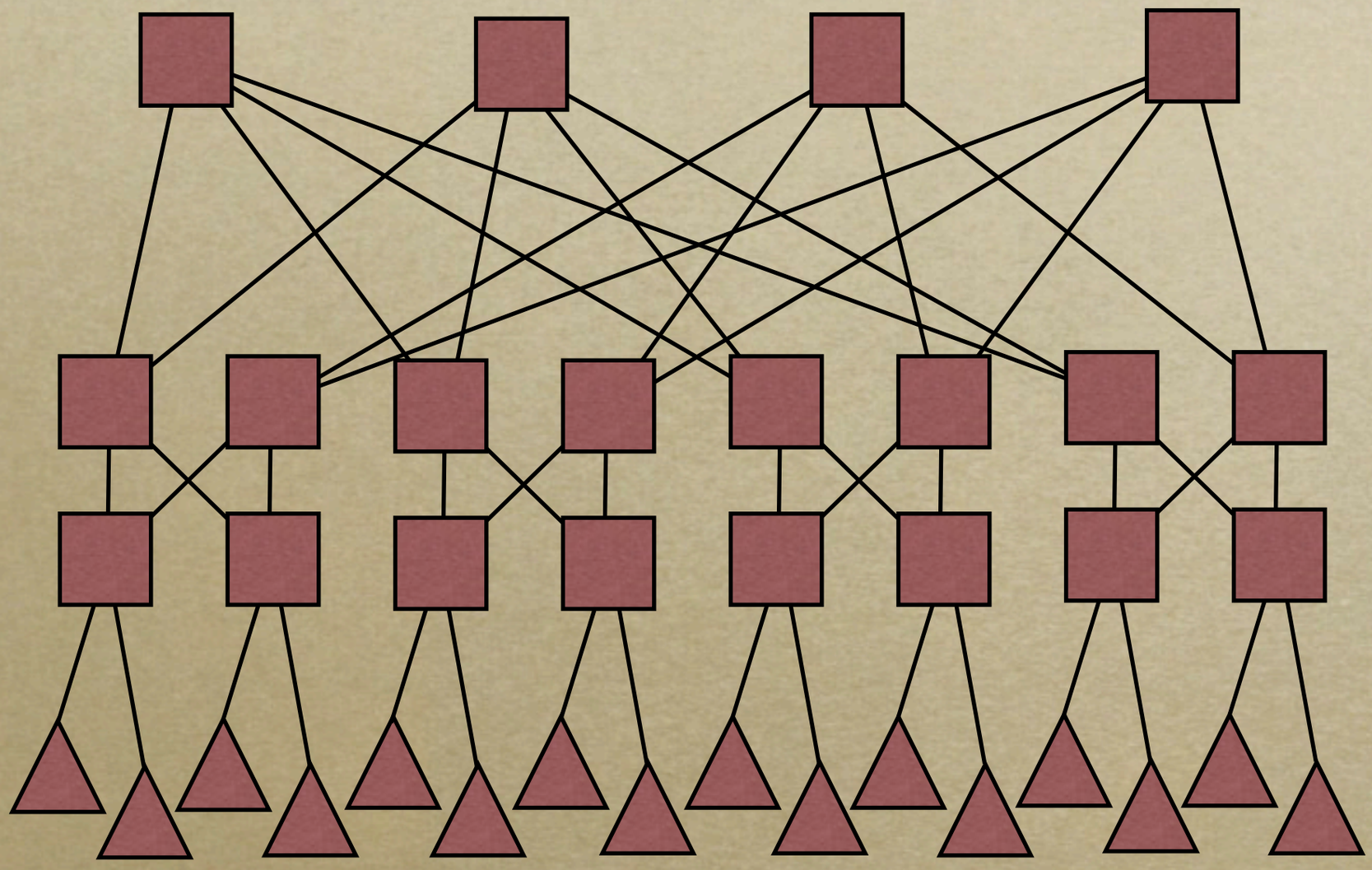
Miswiring models for autonomic naming in a data center fabric

