Power Optimization in Packet Switches through Reinforcement Learning

Advisor: Professor Daniel C. O'Neill

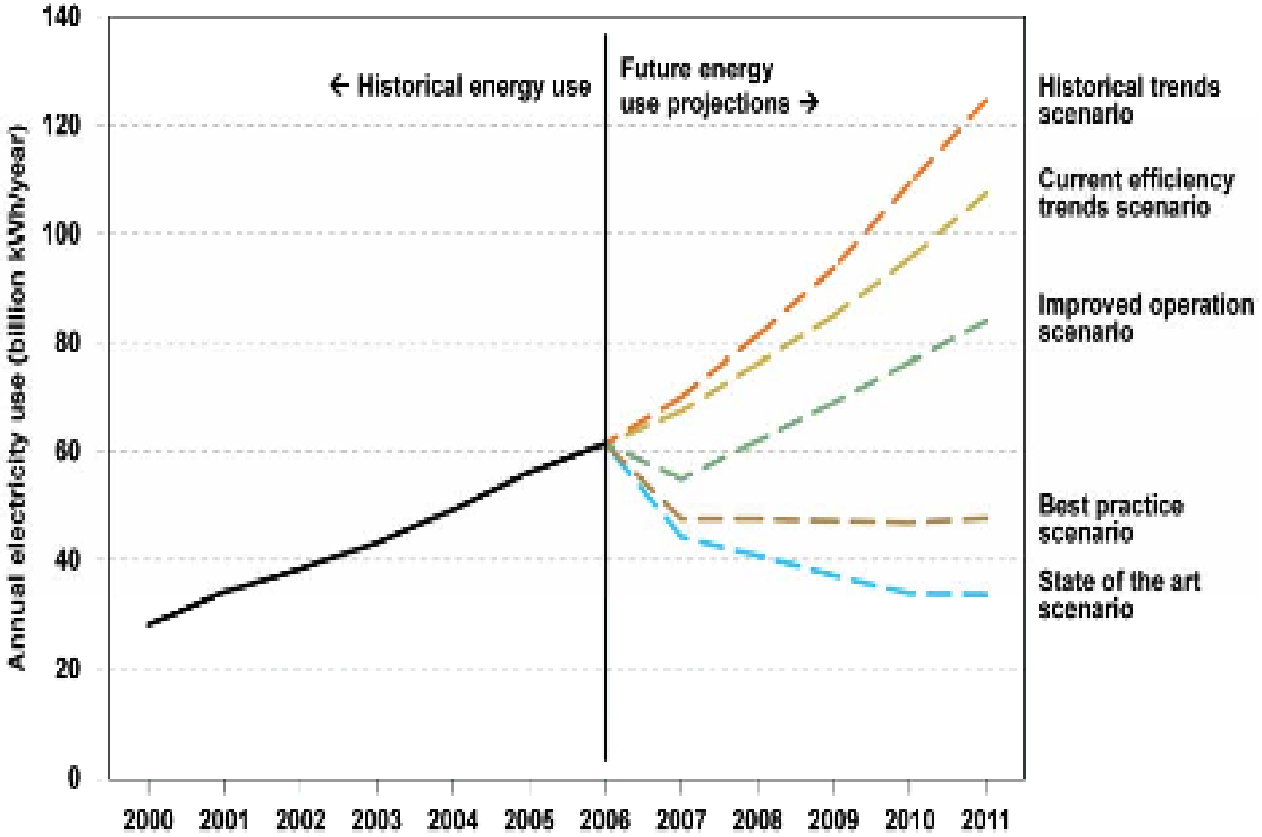
Divya Elayakumar, Stanford University

Agenda

- Motivation
- Switch Model
- Reinforcement Learning
- Power Estimation
- Simulation
- Results

Motivation

US Environmental Protection Agency(EPA) report of electricity use by data centers



