



Étude comparative de méthodes de programmation au sein du passage de message tolérant aux pannes

Michaël Cadilhac

`<michael.cadilhac@lrde.org>`



Sous la direction de Thomas Héroult

8 février 2006



Table des matières



Étude
comparative
dans
MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au
LRI

Le projet

Travail
demandé et
effectué

Conclusion

- 1 Présentation de l'étudiant
- 2 Présentation de l'entreprise
- 3 De l'étudiant dans l'entreprise
- 4 Le projet
- 5 Travail demandé et effectué
- 6 Conclusion



Table des matières



Étude
comparative
dans
MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au
LRI

Le projet

Travail
demandé et
effectué

Conclusion

- 1 Présentation de l'étudiant
- 2 Présentation de l'entreprise
- 3 De l'étudiant dans l'entreprise
- 4 Le projet
- 5 Travail demandé et effectué
- 6 Conclusion



Présentation de l'étudiant «ASV?»



Étude
comparative
dans
MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au
LRI

Le projet

Travail
demandé et
effectué

Conclusion

- Michaël Cadilhac (20 ans),
- Étudiant CSI/LRDE,
- Se destine à un troisième cycle,
- Intéressé par le monde de la recherche.





Présentation de l'étudiant «ASV?»



Étude
comparative
dans
MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au
LRI

Le projet

Travail
demandé et
effectué

Conclusion

- Michaël Cadilhac (20 ans),
- Étudiant CSI/LRDE,
- Se destine à un troisième cycle,
- Intéressé par le monde de la recherche.





Présentation de l'étudiant «ASV?»



Étude
comparative
dans
MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au
LRI

Le projet

Travail
demandé et
effectué

Conclusion

- Michaël Cadilhac (20 ans),
- Étudiant CSI/LRDE,
- Se destine à un troisième cycle,
- Intéressé par le monde de la recherche.





Présentation de l'étudiant «ASV?»



Étude
comparative
dans
MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au
LRI

Le projet

Travail
demandé et
effectué

Conclusion

- Michaël Cadilhac (20 ans),
- Étudiant CSI/LRDE,
- Se destine à un troisième cycle,
- Intéressé par le monde de la recherche.





- 1 Présentation de l'étudiant
- 2 Présentation de l'entreprise**
- 3 De l'étudiant dans l'entreprise
- 4 Le projet
- 5 Travail demandé et effectué
- 6 Conclusion



Présentation de l'entreprise

\$ man LRI



Étude
comparative
dans
MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au
LRI

Le projet

Travail
demandé et
effectué

Conclusion

- Créé en 1974 par l'Université de Paris-Sud,
- 2^{ième} laboratoire d'informatique de France,
- Près de 180 membres, dont 80 chercheurs,
- Sous la direction de Michel Beaudouin-Lafon.





Présentation de l'entreprise

\$ man LRI



Étude
comparative
dans
MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au
LRI

Le projet

Travail
demandé et
effectué

Conclusion

- Créé en 1974 par l'Université de Paris-Sud,
- 2^{ème} laboratoire d'informatique de France,
- Près de 180 membres, dont 80 chercheurs,
- Sous la direction de Michel Beaudouin-Lafon.





Présentation de l'entreprise

\$ man LRI



Étude
comparative
dans
MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au
LRI

Le projet

Travail
demandé et
effectué

Conclusion

- Créé en 1974 par l'Université de Paris-Sud,
- 2^{ème} laboratoire d'informatique de France,
- Près de 180 membres, dont 80 chercheurs,
- Sous la direction de Michel Beaudouin-Lafon.





Présentation de l'entreprise

\$ man LRI



Étude
comparative
dans
MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au
LRI

Le projet

Travail
demandé et
effectué

Conclusion

- Créé en 1974 par l'Université de Paris-Sud,
- 2^{ème} laboratoire d'informatique de France,
- Près de 180 membres, dont 80 chercheurs,
- Sous la direction de Michel Beaudouin-Lafon.





