Runtime resource prediction of component-based applications

Ionut David, Rudolf H. Mak

Research context
The Visual Context Modeling (ViCoMo) project [1]
- addresses multimedia applications (see Fig. 1)
- consisting of pipelines of video processing components
- running on distributed, heterogeneous platforms
- exhibiting high and/or highly fluctuating resource demands

Figure 1. A pipelined video surveillance application

In this project we focus on runtime composition and resource management [2].

Research question
Develop resource reservation strategies for

- processing power
- memory capacity
- network bandwidth

Research contributions
For processing power only:
- State-based utilization and reservation model
- Monitoring method to obtain utilization trajectories
- Reservation strategies: periodic and state-based
- reservation for the next period depends on the utilization state of the current period
- Metrics for resource prediction strategy quality

State-based resource model
For each component and resource, utilization $u$ and reservation $r$ are expressed as a percentage of total resource capacity, and clustered into $n_s$ states of fixed size $S$:

$$\sigma_j = \{ u | (j-1)S \leq u < jS \}, \text{ where } 1 \leq j < n_s$$

Monitoring components for $N$ reservation periods of duration $P$, starting time $t_0$, we obtain:

- utilization trajectory $U_N = \{ u_i | 0 \leq i < N \}$
- reservation trajectory $R_N = \{ r_i | 0 \leq i < N \}$

where $u_i$ and $r_i$ are utilization and reservation in the period starting at time $t_i = t_{i-1} + iP$.  

![Figure 2. Utilization state frequency profile](image)

Fig. 2 displays utilization state frequency profile of a measured trajectory, where $f_i = \# \{ u_i \in \sigma_j | 0 \leq i < N \}/N$

![Figure 3. State transition frequency profile](image)

Fig. 3 shows a state transitions profile, where $\phi(j, t) = \# \{ u_i \in S \wedge u_{i+1} \in t | 0 \leq i < N-1 \} / (N-1)$

Prediction strategies
Strategy $k$ reserves enough resource to accommodate an increase of $k$ states in resource utilization, i.e.,

- $r_{u+1} = \{ \text{max} (j-k, n_s) \} S$ where $u_j \in \sigma_j$

Quality of the strategy is measured by the metrics:

- failure rate $F_k(U_N, R_N) = \# \{ r_j < u_j | 0 \leq j < N \}/N$
- waste $W_k(U_N, R_N) = \{ \sum_{j \leq i < N} \text{max} (0, r_j-u_j) \}/N$

Results
Measurements on a 1 hour video for various $(k, S)$ yield

<table>
<thead>
<tr>
<th>$k$</th>
<th>$P_k$</th>
<th>$W_k$</th>
<th>$F_k$</th>
<th>$R_k$</th>
<th>$W_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.396</td>
<td>2.164</td>
<td>0.397</td>
<td>2.035</td>
<td>0.206</td>
</tr>
<tr>
<td>2</td>
<td>0.397</td>
<td>2.164</td>
<td>0.206</td>
<td>3.041</td>
<td>0.088</td>
</tr>
<tr>
<td>3</td>
<td>0.266</td>
<td>2.032</td>
<td>0.066</td>
<td>6.375</td>
<td>0.002</td>
</tr>
<tr>
<td>4</td>
<td>0.265</td>
<td>2.190</td>
<td>0.064</td>
<td>7.271</td>
<td>0.001</td>
</tr>
<tr>
<td>5</td>
<td>0.158</td>
<td>2.658</td>
<td>0.001</td>
<td>10.932</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Pareto points in red show the failure-waste trade-off

![References](image)


i.david@tue.nl