Vérification de preuves distribuées : compromis temps-espace

Laurent Feuilloley, Pierre Fraigniaud, Juho Hirvonen, Ami Paz, Mor Perry

Vérification de preuves distribuées: compromis temps-espace

Laurent Feuilloley[1] †, Pierre Fraigniaud[2] ‡, Juho Hirvonen[3] §, Ami Paz[4] ¶ et Mor Perry[5]

[1] *Sorbonnes université*
[2] *IRIF, CNRS and University Paris Diderot, France*
[3] *Aalto University*
[4] *IRIF, CNRS and University Paris Diderot, France*
[5] *School of Electrical Engineering, Tel-Aviv University, Israel*

Une preuve distribuée est un mécanisme permettant aux nœuds d'un réseau de décider collectivement et efficacement si le réseau est dans une configuration correcte, par rapport à un certain prédicat. La vérification de ces preuves est généralement faite en un nombre constant de rondes. Dans cet article nous étudions l'impact d'un temps de vérification non-constant sur la taille des preuves.

Mots-clefs : Preuve distribuée, certification distribuée, décision distribuée, compromis temps-espace, redondance

1 Introduction

La tolérance aux pannes est l'un des défis du calcul distribué sur réseau. L'une des réponses à ce défi est le développement d'algorithmes auto-stabilisants : des algorithmes qui convergent vers une configuration correcte à partir d'une configuration arbitraire. Certains de ces algorithmes sont dit silencieux, car les nœuds, quand ils ont atteint une configuration correcte, ne changent plus d'état. Cette propriété ne peut être atteinte que si les nœuds peuvent vérifier localement que la configuration décrite par les états du système est correcte. En général, cela nécessite pour chaque nœud d'avoir en mémoire, non seulement la description de la configuration, mais aussi un champ pour en *certifier* la correction. On parle alors de *preuve distribuée* ou de *certificat distribué*. La contrainte principale sur une preuve distribuée est que sa vérification doit pouvoir être faite localement, c'est-à-dire en n'ayant accès qu'aux voisins dans le réseau ; ceci dans le but de minimiser les ressources en temps et en communication.

La qualité d'une preuve distribuée est mesurée par sa taille, c'est-à-dire par le nombre de bits que chaque nœud utilise pour la stocker. Malheureusement, pour certains prédicats, on peut montrer que cette taille minimum est prohibitive. Diverses modifications du mécanisme de certification ont été explorées pour obtenir des preuves utilisant moins d'espace mémoire. Dans cet article, nous évaluons dans quelle mesure relâcher la contrainte de localité permet d'obtenir des preuves plus petites. En d'autres termes : si l'on permet aux nœuds voulant vérifier l'état du réseau de ne communiquer non pas seulement avec leurs voisins mais à plus grande distance, peut-on, en contrepartie, obtenir une certification moins coûteuse en espace ?

2 Travaux précédents

Le mécanisme de certification évoqué ci-dessus est communément appelé *schéma d'étiquetage de preuve* (ou *proof-labeling scheme* en anglais) [6]. Trouver la taille minimale des preuves pour un prédicat donné est

† Additional funding from ANR Descartes and INRIA project GANG.
‡ Same as for the first authore.
§ Supported by Ulla Tuominen Foundation, additional funding from ANR Descartes.
¶ Supported by Fondation Sciences Mathématiques de Paris, additional funding from ANR Descartes.

un problème crucial du domaine. Citons notamment la certification d'un arbre couvrant qui prend $\Theta(\log n)$ bits de mémoire (où n est la taille du réseau) [6, 3], et la certification d'un arbre couvrant de poids minimum, qui utilise $\Theta(\log^2 n)$ bits [4]. Il est connu que tout prédicat peut être certifié par une *preuve universelle* de $\Theta(n^2)$ bits, et que cette borne est atteinte par certains prédicats [6].

Au-delà de la tolérance aux pannes, les preuves distribuées forment un concept intéressant en soi. En particulier, elles définissent un équivalent de la classe de complexité NP en calcul distribué (voir [1] pour la description d'une théorie de la complexité en calcul distribué).

Le premier article portant sur une vérification en temps non-constant est [5] qui montre qu'en autorisant un nombre polylogarithmique de rondes, on peut avoir des preuves de taille $\Theta(\log n)$ pour l'arbre couvrant de poids minimum. L'article qui motive notre travail est [7], qui montre que pour vérifier l'acyclicité du réseau (et pour la preuve universelle), on peut atteindre un *compromis linéaire*, c'est-à-dire qu'une preuve de taille $f(n)$ dans le modèle classique, peut être transformée en une preuve de taille $O(f(n)/t)$, si on utilise un temps t. La question est alors : peut-on obtenir un compromis linéaire pour tout prédicat ?

3 Résultats

Nous répondons partiellement à la question du compromis linéaire, avec trois types de résultats, qui généralisent ceux de [5] et [7]. Nous montrons que :

— Sur des topologies de réseaux restreintes, comme les arbres, les cycles et les grilles, on peut obtenir un compromis linéaire pour tout prédicat.
— Si, dans le modèle classique, il existe une preuve distribuée optimale qui est uniforme, c'est-à-dire dans laquelle chaque nœud reçoit le même certificat, alors on obtient un compromis (au moins) linéaire.
— Pour une série de prédicats classiques du domaine, dont l'arbre couvrant et l'arbre couvrant minimum, des constructions *ad hoc* permettent un compromis linéaire.

Après une section détaillant le modèle, on développe ces trois résultats, avec des esquisses de preuves. Les preuves complètes, ainsi que des résultats supplémentaires (notamment une borne inférieure par réduction à un problème de complexité de la communication) peuvent être trouvés dans la version complète [2]. La version complète explique aussi en quoi, au-delà du gain en espace, l'étude du compromis améliore notre compréhension de la redondance des preuves distribuées.

4 Modèle et définitions

Le modèle de calcul distribué sur réseau utilisé ici est inspiré du modèle LOCAL [8]. Le réseau est modélisé par un graphe simple où les nœuds sont équipés d'identifiants uniques, sur $O(\log n)$ bits. On considère qu'en temps t, un nœud obtient une vue à distance t, c'est-à-dire qu'il connaît la structure du sous-graphe des nœuds à distance au plus t de lui, ainsi que leur entrées (si il y en a).

Les prédicats que l'on veut vérifier sont par exemple : « le graphe a diamètre k », ou « l'ensemble des arêtes sélectionnées forme un arbre (ou un arbre couvrant minimum) », ou encore « au plus un nœud est sélectionné ». Au cours de la vérification, chaque nœud récupère sa vue en temps t, et donne une unique sortie : *accepte* ou *rejette*. On dit que la configuration (c'est-à-dire le graphe, avec éventuellement des entrées) est acceptée si tous les nœuds acceptent, et qu'elle est rejetée si au moins un nœud rejette.

Une preuve distribué est un étiquetage où chaque nœud u reçoit une chaîne de bits (appelé la *preuve du nœud u*). Ces chaînes sont de même longueur (appelée la *taille de la preuve*), font partie de la vue des nœuds et sont utilisées par ceux-ci pour prendre une décision. Un schéma d'étiquetage de preuve est dit correct si la condition suivante est satisfaite. Il existe une preuve distribuée telle que tous nœuds acceptent, si et seulement si, la configuration respecte le prédicat.

Pour un prédicat donné, dont la taille de preuve optimale est $f(n)$ (pour une vérification classique, en temps 1), on dit que l'on a un *compromis linéaire* s'il existe des preuves de taille $\tilde{O}(f(n)/t)$ pour une vérification en temps t (où $\tilde{O}$ indique que l'on néglige les facteurs polylogarithmiques).

5 Topologies restreintes

On commence par montrer que si la topologie est assez régulière alors un compromis linéaire existe. De plus dans ces cas, on peut transformer de manière automatique un schéma en temps 1 en un schéma en temps t.

