

**Reaching the Unreached**  
***Bridging the technical and  
cultural network architecture  
divide***

Kevin Fall

Intel Research, Berkeley

[kfall@intel.com](mailto:kfall@intel.com)

SIGCOMM/NSDR

27-Aug-2007

Kyoto, Japan

# How We Got Here

- Once upon a time, we had a “digital” telegraphic network that carried messages
- And then came the telephonic network
  - Obsoleted the telegraphic network
- And telex and FAX
- And broadcast TV and radio
- And the personal computer
- And the ARPANET, and Internet, and e-biz

# So What?

- ARPANET came from the US DoD
  - Try packet switching for robust nuclear C<sup>2</sup>
  - Statistical multiplexing as side economic benefit
- Internet too
  - Try to interconnect disjoint networks
  - Goal: IPC across heterogeneous networks
- But IPC is all about the “pipes”
  - So we multiplex at the network layer, but...
  - The virtual circuit model still drives the economics

# Internet Abstractions

- Names – most involve hosts
- Addresses – hosts, ports supporting IPC
- Connections – 2-endpoint IPC channel
- Hosts – things “at the edge” (vague)
- Routers – “not hosts” (also vague)
- Anomalies (middleboxes) – warts
- Byte streams – “water”

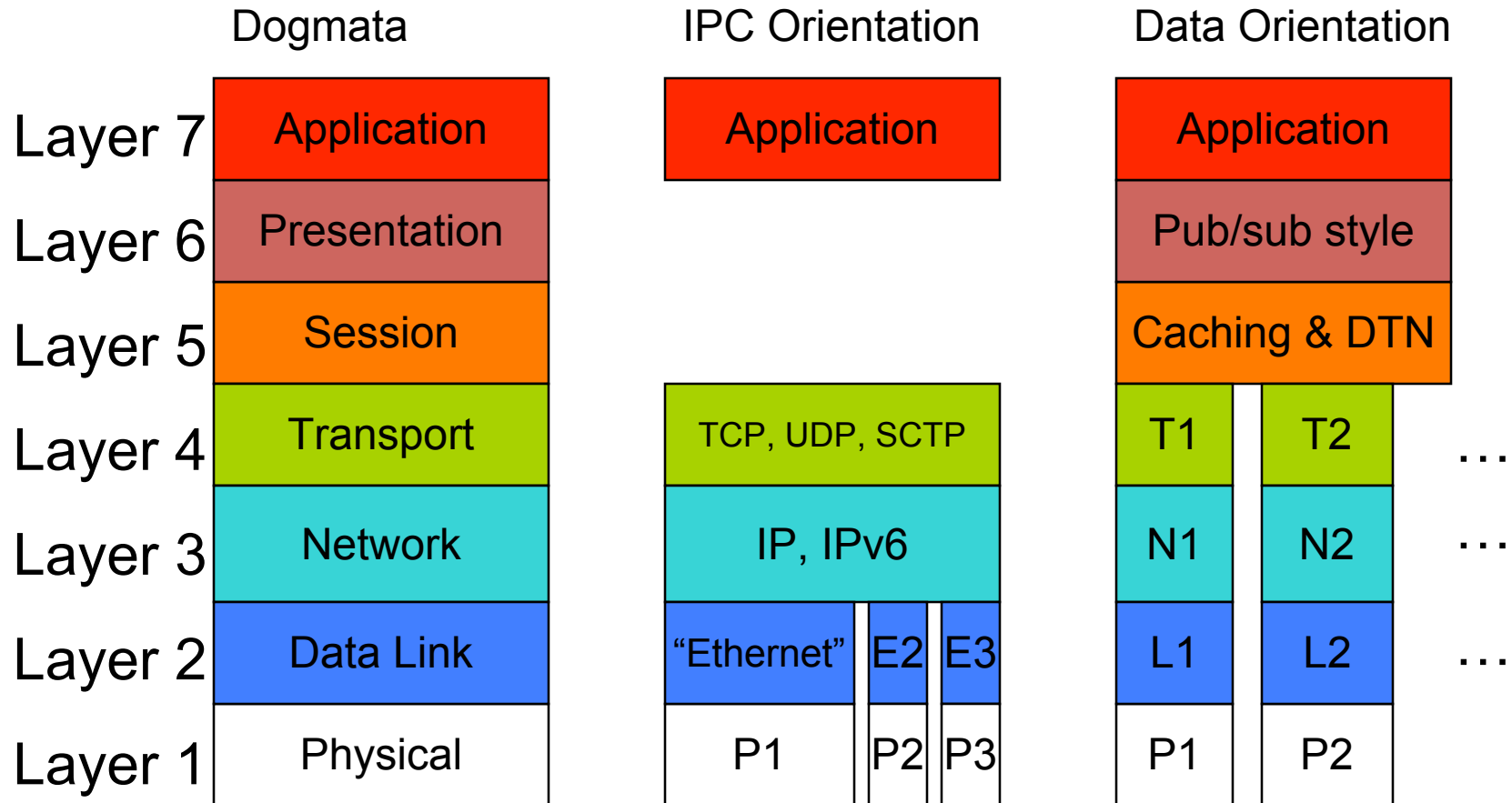
# Back to the Unreached

- Why are people “unreached” by pipes?
  - Used to be technical (interop), now economic
  - It’s packets vs circuits again, this time at a layer up