Présentation de l'entreprise

$\{x, x \in \mathcal{LRI}\}$



Étude comparative dans MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au LRI

Le projet

Travail demandé et effectué

Conclusion

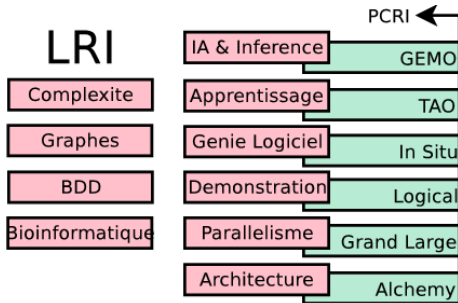


Fig.: Départements du LRI



- 1 Présentation de l'étudiant
- 2 Présentation de l'entreprise
- 3 De l'étudiant dans l'entreprise**
- 4 Le projet
- 5 Travail demandé et effectué
- 6 Conclusion



De l'étudiant dans l'entreprise

Ou l'infiltration épitéenne au LRI



Étude
comparative
dans
MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au
LRI

Le projet

Travail
demandé et
effectué

Conclusion



Projet Grand Large :

- Créé en 2003 et dirigé par Franck Cappello,
- Étudie le calcul dans les systèmes à grande échelle,
- Travail sur deux applicatifs :
 - XtremWeb : Plate-forme de travail pour systèmes distribués,
 - MPICH-V.



De l'étudiant dans l'entreprise

Ou l'infiltration épitéenne au LRI



Étude
comparative
dans
MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au
LRI

Le projet

Travail
demandé et
effectué

Conclusion



Projet Grand Large :

- Créé en 2003 et dirigé par Franck Cappello,
- Étudie le calcul dans les systèmes à grande échelle,
- Travail sur deux applicatifs :
 - XtremWeb : Plate-forme de travail pour systèmes distribués,
 - MPICH-V.



De l'étudiant dans l'entreprise

Ou l'infiltration épitéenne au LRI



Étude
comparative
dans
MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au
LRI

Le projet

Travail
demandé et
effectué

Conclusion



Projet Grand Large :

- Créé en 2003 et dirigé par Franck Cappello,
- Étudie le calcul dans les systèmes à grande échelle,
- Travail sur deux applicatifs :
 - XtremWeb : Plate-forme de travail pour systèmes distribués,
 - MPICH-V.



- 1 Présentation de l'étudiant
- 2 Présentation de l'entreprise
- 3 De l'étudiant dans l'entreprise
- 4 Le projet**
- 5 Travail demandé et effectué
- 6 Conclusion



MPI, MPICH, MPICH-V

Continuez la suite logique



Étude
comparative
dans
MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au
LRI

Le projet

Travail
demandé et
effectué

Conclusion

- MPI : *Message Passing Interface*, standard de communication par passage de message → programmation parallèle à mémoire partagée,
- MPICH : Implémentation de MPI en terme d'*interface* et de *mécanismes de communication*,
- MPICH-V : Implémentation de MPICH pour ressources volatiles (au dessus de TCP/IP).



MPI, MPICH, MPICH-V

Continuez la suite logique



Étude
comparative
dans
MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au
LRI

Le projet

Travail
demandé et
effectué

Conclusion

- MPI : *Message Passing Interface*, standard de communication par passage de message → programmation parallèle à mémoire partagée,
- MPICH : Implémentation de MPI en terme d'*interface* et de *mécanismes de communication*,
- MPICH-V : Implémentation de MPICH pour ressources volatiles (au dessus de TCP/IP).



MPI, MPICH, MPICH-V

Continuez la suite logique



Étude
comparative
dans
MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au
LRI

Le projet

Travail
demandé et
effectué

Conclusion

- MPI : *Message Passing Interface*, standard de communication par passage de message → programmation parallèle à mémoire partagée,
- MPICH : Implémentation de MPI en terme d'*interface* et de *mécanismes de communication*,
- MPICH-V : Implémentation de MPICH pour ressources volatiles (au dessus de TCP/IP).



Si un noeud tombe :

- Le calcul doit rester *juste*,
- Le calcul doit *finir*.

Contraintes supplémentaires pour MPICH-V :

- Seul le noeud tombé redémarrera,
- Les autres noeuds n'interviennent pas dans le redémarrage.



