Disclaimer

- A recent conversation with my three year old daughter:

  D: Daddy, what a big poop.

  Me: Yes sweetheart, daddy's really full of it.

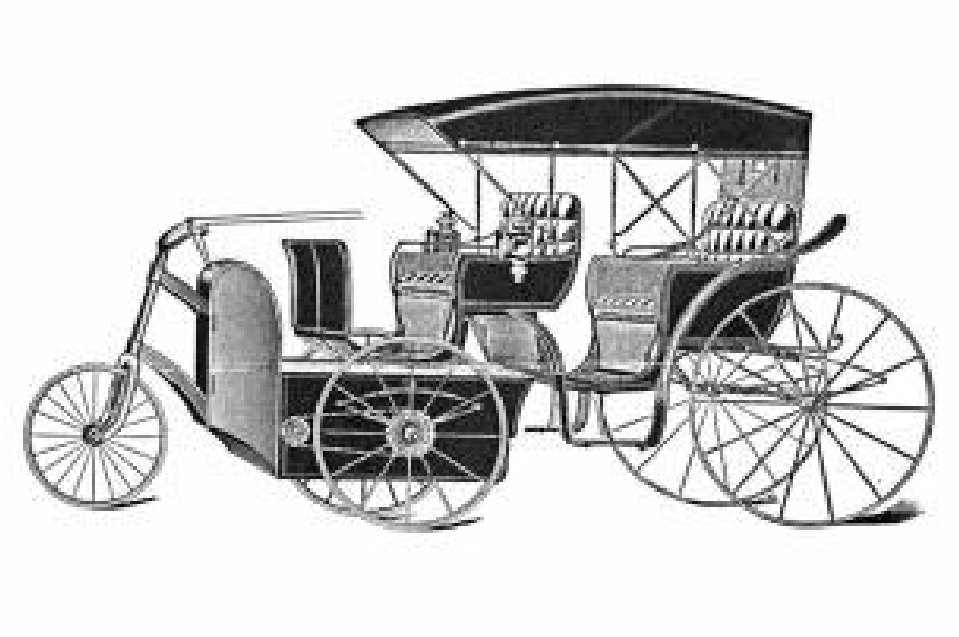
Circuits: the search for a cure

Van Jacobson

31st October 2001
Sigcomm ’01
San Diego, CA

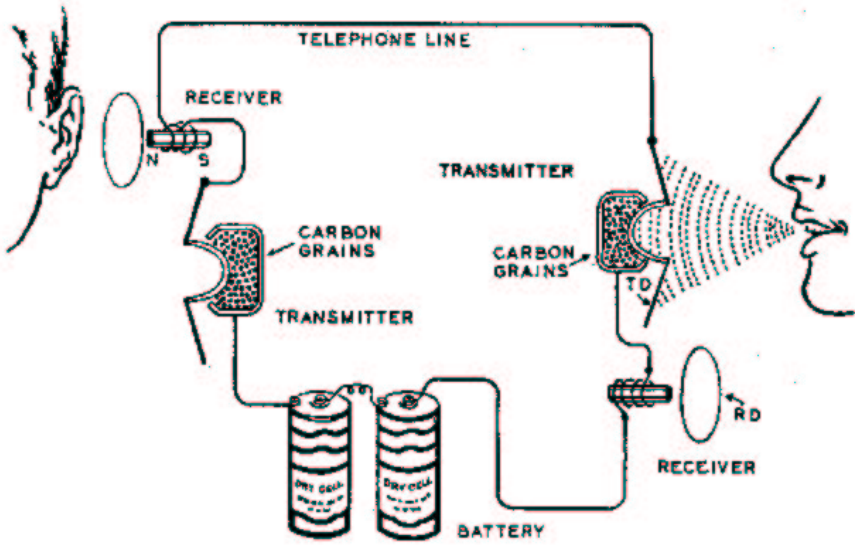
We model the future on the past.

Sometimes that's a mistake.



Early cars were horse carriages with motors and a tiller for steering. The 1898 Phelps Steamer was more user friendly—it steered with reins.

Source of the affliction



- Indoctrination starting with high school science books makes us believe there is something simple and fundamental about the telephony circuit model.

