Benchmarking has always been a critical piece at procuring and accepting new systems, as well as providing concrete data to application developers about expected performance from specific computing platforms. While the community already benefits from many benchmarking suites, the latest improvements in hardware technologies are creating a gap between the metrics captured by these benchmarks and new hardware-level capabilities. An example of such a gap, and the associated response by the High-Performance Computing (HPC) community, is the introduction of the exaflop metric that proved to be more suitable for modern workloads that include machine learning or deep learning algorithms running on Graphical Processing Units (GPUs). That new metric is not meant to replace other metrics such as floating-point operations per second (FLOPS) but instead to complete them.

The focus of the High-Performance Compute Availability (HPCA) group is to provide a benchmark suite for smart networks that are based on devices such as smart switches, Data Processing Units (DPUs), or Network integrated Processing Units (IPUs). These innovative technologies provide a unique opportunity to offload computation to the network, freeing valuable resources on the hosts and enabling new optimizations. For example, smart networks can be used to offload collective operations, such as non-blocking operations from the Message Passing Interface (MPI) standard. Our project is therefore an effort to create and compile a new set of metrics for ranking HPC and Artificial Intelligence (AI) system performance and capabilities, by providing all the necessary benchmarks to capture the benefits enabled by state-of-the-art smart network devices. HPCA is intended as a complement to existing benchmarks, such as HPCC and HPCG.

The current prototype from our group, called OpenHPCA, provides a set of metrics based on MPI. Two existing benchmarks, an extension of an existing benchmark and a brand-new suite have been integrated together. The two existing benchmarks are the Ohio State University (OSU) micro-benchmarks and the Sandia micro-benchmarks (SMB). The modified benchmark is an extension of the OSU micro-benchmarks for non-contiguous memory, which is useful for the evaluation of new capabilities such as User-Mode Memory Registration (UMR). The new benchmark aims at completing the overlapping evaluation provided by both OSU and SMB, by adding a benchmark suite that provides a set of metrics based on a different statistical methodology to compute overlap. The new overlap benchmarks therefore provide a more complete overview of overlapping capabilities when combined with the other metrics. The current implementation of OpenHPCA provides a framework for the execution of all the individual benchmarks and analysis of the results. It also provides a graphical interface that makes it easier for users to visualize and analyze the results.

Our future work includes the development of benchmarks for new programming interfaces targeting state-of-the-art smart networks such as OpenSNAPI. By doing so, we plan to provide a novel set of metrics that capture the benefits of current and future smart network devices and the impact on applications.