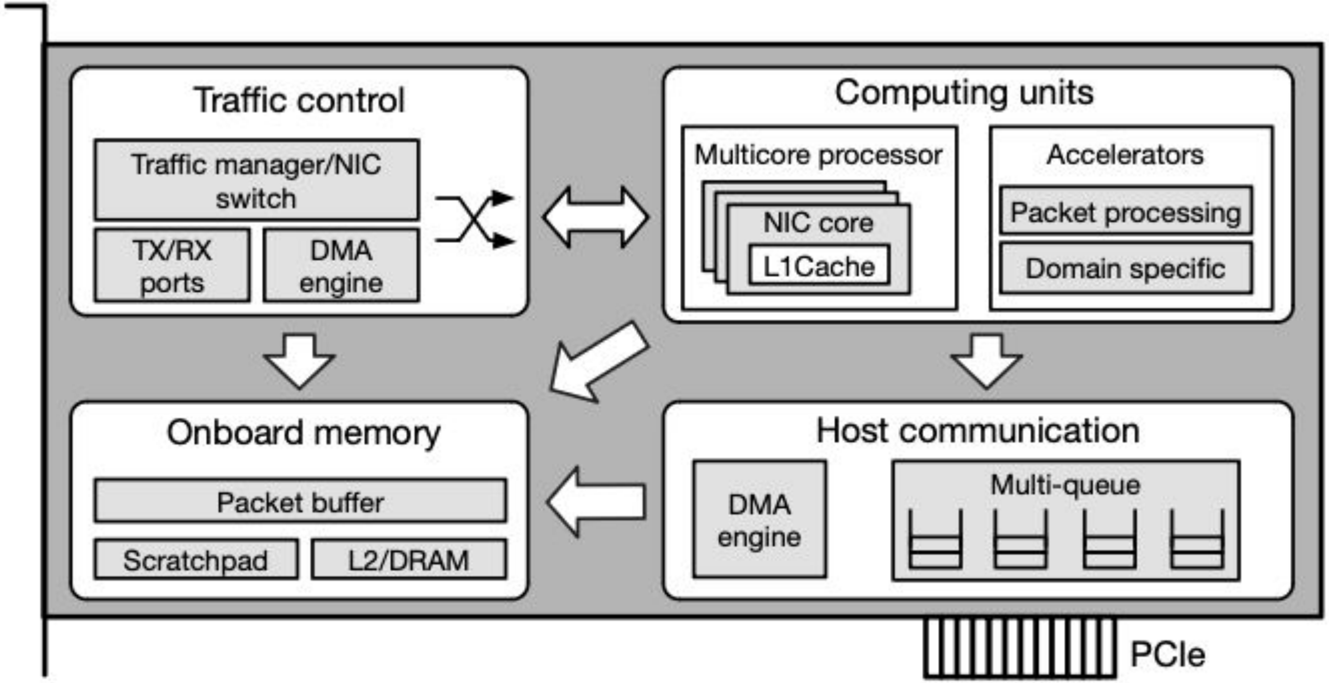
iPipe

Offloading Distributed Applications onto SmartNICs

What is a SmartNIC?


Traffic control
Traffic manager/NIC switch
TX/RX ports
DMA engine
Computing units
Multicore processor
NIC core
L1Cache
Accelerators
Packet processing
Domain specific
Onboard memory
Packet buffer
Scratchpad
L2/DRAM
Host communication
DMA engine
Multi-queue
PCIe

The problem

How do we utilize the **heterogeneous** compute capacity of a SmartNIC **efficiently**

- while exposing **simple programming** abstractions
- in the presence of **varying network traffic**
- for **complex** distributed applications?

The Characterization Study

Observation	Design Implication
Packet size distribution impacts the availability of computing cycles.	Monitor request sizes to adaptively offload tasks to the SmartNIC core.
Wimpy processor on SmartNIC provides opportunity for cheap parallelism.	Offload processes with low IPC without affecting packet processing and execution latency.
SmartNICs perform poorly on tasks with working set exceeding the capacity of L2 cache.	Identify such tasks and migrate/schedule it on the host.
Accelerators are critical resources on a SmartNIC.	Tasks that benefit from domain specific acceleration must be batched and executed on the SmartNIC.

