Enhancing the Applicability of the eST-Miner: Efficient Precision-Guided Implicit Place Avoidance

Felix C. Groß, Lisa L. Mannel, Wil M.P. van der Aalst
Introduction
Idea: eST-Miner

Input: $L$ event log, $\tau \in [0,1]$ noise threshold

Output: Petri net

1. Start with an “empty” Petri net $P$ that consists of
   - A dedicated source place
   - A dedicated target place
   - A transition for each activity contained in the event log

2. Consider all possible places individually

3. Evaluate using token-based replay and add the fitting ones to $P$
\[ L = [(\text{\begin{small}a\end{small}}, \text{\begin{small}•\end{small}}), (\text{\begin{small}b\end{small}}, \text{\begin{small}•\end{small}})] \]

Candidate Tree: \( CT_{\{\begin{small}a, b, •\end{small}\}} \)

\[
\begin{align*}
\text{(a)} & \quad \text{(b)} & \quad \langle \text{•} \rangle \\
\text{(a\text{\begin{small}a\end{small}})} & \quad \text{(a\text{\begin{small}b\end{small}})} & \quad \text{(a\text{\begin{small}•\end{small}})} \\
\text{(b\text{\begin{small}a\end{small}})} & \quad \text{(b\text{\begin{small}b\end{small}})} & \quad \text{(b\text{\begin{small}•\end{small}})} \\
\text{(a\text{\begin{small}b\end{small}})} & \quad \text{(b\text{\begin{small}•\end{small}})} & \quad \text{(\text{\begin{small}•\end{small}}\text{\begin{small}b\end{small}})} \\
\end{align*}
\]

\[ \Rightarrow \text{Traverse the candidate tree in BFS (evaluate each place using TBR)} \]
Traversal Strategy

\[ L = [\langle \text{\textrightarrow}, a, \square \rangle, \langle \text{\textrightarrow}, b, \square \rangle] \]

\[ p = (a|b) \]
Traversal Strategy

\[ L = \langle \langle \sigma_1, \alpha, \rangle, \langle \sigma_2, \beta, \rangle \rangle \]

\[ p = (a|b) \]

\[ a \quad ! \quad b \]

\[ p \text{ is } \textit{underfed} \text{ concerning } \sigma_2 \]

\[ \Rightarrow \text{ all places of the form } a \quad B \quad b \text{ are underfed concerning } \sigma_2 \text{ as well} \]
Traversal Strategy
Traversal Strategy
Traversing Strategy

\[ L = [\langle \text{\textbullet}, a, \text{■} \rangle, \langle \text{\textbullet}, b, \text{■} \rangle] \]

\[ p = (a|b) \]

Diagram showing traversal strategy with states and transitions.
Traversal Strategy

\[ L = \left[ \sigma_1, \sigma_2 \right] \]

\[ p = (a|b) \]

\[ p \text{ is overfed concerning } \sigma_1 \]

⇒ all places of the form

\[ A \]

are overfed concerning \( \sigma_1 \) as well
Traversal Strategy
Traversal Strategy

subtree of red children may contain nodes attached by blue edges
Handling Noise

Local noise threshold $\tau \in [0,1]$

Store places that fit at least $\tau \cdot |L|$ traces

- $\tau = 1$ resulting Petri net can perfectly replay $L$
- $\tau < 1$ fitting places may contradict each other $\Rightarrow$ deadlocks

$\delta$-Variant:

- Only inserts places that do not decrease the fraction of replayable traces by more than $\delta \in [0,1]$
- Guarantees that the resulting Petri net can replay at least $\tau \cdot |L|$ traces
Limiting the Search Space

maximum depth $d$ #connecting activities

$d = 4$

- Improves runtime significantly
- Tradeoff between simplicity, fitness, precision and generalization
Why bother?

- Find complex control-flow structures (e.g., long-term dependencies)

\[ L = [\langle \uparrow, a, c, d, \Box \rangle^{13}, \{ \langle \uparrow, b, c, e, \Box \rangle^{42} \}] \]

- Filter noise and infrequent behavior

\[ L = [\langle \uparrow, x_1, x_2, x_3, y_3, y_2, y_4, \Box \rangle^{13}, \{ \langle \uparrow, x_3, x_2, x_1, y_1, y_2, y_3, \Box \rangle^{13}, \langle \uparrow, x_1, x_2, x_3, y_3, y_2, y_4, \Box \rangle^{13}, \langle \uparrow, x_3, x_2, x_1, y_1, y_2, y_3, \Box \rangle^{13}, \langle \uparrow, x_1, x_2, x_3, y_3, y_2, y_4, \Box \rangle^{13}, \langle \uparrow, x_3, x_2, x_1, y_1, y_2, y_3, \Box \rangle^{13} \}] \]
Motivation
Poor Simplicity, many implicit places

places whose removal does not change the language of the net

\[ L = [\langle ▶, a, c, d, □ \rangle^{13}, \langle ▶, b, c, e, □ \rangle^{42}] \]

eST-Miner (τ = 1)
How the eST-Miner currently deals with implicit places:

(user-definable) noise threshold parameter $\tau$

- $\tau = 1.0$ (region-based)
- $\tau \leq 1.0$ (LP-based)
Current Applicability of the eST-Miner

max. depth $d = 7$

<table>
<thead>
<tr>
<th>event log</th>
<th>#activities</th>
<th>#traces</th>
<th>#trace variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sepsis</td>
<td>18</td>
<td>1050</td>
<td>846</td>
</tr>
</tbody>
</table>

### fastest current IPR-technique (LP)

<table>
<thead>
<tr>
<th>Noise threshold $\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
</tr>
<tr>
<td>0.8</td>
</tr>
<tr>
<td>0.7</td>
</tr>
<tr>
<td>0.6</td>
</tr>
<tr>
<td>0.5</td>
</tr>
</tbody>
</table>

- 7 min
- 12 min
- 15 min
- 34 min
- 1.75 h
- 2.5 h

**Total runtime in s**

0 2000 4000 6000 8000 10000 12000
Current Applicability of the eST-Miner

max. depth $d = 7$

<table>
<thead>
<tr>
<th>event log</th>
<th>#activities</th>
<th>#traces</th>
<th>#trace variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sepsis</td>
<td>18</td>
<td>1050</td>
<td>846</td>
</tr>
</tbody>
</table>

Improvements of up to 82%
Precision-Guided Implicit Place Avoidance
Algorithmic Framework

$p$, fitting with respect to $L$ and $\tau$
Algorithmic Framework

$P$, fitting with respect to $L$ and $\tau$

evaluate $\text{precision}(P \cup \{p\})$
Algorithmic Framework

$p$, fitting with respect to $L$ and $\tau$

evaluate $\text{precision}(P \cup \{p\})$

no increase in precision

discard($p$)

intermediate result

$P$
Algorithmic Framework

\[ P_{\text{potImpl}}((I|O)) = \{(I'|O') \in P \mid O \cap O' \neq \emptyset\} \]
Algorithmic Framework

\[ P_{potImpl}(I|O) = \{(I'|O') \in P \mid O \cap O' \neq \emptyset\} \]
Algorithmic Framework

\[ P_{pot\text{Impl}}((I|O)) = \{(I'|O') \in P \mid O \cap O' \neq \emptyset\} \]
Algorithmic Framework

\[ P_{potImpl}(I|O)) = \{(I'|O') \in P \mid O \cap O' \neq \emptyset\} \]
Algorithmic Framework

\[ P_{potImpl}(I|O) = \{(I'|O') \in P \mid O \cap O' \neq \emptyset\} \]
Example

Event log

\[ L = [⟨úa, a, c, d, □⟩^{13}, ⟨úa, b, c, e, □⟩^{42}] \]

Noise threshold \( τ = 1 \)

Precision abort threshold \( ρ = 1 \)
Example

$L = [⟨\triangleright, a, c, d, □⟩^{13}, ⟨\triangleright, b, c, e, □⟩^{42}]$

1. Build Prefix Automaton and Activity Mapping

$A(·):$ number of times $·$ is allowed (when replaying $L$)

$E(·):$ number of times allowing $·$ is escaping (when replaying $L$)
Example

\[ L = [\langle \uparrow, a, c, d, \blacksquare \rangle^{13}, \langle \uparrow, b, c, e, \blacksquare \rangle^{42}] \]

2. Calculate initial precision

\[ \text{precision} = 1 - \frac{1375}{1650 + 55} \approx 0.194 \]

\[ A(\uparrow) \]

Escaping Activities

Allowed Activities

<table>
<thead>
<tr>
<th></th>
<th>(A(\cdot))</th>
<th>(E(\cdot))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>275</td>
<td>220</td>
</tr>
<tr>
<td>(b)</td>
<td>275</td>
<td>220</td>
</tr>
<tr>
<td>(c)</td>
<td>275</td>
<td>220</td>
</tr>
<tr>
<td>(d)</td>
<td>275</td>
<td>262</td>
</tr>
<tr>
<td>(e)</td>
<td>275</td>
<td>233</td>
</tr>
<tr>
<td>(\blacksquare)</td>
<td>275</td>
<td>220</td>
</tr>
</tbody>
</table>
Example

\[ L = \{\langle \triangleright, a, c, d, \blacksquare \rangle^{13}, \langle \triangleright, b, c, e, \blacksquare \rangle^{42}\} \]

