Computing regular expressions with BSP regular expressions

Thibaut Tachon[1,2]   Frederic Loulergue[1]   Gaetan Hains[2]

Laboratoire d'Informatique Fondamentale d'Orléans[1] & Huawei Technologies Ltd[2]

June 10th, 2016

Préliminaires

Expressions régulières

$$\begin{array}{rcl} r & ::= & r\ r' \\ & | & r + r' \\ & | & r^* \\ & | & a \\ & | & \epsilon \\ & | & \emptyset \end{array}$$

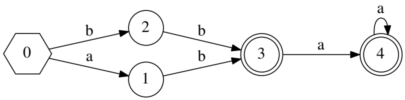
Automates finis

$$A = (Q, \Sigma, \delta, I, F)$$

Q	Ensemble d'états
Σ	Alphabet
δ	Transitions
I	États initiaux
F	États finaux

Exemple

$r = (a + b)ba^*$ $\Longrightarrow$

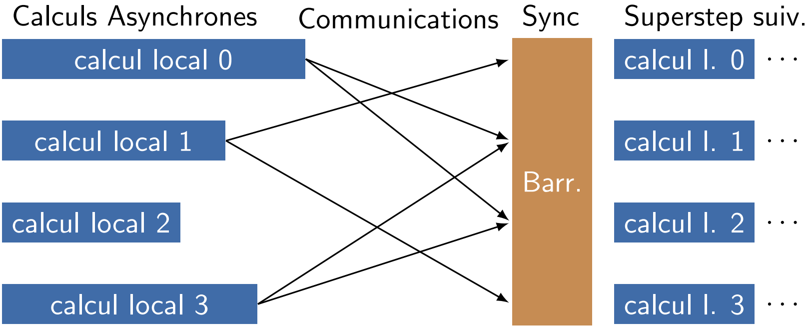


Bulk-Synchronous Parallel

Caractéristiques d'un programme BSP

- Séquence de superstep
- Sans inter blocage ni conflits en écriture
- Simple modèle de coût

Programme avec $p = 4$ processeurs



Langage BSP

Mots et langages "classique"

	Type	Exemple
Symbole :	Σ	a
Mots :	Σ^*	$abccd$
Langage :	$\mathcal{P}\left(\Sigma^*\right)$	$\{a, aab, ab\}$

Mots et langages BSP

	Type	Exemple
Mots vecteur :	$(\Sigma^*)^p$	$\langle ab, a, bbb \rangle$
Mots BSP :	$((\Sigma^*)^p)^*$	$\langle ab, a \rangle \langle bb, aba \rangle$
Langage BSP :	$\mathcal{P}\left(((\Sigma^*)^p)^*\right)$	$\{\langle ab, a \rangle, \langle ab, ba \rangle \langle a, bbb \rangle\}$

Expressions Régulières BSP

Grammaire

$$R ::= R\ R' \mid R + R' \mid R^* \mid \emptyset \mid \epsilon \mid \langle r^0, \ldots, r^{p-1} \rangle$$

Langage

r	$L(r)$
$r\ r'$	$L(r) \cdot L(r')$
$r + r'$	$L(r) \cup L(r')$
r^*	$\bigcup_{i=0}^{\infty} L^i(r)$
a	$\{a\}$
ϵ	$\{\epsilon\}$
$\emptyset$	$\{\}$

R	$L(R)$
$R\ R'$	$L(R) \cdot L(R')$
$R + R'$	$L(R) \cup L(R')$
R^*	$\bigcup_{i=0}^{\infty} L^i(R)$
$\langle r_0, \ldots, r_{p-1} \rangle$	$L(r_0) \times \cdots \times L(r_{p-1})$
ϵ	$\{\epsilon\}$
$\emptyset$	$\{\}$

Automates BSP : Définition

Définition

$$A = (\{Q^i\}_{i\in[p]}, \Sigma, \{\delta^i\}_{i\in[p]}, \{I^i\}_{i\in[p]}, \{F^i\}_{i\in[p]}, \Delta)$$

