Striped Tape Arrays

Ann L. Drapeau

Computer Science Department
University of California
571 Evans Hall
Berkeley, CA 94720

alc@cs.berkeley.edu
Striped Tape Arrays

Ann L. Drapeau
alc@cs.berkeley.edu

Motivation

- Applications require high throughput (100 MB/sec), massive storage (Terabytes, Petabytes)
- Technology Trends
  - Magnetic tape: high capacity, low bandwidth
  - Robots: automatic loading of tape cartridges
- Striping: a technique for increasing throughput
- Issues in striping effectively
- Tape array reliability

Outline

- Introduction to Striping
- Applications
- Tape Technologies
- Robots
- Access Times
  - Drive and Robot Measurements
- Striping Options and Issues
- Reliability Issues
- Summary

Data Striping

- Spread data from individual files across several devices
- Advantages:
  - Increase bandwidth to a single file
  - Reduce latency of large accesses
  - Allows independent "smaller" accesses
  - Easy to incorporate error correction
- Problems:
  - Increase latency of some accesses
  - Synchronization
Do Applications Need Striped Tape?
- Large scientific archives (NASA EOS)
  - High sustained bandwidth (100 MB/s)
  - Total storage very large (Petabytes)
  - Would benefit from striping throughput
- Interactive access to large data sets (Sequoia)
  - Researchers across California
  - Want reasonable response time over network
  - Total storage large (Terabytes)
  - Striping would reduce large access latency

<table>
<thead>
<tr>
<th>Tape Technologies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Technology</strong></td>
</tr>
<tr>
<td>1/4&quot;</td>
</tr>
<tr>
<td>1/2&quot; 3480</td>
</tr>
<tr>
<td>4mm DAT</td>
</tr>
<tr>
<td>8mm Exabyte</td>
</tr>
<tr>
<td>1/2&quot; Metrum VLDS</td>
</tr>
<tr>
<td>Ampex DD2</td>
</tr>
</tbody>
</table>

Linear Recording: 1/4" cartridge, 1/2" 3480
Helical Scan: DAT 4mm, 8mm, 1/2" VLDS, 19mm D2

Tape Tradeoffs: No "Perfect" Drive
- Inexpensive helical scan drives have low bandwidth (DAT, 8mm)
- Inexpensive serpentine drives have moderate bandwidth (1/4")
- High capacity drives have long access times (helical scan, 1/4")
- Drives with short access times are low capacity (1/2" 3480)
  - Moderate price and bandwidth
- High bandwidth drives very expensive (DD2)
  - Bandwidth not high enough
  - Very high capacity
Future Tape Drives (8mm)

- **Bandwidth (MB/sec)**
  - 1990: 1
  - 2000: 6

- **Cartridge Capacity (GBytes)**
  - 1990: 10
  - 2000: 70

- **Source:** Harry C. Hinz, Exabyte Corp.
- **Changes:** increase track density, decrease track width & pitch, reduce tape thickness, increase rotor speed

Robots

- **Large Libraries:**
  - many cartridges, several drives
  - expensive
  - one or more robot arms

- **Carousels**
  - around 50 cartridges, one or two drives
  - moderate cost

- **Stackers**
  - around 10 cartridges, one drive
  - inexpensive

Tape Access Time (Cartridge Switch)

- **Access time =**
  - rewind time +
  - eject time +
  - robot unload +
  - robot load +
  - device load +
  - fast search +
  - transfer time

- Measured three tape drives, one robot:
  Accurate access time models for simulation
Drive Measurements

Drive Load and Eject Times

<table>
<thead>
<tr>
<th></th>
<th>4mm DAT</th>
<th>8mm Exabyte</th>
<th>Metrum VLDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Load Time (sec)</td>
<td>16</td>
<td>35.4</td>
<td>28.3</td>
</tr>
<tr>
<td>Mean Eject Time (sec)</td>
<td>17.3</td>
<td>16.5</td>
<td>3.8</td>
</tr>
</tbody>
</table>

Data Transfer Rates

<table>
<thead>
<tr>
<th></th>
<th>4mm DAT</th>
<th>8mm Exabyte</th>
<th>Metrum VLDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read Rate (MB/sec)</td>
<td>0.17</td>
<td>0.47</td>
<td>1.2</td>
</tr>
<tr>
<td>Write Rate (MB/sec)</td>
<td>0.17</td>
<td>0.48</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Rewind and Search Behavior

<table>
<thead>
<tr>
<th></th>
<th>4mm DAT</th>
<th>8mm Exabyte</th>
<th>Metrum VLDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rewind Startup (sec)</td>
<td>15.5</td>
<td>23</td>
<td>15</td>
</tr>
<tr>
<td>Rewind Rate (MB/sec)</td>
<td>23.1</td>
<td>42.0</td>
<td>350</td>
</tr>
<tr>
<td>Search Startup (sec)</td>
<td>8</td>
<td>12.5</td>
<td>28</td>
</tr>
<tr>
<td>Search Rate (MB/sec)</td>
<td>23.7</td>
<td>36.2</td>
<td>115</td>
</tr>
</tbody>
</table>

- Constant startup
- Approximately linear search/rewind

Tape Access Time Example
(Exabyte EXB8500 Drive, EXB-120 Robot)

- Average Access time =
  - rewind time (1/2 tape) (75 sec) +
  - eject time (17 sec) +
  - robot unload (21 sec) +
  - robot load (22 sec) +
  - device load (35 sec) +
  - fast search (1/2 tape) (84 sec) +
  - transfer time

- Not including data transfer: 4 minutes!

