Log-based Event Gaps Analysis

S. Suriadi, C. Ouyang, W.M.P. van der Aalst, and A.H.M. ter Hofstede

Information Systems School
Queensland University of Technology
Brisbane, Australia

s.suriadi@qut.edu.au

June 13, 2013
About Myself

- PhD in Information Security (till late 2010) - QUT
  - designing privacy-enhancing cryptographic protocols
  - verifying privacy properties with CPN and state space analysis
- Research Assistant, then Postdoc (2007 - 2011) - QUT
  - Web services security
  - Mitigating denial of service attacks
  - Cryptographic client puzzles
- Postdoc (2011 - now) - QUT
  - BPM Group
  - Process mining - ARC Discovery Project - Risk Aware BPM grant (2011-2013)
My Research Interests

- Process mining
  - Event gap analysis (this talk)
    - Working paper
  - Visualization of log
    - Work with Massimiliano - journal submission soon?
  - A framework for 'Root cause analysis' with enriched log
    - Approach formalization - BPI Workshop 2012
  - Quality and reliability of process mining results
    - Inspired by application of process mining (health and finance)
      - CAiSE 2013
    - Data quality - DECRA proposal

- Process mining for information/network security
  - Similarities with traffic flow analysis
  - Opportunities?
Event Gap Analysis

Problem

- Background: originally interested in measuring resources’ workload
  - BPI Workshop 2012 paper: assume start and complete timestamps
  - Realistic?
- Event log may contain limited information (e.g. only ‘complete’ or ‘start’ timestamp)
- May hinder certain type of analyses, e.g. performance analysis
  - how to obtain waiting time of cases/resources?

Question

How can we still obtain a good estimation of the performance of cases and resources from the log with limited event transaction lifecycle information?
An approach to support the analysis of case and resource performance based on the event logs with minimum (timestamp) information.

- Exploits the use of time duration (or gap) between any two adjacent events referring to the same case/resource
- May yield a better estimation of case and resource performance
- Copes with event logs with limited information (e.g. only one event transaction lifecycle timestamp)
- Implemented in ProM
Basic Performance Analysis plug-in
- Estimate the working time of tasks
- Can work with only one event transaction type.
- Task-view only, not considering other dimensions
  - Our approach supports other views (case, resource) and exploits relationships between those views

Resource availability analysis (Nakatumbia’s work)
Use ‘event gaps’ as the basic building blocks for performance analysis

- $\mathcal{E}$ is the set of events
  - Definition of event - see van der Aalst (2011).
- An event log $\mathcal{L} \subseteq \mathcal{E}$ is a set of events.
- Ordered by timestamps, i.e., $e_1 < e_2$ if and only if $time(e_1) < time(e_2)$ and $e_1 \leq e_2$ if and only if $time(e_1) \leq time(e_2)$.

**Gap**

Let $e_1, e_2 \in \mathcal{E}$ with $e_1 < e_2$ : $(e_1, e_2)$ is a gap.

$\mathcal{G} = \{(e_1, e_2) \in \mathcal{E} \times \mathcal{E} \mid e_1 < e_2\}$ is the set of all possible gaps.

$\text{dur}(e_1, e_2) = time(e_2) - time(e_1)$ is the duration of gap $(e_1, e_2)$. 
Given a log, the existence of various types of gaps can be identified and calculated,

Types of gaps:
- Case gap
- Resource gap
- Minimum gap
- Case-Resource gap
- Resource-Case gap

These gaps form the basic building blocks for subsequent (performance) analyses.
Types of Gap

Case Gap

\[ CG(\mathcal{L}) = \{(e_1, e_2) \in \mathcal{G} \cap (\mathcal{L} \times \mathcal{L}) \mid case(e_1) = case(e_2) \land (\nexists e_3 \in \mathcal{L} \ e_1 < e_3 < e_2 \land case(e_3) = case(e_2))\} \]

We use the notation \( cg_{e_j} \) to denote \((e_i, e_j) \in CG(\mathcal{L})\).
Resource Gap

\[ RG(\mathcal{L}) = \{ (e_1, e_2) \in \mathcal{G} \cap (\mathcal{L} \times \mathcal{L}) \mid res(e_1) = res(e_2) \land \nexists e_3 \in \mathcal{L} \ e_1 < e_3 < e_2 \land res(e_3) = res(e_2) \} \].

