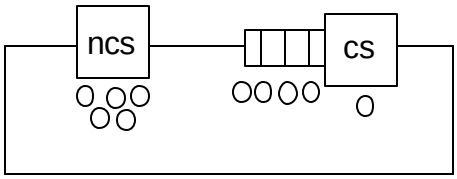
Two-Level Iterative Queuing Modeling of Software Contention

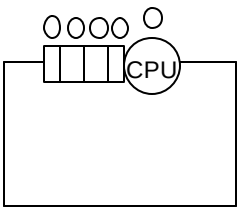
Daniel A. Menascé
Dept. of Computer Science
George Mason University
www.cs.gmu.edu/faculty/menasce.html

In Proc. 2002 IEEE MASCOTS Conference, Forth Worth, TX, October 2002.

Motivation

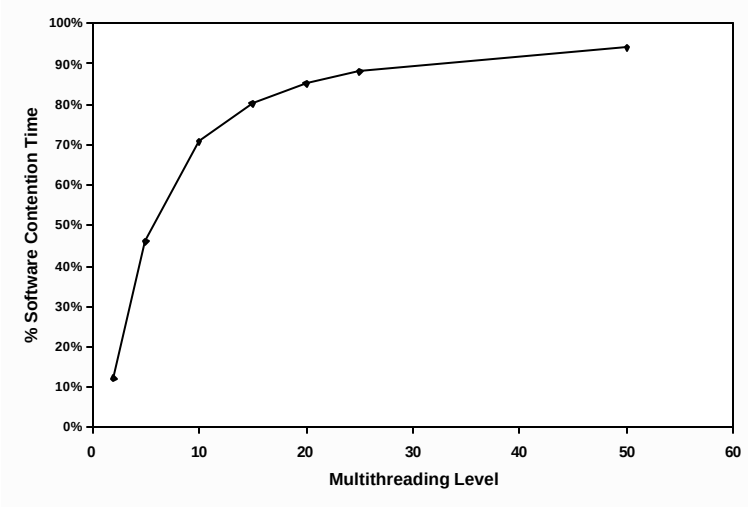


Software View



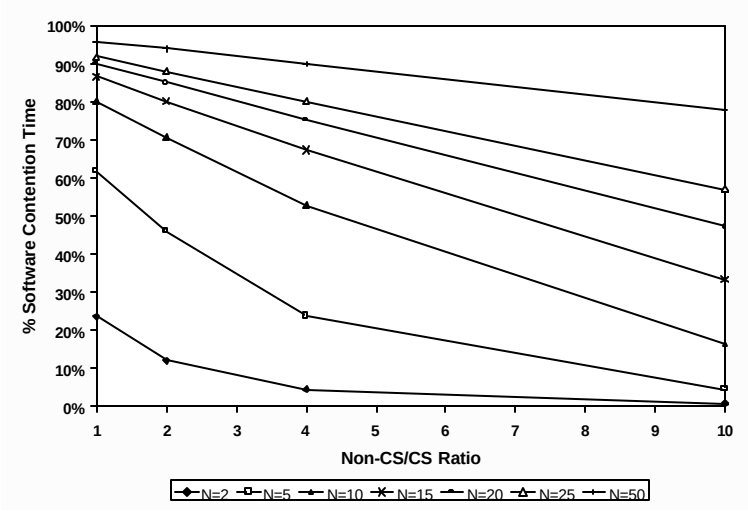
Hardware View

% Software Contention Time vs. Multithreading Level



3

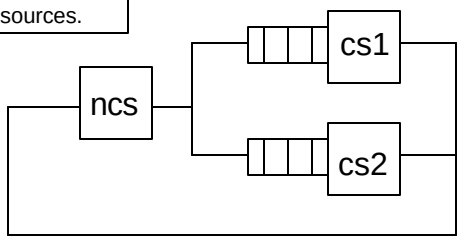
%Sofware Contention Time vs Non-CS/CS Ratio



4

If we just consider the SQN we are ignoring time spent at the software resources due to contention for hardware resources.