Options for Striped Tape

- Within a robot
  + cartridges in stripe kept together
  - few readers, robot arms
  - single point of failure

- Between robots
  + several robot arms used in access
  - harder to keep cartridges together

- Between small robots (stackers)
  + highest proportion arms to readers and cartridges
Striping Issues

- Configuration depends on workload
- Interleave factor crucial:
  - Too small: cartridge switches increase latency
    (Long access times — big penalty)
  - Too big: lose potential parallelism
- Workloads that will benefit from striping
  - Large archives
  - Interactive systems with large avg. request size
- Stripping will hurt performance of some accesses
  - Interleave smaller than average request
  - High load/scarce readers

More Striping Issues

- Striping with improved devices/robots
  - Higher bandwidth drives
    - Bandwidth, aerial density may increase 30X
      by end of decade
    - Less need for striping?
    - Still get throughput benefits
  - Faster access times (drives and robots)
    - faster load, eject, search, rewind, robot arms
    - no rewind before eject
    - cartridge switch penalties reduced
    - striping more effective

Synchronization Issues

- Drives retry after failed writes
  - Bad tape would retry indefinitely
  - Pat Savage (Shell Oil): after write error, retry on
    all tapes in stripe
- If "RAID-5" (large interleaving)
  - Single cassettes may satisfy smaller requests
    independently
  - Large requests spanning several tapes may be out
    of synchronization by minutes
  - Buffer space required to hold stripe units while
    request completes

Reliability Issues: Tape Media

- High rates of raw bit errors
  - before internal ECC
  - one in 10^5 bits
- Dropouts
  - Debris
    - Slicing of tape
    - Particles in atmosphere
    - Start/stop wear
    - Nonhomogeneous Tape Coating
Uncorrectable Bit Error Rates

<table>
<thead>
<tr>
<th>Drive</th>
<th>Bit Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/4&quot;</td>
<td>$10^{-14}$</td>
</tr>
<tr>
<td>4mm DAT</td>
<td>$10^{-15}$</td>
</tr>
<tr>
<td>Exabyte 8mm</td>
<td>$10^{-13}$</td>
</tr>
<tr>
<td>Memtec VLDS</td>
<td>$10^{-13}$</td>
</tr>
<tr>
<td>Ampex DD2</td>
<td>$10^{-12}$</td>
</tr>
</tbody>
</table>

- Error rates after ECC
- Terabyte approximately $10^{13}$ bits
- MSS will contain uncorrectable errors!

- Need Error Correction
  - Easy to implement in striped systems
  - How much?
  - How reliable are error rates?
  - How will ECC affect performance?

- Error Rates Increase with Wear
- Tapes last around 2000 passes
- Severe wear: tape unreadable

- If tapes are rewritten often, need to copy tapes periodically

Reliability: Tape Heads
- Drive design includes tape/head wear
- Accumulate debris
  - tape debris
  - atmosphere
  - tape coating (friction, humidity)
- Wear with tape medium helps clean heads
- Heads last around 2000 hours of tape contact
- Algorithms for
  - Periodic head cleaning
  - Fast replacement on failure

More Reliability Issues
- Other drive problems

Megatape 1991 Repair Statistics (8mm)

<table>
<thead>
<tr>
<th>Repair type</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replace heads</td>
<td>44</td>
</tr>
<tr>
<td>Tape mechanism (reel motors, tape tension, etc.)</td>
<td>21</td>
</tr>
<tr>
<td>Card failure</td>
<td>17</td>
</tr>
<tr>
<td>Other (firmware, power supply, etc.)</td>
<td>14</td>
</tr>
<tr>
<td>No defect found</td>
<td>4</td>
</tr>
</tbody>
</table>

- Robot reliability
- Support hardware
Summary

- Applications want high sustained throughput

- Technology Trends:
  - Tape drives increasing in capacity, bandwidth (currently inadequate)
  - Robots allow automatic handling of cartridges

- Stripping:
  - Increased throughput
  - Reduced latency of large requests

- Stripping configurations:
  - Within or between robots
  - Tradeoffs: ratio of readers, robot arms, tapes

- Striping issues:
  - Interleave factor for best performance
  - Effect of improved drives, robots
  - Synchronization problems

- Reliability Issues:
  - Media Wear
  - Head Wear
  - Other drive failures
  - Robot failures
  - Error correction needed: how much?