- Vecteur d'automate de taille p
- Alphabet commun
- Fonction de synchronization

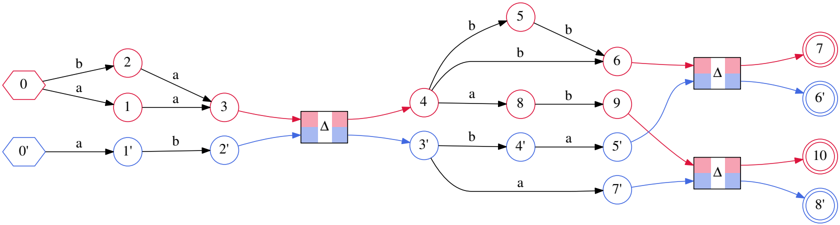
Delta

- $\Delta : \vec{Q} \to \vec{Q}$
- $\vec{Q} = (Q^0 \times ... \times Q^{p-1})$

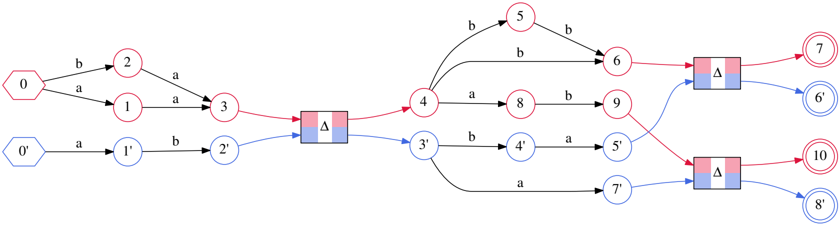
Notation

- $[n] = \{0 \ldots n-1\}$
- $n \in \mathbb{N}$

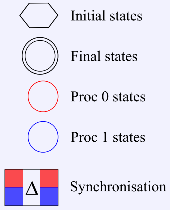
Automates BSP : Exemple ($p = 2$)



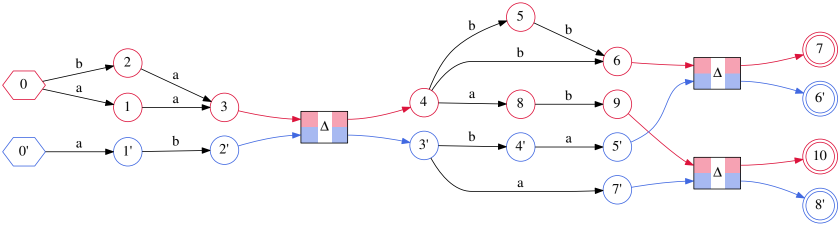
Automates BSP : Exemple ($p = 2$)



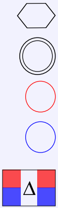
Légende



Automates BSP : Exemple ($p = 2$)



Légende



Initial states

Final states

Proc 0 states

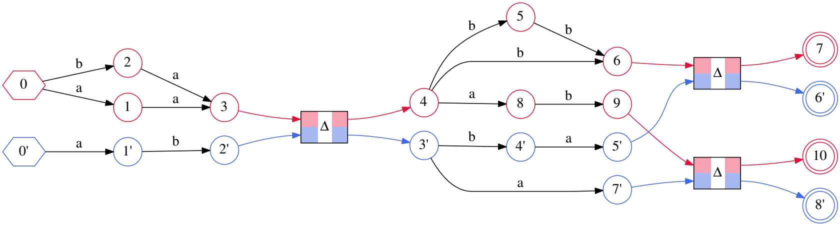
Proc 1 states

Synchronisation

Delta

$$\Delta = \{ \ \langle 3, 2' \rangle \to \langle 4, 3' \rangle \ , \\ \langle 6, 5' \rangle \to \langle 7, 6' \rangle \ , \\ \langle 9, 7' \rangle \to \langle 10, 8' \rangle \ \}$$

Automates BSP : Exemple ($p = 2$)



Légende

- Initial states
- Final states
- Proc 0 states
- Proc 1 states
- Synchronisation

Delta

$$\Delta = \{ \ \langle 3, 2' \rangle \rightarrow \langle 4, 3' \rangle \ , \\ \langle 6, 5' \rangle \rightarrow \langle 7, 6' \rangle \ , \\ \langle 9, 7' \rangle \rightarrow \langle 10, 8' \rangle \ \}$$

Accepte : $\langle aa, ab \rangle \langle ab, a \rangle$?

