AWPS - an architecture for pro-active web performance management

Univ.-Prof. Dr. Gabriele Kotsis (Gabriele.Kotsis@jku.at)
Martin Pinzger (Mpinzger@gmail.com)
Department of Telecooperation
Johannes Kepler University Linz Austria
Motivation

Performance = F (Workload, System)

What is the effect on performance if workload is subject to change?

What is the effect on performance if the system is subject to change? (consider also interdependencies with changes in the workload)

How to design the system to deliver a certain performance for a given workload?
Approaches

- **Performance Measurements**
  - On real world systems
  - On artificial systems
  - Example: DynaTrace (JKU Spin Off)
    - Measurement and analysis tools for web server performance based on the concept of execution paths

- **Performance Modelling**
  - Off line versus online
Automated Web Performance System

- AWPS Concept
- AWPS Environment Interaction
- Case Study
- Conclusion
AWPS Concept

- **Key Characteristics**
  - Automatic
  - Online / „Realtime“
  - Pro-active

- **Three Key Functions**
  - Data Collection
  - Simulation
  - Prediction

Diagram:
- Management Component
- Monitoring Component
- Modeling Component
- Prediction Component
Data Collection - Monitoring
AWPS Concept - Simulation Component

- Model Generation Component
  - Minimum complexity simulation model
  - Maximum complexity simulation model
- Model Comparison Component
- Model Adjustment Component
  - AVG Strategy, Median Strategy, ARMA Strategy
- Model Simulation Component
  - JSIM
Modeling / Simulation Task
AWPS Concept - Simulation Component

Model Generation Component - Example

- **TotalSystemTime** = GlobalOut - GlobalIn
- **WebServerTimeA** = AppServerIn - GlobalIn
- **AppServerTime** = AppServerOut - AppServerIn
- **WebServerTimeB** = GlobalOut - AppServerOut
- **WebServerTime** = WebServerTimeA + WebServerTimeB
Modeling / Simulation Component
Prediction Component
Management Task
AWPS Environment Interaction

- **System Setup**
  - Passive Monitoring Strategy \((\text{Network Sniffing})\)
  - A Multi Class / Single Queue / Multi Server Model is created automatically

- **User Interaction**
  - Main Configuration Site
  - Result Presentation Site
  - Online Observation Site

- **Influence on the Productive System** - Minor
Welcome to the AWPS - System

The AWPS - System and its base components are described in ASE08 Doctoral Symposium Paper "Automated Web Performance Analysis".

Abstract

Performance is a key feature in many systems nowadays. There are several tools on the market that ensure and test for adequate performance. They can be divided into simulation tools and monitoring tools. But not a few automate and combine both approaches. This paper describes a system capable of automatically creating a web performance simulation and conducting trend analysis of the system under test (SUT). To achieve this the system requires input information, like Monitoring Points and Static-Information about the SUT. The system monitors and analyses the SUT and based on this information generates a simulation model of the system. The simulation model is refined stepwise, e.g. by adding or removing connections between the model components or adjusting the parameters until the aimed accuracy is achieved. With the help of the simulation model a prediction module creates an analysis of the SUT, and thereby can give as much information about the current state of the system and potential trends as possible. This predictive information can be used for pro-active server tuning or other performance optimisations. The focus of my PhD thesis is on the adjustment and prediction part of the system described here. For all other parts, already existing tools and techniques will be used where possible. This initial paper outlines the complete system.
# Results of the Analysis

## Overview

In the following table you see the basic results of the executed test runs.

Filter by IdConfRun:

(1739)PERIODIC SPIKE INKL RAND ARMA Stop Size 10 (Inkl 10 Threads)