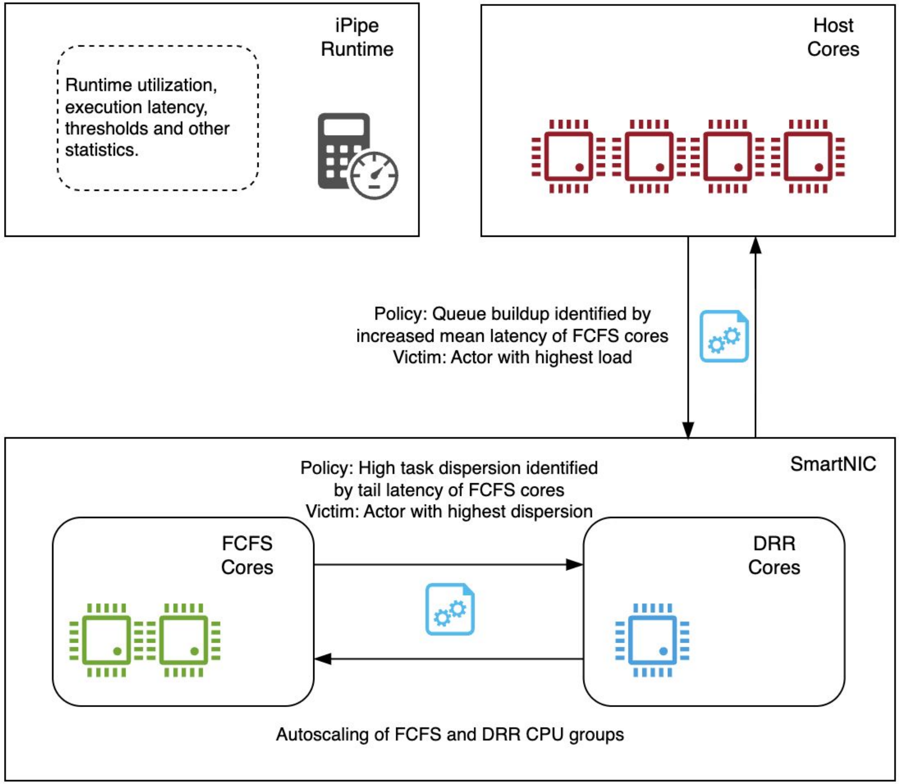
iPipe’s Design

- Actor Programming Model
- Actor Scheduler
- Distributed Memory Objects

Actor Programming Model

- Support compute heterogeneity and hardware parallelism.
- Actors have well-defined state and can be migrated between host and the NIC dynamically.
- Operations:
  - Initialize private state.
  - Trigger execution handler upon receiving a message.
  - Send messages to other actors.

Actor Scheduler

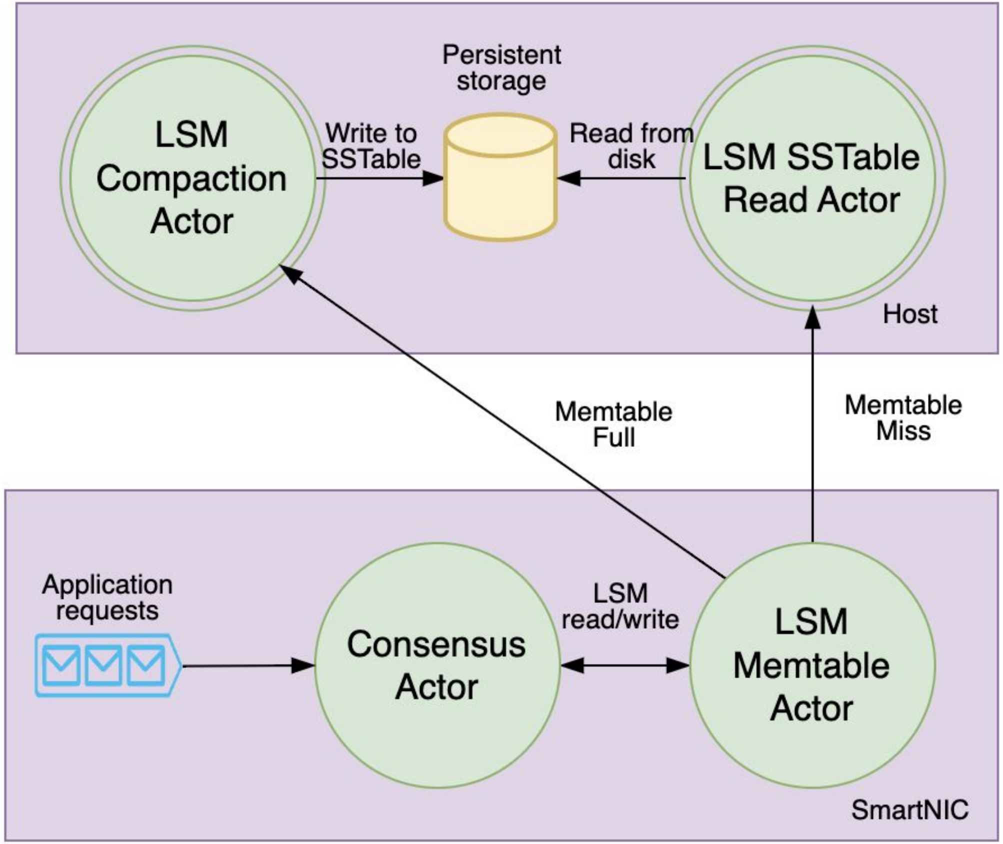


Distributed Memory Objects

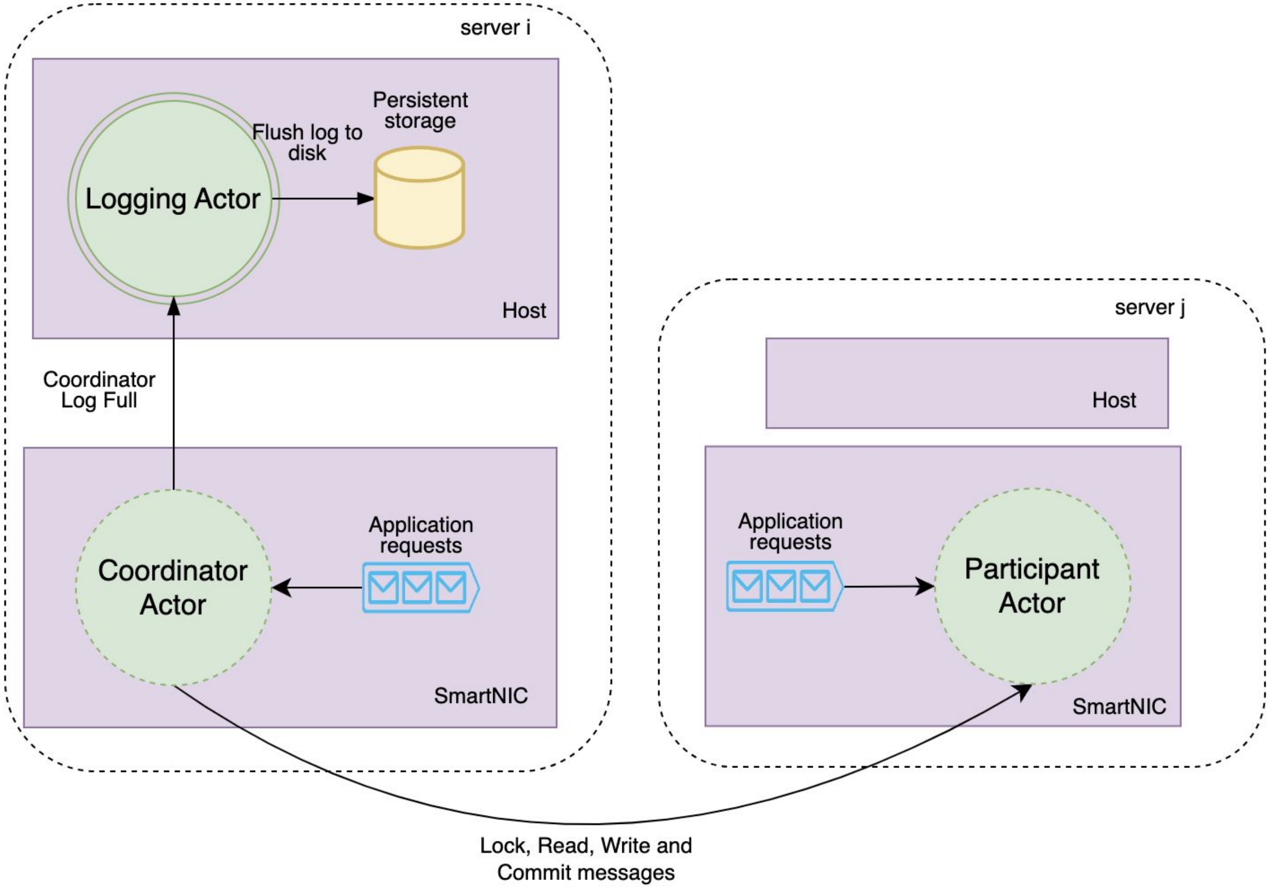
- Abstraction enabling migration of actor state between host and NIC.
- DMO has only one copy tracked by object tables on both sides.
- DMO moved along with actor upon migration.
- Object ID used instead of pointers to transparently migrate objects.

Applications developed using iPipe

Replicated KV Store



Distributed Transactions



Real Time Analytics

- Three Actors: Filter, Counter and Ranker
- Filter worker discards uninteresting rows.
- Counter uses a sliding window and periodically emits tuples to ranker.
- Ranker sorts incoming tuples and emits top-n data.
- Ranker has high load and is migrated to host when network load is high.

Evaluation

Dimension	Results
Host CPU savings when offloading to SmartNIC.	RKV - 3.1, DT - 2.6, RTA - 2.5 cores.
Execution latency savings when offloading to SmartNIC.	RKV - 5.4μs, DT - 28.0μs, RTA - 12.5μs.
iPipe scheduler performance for low and high dispersion workloads.	Low dispersion: Similar to FCFS, better than DRR, High dispersion: Better than both FCFS and DRR.
iPipe Framework overheads (DMO indirection and scheduler overheads).	iPipe consumes around 12% more CPU cycles for RKV leader.
Comparison with Floem, a similar SmartNIC programming framework.	Significant improvement.

Thank you!

Questions?