Timing Behavior Anomaly Detection
for Automatic Failure Detection and Diagnosis

Research visit at Charles University Prague

Matthias Rohr
matthias.rohr@informatik.uni-oldenburg.de

Graduate School TrustSoft, Software Engineering Group
Department of Computing Science, University of Oldenburg

10th of April 2007
Motivation

Failure diagnosis in business-critical software systems is time-consuming and error-prone. Runtime behavior observations are indicative for failure diagnosis.

Matthias Rohr (TrustSoft)
Motivation

Failure diagnosis in business-critical software systems is time-consuming and error-prone. Runtime behavior observations are indicative for failure diagnosis.

Matthias Rohr (TrustSoft)
Motivation

Failure diagnosis in business-critical software systems

- Manual failure diagnosis is time-consuming and error-prone
Failure diagnosis in business-critical software systems

- Manual failure diagnosis is time-consuming and error-prone
- Runtime behavior observations are indicative for failure diagnosis
Motivation

Vision

- Automatic localization of faults through runtime behavior evaluation
Motivation

Approach

- Automatic localization of faults through runtime behavior evaluation
- Automatic detection of timing behavior anomalies in software systems
Approach

- Automatic localization of faults through runtime behavior evaluation
- Automatic detection of timing behavior anomalies in software systems

Research questions:
- How can anomalies be detected in timing behavior?
- How can system usage variations be addresses in timing behavior evaluation?
- What is the relation between software faults and runtime timing behavior?
Outline

1 Foundations
   - Dependability
   - Anomaly Detection
   - Software Performance

2 Creation of the timing behavior profile

3 Fault Localization

4 Evaluation

5 Related work

6 Conclusions
## Threats to dependability

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fault</td>
<td>Root-cause of a failure</td>
</tr>
<tr>
<td>Error</td>
<td>Incorrect system state</td>
</tr>
<tr>
<td>Failure</td>
<td>Deviation from correct system behavior visible to the user</td>
</tr>
</tbody>
</table>

[Avižienis et al., 2004]
Dependability Terminology [Avižienis et al., 2004]

Threats to dependability

- **Fault**: Root-cause of a failure
- **Error**: Incorrect system state
- **Failure**: Deviation from correct system behavior visible to the user

Failure Diagnosis:

- Failure detection
- Identification of faults
- **Fault localization**
Availability

Availability: Common definition (e.g., [Musa et al., 1987])

\[
\text{Availability} = \frac{\text{MTTF}}{\text{MTTF} + \text{MTTR}}
\]

\textbf{MTTF}  Mean Time to Failure
\textbf{MTTR}  Mean Time to Repair
Availability

Availability: Common definition (e.g., [Musa et al., 1987])

\[
\text{Availability} = \frac{\text{MTTF}}{\text{MTTF} + \text{MTTR}}
\]

- **MTTF**: Mean Time to Failure
- **MTTR**: Mean Time to Repair

Two alternative strategies to increase availability

- Increase of mean time to failure (reliability)
- Decrease of mean time to repair
  - Failure diagnosis support
An anomaly is a deviation from “normal” system behavior.
Anomaly Detection (1/2)

- An *anomaly* is a deviation from “normal” system behavior

- Normal system behavior:
  - Static reference values (e.g., mean response time over a day \( \leq T \))
  - Analytical or *statistical* models in dependence to system influences and historical system behavior
Methods to create normal behavior profiles

- Manual specification
- **Automatic profile learning from observations**

Challenges of anomaly detection:
- False alarms
- System usage
- Nonlinear system behavior, modeling uncertainties

Typical application domains:
- Industrial manufacturing, large-scale control systems [Palade et al., 2006]
- Network management [Maxion, 1990]
- Intrusion detection (Security) [Denning, 1987]
Anomaly Detection (2/2)

Methods to create normal behavior profiles

- Manual specification
- **Automatic profile learning from observations**

Challenges of anomaly detection:

- False alarms
- System usage
- Nonlinear system behavior, modeling uncertainties

Typical application domains:

- Industrial manufacturing, large-scale control systems [Palade et al., 2006]
- Network management [Maxion, 1990]
- Intrusion detection (Security) [Denning, 1987]
Anomaly Detection (2/2)

Methods to create normal behavior profiles
- Manual specification
- **Automatic profile learning from observations**