Approach

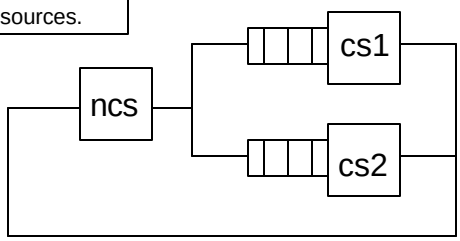


Software QN

5

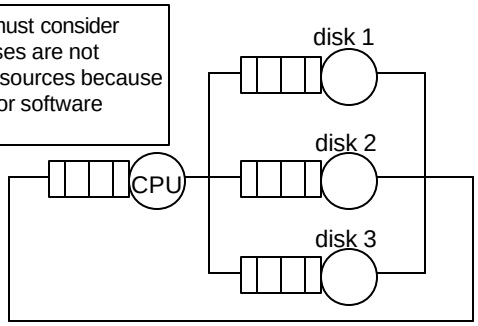
If we just consider the SQN we are ignoring time spent at the software resources due to contention for hardware resources.

Approach



Software QN

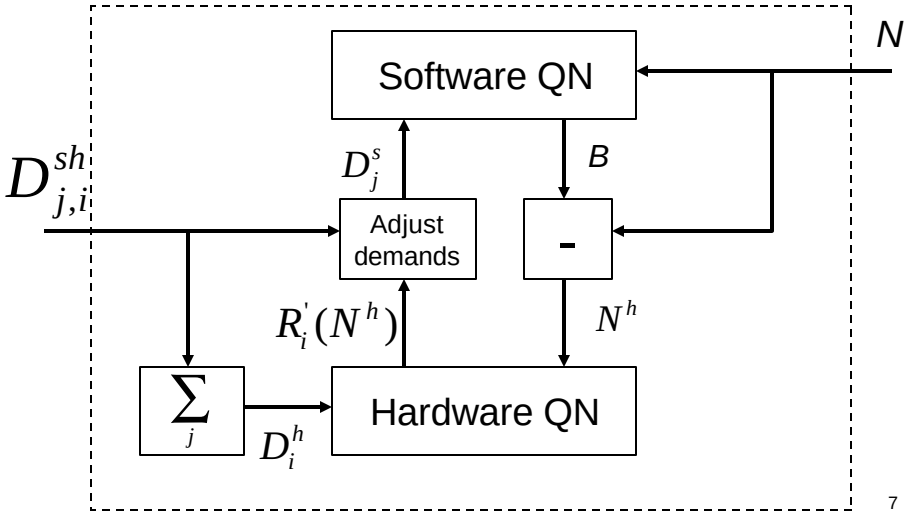
The HQN model must consider that some processes are not using hardware resources because they are blocked for software resources.



Hardware QN

6

SQN-HQN Scheme



7

Input Service Demands

	Software Modules			
Hardware Devices ( )	**NCS**	**CS 1**	**CS 2**	**Hardware Demands**
CPU	0.2000	0.0600	0.0808	0.3408
Disk 1	0.0560	0.0576	0.0000	0.1136
Disk 2	0.0360	0.0000	0.1212	0.1572
Disk 3	0.0360	0.0000	0.0000	0.0360
Software Demands	0.3280	0.1176	0.2020	

sum across a column

sum across a row

8

Service Demands

Hardware Devices (¯)	Software Modules NCS	CS 1	CS 2	Hardware Demands
CPU	0.2000	0.0600	0.0808	0.3408
Disk 1	0.0560	0.0576	0.0000	0.1136
Disk 2	0.0360	0.0000	0.1212	0.1572
Disk 3	0.0360	0.0000	0.0000	0.0360
Software Demands	0.3280	0.1176	0.2020	

D^{s}_{cs1}

(total service time of a software module)

