A Framework for Adaptive Storage I/O on Computational Grids

William Jones (presenter)
Parallel Architecture Research Laboratory
Clemson University, Clemson, South Carolina
http://www.parl.clemson.edu/

Dan Reed, et. al. (authors)
Department of Computer Science
University of Illinois
Introduction

- Computational grids
- Irregular applications
- Complex behavior
- Resource demand / availability
- Introduces complex optimization problems
Difficult

- No single set of policies
- Dynamic performance instrumentation
- Malleable FS policies
- Adaptive control systems
- Application request patterns
- Automatically configure
PPFS II

- Second generation
- Flexible testbed HPIO
- RT adaptive policy control
- Sensors
- Decision procedures
- Policy actuators
Adaptive Components - Cacheing
Cap. vs. xfer rate
Resource contention
Low: \( \rightarrow \) res. times
High: \( \rightarrow \) throughput
M/G/1 and fuzzy logic
Trade-off: Multiple files
Interactive Control

- Adaptive, closed loop software system
- User steering
- Visualize and control
- Sensors and actuators

**Autodriver Actuator Interaction**

- User may enter value for selected actuator and transmit it to the remote process
- Interface may be customized for non-numeric data entry such as pull-down menu choice of LRU or MRU for actuator controlling cache replacement policy
New Access Pattern Classification

- Artificial Neural Nets (qual.) (present)
- Time series analysis (quant.) (future)
- Hidden Markov Models (quant.) (future)

Figure 2: Generalized FFANN Architecture
Arrival Pattern Forecasting

- Predict burstiness
- Data access patterns
- Seasonal trends
- Stochastic analysis
- ARIMA - technique
- PRISM - trace data

Figure 3.1: ARIMA Autocorrelator Architecture
PRISM Interarrival Times

(a) Unfiltered

(b) Filtered
Periodic and Non-periodic

Series: PRISM <- Segment1 <- ACF

Series: PRISM <- Segment1 <- PACF
Forecast

- Shortened PRISM trace
- Omitted 60 read requests
- Good results
- On-line techniques
Hidden Markov Models

- Optimal caching/prefetching policies
- Access patterns fully known
- Impossible to know a priori
- Data dependencies
- Loose client synchronization
- Probabilistic access models
HMM’s

- Markov property
- No history
- Capture the past
- Train on-line or off-line
- Probable sequences
- Non-probable data
Concluding Remarks

✦ Adaptive control
✦ Access pattern classification
✦ Cacheing, prefetching, and striping
✦ ANN’s, HMM’s, ARIMA, M/G/1, fuzzy logic
✦ Progress