

# Archival Storage at NCAR: A Tale Across Twenty Years

Ethan L. Miller

Storage Systems Research Center  
University of California, Santa Cruz

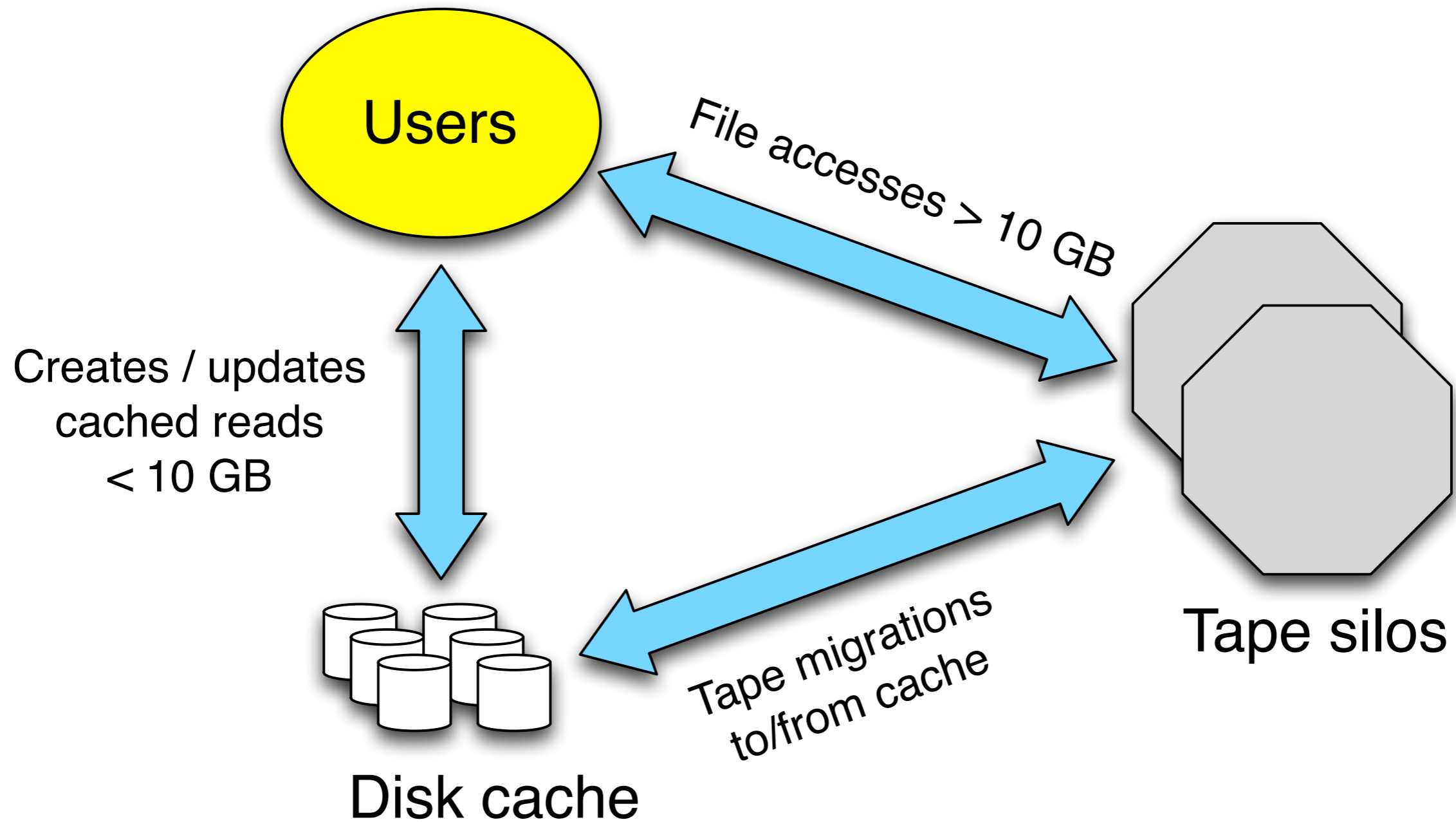