Motivation

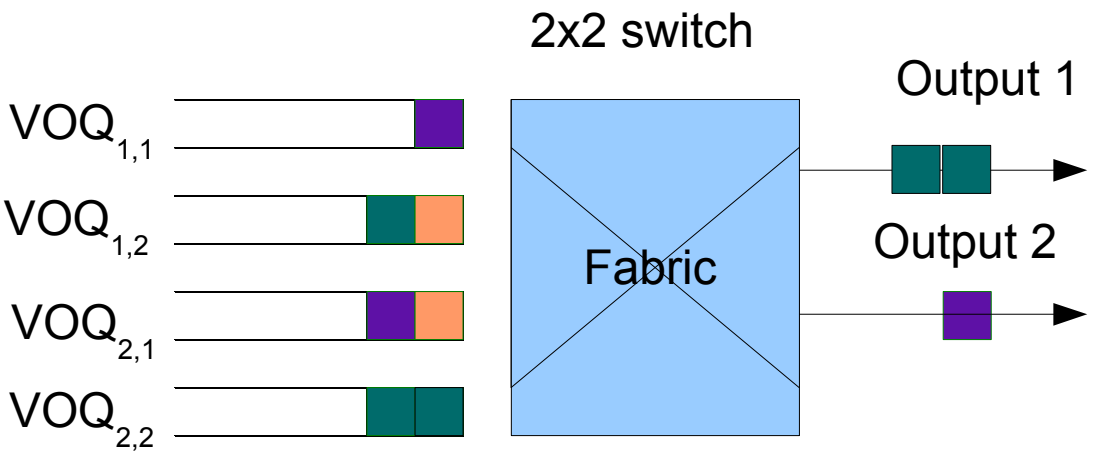
- Recent study by Professor Jonathan Koomey, Stanford shows that electricity used by data centers worldwide has increased by 56% from 2005 to 2010.
- Electricity used by global data centers in 2010 accounts for 1.1-1.5% of total electricity use. In US , it is about 1.7-2.2%.
- Several techniques are being developed and applied at system and circuit level to reduce the power consumption of the servers.
- This has controlled the trend in electricity use that would have doubled otherwise.

Power Consumption of Switches

- High power consumption due to increase in network speed,traffic demands and circuit density.
- When large number of packets are received and routed, the power required to operate the switches increases significantly.
- Power consumption is limited by the physical constraints within the switch, customer and industry standards.
- Scheduling algorithm that determines the packets to service must also balance throughput and power consumption.
- This necessitates power and delay aware scheduling algorithms for packet switches.

Switch Model

- N X N switch N line card processing units(LCP), N^2 virtual output queues(VOQ) and crossbar switching fabric.
- Crossbar constraint - no output simultaneously receives packets from more than one input VOQ and no input simultaneously transfers packets to more than one output.



Definitions

- Backlog state of the switch a time t is denoted by backlog vector $x^t = [x^t_{1,} ... x^t_{q.} .. x^t_Q]$, $Q = N^2$
- Packets are serviced at the beginning of time t denoted by service vector $s^t = [s^t_{1,} ... s^t_{q.} .. s^t_Q]$
- Packets arrive in the VOQs at the end of time t denoted by arrival vector $a^t = [a^t_{1,} ... a^t_{q.} .. a^t_Q]$
- Backlog state of the switch a time t+1 is $x^{t+1} = x^t - s^t + a^t$

Cost Functions

- Power cost in servicing packets at time t is quadratic and denoted by $(s^t)^T R\, s^t$, R is the power cost matrix.
- Backlog cost at time t is quadratic and denoted by $(x^t)^T Q\, x^t$, Q is the identity matrix.
- Total cost = $\Sigma \quad ((s^t)^T R\, s^t + (x^t)^T Q\, x^t)$
- We need a scheduling algorithm that minimizes the average cost over a period of time.
- The algorithm must continuously learn from the load in the switches and minimize the average cost. It can defer or service a packet at time t to minimize the average cost.

Reinforcement Learning(RL)

- The agent interacts with environment and learns through consequences of actions-reward.
- An environment is represented by Markov decision process (MDP) consisting of four tuple (S,A,R,P), where S is state space, A is action space, R is reward model and P stands for transition model.
- At time t the environment is in state $x_t \in S$, the agent takes an action a_t A and receives a scalar reward r_t .
- The agent transitions to the next state x_{t+1}.
- The goal of RL agent, is to find a near optimal policy in a such a way that if the agent follows that policy from any given state then its sum of future reward is maximum .