We use the notation \( rg_{e_j} \) to denote \((e_i, e_j) \in RG(\mathcal{L})\).
Minimum Gap

\[ \text{MinG}(\mathcal{L}) = \{(e_1, e_2) \in \mathcal{G} \cap (\mathcal{L} \times \mathcal{L}) \mid (\text{res}(e_1) = \text{res}(e_2) \lor \text{case}(e_1) = \text{case}(e_2)) \land (\nexists e_3 \in \mathcal{L} \ e_1 < e_3 < e_2 \land (\text{res}(e_3) = \text{res}(e_2) \lor \text{case}(e_3) = \text{case}(e_2)))) \} . \]

We use the notation \( mg_{e_j} \) to denote \((e_i, e_j) \in \text{MinG}(\mathcal{L}) \). A \textit{minimum gap} is the shorter of either a case gap or resource gap.
Resource-Case Gap

\[ RCG(\mathcal{L}) = \{ (e_1, e_2) \in \mathcal{G} \cap (\mathcal{L} \times \mathcal{L}) \mid \exists e_3 \in \mathcal{L} : e_2 < e_3 \land (e_1, e_3) \in RG(\mathcal{L}) \land (e_2, e_3) \in CG(\mathcal{L}) \}. \]

We use the notation \( rcg(e_i, e_j) \) to denote \( (e_i, e_j) \in RCG(\mathcal{L}) \). Exists when resource gap is longer than the corresponding case gap.
Case-Resource Gap

\[ CRG(\mathcal{L}) = \{(e_1, e_2) \in \mathcal{G} \cap (\mathcal{L} \times \mathcal{L}) \mid \exists e_3 \in \mathcal{L} : e_2 < e_3 \land (e_1, e_3) \in CG(\mathcal{L}) \land (e_2, e_3) \in RG(\mathcal{L})\} \].

We use the notation \( crg(e_i, e_j) \) to denote \((e_i, e_j) \in CRG(\mathcal{L}) \). Exists when case gap is longer than the corresponding resource gap.
Interpretation of Gaps

Meaning of Gaps

- Assume:
  - Any two events considered for gaps calculation correspond to work items not executed in parallel.
  - Only consider ‘complete’ event type.
- Case gap ($cg_{e_j}$): expected working time of activity$(e_j)$ in a case?
- Resource gap ($rg_{e_j}$): expected working time of activity$(e_j)$ by a resource?
- Minimum gap ($mg_{e_j}$): the actual working time of activity$(e_j)$?
  - both resource and case are available
- Case-resource gap ($crg_{(e_i,e_j)}$): case waiting time for activity$(e_j)$
- Resource-case gap ($rcg_{(e_i,e_j)}$): resource waiting time for activity$(e_j)$
Event gaps represent basic performance data unit ➔ ‘isolated’, fine-grained

‘Group by’ allows an aggregation of gaps based on certain characteristics, e.g.
- resource,
- activity,
- team,
- null, etc.

By grouping calculated event gaps using different characteristics, we can derive more meaningful results.
### Group By

<table>
<thead>
<tr>
<th>caseID</th>
<th>activity</th>
<th>timestamp</th>
<th>resource</th>
<th>caseResourceGap</th>
<th>minGap</th>
<th>resourceGap</th>
<th>resourceCaseGap</th>
<th>caseGap</th>
</tr>
</thead>
<tbody>
<tr>
<td>341</td>
<td>Repair (Complex)</td>
<td>1970-01-09 16:52:00</td>
<td>SolverC1</td>
<td>720000</td>
<td></td>
<td></td>
<td></td>
<td>720000</td>
</tr>
<tr>
<td>387</td>
<td>Repair (Complex)</td>
<td>1970-01-09 18:01:00</td>
<td>SolverC1</td>
<td>0</td>
<td>300000</td>
<td>4140000</td>
<td>3840000</td>
<td>300000</td>
</tr>
<tr>
<td>369</td>
<td>Repair (Complex)</td>
<td>1970-01-09 22:24:00</td>
<td>SolverC1</td>
<td>0</td>
<td>2040000</td>
<td>15780000</td>
<td>13740000</td>
<td>2040000</td>
</tr>
<tr>
<td>412</td>
<td>Repair (Complex)</td>
<td>1970-01-10 00:44:00</td>
<td>SolverC1</td>
<td>1500000</td>
<td></td>
<td></td>
<td></td>
<td>1500000</td>
</tr>
<tr>
<td>379</td>
<td>Repair (Complex)</td>
<td>1970-01-10 01:35:00</td>
<td>SolverC1</td>
<td>0</td>
<td>180000</td>
<td>3060000</td>
<td>2880000</td>
<td>180000</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>caseID</th>
<th>activity</th>
<th>timestamp</th>
<th>resource</th>
<th>caseResourceGap</th>
<th>minGap</th>
<th>resourceGap</th>
<th>resourceCaseGap</th>
<th>caseGap</th>
</tr>
</thead>
<tbody>
<tr>
<td>354</td>
<td>Test Repair</td>
<td>1970-01-09 16:22:00</td>
<td>Tester1</td>
<td>600000</td>
<td></td>
<td></td>
<td></td>
<td>600000</td>
</tr>
<tr>
<td>351</td>
<td>Analyze Defect</td>
<td>1970-01-09 16:39:00</td>
<td>Tester1</td>
<td>0</td>
<td>420000</td>
<td>1020000</td>
<td>600000</td>
<td>420000</td>
</tr>
<tr>
<td>341</td>
<td>Test Repair</td>
<td>1970-01-09 16:58:00</td>
<td>Tester1</td>
<td>0</td>
<td>360000</td>
<td>1140000</td>
<td>780000</td>
<td>360000</td>
</tr>
<tr>
<td>387</td>
<td>Analyze Defect</td>
<td>1970-01-09 17:32:00</td>
<td>Tester1</td>
<td>0</td>
<td>360000</td>
<td>2040000</td>
<td>1680000</td>
<td>360000</td>
</tr>
<tr>
<td>337</td>
<td>Test Repair</td>
<td>1970-01-09 18:06:00</td>
<td>Tester1</td>
<td>0</td>
<td>480000</td>
<td>2040000</td>
<td>1560000</td>
<td>480000</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

- For each gap type, and for each ‘group by’ on a specific characteristic, we can do three types of analysis:
  - Metrics
  - Decision Tree
  - Gaps Evolution over Time
### Gap Analysis

#### Group By

<table>
<thead>
<tr>
<th>caseID</th>
<th>activity</th>
<th>timestamp</th>
<th>resource</th>
<th>caseResourceGap</th>
<th>minGap</th>
<th>resourceGap</th>
<th>resourceCaseGap</th>
<th>caseGap</th>
</tr>
</thead>
<tbody>
<tr>
<td>341</td>
<td>Repair (Complex)</td>
<td>1970-01-09 16:52:00</td>
<td>SolverC1</td>
<td>720000</td>
<td></td>
<td></td>
<td></td>
<td>720000</td>
</tr>
<tr>
<td>387</td>
<td>Repair (Complex)</td>
<td>1970-01-09 18:01:00</td>
<td>SolverC1</td>
<td>0</td>
<td>300000</td>
<td>4140000</td>
<td>3840000</td>
<td>300000</td>
</tr>
<tr>
<td>369</td>
<td>Repair (Complex)</td>
<td>1970-01-09 22:24:00</td>
<td>SolverC1</td>
<td>0</td>
<td>2040000</td>
<td>1578000</td>
<td>1374000</td>
<td>2040000</td>
</tr>
<tr>
<td>412</td>
<td>Repair (Complex)</td>
<td>1970-01-10 00:44:00</td>
<td>SolverC1</td>
<td>1500000</td>
<td></td>
<td></td>
<td></td>
<td>1500000</td>
</tr>
<tr>
<td>379</td>
<td>Repair (Complex)</td>
<td>1970-01-10 01:35:00</td>
<td>SolverC1</td>
<td>0</td>
<td>180000</td>
<td>3060000</td>
<td>2880000</td>
<td>180000</td>
</tr>
</tbody>
</table>