Candidate place: \( p_1 = (\triangleright|c) \)

\[
\begin{align*}
L &= \{\langle \triangleright, a, c, d, \blacksquare \rangle^{13}, \langle \triangleright, b, c, e, \blacksquare \rangle^{42}\} \\
\mathcal{A}(\cdot) &\quad \mathcal{E}(\cdot) \\
a &\quad 275 \quad 220 \\
b &\quad 275 \quad 220 \\
c &\quad 275 \quad 220 \\
d &\quad 275 \quad 262 \\
e &\quad 275 \quad 233 \\
\blacksquare &\quad 275 \quad 220
\end{align*}
\]
Example

\[ L = [\langle ▶, a, c, d, □ \rangle^{13}, \langle ▶, b, c, e, □ \rangle^{42}] \]

Candidate place: \( p_1 = (▶|c) \)

\[
\text{precision}(P) = 1 - \frac{1210}{1485 + 55} \\
\approx 0.213 (> 0.194)
\]

\[ \Rightarrow \text{add}(p_1) \]

\[ P_{\text{PotImpl}}(p_1) = \emptyset \]
Example

\[ L = [\langle \blacktriangleright, a, c, d, \blacksquare \rangle^{13}, \langle \blacktriangleright, b, c, e, \blacksquare \rangle^{42}] \]

Candidate place: \( p_2 = (a|d) \)

<table>
<thead>
<tr>
<th>( . )</th>
<th>( A(\cdot) )</th>
<th>( E(\cdot) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a )</td>
<td>275</td>
<td>220</td>
</tr>
<tr>
<td>( b )</td>
<td>275</td>
<td>220</td>
</tr>
<tr>
<td>( c )</td>
<td>110</td>
<td>55</td>
</tr>
<tr>
<td>( d )</td>
<td>275</td>
<td>262</td>
</tr>
<tr>
<td>( e )</td>
<td>275</td>
<td>233</td>
</tr>
<tr>
<td>( \blacksquare )</td>
<td>275</td>
<td>220</td>
</tr>
</tbody>
</table>
Example

\[ L = [\langle \triangleright, a, c, d, □ \rangle^{13}, \langle \triangleright, b, c, e, □ \rangle^{42}] \]

Candidate place: \( p_2 = (a|d) \)

\[ \text{precision}(P) = 1 - \frac{961}{1236 + 55} \approx 0.256 (> 0.213) \]

\[ \Rightarrow \text{add}(p_2) \]

\[ P_{\text{PotImpl}}(p_2) = \emptyset \]
Example

\[ L = [\langle \uparrow, a, c, d, \blacksquare \rangle^{13}, \langle \uparrow, b, c, e, \blacksquare \rangle^{42}] \]

Candidate place: \( p_3 = (b|e) \)

<table>
<thead>
<tr>
<th>( \cdot )</th>
<th>( A(\cdot) )</th>
<th>( E(\cdot) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>275</td>
<td>220</td>
</tr>
<tr>
<td>b</td>
<td>275</td>
<td>220</td>
</tr>
<tr>
<td>c</td>
<td>110</td>
<td>55</td>
</tr>
<tr>
<td>d</td>
<td>26</td>
<td>13</td>
</tr>
<tr>
<td>e</td>
<td>275</td>
<td>233</td>
</tr>
<tr>
<td>( \blacksquare )</td>
<td>275</td>
<td>220</td>
</tr>
</tbody>
</table>
Example

\[
L = [(\triangleright, a, c, d, \blacksquare)^{13}, (\triangleright, b, c, e, \blacksquare)^{42}]
\]

Candidate place: \( p_3 = (b|e) \)

\[
\text{precision}(P) = 1 - \frac{770}{1045 + 55} = 0.3 (> 0.256)
\]

\( \Rightarrow \text{add}(p_3) \)

\( P_{PotImpl}(p_3) = \emptyset \)
Example

\[ L = [⟨▶, a, c, d, ■⟩^{13}, ⟨▶, b, c, e, □⟩^{42}] \]

Candidate place: \( p_4 = (c \mid □) \)
Example

\[ L = [\langle ▶, a, c, d, ■ \rangle^{13}, \langle ▶, b, c, e, ■ \rangle^{42}] \]

