

HOCore en Coq : résumé

Aurèle Barrière, sur un article de Petar Maskimovic & Alan Schmitt

22 février 2016

Résumé

Ce document résume l'article [2]. Il présente la formalisation en Coq (l'assistant de preuve) du calcul de processus HOCore. On y présente le calcul et des propriétés, puis on s'intéresse aux enjeux de la formalisation et de la réalisation de preuves en Coq.

Mots-clés : Formalisation ; Calcul ; Processus ; Coq ; HOCore

1 Introduction

L'article résumé ici ([2]) traite de la formalisation en Coq de HOCore, un calcul de processus (similaire au lambda-calcul).

De même que le lambda-calcul donne un formalisme universel pour les fonctions mathématiques, le π -calcul propose un formalisme de l'informatique distribuée ou parallèle : la grammaire contient la composition parallèle de processus. Le π -calcul est donc très intéressant pour étudier des concepts fondamentaux de l'informatique distribuée : sécurité, confidentialité entre processus, comportements de processus en parallèle...

HOCore est une version minimale (c'est à dire qu'elle perdrait en expressivité si on retirait des éléments de sa grammaire) du π -calcul. Ce calcul n'a pas d'intérêt intrinsèque, mais l'étudier et le traduire en Coq va permettre d'en apprendre plus sur les calculs de processus et leurs éventuelles formalisations.

L'enjeu est de contourner les problèmes qui se posent lors de la formalisation d'un tel calcul, puis de construire des preuves sur ce calcul. Certaines de ces preuves ayant déjà été faites à la main avant dans l'article [1], on pourra ainsi comparer les deux méthodes.

On présentera donc une introduction à HOCore, accompagnée d'exemples et de propriétés, comme la décidabilité de l'équivalence de processus. On présentera ensuite les problèmes qui se posent pour formaliser le calcul et ses réductions en Coq. Enfin, on s'intéressera aux conséquences de ce travail.

2 Introduction à HOCore

HOCore est un calcul de processus, semblable au lambda-calcul, utilisé en particulier pour décrire des exécutions distribuées de processus. HOCore est minimal (la syntaxe et les règles de calcul sont très petites) et d'ordre supérieur (un processus peut communiquer un autre processus comme message).

2.1 Syntaxe

Un processus, en HOCORE, suit la grammaire suivante :

$$P ::= a(x).P \mid \bar{a}\langle P \rangle \mid P \parallel P \mid x \mid 0$$

On va distinguer 3 catégories syntaxiques : des canaux sur lesquels émettre et recevoir des messages (dans la grammaire ci-dessus, a est un canal), des variables (remplacées lors de la lecture d'un message, x dans la grammaire) et des processus (P dans la grammaire).

$a(x).P$ correspond à la lecture d'un message sur le canal a pour l'exécution de P , tandis que $\bar{a}\langle P \rangle$ correspond à l'émission du processus P sur le canal a .

Pour deux processus P et Q , le processus $P \parallel Q$ correspond à l'exécution en parallèle de P et Q .

2.2 Sémantique et simplifications

0 est un processus qui ne fait rien.

En HOCORE, l'exécution en parallèle est associative et commutative, et le processus 0 en est l'élément neutre. Elle permet la communication entre processus sur les canaux.

En effet, lorsqu'un processus émet un autre processus ou une variable sur un canal a et qu'en parallèle un autre processus lit sur le même canal a , alors on peut faire la réduction suivante :

$\bar{a}\langle P \rangle \parallel a(x).Q \rightarrow [P/x]Q$, qui signifie que les occurrences de x dans Q sont remplacées par P . On utilisera le symbole \rightarrow pour dire qu'un processus se réduit en un autre.

Grâce à l'associativité et la commutativité de la parallélisation, deux processus peuvent communiquer même s'ils ne sont pas à côté : $\bar{a}\langle P \rangle \parallel \bar{b}\langle Q \rangle \parallel a(x).x \equiv \bar{b}\langle Q \rangle \parallel \bar{a}\langle P \rangle \parallel a(x).x$ qui se réduit en $\bar{b}\langle Q \rangle \parallel P$, où \equiv représente l'égalité de 2 processus à commutativité et associativité de \parallel près.