<table>
<thead>
<tr>
<th>Id</th>
<th>IdConfRun</th>
<th>IdTestRun</th>
<th>SimExCounter</th>
<th>ServerIdentification</th>
<th>Category</th>
<th>AVG-Sim (Sec.)</th>
<th>AVG-Calc-All (Sec.)</th>
<th>AVG-Calc-Sim (Sec.)</th>
<th>AVG-Queue-Length</th>
<th>Nr.Requests</th>
</tr>
</thead>
<tbody>
<tr>
<td>2039427</td>
<td>1739</td>
<td>192</td>
<td>1</td>
<td>MultiServer-All</td>
<td>gif 0</td>
<td>0.0</td>
<td>0.50015</td>
<td>0.4977</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>2039428</td>
<td>1739</td>
<td>192</td>
<td>2</td>
<td>MultiServer-All</td>
<td>gif 0</td>
<td>0.0</td>
<td>0.49966666666667</td>
<td>0.4966</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>2039429</td>
<td>1739</td>
<td>192</td>
<td>3</td>
<td>MultiServer-All</td>
<td>gif 0</td>
<td>0.0</td>
<td>0.50305</td>
<td>0.5123</td>
<td>0</td>
<td>40</td>
</tr>
<tr>
<td>2039430</td>
<td>1739</td>
<td>192</td>
<td>4</td>
<td>MultiServer-All</td>
<td>gif 0.58688995401449</td>
<td>0.50052</td>
<td>0.4994</td>
<td>0.02259659540916813</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>2039431</td>
<td>1739</td>
<td>192</td>
<td>4</td>
<td>MultiServer-All</td>
<td>gif 0.5888995401449</td>
<td>0.50052</td>
<td>0.4904</td>
<td>0.02259659540916813</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td>2039432</td>
<td>1739</td>
<td>192</td>
<td>5</td>
<td>MultiServer-0</td>
<td>gif 0.5584575643211779</td>
<td>0.492833333333333</td>
<td>0.4931</td>
<td>0.01281438159451016</td>
<td>60</td>
<td>0</td>
</tr>
<tr>
<td>2039433</td>
<td>1739</td>
<td>192</td>
<td>5</td>
<td>MultiServer-All</td>
<td>gif 0.5584575643211779</td>
<td>0.492833333333333</td>
<td>0.4931</td>
<td>0.01281438159451016</td>
<td>60</td>
<td>0</td>
</tr>
<tr>
<td>2039434</td>
<td>1739</td>
<td>192</td>
<td>6</td>
<td>MultiServer-0</td>
<td>gif 0.5731647177150283</td>
<td>0.4986285714285714</td>
<td>0.4947</td>
<td>0.014249411680278276</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>2039435</td>
<td>1739</td>
<td>192</td>
<td>6</td>
<td>MultiServer-All</td>
<td>gif 0.5731647177150283</td>
<td>0.4986285714285714</td>
<td>0.4947</td>
<td>0.014249411680278276</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>2039436</td>
<td>1739</td>
<td>192</td>
<td>7</td>
<td>MultiServer-0</td>
<td>gif 0.570354034890223</td>
<td>0.49845</td>
<td>0.4972</td>
<td>0.012140606510588065</td>
<td>80</td>
<td>0</td>
</tr>
<tr>
<td>2039437</td>
<td>1739</td>
<td>192</td>
<td>7</td>
<td>MultiServer-All</td>
<td>gif 0.570354034890223</td>
<td>0.49845</td>
<td>0.4972</td>
<td>0.012140606510588065</td>
<td>80</td>
<td>0</td>
</tr>
<tr>
<td>2039438</td>
<td>1739</td>
<td>192</td>
<td>8</td>
<td>MultiServer-0</td>
<td>gif 0.49750368145471</td>
<td>0.497133333333333</td>
<td>0.4866</td>
<td>0</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>2039439</td>
<td>1739</td>
<td>192</td>
<td>8</td>
<td>MultiServer-All</td>
<td>gif 0.49750368145471</td>
<td>0.497133333333333</td>
<td>0.4866</td>
<td>0</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>2039440</td>
<td>1739</td>
<td>192</td>
<td>9</td>
<td>MultiServer-0</td>
<td>gif 0.5456861503436212</td>
<td>0.49752</td>
<td>0.501</td>
<td>0.006817794779505936</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>2039441</td>
<td>1739</td>
<td>192</td>
<td>9</td>
<td>MultiServer-All</td>
<td>gif 0.5456861503436212</td>
<td>0.49752</td>
<td>0.501</td>
<td>0.006817794779505936</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>2039442</td>
<td>1739</td>
<td>192</td>
<td>10</td>
<td>MultiServer-0</td>
<td>gif 0.5540319738558908</td>
<td>0.4989818181818182</td>
<td>0.5136</td>
<td>0.00691105459259537</td>
<td>110</td>
<td>0</td>
</tr>
<tr>
<td>2039443</td>
<td>1739</td>
<td>192</td>
<td>10</td>
<td>MultiServer-All</td>
<td>gif 0.5540319738558908</td>
<td>0.4989818181818182</td>
<td>0.5136</td>
<td>0.00691105459259537</td>
<td>110</td>
<td>0</td>
</tr>
<tr>
<td>2039444</td>
<td>1739</td>
<td>192</td>
<td>11</td>
<td>MultiServer-0</td>
<td>gif 0.545008483221205</td>
<td>0.4998156666666667</td>
<td>0.5102</td>
<td>0.00484175610645552</td>
<td>120</td>
<td>0</td>
</tr>
</tbody>
</table>
Detail Analysis of the Request Id:ConfRun:1739