Q Learning

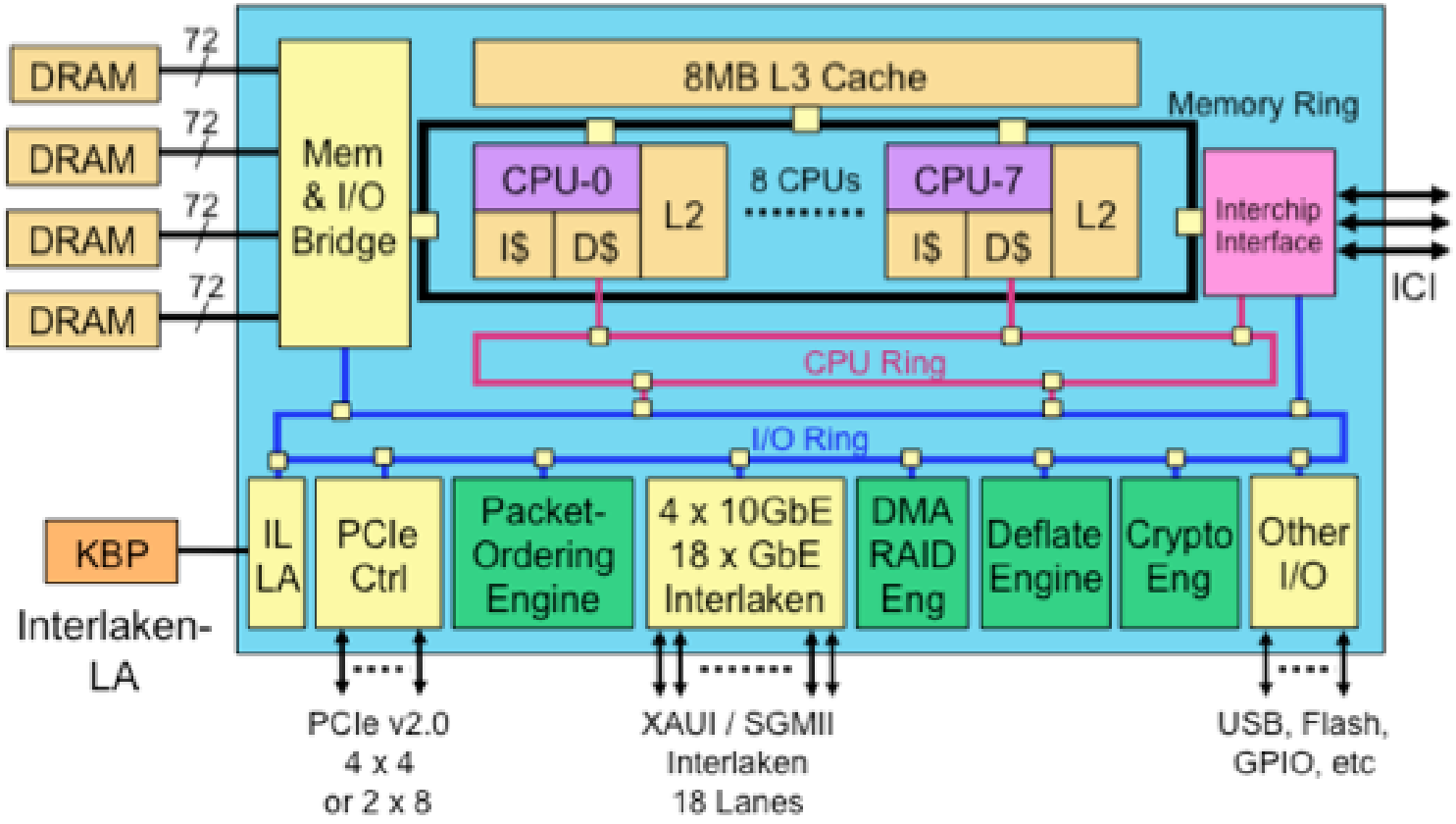
- Simple, sample based , online and incremental learning method
- Does not require access to the model of the environment unlike dynamic programming.
- Q value : $Q(x,a)=E[r_0 + \gamma r_1+\gamma^2 r_2+\cdots]$, γ is discount factor.
- The agent takes greedy actions mostly but also does exploration.
- Q learning update is as follows,

