Web Server Performance Simulation

Andrew Ferrier

Supervisor: Peter Harrison

18th July 2002

Aims / Parts

1. Creation of WS^3 (Web Server Simulation System) — simulates generic systems.
2. Web Serving Guidelines using WS^3 — by evaluating hypothetical models.

Background and Motivation

- Increasing use of web: performance issues.
- Notorious failures (e.g. UK 1901 Census).
- Capacity planning tools tend to be:
  - Flooding tools.
  - General simulation tools/toolkits.
  - GUI tools.
  - Non-web-specific.

Simulation

Simulation uses virtual time-stream; queueing network; statistical distributions to create inter-event times.

- Good at answering specific questions.
- Quick and easy — unlike queueing theory.
- Requires attention to accuracy and general pattern discovery may be difficult.

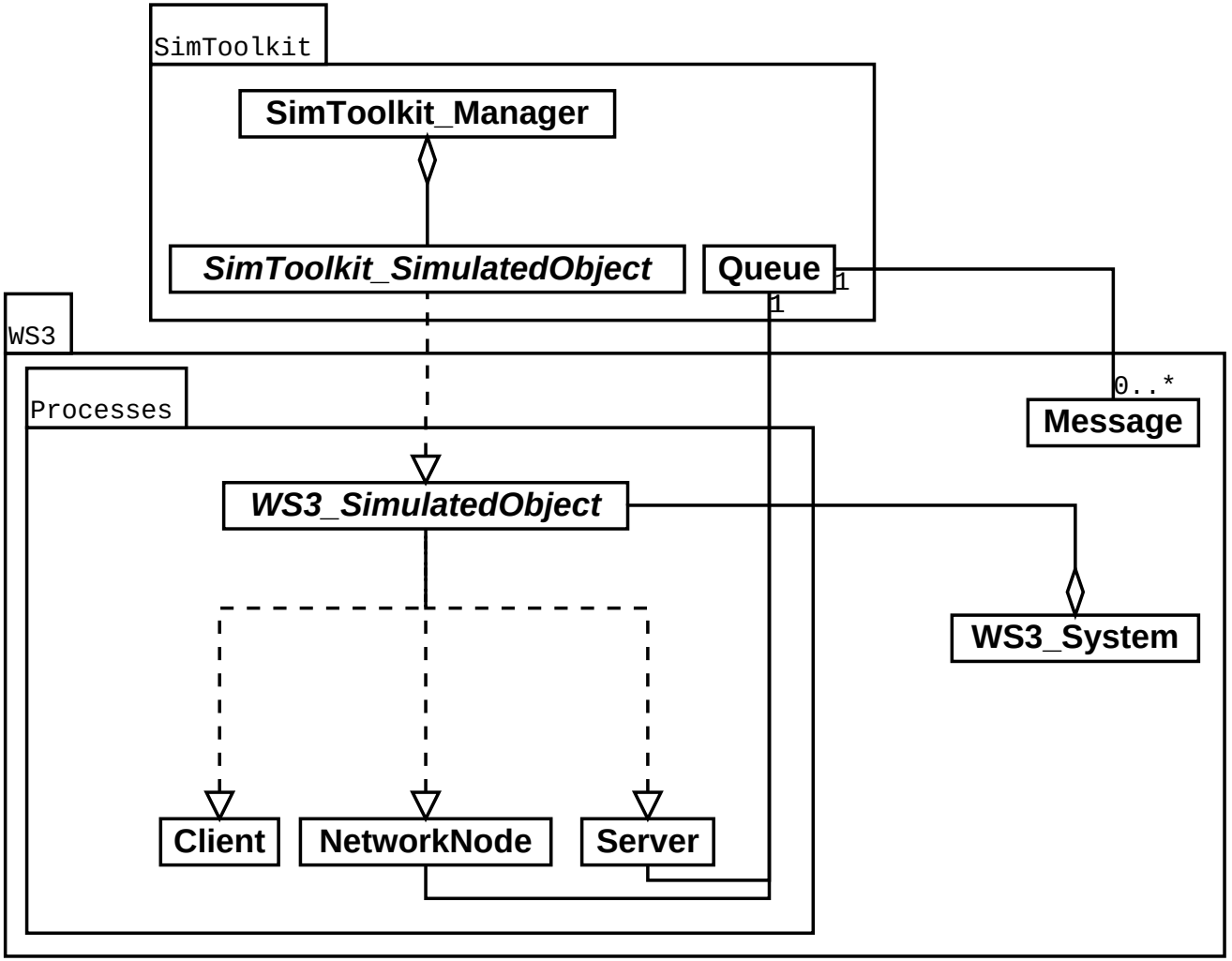
Design of WS^3

Objectives:

- Easy-to-use.
- Unlikely to cover everything so designed for future extension.

Decisions:

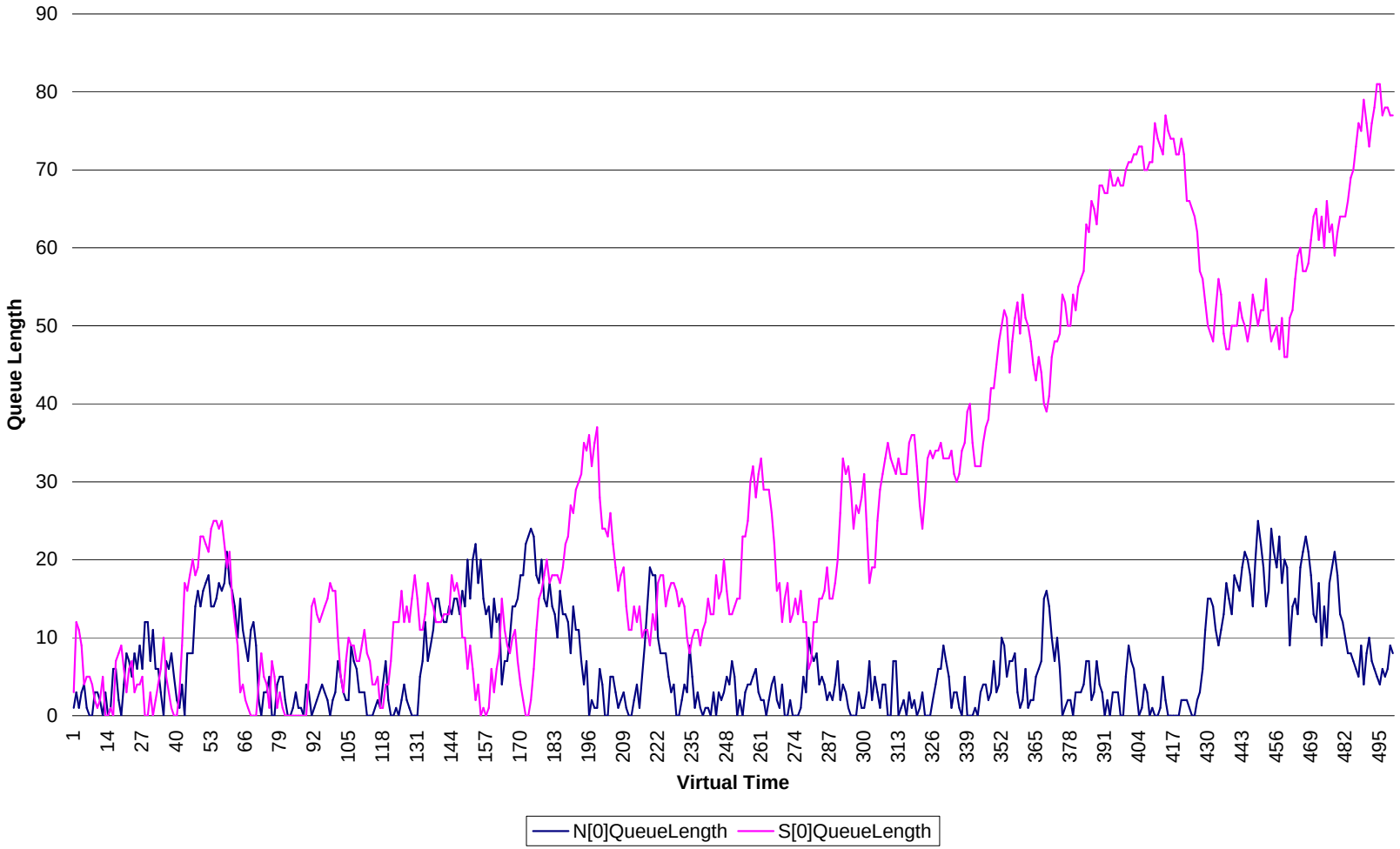
- Java (programming language).
- XML (input file format — for system specification), XML Schema, Apache Xerces Parser.
- Simulation Toolkit (Tony Field) — process based.