Parmi les variables, il faut donc distinguer celles qui sont dites *libres* et celles dites *liées*. Une variable est liée lorsqu'elle peut être changée par la lecture sur un canal. Une variable est donc liée quand elle est dans la portée d'une réception. Ainsi, dans $a(x).P$, les occurrences de x dans P sont liées. On peut rapprocher ces notions à celles de variables *globales* et *locales*. Dans l'exemple de réduction précédent, x est donc une variable liée puisqu'elle est remplacée à la lecture sur le canal a . Par contre, dans le processus $P \parallel x$, x est une variable libre : elle ne va pas être remplacée.

Enfin, dans le processus $\bar{a}\langle P \rangle \parallel (a(x).Q) \parallel x$, qui se réduit donc en $[P/x]Q \parallel x$, il faut distinguer x dans Q qui est une variable liée puisque remplacée par P , et x exécuté en parallèle, libre.

3 Exemples de processus en HOCORE

3.1 Récursivité

On peut se demander s'il est possible de répliquer un processus. Et en effet, HOCORE permet de décrire des exécutions infinies.

En se donnant un processus P , on cherche donc un processus $!P$ tel que $!P \rightarrow P \parallel !P$. On aura ainsi un processus qui se réduit en un processus P en parallèle de lui-même, qui va donc une fois de plus créer une instance de P et ainsi de suite.

On prend

$$!P = (r(x).(x \parallel \bar{r}\langle x \rangle)) \parallel \bar{r}\langle P \parallel r(x).(x \parallel \bar{r}\langle x \rangle) \rangle$$

Montrons que ce processus convient. On va noter $L = r(x).(x \parallel \bar{r}\langle x \rangle)$ et $R = \bar{r}\langle P \parallel r(x).(x \parallel \bar{r}\langle x \rangle) \rangle$, de telle sorte que $!P = L \parallel R$.

On remarque que R émet un processus sur le canal r et qu'en parallèle L reçoit sur le même canal. Les deux processus communiquent donc. Alors pour L , x sera remplacé par tout le processus émis par R . R , une fois le message émis, se réduira en 0.

On a donc

$$!P \rightarrow P \| r(x).(x \| \bar{r}\langle x \rangle) \parallel \bar{r}\langle P \| r(x).(x \| \bar{r}\langle x \rangle) \rangle \| 0$$

Et on constate qu'on a $!P \rightarrow P \| !P \| 0$, et donc $!P \rightarrow P \| !P$ comme on le souhaitait, puisque 0 est l'élément neutre pour la parallélisation.

Il est donc possible, avec HOCore, de répliquer un processus.

3.2 Choix de processus

On aimerait également disposer d'un processus qui choisisse entre deux processus P et Q selon si il est exécuté en parallèle d'un processus \hat{a}_1 ou \hat{a}_2 .

On souhaite donc avoir le processus $(a_1.P \oplus a_2.Q)$ et les processus \hat{a}_1 et \hat{a}_2 tels que

$$(a_1.P \oplus a_2.Q) \| \hat{a}_1 \rightarrow^* P$$

$$(a_1.P \oplus a_2.Q) \| \hat{a}_2 \rightarrow^* Q$$

En HOCore, cela peut se faire ainsi :

$$(a_1.P \oplus a_2.Q) = \bar{a}_1\langle P \rangle \parallel \bar{a}_2\langle Q \rangle$$

$$\hat{a}_1 = a_1(X).(a_2(Y).(X))$$

$$\hat{a}_2 = a_1(X).(a_2(Y).(Y))$$

Ainsi, si on a en parallèle les processus $(a_1.P \oplus a_2.Q)$ et \hat{a}_1 , le premier va émettre P et Q sur les canaux a_1 et a_2 , qui seront lus par le second, qui lui même ne va exécuter après lecture que le processus émis sur a_1 à savoir P .

3.3 Turing Complétude

On a montré qu'on pouvait avoir des comportements infinis ou des choix. On peut aussi encoder, comme en lambda-calcul, des entiers ou faire un test à 0. On a en particulier que HOCore est un calcul Turing-Complet. Cela est montré dans l'article [1] en encodant les machines de Minsky dans HOCore.

Le problème de la terminaison de processus est indécidable pour HOCore.

4 Équivalence de processus

En HOCore, l'équivalence de processus est décidable. Pour comprendre cela, il convient de définir ce que sont deux processus équivalents.

Pour commencer, un contexte C est un processus auquel on a enlevé un processus. Ce n'est donc pas un processus. Mais la donnée d'un contexte C et d'un processus P , noté $C[P]$ est un processus (où le trou de C a été complété par P). Par exemple, $C = P \parallel$ est un contexte. $C[Q] = P \parallel Q$.