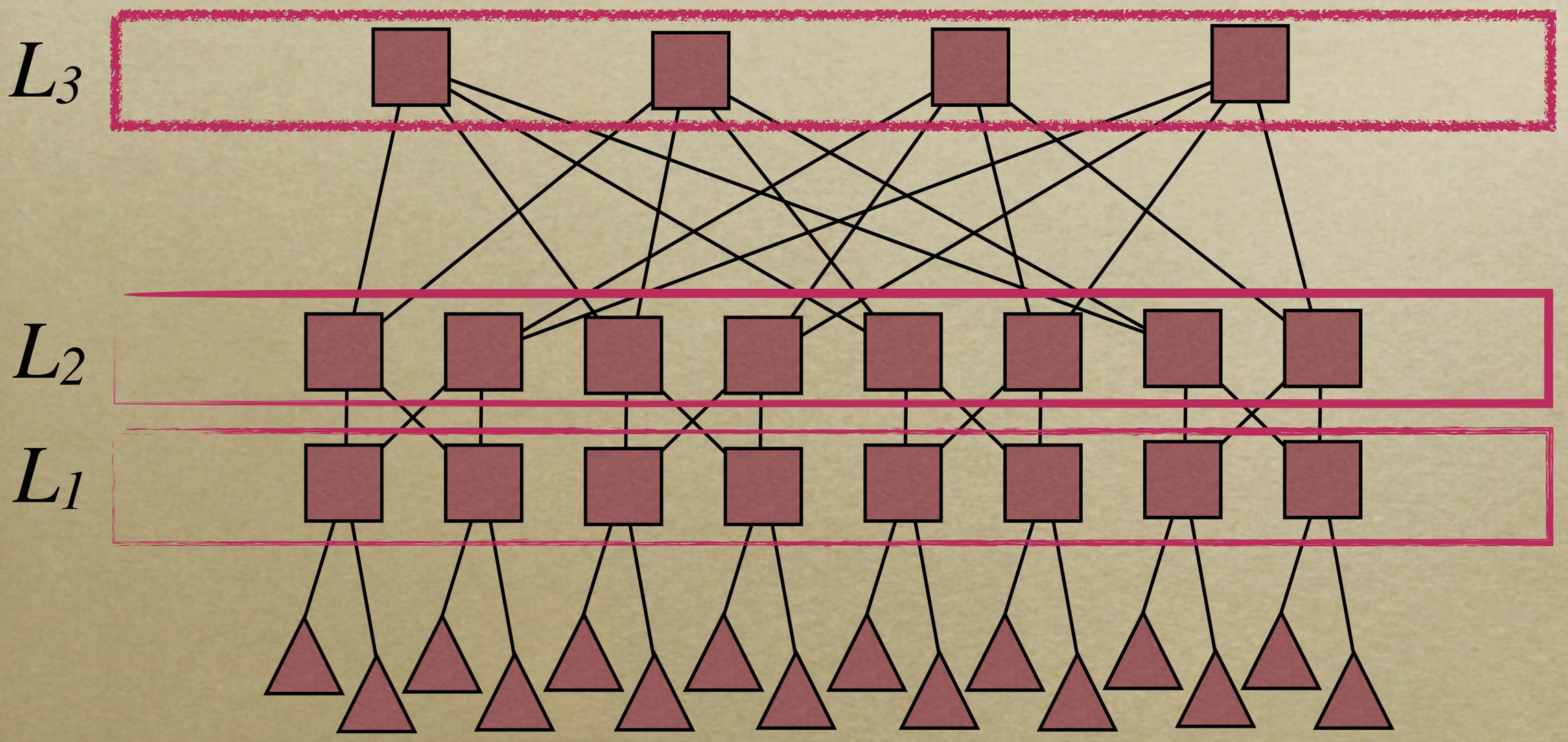
Keith Marzullo, UC San Diego Dept of CSE
with Pardis Miri, Meg Walraed-Sullivan, and Amin Vahdat
... and others of the PortLand project

