debugging/strace

Strace

Basic Usage

- Set up strace and build environment

  ```
  module purge
  module add compiler/gnu
  module add devel/strace
  ```

- Build `stream` benchmark

  ```
  gcc -Ofast -march=native -fopenmp stream.c -o stream -lm
  ```

- Set up OpenMP environment

  ```
  export OMP_NUM_THREADS=4
  export OMP_PROC_BIND=TRUE
  export OMP_PLACES=cores
  ```

- Trace all Linux systemcalls of benchmark stream

  ```
  strace ./stream
  ```

- Strace

  - Filter for `openat` systemcalls
  - Discard standard output
  - Redirect standard error output to standard output to to allow forwarding to `grep`
  - Filter non-successful `openat` by `grep`

  ```
  strace -e openat ./stream 2>&1 1>/dev/null |
      grep -v "No such file or directory"
  ```

- Strace

  - Filter with regular expression for systemcalls containing "open"
  - Discard standard output
  - Show only successful systemcalls

  ```
  strace --trace='/.*open.*' --status=successful ./stream 1>/dev/null
  ```

```
openat(AT_FDCWD, "/etc/ld.so.cache", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/lib64/libm.so.6", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/opt/gcc/12/lib64/libgomp.so.1", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/lib64/libpthread.so.0", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/lib64/libc.so.6", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/lib64/libdl.so.2", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/sys/devices/system/cpu", O_RDONLY|O_NONBLOCK|O_CLOEXEC|O_DIRECTORY)
openat(AT_FDCWD, "/sys/devices/system/cpu/cpu0/topology/core_siblings_list", O_RDONLY)
openat(AT_FDCWD, "/sys/devices/system/cpu/cpu0/topology/thread_siblings_list", O_RDONLY
...
```

Usage scenarios with OpenMPI

- Set up strace and build environment

```
module purge
module add \
    compiler/gnu \
    mpi/openmpi
module add devel/strace
```

- Build `rank_league` benchmark

```
mpicc -O2 -march=native rank_league.c -o rank_league
```

- Strace all MPI ranks to individual files (e.g. for comparison)

```
mpirun -np 4 bash -c \
    'strace -o strace.out.${OMPI_COMM_WORLD_RANK} ./rank_league'
ll -h strace.out.*

-rw-r--r-- 1 bq0742 hk-project-scs 6.8M May  5 09:45 strace.out.0
-rw-r--r-- 1 bq0742 hk-project-scs 6.9M May  5 09:45 strace.out.1
-rw-r--r-- 1 bq0742 hk-project-scs 6.8M May  5 09:45 strace.out.2
-rw-r--r-- 1 bq0742 hk-project-scs 6.8M May  5 09:45 strace.out.3
```

- Strace
  - Only on first MPI rank (e.g. for data reduction)
  - Redirect trace to file

```
mpirun -np 4 bash -c \
    'if [[ ${OMPI_COMM_WORLD_RANK} -eq 0 ]]; then
        exec strace -o strace.out \
            ./rank_league
    else
        exec ./rank_league
    fi'
ll -h strace.out

-rw-r--r-- 1 bq0742 hk-project-scs 7084964 May  5 09:51 strace.out
```