Deux processus P et Q sont dits équivalents si :

- S'il existe une réduction de P à un processus P' , il existe Q' équivalent à P' tel qu'il y a une réduction de Q à Q' .
- P et Q ont les mêmes *observables* : ils émettent des messages sur les mêmes canaux.
- Pour tout contexte C , le contexte complété avec P , $C[P]$, est équivalent à $C[Q]$.

La relation est symétrique. Une des particularités de HOCore (contrairement au π -calcul par exemple), est que les processus ne peuvent pas *cacher* des variables. On peut donc choisir un contexte qui explore le processus et ses actions sur les canaux et les variables.

Remarque : Cela peut paraître surprenant quand on sait qu'en général le problème d'équivalence de fonctions est indécidable. Cependant, il ne s'agit pas d'un paradoxe puisque c'est une équivalence de processus très fine : seuls des processus quasiment identiques sont équivalents. Si on avait choisi l'égalité syntaxique, on aurait également une équivalence décidable. Par exemple, cela ne résout pas le problème de l'arrêt : on sait que si deux processus sont équivalents, soit les deux terminent, soit aucun ne termine mais on ne dispose pas d'un processus indiquant si un autre processus termine.

5 Formalisation en Coq

Un des principaux travaux de l'équipe de recherche a été de formaliser HOCore en Coq (l'assistant de preuve). La syntaxe d'HOCore se traduit simplement en Coq, cependant des problèmes subsistent lorsqu'on s'intéresse à la sémantique : il faut pouvoir reconnaître les variables liées dont le rôle est identique. En effet, deux processus peuvent s'écrire différemment mais être équivalents.

Par exemple, $a(x).a(y).x$ et $a(z).a(y).z$ sont équivalents, mais ne s'écrivent pas rigoureusement de la même manière.

5.1 Alpha-conversion et représentation canonique locale des noms

Dans l'exemple précédent, on parle d'alpha-conversion si deux processus ont des variables liées dont le nom a été changé. Cependant, il faut bien distinguer variables liées et variables libres. On ne peut pas parler d'alpha-conversion pour des variables libres : le contexte peut utiliser ces variables libres et on pourrait alors perdre l'équivalence.

Ainsi, le processus x n'est pas équivalent au processus y . Si on les met dans un certain contexte, par exemple $C = y\|\bar{a}\langle z\rangle\|a(x).()$, on obtient les processus suivant :

$$C[x] = y\|\bar{a}\langle z\rangle\|a(x).(x) \rightarrow y\|0\|z \rightarrow y\|z$$

$$C[y] = y\|\bar{a}\langle z\rangle\|a(x).(y) \rightarrow y\|0\|y \rightarrow y\|y$$

qui ne sont pas équivalents.

Si on veut pouvoir décider, avec Coq par exemple, de l'équivalence de deux processus, cela peut être un problème.

Pour le contourner, on peut entièrement se dispenser des noms de variables : on va disposer d'une fonction qui pour chaque variable va calculer une hauteur de manière identique et indépendante du nom. Cette approche est celle de *représentation canonique locale des noms* introduite dans l'article [3] .

On n'a plus alors d'alpha-conversion possible : les variables ne sont plus que des indices, qui seront égaux dans le cas de processus équivalents.

Pour reprendre l'analogie avec les variables globales et locales d'un programme, on peut dire que deux fonctions dont on a juste changé le nom des variables locales sont équivalentes. Mais si la fonction utilise des variables globales et qu'on modifie leur nom, on perd l'équivalence puisque les fonctions n'auront pas le même comportement dans des programmes qui utilisent ces variables.

5.2 Système de transitions étiquetées

Pour pouvoir analyser un processus, trouver les communications entre les différents processus en parallèles sans modifier complètement sa syntaxe, on introduit une nouvelle forme de sémantique : un système de transitions étiquetées ou LTS (*labeled transition system*).

En effet, notre réduction pour la communication se formalise ainsi :

$$\bar{a}\langle P \rangle \| a(x).Q \rightarrow [P/x]Q$$

Cependant, on sait qu'on a défini la parallélisation comme étant transitive. On peut donc se retrouver dans le cas

$$\bar{a}\langle P \rangle \| R \| S \| T \| U \| a(x).Q \text{ qui se réduit en } 0 \| R \| S \| T \| U \| [P/x]Q.$$

Comment alors repérer la communication entre processus sans changer toute la syntaxe pour se retrouver avec le processus récepteur juste après le processus émetteur ?