For each gap type, and for each ‘group by’ on a specific characteristic, we can do three types of analysis:

- Metrics
- Decision Tree
- Gaps Evolution over Time

---

For each gap type, and for each ‘group by’ on a specific characteristic, we can do three types of analysis:

- Metrics
- Decision Tree
- Gaps Evolution over Time
For each gap type, and for each ‘group by’ on a specific characteristic, we can do three types of analysis:

- Metrics
- Decision Tree
- Gaps Evolution over Time
Gap Analysis

Metrics

Metrics supported: mean, median, and standard deviation

Gap type: resource gap; Group by: resource
### Gap Analysis

#### Decision Tree

```
<table>
<thead>
<tr>
<th>caseID</th>
<th>activity</th>
<th>timestamp</th>
<th>resource</th>
<th>caseResourceGap</th>
<th>minGap</th>
<th>resourceGap</th>
<th>resourceCaseGap</th>
<th>caseGap</th>
</tr>
</thead>
<tbody>
<tr>
<td>341</td>
<td>Repair (Complex)</td>
<td>1970-01-09 16:52:00</td>
<td>SolverC1</td>
<td>720000</td>
<td></td>
<td></td>
<td>720000</td>
<td></td>
</tr>
<tr>
<td>387</td>
<td>Repair (Complex)</td>
<td>1970-01-09 18:01:00</td>
<td>SolverC1</td>
<td>0</td>
<td>300000</td>
<td>4140000</td>
<td>3840000</td>
<td>300000</td>
</tr>
<tr>
<td>369</td>
<td>Repair (Complex)</td>
<td>1970-01-09 22:24:00</td>
<td>SolverC1</td>
<td>0</td>
<td>2040000</td>
<td>1578000</td>
<td>1374000</td>
<td>2040000</td>
</tr>
<tr>
<td>412</td>
<td>Repair (Complex)</td>
<td>1970-01-10 00:44:00</td>
<td>SolverC1</td>
<td>1500000</td>
<td></td>
<td></td>
<td>1500000</td>
<td></td>
</tr>
<tr>
<td>379</td>
<td>Repair (Complex)</td>
<td>1970-01-10 01:35:00</td>
<td>SolverC1</td>
<td>0</td>
<td>180000</td>
<td>3060000</td>
<td>2880000</td>
<td>180000</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>354</td>
<td>Test Repair</td>
<td>1970-01-09 16:22:00</td>
<td>Tester1</td>
<td>600000</td>
<td></td>
<td></td>
<td>600000</td>
<td></td>
</tr>
<tr>
<td>351</td>
<td>Analyze Defect</td>
<td>1970-01-09 16:39:00</td>
<td>Tester1</td>
<td>0</td>
<td>420000</td>
<td>1020000</td>
<td>600000</td>
<td>420000</td>
</tr>
<tr>
<td>341</td>
<td>Test Repair</td>
<td>1970-01-09 16:58:00</td>
<td>Tester1</td>
<td>0</td>
<td>360000</td>
<td>1140000</td>
<td>780000</td>
<td>360000</td>
</tr>
<tr>
<td>387</td>
<td>Analyze Defect</td>
<td>1970-01-09 17:32:00</td>
<td>Tester1</td>
<td>0</td>
<td>360000</td>
<td>2040000</td>
<td>1680000</td>
<td>360000</td>
</tr>
<tr>
<td>337</td>
<td>Test Repair</td>
<td>1970-01-09 18:06:00</td>
<td>Tester1</td>
<td>0</td>
<td>480000</td>
<td>2040000</td>
<td>1560000</td>
<td>480000</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
```