Architecture

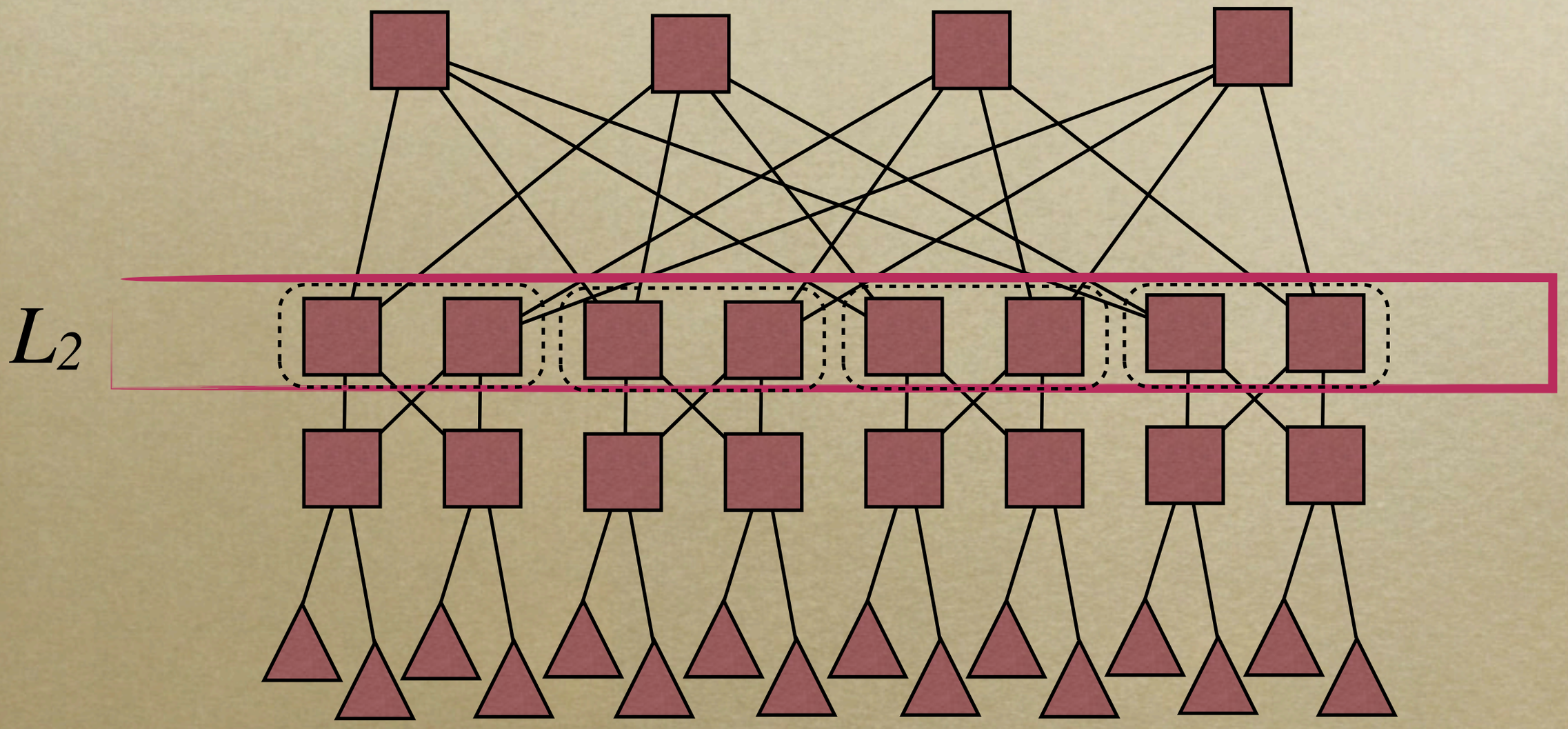


with three levels and k port switches: $5k^2/4$ switches, $k^3/4$ hosts

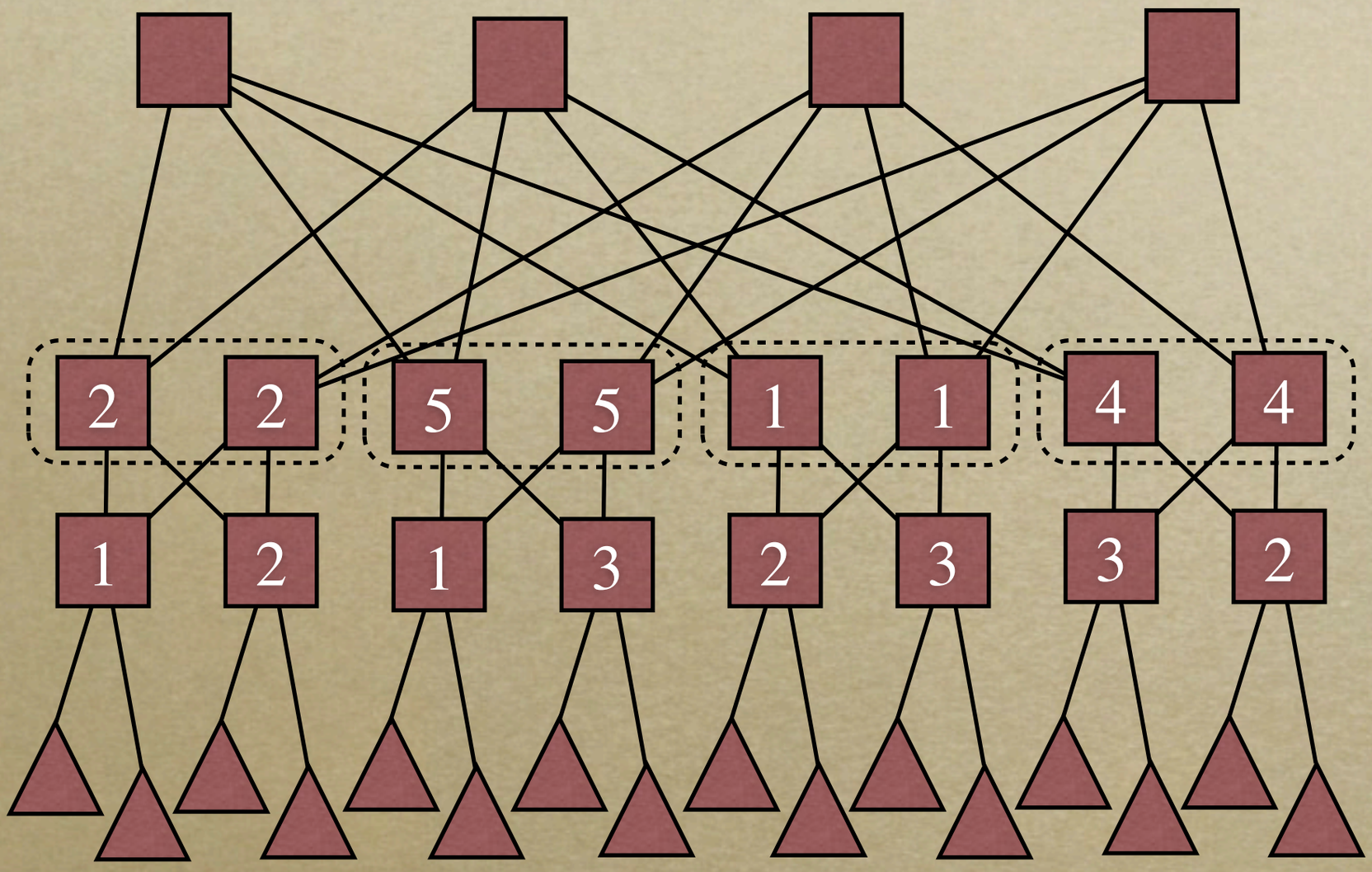
UNRP: Autonomic naming



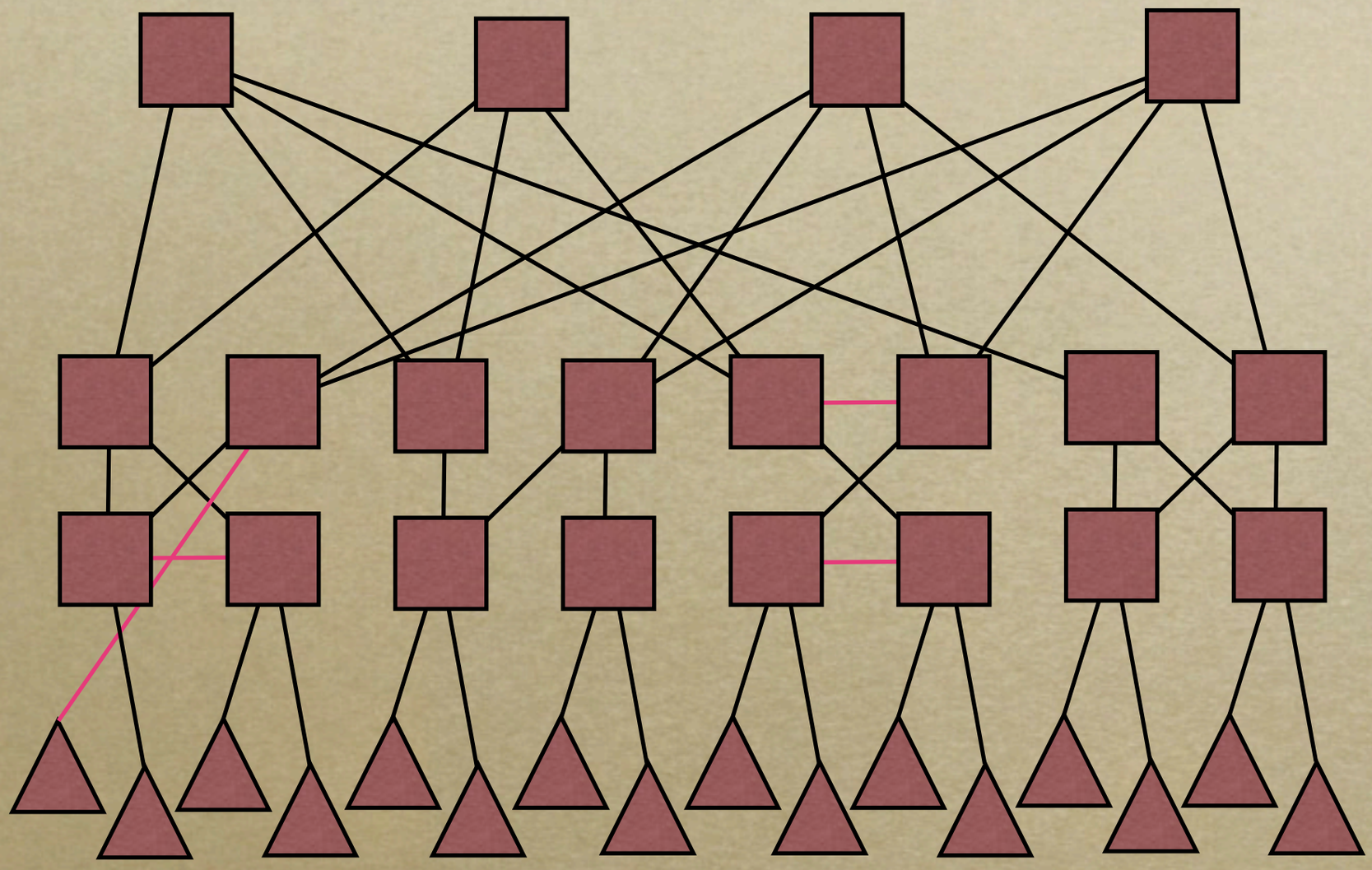
UNRP: Autonomic naming



UNRP: Autonomic naming



Miswiring



Modeling miswirings

- First idea: use a *threshold model*
- Assume graphs are *d-deficient*:
  - In $n \times n$ bipartite graph, each node has at least $n - d$ edges