$D^{sh}_{cs1,disk1}$

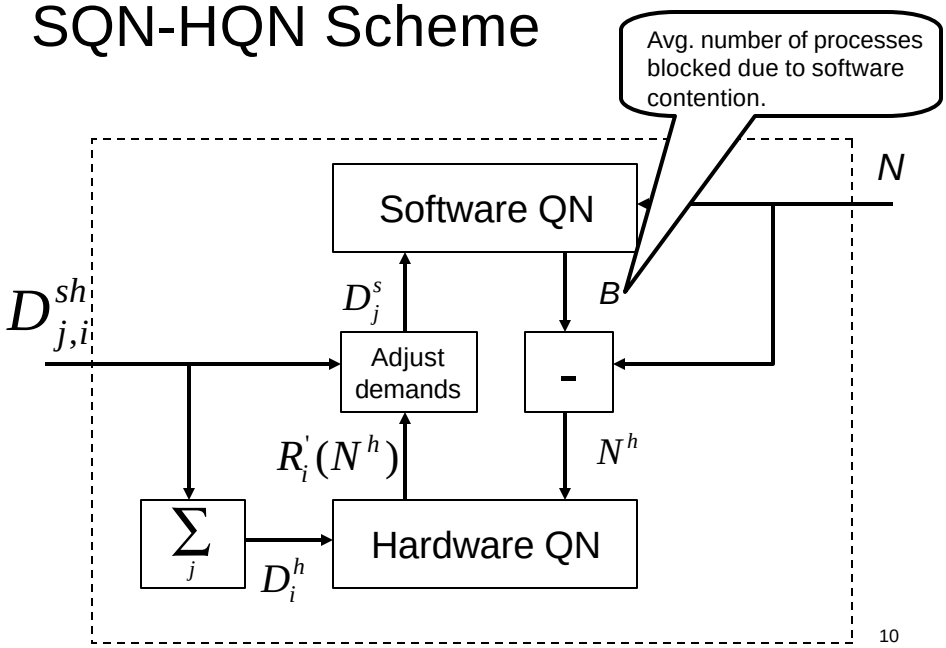
(total service time of a software module at a physical device)

D^{h}_{disk1}

(total service time of the application at a given device)

