E3: Energy-Efficient Microservices on SmartNIC-Accelerated Servers

Ming Liu, Simon Peter, Arvind Krishnamurthy, and Phitchaya Mangpo Phothilimthana
Trend 1: Energy-efficiency now major factor in DC design

✓ Data centers are major energy consumer
✓ Within, servers (CPUs) consume the most energy

Source: United States Data Center Energy Usage Report
Trend 2: the rise of microservices

Online shopping service

- Account Service
- Product Service
- Cart Service
- Order Service

IoT data analytics

- Data 61
- MQTT
- Rule Engine

Netflix

✓ Fine-grained (low memory footprint)
✓ Communication intensive (low latency)

Twitter
Trend 3: recent adoption of SoC SmartNICs in servers

✓ Wimpy multicore processor on NIC
✓ Consume < 30W power
✓ <= 16GB DRAM
✓ Good fit for microservices

Marvell LiquidIO II

Broadcom Stingray

Mellanox Bluefield

Netronome Agilio
Challenges of integrating SmartNICs

✓ How to route and load balance requests?
✓ How to place microservices on a heterogeneous system?
✓ How to avoid SmartNIC overloading?
E3: a microservice execution platform for SmartNIC-accelerated servers with the goal of achieving better energy efficiency at minimum latency cost.
E3: a microservice execution platform for SmartNIC-accelerated servers with the goal of achieving better energy efficiency at minimum latency cost.

**Technique 1:** ECMP-based load balancing at ToR switch, to balance requests among NICs
E3: a microservice execution platform for SmartNIC-accelerated servers with the goal of achieving better energy efficiency at minimum latency cost.

Technique 2: communication-aware microservice placement
E3: a microservice execution platform for SmartNIC-accelerated servers with the goal of achieving better energy efficiency at minimum latency cost.

*Technique 3*: load-aware data-plane orchestrator to avoid SmartNIC overload
Evaluation results

✓ A cluster of commodity servers + LiquidIOII SmartNIC
✓ Compare four cluster setups
  - Homogeneous beefy cluster
  - Homogeneous wimpy cluster
  - Heterogenous beefy+wimpy cluster
  - Super-beefy server cluster
✓ Three applications
  - network functions
  - real time data analytics
  - IoT hub
✓ E3 achieves up to 3X energy efficiency vs. 2nd best solution
Thursday, July 11

Track 1

11:15am ~ 12:35pm