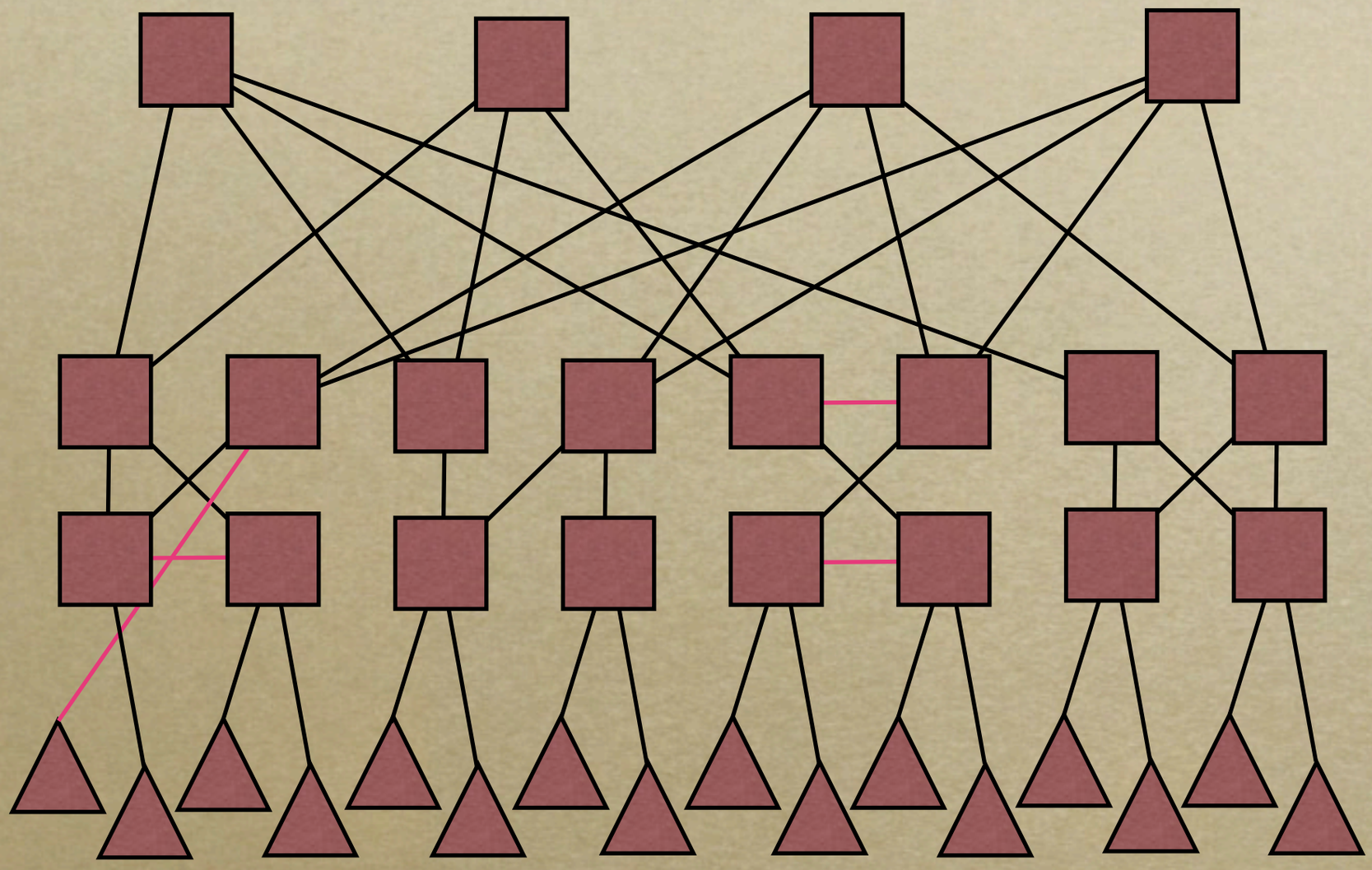
The good and bad with this model

- k-deficiency is a nice mathematical abstraction
  - n and k are large, and so a large d should indicate many failures => unlikely
  - diameter of bipartite graphs are easily bounded in terms of d-deficiency
- ... but we rejected it
  - hard to instantiate
  - issues of fabric startup

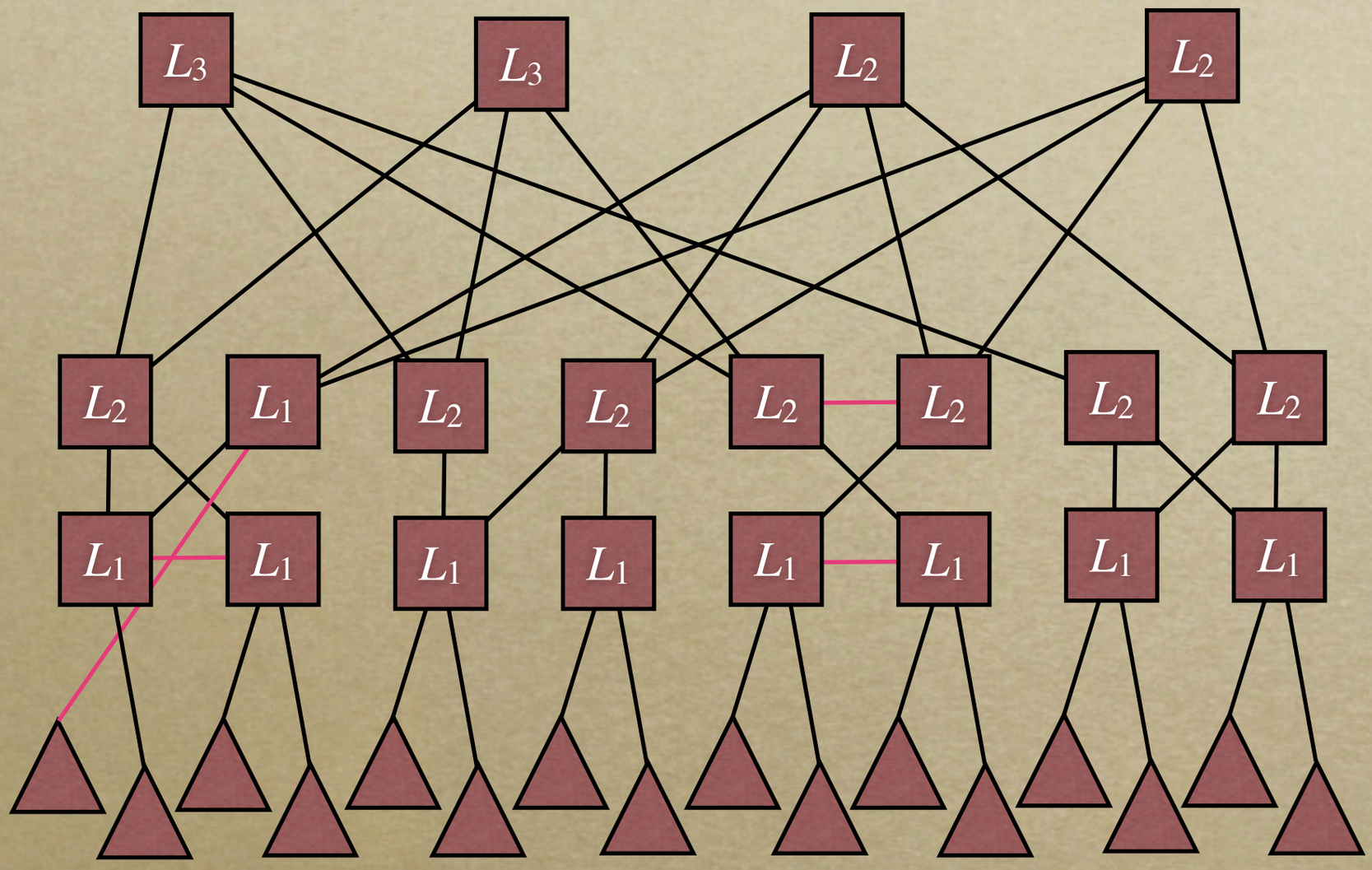
Model we used

- Assume ℓ levels; hosts connected to one (lowest) level
- Assign addresses quickly and using few messages
  - The closer to a fat tree, the better the addresses
  - Two hosts separated by more than $2\ell - 1$ won't be able to communicate with each other
  - Do diagnostics separately
- Transfer risk of loss of reliability from instantiation to quality of immediate result
  - Reminiscent of *graceful degradation*

