Finding The Right Abstraction for a Modular Router: An Axiomatic Basis

S. Keshav and M. Karsten *
University of Waterloo
Waterloo, ON Canada

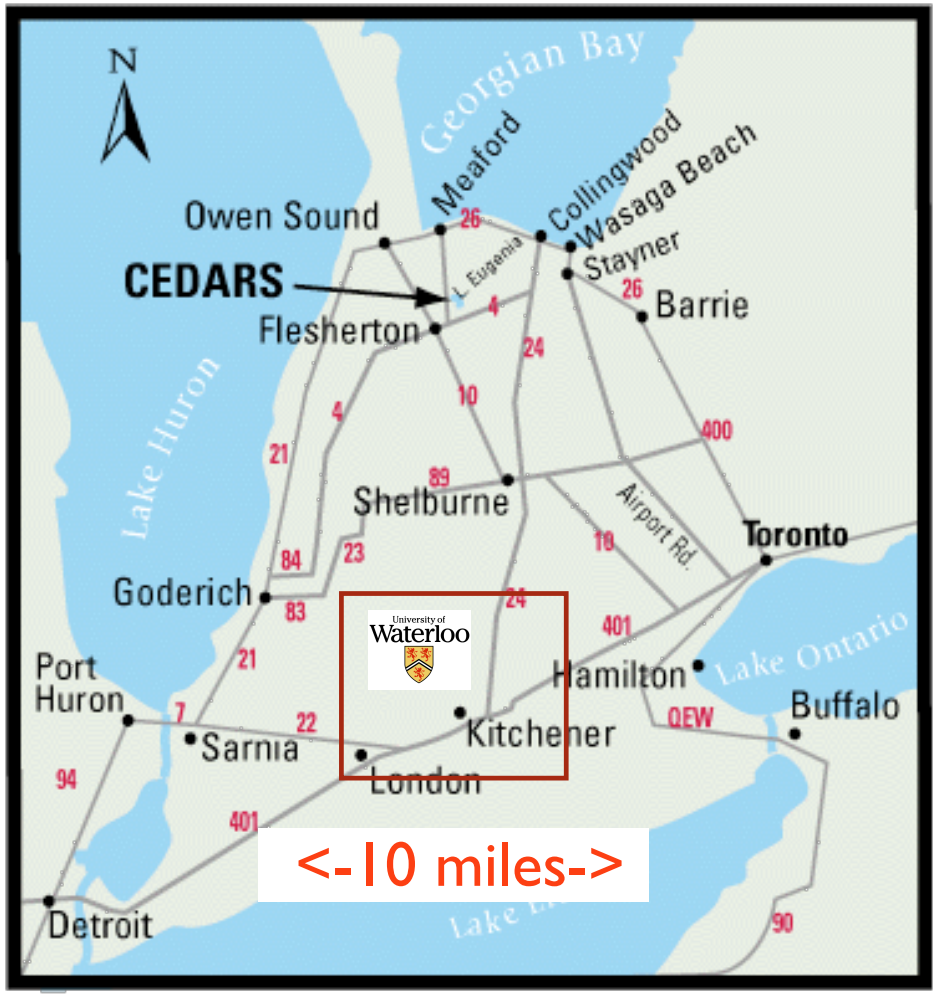
*Joint work with Sanjiva Prasad, IIT Delhi and Omer Beg, UW

Waterloo?

Where is that?


Canada
ARCTIC OCEAN
Ellesmere Island
Arctic Circle
GREENLAND (KALAALLIT NUNAAT) (Denmark)
ICELAND
0 250 Miles
0 250 Kilometers
Queen Elizabeth Islands
Beaufort Sea
Banks Island
Baffin Bay
60° N
ALASKA (U.S.)
Victoria Island
Baffin Island
YUKON TERRITORY
NUNAVUT
Labrador Sea
Gulf of Alaska
Great Bear Lake
Rocky Mountains
Coast Mountains
NORTHWEST TERRITORIES
ATLANTIC OCEAN
Mt. Logan 19,524 ft. (5,951 m)
Great Slave Lake
Hudson Bay
BRITISH COLUMBIA
NEWFOUNDLAND & LABRADOR
ALBERTA
MANITOBA
<-1000 miles->
PACIFIC OCEAN
SASKATCHEWAN
James Bay
Plateau
PRINCE EDWARD ISLAND
NEW BRUNSWICK
Edmonton
Vancouver I.
Calgary
Vancouver
ONTARIO
Laurentian
NOVA SCOTIA
Québec
Winnipeg
Halifax
Montréal
Ottawa
Bay of Fundy
40° N
Toronto
UNITED STATES
120° W
100° W
80° W
60° W
© MAPQUEST.COM