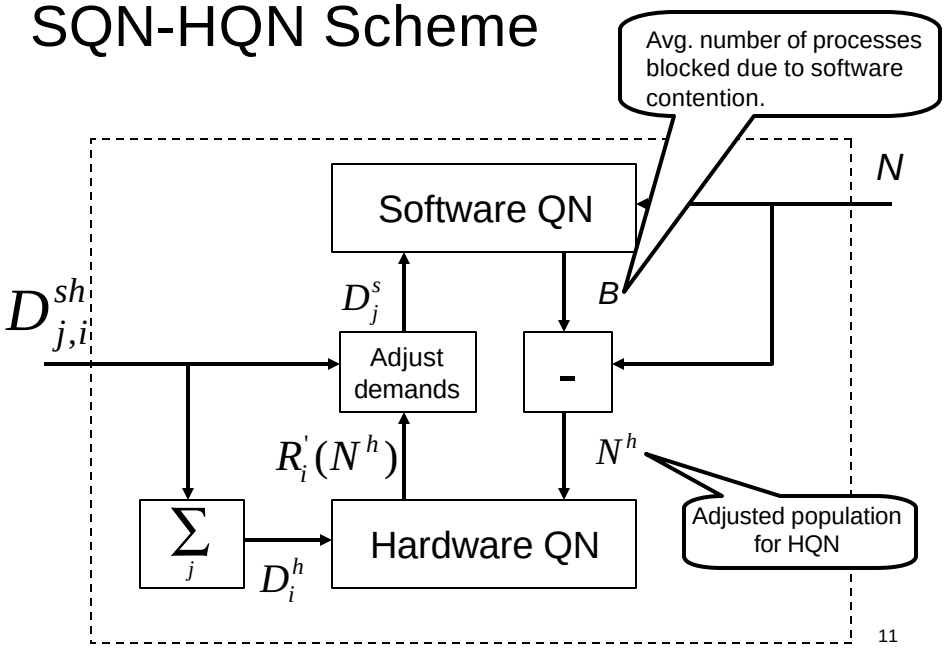
9

SQN-HQN Scheme



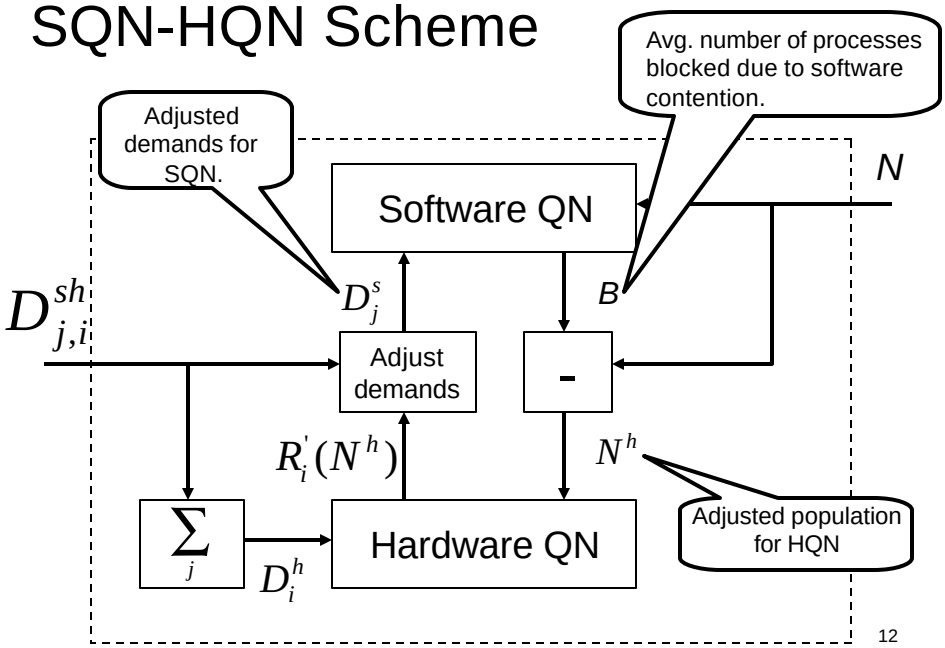
10

SQN-HQN Scheme



11

SQN-HQN Scheme



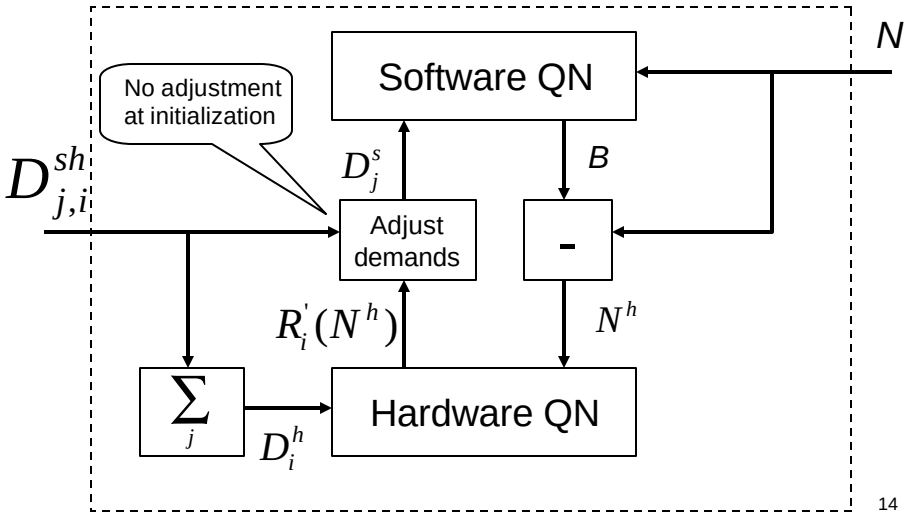
12

Basic Idea

- Iteration between solving the SQN and the HQN.
- Number B of processes blocked due to software contention computed through the SQN.
- Population at HQN is reduced by B.
- Service demands at SQN are adjusted to account for physical contention.