Tolérance aux pannes



Étude
comparative
dans
MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au
LRI

Le projet

Travail
demandé et
effectué

Conclusion

Si un noeud tombe :

- Le calcul doit rester *juste*,
- Le calcul doit *finir*.

Contraintes supplémentaires pour MPICH-V :

- Seul le noeud tombé redémarrera,
- Les autres noeuds n'interviennent pas dans le redémarrage.



MPICH-V et tolérance aux pannes



Étude comparative dans MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au LRI

Le projet

Travail demandé et effectué

Conclusion

Les communications passent par des serveurs qui :

- Sauvegardent les messages,
- Transfèrent les messages.

→ Le « canal-mémoire ».

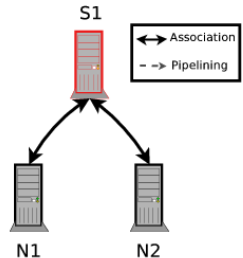


Fig.: Un CM, deux clients



- 1 Présentation de l'étudiant
- 2 Présentation de l'entreprise
- 3 De l'étudiant dans l'entreprise
- 4 Le projet
- 5 Travail demandé et effectué**
- 6 Conclusion



- Exécutions pseudo-parallèle d'un même processus,
- Mémoire partagée,
- Accès concurrents.

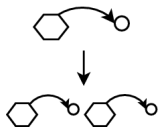


Fig.: *Fork* classique

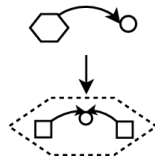


Fig.: *Fork* de thread



- Exécutions pseudo-parallèle d'un même processus,
- Mémoire partagée,
- Accès concurrents.

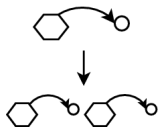


Fig.: *Fork* classique

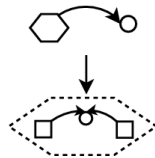


Fig.: *Fork* de thread



- Exécutions pseudo-parallèle d'un même processus,
- Mémoire partagée,
- Accès concurrents.

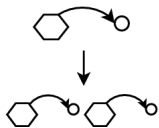


Fig.: *Fork* classique

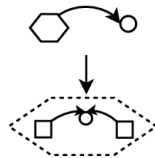


Fig.: *Fork* de thread



Étude comparative : Version 1

Mono-thread



Étude comparative dans MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au LRI

Le projet

Travail demandé et effectué

Conclusion

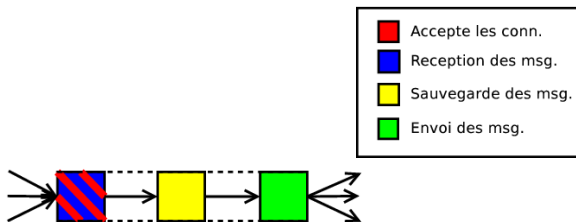


Fig.: Version mono *thread*



Étude comparative : Version 2

Un *thread* par client



Étude comparative dans MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au LRI

Le projet

Travail demandé et effectué

Conclusion

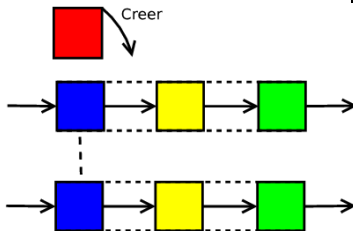
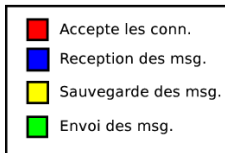


Fig.: Version un *thread* par client



Étude comparative : Version 3

Généralisation des précédentes



Étude comparative dans MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au LRI

Le projet

Travail demandé et effectué

Conclusion

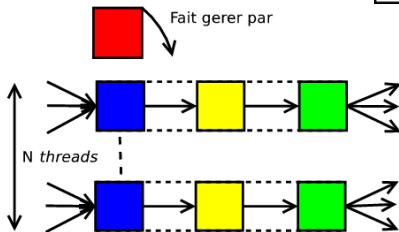
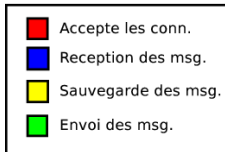


Fig.: Généralisation des précédentes



Étude comparative : Version 4

Threads spécialisés



Étude comparative dans MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au LRI

Le projet

Travail demandé et effectué

Conclusion

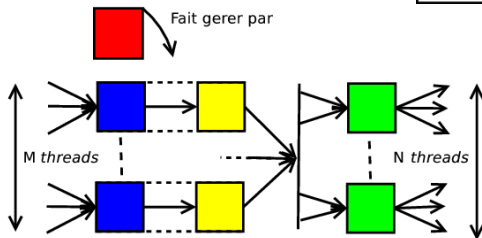
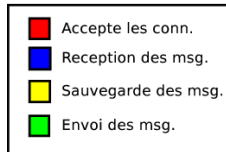


Fig.: Version *threads* spécialisés



- Aucune version ne se distingue des autres dans tous les cas,
- Rédaction d'un rapport européen de recherche (CoreGRID)

many kind of network and communication paradigms, including not only TCP and MPI. This version features the possibility to run simultaneously communication between many pairs, which is mandatory for our experiments.

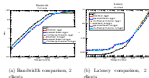


Fig. 4. Bandwidth and latency comparison for all implementations, and two different sizes

Figure 4, 5 and 6 presents bandwidth and latency measurements for respectively 2, 40 and 64 nodes using only one node. Contrary to FI, when using a single-rtw every communication pass through a single call. For 2 nodes, FI and our implementations have the same theoretical maximal bandwidth, but when dealing with 40 or 64 nodes, FI has a maximal bandwidth of around respectively 8 and 32 times the maximal theoretical bandwidth of a single call. The latency for FI for different number of nodes does not change since MPIFI exchanges messages only between Rank pair of nodes (between rank 0 and rank N, 1 and N-1 and so on). When a rtw has to deal with only 2 nodes, every implementation has up to the same performance due to blocking communication the rtw has to find only one communication at a time.

Figure 7 and 8 presents the bandwidth for three sizes of messages and the latency for 4 byte messages of the different rtw implementations when respectively sending and checking the largest and for different number of nodes (from 2 to 64). The bandwidth presented here is the bandwidth as measured by MPIFI, which is the overall bandwidth of the application. So, for 64 nodes, which represents 32 pairs (32 nodes send messages of increasing size and 32 receive them), the theoretical maximal bandwidth should be at least 32 Gb/s. However, since all communication pass through a single rtw, the theoretical maximal bandwidth is limited to 1 Gb/s (limitation of the Capable Ethernet card of the rtw). For small and medium message sizes, the non-threaded implementation has around the same performance as multithreaded implementations the

Fig.: Rapport CoreGRID



- 1 Présentation de l'étudiant
- 2 Présentation de l'entreprise
- 3 De l'étudiant dans l'entreprise
- 4 Le projet
- 5 Travail demandé et effectué
- 6 Conclusion**



Conclusion



Étude
comparative
dans
MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au
LRI

Le projet

Travail
demandé et
effectué

Conclusion

- Travail en laboratoire de recherche,
- Travail sur une problématique de recherche, avec une démarche de recherche,
- Découverte des problématiques du travail distribué,
- Écriture d'un rapport européen de recherche.



Conclusion



Étude
comparative
dans
MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au
LRI

Le projet

Travail
demandé et
effectué

Conclusion

- Travail en laboratoire de recherche,
- Travail sur une problématique de recherche, avec une démarche de recherche,
- Découverte des problématiques du travail distribué,
- Écriture d'un rapport européen de recherche.



Conclusion



Étude
comparative
dans
MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au
LRI

Le projet

Travail
demandé et
effectué

Conclusion

- Travail en laboratoire de recherche,
- Travail sur une problématique de recherche, avec une démarche de recherche,
- Découverte des problématiques du travail distribué,
- Écriture d'un rapport européen de recherche.



- Travail en laboratoire de recherche,
- Travail sur une problématique de recherche, avec une démarche de recherche,
- Découverte des problématiques du travail distribué,
- Écriture d'un rapport européen de recherche.



Questions ?



Étude
comparative
dans
MPICH-V

M. Cadilhac

L'étudiant

L'entreprise

L'étudiant au
LRI

Le projet

Travail
demandé et
effectué

Conclusion

¿ — ?