ONTARIO
KENORA
WINNIPEG
122 miles
LAKE OF THE WOODS
THUNDER BAY
452 miles
INTERNATIONAL FALLS
135 miles
LAKE SUPERIOR
SAULT STE. MARIE
710 miles
CANADA
MINNEAPOLIS
410 miles
U.S.A.
LAKE HURON
TORONTO
1,064 miles
LAKE ONTARIO
BUFFALO
1,200 miles
LAKE MICHIGAN
DETROIT
1,359 miles
LAKE ERIE
CHICAGO
749 miles
<-100 miles->
MILES KILOMETRES


N
Georgian Bay
Meaford
Collingwood
Wasaga Beach
Owen Sound
26
Stayner
CEDARS
L. Eugenia
Barrie
Flesherton
4
24
10
4
21
400
89
Shelburne
Airport Rd.
Toronto
Lake Huron
84
23
18
Goderich
24
83
University of Waterloo
401
Lake Ontario
21
Port Huron
Hamilton
7
22
Kitchener
QEW
Buffalo
Sarnia
London
94
401
<-10 miles->
Detroit
90

Local companies

Seagram
RIM/BlackBerry
MapleSoft
OpenText
ManuLife

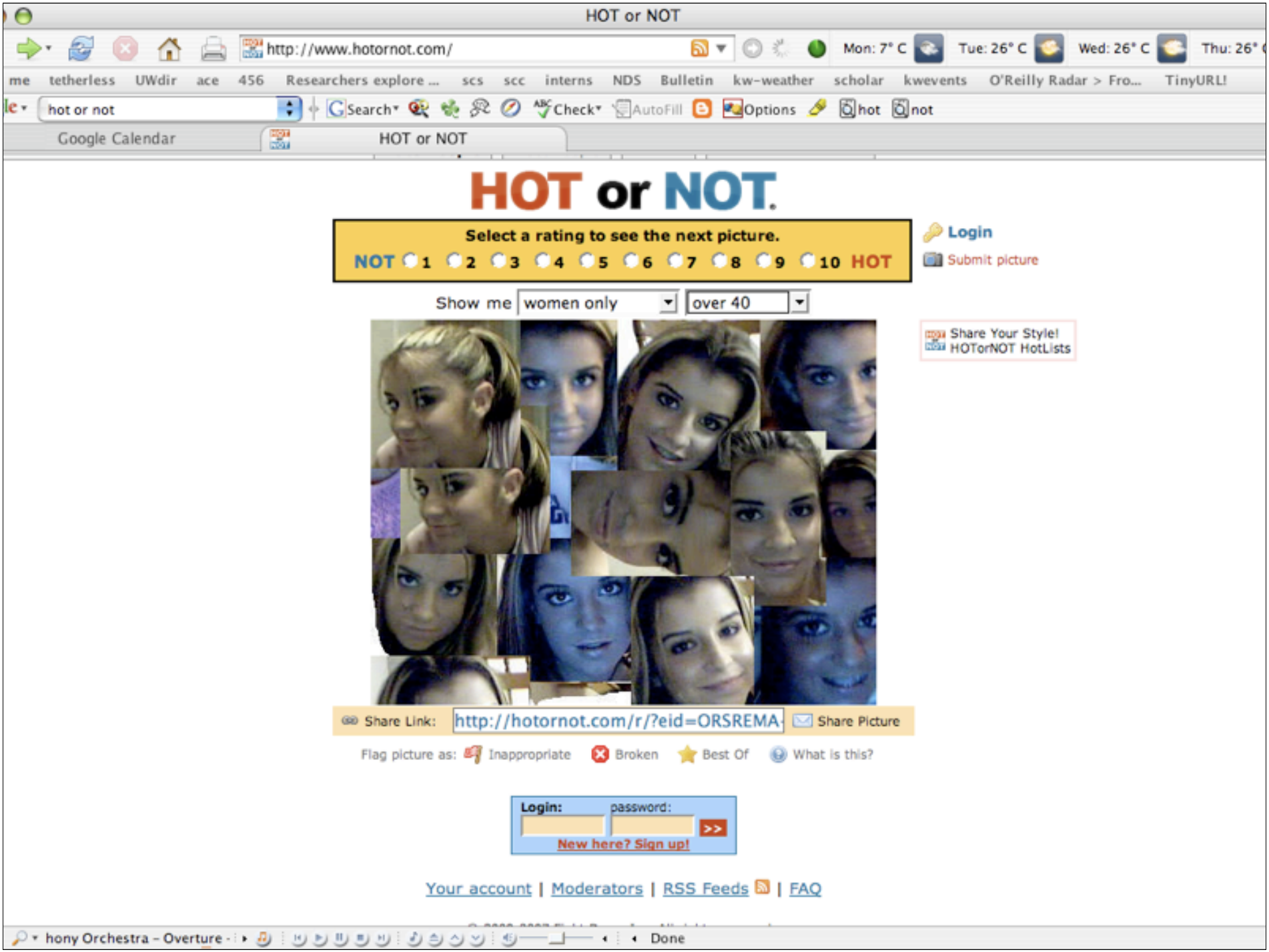
School of CS

77 faculty
~2000 undergrads
~300 grads

Finding The Right Abstraction for a Modular Router: An Axiomatic Basis

S. Keshav and M. Karsten *
University of Waterloo
Waterloo, ON Canada

*Joint work with Sanjiva Prasad, IIT Delhi and Omer Beg, UW


HOT or NOT
http://www.hotornot.com/
Mon: 7° C
Tue: 26° C
Wed: 26° C
Thu: 26° C
me tetherless UWdir ace 456 Researchers explore ... scs scc interns NDS Bulletin kw-weather scholar kwevents O'Reilly Radar > Fro... TinyURL!
hot or not
Search
Check
AutoFill
Options
hot
not
Google Calendar
HOT or NOT
HOT or NOT.
Select a rating to see the next picture.
NOT 1 2 3 4 5 6 7 8 9 10 HOT
Login
Submit picture
Show me women only over 40
Share Your Style!
HOTorNOT HotLists
Share Link:
http://hotornot.com/r/?eid=ORSREMA
Share Picture
Flag picture as:
Inappropriate
Broken
Best Of
What is this?
Login:
password:
>>
New here? Sign up!
Your account | Moderators | RSS Feeds | FAQ
hony Orchestra – Overture -
Done

University of Waterloo

Router or not?