- In fact the telco circuit system is an extremely complex artifact whose evolution was driven by the engineering and economic challenges of the late 1800s and early 1900s.

Some history

1864 J.C. Maxwell presents the fundamental equations of electricity to the Royal Society.

1876 A.G. Bell invents the telephone.

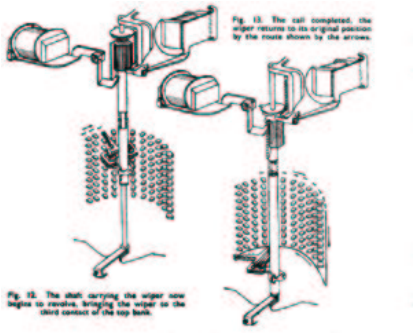
1878 First commercial switchboard starts operation serving 8 lines and 21 telephones.

1881 Bell Telephone patents the “metallic circuit” (two wires from CO to each phone rather than one wire connecting many phones).

1891 First metallic circuits deployed; start of PSTN.

History (cont.)

1891 Almon Strowger, a Kansas City undertaker, patents the first automatic dial system.



1919 Bell starts to switch from human operators to Strowger's "steppers"

1948 Transistor invented.

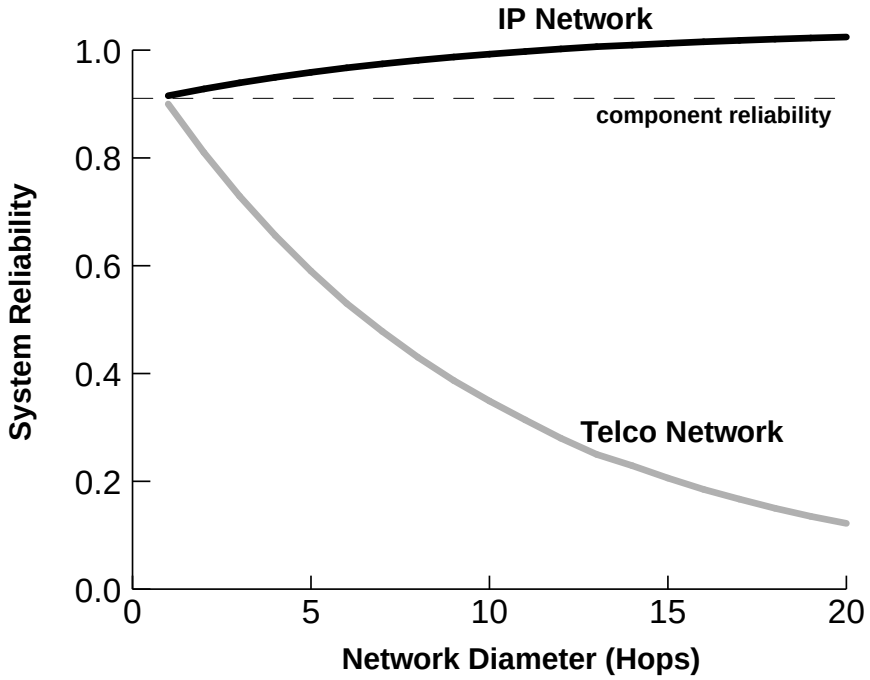
1963 Digital T carrier system introduced.

1965 No.1 ESS introduced.

The telephone system is a profound engineering achievement that has changed the world for the better. But the engineering imperatives of the 1800s are a bad match to the technology of 2001. *In almost every way* the circuit model is a poor fit to today's networking:

- it is structurally unreliable.
- It imposes a symmetry that encourages centralization and monopolization.
- it mandates a small set of globally unique service offerings that don't match existing usage.
- it has an excess of state that vastly complicates the solution of routine operational problems like mis-connection or traffic engineering.