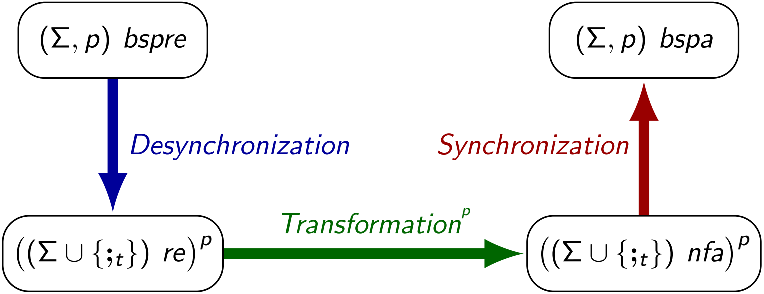
Automates BSP : Reconnaissance

Algorithme de reconnaissance

1. $[\,] \implies \langle q_0^0 \dots q_0^{n-1} \rangle$
2. $\langle w^0, \dots, w^{p-1} \rangle :: suite \implies \forall i \in [p]$, $\delta * (q_0^i, w^i) = q^i$
3. Vecteur d'état n'est pas argument de $\Delta \implies$ rejet
4. Application de Δ : $\langle q^0, \dots q^{p-1} \rangle \rightarrow \langle q'^0, \dots q'^{p-1} \rangle$
5. Pas de mots vecteurs et $\forall i.\ q'^i \in F^i \implies$ acceptation.
6. Pas de mots vecteurs et $\exists i.\ q'^i \notin F^i \implies$ rejet.
7. Encore des mots vecteurs, retour à 2

De BSPRE à BSPA

Schéma global de l'algorithme



Désynchronisation $(\Sigma, p)\ bspre \rightarrow \left((\Sigma \cup \{;_t\})\ re\right)^p$

Pré-traitement

- Annotation des vecteurs selon leurs positions
- Parcours en profondeur de l'arbre syntaxique
- Gauche à droite
- Nombre de vecteurs : S

Désynchronisation (Σ, p) *bspre* $\rightarrow \big((\Sigma \cup \{;_t\})\ re\big)^p$

Pré-traitement

- Annotation des vecteurs selon leurs positions
- Parcours en profondeur de l'arbre syntaxique
- Gauche à droite
- Nombre de vecteurs : S

Exemple

$$\big(\langle a, a + b\rangle + \langle bb, a^*\rangle\big)\langle (a + b)^*, aa\rangle$$

Désynchronisation $(\Sigma, p)\ bspre \to \big((\Sigma \cup \{;_t\})\ re\big)^p$

Pré-traitement

- Annotation des vecteurs selon leurs positions
- Parcours en profondeur de l'arbre syntaxique
- Gauche à droite
- Nombre de vecteurs : S

Exemple

$$\big(\langle a, a + b\rangle + \langle bb, a^*\rangle\big)\langle (a + b)^*, aa\rangle \quad \Longrightarrow$$
$$\big(\langle a, a + b\rangle_0 + \langle bb, a^*\rangle_1\big)\langle (a + b)^*, aa\rangle_2$$

Désynchronisation $\quad (\Sigma, p)\ bspre \rightarrow \left((\Sigma \cup \{;_t\})\ re\right)^p$

Fonction auxiliaire récursive

$$
\begin{aligned}
dsn^i(F \cdot_{BSP} G) &= dsn^i(F)\ .\ dsn^i(G) \\
dsn^i(F +_{BSP} G) &= dsn^i(F) + dsn^i(G) \\
dsn^i(F^{*_{BSP}}) &= dsn^i(F)^* \\
dsn^i(\langle r_0, \ldots, r_{p-1} \rangle_{\boldsymbol{t}}) &= r_i\ .\ ;_{\boldsymbol{t}} \\
dsn^i(\epsilon_{BSP}) &= \epsilon \\
dsn^i(\emptyset_{BSP}) &= \emptyset
\end{aligned}
$$

Fonction principale

$$Dsn(R) = \langle dsn^0(R), \ldots, dsn^{p-1}(R) \rangle$$

Désynchronisation $(\Sigma, p)\ bspre \to \left((\Sigma \cup \{;_t\})\ re\right)^p$

Fonction auxiliaire récursive

$$
\begin{aligned}
dsn^i(F \cdot_{BSP} G) &= dsn^i(F)\ .\ dsn^i(G)\\
dsn^i(F +_{BSP} G) &= dsn^i(F) + dsn^i(G)\\
dsn^i(F^{*_{BSP}}) &= dsn^i(F)^*\\
dsn^i(\langle r_0, \ldots, r_{p-1}\rangle_{\boldsymbol{t}}) &= r_i\ .\ ;_{\boldsymbol{t}}\\
dsn^i(\epsilon_{BSP}) &= \epsilon\\
dsn^i(\emptyset_{BSP}) &= \emptyset
\end{aligned}
$$

Fonction principale

$$Dsn(R) = \langle dsn^0(R), \ldots, dsn^{p-1}(R)\rangle$$

$$\boxed{\Sigma_{dsn} = \Sigma \cup \Big(\bigcup_{t \in [S]} ;_t\Big)}$$

Brüggemann-Klein $\left((\Sigma \cup \{;_t\})\ re\right)^p \rightarrow \left((\Sigma \cup \{;_t\})\ nfa\right)^p$

Principes

- Amélioration de Glushkov
- Pré-calcul : forme normale de fermeture

Caractéristiques

- Complexité en temps optimale : quadratique
- Complexité en espace : linéaire
- Pas d'ϵ-transitions.

$$\forall i \in [p],\ A^i = (Q^i, \Sigma_{dsn}, \delta^i_{dsn}, I^i, F^i)$$

Synchronisation : Fonctions auxiliaires

Vecteurs arguments de Δ

$$src^t(\delta_{dsn}) \;=\; \Big\{ q \mid \big((q, ;_t) \to q'\big) \in \delta_{dsn} \Big\}$$

$$\nu^t_{src}(\delta^p_{dsn}) \;=\; src^t(\delta^0_{dsn}) \times \cdots \times src^t(\delta^{p-1}_{dsn})$$

Vecteurs sorties de Δ

$$dst^t(\delta_{dsn}) \;=\; \Big\{ q' \mid \big((q, ;_t) \to q'\big) \in \delta_{dsn} \Big\}$$

$$\nu^t_{dst}(\delta^p_{dsn}) \;=\; dst^t(\delta^0_{dsn}) \times \cdots \times dst^t(\delta^{p-1}_{dsn})$$

$$Syn(\ (\{Q^i\}_{i\in[p]}, \Sigma_{dsn}, \{\delta^i_{dsn}\}_{i\in[p]}, \{I^i\}_{i\in[p]}, \{F^i\}_{i\in[p]}))$$

$$= (\{Q^i\}_{i\in[p]}, \Sigma, \{\delta^i\}_{i\in[p]}, \{I^i\}_{i\in[p]}, \{F^i\}_{i\in[p]}, \Delta)$$

où **soit** $arg^t = \nu^t_{src}(\{\delta^i_{dsn}\}_{i\in[p]})$ **et** $img^t = \nu^t_{dst}(\{\delta^i_{dsn}\}_{i\in[p]})$

$$\Delta = \bigcup_{t\in[S]} \left(\bigcup_{\vec{q}\in arg^t} \ \bigcup_{\vec{q}'\in img^t} \left(\vec{q} \to \vec{q}'\right)\right)$$

$$\delta^i = \delta^i_{dsn} \setminus \bigcup_{t\in[S]} ((q, \boldsymbol{;_t}) \to q') \in \delta^i_{dsn}$$

$$\Sigma = \Sigma_{dsn} \setminus \bigcup_{t\in[S]} \boldsymbol{;_t}$$

Exemple

BSPRE

$\langle (a+b)a \text{ , } ab \rangle_0 \left(\langle (\epsilon + b)b \text{ , } ba \rangle_1 + \langle ab \text{ , } a \rangle_2 \right)$

