DMTCP

Dr. Gene Cooperman
Professor
Northeastern University

Why Use Checkpoint-Restart?

- My application takes 1 week to run, and I only have a 48-hour batch allocation slot.
- There are upcoming shutdowns (either planned or unplanned; due to electricity shutdown, building work, etc.).
- Computer nodes sometimes crash..

DMTCP: What does it stand for?

- DMTCP:
  Distributed MultiThreaded CheckPointing

- Current and past industrial support from:

- Academic support from:

So, what is Checkpoint-Restart/DMTCP?

Checkpoint-Restart is the ability to save a running application to disk (and restart it).

- DMTCP is a widely used C-R package for HPC.
- 120 publications by others using DMTCP:
  http://dmtcp.sourceforge.net/publications.html
- Periodic ckpt or under program control.
- Runtime overhead typically less than 1%.
- End-user plugins for added flexibility.
- Leveraging support at NERSC Supercomputer Center (best practices)

Philosophy of DMTCP

- ***Open Source***: LGPL-v3 (non-contagious license)
  - ➔ Considering switching to Apache license

- ***Transparent*** checkpointing:
  - No modification to application binaries; no modification to Linux kernel or system libraries

- ***Long-term Standards*** for **robustness**:
  - Mostly built on top of **POSIX** syscalls

- ***Synergy*** between **academia** (publishing) and **industry** (impact on current practices)

Long-Term Research Questions

FUNDAMENTAL QUESTION:

What are the limits of
transparent checkpointing*?*

- What about MPI over InfiniBand or TCP?
  (Yes, DMTCP: last 7 years)
- And newer networks? (CRAY GNI, HPE Slingshot, Intel Omni-Path, InfiniBand extensions, …):
  (Yes, DMTCP/MANA: last two years)
- What about CUDA? (for NVIDIA GPUs)
  (Yes, DMTCP/CRAC: last year)

***EXTENDING DMTCP:** **plugins, virtual ids, split processes (see later)**

Checkpointing to Stable Storage

The ***time to checkpoint*** is a long-standing question.
The time can be decomposed into:

1) Saving kernel/network status: a few seconds

2) Saving the memory image to stable storage:
   depends on bandwidth of storage

EXAMPLE: storage to disk: ~100 MB/s
storage to SSD: ~500 MB/s
storage to Optane: ~5 GB/s

Big Data Applications:
Is there a CoW (copy-on-write) filesystem for snapshot?

