MPI Reproducibility for Debugging

Reproducibility of High Performance Codes and Simulations – Tools, Techniques, Debugging

Kento Sato

November 19th, 2015
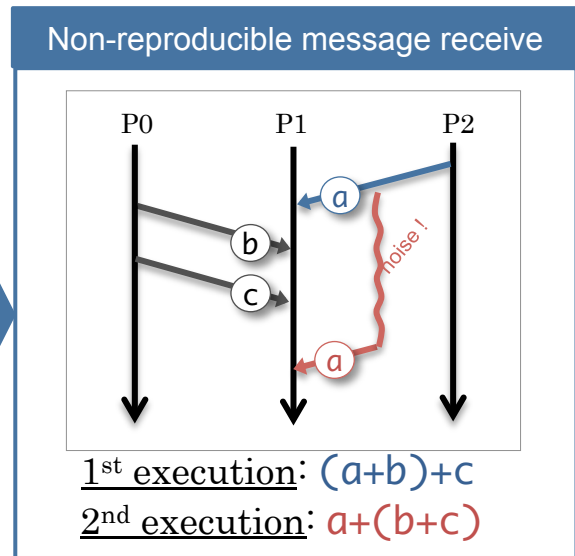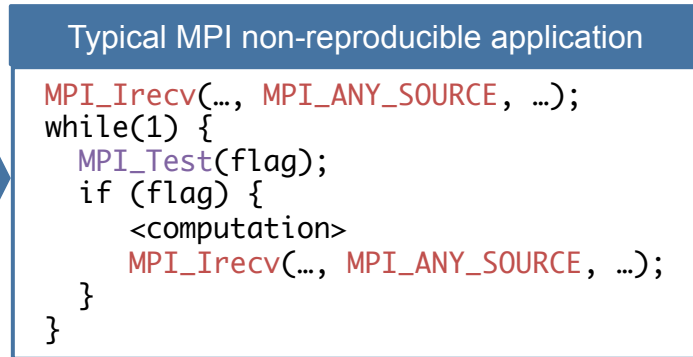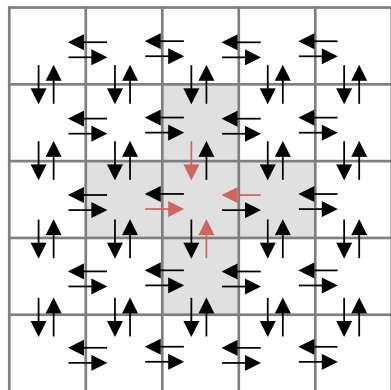
Lawrence Livermore National Laboratory

Q1: How do you define "Reproducibility" ?

➔ A1: In terms of MPI, if "message receive orders" consistent across different executions, I define it as "Reproducible"

Q2: Where reproducibility has been or could be an issue ?

➔ A2: Without MPI reproducibility, application developers will spend more time for debugging

Non-reproducible message receive

Typical MPI non-reproducible application

```
MPI_Irecv(…, MPI_ANY_SOURCE, …);
while(1) {
  MPI_Test(flag);
  if (flag) {
      <computation>
      MPI_Irecv(…, MPI_ANY_SOURCE, …);
  }
}
```

1st execution: $(a+b)+c$
2nd execution: $a+(b+c)$
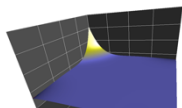
CASE STUDY: Monte Carlo Simulation (MCB)

- Non-reproducible control flow
  - Successful run, seg-fault or hang

- Non-reproducible numerical results
  - Floating-point arithmetic is "NOT" necessarily associative

$$(a+b)+c \neq a+(b+c)$$

➡ Developers need to do debug runs until the same bug manifests

➡ Running as intended ?
Application bugs ?
Silent data corruption happened ?

```
$ diff result_run1.out result_run2.out
result_run1.out:< IMC E_RR_total  -3.314023409e-05   -8.302693774e-08   2.9153322360e-08   -4.819850656e-06   2.311382122e-06
result_run2.out:> IMC E_RR_total  -3.314023410e-05   -8.302693776e-08   2.9153322360e-08   -4.819850657e-06   2.311382121e-06
```

| 09e-05 | 74e-08 | 56e-06 | 22e-06 |
| 10e-05 | 76e-08 | 57e-06 | 21e-06 |

MCB: Monte Carlo Benchmark

Lawrence Livermore National Laboratory
LLNL-PRES-679419

NNSA
National Nuclear Security Administration
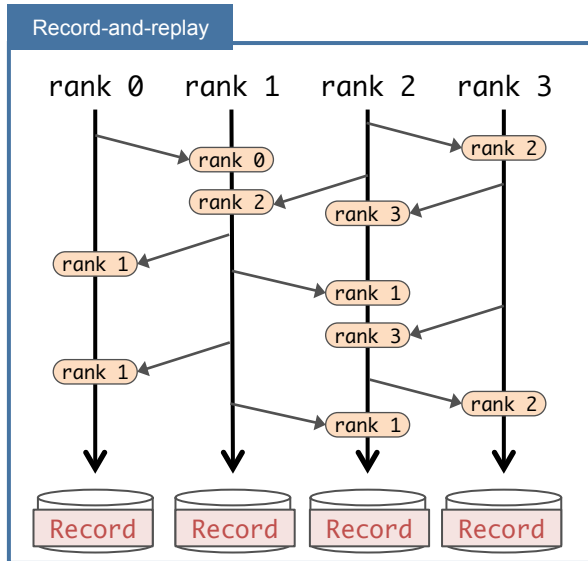
Q3: What solutions do you envision ?

➔ A3: Record-and-replay

- Traces and records message receive orders in a run, and replays the orders in successive runs for debugging
  - Record-and-replay can reproduce a target control flow
  - Developers can focus on debugging a particular control flow

Although record-and-replay has several challenges, this technique can reproduce a particular control flows and numerical results, thereby can reduce cost for debugging non-reproducible MPI applications