$$Q_{t+1}(x_t, a_t) = Q_t(x_t, a_t)+\alpha_t [r_t + \max_{a'} Q_t(x_{t+1}, a') - Q_t(x_t, a_t)]$$

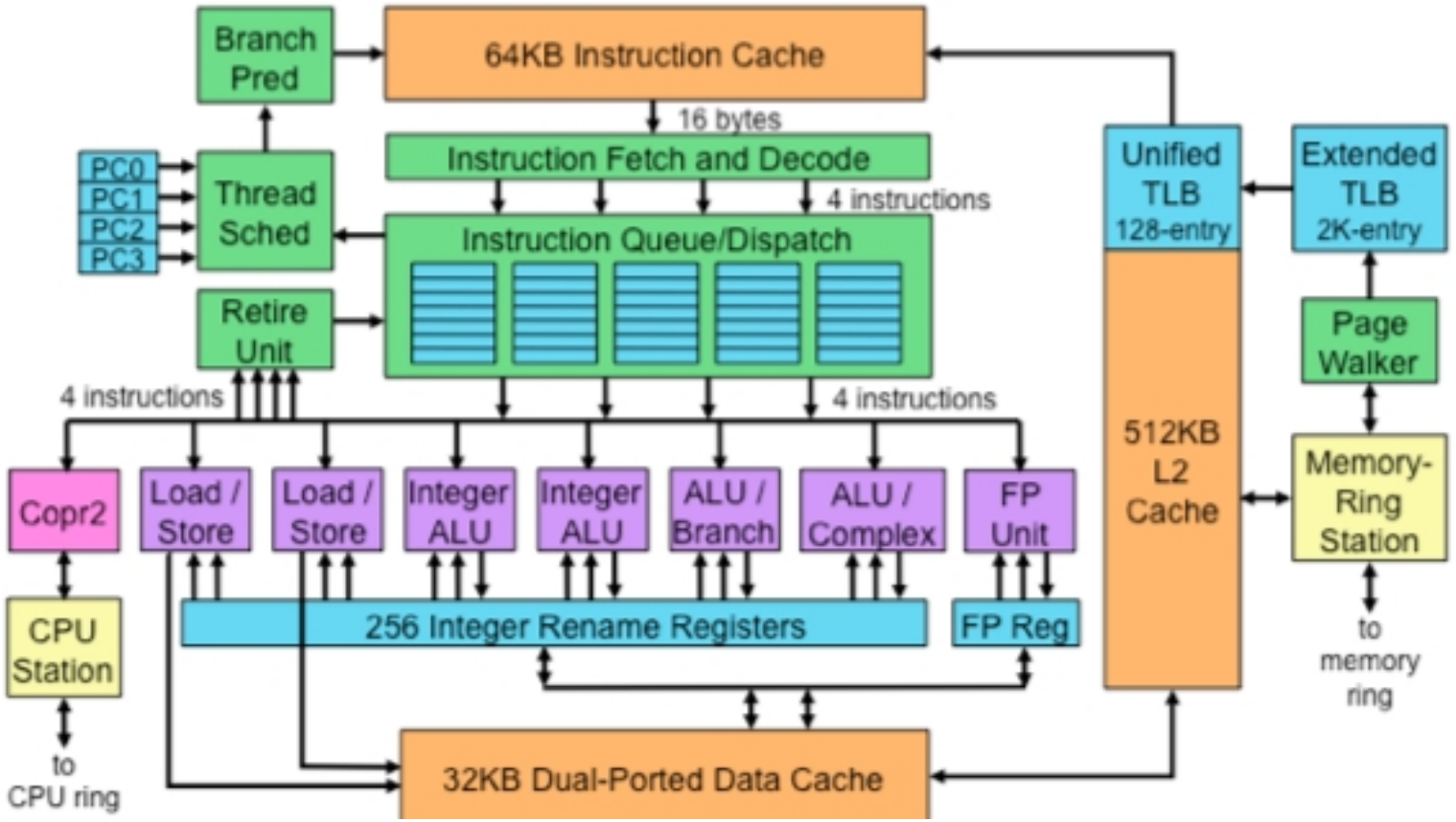
Network Processors

- Power cost : $(s^t)^T R\, s^t$. How to find the power cost matrix R ?
- Network processors are programmable chips like general purpose processors but they are optimized for packet processing.
- They are programmable as a CPU but as fast as an ASIC.
- They perform functions such as pattern matching,key lookups, computation,bit manipulation,queue management..
- Used in switches, routers, firewalls, VOIP bridges..
- Approach: Analyze a state of art network processor to arrive at the values for cost matrix R.