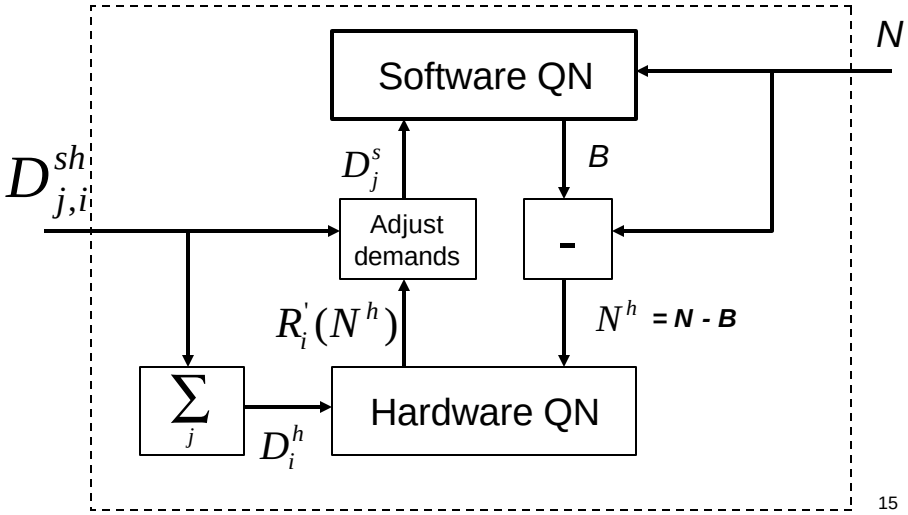
13

SQN-HQN: Initialization



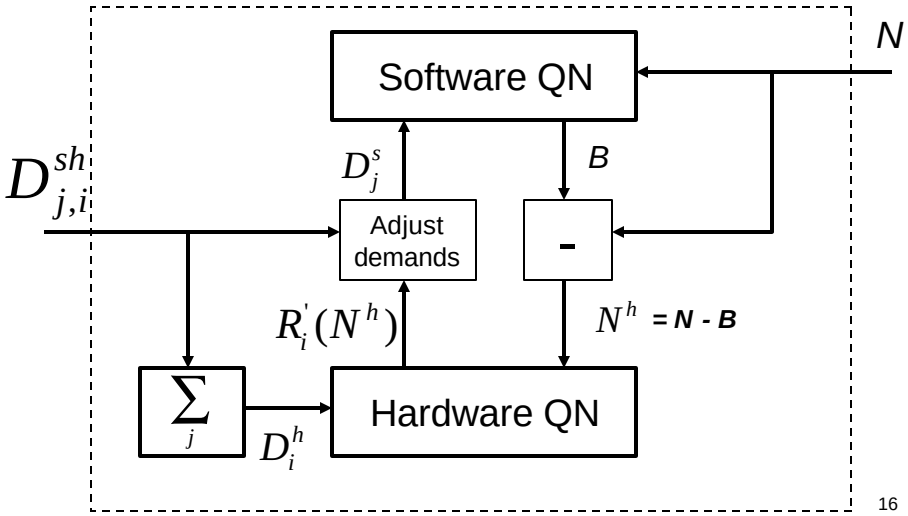
14

SQN-HQN: Solve SQN – No Hardware Contention



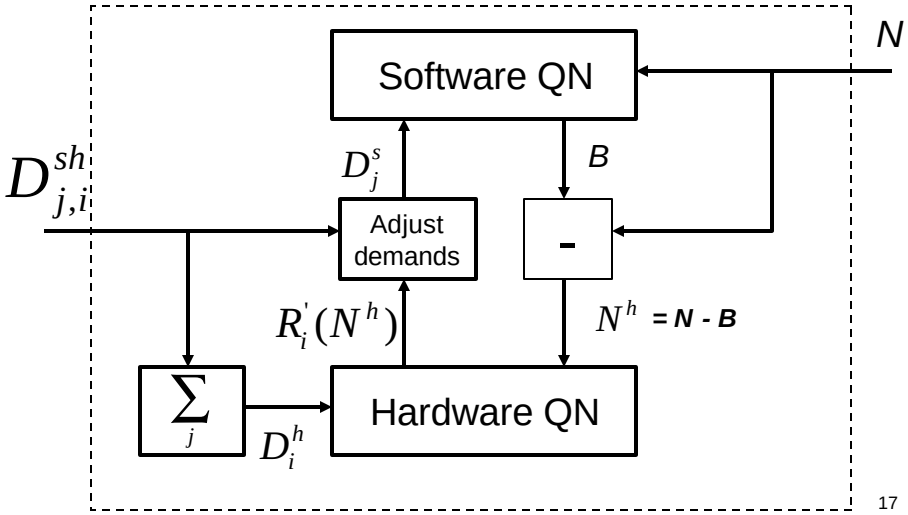
15

SQN-HQN: Solve HQN



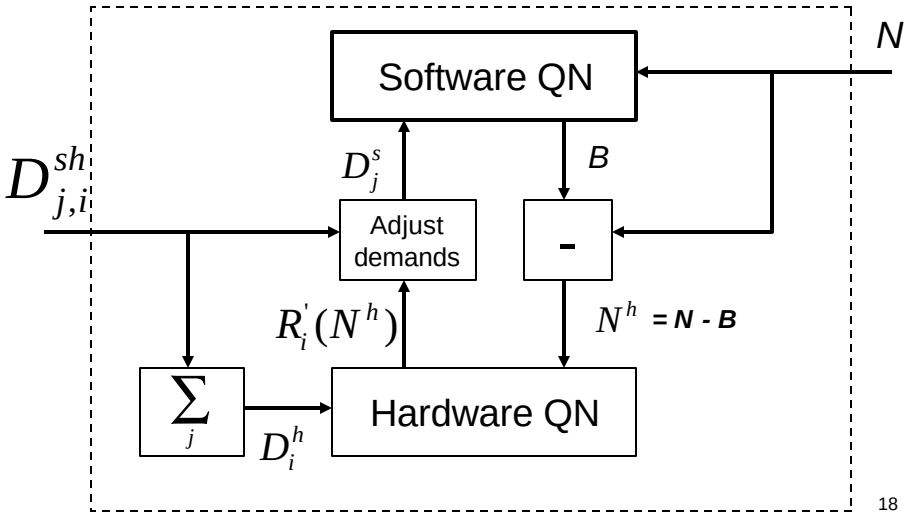
16

SQN-HQN: Adjust demands for SQN



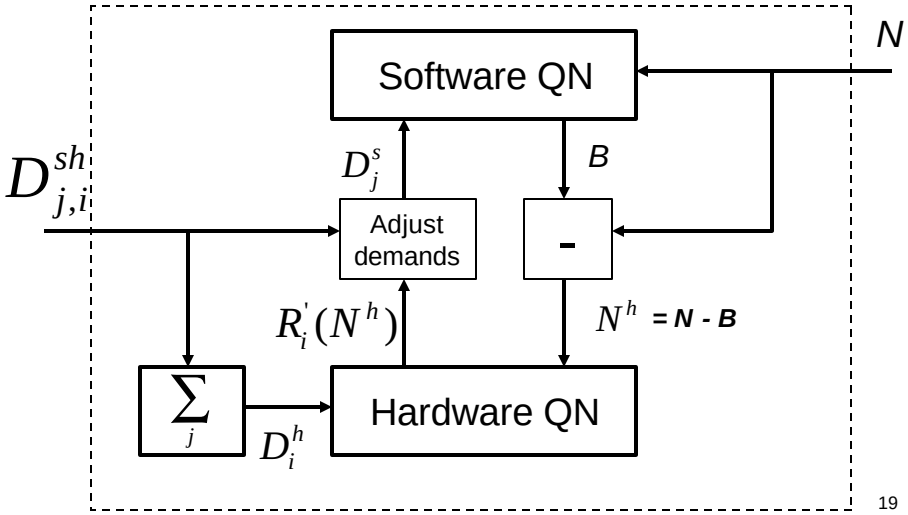
