Memory Trace Analysis using Machine Learning

LA-UR-20-25819

Arm instruction emulator (ArmIE) provides its users with the capability to compile SVE code with the Arm Compiler and run the SVE binary without SVE-enabled hardware. Memory traces are produced in a binary file during the emulator’s run time. These files contain information about each trace including whether an access was a read or a write, and what memory address was accessed. Using Python libraries like Pandas and Pyspark, it was possible to decode, and plot the data to get a general idea of what the data looks like. After this, early stages of machine learning techniques were applied to the data using libraries like Keras, Tensor Flow, and Scikit-learn. Using decision trees, neural networks, and clustering techniques, models are generated that can predict with a certain degree of accuracy what address was accessed, as well as whether it was a read or write. This talk will encompass the methods and tools mentioned above and how they were applied to accomplish these goals, as well as discuss the next steps moving forward for the coming months.