# How has HPC archive usage changed over the past two decades?

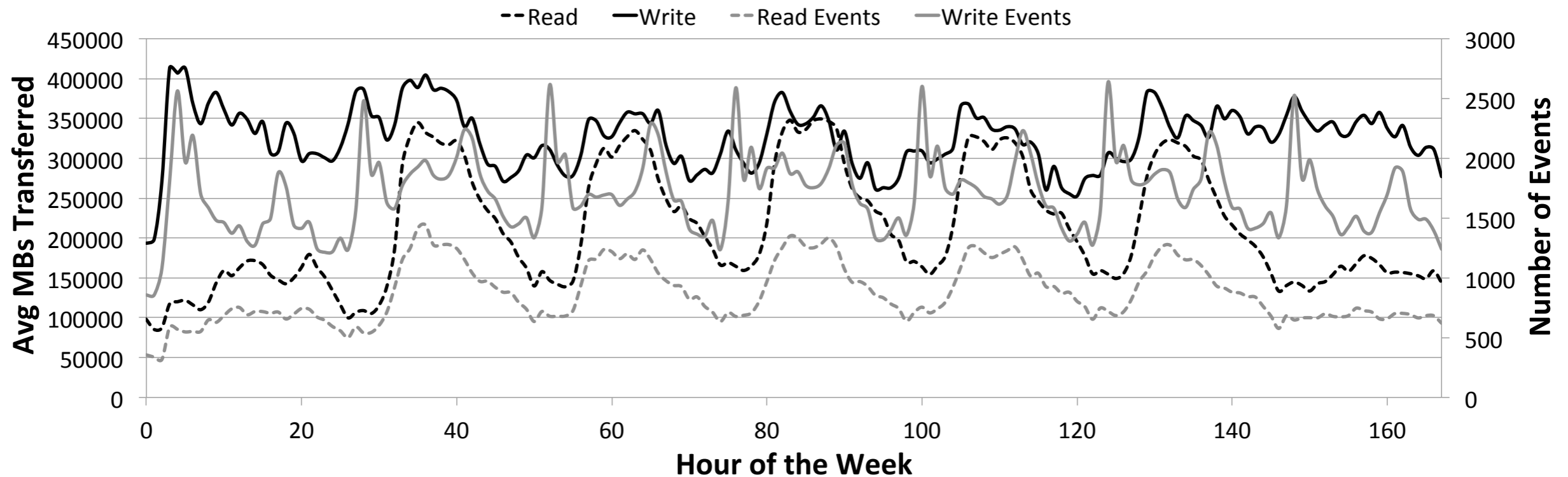
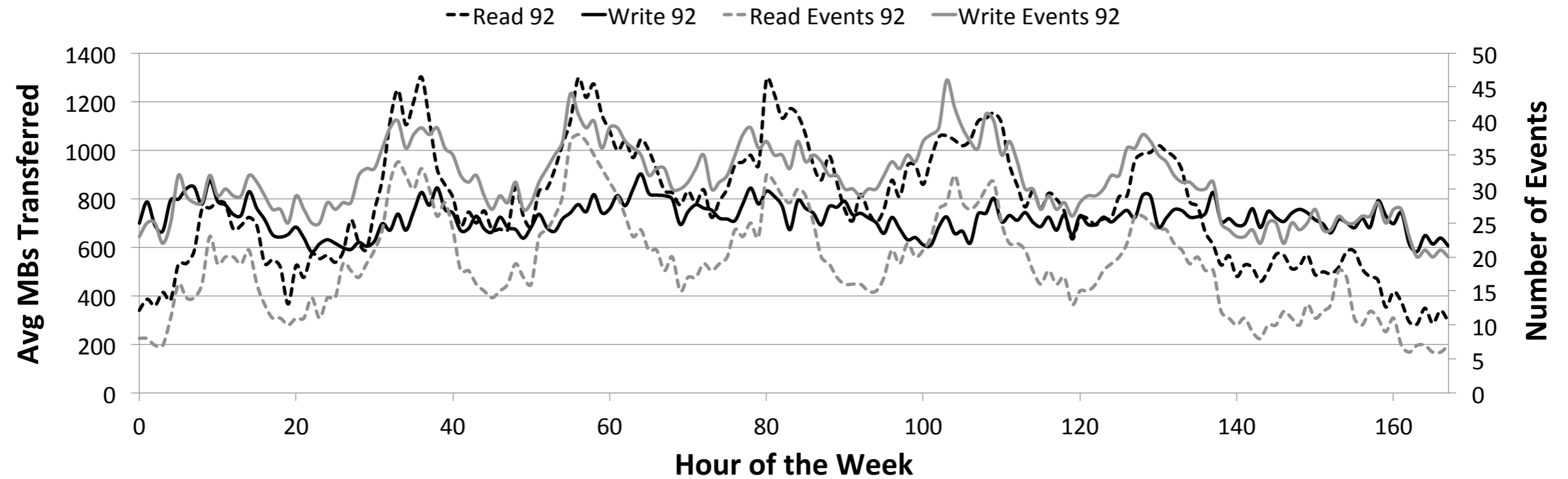
---

