Dynamics and Intrinsic Unpredictability of Load Balancing

Greg Robinson
Physics 256
1 June 2010

Motivation

Load balancing is everywhere!

Critical services use load balanced clusters.

- Satellite Ground Systems: GPS, emergency communication...
- Stock exchanges
- Search engines
- Logistics optimization (FedEx, large industrial chains...)

Algorithms often pulled out of a hat... or worse.

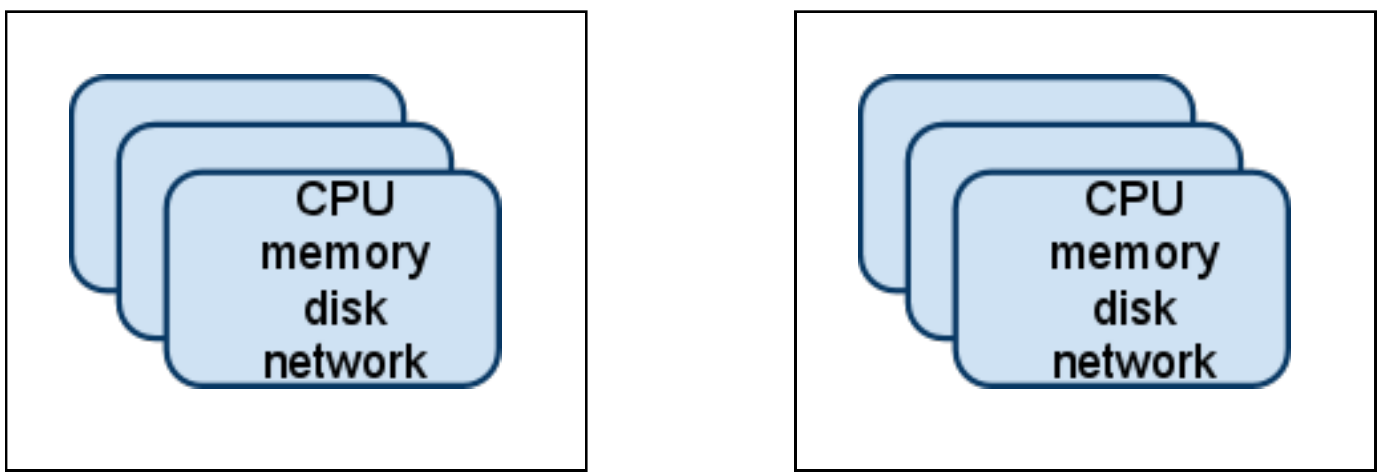
"Heuristics" are used to "tune" algorithms, but rigorous analysis is exceedingly rare.

Insight into relativistic dynamics.

Load Balancing: *dissipation of computational load to maximize efficient utilization of resources*

Types:

- Network resources
  - External bandwidth, intra-cluster, intra-rack...
- "Local" resources
  - CPU, memory, hard disk...
  - Can be abstract: intra-cluster bandwidth is a local resource of a cluster



The Model

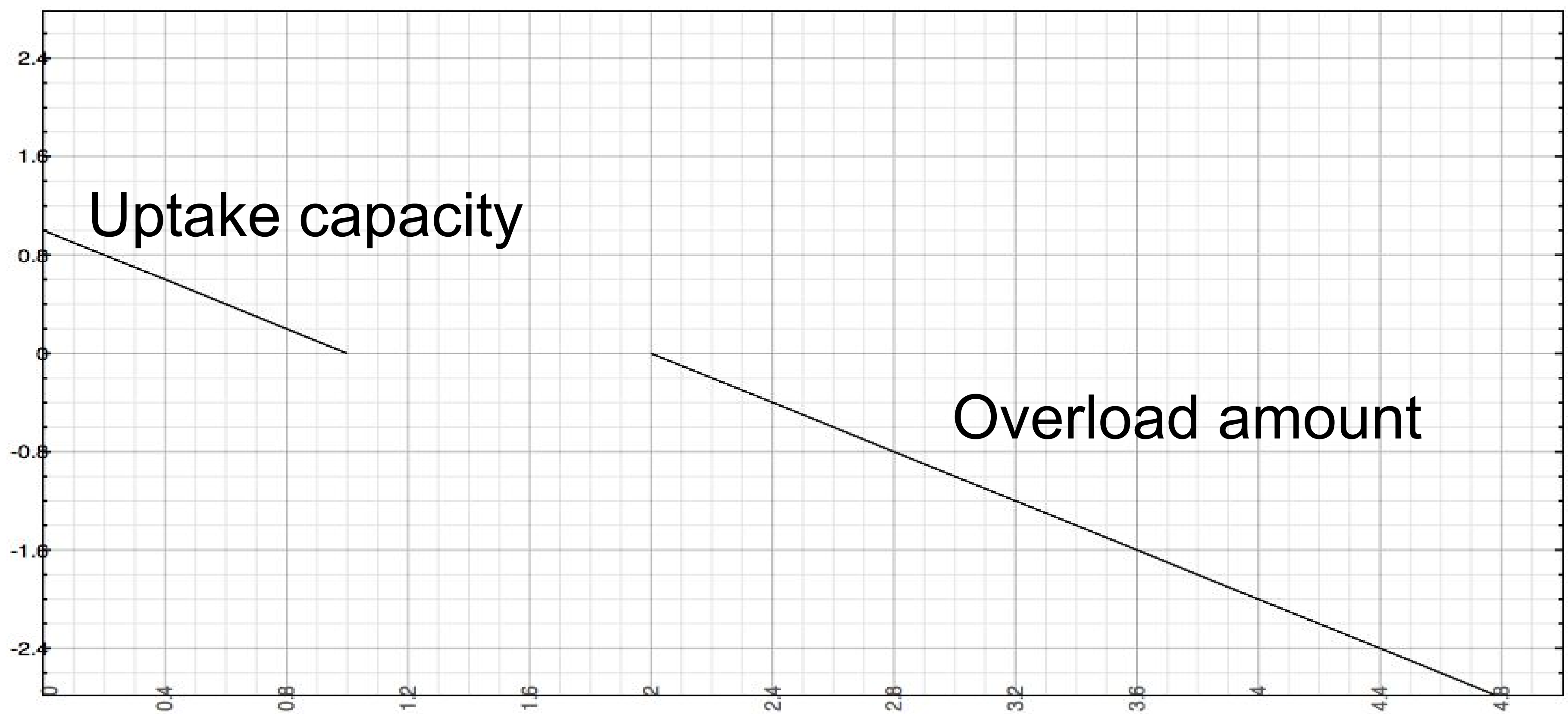
Each cluster has a real-valued CPU load.