Théorème 1 *Dans les chemins, les arbres, les grilles, les cycles et les tores, tout prédicat a un compromis linéaire.*

Esquisse de preuve. L'intuition de la preuve est la suivante, dans le cas d'un chemin. Prenons une preuve distribuée pour une vérification en temps 1, et soit $f(n)$ sa taille. Effaçons toutes les preuves, à part celles qui sont portées par des nœuds à distance 0 modulo t d'une extrémité fixée du chemin. Considérons la vue d'un sommet en temps t après cette opération. Ou bien ce sommet a gardé sa preuve, ou bien il peut voir un sommet à sa droite et un sommet à sa gauche ayant gardé leurs preuves. On montre alors que grâce à ses « ancres » ayant gardé leurs preuves, chaque nœud peut deviner une preuve pour lui-même, et que l'ensemble de ces preuves devinées (en plus de celles qui ont été gardées) forment collectivement une preuve distribuée correcte. À ce niveau, nous obtenons un vérificateur en temps t avec des preuves qui *en moyenne* sont de taille $O(f(n)/t)$.

Nous devons maintenant obtenir des preuves dont la taille est *uniformément* $O(f(n)/t)$. L'idée est de découper les preuves restantes en blocs en longueur $O(f(n)/t)$ et de répartir les blocs entre les nœuds proches de manière à ce que : (1) chaque nœud ne reçoivent qu'un nombre constant de blocs, et (2) chaque nœud puisse reconstruire correctement les preuves restantes à partir de sa vue à distance t (ou détecter que la preuve est mal formée). □

Les deux contraintes de cette technique de preuve sont (1) d'avoir un petit ensemble de nœuds à partir desquels on peut reconstruire des preuves complètes, et (2) d'avoir assez de structure dans la topologie pour permettre une répartition des preuves avec la possibilité de reconstruction. Ces contraintes expliquent que l'on se place dans des classes de graphes très régulières.

6 Preuves uniformes

On appelle preuve uniforme, une preuve distribuée (en temps 1) telle que tous les nœuds reçoivent la même preuve. Pour certains prédicats, une preuve uniforme est optimale. C'est le cas pour le prédicat qui spécifie que le graphe est symétrique (c'est-à-dire qu'il est isomorphe à lui-même, de manière non-triviale), qui est aussi un exemple de prédicat qui demande une preuve de taille $\Omega(n^2)$. C'est aussi le cas pour le prédicat « au plus un nœud est sélectionné ». De manière plus générale, dans une preuve distribuée, il existe souvent une partie qui est commune à tous les nœuds et une partie plus spécifique, et le théorème suivant peut être appliqué à la partie uniforme.

Théorème 2 *Les prédicats ayant une preuve optimale uniforme, ont un compromis linéaire.*

Esquisse de preuve. Le principe général est le même que pour la seconde partie de la preuve du Théorème 1 : on découpe la preuve en blocs (comme les preuves des nœuds sont toutes identiques, on peut parler de « la » preuve) et on attribue à chaque nœud un certain nombre de blocs de taille $O(f(n)/t)$. Une fois encore, on doit s'assurer que chaque nœud puisse obtenir tous les blocs pertinents en ne regardant que son voisinage à distance t. Ici la topologie n'est pas restreinte, donc on ne peut pas profiter d'une structure régulière. On utilise alors une méthode probabiliste pour montrer qu'une telle attribution de blocs existe toujours. Notons que la démonstration est ainsi non-constructive, en ce qui concerne les preuves. □

Le vrai résultat est en fait plus fort : si la taille des boules de rayon t est bornée inférieurement par une fonction $b(t)$, alors la taille des preuves obtenue est dans $\tilde{O}(f(n)/b(t))$. Ainsi si le nombre de nœuds augmente de manière exponentielle (par exemple dans le centre d'un arbre Δ-régulier), alors on a un compromis exponentiel.