- TCP/IP supports heterogeneity
  - But it’s a “virtual” circuit network
  - So, even w/packets, It’s pseudo-synchronous
    - And when we’re driven by VOIP constraints, then...
    - And thus not cost effective for the remotest places
- Consider the benefits of breaking this mold

# “Packets” at the service layer

- It's not about plain water (IPC), but data
  - Not so new... (message switching; ADU's)
    - But there exist other factors now (economic)
  - Move (stat) mux'ing up to service layer
- Consequences
  - Storage: tailor connectivity to right time/place
  - Embrace even more heterogeneity
  - Re-consider application interface

# One direction...



# The Unreached Again

- By next year, 50% of the population will live in (connected) urban environments [UN]
- About 1.1B Internet users today [IWS]
- Yet, many of these are *culturally* unreached
  - Due to parts of the architecture
  - Due to language

# Lingua Internetica

<http://global-reach.biz/globstats/evol.html>

# The Architectural Cultural Divide

- Much of the infrastructure is English/ASCII
  - Protocol descriptions, DNS, (IETF generally)
  - characters, IDs, searching, comparison
  - Display / rendering fonts
- Just separate the plumbing from the water
  - Just switch everything to unicode, no problem!

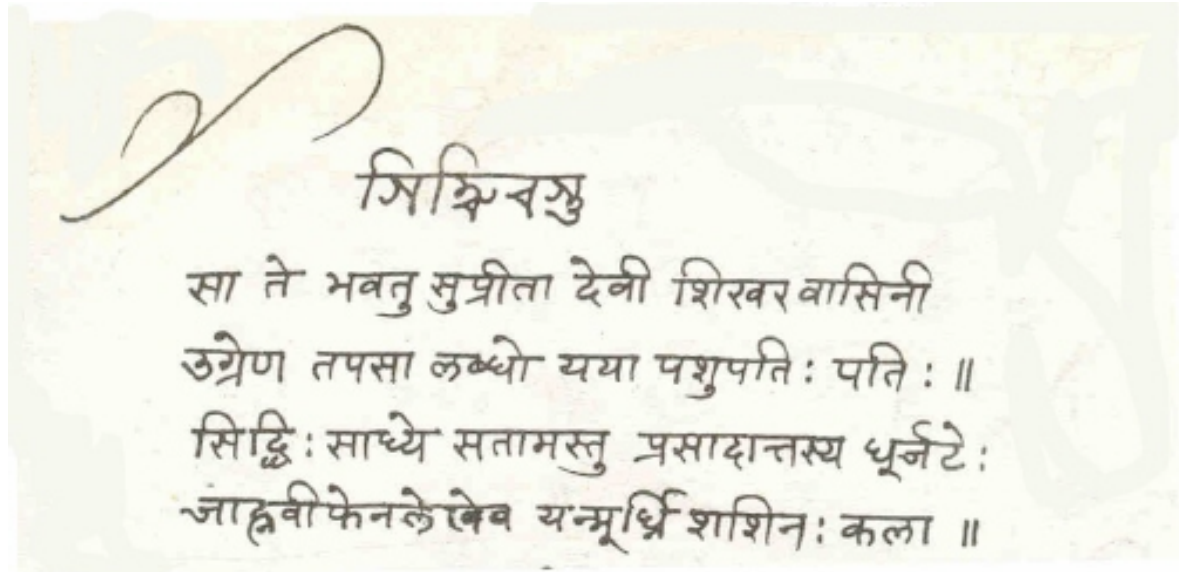
# Unicode, you say?