Dynamic:

- Piecewise linear function of load.
- Fully conservative
- Does not prevent negative load (why?)
- Distributes overload across all under-loaded clusters
- Weighted by uptake capacity
- NOT instantaneous

$$\omega_i = \gamma\left(l_i - \frac{1}{2}(\Upsilon + \Omega)\right)\mathcal{H}(l_i - \Omega)$$

$$\upsilon_i = \frac{\frac{1}{2}(\Upsilon + \Omega) - l_i}{\Omega}\mathcal{H}(\Upsilon - l_i)\sum_j \omega_j$$



Instantaneous Dynamic

CAACA CAACA BBBBB ACCAC ACCAC ACCAC
ACCAC ACCAC BBBBB CAACA CAACA CAACA
CAACA CAACA BBBBB ACCAC ACCAC ACCAC
ACCAC BCCAC CAACB CAACA CAACA CAACA
CBBCA BCCBB ACCAC ACCAC ACCAC ACCAC
BCCAC BBBBB BBBCB BAACB BAACB BAACB
BBBCB BCCBB BCCAB BCCAB BCCBB BCCBB.....

Instantaneous Dynamic

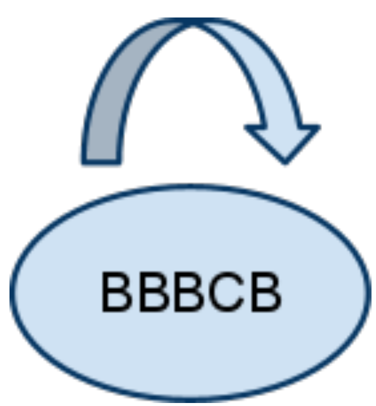
BBBCB BCCBB BCCAB BCCAB BCCBB BCCBB BCCBB
BBBCB BCCBB BCCAB BCCAB BCCBB BCCBB BCCBB
BBBCB BCCBB BCCAB BCCAB BCCBB BCCBB BCCBB
BBBCB BCCBB BCCAB BCCAB BCCBB BCCBB BCCBB
BBBCB BCCBB BCCAB BCCAB BCCBB BCCBB BCCBB
BBBCB BCCBB BCCAB BCCAB BCCBB BCCBB BCCBB
BBBCB BCCBB BCCAB BCCAB BCCBB BCCBB BCCBB
BBBCB BCCBB BCCAB BCCAB BCCBB BCCBB BCCBB

.....

Instantaneous Dynamic

BBBCB BCCBB BCCAB BCCAB BCCBB BCCBB BCCBB
BBBCB BCCBB BCCAB BCCAB BCCBB BCCBB BCCBB
BBBCB BCCBB BCCAB BCCAB BCCBB BCCBB BCCBB
BBBCB BCCBB BCCAB BCCAB BCCBB BCCBB BCCBB
BBBCB BCCBB BCCAB BCCAB BCCBB BCCBB BCCBB
BBBCB BCCBB BCCAB BCCAB BCCBB BCCBB BCCBB
BBBCB BCCBB BCCAB BCCAB BCCBB BCCBB BCCBB
BBBCB BCCBB BCCAB BCCAB BCCBB BCCBB BCCBB

.....



Non-instantaneous

Same as instantaneous case, but make changes happen fixed-N timesteps in the future.

**Locking: any cluster with a change scheduled is "locked". It then plays no role in load balancing decisions until the scheduled change occurs.

Non-instantaneous

Boring for some parameter ranges and initial conditions...

BBBCB BBBCB BBBCB BBBCB BBBCB........

Non-instantaneous

More interesting for others!

Such as:
BCCCB BAACB BAACB BCCCB BCCAB BCCAB
BCCCB BAACB BAACB BCCCB BCCAB BCCAB
BCCCB BAACB BAACB BCCCB BCCAB BCCAB
BCCCB BAACB BAACB BCCCB BCCAB BCCAB
BCCCB BAACB BAACB BCCCB BCCAB BCCAB
BCCCB BAACB BAACB BCCCB BCCAB BCCAB
BCCCB BAACB BAACB BCCCB BCCAB BCCAB...

Periodic!

Some Vague Findings

Algorithm is stable without time delay.

Algorithm periodic even with time delay (not proven)

More periodic orbits exist for larger N, and period is not necessarily bounded by N

Sub-orbits

- For N>=4, there can be orbits which are independently periodic.

No aperiodicity yet discovered for constant N.

All versions of algorithm are irreversible: there are uncountable infinite ways to achieve equilibrium.

Stochastic Delays

Add move to queue between 1 and N timesteps in the future, according to some probability distribution.

Making delays non-constant is **much** harder to explain. (Finally!)

A periodic orbit for one N is not necessarily period for another N, but there may still be some structure to this (working on it!)