Core router — Yes

Edge router — Yes

Enterprise router — Yes

Ethernet switch — Yes

Load balancer — Yes

NAT box — Yes

Firewall — Yes

Spam filter — Yes

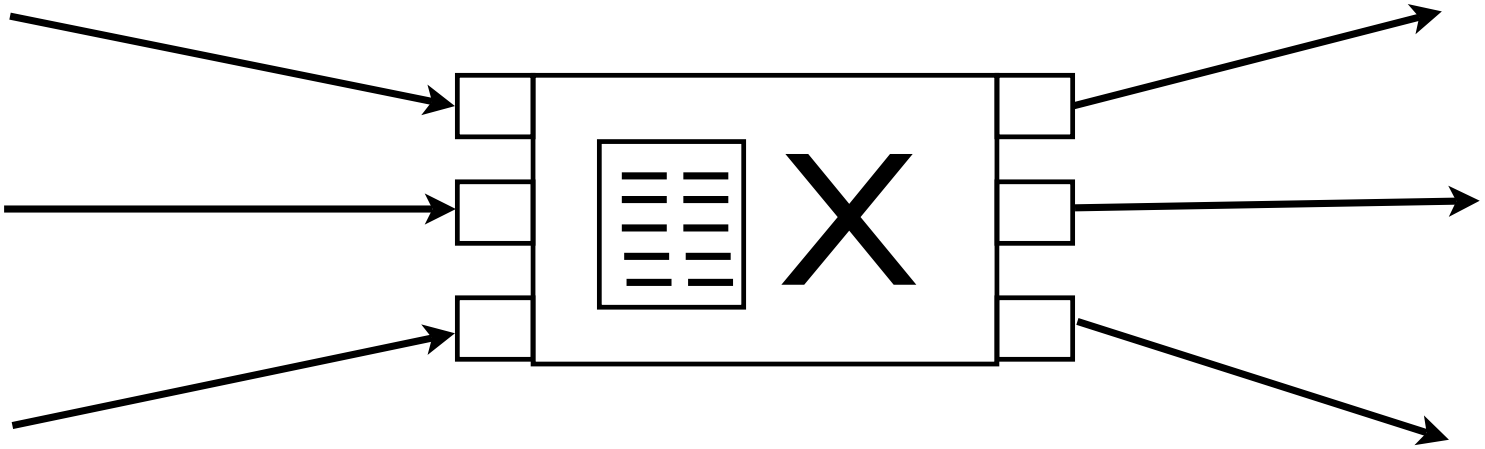
IP layer — Yes

Majordomo — Yes

Orkut member — Yes

Simplest possible router

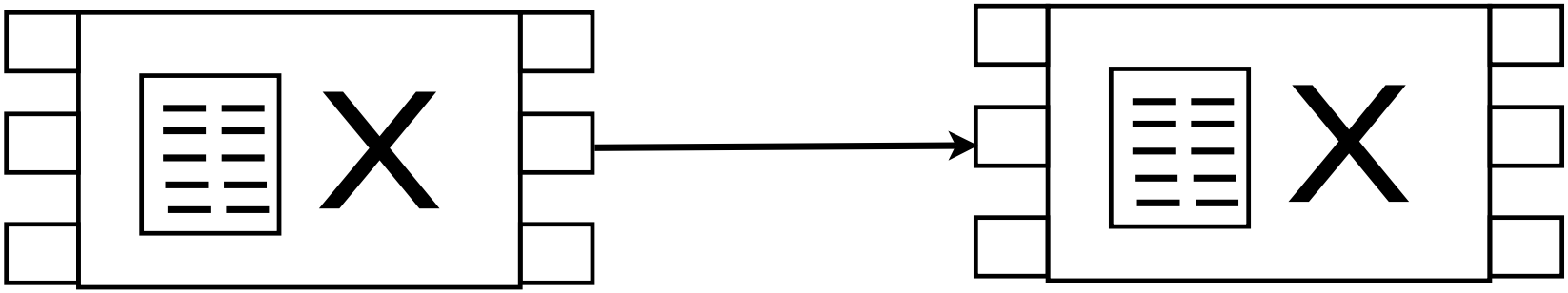
University of Waterloo



Abstract Switching Element (ASE)

