Distributed Information Processing

24th Lecture

Eom, Hyeonsang (엄현상)
Department of Computer Science & Engineering
Seoul National University
Outline

- Dynamic Adaptation
  - Introduction
  - Maximizing Speedup through Self-Tuning of Processor Allocation
- Q&A
Introduction

- **Adaptivity**
  - Adaptation of Applications to Changes in Their Execution Environment
    - Changes in computational load
    - Changes in network performance

- **Application-Specific Framework**
  - Working Cooperatively with:
    - Measurement and monitoring
    - Alternative evaluation and selection
    - Performance-driven scheduling

Support at Different Levels

- Prior Identification or Provisioning
- Resource Reservation
Introduction (Cont’d)

- Changes in Computational Load
  - Obtaining Additional Resources
  - Asking Other Components to Adapt for Improvement
  - Relocating Computation
  - Reducing Requirements in Areas of Little Interest
Introduction (Cont’d)

Changes in Network Performance

- Controlling Error
- Making a Bandwidth Reservation
- Making Other Links Ask for More Bandwidth
- Relocating Communication
- Applying Compression
Introduction (Cont’d)

- Adapting Applications
  - Bandwidth Adaptation Approaches
    - Resource reservation (problematic)
      - Consumption of large memory for storing flowspecs
      - Low utilization for guaranteed services
      - Not being supported by nonswitched Ethernets & wireless LANs
      - Need for deploying policy control, security, and charging mechanisms
  - Adaptation of applications’ requirements

- Bandwidth Adaptation Requirements
  - QoS Measurements
    - E.g., RTCP (Control Protocol) in RTP for continuous media
Introduction (Cont’d)

- **Delay Adaptation**
  - Goal of Using Large Playout Buffers
    - Conversion of a variable delay into a fixed delay
    - Starvation prevention
  - Support for Using Large Playout Buffers
    - Variable buffer requirement estimation
    - Fixed Buffer

- **Loss Adaptation**
  - Retransmission (Limited by Delay Tolerance)
  - Redundant Transmission
  - Interleaving
  - Forward Error Correction (for Perfect Reconstruction with Redundant Parity Packets)
Application Adaptation

Goal

- Performance Contract
  - Quantified expectations between application performance demands and resource service capabilities

Techniques

- Contract Monitoring
  - Verifying, detecting when, and diagnosing why
- Adaptive Control
  - Adapting to a new resource regime

As a Form of Service-Level Agreements

C.f., Migrating to Other Resources vs Adjusting Contract Parameters Dynamically
Resource and Performance Variability

Sources of Variability in Performance and Availability
- Contention
- No Support for Reservation
- Failure and Preemption

Adapting the Execution for High Performance in a Shared Environment
- Relocating Resources
- Changing Application Behavior
Variability (Cont’d)

Problem of “Static” Performance Models

- Working Only under Ideal Conditions
  - Computational speeds
  - Network latency and bandwidth
  - I/O speed

- Not Working under Dynamic Conditions

Adapting Dynamically to Changing Conditions

- Acquiring New Resources
- Reducing Solution Resolutions
- Switching to Alternatives
Instrumentation and Metrics

Instrumentation

- Automatic
- Minimizing Perturbation and Intrusion
- To Be Inserted at the Proper Level

Metrics

- To Be Selected Appropriately
- Considering Measurement Uncertainty and Temporal Variability
  - Tradeoff between the length of measurement interval and adaptability

E.g., for Contract Verification and Validation
Adaptive Control Example

Real-Time (Runtime) Monitoring

- Instrumentation
  - E.g., inserting “sensors”
- Periodic Transmission of Sensor Data
- Analysis of the Data
  - E.g., evaluation of the rule base with the data, and detection of contract violation
- Notification of the Result
  - E.g., distributing the result via the sensors

Issues: e.g., Assessing Temporal Variability and Contract Violation
Adaptive Control Example (Cont’d)

- Remediation
  - Halting the Execution
  - Migrating the Workload
    - At different levels
    - Comparing the benefit & cost
  - Restarting the Application

Requiring Control Stability and Rescheduling Mechanisms

Requiring the Support Such as Checkpointing
References


[Foster03] I. Foster and C. Kesselman (Editors), The GRID 2: Blueprint for a New Computing Infrastructure, Morgan Kaufmann Publishers, November 2003