Candidate place: \( p_4 = (c|■) \)

\[ p_4 \text{ precision}(P) = 1 - \frac{605}{880+55} \approx 0.353 (> 0.3) \]

\( ⇒ add(p_4) \)

\[ P_{PotImpl}(p_4) = \emptyset \]
Example

\[ L = [\{\uparrow, a, c, d, \blacklozenge\}^{13}, \{\uparrow, b, c, e, \blacklozenge\}^{42}] \]

Candidate place: \( p_5 = (\uparrow|\blacklozenge) \)

\[
\begin{align*}
A & : \begin{array}{c} a \ 275 \\ b \ 275 \\ c \ 110 \\ d \ 26 \\ e \ 84 \\ \blacklozenge \ 110 \end{array} \\
E & : \begin{array}{c} a \ 220 \\ b \ 220 \\ c \ 55 \\ d \ 13 \\ e \ 42 \\ \blacklozenge \ 55 \end{array}
\end{align*}
\]

\[
\text{precision}(P) = 0.353
\]

\[ \Rightarrow \text{discard}(p_5) \]
Example

\[ L = [⟨\triangleright, a, c, d, ■⟩^{13}, ⟨\triangleright, b, c, e, ■⟩^{42}] \]

Candidate place: \( p_5 = (\triangleright | ■) \)

\[
\begin{array}{c|c|c|c|c|c|}
\cdot & A(\cdot) & E(\cdot) \\
\hline
a & 275 & 220 \\
b & 275 & 220 \\
c & 110 & 55 \\
d & 26 & 13 \\
e & 84 & 42 \\
\text{■} & 110 & 55 \\
\end{array}
\]
Example

\[ L = [\langle \triangleright, a, c, d, \blacksquare \rangle^{13}, \langle \triangleright, b, c, e, \blacksquare \rangle^{42}] \]

Candidate place: \( p_6 = (\triangleright|a, b) \)

\[
\begin{array}{cccc}
\cdot & A(\cdot) & E(\cdot) \\
\hline
a & 275 & 220 \\
b & 275 & 220 \\
c & 110 & 55 \\
d & 26 & 13 \\
e & 84 & 42 \\
\blacksquare & 110 & 55 \\
\end{array}
\]
**Example**

\[ L = [\langle \uparrow, a, c, d, \blacksquare \rangle^{13}, \langle \uparrow, b, c, e, \blacksquare \rangle^{42}] \]

Candidate place: \( p_6 = (\uparrow | a, b) \)

\[ \begin{array}{c|c|c}
\cdot & A(\cdot) & E(\cdot) \\
\hline
a & 275 & 220 \\
b & 275 & 220 \\
c & 110 & 55 \\
d & 26 & 13 \\
e & 84 & 42 \\
\blacksquare & 110 & 55 \\
\end{array} \]

\[
\text{precision}(P) = 1 - \frac{165}{440 + 55} \\
\approx 0.667 (> 0.353)
\]

\[
\Rightarrow \text{add}(p_6)
\]

\[ P_{PotImpl}(p_6) = \emptyset \]
Example

\[ L = [\langle a, c, d, \blacklozenge \rangle^{13}, \langle b, c, e, \blacklozenge \rangle^{42}] \]

Candidate place: \( p_7 = (a, b | c) \)

<table>
<thead>
<tr>
<th></th>
<th>( A(\cdot) )</th>
<th>( E(\cdot) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>55</td>
<td>0</td>
</tr>
<tr>
<td>b</td>
<td>55</td>
<td>0</td>
</tr>
<tr>
<td>c</td>
<td>110</td>
<td>55</td>
</tr>
<tr>
<td>d</td>
<td>26</td>
<td>13</td>
</tr>
<tr>
<td>e</td>
<td>84</td>
<td>42</td>
</tr>
<tr>
<td>( \blacklozenge )</td>
<td>110</td>
<td>55</td>
</tr>
</tbody>
</table>
Enhancing the Applicability of the eST-Miner: Efficient Precision-Guided Implicit Place Avoidance

Example

\[ L = [\langle \text{�} , a, c, d, \text {■} \rangle^{13}, \langle \text{�} , b, c, e, \text{■} \rangle^{42}] \]

Candidate place: \( p_7 = (a, b | c) \)

\[ \text{precision}(P) = 1 - \frac{110}{385+55} = 0.75 (> 0.667) \]

\[ \Rightarrow \text{add}(p_7) \]

\[ P_{\text{PotModal}}(p_7) = \{ (\text{�} | c) \} \]
Example

\[ L = [\langle \triangleright, a, c, d, \blacksquare \rangle^{13}, \langle \triangleright, b, c, e, \blacksquare \rangle^{42}] \]

