Vers une meilleure identification d'acteurs de Bitcoin par apprentissage supervisé

Rafael Ramos Tubino*, Rémy Cazabet*
Céline Robardet**

*Univ Lyon, Université Lyon 1, CNRS, LIRIS UMR5205, F-69622 France
**Univ Lyon, INSA Lyon, CNRS, LIRIS UMR5205, F-69621 France
prenom.nom@liris.cnrs.fr

Résumé. Bitcoin est la crypto-monnaie la plus largement répandue et la plus étudiée. De par sa nature décentralisée, les données de transactions sont librement accessibles et peuvent être analysées. La première étape de la plupart des analyses consiste à regrouper les adresses anonymes en agrégats supposés correspondre à des acteurs. Dans cet article, nous proposons une nouvelle méthode pour réaliser ces agrégats à base d'apprentissage automatique. Notre approche repose sur la construction d'un jeu de données d'apprentissage dont la variable de classe est obtenue par une vérité de terrain calculée a posteriori. Ce jeu de données est utilisé pour identifier les adresses de change des transactions, adresses appartenant au donneur d'ordre de la transaction. Cela nous permet d'augmenter le nombre d'adresses découvertes appartenant à un même acteur. Nous montrons expérimentalement la pertinence de cette méthode en comparaison des heuristiques habituellement utilisées à l'aide d'un critère de validation externe.

1 Introduction

Présentée en 2008 et créée l'année suivante, Bitcoin (BTC) est une monnaie numérique basée sur le principe de la chaîne de blocs (Blockchain). Sa spécificité tient à plusieurs caractéristiques. Tout d'abord, Bitcoin est une monnaie **décentralisée**, c'est-à-dire que le système fonctionne sans autorité centrale, ni administrateur unique. Elle est gérée par un réseau de nœuds validateurs ouvert à tous, ordinateurs mettant à contribution leur puissance de calcul informatique afin de vérifier, de sécuriser et d'inscrire les transactions dans la blockchain. Ces nœuds assurent **la fiabilité** du système, puisque chaque transaction doit être validée par la majorité des nœuds validateurs pour être enregistrée. Enfin, **l'anonymat** est garanti par le chiffrement des identités des bénéficiaires et des donneurs d'ordre, même si toutes les transactions effectuées sont enregistrées dans un registre public. C'est ce dernier point qui rend la blockchain pseudo-anonyme, car les bénéficiaires et donneurs d'ordre des transactions sont identifiés par leur adresse publique. Il est donc possible de tracer toutes les transactions impliquant une même adresse. Afin d'accroître l'anonymat, un même acteur change donc régulièrement d'adresse. Nous définissons un **acteur** comme une personne, un groupe de personnes, une entreprise, ou une quelconque entité qui détient un ensemble de clés privés, et donc un ensemble