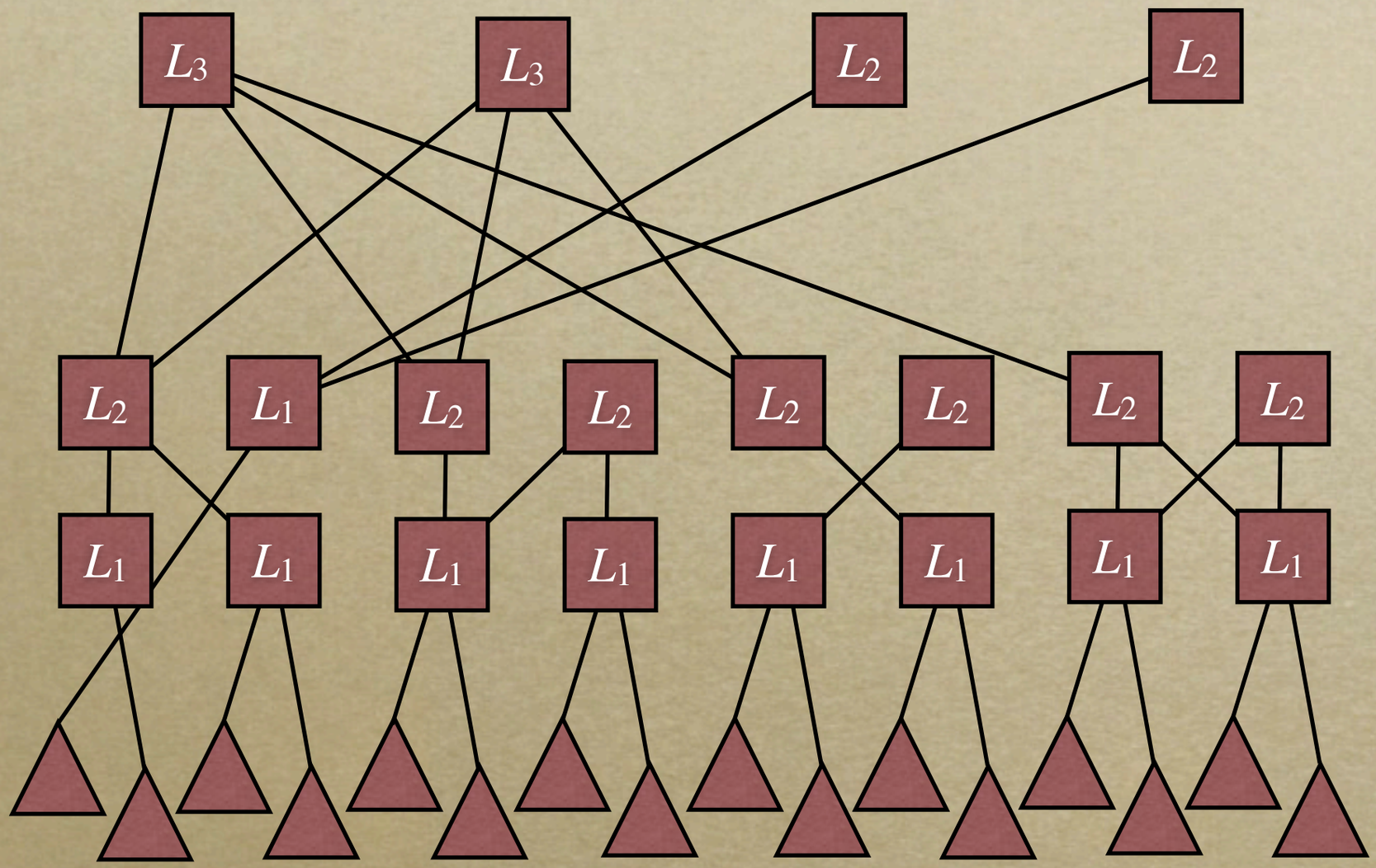
In action



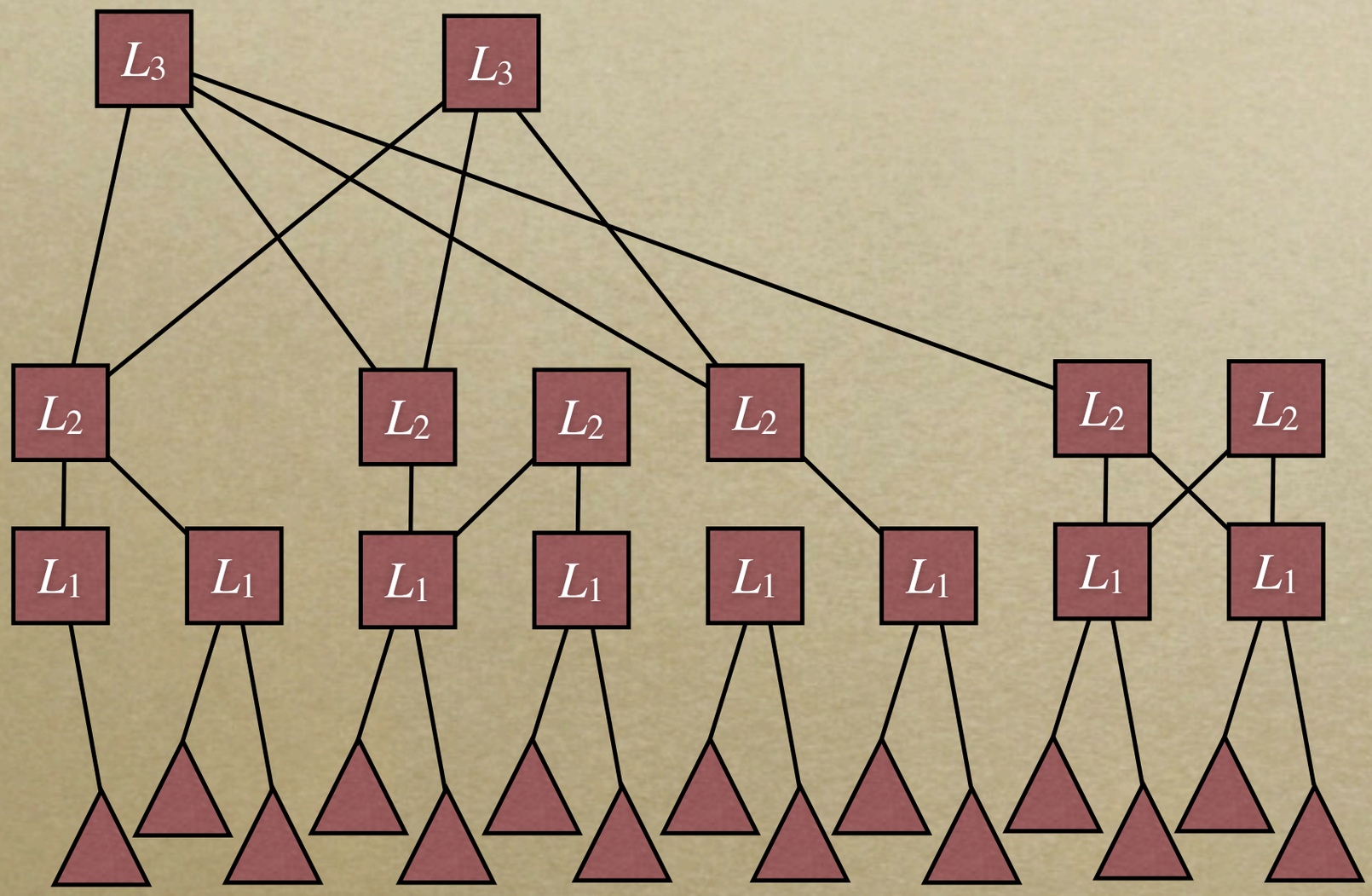
In action



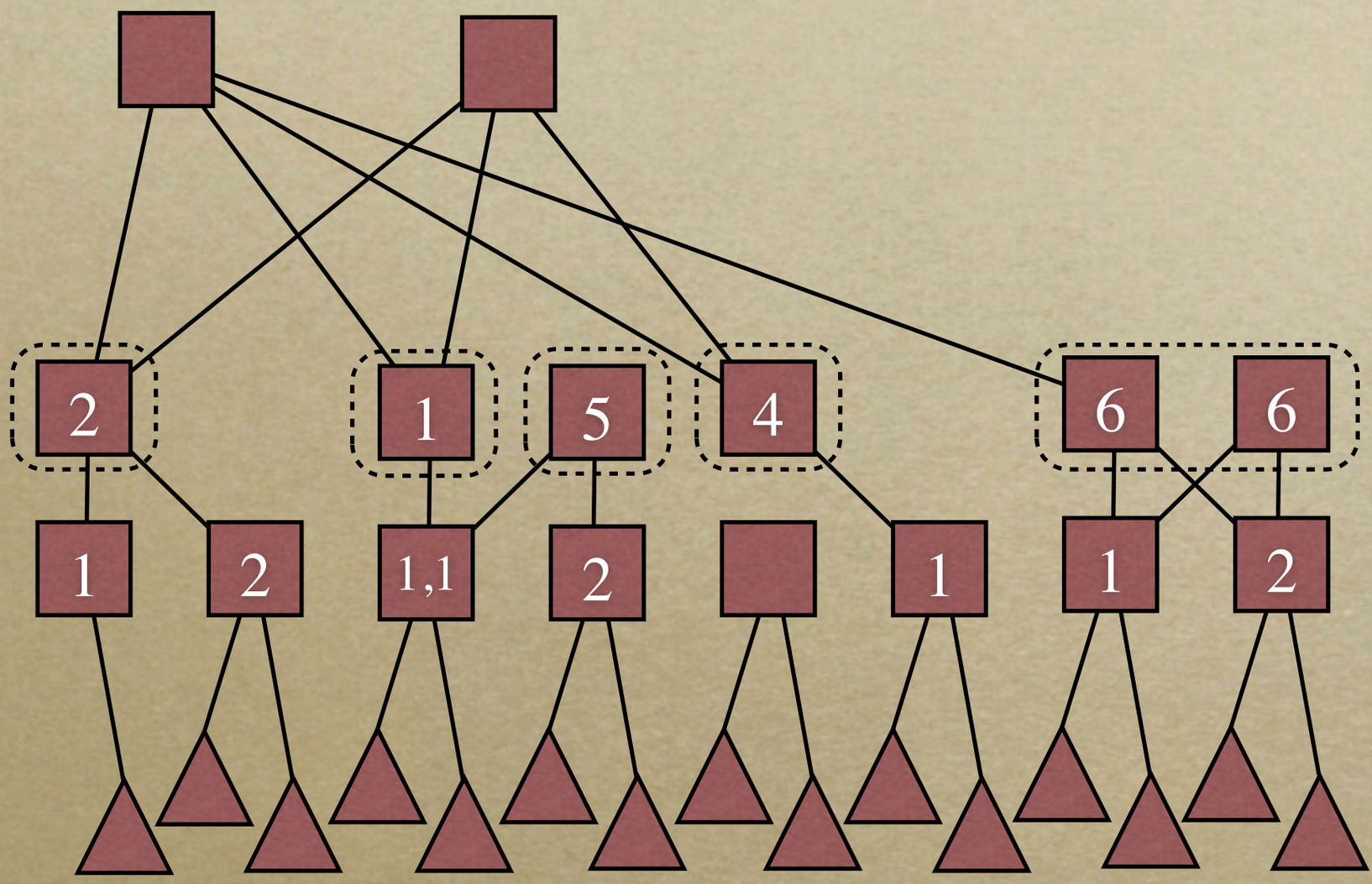
In action