Candidate place: \( p_7 = (a, b | c) \)

Pot. implicit place: \( p_1 = (\triangleright | c) \)

\[ \text{precision}(P) = 0.75 \]

\[ \Rightarrow \text{revoke}(p_1) \]
Example

\[ L = [\langle \blacktriangleright, a, c, d, \blacksquare \rangle^{13}, \langle \blacktriangleright, b, c, e, \blacksquare \rangle^{42}] \]

Candidate place: \( p_8 = (c|d, e) \)

\[ L = [\langle \blacktriangleright, a, c, d, \blacksquare \rangle^{13}, \langle \blacktriangleright, b, c, e, \blacksquare \rangle^{42}] \]

<table>
<thead>
<tr>
<th>( \cdot )</th>
<th>( A(\cdot) )</th>
<th>( E(\cdot) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a )</td>
<td>55</td>
<td>0</td>
</tr>
<tr>
<td>( b )</td>
<td>55</td>
<td>0</td>
</tr>
<tr>
<td>( c )</td>
<td>55</td>
<td>0</td>
</tr>
<tr>
<td>( d )</td>
<td>26</td>
<td>13</td>
</tr>
<tr>
<td>( e )</td>
<td>84</td>
<td>42</td>
</tr>
<tr>
<td>( \blacksquare )</td>
<td>110</td>
<td>55</td>
</tr>
</tbody>
</table>
Example

\[ L = [\langle \triangleright, a, c, d, \blacksquare \rangle^{13}, \langle \triangleright, b, c, e, \blacksquare \rangle^{42}] \]

Candidate place: \( p_8 = (c|d, e) \)

\[
\begin{align*}
L &= \{(\triangleright, a, c, d, \blacksquare)\}^{13}, \{(\triangleright, b, c, e, \blacksquare)\}^{42} \\
p_8 &= (c|d,e) \\
\text{precision}(P) &= 1 - \frac{55}{330+55} \\
&\approx 0.857 (> 0.75) \\
\Rightarrow &\text{add}(p_8) \\
P_{PotImp}(p_8) &= \{(a|d), (b|e)\}
\end{align*}
\]
Example

\[ L = [(\mathbb{I}, a, c, d, \bullet)^{13}, (\mathbb{I}, b, c, e, \bullet)^{42}] \]

Candidate place: \( p_8 = (c|d, e) \)

Pot. implicit place: \( p_2 = (a|d) \)

<table>
<thead>
<tr>
<th></th>
<th>( A(\cdot) )</th>
<th>( E(\cdot) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a )</td>
<td>55</td>
<td>0</td>
</tr>
<tr>
<td>( b )</td>
<td>55</td>
<td>0</td>
</tr>
<tr>
<td>( c )</td>
<td>55</td>
<td>0</td>
</tr>
<tr>
<td>( d )</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>( e )</td>
<td>42</td>
<td>0</td>
</tr>
<tr>
<td>( \bullet )</td>
<td>110</td>
<td>55</td>
</tr>
</tbody>
</table>
Example

\[ L = [\langle \uparrow, a, c, d, \square \rangle^{13}, \langle \uparrow, b, c, e, \square \rangle^{42}] \]

**Candidate place:** \( p_8 = (c | d, e) \)

**Pot. implicit place:** \( p_2 = (a | d) \)

\[
\text{precision}(P) = 1 - \frac{97}{372+55} \\
\approx 0.787 (< 0.857)
\]

\( \Rightarrow \text{keep}(p_2) \)
Example

\[ L = [\langle \triangleright, a, c, d, \blacksquare \rangle^{13}, \langle \triangleright, b, c, e, \blacksquare \rangle^{42}] \]

Candidate place: \( p_8 = (c|d,e) \)
Example

\[ L = [\langle \uparrow, a, c, d, \blacksquare \rangle^{13}, \langle \uparrow, b, c, e, \blacksquare \rangle^{42}] \]

Candidate place: \( p_8 = (c|d, e) \)

Pot. implicit place: \( p_3 = (b|e) \)