### Case Gap Analysis

```
- <= 540000: Low (2220, 421008.8)
- > 540000: High (115358/508.39)
```
### Gap Analysis

#### Decision Tree

<table>
<thead>
<tr>
<th>caseID</th>
<th>activity</th>
<th>timestamp</th>
<th>resource</th>
<th>caseResourceGap</th>
<th>minGap</th>
<th>resourceGap</th>
<th>resourceCaseGap</th>
<th>caseGap</th>
</tr>
</thead>
<tbody>
<tr>
<td>341</td>
<td>Repair (Complex)</td>
<td>1970-01-09 16:52:00</td>
<td>SolverC1</td>
<td>720000</td>
<td>0</td>
<td>4140000</td>
<td>13740000</td>
<td>720000</td>
</tr>
<tr>
<td>387</td>
<td>Repair (Complex)</td>
<td>1970-01-09 18:01:00</td>
<td>SolverC1</td>
<td>300000</td>
<td>0</td>
<td>1578000</td>
<td>2040000</td>
<td>300000</td>
</tr>
<tr>
<td>369</td>
<td>Repair (Complex)</td>
<td>1970-01-09 22:24:00</td>
<td>SolverC1</td>
<td>2040000</td>
<td>0</td>
<td>3060000</td>
<td>2880000</td>
<td>2040000</td>
</tr>
<tr>
<td>412</td>
<td>Repair (Complex)</td>
<td>1970-01-10 00:44:00</td>
<td>SolverC1</td>
<td>1500000</td>
<td>0</td>
<td>13740000</td>
<td>13740000</td>
<td>1500000</td>
</tr>
<tr>
<td>379</td>
<td>Repair (Complex)</td>
<td>1970-01-10 01:35:00</td>
<td>SolverC1</td>
<td>180000</td>
<td>0</td>
<td>3060000</td>
<td>2880000</td>
<td>180000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>caseID</th>
<th>activity</th>
<th>timestamp</th>
<th>resource</th>
<th>caseResourceGap</th>
<th>minGap</th>
<th>resourceGap</th>
<th>resourceCaseGap</th>
<th>caseGap</th>
</tr>
</thead>
<tbody>
<tr>
<td>354</td>
<td>Test Repair</td>
<td>1970-01-09 16:22:00</td>
<td>Tester1</td>
<td>600000</td>
<td>0</td>
<td>1020000</td>
<td>600000</td>
<td>600000</td>
</tr>
<tr>
<td>351</td>
<td>Analyze Defect</td>
<td>1970-01-09 16:39:00</td>
<td>Tester1</td>
<td>420000</td>
<td>0</td>
<td>780000</td>
<td>360000</td>
<td>420000</td>
</tr>
<tr>
<td>341</td>
<td>Test Repair</td>
<td>1970-01-09 16:58:00</td>
<td>Tester1</td>
<td>360000</td>
<td>0</td>
<td>1140000</td>
<td>780000</td>
<td>360000</td>
</tr>
<tr>
<td>387</td>
<td>Analyze Defect</td>
<td>1970-01-09 17:32:00</td>
<td>Tester1</td>
<td>360000</td>
<td>0</td>
<td>2040000</td>
<td>1680000</td>
<td>360000</td>
</tr>
<tr>
<td>337</td>
<td>Test Repair</td>
<td>1970-01-09 18:06:00</td>
<td>Tester1</td>
<td>480000</td>
<td>0</td>
<td>2040000</td>
<td>1560000</td>
<td>480000</td>
</tr>
</tbody>
</table>