- Universal character set (to 1M characters)
  - An evolving standard (v 5.0 current, 1 year ago)
    - But really aka ISO 10646 and UCS
  - Plane 0 (0x0000-0xFFFF) : basic multilingual plane
    - Most of world's languages, incl CJK(V)
    - Many symbols (money, dingbats, etc)
  - Plane 1 (0x10000-0x1FFFF) supplementary MP
    - Historic scripts, musical and math symbols
  - Plane 2: “unified” HAN (CJK)
- A bit complicated, but look what you can do...

# Unicode examples

འཇམ་དབྱངས་

Phagspa (Tibet, 1200s)



Tirhutā (Bihar area)

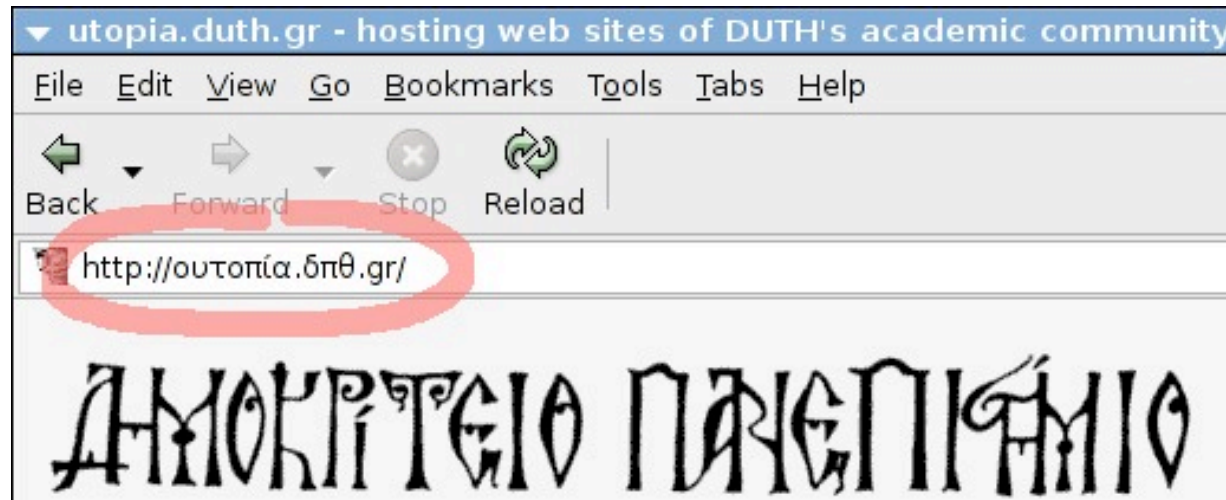
0 1 2 3 4 5 6 7 8 9

Limbu (India, Nepal)

- Fine, but you need to *encode* unicode

# Unicode in Domain Names

- IDNA – Int’l Domain Names in Apps [RFC3490]



- *Bücher.ch* becomes *xn--bcher-kva.ch* [punycode]
- Beware IDN homoglyph attacks

# Unicode in URL/URI/IRIs

- Good ol' days: simple URL's
  - Could use %-escapes for non-ASCII
- URLs extended to URI's
  - More formal structure, schemes, etc.
  - Need way to encode them: UTF-8 [RFC3986]
- Dealing with internationalized strings
  - “Stringprep” [RFC3454] and “Nameprep” [RFC3491]
  - Collation registry [RFC 4790]
- IRI's define comparisons and “Bidi IRIs”
  - Like URI's except non-URI chars ok where %xx used

# IRI example [RFC 3987]

For example,

"http://r&sum&.example.org"

may be converted to

"http://xn--rsum-bpad.example.org"

instead of

"http://r%C3%A9sum%C3%A9.example.org".

# Conclusions

- Some people are unreached
  - Technical reasons (plumbing problems)
  - Cultural reasons (water problems)
- Architecture switching from IPC to data
  - Better matching of user intent to network
  - Opportunity for better cost / performance
- The architectural cultural divide is there
  - Not easily displaced
  - Complicated and poorly understood