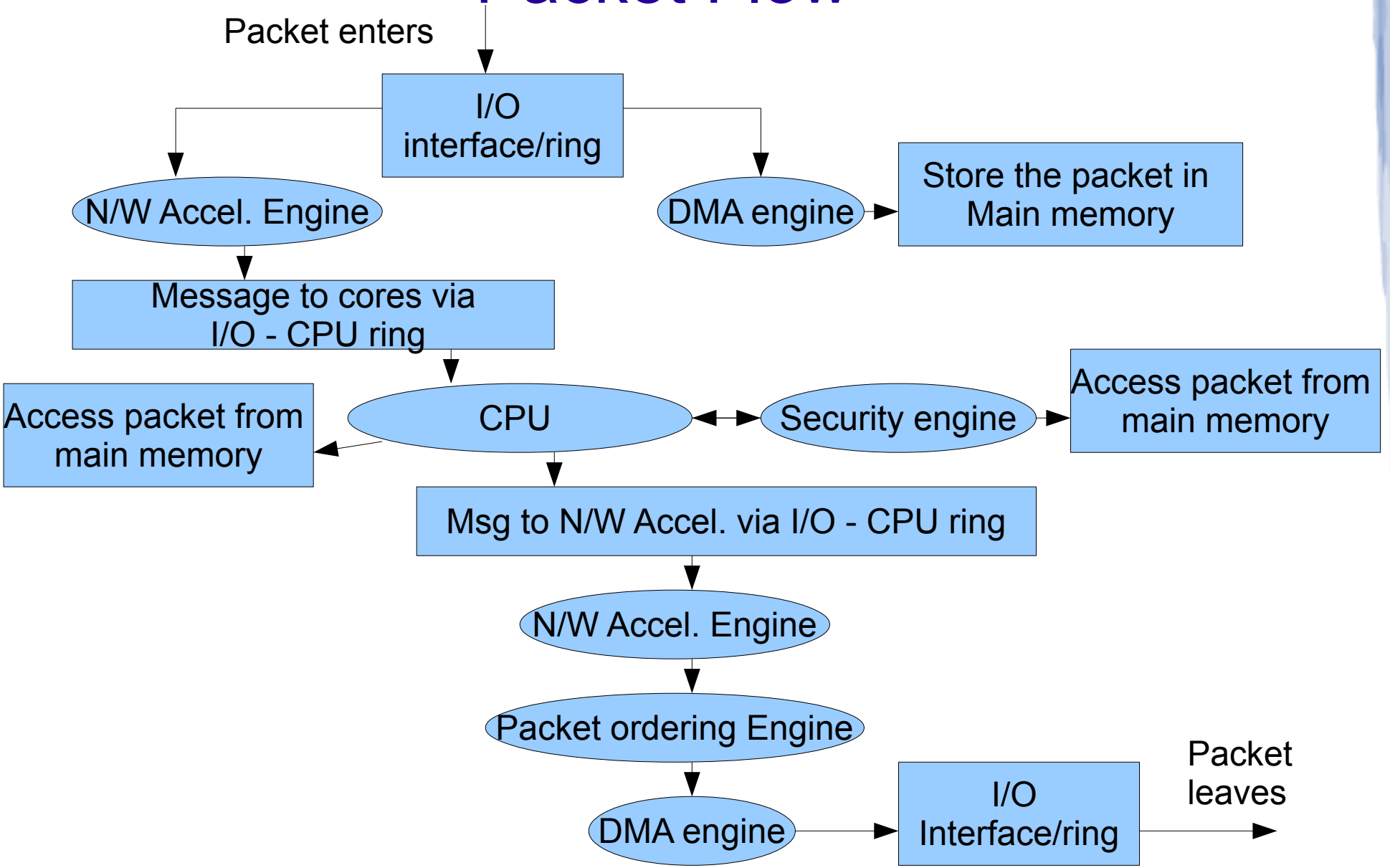
Netlogic - XLP832



CPU


Branch Pred
64KB Instruction Cache
16 bytes
Instruction Fetch and Decode
4 instructions
PC0
PC1
PC2
PC3
Thread Sched
Instruction Queue/Dispatch
Retire Unit
4 instructions
4 instructions
Copr2
Load / Store
Load / Store
Integer ALU
Integer ALU
ALU / Branch
ALU / Complex
FP Unit
CPU Station
256 Integer Rename Registers
FP Reg
to CPU ring
32KB Dual-Ported Data Cache
Unified TLB 128-entry
512KB L2 Cache
Extended TLB 2K-entry
Page Walker
Memory-Ring Station
to memory ring

Packet Flow


Packet enters
I/O interface/ring
N/W Accel. Engine
DMA engine
Store the packet in Main memory
Message to cores via I/O - CPU ring
Access packet from main memory
CPU
Security engine
Access packet from main memory
Msg to N/W Accel. via I/O - CPU ring
N/W Accel. Engine
Packet ordering Engine
DMA engine
I/O Interface/ring
Packet leaves

Dynamic Power Calculation

- 3 types of packets - Ack(48B) ,Data(512B), Video(1500B)
- **CPU**
  4 way multi threaded 64 bits MIPS core
  4W/1Ghz, Power/instruction, mostly arithmetic,loads/stores

Functional units	Instructions / packet	Power /packet(W)
ACK	18	1.608E-008
Data	93	8.308E-008
Video	18	1.608E-008

- **Security engine**