Le principe du LTS est d'indiquer le comportement de chaque processus et les réductions possibles avec celui-ci. Ainsi on va indiquer pour chaque processus émetteur ou récepteur le canal sur lequel on émet ou on reçoit, et quel processus est émis. Il sera alors simple de détecter les communications possibles : pour un nombre quelconque de processus en parallèle, on les exprime avec le système de transitions étiquetées, puis on regarde les étiquettes pour voir si une réduction est possible. On dispose également d'une étiquette τ pour indiquer une réduction interne (qui n'émet ni ne reçoit aucun processus).

On a par exemple les règles suivantes :

$$\text{OUT } \bar{a}\langle P \rangle \xrightarrow{\bar{a}\langle P \rangle} 0$$

$$\text{IN } a(x).Q \xrightarrow{a(P)} [P/x]Q$$

$$\text{TAU1 Si } P \xrightarrow{\bar{a}\langle R \rangle} P' \text{ et } Q \xrightarrow{a(R)} Q' \text{ alors } P \| Q \xrightarrow{\tau} P' \| Q'$$

Sans ce système, le seul principe d'induction disponible pour faire des preuves est sur la syntaxe des processus. Ici, on va pouvoir se permettre de raisonner par induction sur les transitions.

Ce système permet alors de définir des techniques de preuves pour caractériser l'équivalence.

5.3 Correction de preuves

Un des avantages de formaliser avec Coq HOCore a été de repérer des fautes dans des démonstrations de l'article [1].

Par exemple, une des preuves raisonne inductivement sur des tailles de processus mais en utilisant une structure différente de HOCore.

Dans une autre preuve, on affirme implicitement que la décomposition première d'un processus en forme normale reste en forme normale alors qu'il y a des contre-exemples.

Certaines erreurs peuvent amener à redéfinir une notion pour pouvoir rester cohérent avec le reste des travaux.

Ces erreurs sont faciles à commettre à la main lorsque la complexité de la preuve en cache les subtilités, et refaire ces preuves en Coq garantit leur validité. Il s'agit cependant d'une grande partie du travail à effectuer : si la formalisation de Coq a nécessité 4000 lignes de code, les preuves s'étendent sur 22000 lignes.

Le développement est disponible à l'adresse www.irisa.fr/celtique/aschmitt/research/hocore/.

6 Conclusion

Le travail réalisé a donc consisté à formaliser la syntaxe d'HOCORE en Coq, puis sa sémantique en permettant à l'assistant de preuve d'effectuer des réductions. On doit alors utiliser des procédés comme par exemple l'approche canonique locale des noms pour reconnaître les variables locales identiques. On arrive ainsi à avoir une procédure pour montrer l'équivalence de deux processus. Ensuite, de nombreuses preuves sur ce calcul ont été faites en Coq. La vérification de Coq a permis de détecter des erreurs dans des preuves faites à la main, ou des définitions mal posées. Ce travail a également amélioré les connaissances sur HOCORE et corrigé les erreurs dans les preuves.

Il reste cependant des preuves à traduire en Coq et des résultats à formaliser.

À la connaissance des auteurs, il s'agit de la première formalisation d'un π -calcul d'ordre supérieur.

Références

- [1] Ivan Lanese, Jorge A Pérez, Davide Sangiorgi, and Alan Schmitt. On the expressiveness and decidability of higher-order process calculi. *Information and Computation*, 209(2) :198–226, February 2011. Extended abstract presented at *Logic in Computer Science (LICS)*, 2008. Coq formalization available at <http://www.irisa.fr/celtique/aschmitt/research/hocore/>.
- [2] Petar Maksimović and Alan Schmitt. Hocore in coq. In Christian Urban and Xingyuan Zhang, editors, *Proceedings of the 6th conference on Interactive Theorem Proving (ITP 2015)*, volume 9236 of *Lecture Notes in Computer Science*, pages 278–293. Springer, August 2015.
- [3] Randy Pollack, Masahiko Sato, and Wilmer Ricciotti. A canonical locally named representation of binding. *Journal of Automated Reasoning*, pages 1–23, May 2011. 10.1007/s10817-011-9229-y.
- [4] Damien Pous and Alan Schmitt. De la kam avec un processus d'ordre supérieur. In *Actes des 25èmes Journées Francophones des Langages Applicatifs*, January 2014.