Exemple

BSPRE

$\langle (a+b)a\ ,\ ab\rangle_0 \left(\langle (\epsilon+b)b\ ,\ ba\rangle_1 + \langle ab\ ,\ a\rangle_2\right)$

Desynchronisation

$\langle (a+b)\, a\, ;_0 \left((\epsilon+b)\, b\, ;_1) + (a\, b\, ;_2)\right)\ ,\ a\, b\, ;_0 \left((b\, a\, ;_1) + (a\, ;_2)\right)\rangle$

Exemple

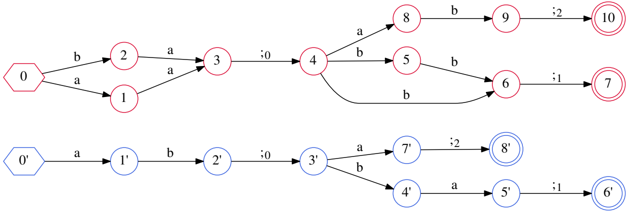
BSPRE

$$\langle (a+b)a\ ,\ ab\rangle_0\ \big(\langle(\epsilon+b)b\ ,\ ba\rangle_1\ +\langle ab\ ,\ a\rangle_2\big)$$

Desynchronisation

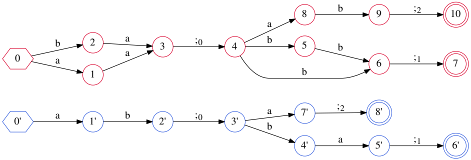
$$\langle (a+b)\,a\,;_0\,\big(((\epsilon+b)\,b\,;_1)+(a\,b\,;_2)\big)\ ,\ a\,b\,;_0\,\big((b\,a\,;_1)+(a\,;_2)\big)\rangle$$

Transformation

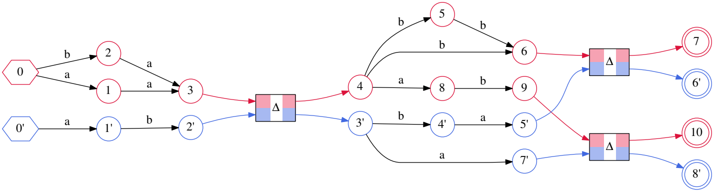


Exemple

Transformation



Synchronisation



de RE à BSPRE

POSIX

```
r = .*<tag>.*Valiant.*</tag>.*
```

Classique

$$r = \sigma^* < \text{tag} > \sigma^* \text{ Valiant } \sigma^* < /\text{tag} > \sigma^*$$

où $\sigma = \forall \sigma_i \in \Sigma, (\sigma_0 + \sigma_1 + \dots)$

Pre-calculs

Découpage de RE

$$\phi(r) = (r_0, r_1, r_2, r_3)$$

$$r = \underbrace{\sigma^* < \mathsf{tag} >}_{r_0} \underbrace{\sigma^* \ \mathsf{Valiant}}_{r_1} \ \underbrace{\sigma^* < /\mathsf{tag} >}_{r_2} \ \underbrace{\sigma^*}_{r_3}$$

Pre-calculs

Découpage de RE

$$\phi(r) = (r_0, r_1, r_2, r_3)$$

$$r = \underbrace{\sigma^* < \text{tag} >}_{r_0} \underbrace{\sigma^* \text{ Valiant }}_{r_1} \underbrace{\sigma^* < /\text{tag} >}_{r_2} \underbrace{\sigma^*}_{r_3}$$

Distribution de l'input

$$\begin{aligned} D(w) &= D(u_0 \dots u_{n-1}) \\ &= (u_0, 0 \bmod p) \dots (u_{n-1}, (n-1) \bmod p) \\ &= \langle u_0 u_p \dots \ , \ \dots\dots \ , \ u_{p-1} u_{2p-1} \dots \rangle \end{aligned}$$

Pre-calculs

Découpage de RE

$$\phi(r) = (r_0, r_1, r_2, r_3)$$

$$r = \underbrace{\sigma^* < \text{tag} >}_{r_0} \underbrace{\sigma^* \text{ Valiant }}_{r_1} \underbrace{\sigma^* < /\text{tag} >}_{r_2} \underbrace{\sigma^*}_{r_3}$$