- 1993: studied archival storage workload at NCAR
  - Archive held 25 TB of data (!)
- 2011: studies archival storage workload at NCAR
  - How are people using archives now?
  - What's changed?
  - What's stayed the same?
- Examine
  - Usage patterns
  - File sizes
  - Performance

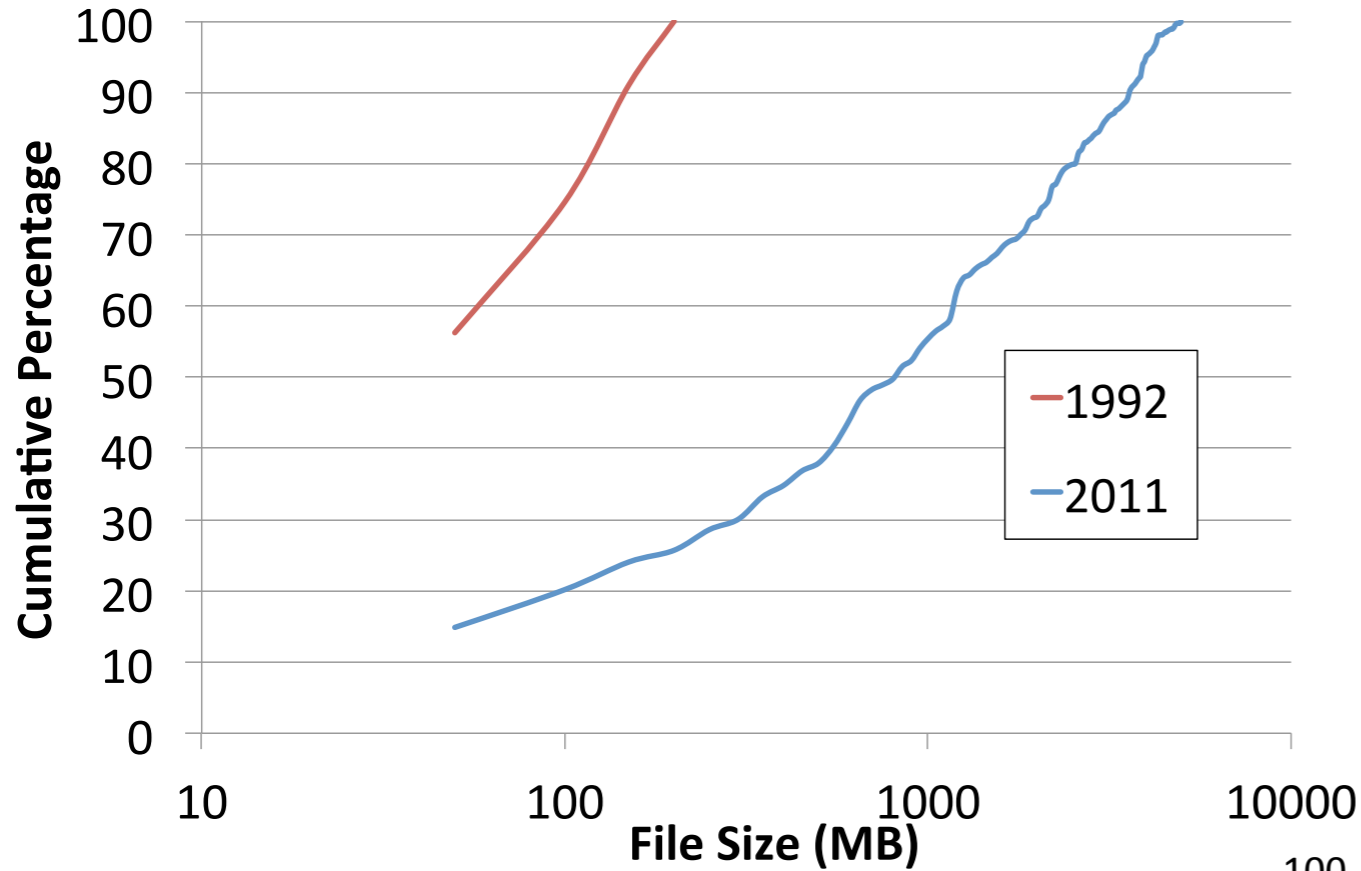
# NCAR archival storage system



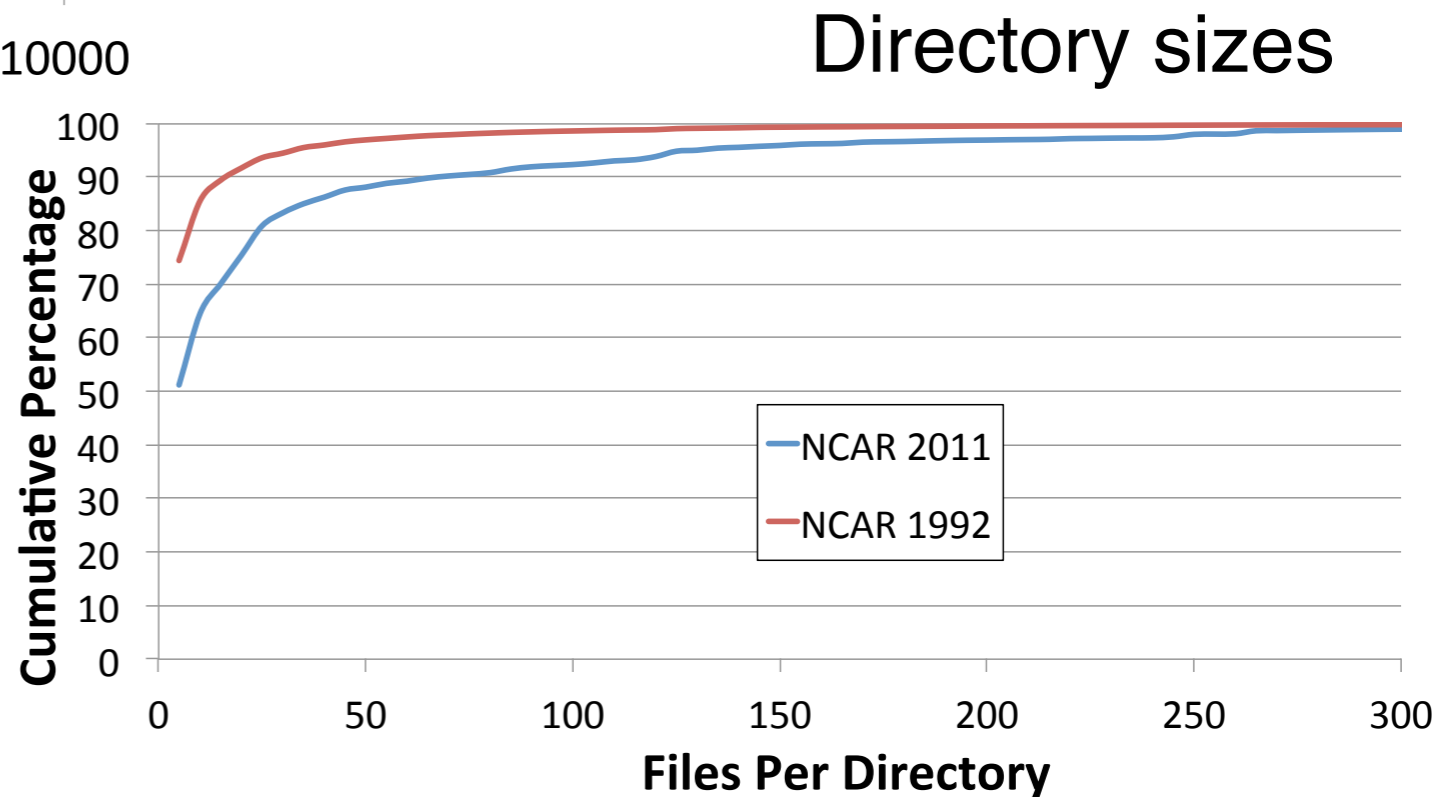
# Trends in overall system activity



# File & directory sizes

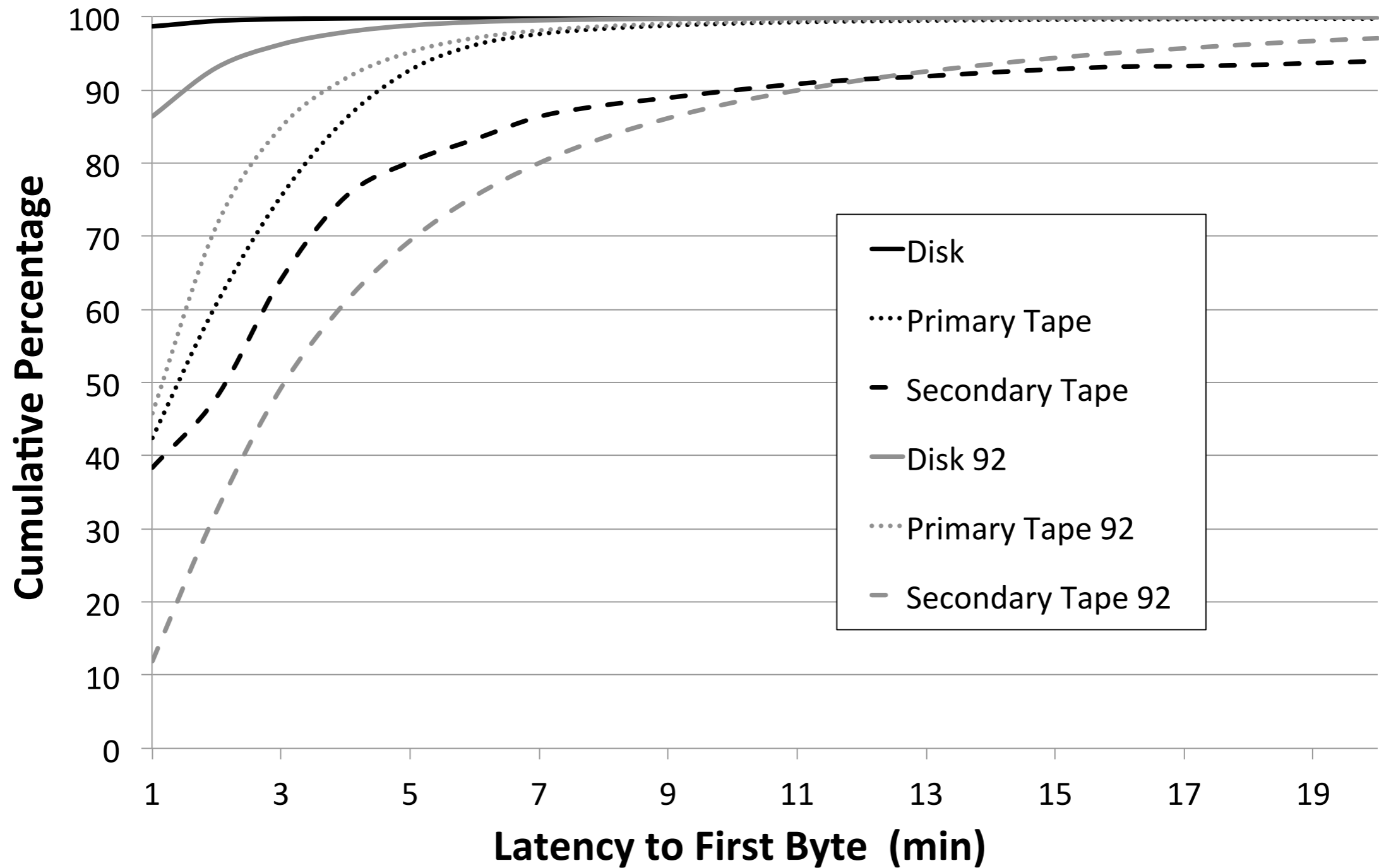