Settings

1739:PERIODIC SPIKE URL: RRDH-D-RRH Step Size: 15 min 10

Diagram

In the following diagram you see the basic results of the executed test run, in comparison.
System Status

Environment:

- Web Server
  - OK

- PostgreSQL Database Server 8.3
  - OK

- Socket Communication Service
  - OK

Select Conf Run to Monitor Online:

(1740)PERIODIC SPIKE INKL RAND ARMA Step Size 1 (Incl 10 Threads)
System Status ONLINE - Monitor
Analysed System Categories

- Feasibility of the Approach
  - Synthetic System Offline
    e.g. strictly increasing, constant, complex function

- Real-time Feasibility
  - Synthetic System Online
    e.g. by control invoice concerning calculation time consumption

- Representative Test
  - Productive System Offline
    TK-Website offline analysis (normal load / synthetic load),
    GoSpace offline analysis (synthetic load)

- Representative Realtime Test
  - Productive System Online / Field Test
Case Study

The case study was done on a two tier web application, which provides as functionality a web page where you can search and book space flights.

\[
\text{TotalSystemTime} = \text{GlobalOut} - \text{GlobalIn}
\]

\[
\text{WebServerTime} = \text{AppServerIn} - \text{GlobalIn}
\]

\[
\text{AppServerTime} = \text{GlobalOut} - \text{AppServerIn}
\]
Case Study - Results - Overview

- Analysis of difference between simulation and reference data
  - T-Test
  - Mean Error / Variation

- Correlation between accuracy and simulation runs
  - Step Size 1000 (Median, AVG, ARMA, ARMA G.)
  - Step Size 100 (Median, AVG, ARMA)

- Realtime / Online Capability
**Case Study - Results - T-Test**

| Strategy | Step Size 1000 | | | Step Size 100 | | | Step Size 10 | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | image | special | image | special | image | special | |
| AVG | Yes | No (0.980230) | Yes | Yes (0.935309) | Yes | Yes (0.789822) | |
| Median | Yes | Yes (0.758424) | Yes | Yes (0.335950) | Yes | Yes (0.002655) | |
| ARMA | Yes | No (0.982876) | Yes | No (0.990521) | Yes | No (0.998247) | |
| ARMA G. | Yes | No (0.992849) | Yes | Yes (0.289593) | Yes | Yes (0.000000) | |

Significant difference between simulation and reference data. The value in the brackets represents the double sided t-Test value.
Case Study - Results - Observed Error

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Step Size 1000</th>
<th></th>
<th>Step Size 100</th>
<th></th>
<th>Step Size 10</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>variance</td>
<td>median</td>
<td>variance</td>
<td>median</td>
<td>variance</td>
</tr>
<tr>
<td>AVG</td>
<td>0.002228</td>
<td>0.000002</td>
<td>0.005419</td>
<td>0.000013</td>
<td>0.023877</td>
<td>0.000514</td>
</tr>
<tr>
<td>Median</td>
<td>0.017624</td>
<td>0.000052</td>
<td>0.017543</td>
<td>0.000084</td>
<td>0.020431</td>
<td>0.001013</td>
</tr>
<tr>
<td>ARMA</td>
<td>0.005091</td>
<td>0.000012</td>
<td>0.006319</td>
<td>0.000022</td>
<td>0.023590</td>
<td>0.000568</td>
</tr>
<tr>
<td>ARMA G.</td>
<td>0.006824</td>
<td>0.000019</td>
<td>0.088866</td>
<td>0.055627</td>
<td>0.230810</td>
<td>0.060259</td>
</tr>
</tbody>
</table>