Distribution de l'input

$$\begin{aligned} D(w) &= D(u_0 \dots u_{n-1}) \\ &= (u_0, 0 \bmod p) \dots (u_{n-1}, (n-1) \bmod p) \\ &= \langle u_0 u_p \dots \ , \ \dots\dots \ , \ u_{p-1} u_{2p-1} \dots \rangle \end{aligned}$$

Forme de BSPRE

$$R = R_0 \ R_1 \ R_2 \ R_3 \qquad \textbf{où } \forall i \in \{0 \dots 3\}, \bigcup_{w_j \in L(r_i)} \{D(w_j)\} = L(R_i)$$

BSPRE avec $p = 3$

Reconnaitre "<tag>"

$$
\begin{aligned}
&D(< tag >)\\
=&(<, 0 \,\%\, p)\ ,\ (t, 1 \,\%\, p)\ ,\ (a, 2 \,\%\, p)\ ,\ (g, 3 \,\%\, p)\ ,\ (>, 4 \,\%\, p)\\
=&(<, 0)\ ,\ (t, 1)\ ,\ (a, 2)\ ,\ (g, 0)\ ,\ (>, 1)\\
=&\langle < g\ ,\ t >\ ,\ a \rangle
\end{aligned}
$$

BSPRE avec $p = 3$

Reconnaitre "<tag>"

$$
\begin{aligned}
&D(< tag >)\\
=&(<, 0 \% p) , (t, 1 \% p) , (a, 2 \% p) , (g, 3 \% p) , (>, 4 \% p)\\
=&(<, 0) \quad , (t, 1) \quad , (a, 2) \quad , (g, 0) \quad , (>, 1)\\
=&\langle < g \; , \; t > \; , a \rangle
\end{aligned}
$$

BSPRE correspondant à r_0 $\qquad$ $r_0 = \sigma^* < tag >$

$$
R_0 = \begin{cases}
\langle \; \sigma^* < \mathrm{g} \; , \; \sigma^* \; \mathrm{t} > \; , \; \sigma^* \; \mathrm{a} \; \rangle \\
\qquad + \\
\langle \; \sigma^* \; \mathrm{a} \; , \; \sigma^* < \mathrm{g} \; , \; \sigma^* \; \mathrm{t} > \; \rangle \\
\qquad + \\
\langle \; \sigma^* \; \mathrm{t} > \; , \; \sigma^* \; \mathrm{a} \; , \; \sigma^* < \mathrm{g} \; \rangle
\end{cases}
$$

BSPRE

BSPRE correspondant à r_1, r_2 et r_3

$$r_1 = \sigma^* Valiant$$

$$R_1 = \begin{cases} \langle\, \sigma^*\text{Vit} \;,\; \sigma^*\text{aa} \;,\; \sigma^*\text{ln} \,\rangle \\ \qquad\qquad + \\ \langle\, \sigma^*\text{ln} \;,\; \sigma^*\text{Vit} \;,\; \sigma^*\text{aa} \,\rangle \\ \qquad\qquad + \\ \langle\, \sigma^*\text{aa} \;,\; \sigma^*\text{ln} \;,\; \sigma^*\text{Vit} \,\rangle \end{cases}$$

BSPRE

BSPRE correspondant à r_1, r_2 et r_3

$$r_1 = \sigma^* \mathit{Valiant} \qquad\qquad r_2 = \sigma^* < / \mathit{tag} >$$