Challenges of anomaly detection:
- False alarms
- System usage
- Nonlinear system behavior, modeling uncertainties

Typical application domains:
- Industrial manufacturing, large-scale control systems [Palade et al., 2006]
- Network management [Maxion, 1990]
- Intrusion detection (Security) [Denning, 1987]
Influences to software timing behavior:

- **System architecture:**
  - Hardware resource capacity
  - Software design

- **System usage:** [cp. Sabetta and Koziolek, 2007]:
  - Workload intensity (e.g., number of active users)
  - Service demand characteristics (e.g., individual request parameters)

- **System state**
  - Performance tuning (e.g., caching, load balancing), ...
  - Server virtualization
Outline

1. Foundations

2. Creation of the timing behavior profile
   - Instrumentation
   - Monitoring
   - Analysis of Execution Sequences
   - Analysis of Workload Intensity

3. Fault Localization

4. Evaluation

5. Related work

6. Conclusions
Timing behavior anomalies:

- Deviations from normal timing behavior (here: response times) of operations of a software system
- e.g., exceptional high or low response times
Timing behavior anomalies:
- Deviations from normal timing behavior (here: response times) of operations of a software system
- e.g., exceptional high or low response times

Relation between software faults and timing behavior anomalies
Timing behavior anomalies:
- Deviations from normal timing behavior (here: response times) of operations of a software system
- e.g., exceptional high or low response times

Relation between software faults and timing behavior anomalies:
- Software faults tend to cause timing behavior anomalies [Kao et al., 1993]
- Successful fault localization based on timing behavior anomalies [Agarwal et al., 2004]
- Response times in enterprise resource planning systems (ERP) are often log-normally distributed [Mielke, 2006]
Overview

Timing behavior anomaly detection for failure diagnosis

Initial activities
- Instrumentation for Monitoring
- Monitoring
- Creation of the timing behavior profile

Continuous activities
- Monitoring
- Update of timing behavior profile
- Timing behavior profile

Activities during failure diagnosis
- Log:
  - response times
  - execution sequences
- Timing behavior profile
- Anomaly detection
- Anomaly analysis
- Diagnosis report
Overview

Timing behavior anomaly detection for failure diagnosis

Initial activities

- Instrumentation for Monitoring
- Monitoring
- Creation of the timing behavior profile
- Timing behavior profile

Continuous activities

- Monitoring
- Update of timing behavior profile
- Timing behavior profile

Activities during failure diagnosis

- Log:
  - response times
  - execution sequences
- Timing behavior profile
- Anomaly detection
- Anomaly analysis
- Diagnosis report
Overview

Timing behavior anomaly detection for failure diagnosis

Initial activities

Instrumentation for Monitoring

Monitoring

Creation of the timing behavior profile

Timing behavior profile
Timing behavior anomaly detection for failure diagnosis

**Initial activities**

- Instrumentation for Monitoring
- Monitoring
- Creation of the timing behavior profile

**Execution sequence analysis**

- Operation analysis
- Execution sequence analysis
- Workload intensity analysis

**Timing behavior profile**
Creation of the timing behavior profile

Instrumentation

Instrumentation for Monitoring

Operation analysis

Execution sequence analysis

Worload intensity analysis

Monitoring of Response times (Start and end of an operation execution)

Execution sequences of operations for each thread

Instrumentation challenges:
- Measurement metrics (Focke et al., 2007a)
- Maintainable integration of measurement logic (Focke et al., 2007b)

Matthias Rohr (TrustSoft)
Monitoring of

- **Response times** (Start and end of an operation execution)
- **Execution sequences** of operations for each thread
Monitoring of
- **Response times** (Start and end of an operation execution)
- **Execution sequences** of operations for each thread

Instrumentation challenges:
- (Measurement metrics)
- Number and position of measurement points [Focke et al., 2007a]
- Maintainable integration of measurement logic [Focke et al., 2007b]
Operation sequence analysis

Worload intensity analysis

Instrumentation
for Monitoring

Monitoring

Execution sequence analysis

Creation of the timing behavior profile

Monitoring

Operation analysis

Trace reconstruction of execution sequences from monitoring log:

Operations: O = {a, b, c}

Executions with TraceID 1:

E_1 = {a, b, c_1, c_2}

Execution sequence t_1 = (a, a_c_1, c_1_a, a_b, b_c_2, c_2_b, b_a, a)