17

SQN-HQN: Solve SQN Again



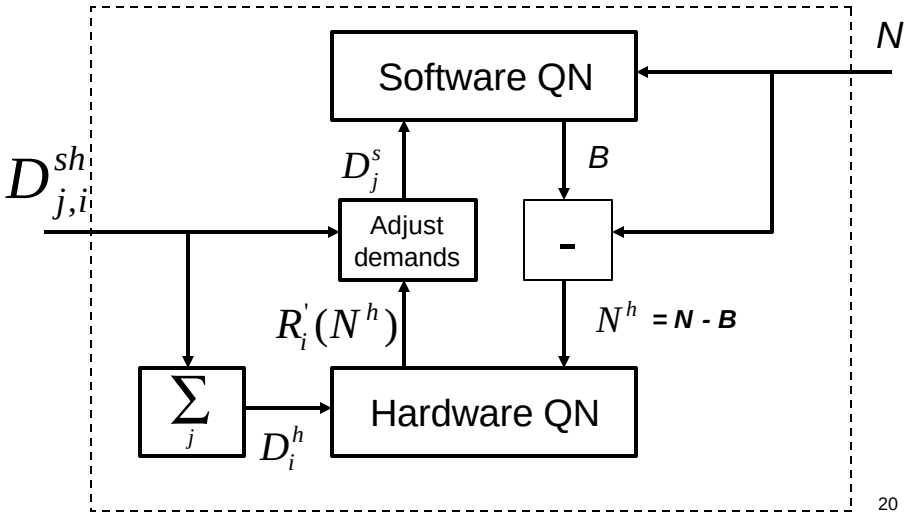
18

SQN-HQN: Solve HQN



19

SQN-HQN: Adjust demands for SQN



20

and so on …

Convergence is checked on absolute relative error on the number of blocked processes in the SQN.

Adjustment of SQN Demands

- Single class case:

residence time at device i.

