



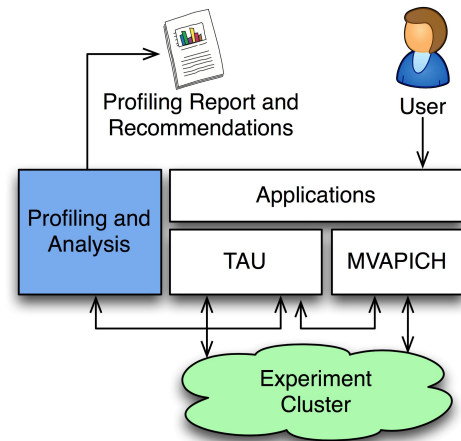
# SI2-SSI: Collaborative Research: A Software Infrastructure for MPI Performance Engineering: Integrating MVAPICH and TAU via the MPI Tools Interface

Award #: ACI-1450440 & ACI-1450471

Co-PIs: D. K. Panda, Sameer Shende  
Institutions: Ohio State University, University of Oregon

## Research Challenges

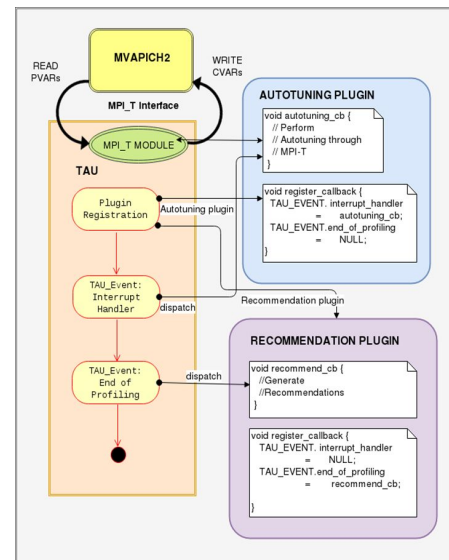
Creating an MPI programming infrastructure that can integrate performance analysis capabilities more directly, through the MPI Tools Information Interface, monitor Performance metrics during run time, and deliver greater optimization opportunities for scientific applications.



## Enabling Runtime Control

- TAU defines a plugin API to deliver access control to the internal plugin map
- User can specify a regular expression to control plugins executed for a class of named states at runtime

## Plugin Infrastructure



- Fully-customizable plugin infrastructure based on callback event handler registration for salient states inside TAU:
  - Function Registration / Entry / Exit
  - Phase Entry / Exit
  - Atomic Event Registration / Trigger
  - Init / Finalize Profiling
  - Interrupt Handler
  - MPI\_T
- Application can define its own "trigger" states and associated plugins

## Enhanced MPI\_T Support in MVAPICH2

- Added MPI\_T PVARs for various MPI collectives (Bcast, Reduce, Allreduce etc) to measure
- Added PVARs and CVARs for host and device based MPI operations
- Added MPI\_T PVAR timers to measure the time taken and the number of calls pertaining to various collective algorithms (allreduce, barrier, reduce etc)
- Added support for dynamic MPI\_T PVAR counter arrays where each index in the array represents a counter for a "bucket" or a user specified message range
- Added new CVARs that can be tuned at run-time

## Phase-based Recommendation

- MiniAMR: Benefits from hardware offloading using SHArP hardware offload protocol supported by MVAPICH2 for MPI\_Allreduce operation
- Recommendation Plugin:
  - Registers callback for "Phase Exit" event
  - Monitors message size through PMPI interface
  - If message size is low and execution time inside MPI\_Allreduce is significant, a recommendation is generated on ParaProf (TAU's GUI) for the user to set the CVAR enabling SHArP

### Per-thread, Per-phase Recommendation Generated as Metadata on ParaProf

Name	Value
TAU_MEMORY_PROTECT_BELOW	off
TAU_MEMORY_PROTECT_FREE	off
TAU_MPI_T_ENABLE_USER_TUNING_POLICY	off
TAU_OPENMP_RUNTIME	on
TAU_OPENMP_RUNTIME_EVENTS	off
TAU_OPENMP_RUNTIME_STATES	off
TAU_OUTPUT_GUIDE_CSZ	off
TAU_PAPI_MULTIPLEXING	off
TAU_PROFILE	off
TAU_PROFILE_FORMAT	profile
TAU_RECOMMENDATION_PHASE_ALLOCATE	MPI_T_RECOMMEND_SHARP_USAGE: No performance benefit foreseen with SHArP usage
TAU_RECOMMENDATION_PHASE_REALLOCATE	MPI_T_RECOMMEND_SHARP_USAGE: You could see potential improvement in performance by enabling MV2_ENABLE_SHARP in MVAPICH version 2.3a and above
TAU_RECOMMENDATION_PHASE_SHARP	MPI_T_RECOMMEND_SHARP_USAGE: You could see potential improvement in performance by enabling MV2_ENABLE_SHARP in MVAPICH version 2.3a and above
TAU_RECOMMENDATION_PHASE_SHARP_INIT	MPI_T_RECOMMEND_SHARP_USAGE: No performance benefit foreseen with SHArP usage
TAU_RECOMMENDATION_PHASE_PROFILE	MPI_T_RECOMMEND_SHARP_USAGE: You could see potential improvement in performance by enabling MV2_ENABLE_SHARP in MVAPICH version 2.3a and above
TAU_REGION_ACCESSORS	off
TAU_SAMPLING	off