Abstract

The detection and suspension of fake accounts (Sybils) is a major challenge in Online Social Networks (OSNs). These accounts undermine the services of OSNs and exploit trusted users' accounts for their benefit. Existing social-graph-based approaches restrict Sybil accounts by relying on the structure of links between suspicious and trusted (non-Sybil) accounts.

However, Sybil users continue to acquire connections by sending unsolicited friend requests (friend spam) to non-Sybil users. Therefore, new solutions are required. Rejecto is an innovative system developed by researchers of Cyprus University of Technology (CUT) and Duke University in collaboration with Facebook. It is based on the observation that non-Sybil users tend to reject friend spam. Leveraging this insight, Rejecto achieves accurate detection and further restricts Sybil users.

Rejecto uses and extends the Kernighan technique, a graph partitioning algorithm. Hence, Rejecto can be treated as a graph algorithm, which enables a wider range of efficient tools to be used for its deployment. Among them we distinguish Apache Giraph, an open source framework that comes to fill the gap of existing tools for large scale graph processing. Giraph is based on the Pregel model that allows the users to implement graph algorithms in an intuitive way, adding scalability, resilience and fault tolerance.

In this thesis project, we: i) present our study on the architecture of popular parallel processing paradigms; ii) survey existing social-graph-based approaches; iii) describe the architecture of Apache Giraph; iv) expose the reasons why Rejecto can benefit from an implementation under this open source project.