$$R_1 = \begin{cases} \langle\, \sigma^*\mathsf{Vit} \;,\; \sigma^*\mathsf{aa} \;,\; \sigma^*\mathsf{ln} \,\rangle \\ \qquad\qquad + \\ \langle\, \sigma^*\mathsf{ln} \;,\; \sigma^*\mathsf{Vit} \;,\; \sigma^*\mathsf{aa} \,\rangle \\ \qquad\qquad + \\ \langle\, \sigma^*\mathsf{aa} \;,\; \sigma^*\mathsf{ln} \;,\; \sigma^*\mathsf{Vit} \,\rangle \end{cases} \quad R_2 = \begin{cases} \langle\, \sigma^* < \mathsf{a} \;,\; \sigma^*/\mathsf{g} \;,\; \sigma^*\mathsf{t} > \,\rangle \\ \qquad\qquad + \\ \langle\, \sigma^*\mathsf{t} > \;,\; \sigma^* < \mathsf{a} \;,\; \sigma^*/\mathsf{g} \,\rangle \\ \qquad\qquad + \\ \langle\, \sigma^*/\mathsf{g} \;,\; \sigma^*\mathsf{t} > \;,\; \sigma^* < \mathsf{a} \,\rangle \end{cases}$$

BSPRE

BSPRE correspondant à r_1, r_2 et r_3

$$r_1 = \sigma^* Valiant$$

$$R_1 = \begin{cases} \langle\, \sigma^*\mathrm{Vit} \;,\; \sigma^*\mathrm{aa} \;,\; \sigma^*\mathrm{In} \,\rangle \\ + \\ \langle\, \sigma^*\mathrm{In} \;,\; \sigma^*\mathrm{Vit} \;,\; \sigma^*\mathrm{aa} \,\rangle \\ + \\ \langle\, \sigma^*\mathrm{aa} \;,\; \sigma^*\mathrm{In} \;,\; \sigma^*\mathrm{Vit} \,\rangle \end{cases}$$

$$r_2 = \sigma^* < /tag >$$

$$R_2 = \begin{cases} \langle\, \sigma^* < \mathrm{a} \;,\; \sigma^*/\mathrm{g} \;,\; \sigma^*\mathrm{t} > \,\rangle \\ + \\ \langle\, \sigma^*\mathrm{t} > \;,\; \sigma^* < \mathrm{a} \;,\; \sigma^*/\mathrm{g} \,\rangle \\ + \\ \langle\, \sigma^*/\mathrm{g} \;,\; \sigma^*\mathrm{t} > \;,\; \sigma^* < \mathrm{a} \,\rangle \end{cases}$$

$$r_3 = \sigma^*$$

$$R_3 = \langle \sigma^* \;,\; \sigma^* \;,\; \sigma^* \rangle$$

BSPRE

BSPRE correspondant à r_1, r_2 et r_3

$$r_1 = \sigma^* \mathit{Valiant}$$

$$R_1 = \begin{cases} \langle\, \sigma^*\mathrm{Vit} \;,\; \sigma^*\mathrm{aa} \;,\; \sigma^*\mathrm{In} \,\rangle \\ \qquad\qquad + \\ \langle\, \sigma^*\mathrm{In} \;,\; \sigma^*\mathrm{Vit} \;,\; \sigma^*\mathrm{aa} \,\rangle \\ \qquad\qquad + \\ \langle\, \sigma^*\mathrm{aa} \;,\; \sigma^*\mathrm{In} \;,\; \sigma^*\mathrm{Vit} \,\rangle \end{cases}$$

$$r_2 = \sigma^* < / \mathit{tag} >$$

$$R_2 = \begin{cases} \langle\, \sigma^* < \mathrm{a} \;,\; \sigma^*/\mathrm{g} \;,\; \sigma^*\mathrm{t} > \,\rangle \\ \qquad\qquad + \\ \langle\, \sigma^*\mathrm{t} > \;,\; \sigma^* < \mathrm{a} \;,\; \sigma^*/\mathrm{g} \,\rangle \\ \qquad\qquad + \\ \langle\, \sigma^*/\mathrm{g} \;,\; \sigma^*\mathrm{t} > \;,\; \sigma^* < \mathrm{a} \,\rangle \end{cases}$$

$$r_3 = \sigma^*$$

$$R_3 = \langle \sigma^* \,,\, \sigma^* \,,\, \sigma^* \rangle$$

Rappel

$$R = R_0 \;\; R_1 \;\; R_2 \;\; R_3$$

Conclusion

Résumé

- BSPRE → BSPA
- RE → BSPRE

Futur Travaux

- Automatiser dernière étape
- Déterminiser et minimiser BSPA
- Modéliser les communications