<table>
<thead>
<tr>
<th>( \cdot )</th>
<th>( A(\cdot) )</th>
<th>( E(\cdot) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a )</td>
<td>55</td>
<td>0</td>
</tr>
<tr>
<td>( b )</td>
<td>55</td>
<td>0</td>
</tr>
<tr>
<td>( c )</td>
<td>55</td>
<td>0</td>
</tr>
<tr>
<td>( d )</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>( e )</td>
<td>42</td>
<td>0</td>
</tr>
<tr>
<td>( \blacksquare )</td>
<td>110</td>
<td>55</td>
</tr>
</tbody>
</table>
Example

\[ L = [\langle \triangleright, a, c, d, \blacklozenge \rangle^{13}, \langle \triangleright, b, c, e, \blacklozenge \rangle^{42}] \]

Candidate place: \( p_8 = (c | d, e) \)

Pot. implicit place:

\[ p_3 = (b | e) \]

\[
\text{precision}(P) = 1 - \frac{68}{343+55} \\
\approx 0.829 (< 0.857)
\]

\( \Rightarrow \text{keep}(p_3) \)
Example

\[ L = [\langle ▶, a, c, d, □ \rangle^{13}, \langle ▶, b, c, e, □ \rangle^{42}] \]

Candidate place: \( p_8 = (c | d, e) \)

\[
\begin{array}{cccccccccccccc}
a & 13 & 13 & d & 13 & □ \\
b & 42 & 42 & e & 42 & □ \\
\end{array}
\]

\[
\begin{array}{cccccc}
\text{start} & 55 & ▶ & 55 \\
\text{a} & 55 & ▶ & \text{b} & 42 & ▶ & \text{c} & 42 & ▶ & \text{d} & 13 & ▶ & \text{e} & 42 & ▶ & □ & 110 & 55
\end{array}
\]

\[
\begin{array}{|c|c|c|}
\hline
\cdot & A(\cdot) & E(\cdot) \\
\hline
a & 55 & 0 \\
b & 55 & 0 \\
c & 55 & 0 \\
d & 13 & 0 \\
e & 42 & 0 \\
□ & 110 & 55 \\
\hline
\end{array}
\]
Example

\[ L = [\langle \uparrow, a, c, d, \Box \rangle^{13}, \langle \uparrow, b, c, e, \Box \rangle^{42}] \]

Candidate place: \( p_9 = (d, e|\Box) \)
**Example**

\[L = [\langle \text{▶}, a, c, d, \text{□} \rangle^{13}, \langle \text{▶}, b, c, e, \text{□} \rangle^{42}]\]

Candidate place: \(p_9 = (d, e|\text{□})\)

\[\begin{align*}
A(\cdot) & \quad E(\cdot) \\
\text{a} & \quad 55 \quad 0 \\
\text{b} & \quad 55 \quad 0 \\
\text{c} & \quad 55 \quad 0 \\
\text{d} & \quad 13 \quad 0 \\
\text{e} & \quad 42 \quad 0 \\
\text{□} & \quad 440 \quad 55 \quad 55 \quad 0
\end{align*}\]

\[\text{precision}(P) = 1 - \frac{0}{275+55} \approx 1 (> 0.857)\]

\[\Rightarrow \text{add}(p_9)\]

\[P_{\text{PotImpl}}(p_9) = \{(c|\text{□})\}\]
Example

\[ L = [(\uparrow, a, c, d, \square)^{13}, (\uparrow, b, c, e, \square)^{42}] \]

Candidate place: \( p_9 = (d, e|\square) \)

Pot. implicit place: \( p_4 = (c|\square) \)

\[ \text{precision}(P) = 1 \]

\[ \Rightarrow \text{revoke}(p_4) \]

\[ \text{precision} \geq \rho \Rightarrow \text{prematurely abort!} \]
How do we do this efficiently?
Approximating ETC-Precision efficiently

*Observation:*

If \( p = (I|O) \) is added to / removed from the intermediate result, we only have to reevaluate \( E(o) \) and \( A(o) \), for all \( o \in O \).
Approximating ETC-Precision efficiently

“When is an activity allowed / reflected / escaping during replay?”

Calculate the marking histories of its input places

\[ H_L(p) \]

Observation:

\[ H_L(p) \] is place-dependent

⇒ only replay \( L \) once per (new) place

⇒ store \( H_L(p) \) for all \( p \in P \)
Experimental Results
Implementation

Composition

Variant: Standard

- apply concurrent implicit place removal
- ETC-based Composer

ρ = 1.0

Image: promtools.org (accessed: 27.09.23)
Experimental Setup

- PADS HPC cluster (AMD Threadripper CPU: 12x3,5 GHz, 128 GB RAM)
- Six (real-life and artificial event logs)