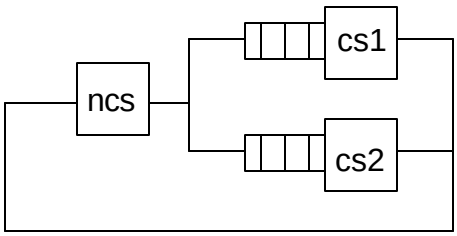
$$D_j^s \leftarrow \sum_i \frac{D_{j,i}^{sh}}{D_i^h} \times R_i^{'}(N^h)$$

- Multiple class case:

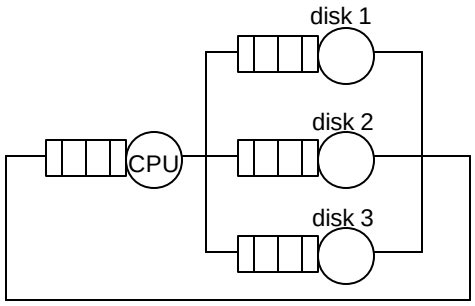
residence time at device i. for class r.

$$D_{j;r}^s \leftarrow \sum_i \frac{D_{j;i,r}^{sh}}{D_{i;r}^h} \times R_{i;r}^{'}(\vec{N}^h)$$

Example



Software QN



Hardware QN

23

Comparison with other approaches

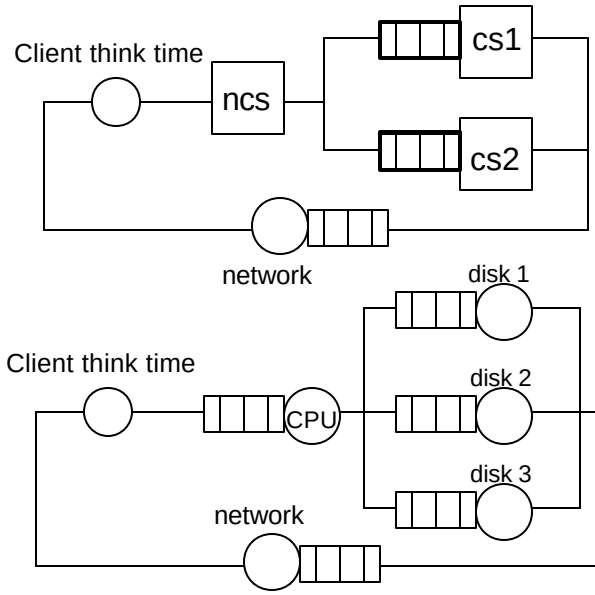
			Absolute % Error		
N	**SQN-HQN**	**GB**	**SQN_HQN**	**ASM**	**ASPA**
1	1.544	1.54	0.27	0.00	
2	2.088	2.11	1.06	4.60	
3	2.317	2.37	2.22	5.89	4.2
4	2.428	2.49	2.49	5.99	2.8
5	2.487	2.56	2.86	5.87	2.0
6	2.521	2.60	3.05	5.78	1.5
7	2.541	2.62	3.00	5.75	1.5
8	2.555	2.63	2.86	5.77	1.1

GB: global balance equations
ASM: Aggregate Server Method [Agrawal and Buzen 1983]
ASPA: Avg. Subsystem Population Approximation [Jacobson & Lazowska 1983]

SQN is consistently pessimistic.
ASPA is much more complex to implement.

24

Modeling Non-Software Resources



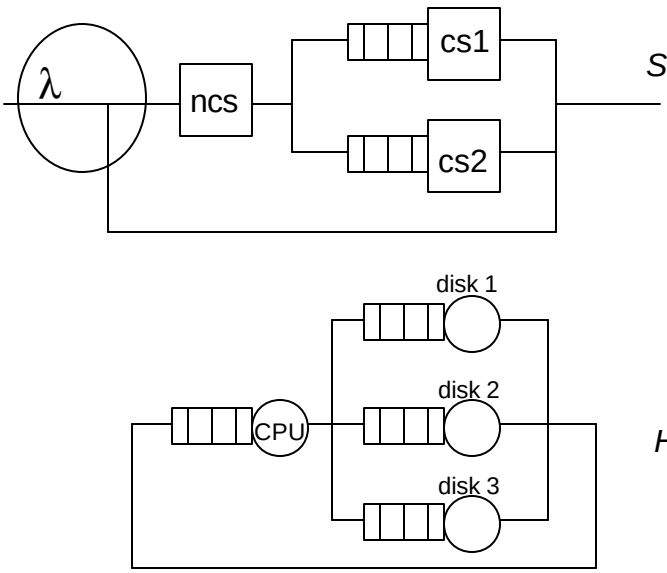
Software QN

B is the avg. no. of processes in the software resource waiting lines.

