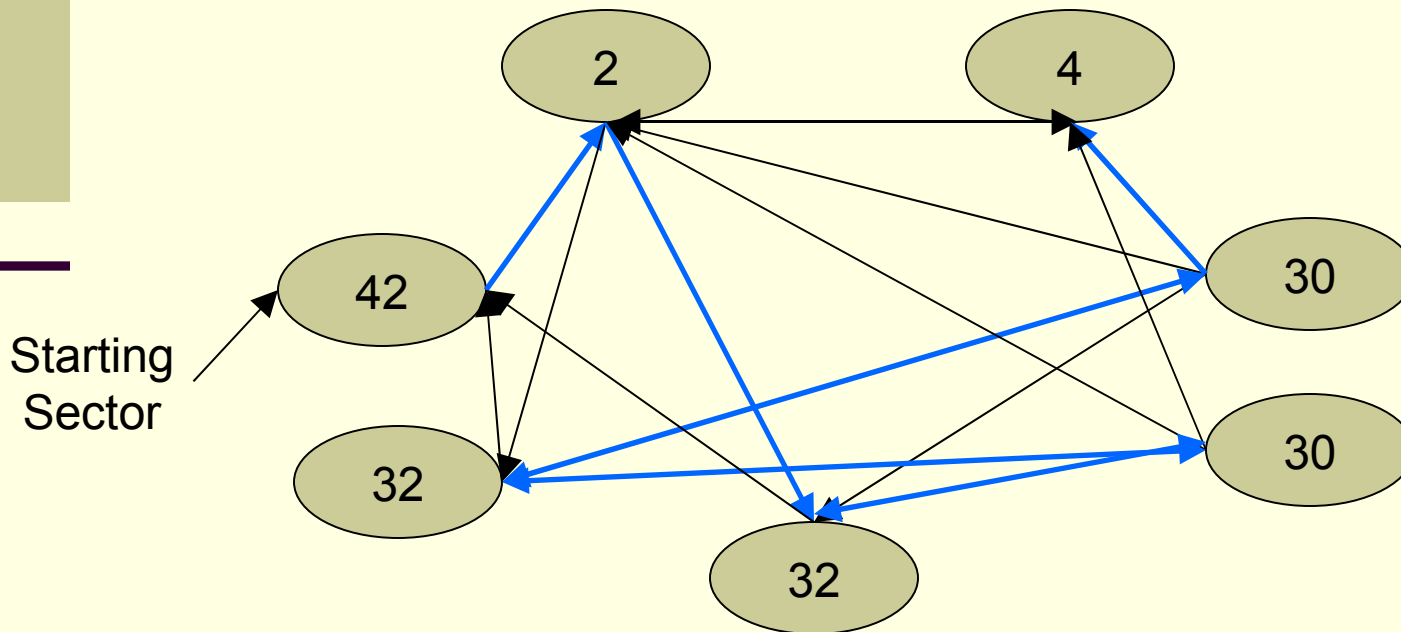
- Our poster presents a synthetic workload generation technique called “Jump Distance”
- Each jump distance approximates distance traveled by disk head
- Goal: Synthetic and target workloads make disk heads move the same amount

Distribution of Observed Jump Distances

Frequency:	1	2	1	1
Jump Distance:	-3	1	3	4

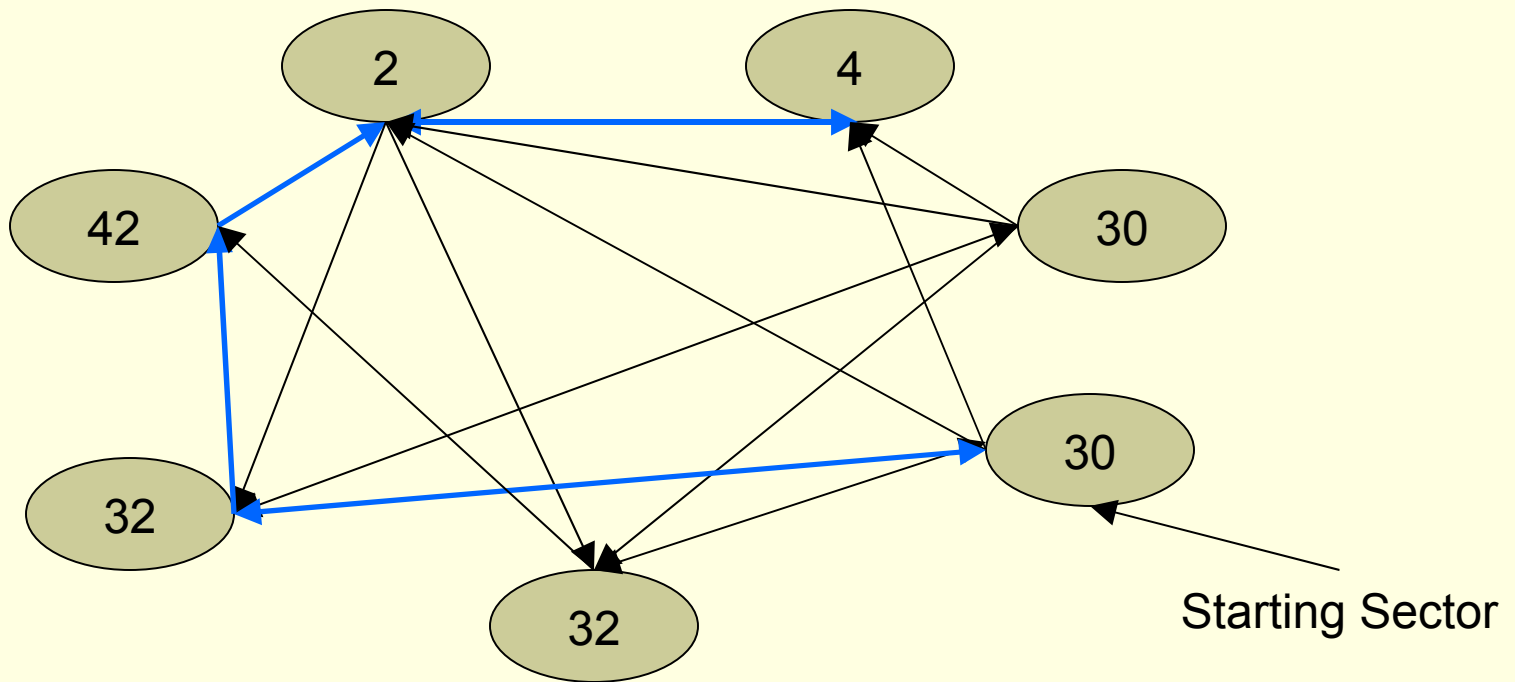
Solution

- Perform depth-first search for Hamiltonian Path
- Complete path represents synthetic disk access pattern with identical jump distance and location distributions



Solution #2

- Partial path completed randomly for an approximate solution



Results

- Partial paths average just under 80% of maximum length
- Resulting Jump Distance distributions not perfectly accurate, but reasonably accurate
- Results in reasonably accurate synthetic workloads --- if workload behavior dominated by disk head movement