<table>
<thead>
<tr>
<th>type</th>
<th>name</th>
<th>#activities</th>
<th>#traces</th>
<th>#variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>real</td>
<td>HD2017</td>
<td>16</td>
<td>4580</td>
<td>226</td>
</tr>
<tr>
<td></td>
<td>RTFM</td>
<td>13</td>
<td>150370</td>
<td>231</td>
</tr>
<tr>
<td></td>
<td>Sepsis</td>
<td>18</td>
<td>1050</td>
<td>846</td>
</tr>
<tr>
<td>artificial</td>
<td>Repair</td>
<td>10</td>
<td>1104</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>Reviewing</td>
<td>16</td>
<td>100</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td>Teleclaims</td>
<td>13</td>
<td>3512</td>
<td>12</td>
</tr>
</tbody>
</table>
Suitability for Implicit Place Removal \( (d = 4) \)

Remaining implicit places (missed by the ETC-based Composer) in relation to all fitting places:

No remaining implicit places for \( \tau = 1 \)

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>HD2017</th>
<th>RTFM</th>
<th>Sepsis</th>
<th>Repair</th>
<th>Reviewing</th>
<th>Teleclaims</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0/20</td>
<td>0/32</td>
<td>0/29</td>
<td>0/21</td>
<td>0/88</td>
<td>0/82</td>
</tr>
<tr>
<td>0.9</td>
<td>14/4390</td>
<td>3/567</td>
<td>15/543</td>
<td>5/59</td>
<td>0/88</td>
<td>3/103</td>
</tr>
<tr>
<td>0.8</td>
<td>32/5282</td>
<td>11/846</td>
<td>15/978</td>
<td>4/63</td>
<td>0/88</td>
<td>3/238</td>
</tr>
<tr>
<td>0.7</td>
<td>40/8125</td>
<td>20/989</td>
<td>18/1521</td>
<td>5/104</td>
<td>0/88</td>
<td>3/372</td>
</tr>
<tr>
<td>0.6</td>
<td>43/8642</td>
<td>10/1452</td>
<td>33/2987</td>
<td>7/127</td>
<td>0/88</td>
<td>3/414</td>
</tr>
<tr>
<td>0.5</td>
<td>47/8642</td>
<td>5/2535</td>
<td>42/3614</td>
<td>8/252</td>
<td>2/681</td>
<td>2/1035</td>
</tr>
<tr>
<td>AVG</td>
<td>29/5850</td>
<td>8/1070</td>
<td>21/1611</td>
<td>5/104</td>
<td>0/187</td>
<td>2/374</td>
</tr>
</tbody>
</table>

Only few implicit places left for \( \tau < 1 \), solve ILP to remove remaining in PP

(fast, since only called once and few places as input)
Suitability for Implicit Place Removal \( (d = 4) \)

\[
\begin{align*}
\text{Difference in Alignment-based Fitness} &
\text{to Standard eST-Miner} \\
\text{Difference in ETC-Precision} &
\text{to Standard eST-Miner}
\end{align*}
\]

⇒ (almost) no difference in quality
Problem: Overly Complex Models

Difference in Simplicity to Standard eST-Miner

\[ \tau \]

\[
\begin{align*}
\text{absolute difference in \#arcs} \\
1 & 0.9 & 0.8 & 0.7 & 0.6 & 0.5 \\
10 & 0 & 0 & 0 & 0 & 0 \\
20 & 10 & 10 & 10 & 10 & 10 \\
30 & 20 & 20 & 20 & 20 & 20 \\
40 & 30 & 30 & 30 & 30 & 30 \\
50 & 40 & 40 & 40 & 40 & 40 \\
\end{align*}
\]

- HD2017
- RTFM
- Sepsis
- Repair
- Reviewing
- Teleclaims

\((d = 4)\)
Reason: Proposal order matters

\[ L = [\langle \triangleright, a, c, d, \blacksquare \rangle^{13}, \langle \triangleright, b, c, e, \blacksquare \rangle^{42}] \]  

(previous example, different proposal order)

Final Model:

![Diagram showing a directed graph with nodes labeled a, b, c, d, e, and an initial event.]
Performance Comparison of IPR-Techniques

(user-definable) noise threshold parameter $\tau$

$\tau = 1.0$

• ETC-based composer \( (ETC) \)
• region-based IPR \( (Region) \)
• LP-based IPR \( (LP) \)
Performance Comparison of IPR-Techniques (Real-Life Event Logs, $\tau = 1, d = 4$)

For $\tau = 1$, all techniques have comparable runtime.
Performance Comparison of IPR-Techniques (Artificial Event Logs, $\tau = 1$, $d = 4$)