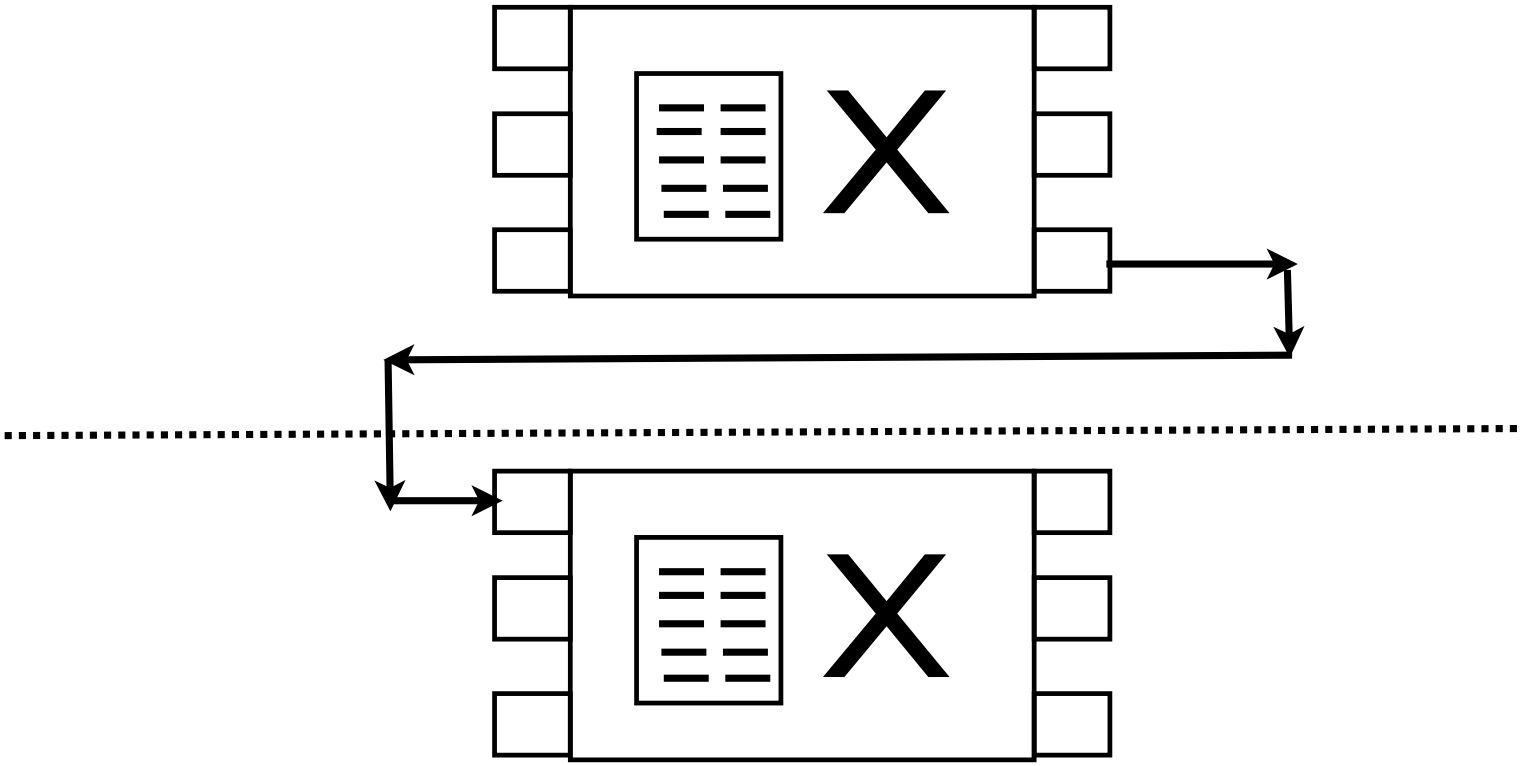
Direct connection

University of Waterloo

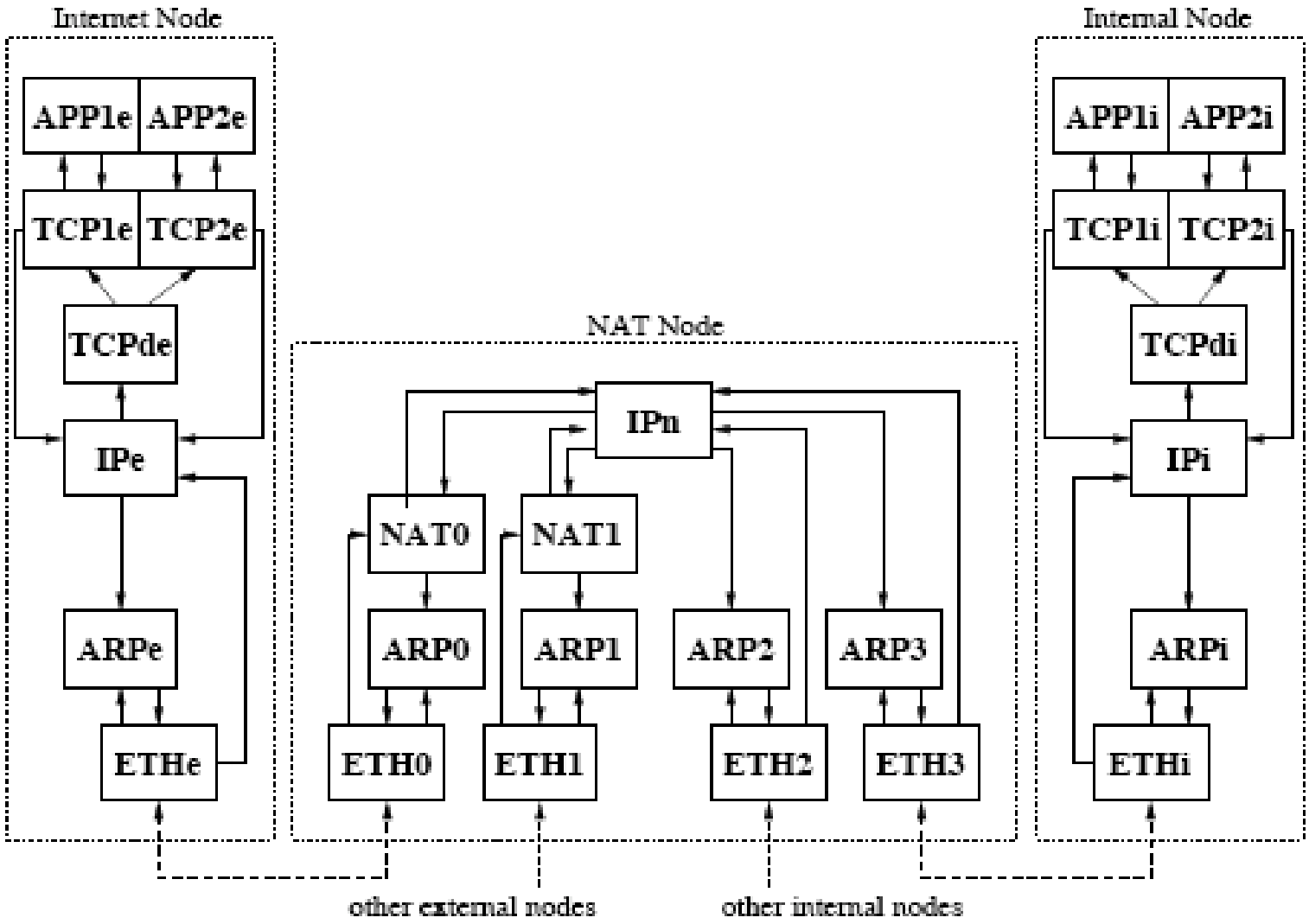


Another direct connection

University of Waterloo

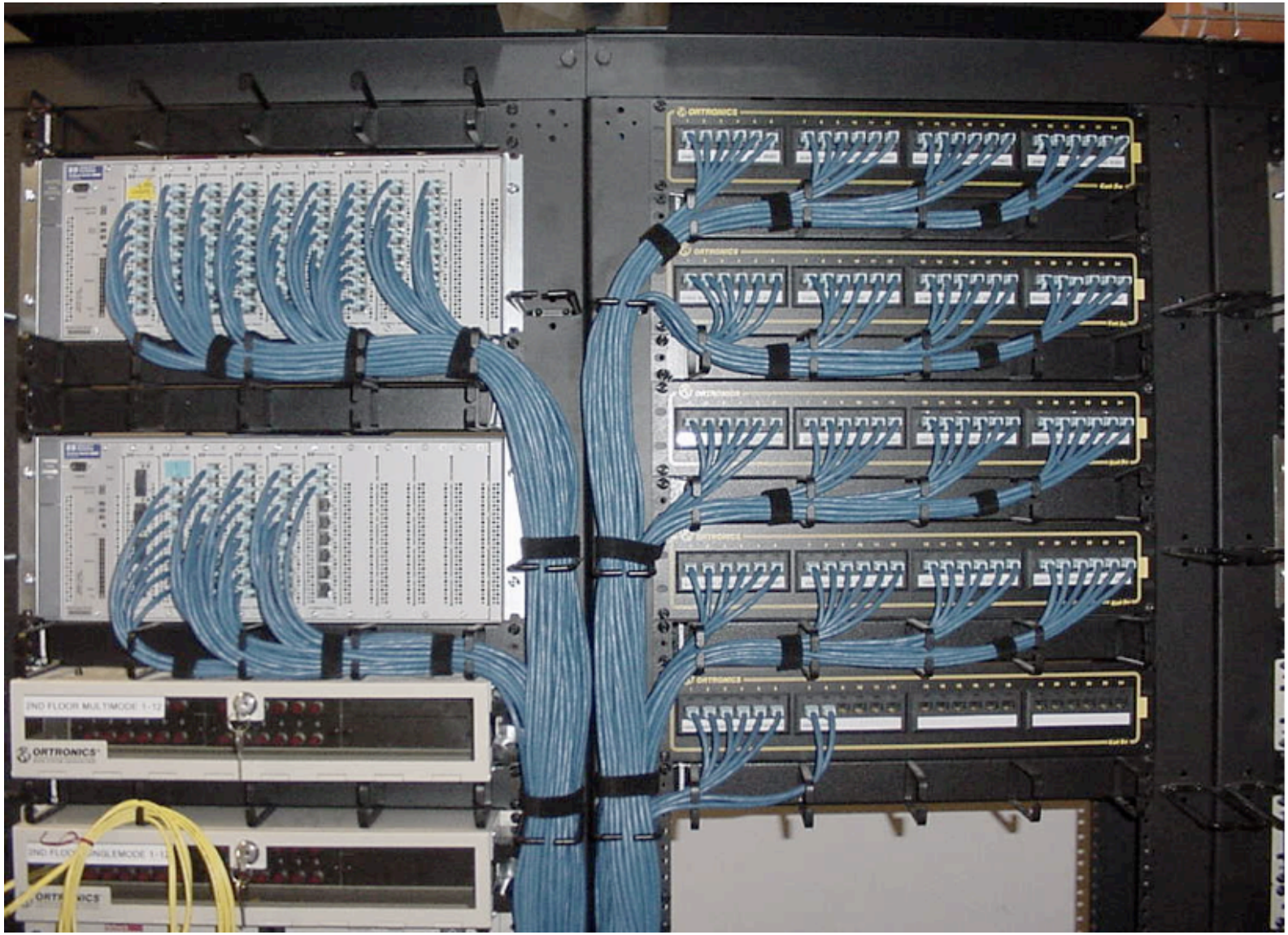


Combining ASEs



Theory

University of Waterloo



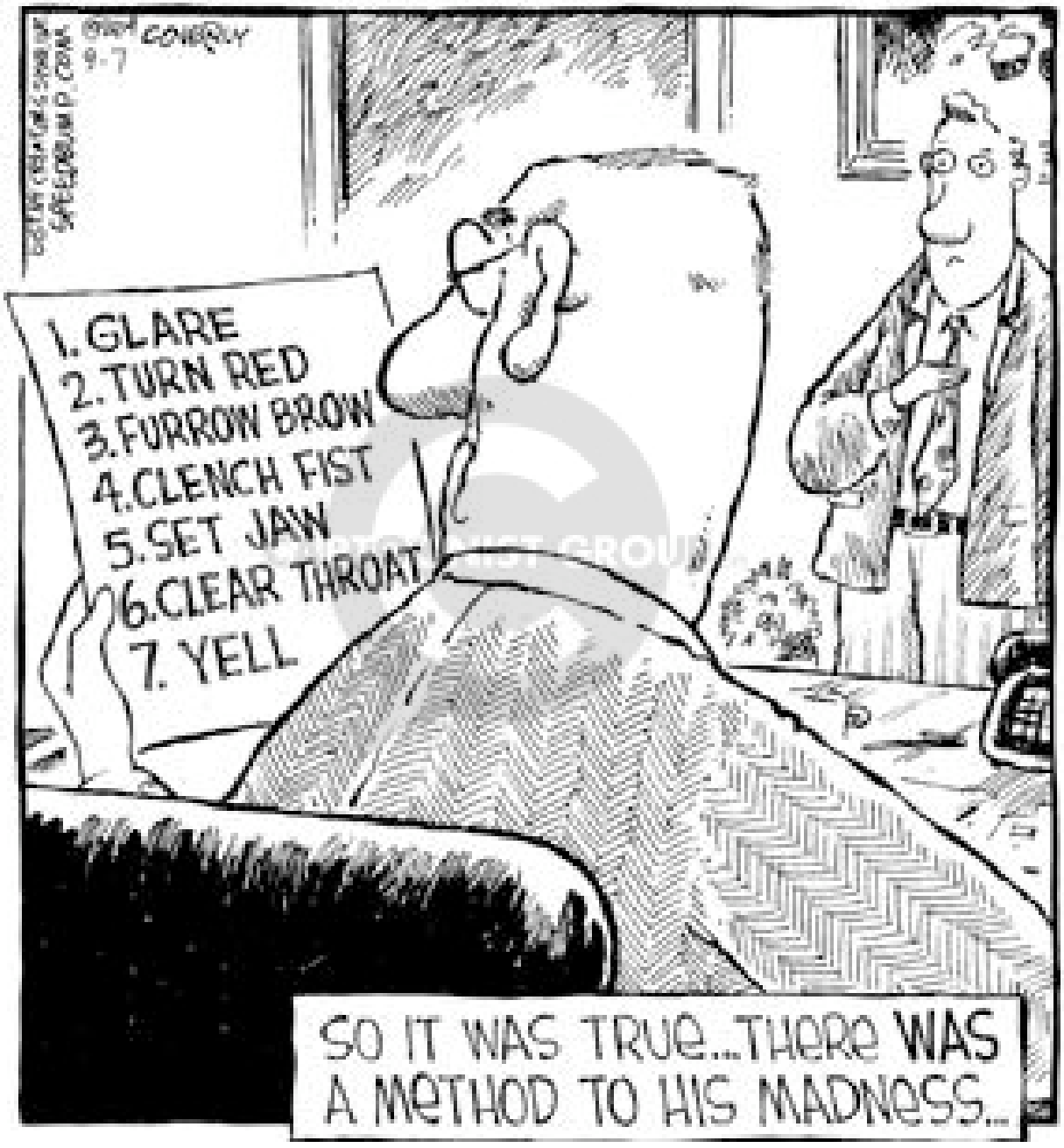
Reality

Reality

University of Waterloo

- Many layers: Socket, Session, TCP, MPLS, IP, PPP, VLAN, ATM, MAC
- Middleboxes: Firewalls, NATs, shapers, load balancers, DNS redirectors...
- Cross-layered designs

Yet... it works!



Axioms

University of Waterloo

-> = 'leads to'

- Direct communication

  m@output port -> m@input port

- Simple switching

  if p is an an ASE's switching table with a translation p',
  pm@input port -> p'm@output port

- Transitivity

  if m@input port of ASE A, and tables at A, B, C, ... , K are set up properly, m will reach output port of K

Concepts

University of Waterloo

- Name

  if pm@input port of ASE A, and tables at A, B, C, ... , K are set up properly, m will reach output port of K, and *p is its name*

- Address

  if two ASEs send a message with a given name to the same destination, the name is also an address

- Name scope

  set of ASEs where a name leads to the same destination

- Routing

  process of maintaining consistent forwarding within a particular naming scope

Data plane primitives

University of Waterloo

- *push*
- *pop*
- *swap*
- *send*
- *receive*
- *copy*

Control plane primitives

University of Waterloo

- Essentially manipulate the state table
- *update the state table*
- *get label from a control message header*
- *set label in a control message header*