Hardware QN

25

Open QN at the Software Level

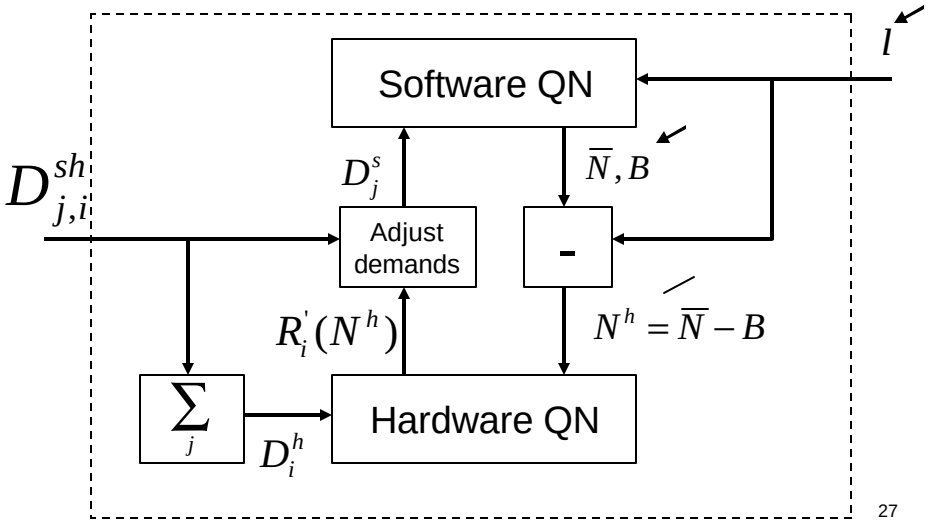


Software QN

Hardware QN

26

SQN-HQN Scheme: Open SQN



27

Results of Iterations for Open SQN Case

					Adjusted SQN Demands		
Iteration	**Nh**	**Resp. Time**	**Ns**	**B**	**NCS**	**CS1**	**CS2**
0	-				0.3280	0.1176	0.2020
1	1.295	0.821	1.641	0.346	0.3662	0.1302	0.2235
2	1.440	0.946	1.893	0.453	0.3858	0.1365	0.2342
3	1.513	1.014	2.028	0.515	0.3958	0.1397	0.2396
4	1.550	1.050	2.100	0.549	0.4010	0.1414	0.2424
5	1.570	1.069	2.137	0.568	0.4037	0.1423	0.2438
6	1.580	1.079	2.157	0.577	0.4051	0.1427	0.2446
7	1.585	1.084	2.167	0.582	0.4059	0.1430	0.2450
8	1.588	1.086	2.173	0.585	0.4062	0.1431	0.2452
9	1.589	1.088	2.175	0.587	0.4064	0.1432	0.2453
10	1.590	1.088	2.177	0.587			

28

Related Work

- [Agrawal&Buzen, 1983]: approximate iterative technique. Aggregate servers represent serialized processing. Single class presentation.
- [Thomasian, 1983]: two-level model (QN+MC). Does not generalize to multiple classes.
- [Jacobson&Lazowska, 1983]: transform network to one with population constraints and solve through approximation.
- [Llado&Harrison, 2000] Modeling of Enterprise JavaBean Server.
- [Reeser&Hariharan, 2000] Web server software contention. Single class.
- [Ramesh&Perros,2000] Multi-layer client-server QNs.
- [Kahkipuro,2000] Performance modeling framework for CORBA-based systems. QNs with simultaneous resource possession.
- [Rolia&Sevcik, 1995] [Woodside, 1986] Method of Layers. LQNs.
- [Petriu et al, 2000] [Woodside et al., 1995] Stochastic rendez-vous.
- [Franks&Woodside, 1998] Parallel operations in multi-layer model.

Concluding Remarks

- Simple approach.
- Open, closed, and multiclass QNs can be used at the SQN.
- SQNs can include non-software resources that are not mapped to hardware resources.
- HQNs are closed and can be multiclass.
- Any technique can be used to solve the SQN and HQN. This includes any known approximation to multiple-server devices, priorities, simultaneous resource possession, etc.