Reliability

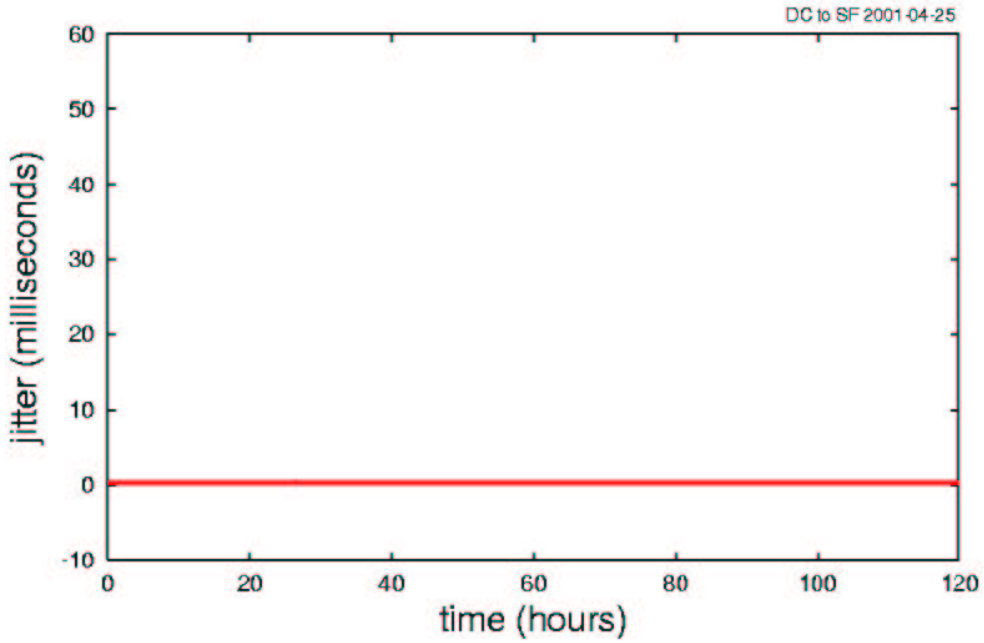


Circuit systems require very high component reliability since system reliability decreases exponentially with number of elements in series.

IP networks can be created from very low reliability components since alternate paths cause system reliability to increase exponentially with system size.

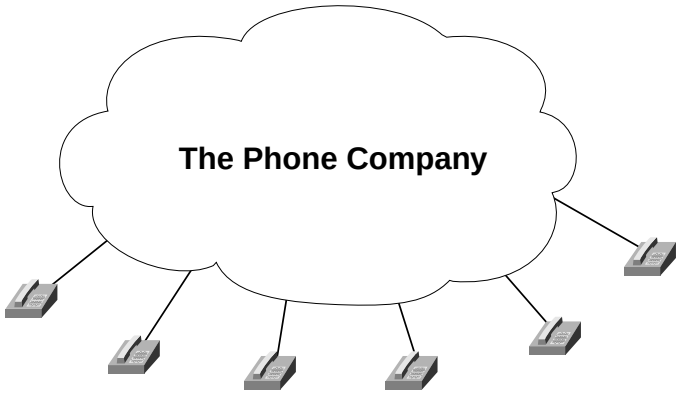
Best Effort ≠ 'Who Knows?'

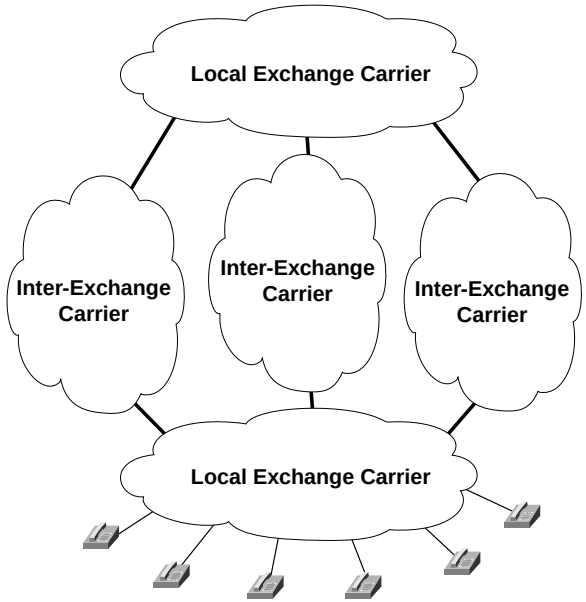
We recently measured the busiest transcontinental (SF to DC) core path of a large, tier-1 ISP. Our test rig sent randomly sized probe packets at exponential intervals (1ms avg.) with departure and arrival hardware time-stamped to 20us accuracy. This is the data from one week:



69 million probe packets were sent, zero were lost, worst case jitter < 700us.
(see Casner, Alaettinoglu and Kuan talk at NANOG 22, May 2001).