For $\tau = 1$, all techniques have comparable runtime
Performance Comparison of IPR-Techniques

(user-definable) noise threshold parameter $\tau$

$\tau < 1.0$

- ETC-based composer ($ETC$)
- region-based IPR ($Region$)
- LP-based IPR ($LP$)
Performance Comparison of IPR-Techniques (Real-Life Event Logs, $\tau < 1, d = 4$)

average improvements of 10 seconds (30%)
Performance Comparison of IPR-Techniques (Real-Life Event Logs, \( \tau < 1, d = 7 \))

initial example

average improvements of 21 minutes (50%)
Performance Comparison of IPR-Techniques (Artificial Event Logs, $\tau < 1, d = 4$)

Repair

Reviewing

Teleclaims

premature abort
Performance Comparison of IPR-Techniques (Artificial Event Logs, $\tau < 1, d = 7$)

Improvements of up to 99% (12 seconds)
Limitations

For bigger event logs, we are still unable to discover process models efficiently ($d = 4$):

- $\tau = 1$: 30 min
- $\tau < 1$: >2 h

<table>
<thead>
<tr>
<th>name</th>
<th>#activities</th>
<th>#traces</th>
<th>#variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPI-Challenge 2019</td>
<td>44</td>
<td>251.734</td>
<td>11.973</td>
</tr>
</tbody>
</table>
eST-Miner in Process Discovery
eST-Miner in Process Discovery

Our improvements in performance enable us to compare the eST-Miner to other process discovery techniques. Grid search for the "best model" (considering F1-score) across all parameters of:

- eST-Miner ($\delta$-Variant)
- Alpha Miner
- Inductive Miner
- Heuristic Miner
- Split Miner
- Region-based Algorithms too slow (not practically usable)
eST-Miner in Process Discovery

<table>
<thead>
<tr>
<th>log</th>
<th>eST</th>
<th>Alpha</th>
<th>Heuristic</th>
<th>Inductive</th>
<th>Split</th>
</tr>
</thead>
<tbody>
<tr>
<td>HD2017</td>
<td>0.983</td>
<td>0.354</td>
<td>0.952</td>
<td>0.916</td>
<td>0.962</td>
</tr>
<tr>
<td>RTFM</td>
<td>0.966</td>
<td>-</td>
<td>0.974</td>
<td>0.895</td>
<td><strong>0.979</strong></td>
</tr>
<tr>
<td>Sepsis</td>
<td>0.731</td>
<td>-</td>
<td><strong>0.829</strong></td>
<td>0.724</td>
<td>-</td>
</tr>
<tr>
<td>Repair</td>
<td>0.826</td>
<td>0.310</td>
<td>0.757</td>
<td>0.820</td>
<td><strong>0.848</strong></td>
</tr>
<tr>
<td>Reviewing</td>
<td>0.809</td>
<td>0.877</td>
<td><strong>0.999</strong></td>
<td>0.877</td>
<td>0.959</td>
</tr>
<tr>
<td>Teleclaims</td>
<td>0.962</td>
<td>-</td>
<td>0.976</td>
<td>0.959</td>
<td><strong>1</strong></td>
</tr>
</tbody>
</table>

| % of models that have no possible firing sequence | 0% | 50% | 44.44% | 0% | 39.1% |

not all discovered models satisfy the requirements to compute alignment-based fitness (within 2 hours)
eST-Miner in Process Discovery

- Discovers models in competitive runtime for mid-sized, real-live event logs ($d = 4$)
  - Real-life event logs: < 13s
  - Artificial event logs: < 2s
- Discovers high-quality models (balances well between fitness and precision)
- Can handle noise and infrequent behavior
- Discovers complex control-flow structures
- Provides guarantees:
  - $\delta$-Variant: models can replay at least $\tau \cdot |L|$ traces
Limitations & Future Work

Limitations:

- Poor performance on logs with many different activities
- Some implicit places remain undetected ($\tau < 1$)
- Some models are overly complex

Future Work: ETC-based Composer

- Usage for conformance checking (incorporate non-fitting log traces instead of using alignments)
- Usage for other process discovery techniques
- Usage of information on precision for the eST-Miner:
  - Prune the search space
  - Guide the search
Thank You For Your Attention!

Key Takeaways: ETC-based Composer

- Efficiently (re)calculates precision of (expanding) Petri nets while incorporating non-fitting log traces
- Place classification: precision-guided implicit place avoidance of (expanding) Petri nets
- Extends the discovery with information on precision
- Exemplary use case: eST-Miner + ETC-based Composer → competitive choice for process discovery