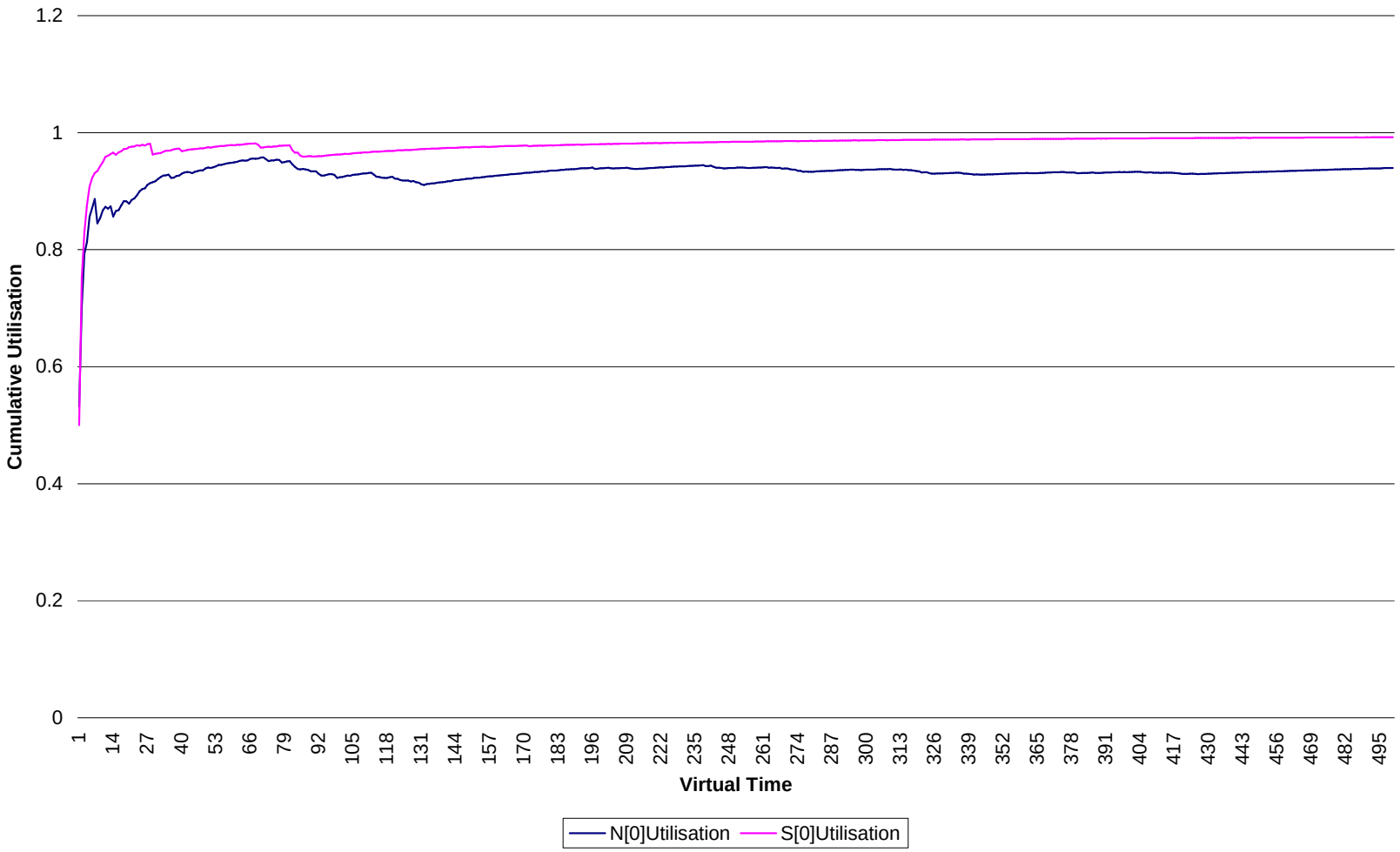

SimToolkit
SimToolkit_Manager
SimToolkit_SimulatedObject
Queue
1
1
WS3
Processes
0..*
Message
WS3_SimulatedObject
WS3_System
Client
NetworkNode
Server

Implementation Issues

Implementation mostly straightforward. However:

- Randomness — $x_n = f(x_{n-1})$.
- Equilibrium.
- Added features:
  - Routing System Changed.
  - Server Multithreading and Multiprocessors.
  - Tracing / Data Dumping.
  - Queue Length Capping.
  - Others (see report).


90
80
70
60
50
40
30
20
10
0
Queue Length
1
14
27
40
53
66
79
92
105
118
131
144
157
170
183
196
209
222
235
248
261
274
287
300
313
326
339
352
365
378
391
404
417
430
443
456
469
482
495
Virtual Time
N[0]QueueLength
S[0]QueueLength


1.2
1
0.8
0.6
0.4
0.2
0
Cumulative Utilisation
1
14
27
40
53
66
79
92
105
118
131
144
157
170
183
196
209
222
235
248
261
274
287
300
313
326
339
352
365
378
391
404
417
430
443
456
469
482
495
Virtual Time
N[0]Utilisation
S[0]Utilisation

Testing / Evaluation Distinction

- Testing — TST — Validity test cases: part of software creation.
- Evaluation — CNC — Hypothetical queueing systems: used to discover guidelines: second part of project.

  Also discussed accuracy, speed of simulation, etc. (in report)

Testing

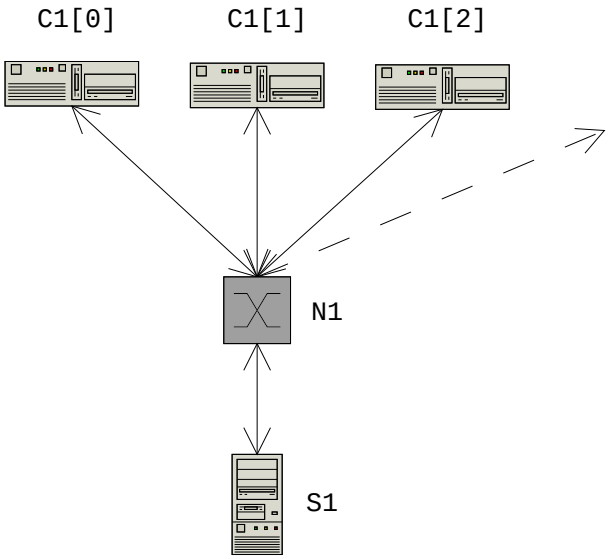
Test cases for:

- Invalid XML / Invalid XML for WS^3.
- Simple test cases — checked with queueing theory
- Boundary condition cases (large names, unusual parameters etc.)
- Others (see report)

Evaluation

Second (and smaller) part of project.

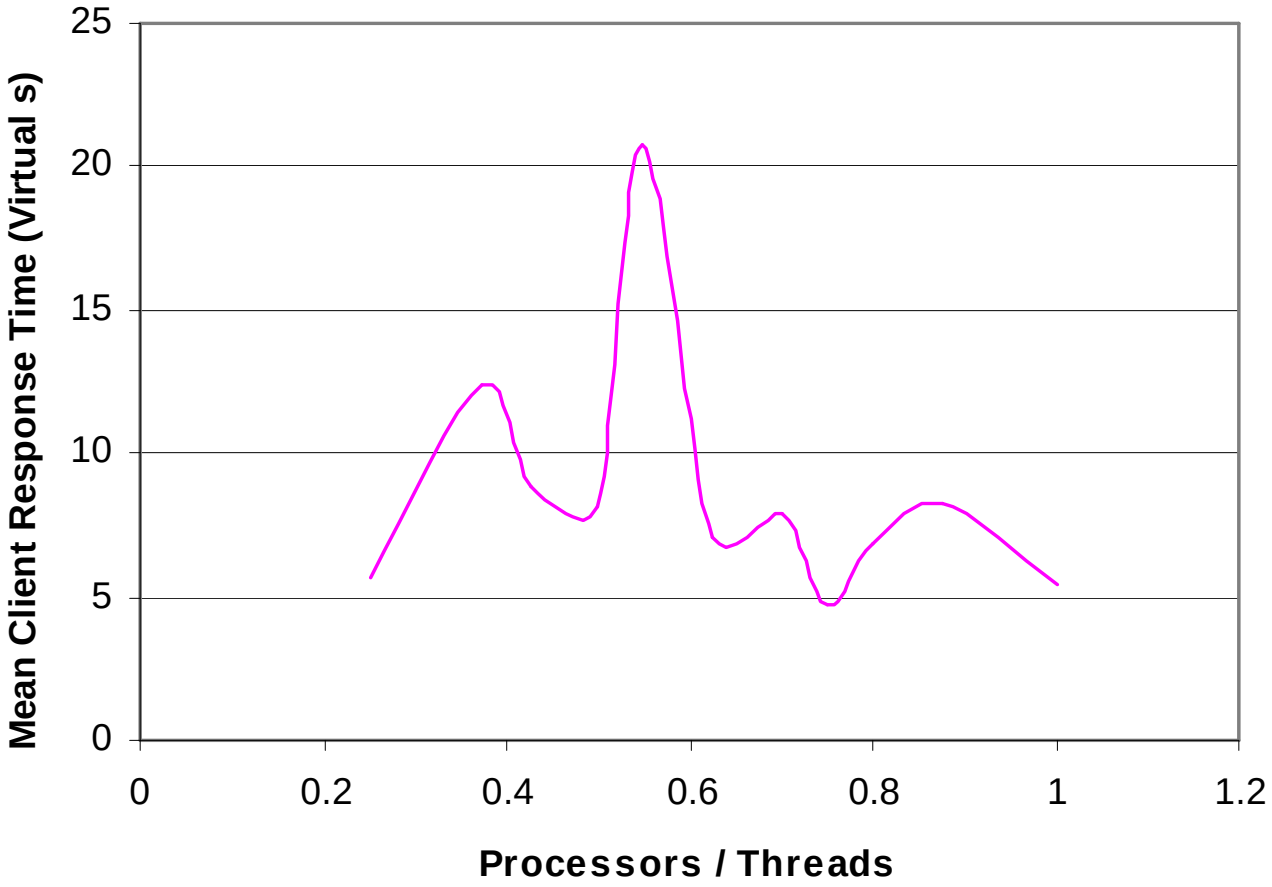
One unusual and interesting example from report. Purpose was to examine ratio: $\frac{Processors}{Threads}$ on a single server.



Evaluation Procedure

- 50 clients. Varied $\frac{Processors}{Threads}$ ratio. Ran simulations (equilibrium-adjusted).
- Plotted server utilisation, client response time, other parameters.

Most interesting pattern discovered was...


Mean Client Response Time (Virtual s)
25
20
15
10
5
0
0
0.2
0.4
0.6
0.8
1
1.2
Processors / Threads

Conclusion

1. Created WS^3 — many features: clients, network nodes, multi-processor and multi-threaded servers, 8 different statistical distributions, different queue lengths, network node message dropping etc. . .
2. Evaluated WS^3 — hypothetical system evaluation, discussed speed, accuracy etc.

Future Extensions

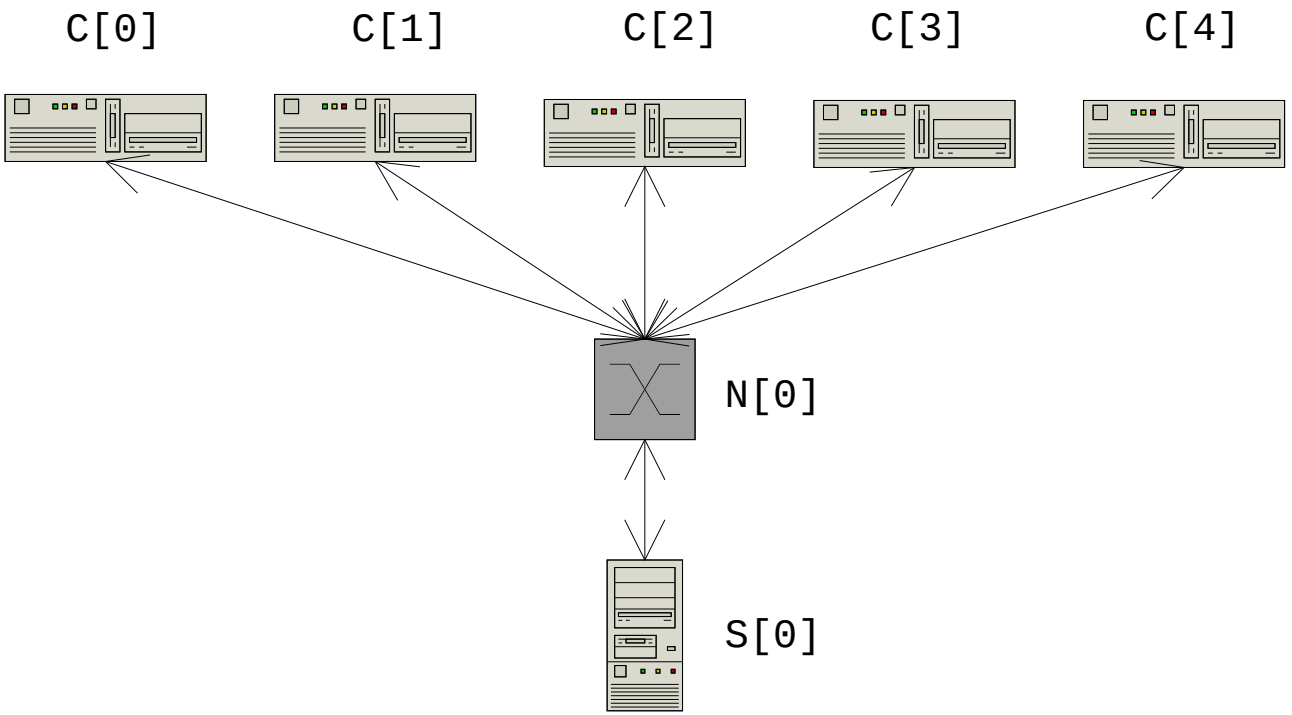
- Features for WS^3: an empirical distribution, time-based demand variation, different types of requests, etc. . .
- Graphical interface for industrial use.
- More analysis of: different hypothetical models, accuracy.

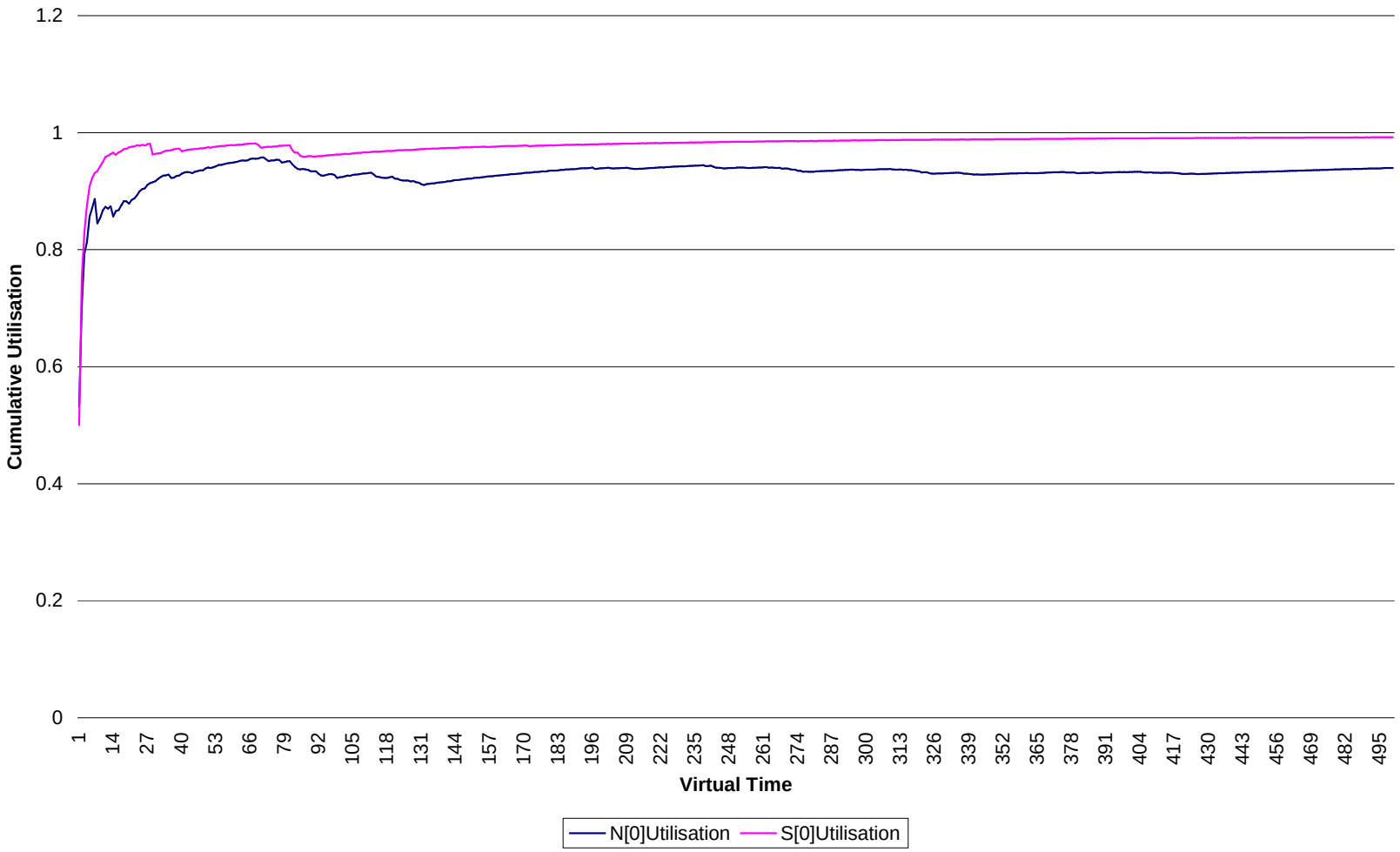
Any Questions?

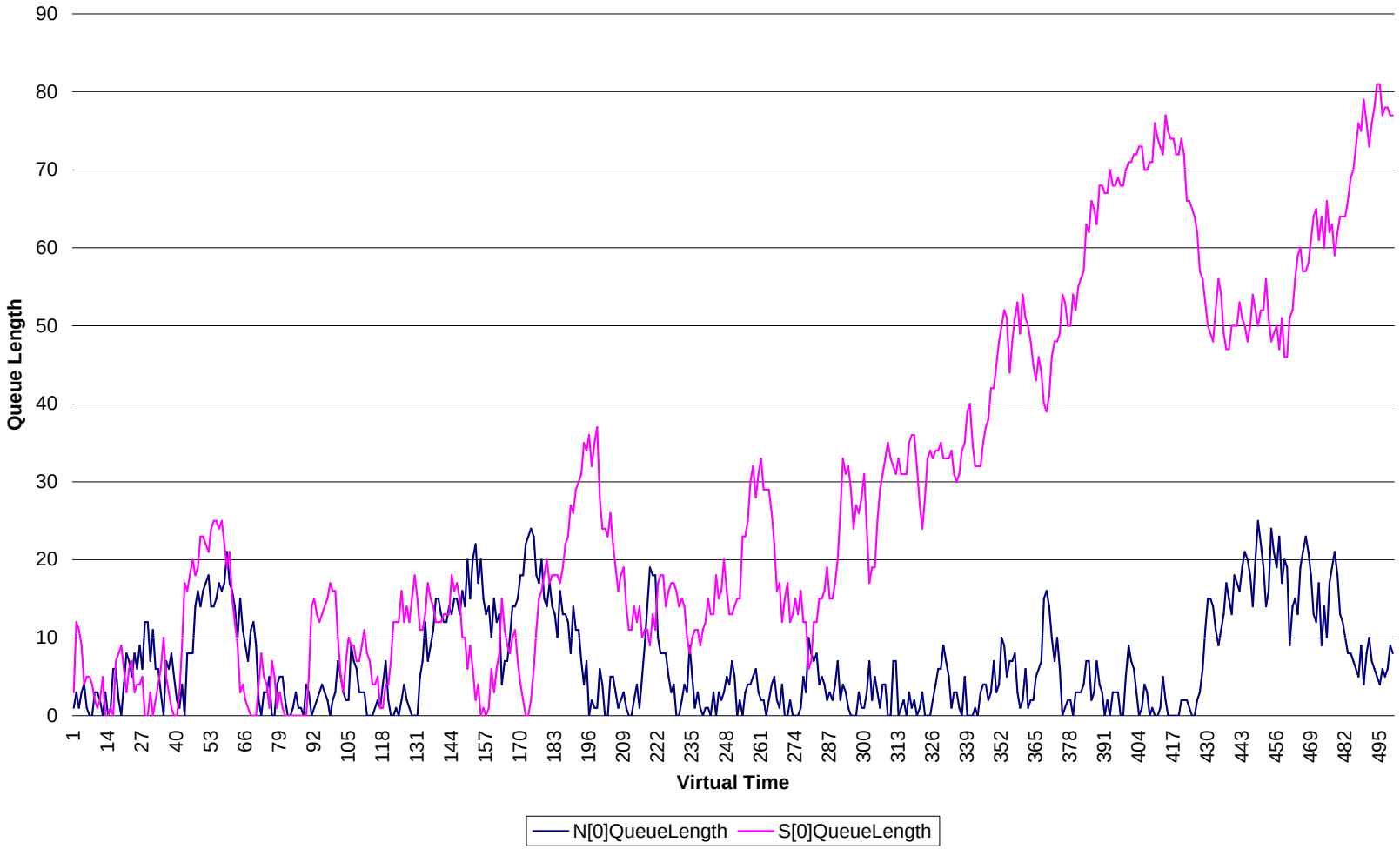
andrew@new-destiny.co.uk
http://www.new-destiny.co.uk/andrew/project/

DoC: ajf98

Thanks to: Peter Harrison, Uli Harder, and Tony Field


C[0]
C[1]
C[2]
C[3]
C[4]
N[0]
S[0]


1.2
1
0.8
0.6
0.4
0.2
0
Cumulative Utilisation
1
14
27
40
53
66
79
92
105
118
131
144
157
170
183
196
209
222
235
248
261
274
287
300
313
326
339
352
365
378
391
404
417
430
443
456
469
482
495
Virtual Time
N[0]Utilisation
S[0]Utilisation


90
80
70
60
50
40
30
20
10
0
Queue Length
1
14
27
40
53
66
79
92
105
118
131
144
157
170
183
196
209
222
235
248
261
274
287
300
313
326
339
352
365
378
391
404
417
430
443
456
469
482
495
Virtual Time
N[0]QueueLength
S[0]QueueLength