- + a few other minor details

Universal forwarding loop

University of Waterloo

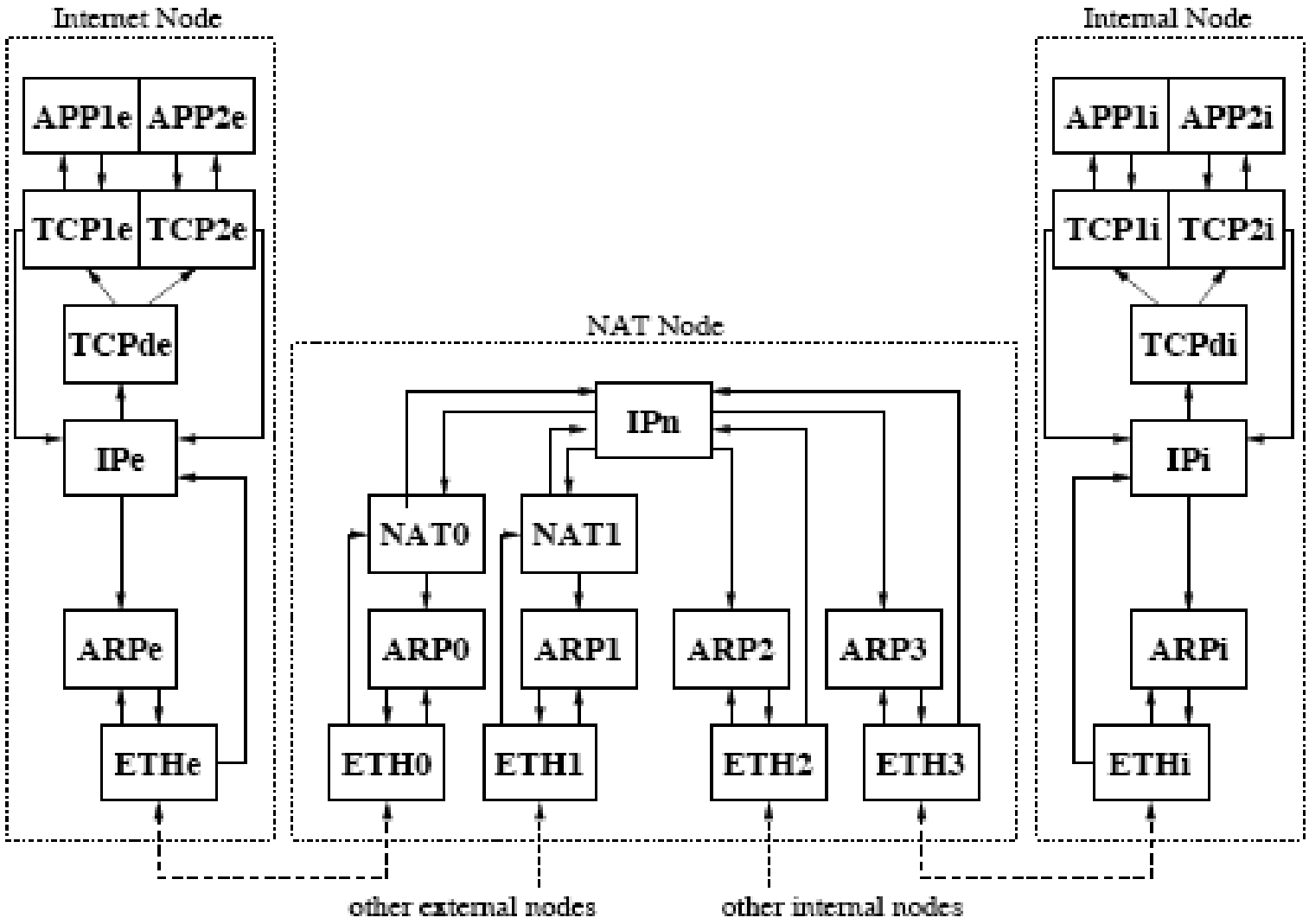
```
1   process(ase prev, message msg) {
2     bool setup = (ctl(msg) == SETUP
3                || prev in this->SETUP_ASE);
4     name lin, lout;
5     if (setup) lin = lout = getlabel(msg);
6     name n = pop(msg);
7     {<ase, name>} S = lookup(prev, n);
8     if (!S && this->RESOLVE_ASE) {
9       resolve(n); // wait for S update
10      S = lookup(prev, n);
11    }
12    for each <ase, name> s_i in S {
13      if (s_i.ase == this) { // local
14        if (ctl(msg) == RLOOKUP) {
15          respond(prev, msg, n, s_i.name);
16        } else if (ctl(msg) == RUPDATE) {
17          rupdate(msg);
18        } else {
19          // other local control activity
20        }
21      } else { // forward
22        message outmsg = copy(msg);
23        push(outmsg, s_i.name);
24        if (setup) {
25          if (VC) lin = local_name(prev, n);
26          update(s_i.ase, lin, prev, lout);
27          setlabel(outmsg, lin);
28        }
29        send(s_i.ase, outmsg);
30      }
31    }
32  }
```

What can we do with this?

University of Waterloo

- Can build any forwarding system as a composition of specializations of the universal forwarding loop
- Can formally verify the correctness of any router using Hoare logic

NAT


Internet Node
APP1e
APP2e
TCP1e
TCP2e
TCPde
IPe
ARPe
ETHe
NAT Node
IPn
NAT0
NAT1
ARP0
ARP1
ARP2
ARP3
ETH0
ETH1
ETH2
ETH3
Internal Node
APP1i
APP2i
TCP1i
TCP2i
TCPdi
IPi
ARPi
ETHi
other external nodes
other internal nodes

Building a router