31. KHADEEJA SHAGUFTA K V

Defending APT Attack using Reuse in SPARK

Advanced Persistent Threat (APT) attack, an attack option in recent years, poses serious threats to the security of governments and enterprises data due to its advanced and persistent attacking characteristics. To address this issue, a security policy of big data analysis has been proposed based on the analysis of log data of servers and terminals in Spark.I propose a scheduling optimization technique based on the reuse of datasets to improve Spark performance. In this technique, we define and formulate the reuse degree of Directed Acyclic Graphs (DAGs) in Spark based on Resilient Distributed Datasets (RDDs). Then, I define a global optimization function to obtain the optimal DAG sequence, that is, the sequence with the least execution time. To implement the global optimization function, we further propose a novel cost optimization algorithm based on the traditional Genetic Algorithm (GA).