7 Prédicats classiques

Dans cette section, on considère des prédicats classiques, pour lesquels on prouve un compromis linéaire par des méthodes *ad hoc*. En plus des prédicats définis plus haut, on s'intéresse au diamètre (≪ le graphe a-t-il diamètre k ?≫) et au couvreur additif (*additive spanner*)(≪ les arêtes sélectionnées forment-elles un couvreur de terme additif β ? ≫).

Théorème 3 *Les prédicats suivants ont un compromis linéaire : arbre couvrant, arbre couvrant de poids minimum, diamètre et couvreur additif.*

Esquisse de preuve. Pour l'arbre couvrant, la preuve distribuée en temps 1 consiste à donner à chaque nœud l'identifiant de la racine, ainsi que la distance du nœud à la racine. Comme l'identifiant est le même pour tous les nœuds, on peut utiliser la technique du Théorème 2 pour le répartir entre les nœuds. Pour les distances, la stratégie est similaire à celle du Théorème 1 : (1) on sélectionne un sous-ensemble de nœuds dont on conserve les distances, en s'assurant que chaque nœud peut recalculer sa propre distance, et (2) on répartit les bits de ses distances sur les nœuds proches. Cela demande plus de techniques que précédemment car la topologie n'est plus restreinte, mais on peut utiliser le fait que la structure à vérifier est supposée être un arbre. Pour l'arbre couvrant minimum, le schéma classique utilise une pile de $O(\log n)$ certificats d'arbre couvrant, et en substance, on utilise la technique ci-dessus sur chacun d'eux.

Enfin pour le diamètre (respectivement le couvreur additif), le schéma classique consiste à donner à chaque nœud u, un vecteur établissant pour chaque nœud v, la distance de u à v dans le graphe (respectivement dans le sous-graphe des arêtes sélectionnées). Dans un schéma en temps t, on ne donne à chaque nœud, qu'un sous-ensemble des distances, tel que la propriété suivante soit vérifiée. Pour tout couple de sommets u et v, il existe un nœud dans le voisinage de u, qui est sur le plus court chemin du u à v, et qui a dans sa liste la distance à v. Ainsi, chaque nœud peut retrouver toutes les distances, en inspectant son voisinage. Comme pour la preuve du Théorème 2, la preuve utilise une méthode probabiliste. □

8 Conclusion

La question qui motivait cet article était de savoir si tous les prédicats ont un compromis linéaire. Il a été montré que dans plusieurs cas importants et pour les prédicats classiques, la réponse est positive. Il reste désormais à répondre à cette question en toute généralité.

Références

[1] Laurent Feuilloley and Pierre Fraigniaud. Survey of distributed decision. *Bulletin of the EATCS*, 119, 2016.

[2] Laurent Feuilloley, Pierre Fraigniaud, Juho Hirvonen, Ami Paz, and Mor Perry. Redundancy in distributed proofs. In *DISC 2018*, pages 24 :1–24 :18, 2018.

[3] Mika Göös and Jukka Suomela. Locally checkable proofs in distributed computing. *Theory of Computing*, 12(1) :1–33, 2016.

[4] Amos Korman and Shay Kutten. Distributed verification of minimum spanning trees. *Distributed Computing*, 20 :253–266, 2007.

[5] Amos Korman, Shay Kutten, and Toshimitsu Masuzawa. Fast and compact self-stabilizing verification, computation, and fault detection of an MST. *Distributed Computing*, 28(4) :253–295, 2015.

[6] Amos Korman, Shay Kutten, and David Peleg. Proof labeling schemes. *Distributed Computing*, 22(4) :215–233, 2010.

[7] Rafail Ostrovsky, Mor Perry, and Will Rosenbaum. Space-time tradeoffs for distributed verification. In *SIROCCO 2017*, pages 53–70, 2017.

[8] David Peleg. *Distributed Computing : A Locality-Sensitive Approach*. Discrete Mathematics and Applications. SIAM, Philadelphia, 2000.