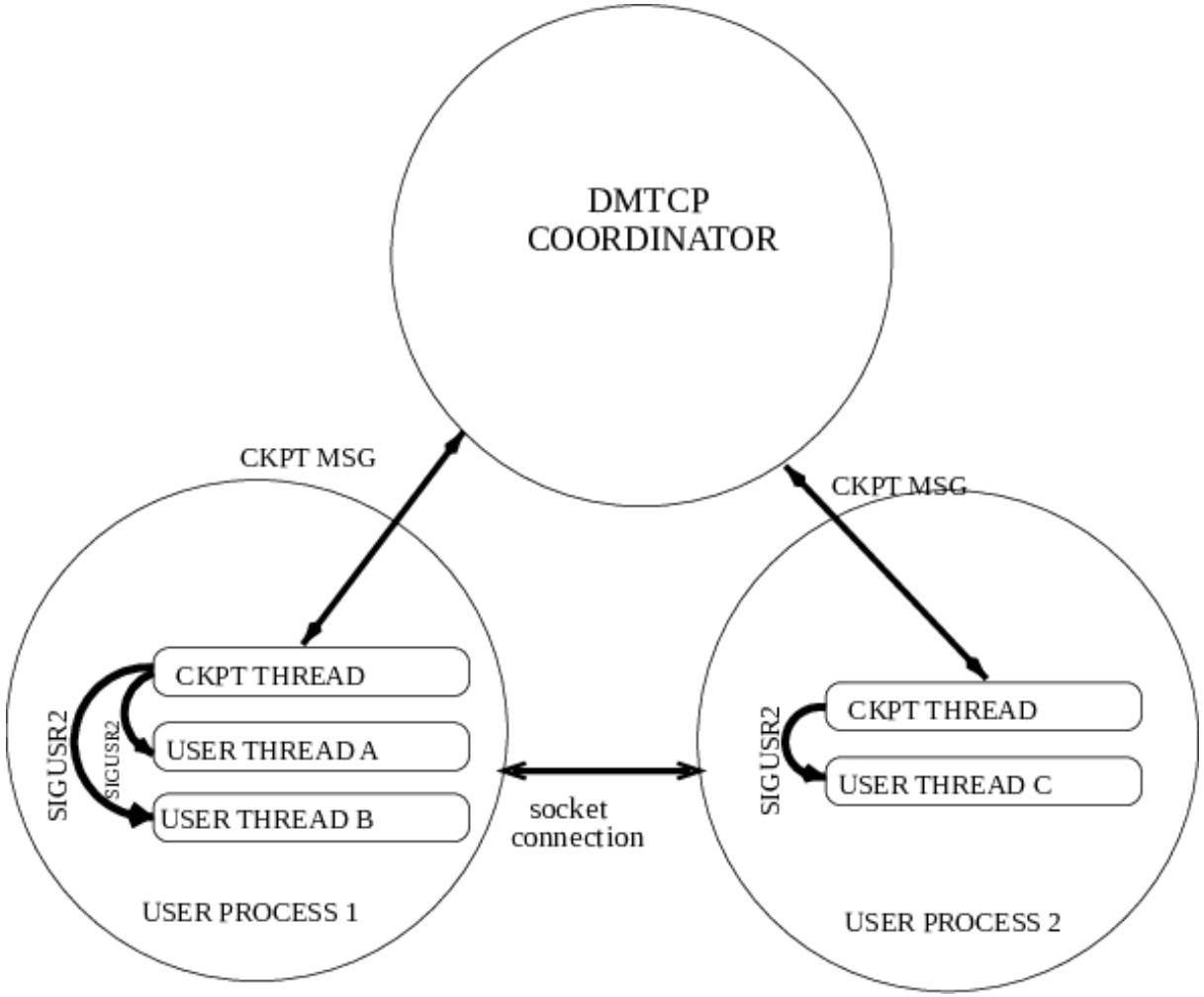
Persistent Storage/Memory and HPC

Newer persistent memories in clusters are used in at least three application-visible ways:

1) **Substitute for RAM:** provides persistence in database applications
2) **Local burst-buffer:** serves to accelerate writes to secondary storage
3) **Intermediate level in the storage hierarchy:**
   fast local-node storage or shared storage among nearby nodes