#### Event Gaps Analysis

- **Low (2220.42/1008.8)**
- **High (1153.56/508.39)**
### Gap Analysis

#### Decision Tree

<table>
<thead>
<tr>
<th>caseID</th>
<th>activity</th>
<th>timestamp</th>
<th>resource</th>
<th>caseResourceGap</th>
<th>minGap</th>
<th>resourceGap</th>
<th>resourceCaseGap</th>
<th>caseGap</th>
</tr>
</thead>
<tbody>
<tr>
<td>341</td>
<td>Repair (Complex)</td>
<td>1970-01-09 16:52:00</td>
<td>SolverC1</td>
<td>720000</td>
<td>0</td>
<td>4140000</td>
<td>3840000</td>
<td>720000</td>
</tr>
<tr>
<td>387</td>
<td>Repair (Complex)</td>
<td>1970-01-09 18:01:00</td>
<td>SolverC1</td>
<td>0</td>
<td>500000</td>
<td>15780000</td>
<td>13740000</td>
<td>0</td>
</tr>
<tr>
<td>369</td>
<td>Repair (Complex)</td>
<td>1970-01-09 22:24:00</td>
<td>SolverC1</td>
<td>0</td>
<td>2040000</td>
<td>150000</td>
<td>150000</td>
<td>0</td>
</tr>
<tr>
<td>412</td>
<td>Repair (Complex)</td>
<td>1970-01-10 00:44:00</td>
<td>SolverC1</td>
<td>1500000</td>
<td>0</td>
<td>20400000</td>
<td>2040000</td>
<td>1500000</td>
</tr>
</tbody>
</table>

- **Response Variables**
  - caseID: 341
  - activity: Repair (Complex)
  - timestamp: 1970-01-09 16:52:00
  - resource: SolverC1
  - caseResourceGap: 720000
  - minGap: 0
  - resourceGap: 4140000
  - resourceCaseGap: 3840000
  - caseGap: 720000

- **Response Variables**
  - caseID: 387
  - activity: Repair (Complex)
  - timestamp: 1970-01-09 18:01:00
  - resource: SolverC1
  - caseResourceGap: 0
  - minGap: 0
  - resourceGap: 15780000
  - resourceCaseGap: 13740000
  - caseGap: 0

- **Response Variables**
  - caseID: 369
  - activity: Repair (Complex)
  - timestamp: 1970-01-09 22:24:00
  - resource: SolverC1
  - caseResourceGap: 0
  - minGap: 0
  - resourceGap: 2040000
  - resourceCaseGap: 1800000
  - caseGap: 0

- **Response Variables**
  - caseID: 412
  - activity: Repair (Complex)
  - timestamp: 1970-01-10 00:44:00
  - resource: SolverC1
  - caseResourceGap: 1500000
  - minGap: 0
  - resourceGap: 20400000
  - resourceCaseGap: 2040000
  - caseGap: 1500000

- **Response Variables**
  - caseID: 354
  - activity: Test Repair
  - timestamp: 1970-01-09 16:22:00
  - resource: Tester1
  - caseResourceGap: 600000
  - minGap: 0
  - resourceGap: 1020000
  - resourceCaseGap: 600000
  - caseGap: 600000

- **Response Variables**
  - caseID: 351
  - activity: Analyze Defect
  - timestamp: 1970-01-09 16:39:00
  - resource: Tester1
  - caseResourceGap: 0
  - minGap: 0
  - resourceGap: 1140000
  - resourceCaseGap: 780000
  - caseGap: 360000

- **Response Variables**
  - caseID: 341
  - activity: Test Repair
  - timestamp: 1970-01-09 16:58:00
  - resource: Tester1
  - caseResourceGap: 0
  - minGap: 0
  - resourceGap: 2040000
  - resourceCaseGap: 1680000
  - caseGap: 360000

- **Response Variables**
  - caseID: 337
  - activity: Test Repair
  - timestamp: 1970-01-09 18:06:00
  - resource: Tester1
  - caseResourceGap: 0
  - minGap: 0
  - resourceGap: 2040000
  - resourceCaseGap: 1560000
  - caseGap: 480000

---

The decision tree illustrates the classification of case gaps into two categories: Low and High. The tree structure helps in understanding the relationship between case gaps and their resource gaps, enabling a more informed decision-making process.
Gap Analysis

