Réduction de la taille de stockage des blockchains

Benjamin LOISON, Emmanuelle ANCEAUME, IRISA

2 août 2021

Le contexte général

Les blockchains est un concept récent en informatique, défini pour sa célèbre application dans Bitcoin par S. Nakamoto en 2008. Bitcoin étant une blockchain permettant des échanges de la monnaie virtuelle Bitcoin de manière décentralisée, c'est-à-dire sans intervention d'états ou de banques. Il est intéressant de remarquer que les blockchains est un des rares domaines où la pratique à une avance importante sur la théorie, par exemple Bitcoin a été démontré sûr sous certaines conditions seulement en 2014 par Juan A. Garay.
Une des grandes difficultés pour les blockchains est le passage à l'échelle, c'est-à-dire le fait de maintenir la stabilité et une interaction aisée avec la blockchain même si le nombre d'utilisateurs augmente d'un ordre de grandeur. Dans le cas de Bitcoin afin de vérifier les transactions, que ce soit en tant que mineur (utilisateur sécurisant le réseau) ou en tant que full node (utilisateur vérifiant le réseau), ceux-ci doivent vérifier que chaque transaction du réseau est légitime et correcte. Cependant pour se faire ils doivent retracer la provenance de l'argent dans tout l'historique de Bitcoin qui pèse 338 Go actuellement en 2021. Cette quantité de données stockée de manière linéaire en l'utilisation de la blockchain en plus de ralentir l'initialisation des mineurs et full nodes (qui doivent alors télécharger sur le réseau peer-to-peer Bitcoin l'intégralité de la blockchain), cela empêche les utilisateurs lambdas utilisant par exemple leur téléphone de vérifier le réseau.
Le logiciel de référence Bitcoin Core pour les noeuds qui constituent le réseau Bitcoin propose une option d'élagage qui supprime l'historique de Bitcoin pour ne garder que l'état courant qu'une fois le téléchargement entier de la blockchain de Bitcoin effectué. En pratique jusqu'à maintenant les smartphones se basait sur la technique du Simple Payment Verification (SPV) qui consiste à dépendre de full nodes et d'avoir à attendre un certain nombre de confirmations dans la blockchain pour s'assurer que le paiement effectué depuis son smartphone est bien pris en compte pour de bon par le réseau.

Le problème étudié

Notre solution permet de réduire la consommation en bande passante et aussi en stockage, bien que des solutions sur ce dernier point existe déjà. Effectivement à part la consommation en électricité très élevé de Bitcoin, équivalente à un pays comme Israël, à cause de sa preuve de travail nécessitant la compétition de très nombreux processeurs, un problème majeur pour quelqu'un voulant participer au protocole est l'initialisation et le stockage. En pratique l'initialisation dure 10 jours avec une connexion fibre 1 Go/s puisque le téléchargement de la blockchain de Bitcoin grâce aux noeuds est très lent. En plus des plus de 340 Go que le noeud initialisant doit allouer pour conserver la blockchain de Bitcoin. Ces deux points découragent de nombreux amateurs, alors que justement Bitcoin se veut être une cryptomonnaie décentralisée sécurisée par la participation au protocole de tout le monde.

La contribution proposée

Une idée peut alors être de stocker uniquement un état courant vérifié du montant monétaire appartenant à chaque utilisateur, ainsi au lieu de parcourir toute l'histoire de la provenance de l'argent, on peut simplement vérifier le solde du compte. Cela permet notamment de se débarasser de l'historique des transactions et donc de la majeur partie de la blockchain de Bitcoin tout en gardant le même niveau de sécurité. Mes travaux se base sur cette idée ingénieuse de l'article "Mining in Logarithmic Space" qui a toutefois ses limites que l'on va essayer de repousser, puisque effectivement cet article ne traite pas le cas d'une difficulté croissante pour les mineurs ce qui est le cas dans Bitcoin. En pratique après application numérique, on transformera la blockchain de Bitcoin de 338 Go à 4.2 Go, ce qui permet notamment alors à un téléphone moderne de vérifier aisément le réseau Bitcoin.

Les arguments en faveur de sa validité

Le cas particulier de Bitcoin par rapport au papier "Mining in Logarithmic Space", traitant les blockchains de manière générale, profite des preuves de ce dernier et de celles inhérentes au papier "The Bitcoin Backbone Protocol with Chains of Variable Difficulty". Tant que l'hypothèse selon laquelle la majorité des noeuds participants protocole sont honnêtes est vérifiée, la sécurité de cette approche est garantie. Toutefois en pratique si on implémente cette approche toute en conservant l'ancienne en fonctionnement, on remarque que le haché de l'état actuel n'est vérifié que par les noueds éxécutant le nouveau protocole, cependant on pourra remarquer que seul le dernier haché est considéré dans notre approche et que donc celui-ci est le fruit du concensus de la majorité des noeuds éxécutant le nouveau protocole et que donc puisque les noeuds déjà initialisés passant au nouveau protocole peuvent vérifier ce haché de manière indépendante, on peut espérer que notre contribution garantie la correction des données partagées par la blockchain.

Le bilan et les perspectives

Par rapport à l'alternative d'élagage proposée par Bitcoin Core, notre contribution rétrocompatible permet d'une part de ne pas avoir besoin de stocker, même temporairement, l'entièreté de la blockchain de Bitcoin et permet surtout d'autre part de ne même pas avoir à télécharger l'entièreté de la blockchain mais seulement l'état courant et quelques blocs bien sélectionnés assurant l'authenticité de l'état courant reçu.
De cette manière si une telle approche était utilisée à l'initialisation des plus de 10 000 noeuds Bitcoin, on pourrait économiser plus de 3 000 To de bande passante.
Pour permettre aux nouveaux noeuds s'initialisant de bénéficier de cette initialisation rapide et légère il faudrait proposer une modification de Bitcoin Core implémentant notre approche et une modification d'un des logiciels utilisant pour miner du Bitcoin pour rajouter le haché de l'état courant dans les blocs de la blockchain afin de profiter pleinement de l'efficacité de notre approche.
Cette approche ne peut être généralisée qu'aux blockchains utilisant la preuve de travail.