Genetic Algorithm-Based Adversarial Attack on SybilSCAR

Heechan Yang
heechan.yang@kaist.ac.kr
KAIST
Daejeon, South Korea

Yoonha Bahng
ybahng@kaist.ac.kr
KAIST
Daejeon, South Korea

Abstract

Sybil attack is a threat to online social network platforms enabling attackers to perform malicious activities, spread fake news, and more. SybilSCAR [2] is a tool that identifies sybil accounts considering relations of accounts within the network. We believe that by analyzing attributes of an adversarial attack on SybilSCAR, we can learn ways to improve SybilSCAR's performance on correctly identifying sybil accounts. Hence, in this work, we aim to leverage an evolutionary algorithm to construct a set of adversarial graphs as an attack on SybilSCAR and study the attributes observed in these graphs.

ACM Reference Format:
Heechan Yang and Yoonha Bahng. 2024. Genetic Algorithm-Based Adversarial Attack on SybilSCAR. In . ACM, New York, NY, USA, 1 page. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

In online social networks (OSNs), Sybil accounts manipulate the network, disseminate misinformation, and conduct fraudulent activities. These attacks pose a significant threat to the reliability of OSNs. To address this issue, previous work such as SybilSCAR has proposed algorithms to detect Sybil accounts.

SybilSCAR attempts to identify Sybil accounts by analyzing the structural relationships between accounts. While Sybil detection mechanisms such as SybilSCAR have proven effective in identifying malicious accounts, adversaries may find ways to evade detection by exploiting vulnerabilities, such as modifying the structure of the account graphs.

In this work, we aim to explore the limitations of Sybil detection algorithms by constructing adversarial graphs using a genetic algorithm. These adversarial graphs are designed to defeat Sybil detection algorithms. By generating and analyzing a series of adversarial graphs, we can observe how well existing detection algorithms perform under various attack scenarios.

2 Background

Genetic Algorithm (GA)[1] is a search based method to automatically find an optimal solution for hard-to-solve problems based on evolutionary selection. Given an initial population, GA iterates N numbers of generations (N being the available budget) on this population. At each iteration, individuals are given a fitness to show how fit they are to the given environment. Based on these fitness scores, individuals with low fitness scores are killed while individuals with higher fitness scores get the chance to pass their genes to the next generation. By conducting this iteration N numbers of generations, the GA finds an optimal set of individuals that best fit the given constraint or environment.

3 Proposal

SybilSCAR is a tool that classifies Sybil (fake) accounts based on a given graph (network of accounts). However, attackers manipulate graphs to disturb such tools to correctly detect a sybil account. Therefore, we aim to leverage GA to find an optimal graph in which SybilSCAR fails to correctly detect sybil accounts. By doing so, we can analyze what attributes make these attacks invincible to SybilSCAR and contribute in ways to improve the performance of sybil account detection.

4 Expected Outcomes

By conducting this research, we aim to identify the graph structures that are effective at evading detection by SibilSCAR and other Sybil detection algorithms. We expect that the adversarial graphs will progressively evolve to exploit weaknesses in the detection algorithms.

The mutation of the generic algorithm will disrupt the structural features of the network that Sybil detection algorithms rely on, thereby making the adversarial graphs harder to detect.

Ultimately, this research would provide valuable insights how Sybil detection mechanisms can be improved by identifying the structural patterns that adversarial graphs may use to evade current detection algorithms.

References

[1] Melanie Mitchell. 1998. *An introduction to genetic algorithms.* MIT press.
[2] Binghui Wang, Le Zhang, and Neil Zhenqiang Gong. 2017. SybilSCAR: Sybil detection in online social networks via local rule based propagation. In *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*. 1–9. https://doi.org/10.1109/INFOCOM.2017.8057066

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn