Une nouvelle approche pour la détection d'anomalies dans les flux de graphes hétérogènes

Abd Errahmane Kiouche[*,**], Karima Amrouche[*]
Hamida Seba[**], Sofiane Lagraa[***]

[*]Laboratoire de la Communication dans les Systèmes Informatiques(LCSI)
École nationale Supérieure d'Informatique, BP 68M, 16309, Oued-Smar, Alger, Algérie.
http://www.esi.dz
[**]Université de Lyon, CNRS, Université Lyon 1,
LIRIS, UMR5205, F−69622 Lyon, France.
[***]SnT, Université du Luxembourg, L-1855 Luxembourg

Résumé. Nous proposons dans ce travail une nouvelle approche de détection d'anomalies dans un flux de graphes hétérogènes orientés et étiquetés. Notre approche utilise une nouvelle représentation des graphes par des vecteurs. Cette représentation est flexible et permet de mettre à jour les vecteurs de graphes de manière incrémentale à fur et à mesure de l'arrivée de nouvelles arêtes. Elle est applicable à n'importe quel type de graphes et optimise l'espace mémoire utilisé. De plus, elle permet la détection d'anomalies en temps réel.

1 Introduction

La détection d'anomalies est un domaine de recherche très actif traité par plusieurs communautés scientifiques telles que : la sécurité informatique, la médecine, l'industrie et la finance. De façon générale, ce problème consiste à détecter les données qui sont significativement différentes des données bénignes ou normales. De nos jours, les données sont de plus en plus représentées par les graphes car ces derniers ont la faculté de modéliser les interactions complexes de façon simple et intuitive. Un graphe $G = (V, E)$ est un outil de représentation de données formé d'un ensemble de sommets V et d'un ensemble de liens (arêtes) E entre les sommets. Lorsque les données sont représentées par des graphes, le problème de détection d'anomalies revient à repérer les graphes qui sont différents des graphes correspondants aux objets normaux observés par le système. De plus, les graphes en flux (graph stream) sont de plus en plus utilisés. En effet, dans la plupart des applications de surveillance en temps réel, la structure complète des graphes n'est pas connue, car les graphes grandissent et évoluent au fil du temps. De même, lorsque les graphes sont trop volumineux pour être chargés entièrement en mémoire centrale, les traiter dans le modèle de flux de données où le flux est en général une séquence d'arêtes est une nécessité. La détection d'anomalies dans un flux d'arêtes pose plusieurs défis comme le traitement incrémental des arêtes, la gestion de l'espace mémoire occupé par le flux et la détection des anomalies en temps réel.

Dans ce travail, nous nous intéressons au problème de la détection d'anomalies dans un flux de graphes hétérogènes et étiquetés. Notre application principale est la sécurité des systèmes