

Benchmarks pour les projets de BWare

David Delahaye et Catherine Dubois

David.Delahaye@cnam.fr
dubois@ensiie.fr

Cedric / Cnam
Inria / Cedric / Ensiie

Plénière BWare

Cnam
24 juin 2013



Projets

- Projets donnés par ClearSy : basic, DAB, p4, p9 (pas p20) ;
- Projet donné par Mitsubishi : RCS3 (bientôt) ;
- Pour le moment : un total de 10 602 OP.

Prouveurs

- Prouveurs du projet :
 - ▶ mp (4.0) ;
 - ▶ Alt-Ergo (0.95) ;
 - ▶ Zenon (0.7.1) ;
 - ▶ iProver Modulo (0.7).
- Prouveurs considérés plus tard :
 - ▶ Super Zenon (superdéduction) ;
 - ▶ Zenon-B (extension à B avec superdéduction) ;
 - ▶ Zenon Modulo (déduction modulo).

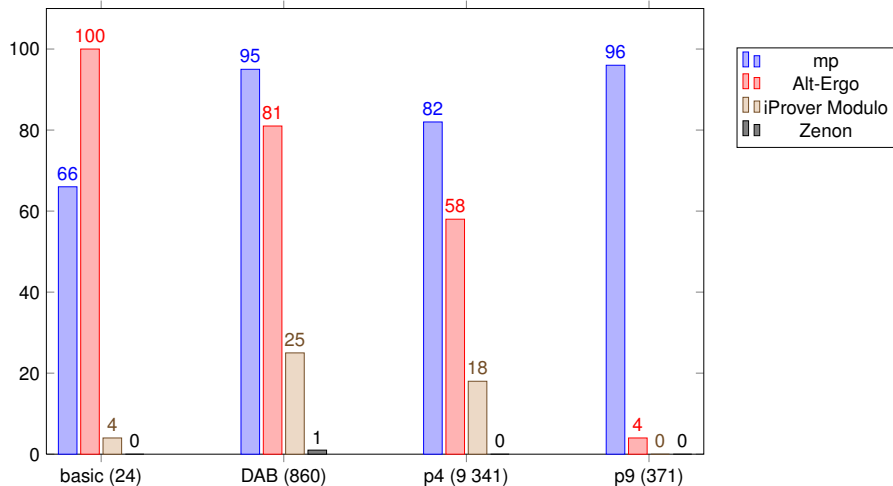
Résumé compact

Nb OP total : 10 602.

	mp	Alt-Ergo	iProver Modulo	Zenon
Nb OP prouvées	8 947	6 191	1 965	17
Proportion	84%	58%	19%	0,2%

- « Timeout » court (30 s) ;
- Machine Intel Xeon (24 coeurs) ;
- Mémoire utilisable (24 Go).

Résumé par projet



Difficultés concernant mp

- « Timeout » uniquement global (pas par OP) ;
- Interprétation des sorties de krt (version 4.0) :

```
PRV.0.II.NN.101["Arithmetic_check"]
PRV.1.II.NN.105["0","8"]
PRV.6.II.NN.107["AssertionLemmas"]
PRV.8.II.NN.109["1"]
PRV.8.II.NN.109["2"]
PRV.8.II.NN.109["3"]
PRV.7.II.NN.108["4"]
PRV.7.II.NN.108["5"]
PRV.7.II.NN.108["6"]
PRV.7.II.NN.108["7"]
PRV.7.II.NN.108["8"]
PRV.2.II.NN.106["novar"]
PRV.4.II.NN.103["AssertionLemmas","5","3"]
PRV.3.II.NN.102["Arithmetic_check","5","3"]
```

nb OP 108 (8 947) + nb OP 109 (1 699) = 10 646 \neq 10 602.

- Langage de traces difficile à analyser.

Difficultés concernant Why3

- « Timeout » dépassé dans certains cas (ici 30 s) :

```
File Lecteur_imp.why:
```

```
Lecteur_imp.why B_translation Initialisation_1 : Timeout (345.54s)
Lecteur_imp.why B_translation Initialisation_2 : Timeout (372.39s)
Lecteur_imp.why B_translation Initialisation_3 : Timeout (365.93s)
Lecteur_imp.why B_translation initialiser_lecteur_1 : Timeout (381.05s)
Lecteur_imp.why B_translation initialiser_lecteur_2 : Timeout (372.42s)
Lecteur_imp.why B_translation initialiser_lecteur_3 : Timeout (381.54s)
Lecteur_imp.why B_translation fermer_lecteur_1 : Timeout (372.85s)
Lecteur_imp.why B_translation scruter_lecteur_1 : Timeout (388.59s)
Lecteur_imp.why B_translation scruter_lecteur_2 : Timeout (398.91s)
Lecteur_imp.why B_translation scruter_lecteur_3 : Timeout (402.35s)
...
```

- Traçabilité des OP entre fichiers B et fichiers Why3 ;
- Difficulté pour faire des analyses croisées et des tests de couverture.

Comportement des prouveurs

- Très bonnes performances de mp (dédié à la théorie de B) ;
- Performances très raisonnables d'Alt-Ergo (et des autres solveurs SMT) ;
- Faibles performances pour le premier ordre (iProver Modulo, Zenon).

Améliorations pour le premier ordre

- Raisonnement modulo la théorie de B ;
- Intégration d'autres théories (arithmétique) ;
- Gestion des hypothèses (surtout pour Zenon).

Analyse de mp

- Gros cœur de preuve (validé) ;
- Gestion du contexte (fort nombre d'hypothèses) ;
- Comparaison avec pp (vérification des règles).

Superdédution

- Application de Zenon-B (superdédution + B) ;
- « Driver » Why3 à écrire (TPTP + opérateurs) ;
- Complétion à attendre (encore assez peu d'opérateurs) ;
- Utilisation de Super Zenon pour générer les sur-règles.

Dédution modulo

- Théorie de B modulo (G. Burel, D. Delahaye, C. Dubois, O. Hermant) ;
- Zenon Modulo (P. Halmagrand, exposé à venir) ;
- « Backend » Dedukti pour Zenon Modulo (F. Gilbert) ;
- « Backend » Dedukti pour iProver Modulo (G. Burel, exposé à venir).

Améliorations pour Zenon

- Traitement de l'égalité (CC, Zipperposition) ;
- Arithmétique (fait partie des tâches de BWare).