Mean values and Variance values in seconds referring to the delta error, for the special offers (do?action=special) request class.
Case Study - Results - Observed Error Correlation to the Simulation Runs

Delta Error of the Special Request (Step Size 1000)
Case Study - Results - Observed Error Correlation to the Simulation Runs

The Visualization of ARMA G. was skipped because of the high fluctuation.
Case Study - Results - Percentile 0.99
Case Study - Results - Realtime

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Step Size 1000</th>
<th>Step Size 100</th>
<th>Step Size 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>AVG</td>
<td>244 sec.</td>
<td>511 sec.</td>
<td>2659 sec.</td>
</tr>
<tr>
<td>Median</td>
<td>489 sec.</td>
<td>541 sec.</td>
<td>3243 sec.</td>
</tr>
<tr>
<td>ARMA</td>
<td>339 sec.</td>
<td>945 sec.</td>
<td>6706 sec.</td>
</tr>
<tr>
<td>ARMA G.</td>
<td>288 sec.</td>
<td>827 sec.</td>
<td>2947 sec.</td>
</tr>
</tbody>
</table>

Calculation time consumption. Values below 3600 sec. mean that the adjustment method is online-capable.
Conclusion

- AWPS works as expected and provides representative results
- Simulation model generation process works autonomously and is sufficiently fault-tolerant
- Strategies for the adjustment of the simulation model work accurately

- Functionality should be enhanced e.g. adaptive scenario generation
- Additional case studies e.g. productive system under high (real) load
The Team

- **Head of Department**
  - Univ.Prof. Dr. Gabriele Kotsis

- **Research and Teaching Assistants**
  - Dipl.Ing Kerstin Altmanninger
  - Dipl.Ing. Sabine Bachmayer
  - Dr. Ismail Khalil Ibrahim
  - Dr. Wieland Schwinger

- **Project Researchers and PhD Students**
  - Stefan Mitsch
  - Martin Pinzger
  - Wolfgang Pointner
  - Martin Wischenbart
  - Elena Zanzani

- **Secretary**
  - Angela Kohl

- **Technical Support**
  - Fabian Mergl
Overview of Research

Distributed & Mobile Computing
- grid computing
- internet computing
- wireless networks
- sensor networks
- ubiquitous web applications
- agent-based computing
- mobile communication
- mobile multimedia
- mobile computing
- intelligent systems
- cooperation
- mobile information management
- model driven engineering
- conceptual modeling
- semantic modeling

Media & Interaction
- personalization
- non-standard HCI
- multimedia
- interaction
- mobile information management
- performance evaluation
- simulation
- usability

Modelling & Evaluation

TK
Ubiquitous Communication Management - HERMES

- **Vision**: support users’ communication needs
  - **Gain knowledge** of users’ communication needs, recognize behavior patterns and learn
  - **Provide appropriate communication tool support** based on precedent data mining and resultant knowledge
  - **Execute** appropriate communication-related actions
HERMES Architecture Overview

- **Operation starts with Sensors**
  - Specialized data mining components
  - Responsible for gathering data of interest for any communication management-related task
  - Receive new data and publish this to all other components via events
  - Published events are routed to interested recipients, usually Rule Bases

- **Rule Bases** are regarded as brains
  - Represent key idea of a ubiquitous communication management system: ability to learn from previous experience in the domain of communications
  - Supposed to analyze data contained in events and take appropriate communication-related actions
HERMES Architecture Overview

- Possibilities to interact with environment
  - Send Inputs/Outputs to IO Services
  - Send actions to Tool Services

- Inputs/Outputs wrap actual values
  - Generic approach enables to switch between Input/Output methods and to develop further methods for different devices
  - Numerous predefined Inputs/Outputs to use out of the box

- Actions are communication-related intents
  - Usually passed from Rule Bases to Tools
  - Series of predefined Actions

- Tools are responsible for executing received communication-related Actions
  - May forward Actions to existing communication tools and interact with those
HERMES Architecture Overview

- Knowledge of users’ communication needs supposed to be persistent in embedded database

- Communication between framework-based applications powered by XMPP communication platform
  - Primarily intended for simultaneous knowledge replication

- Framework’s structure enables framework-based applications to run in distributed heterogeneous networks at the same time
Collaborative Streaming Media

