

Assignment 1 : Part A - Summary

Title: Mitigating Cold Start Problem in Serverless Computing: A Reinforcement Learning Approach

Authors: P. Vahidinia, B. Farahani, and F. S. Aliee

Journal: IEEE Internet of Things Journal

Summary:

The paper aims to address the cold start problem in serverless computing and propose a reinforcement learning approach to mitigate its impact [1]. Cold start delay is a significant challenge in serverless computing, directly affecting performance and customer satisfaction [1]. The authors aim to reduce this delay in real-time applications by determining the idle-container window and optimizing the number of prewarmed containers [1].

The central idea proposed by the authors is a two-layer approach that combines reinforcement learning and long short-term memory (LSTM) prediction [1]. In the first layer, a reinforcement learning algorithm is used to discover the function invocation pattern and determine the idle-container window [1]. This helps understand when and how long the containers should be kept warm [1]. In the second layer, LSTM prediction is employed to estimate the next invocation time, which helps determine the required number of prewarmed containers [1].

To evaluate their approach, the authors conducted experiments and compared their results with the OpenWhisk platform [1]. The evaluations showed that it is possible to determine the idle-container window by discovering the pattern of invocations, leading to a reduction in the number of cold start occurrences and control over memory consumption [1]. The proposed approach achieved comparable results to the OpenWhisk platform regarding the number of cold start delays [1]. Additionally, the proposed approach's cold start occurrence reduction layer showed improvements in memory consumption time for different average invocations per hour [1].

In conclusion, the authors proposed a reinforcement learning approach to mitigate the cold start problem in serverless computing [1]. By discovering the invocation pattern and optimizing the idle-container window, the proposed approach effectively reduces cold-start delays and controls memory consumption [1]. The experimental results demonstrate the feasibility and effectiveness of the approach in improving performance and customer satisfaction in real-time serverless applications [1].

Overall, the paper presents a novel approach to addressing the cold start problem in serverless computing and provides valuable insights into reducing cold start delays through reinforcement learning and predictive modeling techniques [1].

Citations:

[1] P. Vahidinia, B. Farahani, and F. S. Aliee, "Mitigating Cold Start Problem in Serverless Computing: A Reinforcement Learning Approach," in IEEE Internet of Things Journal, doi: 10.1109/JIOT.2022.3165127, 2022. [Online]. Available: <https://ieeexplore-ieee-org.ezproxy.library.dal.ca/document/9749611>. [Accessed: 7 June 2023].