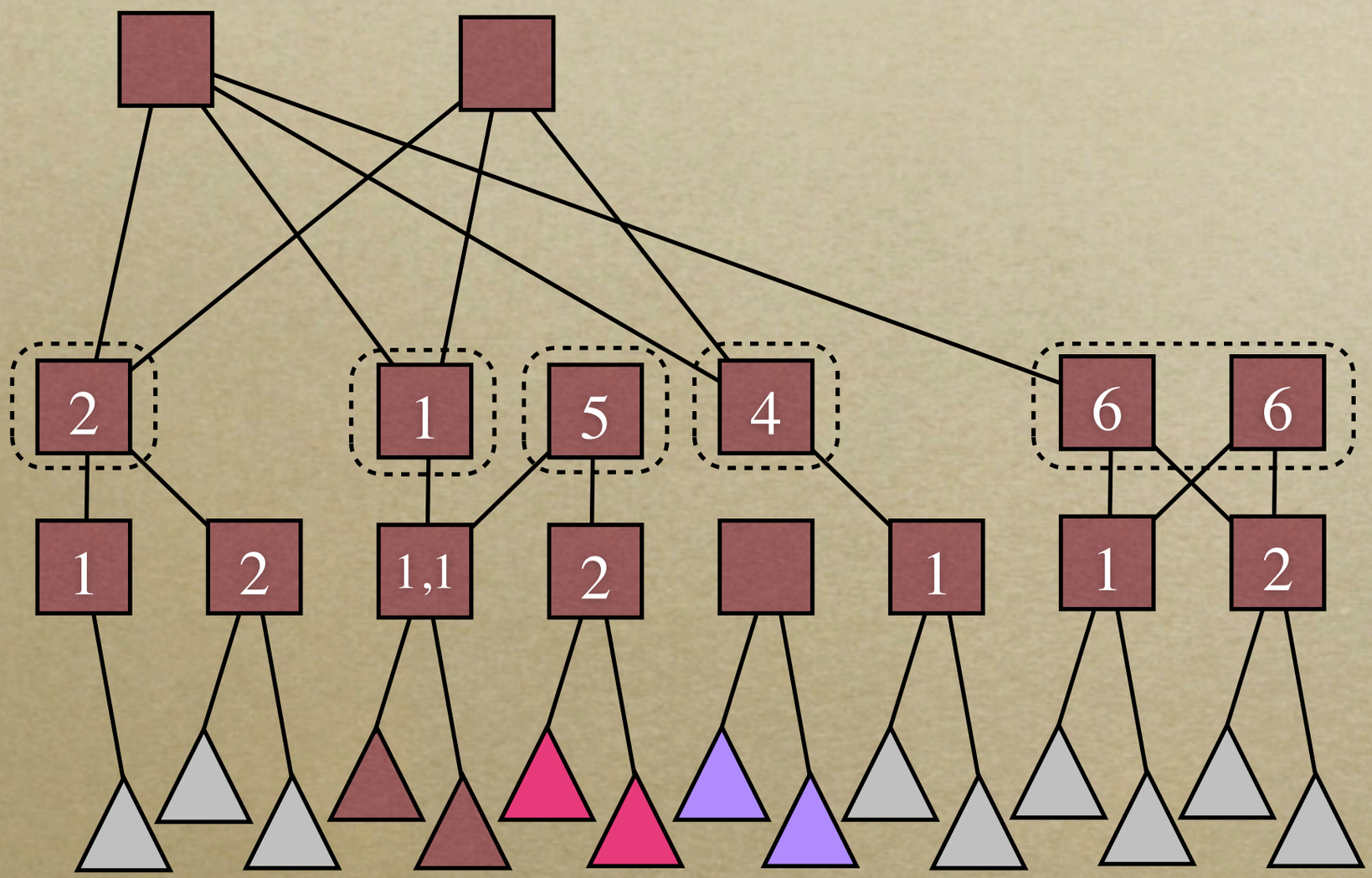
In action



In action



In action



Current work and questions

- Allowing for hosts to be connected at any level
  - Level assignment becomes harder
  - Miswiring model still appropriate?
- Diagnostics
  - Can we be more precise about *the closer to a fat tree, the better the addresses*?