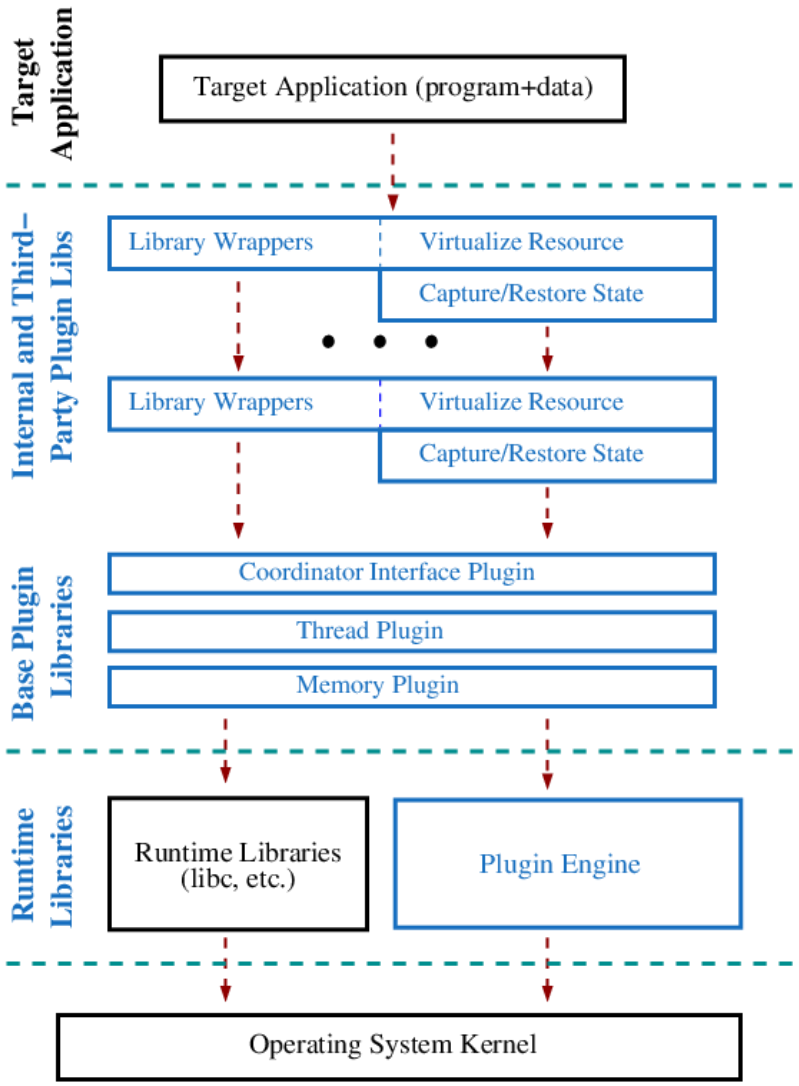
DMTCP plugins *are used to recognize these situations during ckpt; and then restore during restart.*

DMTCP Internal Architecture



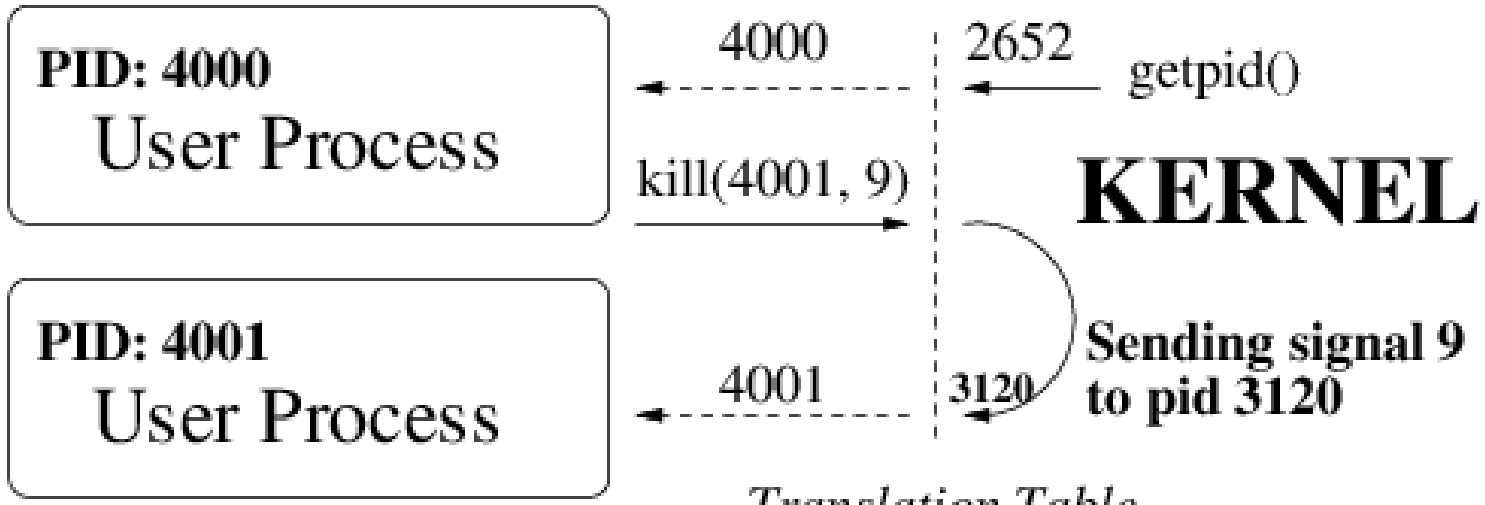
Plugins, Virtual Ids, and Split Process:

Pt. 1: Plugins



Plugins, Virtual Ids, and Split Process

Part 2: Virtual Ids: **PID Plugin**

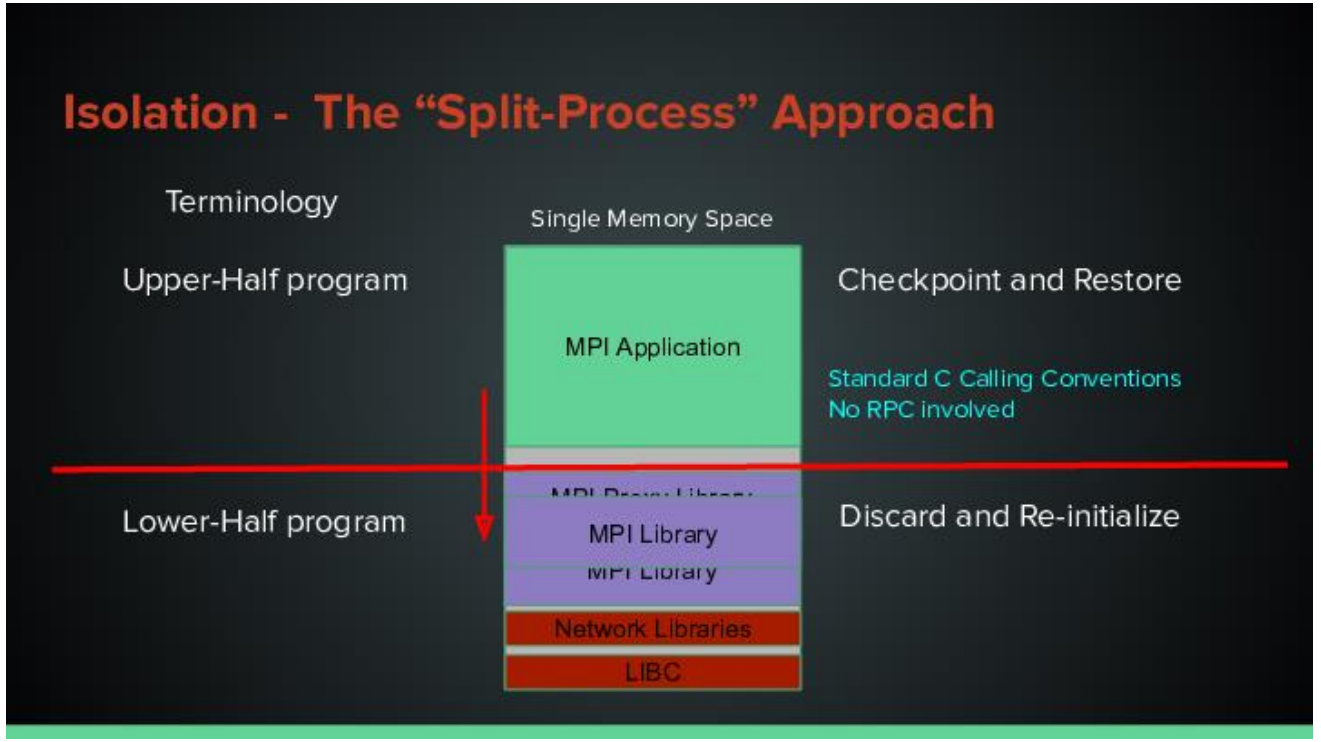


Translation Table

Virt. PID	Real PID
4000	2652
4001	3120

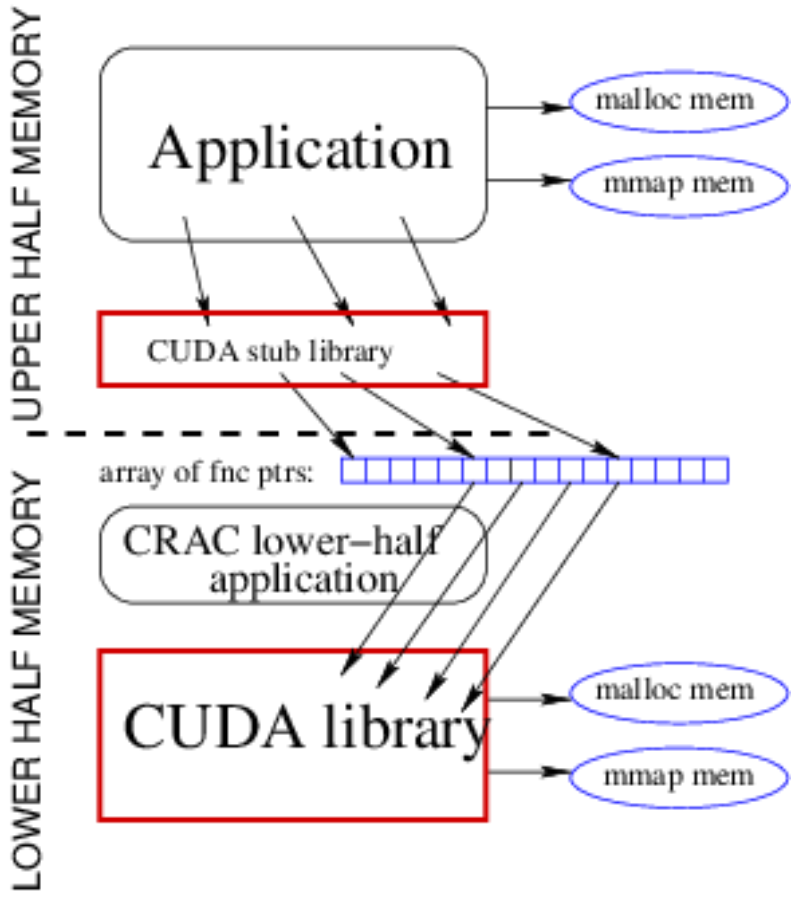
Plugins, Virtual Ids, and Split Process

Part 3a: Split Processes: **MANA Plugin** for MPI



Plugins, Virtual Ids, and Split Process

Part 3b:
Split Processes:
CRAC Plugin
for CUDA



Acknowledgment

THANKS TO THE MANY STUDENTS AND OTHERS WHO HAVE CONTRIBUTED TO DMTCP OVER THE YEARS:

Jason Ansel, Kapil Arya, Alex Brick, Jiajun Cao, Prashant Chouhan, Tyler Denniston, Xin Dong, Younes El Idrissi Yazami, William Enright, Rohan Garg, Twinkle Jain, Samaneh Kazemi, Jay Kim, Gregory Kerr, Apoorve Mohan, Neil Resnik, Manuel Rodríguez Pascual, Artem Y. Polyakov, Michael Rieker, Praveen S. Solanki, Ana-Maria Visan

If you want to hear more about checkpointing, consider the upcoming **SuperCheck'21** conference: Feb. 4-5, 2021
First Int. Symp. on Checkpointing for Supercomputing

*This work was partially supported by National Science Foundation Grant OAC-1740218 and grants from Intel Corporation, and Raytheon Technologies.