Speculative Virtual Machine Streaming

Mizuki OKA[1] Richard POTTER[2] Kazuhiko KATO[1,2]

Our research concerns the problem of optimization of virtual machine migration over wide area networks without assuming the pre-copying of a base image so that users can use a pre-configured Linux virtual machine as quickly as possible. One approach is to retrieve needed blocks on demand. However, this approach could cause significant user latency due to available bandwidth and/or network latency. Our goal is thus to speculatively stream needed blocks so that users do not experience latency.

In order to eliminate user latency, the system should download multiple possible courses of user actions in advance. This is only possible when high bandwidth is available. This leads us to an important issue: the scheduling of the order of download for the predicted blocks to eliminate the latency as much as possible. In order to explore this issue, we developed a system that learns previous users activity and dynamically schedules the most probable order of blocks to download taking into account the user's current state. Given a tree of access pattern model of the previous users and a metric (e.g., cache hit ratio), various streaming orders, such as breadth-first, depth-first, or hybrid, are simulated and the system chooses the optimal order based on the current user's bandwidth and the predicted time for blocks to be used.

The results are the system that adjusts to variations in user behavior and network performance by changing the order of blocks to download.

1 Graduate School of Systems and Information Engineering, Department of Computer Science, University of Tsukuba
2 Japan Science and Technology Agency (JST)