Preserving QoS of E-commerce Sites Through Self-Tuning: A Performance Model Approach

Daniel Menascé and Daniel Barbará
E-Center for E-Business
George Mason University
Fairfax, VA,

and

Ronald Dodge Jr
United States Army
Forth Leavenworth, KS

© 2001 Menascé, Barbará, and Dodge. All Rights Reserved.
workload $\rightarrow$ E-commerce Site $\rightarrow$ QoS

response time
throughput
probability of rejection
Outline of the Talk

- The need for dynamic QoS control.
- Definition of a QoS metric.
- Discussion of the approach.
- Experimental results.
- Concluding remarks.
Dynamic QoS Control: Motivation

- E-commerce sites are complex and composed of multiple tiers.
Multi-tier Architecture
Dynamic QoS Control: Motivation

- E-commerce sites are complex and composed of multiple tiers.
- The workload presents short-term variations with high peak-to-average ratios.
E-commerce Bursty Workloads

60 sec

1 sec

© 2001 Menascé, Barbará, and Dodge. All Rights Reserved.
Dynamic QoS Control: Motivation

- E-commerce sites are complex and composed of multiple tiers.
- The workload presents short-term variations with high peak-to-average ratios.
- Many software and hardware parameters influence the performance of e-commerce sites.

Manual reconfiguration is not an option!
Dynamic QoS Control for E-commerce

- Definition of a combined QoS metric.
QoS Deviation

- Relative difference between the observed QoS value and the QoS goal.

Response time deviation:

$$\Delta QoS_R = \frac{R_{\text{max}} - R_{\text{measured}}}{R_{\text{max}}}$$
QoS Deviation

- Relative difference between the observed QoS value and the QoS goal.

**Response time deviation:**

\[ \Delta QoS_R = \frac{R_{\text{max}} - R_{\text{measured}}}{R_{\text{max}}} \]

**Throughput deviation:**

\[ \Delta QoS_X = \frac{X_{\text{measured}} - X_{\text{min}}}{X_{\text{min}}} \]
QoS Deviation

- Relative difference between the observed QoS value and the QoS goal.

  **Response time deviation:**
  \[
  \Delta QoS_R = \frac{R_{\text{max}} - R_{\text{measured}}}{R_{\text{max}}}
  \]

  **Throughput deviation:**
  \[
  \Delta QoS_X = \frac{X_{\text{measured}} - X_{\text{min}}}{X_{\text{min}}}
  \]

  **Probability of rejection deviation:**
  \[
  \Delta QoS_{\text{Prej}} = \frac{P_{\text{MAX}}^{\text{Prej}} - P_{\text{measured}}^{\text{Prej}}}{P_{\text{MAX}}^{\text{Prej}}}
  \]
QoS Deviation

- Relative difference between the observed QoS value and the QoS goal.

  Response time deviation:

\[
\Delta QoS_R = \frac{R_{\text{max}} - R_{\text{measured}}}{R_{\text{max}}}
\]

  Throughput deviation:

\[
\Delta QoS_X = \frac{X_{\text{measured}} - X_{\text{min}}}{X_{\text{min}}}
\]

  Probability of rejection deviation:

\[
\Delta QoS_{\text{Prej}} = \frac{P_{\text{MAX}}^{\text{Prej}} - P_{\text{measured}}^{\text{Prej}}}{P_{\text{MAX}}^{\text{Prej}}}
\]

A negative deviation means that the QoS level for the metric has not been met.

© 2001 Menascé, Barbará, and Dodge. All Rights Reserved.
QoS Metric

- The QoS metric is defined as a linear combination of QoS deviations.
- The weights are determined by management based on the relative importance of each metric.

\[
QoS_v = (\Delta QoS_R \times W_R) + (\Delta QoS_{\text{Prej}} \times W_{\text{Prej}}) + (\Delta QoS_X \times W_X)
\]

determined by management
Dynamic QoS Control for E-commerce

- Definition of a combined QoS metric.
- Use of hill-climbing techniques combined with predictive queuing models.
Heuristic Optimization Approach

QoS_0: observed QoS value for current configuration
QoS_1, QoS_2, and QoS_3 are determined by a predictive queuing model of the site.

\[ QoS = f(\vec{W}, c_1, c_2, \ldots, c_m) \]
E-commerce Site Queuing Model
Heuristic Optimization Approach

- Parameters for the queuing model are collected dynamically from the site.
- The QoS values for surrounding points are calculated.
- The path with the greatest QoS gain is chosen. If no configuration improves the QoS or a limit is reached, the search ends.
Dynamic QoS Control for E-commerce

- Definition of a combined QoS metric.
- Use of hill-climbing techniques combined with predictive queuing models.
- Implemented a TPC-W site in a multi-tier architecture.
Prototype Configuration

QoS Controller

Database Server

Web Server

Application Server

100 Mbps Hub

Workload Generator

Workstation

TPC-W site

© 2001 Menascé, Barbará, and Dodge. All Rights Reserved.
Dynamic QoS Control for E-commerce

- Definition of a combined QoS metric.
- Use of hill-climbing techniques combined with predictive queuing models.
- Implemented a TPC-W site in a multi-tier architecture.
- Implemented a QoS Controller on a dedicated machine and evaluated the approach.
Dynamic QoS Controller

Arriving Request → E-commerce System

Configuration Instructions → System Performance Data

Arrival Process Data → QoS Metric Data

Desired QoS Levels

Completed Request

Controller
Experiment Results

Arrival Rate (requests/sec) vs. Time (Controller Intervals)
Results of QoS Controller
Experiment Results

Arrival rate

QoS is not met!
Concluding Remarks

- Performance models are useful to guide optimization techniques to dynamically control the QoS of e-commerce sites.

- Work in progress:
  - Different search techniques.
  - Use of workload forecasting techniques.
  - Blind vs. informed search.
  - Other configuration parameters.