<table>
<thead>
<tr>
<th>caseID</th>
<th>activity</th>
<th>timestamp</th>
<th>resource</th>
<th>caseResourceGap</th>
<th>minGap</th>
<th>resourceGap</th>
<th>resourceCaseGap</th>
<th>caseGap</th>
</tr>
</thead>
<tbody>
<tr>
<td>341</td>
<td>Repair (Complex)</td>
<td>1970-01-09 16:52:00</td>
<td>SolverC1</td>
<td>720000</td>
<td></td>
<td></td>
<td></td>
<td>720000</td>
</tr>
<tr>
<td>387</td>
<td>Repair (Complex)</td>
<td>1970-01-09 18:01:00</td>
<td>SolverC1</td>
<td>0</td>
<td>300000</td>
<td>4140000</td>
<td>3840000</td>
<td>300000</td>
</tr>
<tr>
<td>369</td>
<td>Repair (Complex)</td>
<td>1970-01-09 22:24:00</td>
<td>SolverC1</td>
<td>0</td>
<td>2040000</td>
<td>15780000</td>
<td>13740000</td>
<td>2040000</td>
</tr>
<tr>
<td>369</td>
<td>Repair (Complex)</td>
<td>1970-01-10 00:44:00</td>
<td>SolverC1</td>
<td>0</td>
<td>0</td>
<td>1500000</td>
<td>1500000</td>
<td>1500000</td>
</tr>
<tr>
<td>369</td>
<td>Repair (Complex)</td>
<td>1970-01-10 01:35:00</td>
<td>SolverC1</td>
<td>0</td>
<td>2040000</td>
<td>18000000</td>
<td>16800000</td>
<td>1800000</td>
</tr>
</tbody>
</table>

Response Variables

- 354 Test Repair 1970-01-09 16:22:00 Tester1 600000 420000 1020000 600000 420000
- 351 Analyze Defect 1970-01-09 16:39:00 Tester1 0 420000 1020000 600000 420000
- 341 Test Repair 1970-01-09 16:58:00 Tester1 0 360000 1140000 780000 360000
- 387 Analyze Defect 1970-01-09 17:32:00 Tester1 0 360000 2040000 1680000 360000
- 337 Test Repair 1970-01-09 18:06:00 Tester1 0 480000 2040000 1560000 480000

Decision Tree

- caseGap
  - <= 540000
    - Low (2220.42/1008.8)
  - > 540000
    - High (1153.68/508.39)
Gap Analysis

Decision Tree

<table>
<thead>
<tr>
<th>caseID</th>
<th>activity</th>
<th>timestamp</th>
<th>resource</th>
</tr>
</thead>
<tbody>
<tr>
<td>341</td>
<td>Repair (Complex)</td>
<td>1970-01-09 16:52:00</td>
<td>SolverC1</td>
</tr>
<tr>
<td>387</td>
<td>Repair (Complex)</td>
<td>1970-01-09 18:01:00</td>
<td>SolverC1</td>
</tr>
<tr>
<td>369</td>
<td>Repair (Complex)</td>
<td>1970-01-09 22:24:00</td>
<td>SolverC1</td>
</tr>
<tr>
<td>412</td>
<td>Repair (Complex)</td>
<td>1970-01-10 00:44:00</td>
<td>SolverC1</td>
</tr>
<tr>
<td>379</td>
<td>Repair (Complex)</td>
<td>1970-01-09 16:58:00</td>
<td>SolverC1</td>
</tr>
<tr>
<td>354</td>
<td>Test Repair</td>
<td>1970-01-09 16:22:00</td>
<td>Tester1</td>
</tr>
<tr>
<td>351</td>
<td>Analyze Defect</td>
<td>1970-01-09 16:39:00</td>
<td>Tester1</td>
</tr>
<tr>
<td>341</td>
<td>Test Repair</td>
<td>1970-01-09 16:58:00</td>
<td>Tester1</td>
</tr>
<tr>
<td>387</td>
<td>Analyze Defect</td>
<td>1970-01-09 17:32:00</td>
<td>Tester1</td>
</tr>
<tr>
<td>337</td>
<td>Test Repair</td>
<td>1970-01-09 18:06:00</td>
<td>Tester1</td>
</tr>
</tbody>
</table>