  Similar to Netlogic's AU1550 security network processor
  500mW/400Mhz,40Gbps of encryption/decryption

Packet type	Bytes to process	Time to process/packet(ns)	Power/packet (nW)
Video	1452	290.4	145.2

Dynamic Power Calculation

- **Caches and memory**
  Power/access or power/byte

	% of total power	dynamic power (W)	power (W)/ access	power (mW)/byte accessed	accesses in 100 cycles	bytes accessed in 100 cycles	bytes/acces s
I cache	10.5%	14	0.14	8.75000	100	1600	16
dcache	3.0%	4	0.025	3.12500	160	1280	8
L2	4.2%	5.6	0.28	8.75000	20	640	32
L3	0.6%	0.8	0.8	6.25000	1	128	128

	Total power (mW)/ access	Read power (mW)/acces s	Activate power(mW)/a ccess	Read power (mW)/byte accessed	bytes/acces s
Main memory	324	50	274	3.12500	16

- **Network Accelerator/Packet Ordering Engine**
  Many in-order 32 bit MIPS cores
  100mW/1Ghz
- **I/O ring**
  Transfers 64B/clk cycle.
  Bus width:256, data rate: 50 Gbps,8 lanes at 6.25Gbps each

Total Dynamic Power

- Design units that have same power consumption for all packets such as Network accelerator engine,packet ordering engine, rings are not included in the calculation.
- Power consumption of DMA engine and memory controller are assumed negligible.
- Memory accesses dominate the power consumption.