Symmetry

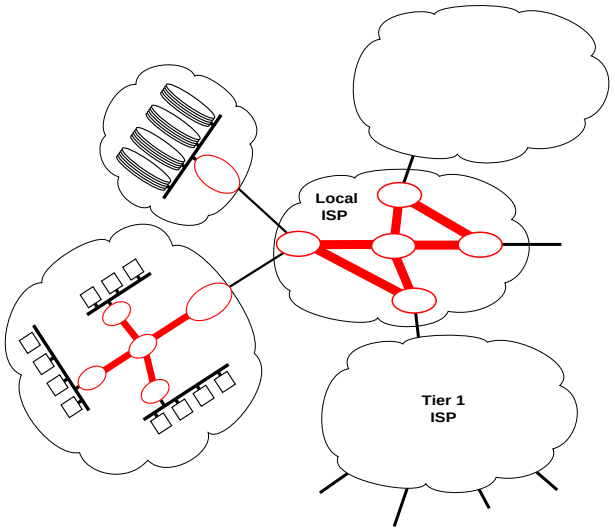




TelCos evolved from a single monopoly carrier to local monopolies with competitive inter-exchange (long distance) carriers. But all the parties deal in the same unit of service — a “call” (a brokering economic model). This sameness means there is a well defined *global* partition function which can be used to give a global meaning to “fairness”.

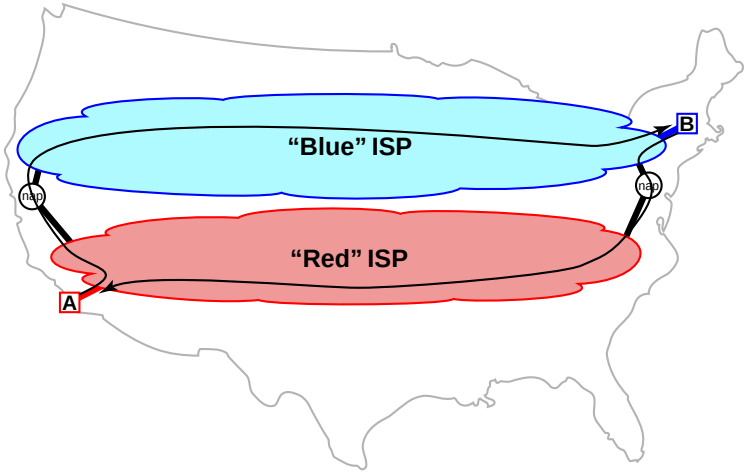
Traffic asymmetry and “fairness”

Internet service is typically based on bilateral agreements on the amount of data allowed to cross a mutual border (a wholesale-retail economic model). Since the unit of service is different everywhere there is no global partition function to support a global “fairness”.



Internet topological asymmetry

The internet model has no symmetry: In bi-directional communication the two directions almost always follow different paths. This is a deliberate engineering decision ("early out") that follows from the open competition of ISPs:



There's also a 10:1 to 100:1 difference between the data sent each direction so a web hosting ISP and cable modem ISP see very different backbone loads *from the same transactions.* Is that "unfair"? To whom?