<table>
<thead>
<tr>
<th>Operation</th>
<th>TraceID</th>
<th>t_in</th>
<th>t_out</th>
<th>t_out − t_in</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1</td>
<td>0000</td>
<td>0150</td>
<td>150</td>
</tr>
<tr>
<td>c</td>
<td>1</td>
<td>0030</td>
<td>0050</td>
<td>20</td>
</tr>
<tr>
<td>b</td>
<td>1</td>
<td>0060</td>
<td>0140</td>
<td>80</td>
</tr>
<tr>
<td>c</td>
<td>1</td>
<td>0090</td>
<td>0130</td>
<td>40</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>c</td>
<td>2</td>
<td>0340</td>
<td>0358</td>
<td>18</td>
</tr>
<tr>
<td>c</td>
<td>2</td>
<td>0400</td>
<td>0437</td>
<td>37</td>
</tr>
</tbody>
</table>

Matthias Rohr (TrustSoft)
Trace reconstruction of execution sequences from monitoring log:

- Operations: \( O = \{a, b, c\} \)
- Executions with TraceID 1: \( E^1 = \{a, b, c_1, c_2\} \)
Trace reconstruction of execution sequences from monitoring log:

- Operations: \( O = \{a, b, c\} \)
- Executions with TraceID 1: \( E^1 = \{a, b, c_1, c_2\} \)
- Execution sequence \( t_1 = (a, ac_1, c_1 a, ab, bc_2, c_2 b, ba, a) \)
Trace reconstruction of execution sequences from monitoring log:

- Operations: \( O = \{a, b, c\} \)
- Executions with TraceID 1: \( E^1 = \{a, b, c_1, c_2\} \)
- Execution sequence \( t_1 = (\$a, ac_1, c_1 a, ab, bc_2, c_2 b, ba, a\$) \)
Associate response times to operations:

- $RT := \text{all response times}$
- $RT(o) := \text{response times of one operation } o$

$$\text{All response times}$$

$$RT = (150, 80, \ldots)$$

$$\text{Response time per operation}$$

$$RT(a) = (150, \ldots)$$

$$RT(b) = (80, \ldots)$$

$$RT(c) = (40, 20, 37, 18, \ldots)$$
Associate response times to operations:

- \( RT := \) all response times
- \( RT(o) := \) response times of one operation \( o \)

All response times
\[
RT = (150, 80, \ldots)
\]

Response time per operation
\[
RT(a) = (150, \ldots) \\
RT(b) = (80, \ldots) \\
RT(c) = (40, 20, 37, 18, \ldots)
\]

Statistical description of \( RT(o) = (rt_1, \ldots, rt_n) \)

- **Probability density functions**, histograms
- Location parameters: Mean, Median, Mode
Creation of the timing behavior profile

Analysis of Execution Sequences

Instrumentation for Monitoring

Operation analysis

Execution sequence analysis

Workload intensity analysis

:Bookshop

:CRM

:Catalog

Prob. density

Response time

Matthias Rohr (TrustSoft)
Creation of the timing behavior profile

Analysis of Execution Sequences

Instrumentation for Monitoring

Monitoring

Operation analysis

Execution sequence analysis

Workload intensity analysis

Instrumentation for Monitoring

Execution sequence analysis

Workload intensity analysis

Prob. density

Response time

:Bookshop

:CRM

:Catalog

0

Matthias Rohr (TrustSoft)

Timing Behavior Anomaly Detection
Creation of the timing behavior profile

Analysis of Execution Sequences

Instrumentation for Monitoring
Monitoring
Operation analysis
Execution sequence analysis
Worload intensity analysis

Response time (ms): Bookshop : CRM : Catalog

Separation to achieve "trace-aware" timing behavior evaluation

Prob. density
Prob. density

Matthias Rohr (TrustSoft)
Prefix of an execution sequence \( t \in T \) of an execution \( e \in E \):

\[
p : T \times E \rightarrow T ; (t, e) \mapsto (m_j)_{j=1}^{i}
\]

with \( m_j \) as pair \((e', e)\).