Packet type	Network I/O bus/ring(W)	Packet storage in memory (I/O ring->Mem ctl & I/O bridge->main memory)(W)	Core processing (W)	Core accesses memory(L1-L2-L3-main memory) memory ring(W)	Security engine (encryption/ decryption) (W)	Security engine accesses the memory(W)	Access memory to transfer packet out (I/O ring->Mem ctl & I/O bridge->main memory)(W)	Network I/O bus(W)	Total power consumption/ packet (W)
ACK	3.2242E-010	0.424	1.608E-008	1.144	0	0	0.424	3.2242E-010	1.992000017
Data	3.4391E-009	1.874	8.308E-008	5.349	0	0	1.874	3.4391E-009	9.09700009
Video	1.0075E-008	4.9615	1.608E-008	1.144	1.452E-007	4.9615	4.9615	1.0075E-008	16.02850018

Simulation – 2 x 2 Switch

- 2 input ports each with 2 VOQs and 2 output ports
- Each VOQ can have at most 2 packets.No more than 2 packets can arrive at any time in a VOQ.
- Switch follows crossbar constraint and can service at most 2 packets from each port at a time.
- Supports 3 types of packets on each VOQ, but there can be be only packets of same type in a VOQ at a time.
- MDP M= (S,A,R,P) that represents the switch is developed. There are 2401 states, each can have 25 actions.
- Transition probabilities that represent P are developed based on the linear equation $x^{t+1} = x^t - s^t + a^t$
- Reward is represented by the total cost(power and backlog).
- Simulations are done in Matlab

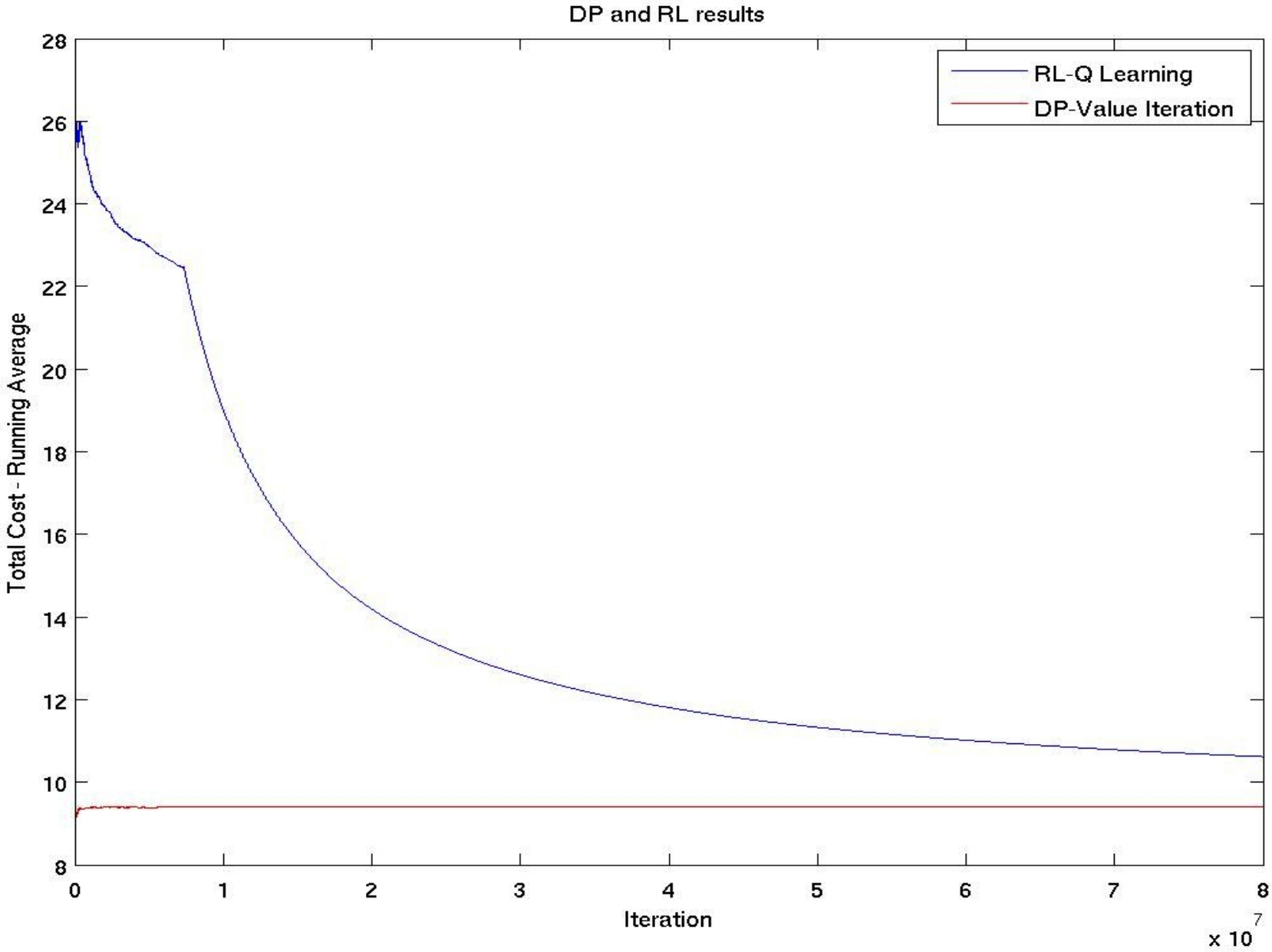
Dynamic Programming(DP)