Archive space consumed by file size

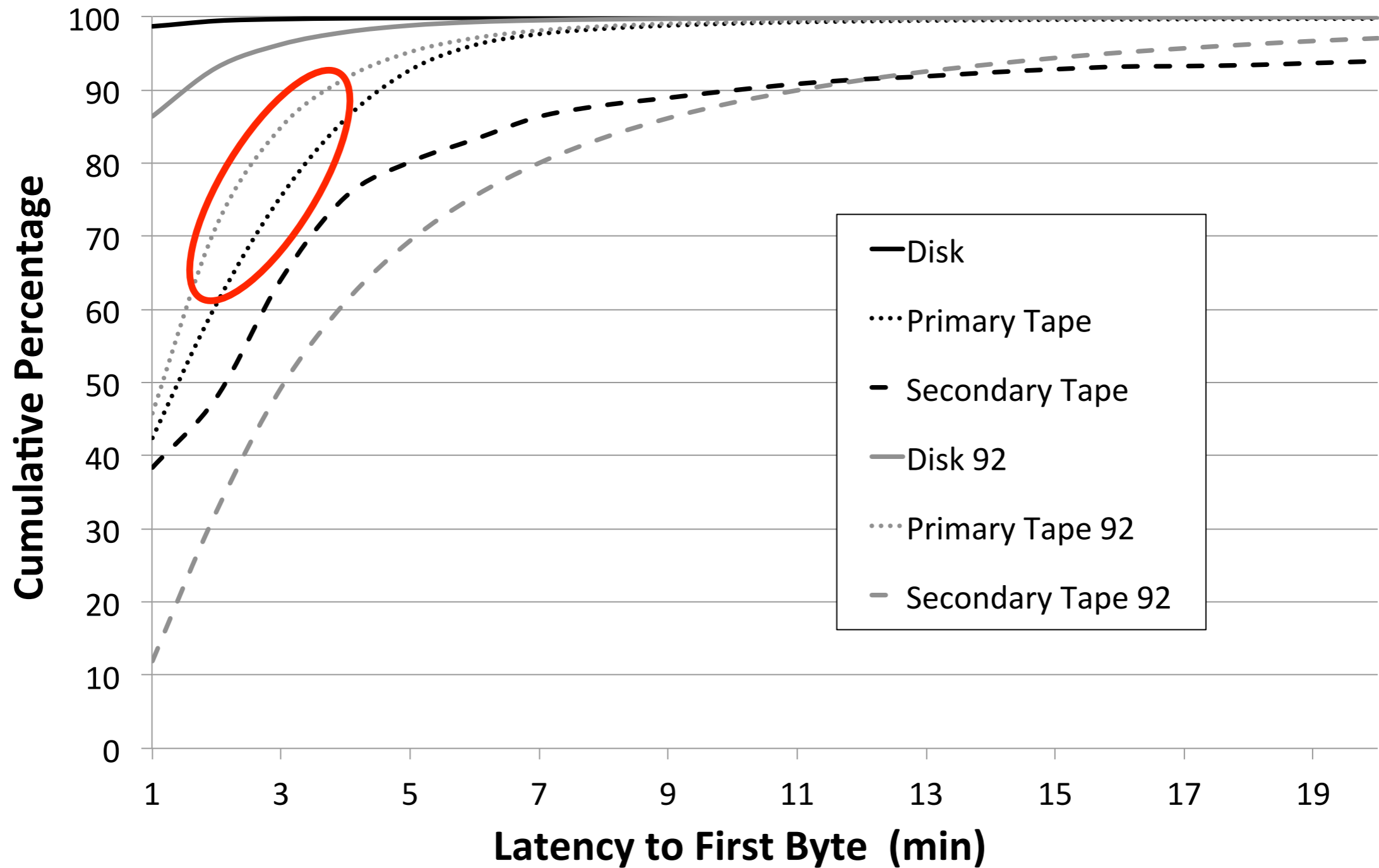


Directory sizes

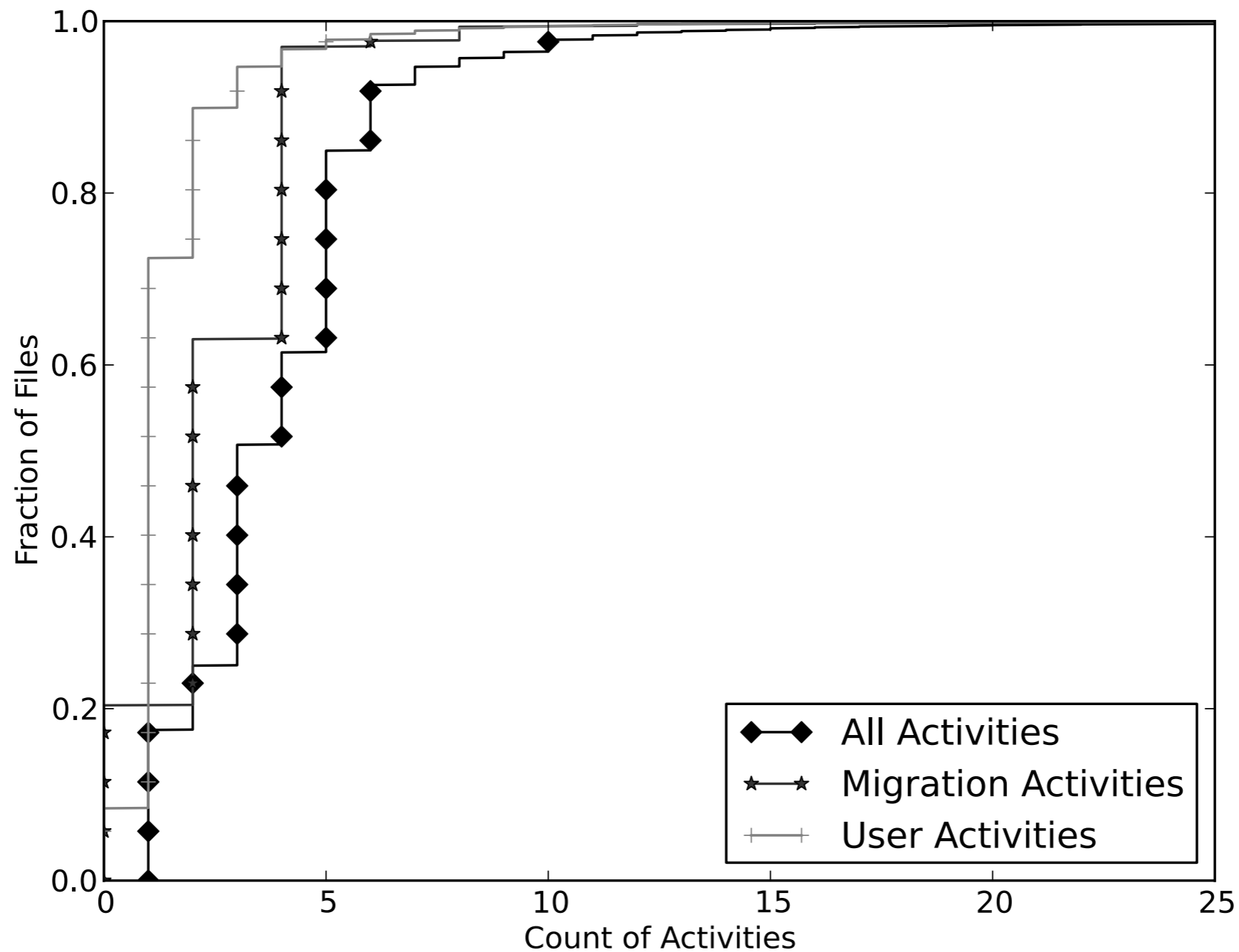
# Performance



# Performance

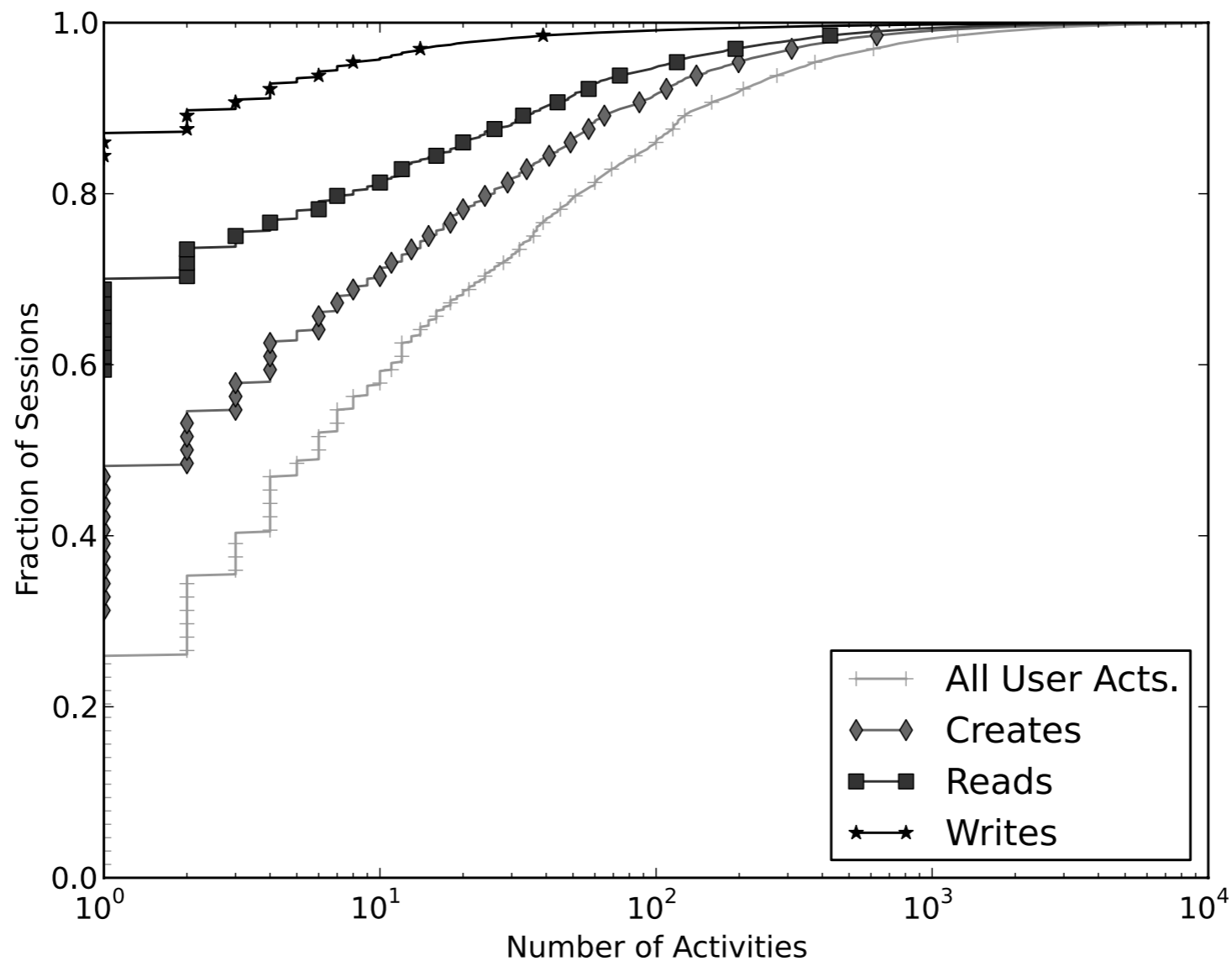


# How often are files accessed?





# How big are user sessions?



- User sessions were denoted by actions within 15 minutes of another action

# So what did we learn?

---

- Over the past two decades
  - Files got bigger
  - Directories have only moderately more files
  - Most (but not all!) latencies got better
  - File activity over the week is still cyclical, but...
- Read-write ratio has gone down by a factor of 4 (!)
- In the most recent study,
  - Most files that are accessed at all have only a few accesses
  - Many user sessions access fewer than 100 files

# Questions?

---

## Thanks to:

Ian Adams

Joel Frank

Gene Harano

Brian Madden

Daniel Rosenthal

Mark Storer

Funded by  and  industrial sponsors.