Specialized μservers for the data center

MICROSERVER n. An application-specific server implemented as a small appliance

1 Motivation – Optimizing scale-out workloads in the datacenter

Main-memory key-value stores deployed in most datacenters
- Network-heavy: little computation, high concurrency
- Random access: multi-level cache poorly utilized
Modern CPUs and architectures are optimized for the opposite!

2 Implementation – From single instance to distributed service

We use Xilinx VC709s for prototyping:
- 4x 10Gbps interfaces,
- 8GB memory

3 Demo use-cases

a) Benchmarking with memaslap

Memaslap is a C++ benchmarking client for memcached.
It offers many parameters for customized load generation:
- Fraction of set and get commands (write vs. read)
- Different key and value sizes
- Number of clients simulated on the machine

We demonstrate the functionality of our key-value store, and measure both throughput and round trip times.
A single machine is not able to saturate the microservers (see results below).

b) Integration with a database

OLTP-bench is a collection of database benchmarks written in Java.
We use the Twitter benchmark to demonstrate the benefits of caching results in main memory → increased throughput.

procedure get-tweets-of-user(userID)
if not found then
    run query on DB
set(userID, tweets) on microserver

4 Performance of a node compared to related work

Performance (MOPS)

<table>
<thead>
<tr>
<th>Our system</th>
<th>FPGA Off-the-shelf</th>
<th>ARM Off-the-shelf</th>
<th>Atom Server</th>
<th>Atom ARMv8 @2.4GHz</th>
<th>Atom ARMv8 @2.4GHz</th>
</tr>
</thead>
<tbody>
<tr>
<td>10Gbps line-rate</td>
<td>11.5</td>
<td>4.3</td>
<td>2.8</td>
<td>2.6</td>
<td>9.4</td>
</tr>
</tbody>
</table>

Power (Watts)

<table>
<thead>
<tr>
<th>Our system</th>
<th>FPGA Off-the-shelf</th>
<th>ARM Off-the-shelf</th>
<th>Atom Server</th>
<th>Atom ARMv8 @2.4GHz</th>
<th>Atom ARMv8 @2.4GHz</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>25</td>
<td>17.5</td>
<td>66</td>
<td>75</td>
<td>66</td>
</tr>
</tbody>
</table>

Acknowledgements: This work is funded in part through the Microsoft Joint Research Center MSR-ETHZ-EPFL and an equipment donation from Xilinx.