- Applied to validate average reward from Q-learning.
- Unlike RL, DP requires knowledge of the model - transition probabilities,P and reward,R for a given state and action.
- RL needs to construct sample trajectories and learn from these samples while DP does not need to do this. P and R are not available in real world problems.
- Value iteration
  - For a state-action pairs, given R and P
  Q-value update for iteration t is
  $Q_{t+1}(x, a) = \sum_{x'} P(x'|x,a) \, (R(x,a,x') + \max_{a'} Q_t(x', a'))$
  - It synchronously updates Q values for all state-action pairs in a iteration.
  - It repeats this until the Q values stop changing beyond a threshold.

Interesting observations

- Varying step size
  - Large step size initially(exploration phase) and small step size when the algorithm is about to converge(exploitation)
- Exploration policy
  - Greedy mostly, but also sometimes take random actions or actions that have not been hit frequently.
- Stochastic environment
  - There can be any arrival pattern(not exceeding 2 packets in a port) at a time.
  - With large state space and uniform probabilities for all the arrival patterns, it is harder to find optimal solution.

Results


DP and RL results
RL-Q Learning
DP-Value Iteration
Total Cost - Running Average
Iteration
x 10^7
28
26
24
22
20
18
16
14
12
10
8
0
1
2
3
4
5
6
7
8

References

- Lykomidis Mastroleon, Daniel O' Neill, Benjamin Yolken and Nicholas Bambos ,“Power and Delay Aware Management of Packet Switches” , IEEE Transactions on Computers, October 2011.
- Jonathan G. Koomey , “ Growth in Data center electricity use 2005 t0 2010”, August 2011.
- XLP832 Processor,Product brief, Netlogic Microsystems. http://www.netlogicmicro.com/Products/ProductBriefs/MultiCore/XLP832.htm
- “Netlogic broadens XLP family, Multithreading and four way issue with one to eight CPU cores” , Microprocessor Report ,July 2010.
- “Specifying Power consumption ” ,Freescale Semiconductors. Document no : AN2436
- “PowerPC MPC7455 I/O Power Evaluation” ,Freescale Semiconductors,Inc. .2004.
- “RMI Alchemy AU1550 Processor,security network processor”, Raza Microelectronics,Inc..
- US Environmental Protection Agency,“Report to Congress on Server and data center energy efficiency”, August 2007.
- Stranger in an ARM world ,Ingenic designs MIPS CPU for JZ4700 Mobile Processor”,Microprocessor Report, 2012.
- “Calculating memory system power for DDR SDRAM” Vol 10. Issue 2, Micron 2Q01.
- Reinforcement learning book - http://webdocs.cs.ualberta.ca/~sutton/book/ebook

Thank you!

My sincere thanks to

- Professor Daniel C. O'Neill,Department of EE, Stanford.
- Hamid Reza Maei,Department of EE, Stanford.
- Wayne Yamamoto, MIPS Technologies.

THANK YOU ALL!!