Case Resource Gap: 720000
Min Gap: 0
Resource Gap: 150000
Resource Case Gap: 150000
Case Gap: 180000

Response Variables:
- Low (2220.42/1008.8)
- High (1153.58/508.39)

Predictor Variables:
- 720000
- 300000
- 4140000
- 3840000
- 2040000
- 15780000
- 13740000
- 2040000
- 180000
- 600000
- 420000
- 1020000
- 600000
- 420000
- 360000
- 1140000
- 780000
- 360000
- 360000
- 2040000
- 1680000
- 360000
- 480000
- 2040000
- 1560000
- 480000
Gap Analysis
Evolution of Gaps

Event gaps per resource

Resource gap (or any other type of gap), group by resource.
Putting them all together

<table>
<thead>
<tr>
<th>Gap Type</th>
<th>Group By</th>
<th>Analysis</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case</td>
<td>Resource</td>
<td>Metrics</td>
<td>?</td>
</tr>
<tr>
<td>Resource</td>
<td>Activity</td>
<td>Decision Tree</td>
<td>?</td>
</tr>
<tr>
<td>Minimum</td>
<td>Case</td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>Case-Resource</td>
<td>Team</td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>Resource-Case</td>
<td>Null</td>
<td>Evolution</td>
<td>?</td>
</tr>
</tbody>
</table>

- For each analysis type, we can use 25 different ‘grouping’:
  - 5 gap types $\times$ 5 ‘group by’
- Interpretation for each grouping?
- All useful? Probably not...
ProM Plug-in

- Has been implemented as a ProM plug-in “Event Gap Analysis”
- Input: XES/MXML log
- Configuration:
  - Period of time to use
  - Types of event gap
  - Event transaction lifecycle to consider
Configuration Panel

Specify a specific time range for gap analysis

Choose a specific event lifecycle:
schedule, allocate, start, complete

Select the type of event gaps for analysis:
case gap, resource gap, minimum gap,
case resource gap, resource case gap
ProM Plug-in

ProM Plug-in Visualization Panel

Select a gap type for result visualization

Select how the chosen gaps should be grouped for result visualization purposes.

Select a particular analysis type for visualization: metrics, decision tree, or gaps evolution

If 'decision tree' analysis is selected, then users need to select relevant predictor variables

If 'gap evolution' analysis is selected, then users need to specify the relevant window size

Result Visualization Panel
ProM Plug-in

ProM Plug-in
Visualization Panel

Select a gap type for result visualization.

Select how the chosen gaps should be grouped for result visualization purposes.

Select a particular analysis type for visualization: metrics, decision tree, or gaps evolution.

If 'decision tree' analysis is selected, then users need to select relevant predictor variables.

If 'gap evolution' analysis is selected, then users need to specify the relevant window size.

Result Visualization Panel

Tool Demo
Validation - Application to Logs

- Still need to be properly conducted
- Preliminary attempts with Suncorp data
- Results
  - Evolution graph shows peaks consistent with business hours
  - Distinct differences between resources
  - Decision tree: unclear
    - Make sense to correlate one gap (e.g. case gap) with another gap (e.g. resource gap)?
- Challenges:
  - Unrealistic short gap durations (in seconds/miliseconds)
  - May be due to log quality (imprecise recording of events)
Summary

- Event gaps analysis may provide useful insights about process behaviours, especially performance
- Conceptually straight-forward, potentially useful
- Has been implemented as a ProM plug-in
- Next steps:
  - Proper validation using a variety of logs
  - Suncorp, hospitals, BPI 2013 challenge
  - Exploit certain configurations of event gaps analysis (i.e. basic gaps + group by + analysis type)
    - to address typical business analysis questions,
    - leading to tool improvement to suit domain (business) users.
  - Publication strategy: conference then journal
- Collaboration sought
Log-based Event Gaps Analysis

S. Suriadi, C. Ouyang, W.M.P. van der Aalst, and A.H.M. ter Hofstede

Information Systems School
Queensland University of Technology
Brisbane, Australia

s.suriadi@qut.edu.au

June 13, 2013