<table>
<thead>
<tr>
<th>Actions</th>
<th>Type</th>
<th>Mode</th>
<th>Architecture</th>
<th>Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create Streaming Session, Watch</td>
<td>Single</td>
<td>Synchronous</td>
<td>Client-Server, CDN</td>
<td>Hierarchical Collaborative Multicast [16], CoolStreaming [21], DISCOVR [14], COSMOS [13], [12], STARCast [19], [20]</td>
</tr>
<tr>
<td>Data Acquisition / Distribution</td>
<td>Collaborative</td>
<td>Synchronous / Collaborative</td>
<td>Peer-to-Peer</td>
<td>Zync [18], [15], Comodin [7], [6], [4], [3], [5], coStream [8], [11], [10], [9]</td>
</tr>
<tr>
<td>Watch, Control, Edit the Stream</td>
<td>Shared / Collaborative</td>
<td>Collaborative</td>
<td>Client-Server, CDN</td>
<td>Zync [18], [15], Comodin [7], [6], [4], [3], [5], coStream [8], [11], [10], [9]</td>
</tr>
<tr>
<td>Invite / Join Session, Communicate, (Automated) Awareness</td>
<td>Collaborative</td>
<td>Synchronous</td>
<td>Client-Server, CDN</td>
<td>CWaCTool [17]</td>
</tr>
<tr>
<td>Create, Annotate</td>
<td>Collaborative</td>
<td>Asynchronous</td>
<td>Peer-to-Peer</td>
<td></td>
</tr>
</tbody>
</table>
Collaborative Streaming Architecture
Thank you for your attention!
Case Study 1/5
Automatic Simulation Model Parameter Adjustment

- **Case Study Setup**
  - Test - Web - Sites
  - Use of Recorded Data

- **Type D**
  - Recorded Base-data

- Simulation Parameter Calculation Base-data
- Simulation Execution Base-data

- a) Constant Response Time
  - Response time / Request

- b) Sawtooth Response Time
  - Response time / Request

- c) Strictly Increasing Response Time
  - Response time / Request

* By Request Number the ID of a request is meant, the number of requests sent to the SUT in one second is assumed constant.
Case Study 2/5
Automatic Simulation Model Parameter Adjustment

Executed Case Studies
- 1 Minute
  - Constant Response Time (ARMA, AVG, Binary)
  - Sawtooth Response Time (ARMA, AVG, Binary)
  - Strictly Increasing Response Time (ARMA, AVG, Binary)
- 15 Minutes
  - Constant Response Time (ARMA, AVG, ARMA G., Binary)
  - Sawtooth Response Time (ARMA, ARMA G., AVG, Binary)
  - Strictly Increasing Response Time (ARMA G., AVG, ARMA, Binary)
- 60 Minutes
  - Constant Response Time (ARMA, AVG, ARMA G., Binary)
  - Sawtooth Response Time (ARMA G., AVG, ARMA, Binary)
  - Strictly Increasing Response Time (ARMA G., Binary, AVG, ARMA)
Case Study 3/5 15 Minutes Sawtooth 1/3
Automatic Simulation Model Parameter Adjustment
Case Study 3/5  15 Minutes Sawtooth 2/3
Automatic Simulation Model Parameter Adjustment
Case Study 3/5  15 Minutes Sawtooth 3/3
Automatic Simulation Model Parameter Adjustment
Case Study 4/5 60 Minutes Sawtooth 1/3
Automatic Simulation Model Parameter Adjustment
Case Study 4/5 60 Minutes Sawtooth 2/3
Automatic Simulation Model Parameter Adjustment

![Chart showing AVG response times over number of requests with a sawtooth pattern.](chart.png)
Case Study 4/5 60 Minutes Sawtooth 3/3
Automatic Simulation Model Parameter Adjustment
Case Study 5/5
Automatic Simulation Model Parameter Adjustment

Which approach performs best for varying sample size?

- 1 Minute
  - (1) ARMA; (2) AVG

- 15 Minutes
  - (1) ARMA and ARMA G.; (2) AVG

- 60 Minutes
  - (1) AVG; (2) ARMA G.

- Global Ranking
  - ARMA G. (1,500)
  - AVG (1,555)
  - ARMA (1,666)
  - Binary (3,111)