Example: Prefix of an execution sequence \( t_1 = (a, ac_1, c_1a, ab, bc_2, c_2b, ba)\)
Example: Prefix of an execution sequence

\[ p(t_1, c_1) \]
\[ t_1 = (a, ac_1, c_1a, ab, bc_2, c_2b, ba, a$) \]
Example: Prefix of an execution sequence

\[ p(t_1,c_1) \]

\[ t_1 = (\$a, ac_1, c_1a, ab, bc_2, c_2b, ba, a$) \]
Prefix of an execution sequence $t \in T$ of an execution $e \in E$:

$$p : T \times E \rightarrow T; (t, e) \mapsto (m_i)_{i=1}^j$$

with $m_j$ as pair $(e', e)$.

**Example: Prefix of an execution sequence**

$$t_1 = (\{a, ac_1, c_1a, ab, bc_2, c_2b, ba, a\}$

$$p(t_1, c_2)$$
Distinction of **response times** based on prefixes

The timing behavior observations of an operation $o$ are distinguished based on their prefix $p$, denoted $RT_p = (rt_1, \ldots, rt_n)$. 

$$
RT(a, ac_1) = (40, 37, \ldots) \\
RT(a, ac_1, c_1a, ab, bc_2) = (20, 18, \ldots)
$$
Distinction of *response times* based on prefixes

The timing behavior observations of an operation $o$ are distinguished based on their prefix $p$, denoted $RT_p = (rt_1, \ldots, rt_n)$.

Example:

$$t_1 = t_2 = (a, ac_1, c_1 a, ab, bc_2, c_2 b, ba, a)$$

$$RT_{a, ac_1} = (40, 37, \ldots)$$

$$RT_{a, ac_1, c_1 a, ab, bc_2} = (20, 18, \ldots)$$

<table>
<thead>
<tr>
<th>O</th>
<th>TID</th>
<th>...</th>
<th>$t_{out} - t_{in}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1</td>
<td>...</td>
<td>150</td>
</tr>
<tr>
<td>c</td>
<td>1</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>b</td>
<td>1</td>
<td></td>
<td>80</td>
</tr>
<tr>
<td>c</td>
<td>1</td>
<td></td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>2</td>
<td></td>
<td>18</td>
</tr>
<tr>
<td>c</td>
<td>2</td>
<td></td>
<td>37</td>
</tr>
</tbody>
</table>
All response times

\[ RT \]

Response times per operation

\[ RT(a) = (150, \ldots ) \]
\[ RT(b) = (80, \ldots ) \]
\[ RT(c) = (20, 40, 18, 37, \ldots ) \]

Distinction based on prefix

\[ RT_{p1} = (150, \ldots ) \]
\[ RT_{p2} = (80, \ldots ) \]
\[ RT_{p3} = (20, 18, \ldots ) \]
\[ RT_{p4} = (40, 37, \ldots ) \]

\[ p3 = (a, ac) \]
\[ p4 = (a, ac, c, ab, bc) \]
The workload intensity during an execution influences the response times.
- The workload intensity during an execution influences the response times.
- What is the expected response time distribution of an operation for a particular workload intensity?
The workload intensity during an execution influences the response times.

What is the expected response time distribution of an operation for a particular workload intensity?

Metric for workload intensity $w(e)$:

- Average number of active application threads during the operation execution $e$
Process:

- Determine the workload intensity for each execution monitored
Process:

- Determine the workload intensity for each execution monitored

\[ RT_p = (rt_1, ..., rt_n) \] is extended to

\[ RT'_p = ((rt_1, w_1), \ldots, (rt_n, w_n)) \]
Process:

- Determine the workload intensity for each execution monitored

\[ RT_p = (rt_1, \ldots, rt_n) \] is extended to

\[ RT'_p = ((rt_1, w_1), \ldots, (rt_n, w_n)) \]

- Approximation of normalized probability density functions

\[ f_{RT_p}^w : \mathbb{R} \rightarrow [0, 1]; rt \mapsto f_{RT_p}^w (rt) \]
Process:

- Determine the workload intensity for each execution monitored

\[ RT_p = (rt_1, ..., rt_n) \] is extended to

\[ RT'_p = ((rt_1, w_1), \ldots, (rt_n, w_n)) \]

- Approximation of normalized probability density functions

\[ f^w_{RT_p} : \mathbb{R} \rightarrow [0, 1]; rt \mapsto f^w_{RT_p}(rt) \]

- Example: Approximated normal distributions for response times in dependence to the workload intensity \( w \) (normalized to [0, 1])
Process:

- Determine the workload intensity for each execution monitored
- \( RT_p = (rt_1, ..., rt_n) \) is extended to \( RT'_p = ((rt_1, w_1), ..., (rt_n, w_n)) \)
- Approximation of normalized probability density functions \( f^{w}_{RT_p} : \mathbb{R} \rightarrow [0, 1]; \ rt \mapsto f^{w}_{RT_p}(rt) \)
- Example: Approximated normal distributions for response times in dependence to the workload intensity \( w \) (normalized to \( [0, 1] \))
Process:

- Determine the workload intensity for each execution monitored

\[ RT_p = (rt_1, ..., rt_n) \] is extended to

\[ RT'_p = ((rt_1, w_1), \ldots, (rt_n, w_n)) \]

- Approximation of normalized probability density functions

\[ f^w_{RT_p} : \mathbb{R} \rightarrow [0, 1]; \ rt \rightarrow f^w_{RT_p}(rt) \]

- Example: Approximated normal distributions for response times in dependence to the workload intensity \( w \) (normalized to \([0, 1]\))
Process:

- Determine the workload intensity for each execution monitored
- $RT_p = (rt_1, ..., rt_n)$ is extended to $RT'_p = ((rt_1, w_1), ..., (rt_n, w_n))$
- Approximation of normalized probability density functions $f_{RT_p}^w : \mathbb{R} \rightarrow [0, 1]; rt \mapsto f_{RT_p}^w(rt)$
- Example: Approximated normal distributions for response times in dependence to the workload intensity $w$ (normalized to $[0, 1]$)
**Creation of the timing behavior profile in summary**

<table>
<thead>
<tr>
<th>All monitored response times</th>
<th>Response times per operation</th>
<th>Distinction based on prefix</th>
<th>Modeling of workload intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT(a) = (150, ...)</td>
<td>RT(p1) = (150, ...)</td>
<td>f_{RT_{p1}}(rt)</td>
<td></td>
</tr>
<tr>
<td>RT(b) = (80, ...)</td>
<td>RT(p2) = (80, ...)</td>
<td>f_{RT_{p2}}(rt)</td>
<td></td>
</tr>
<tr>
<td>RT(c) = (20, 40, 18, 37, ...)</td>
<td>RT(p3) = (20, 18, ...)</td>
<td>f_{RT_{p3}}(rt)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RT(p4) = (40, 37, ...)</td>
<td>f_{RT_{p4}}(rt)</td>
<td></td>
</tr>
<tr>
<td>p1 = ($a$)</td>
<td>p3 = ($a, a_1$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>p2 = ($a, a_1, c_1, a, b$)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p4 = ($a, a_1, c_1, a, b, c_2$)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Timing behavior profile**

The timing behavior profile consists of a function $f_{RT_p}^w$ for each prefix of the monitoring data. The values $f_{RT_p}^w(rt) \in [0, 1]$ describe how “normal” a response time $rt$ is under consideration of a workload intensity $w$ and a prefix $p$. 
Outline

1. Foundations
2. Creation of the timing behavior profile
3. Fault Localization
4. Evaluation
5. Related work
6. Conclusions
Overview Fault Localization

Initial Activities

- Instrumentation for Monitoring
- Monitoring
- Creation of Timing behavior profile
- Timing behavior profile

Continuous activities

- Monitoring
- Update of Timing behavior profile
- Timing behavior profile

Activities during diagnosis after detection of a failure

- Monitoring Log of some time period before the failure:
  - Response times
  - Execution sequences
- Timing behavior profile
- Anomaly detection
- Anomaly analysis
- Diagnosis report
After *detection* of a failure at time $t_a$:

- Determination of response times, prefixes and workload intensities (for each execution) for the time period $[t_a - \delta, t_a]$: 
  
<table>
<thead>
<tr>
<th>TraceID</th>
<th>in</th>
<th>out</th>
<th>out - in</th>
<th>prefix</th>
<th>workload intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalog.getBook(..)</td>
<td>121</td>
<td>1182</td>
<td>1201</td>
<td>19</td>
<td>p17</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bookshop.query(..)</td>
<td>131</td>
<td>1195</td>
<td>1221</td>
<td>26</td>
<td>p41</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
  
Anomaly detection through computation of $1 - f_w^{RT_p}(rt)$: 

