

METHODS AND APPARATUSES FOR EXECUTING GPU TASK IN CONFIDENTIAL COMPUTE ARCHITECTURE

ABSTRACT

A graphics processing unit (GPU) task is executed in a confidential compute architecture. GPU software in a non-secure world configures, based on task code and a cache description of a GPU task, a stub data structure including cache areas allocated based on the cache description and metadata indicating each cache area. In a realm segment in a memory, a root world root monitor configures a real data structure corresponding to the stub data structure, and stores to-be-processed confidential data. The root monitor updates a granule protection table (GPT) table so that based on an updated GPT table, a target segment storing the metadata and the task code is accessible to a GPU and has realm world permission for all other objects. The root monitor modifies a target mapping relationship so that the GPU executes the GPU task by using the target segment and the real data structure.

FIG. 3