2

- Strace
    - Only on first MPI rank
    - Filter for `openat` systemcalls,
    - Show only successful systemcalls
    - grep for loaded dynamic libraries

```
mpirun -np 4 bash -c \
    'if [[ ${OMPI_COMM_WORLD_RANK} -eq 0 ]]; then
        exec strace -e openat --status=successful -o"| grep [.]so" \
            ./rank_league
    else
        exec ./rank_league
    fi'
```

```
openat(AT_FDCWD, "/software/all/mpi/openmpi/4.1_gnu_11/lib64/libmpi.so.40", O_RDONLY|O_
openat(AT_FDCWD, "/usr/lib64/libc.so.6", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/liblustreapi.so.1", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libgpfs.so", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/software/all/mpi/openmpi/4.1.5_gnu_11/lib64/libopen-rte.so.40", O_RD
openat(AT_FDCWD, "/software/all/mpi/openmpi/4.1.5_gnu_11/lib64/libopen-pal.so.40", O_RD
openat(AT_FDCWD, "/usr/lib64/libucp.so.0", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libuct.so.0", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libucs.so.0", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libnuma.so.1", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libucm.so.0", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libz.so.1", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libdl.so.2", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libpthread.so.0", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libpmi2.so.0", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libpmi.so.0", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/librt.so.1", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libutil.so.1", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/opt/hwloc/2.7/lib/libhwloc.so.15", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libm.so.6", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libudev.so.1", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libpciaccess.so.0", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libxml2.so.2", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libevent_core-2.1.so.6", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libevent_pthreads-2.1.so.6", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libreadline.so.7", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libresolv.so.2", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/slurm/libslurm_pmi.so", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libmount.so.1", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libgcc_s.so.1", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/liblzma.so.5", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libcrypto.so.1.1", O_RDONLY|O_CLOEXEC) = 3
```

```
openat(AT_FDCWD, "/usr/lib64/libtinfo.so.6", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libblkid.so.1", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libuuid.so.1", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libselinux.so.1", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/usr/lib64/libpcre2-8.so.0", O_RDONLY|O_CLOEXEC) = 3
openat(AT_FDCWD, "/etc/ld.so.cache", O_RDONLY|O_CLOEXEC) = 15
openat(AT_FDCWD, "/etc/ld.so.cache", O_RDONLY|O_CLOEXEC) = 15
openat(AT_FDCWD, "/usr/lib64/ucx/libuct_cuda.so.0", O_RDONLY|O_CLOEXEC) = 16
openat(AT_FDCWD, "/etc/ld.so.cache", O_RDONLY|O_CLOEXEC) = 16
openat(AT_FDCWD, "/usr/lib64/ucx/libuct_cuda.so.0", O_RDONLY|O_CLOEXEC) = 16
openat(AT_FDCWD, "/etc/ld.so.cache", O_RDONLY|O_CLOEXEC) = 16
openat(AT_FDCWD, "/usr/lib64/ucx/libuct_ib.so.0", O_RDONLY|O_CLOEXEC) = 16
openat(AT_FDCWD, "/etc/ld.so.cache", O_RDONLY|O_CLOEXEC) = 16
openat(AT_FDCWD, "/lib64/libibverbs.so.1", O_RDONLY|O_CLOEXEC) = 16
openat(AT_FDCWD, "/lib64/libmlx5.so.1", O_RDONLY|O_CLOEXEC) = 16
openat(AT_FDCWD, "/lib64/libnl-route-3.so.200", O_RDONLY|O_CLOEXEC) = 16
openat(AT_FDCWD, "/lib64/libnl-3.so.200", O_RDONLY|O_CLOEXEC) = 16
openat(AT_FDCWD, "/usr/lib64/ucx/libuct_rdmacm.so.0", O_RDONLY|O_CLOEXEC) = 16
openat(AT_FDCWD, "/etc/ld.so.cache", O_RDONLY|O_CLOEXEC) = 16
openat(AT_FDCWD, "/lib64/librdmacm.so.1", O_RDONLY|O_CLOEXEC) = 16
openat(AT_FDCWD, "/usr/lib64/ucx/libuct_cma.so.0", O_RDONLY|O_CLOEXEC) = 16
openat(AT_FDCWD, "/usr/lib64/ucx/libuct_knem.so.0", O_RDONLY|O_CLOEXEC) = 16
openat(AT_FDCWD, "/usr/lib64/ucx/libucm_cuda.so.0", O_RDONLY|O_CLOEXEC) = 28
openat(AT_FDCWD, "/etc/ld.so.cache", O_RDONLY|O_CLOEXEC) = 28
openat(AT_FDCWD, "/usr/lib64/ucx/libucm_cuda.so.0", O_RDONLY|O_CLOEXEC) = 28
openat(AT_FDCWD, "/etc/ld.so.cache", O_RDONLY|O_CLOEXEC) = 28
```