<table>
<thead>
<tr>
<th>TraceID</th>
<th>in</th>
<th>out</th>
<th>out - in</th>
<th>prefix</th>
<th>workload intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalog.getBook(..)</td>
<td>121</td>
<td>1182</td>
<td>1201</td>
<td>19</td>
<td>p2</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bookshop.query(..)</td>
<td>131</td>
<td>1195</td>
<td>1221</td>
<td>26</td>
<td>p41</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$1 - f^{RT_p}(19) = 0.75$

$1 - f^{RT_p}(17) = 0.21$
After *detection* of a failure at time $t_a$:

- Determination of response times, prefixes and workload intensities (for each execution) for the time period $[t_a - \delta, t_a]$:

<table>
<thead>
<tr>
<th>O</th>
<th>TraceID</th>
<th>$t_{in}$</th>
<th>$t_{out}$</th>
<th>$t_{out} - t_{in}$</th>
<th>$p$</th>
<th>$w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalog.getBook(..)</td>
<td>121</td>
<td>1182</td>
<td>1201</td>
<td>19</td>
<td>p17</td>
<td>17</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bookshop.query(..)</td>
<td>131</td>
<td>1195</td>
<td>1221</td>
<td>26</td>
<td>p41</td>
<td>21</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Activities of Fault Localization (1/2)

After *detection* of a failure at time $t_a$:

1. Determination of response times, prefixes and workload intensities (for each execution) for the time period $[t_a - \delta, t_a]$:

```
<table>
<thead>
<tr>
<th>O</th>
<th>TraceID</th>
<th>$t_{in}$</th>
<th>$t_{out}$</th>
<th>$t_{out} - t_{in}$</th>
<th>p</th>
<th>w</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalog.getBook(..)</td>
<td>121</td>
<td>1182</td>
<td>1201</td>
<td>19</td>
<td>p17</td>
<td>17</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bookshop.query(..)</td>
<td>131</td>
<td>1195</td>
<td>1221</td>
<td>26</td>
<td>p41</td>
<td>21</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

2. Anomaly detection through computation of $1 - f_{RT_p}^w(rt)$:
Activities of Fault Localization (1/2)

After detection of a failure at time $t_a$:

1. Determination of response times, prefixes and workload intensities (for each execution) for the time period $[t_a - \delta, t_a]$:

<table>
<thead>
<tr>
<th>O</th>
<th>TraceID</th>
<th>$t_{in}$</th>
<th>$t_{out}$</th>
<th>$t_{out} - t_{in}$</th>
<th>$p$</th>
<th>$w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalog.getBook(..)</td>
<td>121</td>
<td>1182</td>
<td>1201</td>
<td>19</td>
<td>p17</td>
<td>17</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bookshop.query(..)</td>
<td>131</td>
<td>1195</td>
<td>1221</td>
<td>26</td>
<td>p41</td>
<td>21</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. Anomaly detection through computation of $1 - f_{RT_p}^w(rt)$:

<table>
<thead>
<tr>
<th>O</th>
<th>TraceID</th>
<th>$t_{in}$</th>
<th>$t_{out}$</th>
<th>$t_{out} - t_{in}$</th>
<th>$p$</th>
<th>$w$</th>
<th>$1 - f_{RT_p}^w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalog.getBook(..)</td>
<td>121</td>
<td>1182</td>
<td>1201</td>
<td>19</td>
<td>p2</td>
<td>17</td>
<td>$1 - f_{RT_p2}^w(19) = 0.75$</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bookshop.query(..)</td>
<td>131</td>
<td>1195</td>
<td>1221</td>
<td>26</td>
<td>p41</td>
<td>21</td>
<td>$1 - f_{RT_p41}^w(17) = 0.21$</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Anomaly analysis: Aggregation of many anomaly values

- Mean degree of anomaly for each operation / component / deployment context
- Analysis of anomalies in combination with component dependency graphs
- Neural networks [Stransky, 2006]
- Event correlation techniques [Steinder and Sethi, 2004]
Activities of fault localization (2/2)

3. **Anomaly analysis: Aggregation of many anomaly values**
   - Mean degree of anomaly for each operation / component / deployment context
   - Analysis of anomalies in combination with component dependency graphs
   - Neural networks [Stransky, 2006]
   - Event correlation techniques [Steinder and Sethi, 2004]

4. **Presentation of results (diagnosis report)**

---

![Diagram showing components and probabilities](attachment:diagram.png)
Outline

1. Foundations
2. Creation of the timing behavior profile
3. Fault Localization
4. Evaluation
   - Lab studies
   - Field studies
5. Related work
6. Conclusions
Evaluation goals:
- Proof of concept: Failure diagnosis for injected faults
- Efficiency of anomaly detection and anomaly analysis

Setting:
- Generation of artificial (probabilistic) system usage
- Fault injection
- Example applications:
  - Sun Java PetStore Demo Application, (and reimplementations)
  - (Rubis Benchmark)
  - TPC-App Benchmark
Evaluation goals:

- Applicability in real world systems
  - Complex system usage
  - Long execution sequences
- Effectiveness of anomaly detection
Evaluation – Field studies

Evaluation goals:

- Applicability in real world systems
  - Complex system usage
  - Long execution sequences
- Effectiveness of anomaly detection

Field study in progress:

- Evaluation of 12 month timing behavior data from a customer portal of a middle-size telecommunication company (only highly aggregated response times)
Evaluation goals:
- Applicability in real world systems
  - Complex system usage
  - Long execution sequences
- Effectiveness of anomaly detection

Field study in progress:
- Evaluation of 12 month timing behavior data from a customer portal of a middle-size telecommunication company (only highly aggregated response times)

Field studies in preparation:
- Telecommunication system of Siemens
Evaluation – Field studies

Evaluation goals:
- Applicability in real world systems
  - Complex system usage
  - Long execution sequences
- Effectiveness of anomaly detection

Field study in progress:
- Evaluation of 12 month timing behavior data from a customer portal of a middle-size telecommunication company (only highly aggregated response times)

Field studies in preparation:
- Telecommunication system of Siemens
- E-learning management platform StudIP
Related work

Failure diagnosis based on analysis of timing behavior:

Related work

Failure diagnosis based on analysis of (component) execution sequences:

Failure diagnosis based on multiple runtime behavior metrics:
Related work

Failure diagnosis based on analysis of timing behavior:
- [Agarwal et al., 2004] Response time analysis in the context of average historic response times and SLA violations
- [Diaconescu and Murphy, 2005]: Anomalies as violations of relative threshold values (based on historic average)

Failure diagnosis based on analysis of (component) execution sequences:

Failure diagnosis based on multiple runtime behavior metrics:
Related work

Failure diagnosis based on analysis of timing behavior:

- [Agarwal et al., 2004] Response time analysis in the context of average historic response times and SLA violations
- [Diaconescu and Murphy, 2005]: Anomalies as violations of relative threshold values (based on historic average)

Failure diagnosis based on analysis of (component) execution sequences:

- [Kiciman and Fox, 2005; Kiciman, 2005; Aguilera et al., 2003; Barham et al., 2004]: Component interaction probabilities and component dependency graphs

Failure diagnosis based on multiple runtime behavior metrics:
Related work

Failure diagnosis based on analysis of timing behavior:
- [Agarwal et al., 2004] Response time analysis in the context of average historic response times and SLA violations
- [Diaconescu and Murphy, 2005]: Anomalies as violations of relative threshold values (based on historic average)

Failure diagnosis based on analysis of (component) execution sequences:
- [Kiciman and Fox, 2005; Kiciman, 2005; Aguilera et al., 2003; Barham et al., 2004]: Component interaction probabilities and component dependency graphs

Failure diagnosis based on multiple runtime behavior metrics:
- [Cohen et al., 2005]: Monitoring and evaluation of 62 platform metrics for failure diagnosis, response times of diagnosis of performance problems
- [Salfner and Malek, 2005]: Prediction of failures based on runtime behavior monitoring (for rejuvenation)
Conclusions

- New approach to the detection of timing behavior anomalies for the localization of faults
- Improvement of timing behavior analysis:
  - Workload intensity awareness
  - Awareness of service demand characteristics
- Anomaly detection is used to increase availability and reliability of enterprise-scale software systems
Conclusions

- New approach to the detection of timing behavior anomalies for the localization of faults
- Improvement of timing behavior analysis:
  - Workload intensity awareness
  - Awareness of service demand characteristics
- Anomaly detection is used to increase availability and reliability of enterprise-scale software systems
- Empirical evaluation requires much effort (fault injection & complex usage)


I. Cohen, S. Zhang, M. Goldszmidt, J. Symons, T. Kelly, and A. Fox. Capturing, indexing, clustering, and retrieving system history. In *SOSP ’05:


