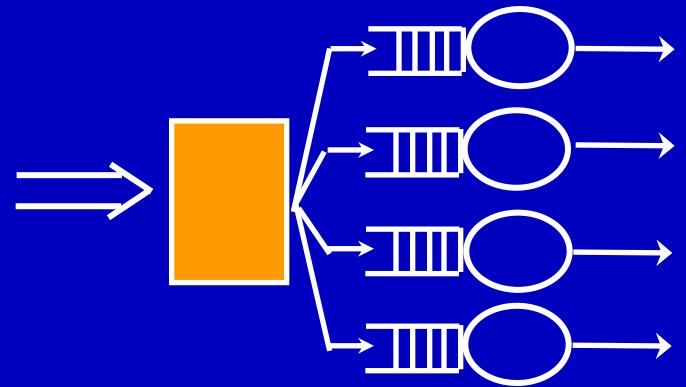


Surprising results on task assignment for high-variability workloads

Mor Harchol-Balter, CMU, Computer Sci.

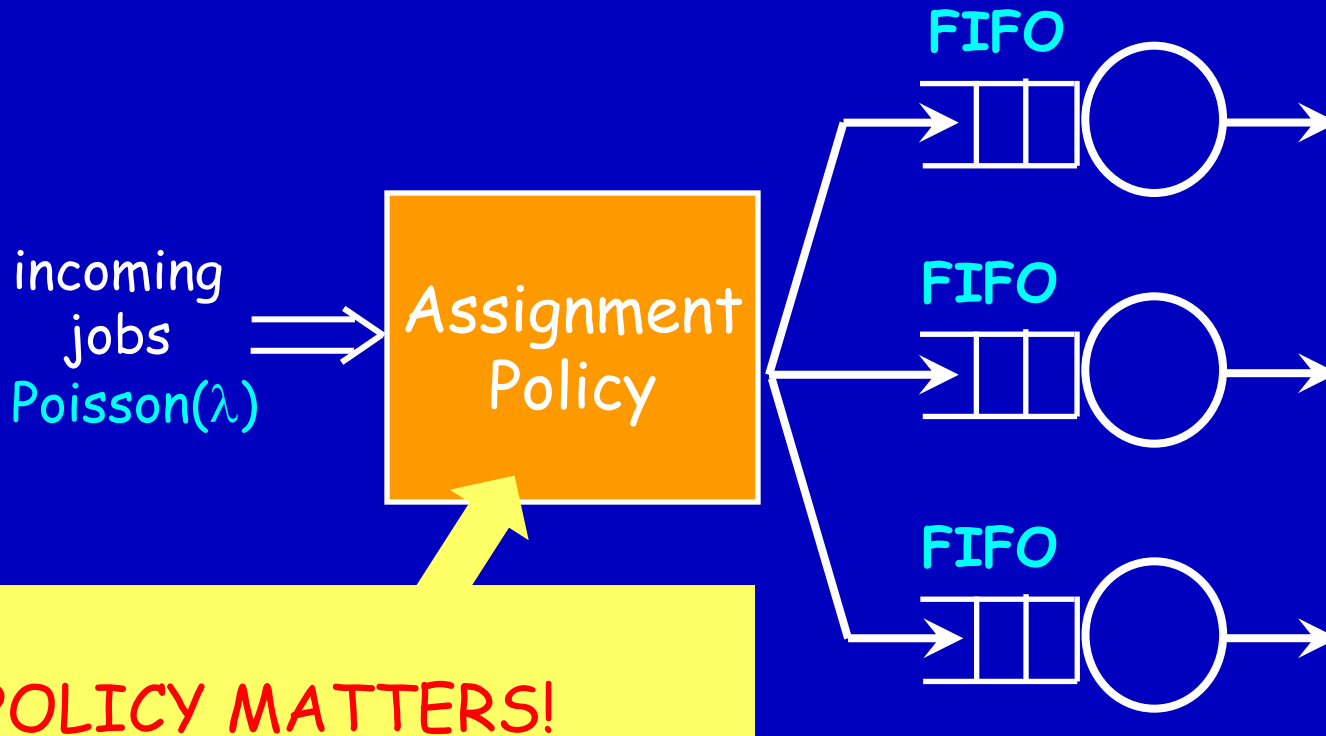
Alan Scheller-Wolf, CMU, Tepper Business

Andrew Young, Morgan Stanley



Server farm model

Goal: Minimize mean response time: $E[T]$



n servers

general i.i.d.
job sizes $\sim X$

$$C^2 = \frac{\text{var}(X)}{E[X]^2}$$

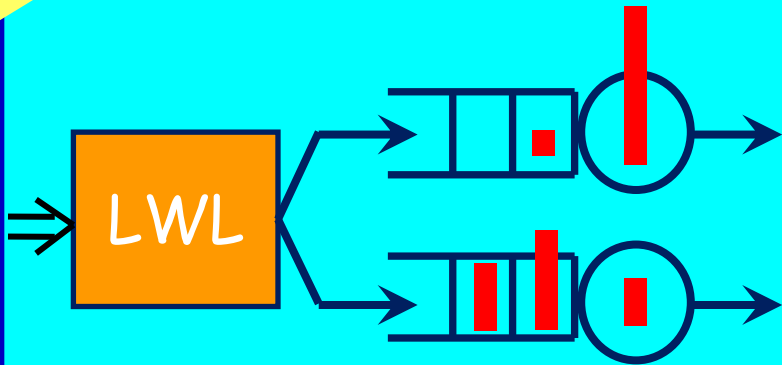
$$\rho = \lambda E[X] \leq n$$

Good Answers

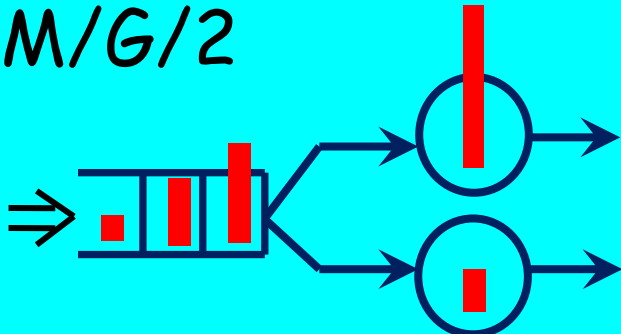
+High throughput

LWL (Least Work Left)

Select job to host with least remaining work.



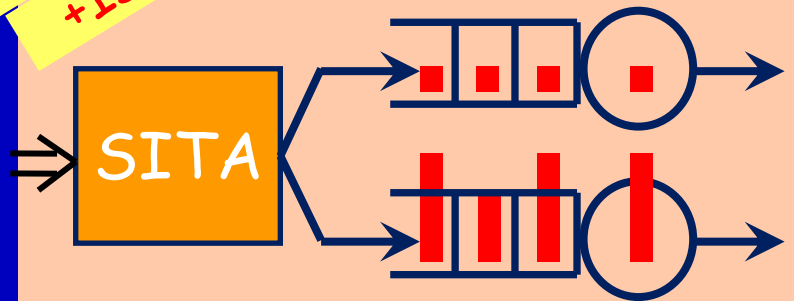
M/G/2



+Protects against high variability
+Isolation for smalls

SITA (Short Interval)

Split on size



Prior Work on SITA

SITA in Practice

- Supercomputing Centers
[Hotovy, Schneider, O'Donnell 96]
[Schroeder, Harchol-Balter 00]
- Manufacturing Centers
[Buzacott, Shanthikumar 93]
- File Server Farms
[Cardellini, Colajanni, Yu 01]
- Supermarkets

Optimizing SITA cutoffs

- [Harchol-Balter, Crovella, Murta 98]
- [Bachmat, Sarfati 08]
- [Sarfati 08]
- [Harchol-Balter, Vesilo 08]

SITA variants

- [Harchol-Balter 00]
- [Harchol-Balter 02]
- [Thomas 08]
- [Tari, Broberg, Zomaya, Baldoni 05]
- [Fu, Broberg, Tari 03]

SITA vs. LWL

- [Broberg, Tari, Zeephongchai 03]
- [Harchol-Balter, Crovella, Murta 99]
- [Cianci, Shinjo 99]
- [Tari, Broberg, Zomaya, Baldoni 05]
- [Thomas 08]

All conclude SITA far superior for high variability

In search of a proof of SITA's total dominance.

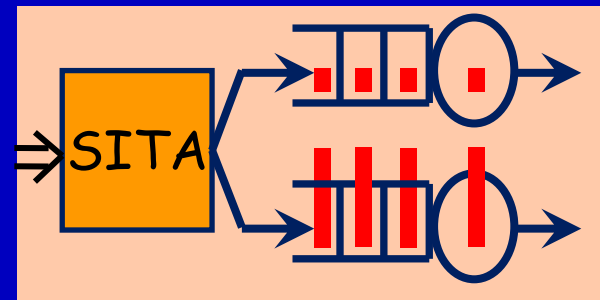
OK, so not optimal, but definite win for high variability.

Should at least beat all commonly used policies when variability is high enough.

Months later

Years later

Can't prove anything because it's not true!



Alternative Title:

The TRUTH about SITA,
under very high job size variability

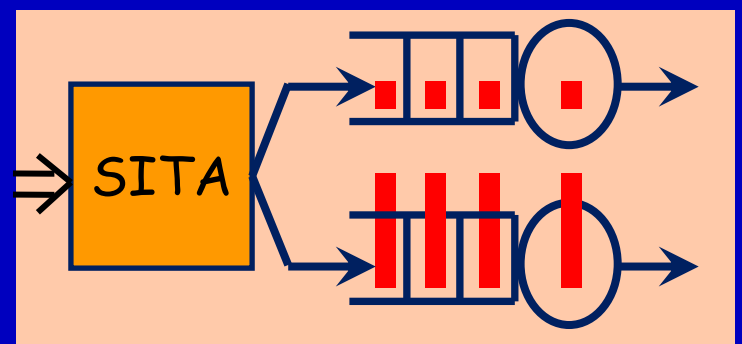
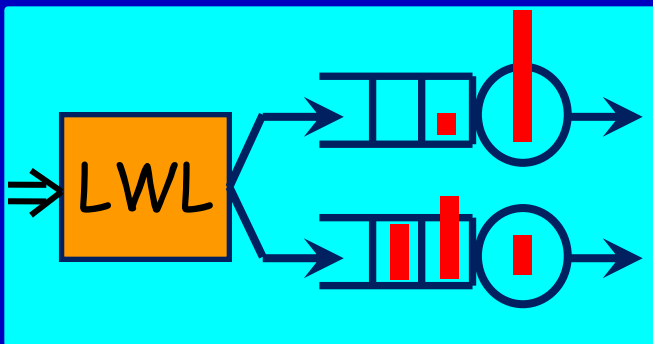
$$C^2 = \frac{\text{var}(X)}{E[X]^2} \rightarrow \infty \quad \text{while} \quad E[X]: \textit{fixed}$$

Q: In this talk we will show ...

as $C^2 \rightarrow \infty$

- a) SITA diverges & LWL diverges?
- b) SITA converges & LWL diverges?
- c) SITA diverges & LWL converges?
- d) SITA converges & LWL converges?

A: All of the above




Q: In this talk we will show ...

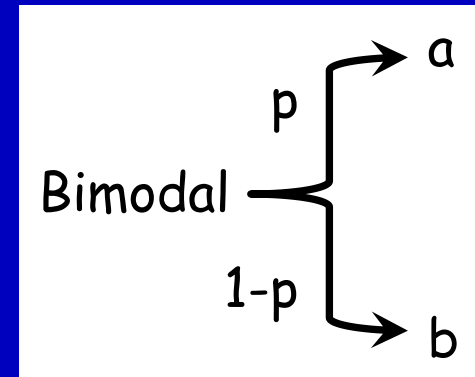
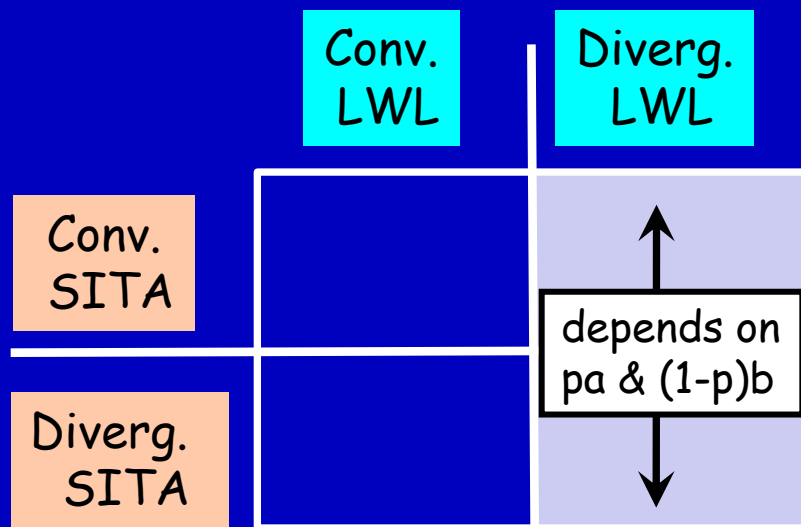
as $C^2 \rightarrow \infty$

	Convergent LWL	Divergent LWL
Convergent SITA	✓	✓
Divergent SITA	✓	✓

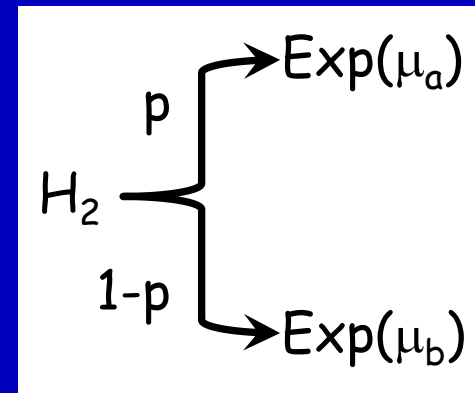
Looking for simple job size distributions to illustrate each.



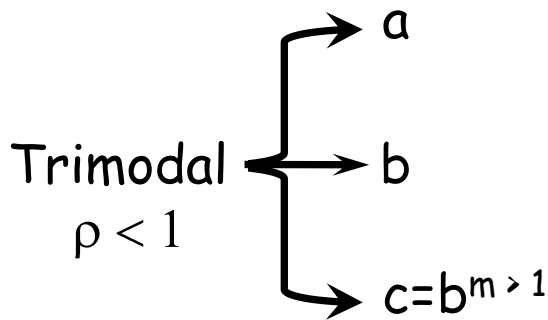
Results (2 server system)



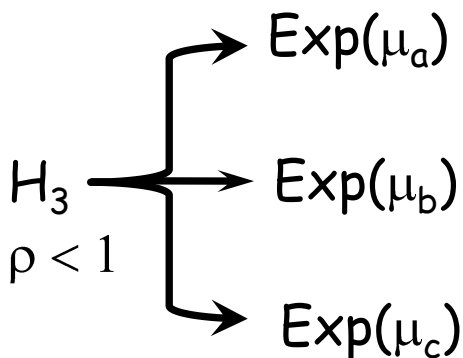
or



Results (2 server system)



or



Conv.
SITA

Diverg.
SITA

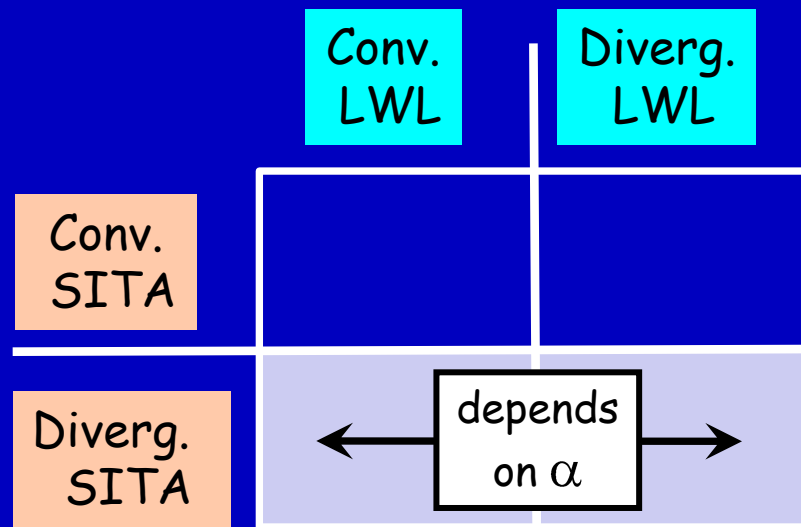
Conv.
LWL

Diverg.
LWL

depends
on m

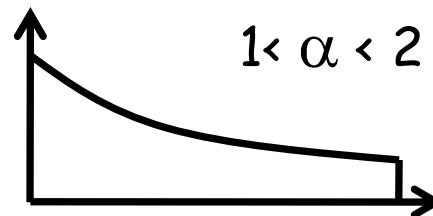


Results (2 server system)



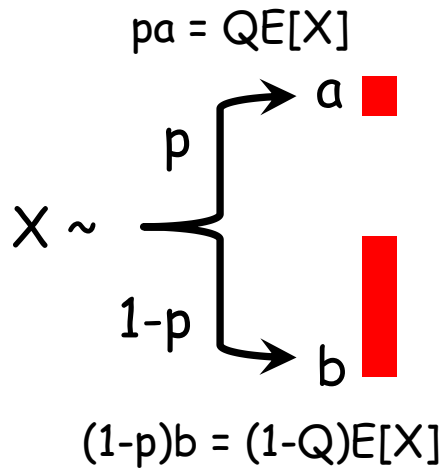
Bounded Pareto(α)

$$1 < \alpha < 2$$



Bimodal Results

	Conv. LWL	Diverg LWL
Conv. SITA		↑ depends ρ_a & ρ_b ↓
Diverg SITA		

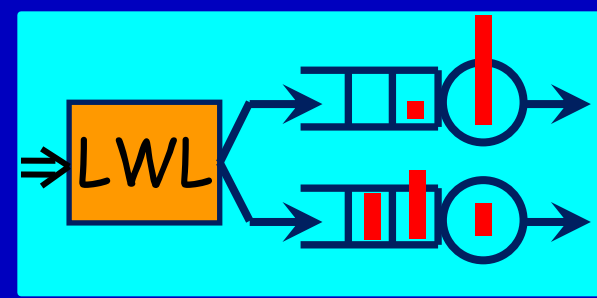


Lemma: As $C^2 \rightarrow \infty$, but $E[X]$, Q : const,
 a 's get little smaller $\rightarrow QE[X]$
 b 's get much bigger $\rightarrow \infty$
 $p \rightarrow 1$

THM: If $\rho_a < 1$ & $\rho_b < 1$
 \rightarrow Convergent SITA

THM: LWL always diverges.

Understanding LWL



Isn't LWL always bad for high C^2 ?

It depends ...

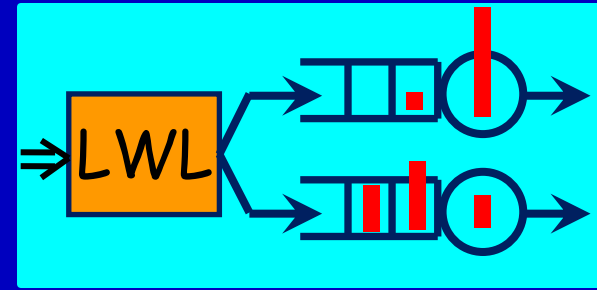
But shorts stuck behind longs, so $E[T] \rightarrow \infty$

Need 2 longs for this to be a problem!

So we need: $\Pr\{2 \text{ longs}\} * E[T | 2 \text{ longs}]$?

Suffices to just look at $E[X^{3/2}]$.

Understanding LWL



Thm: [Scheller-Wolf, Sigman 97], [Scheller-Wolf, Vesilo 06] (2 SERVERS)

$$\text{If } E[X^{3/2}] < \infty \ \& \ \rho < 1 \Rightarrow E[T]^{LWL} < \infty$$

(\Leftarrow usually)

1 spare server

$C^2 \rightarrow \infty$

Thm:

$$\text{If } \left\{ \begin{array}{l} E[X^{3/2}]: \text{bounded} \\ \text{while } C^2 \rightarrow \infty \end{array} \right\} \ \& \ \rho < 1 \Rightarrow \text{LWL converges}$$

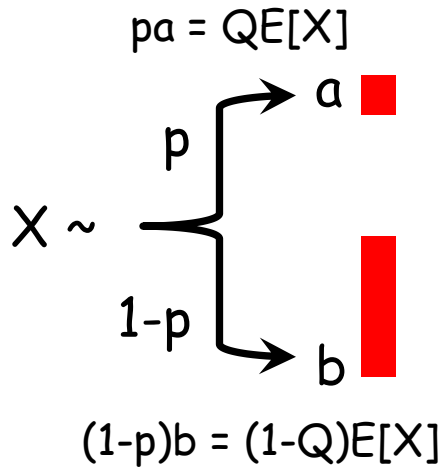
(\Leftarrow usually)

I can make both happen!



Bimodal Results

	Conv. LWL	Diverg LWL
Conv. SITA		↑ depends ρ_a & ρ_b ↓
Diverg SITA		



Lemma: As $C^2 \rightarrow \infty$, but $E[X]$, Q : const,
 $a \rightarrow QE[X]$, $b \rightarrow \infty$, $p \rightarrow 1$

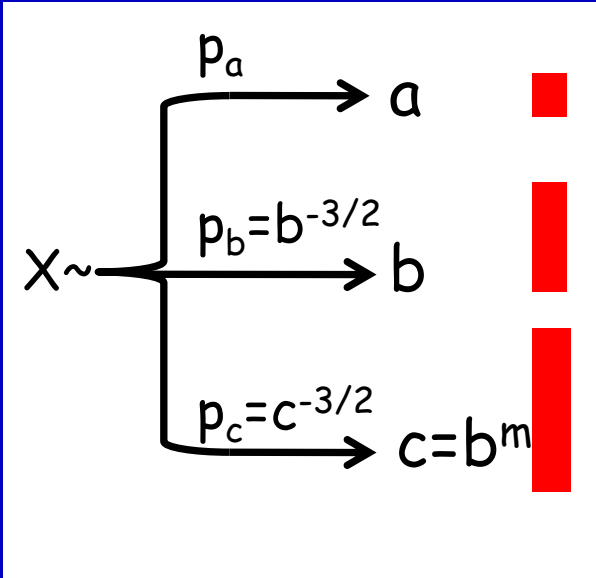
THM: LWL always diverges.

$$\begin{aligned}
 E[X^{3/2}] &= pa^{3/2} + (1-p)b^{3/2} \\
 &= QE[X]\sqrt{a} + (1-Q)E[X]\sqrt{b} \\
 &\rightarrow \infty \text{ (as } C^2 \rightarrow \infty)
 \end{aligned}$$

THM: If $\rho_a < 1$ & $\rho_b < 1$
 \rightarrow Convergent SITA

Trimodal Results

	Conv. LWL	Diverg LWL
Conv. SITA	↑ depends on m ↓	
Diverg SITA		

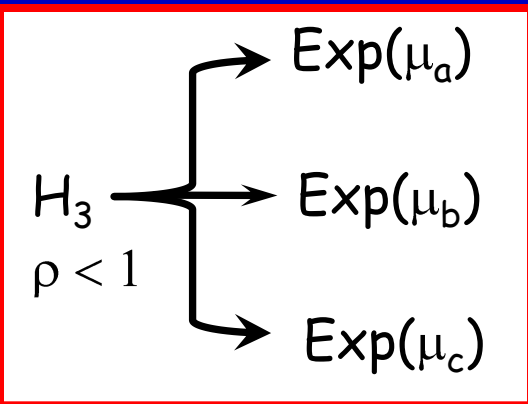
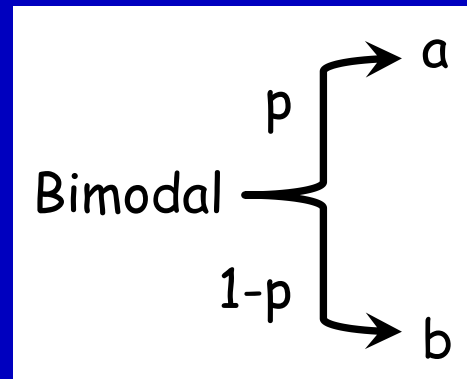
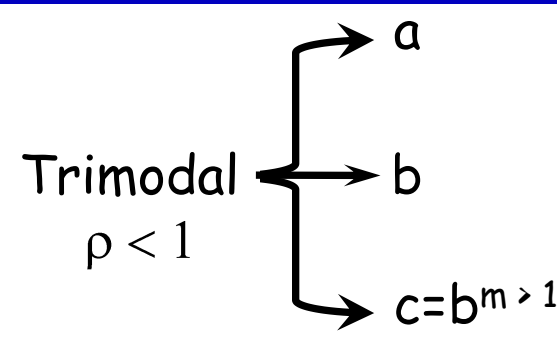


Lemma: As $C^2 \rightarrow \infty$, but $E[X]: \text{const}$,
 $a \rightarrow E[X]$
 $b \rightarrow \infty, c \rightarrow \infty$
 $p_a \rightarrow 1$

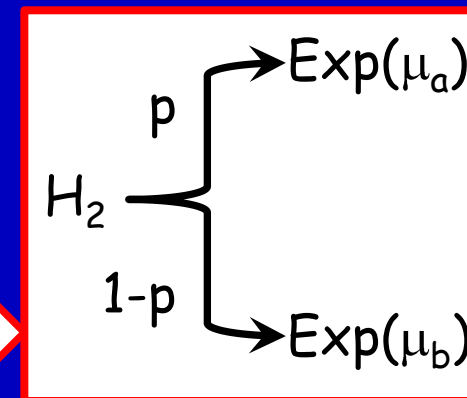
THM: If $m \leq 3$, SITA converges
 If $m > 3$, SITA diverges

THM: LWL always converges for $p < 1$
 $E[X^{3/2}] = p_a a^{3/2} + p_b b^{3/2} + p_c c^{3/2}$
 $\rightarrow E[X]^{3/2} + 1 + 1$

Results (2 server system)



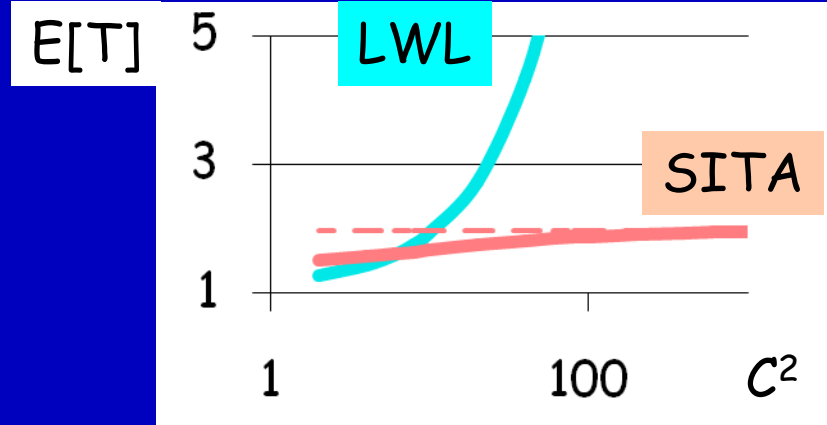
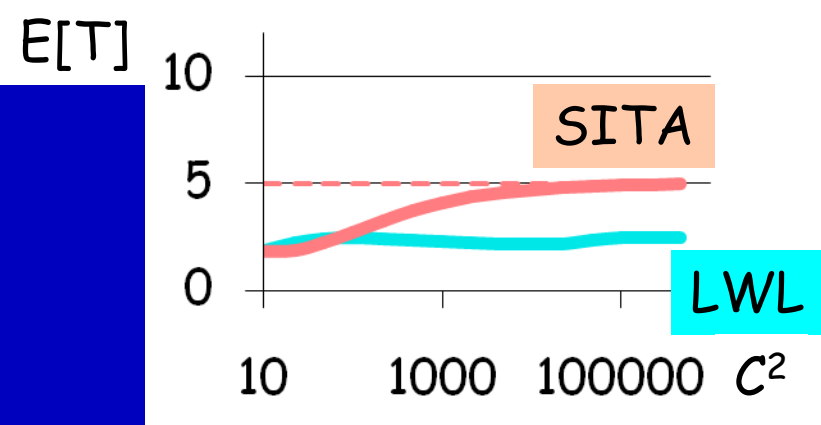
	Conv. LWL	Diverg. LWL
Conv. SITA	✓✓	✓✓
Diverg. SITA	✓✓	✓✓



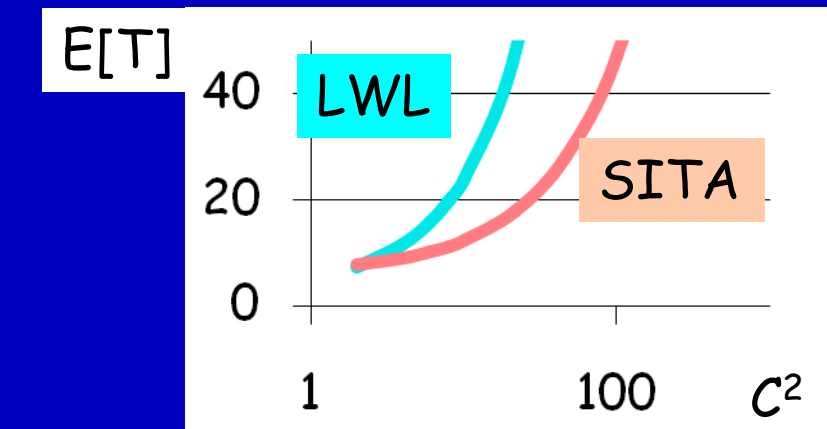
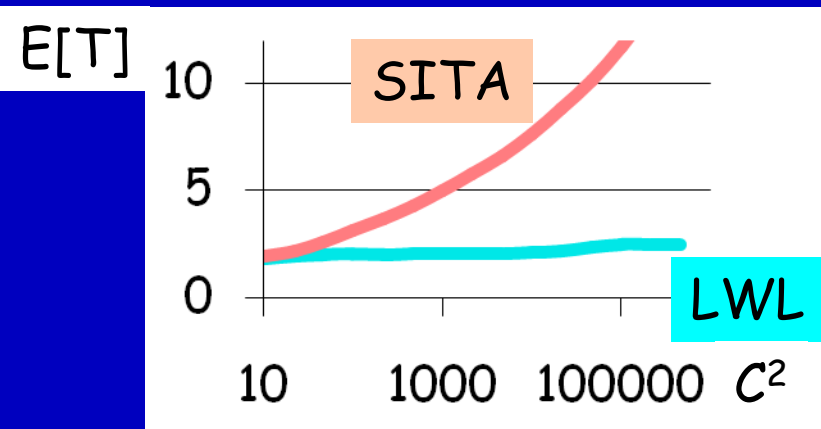
Way more complex, because job types overlap!



"Separation in the limit"



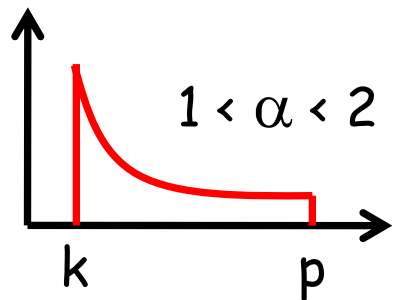
	Conv. LWL	Diverg. LWL
Conv. SITA	✓	✓
Diverg. SITA	✓	✓



Bounded Pareto (2 server system)

	Conv. LWL	Diverg LWL
Conv. SITA		
Diverg SITA	← depends on α →	

$X \sim$ Bounded
Pareto (k, p, α)



Lemma: As $C^2 \rightarrow \infty$, but $E[X]$, α : const,
 $k \rightarrow (\alpha - 1)/\alpha \cdot E[X]$
 $p \rightarrow \infty$

THM: SITA always
diverges.

THM: If $\alpha > 3/2$ and $\rho < 1$,
then LWL converges.
Else LWL diverges.

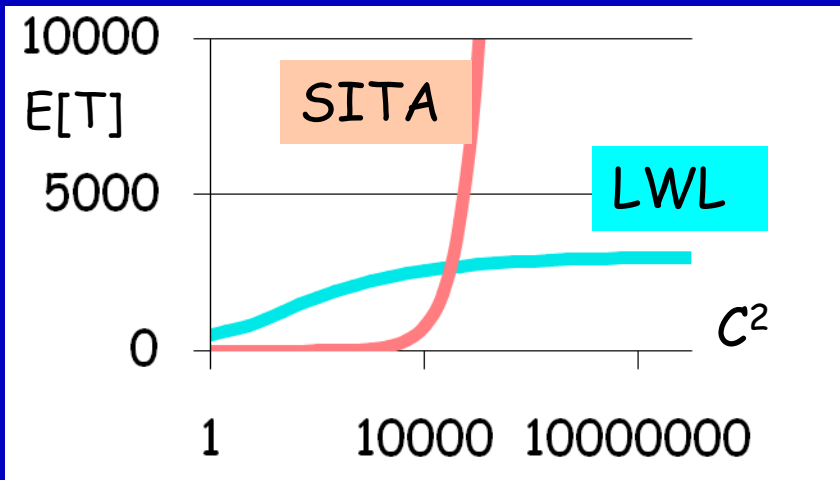
Extends to $n > 2$ servers when $\rho < n - 1$

Bounded Pareto Results

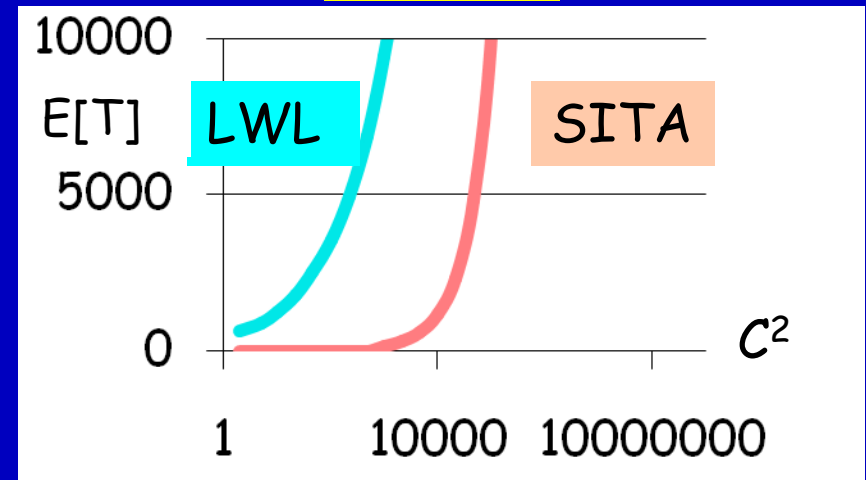
Why was this not noticed?



$\alpha = 1.6$

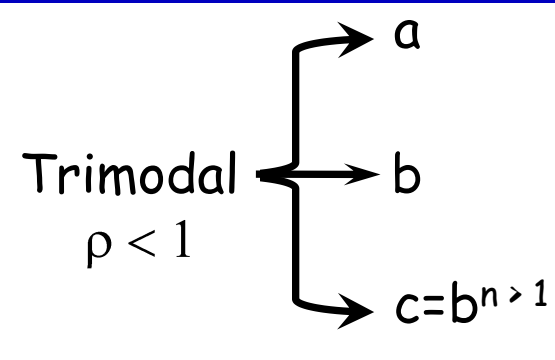


$\alpha = 1.4$

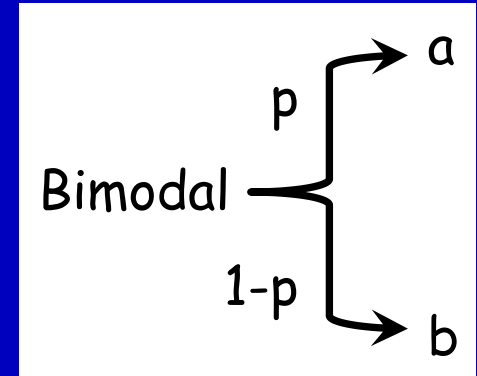
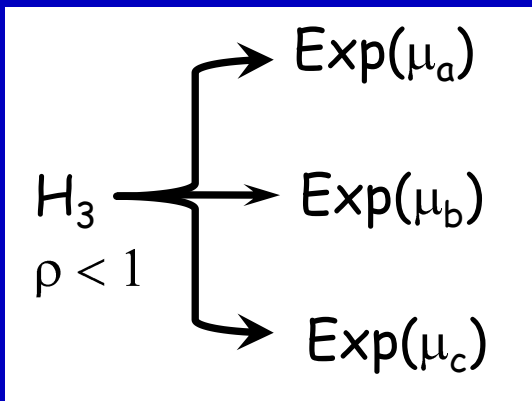


	Conv. LWL	Diverg. LWL
Conv. SITA		
Diverg. SITA	✓	✓

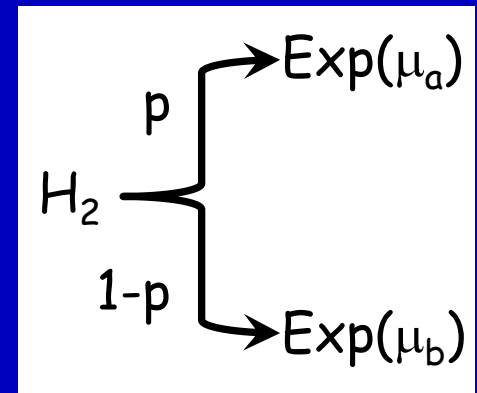
Summary



or



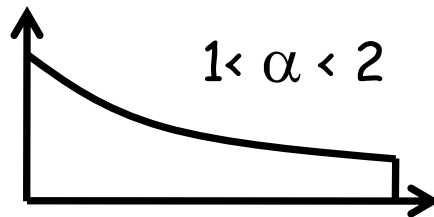
or



	Conv. LWL	Diverg. LWL
Conv. SITA	✓✓	✓✓
Diverg. SITA	✓✓✓	✓✓✓

Bounded Pareto(α)

$$1 < \alpha < 2$$



Old Nursery Rhyme

	Conv. LWL	Diverg. LWL
Conv. SITA	✓	✓
Diverg. SITA	✓	✓



When SITA is good, it is very, very good
But when it is bad, it is horrid.

Epilogue ...

Where did SITA go wrong?

SITA designed to keep shorts from getting stuck behind longs. Isn't that good?

But stringent segregation of shorts & longs can lead to underutilization of servers.

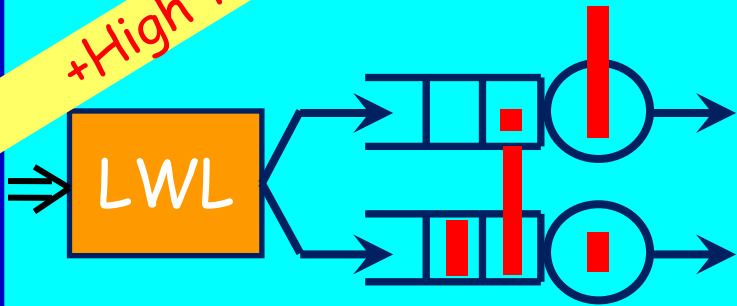
Also, for some distributions, can't subdivide to avoid infinite variability.



Must be a better way...

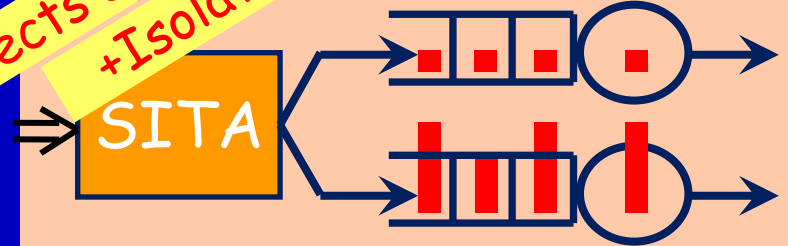
LWL Send jobs to host with least remaining work.

+High throughput

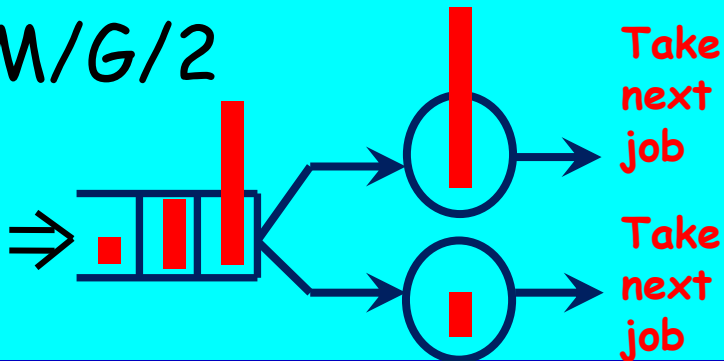


SITA Split jobs based on size cutoff for smalls

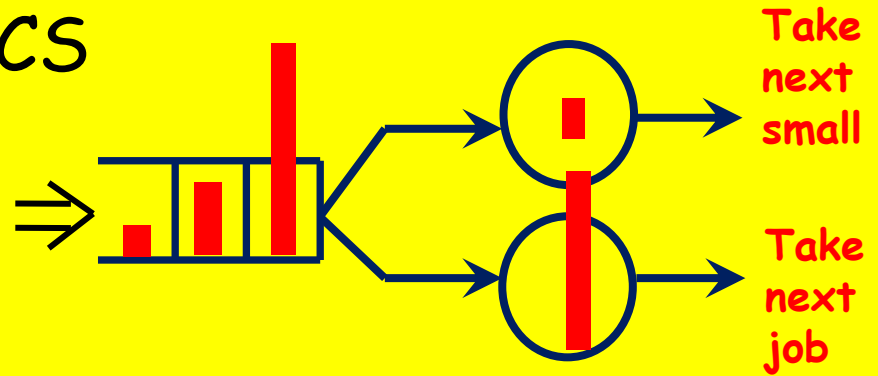
+Protects against high variability
+Isolation for smalls



M/G/2

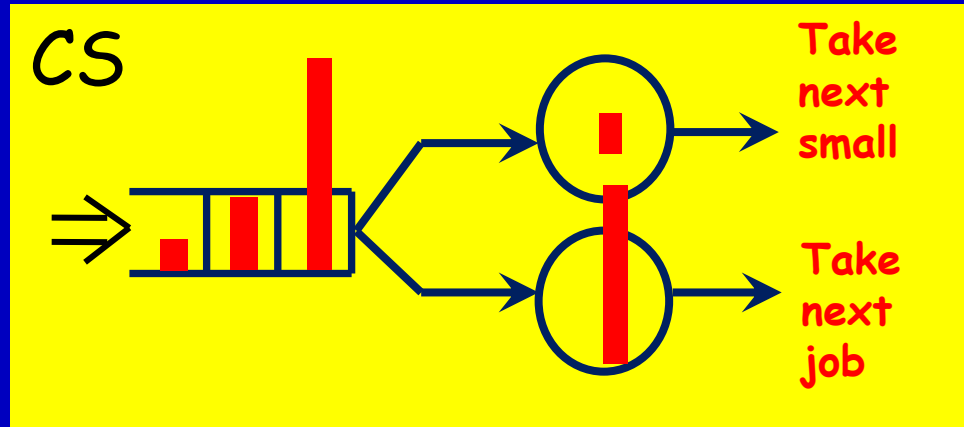


CS





Must be a better way...



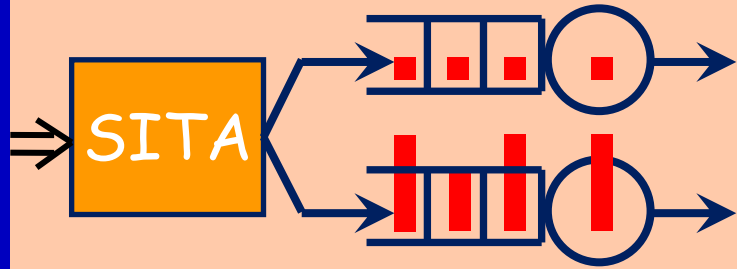
WIN/WIN!

Shorts have isolation from longs
And server utilization is high

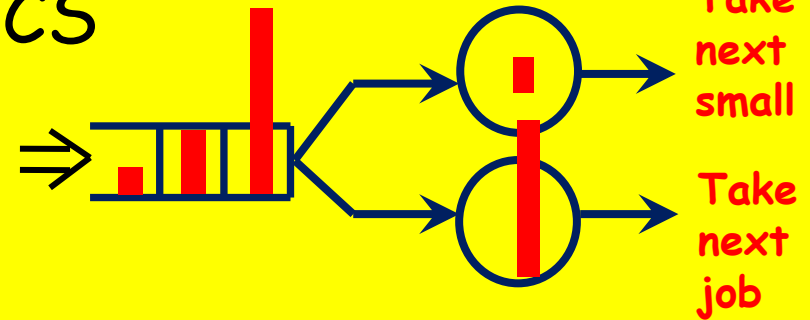
WRONG!

Thm: Whenever SITA diverges, CS diverges too.

SITA Split jobs by size.



CS

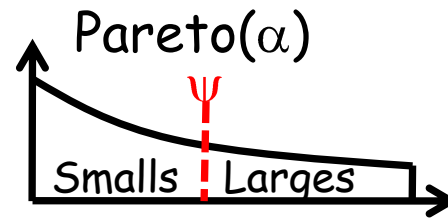


Thm: Whenever SITA diverges, CS diverges too.

PROOF: There are 2 reasons why SITA diverges under given ψ :

(1) Any way of slicing leads to

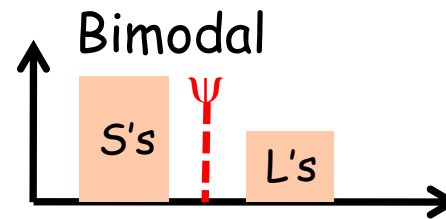
$$\rho_L E \left[K_L^2 \right] \rightarrow \infty$$



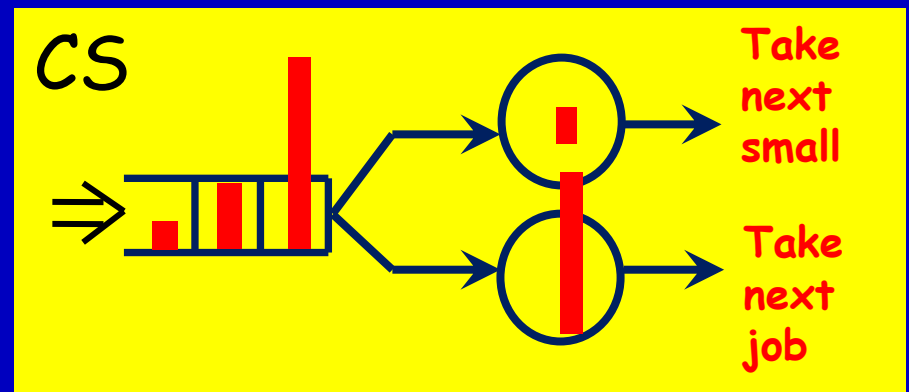
CS can't help.

(2) There is a way of slicing away variability, but it forces

$$\rho_s > 1$$



CS should help, but doesn't.



Thm: Whenever SITA diverges, CS diverges too.

Small job sees L of age $\sim L_e$.

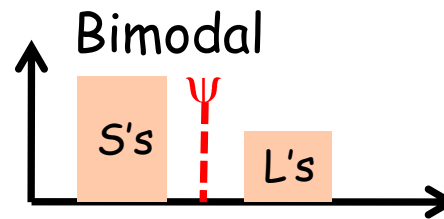
→ Small server has been in overload for $\sim L_e$ time

→ Small sees $\sim L_e$ work → experiences L_e delay.

→ Delay of small $\rightarrow \infty$ as $C^2 \rightarrow \infty$

(2) There is a way of slicing away variability, but it forces

$$\rho_s > 1$$



CS should help, but doesn't.

Conclusion

Maybe isolating short jobs is not the panacea for high-variability workloads after all ...