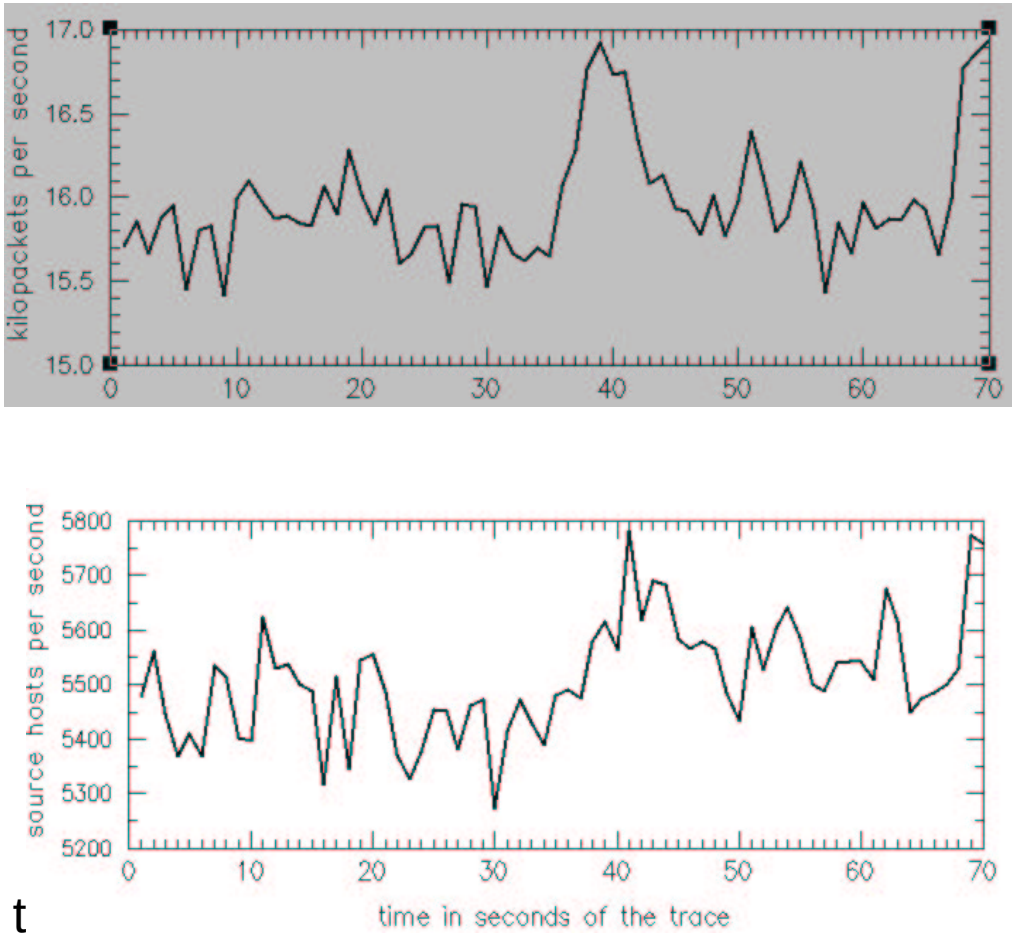
Temporal asymmetry

Net usage is very bursty: a Bell Labs study found home users weren't happy with <1Mb/s access but *averaged* <40 b/s over a week.

This huge disparity between peak and average usage gives IP a large multiplexing gain compared to a telco TDM system.

But with per-user traffic rates varying over 4 orders of magnitude, it's hard to pick a time interval over which to compare different user demands.

Calls??? Where are the “Flows”?



t

If kpps data is plotted vs. hosts per sec data, result is a straight line with slope 3. Same number results from just looking at average behavior: 16K packets per second / 5500 hosts per second = 3 packets per host per second.

(1997 NLANR FIX-W data from http://www.nlanr.net/NA/).

Where are the “Conversations”?

- 1999 study of NASA Ames internet exchange from CAIDA shows more than 90% of traffic is web (http) and most of the remainder is mail, netnews and ftp.
  (data at http://www.caida.org/outreach/papers/AIX0005/)

- The communication model for this kind of traffic is not a conversation (a dialog between two parties) but rather a dissemination (user wants the data associated with some URL but doesn’t care who gives it to them).

- Poor fit between dissemination and a circuit’s conversation model amplifies scaling and traffic control problems. E.g., the difficulty of deploying web caching.

Most of the preceding mismatch between the circuit model and networking use, traffic and economics has been well documented and obvious for several years.

So why do people keep trying to turn back the clock and impose circuits on the Internet?

- The main reason seems to be to take advantage of analysis and control techniques that are well developed for voice traffic over circuits but have no equivalent for data traffic over a packet net.

- The most prominent current example of this is “traffic engineering”.

Traffic Engineering (TE)

- When the Internet was being created, the main concern was *reachability.* Everything needed to talk to everything else but no one particularly cared what path the bits took.
- As the Internet became commercialized, more and more people tried to make money by moving bits around.
- Initially it was sufficient to sell connectivity but today everything is connected and alternative carriers have to distinguish themselves on quantity of bits moved or delivery quality.
- This implies that controlling the path the bits take is important to an ISP's bottom line.
- Unfortunately, Internet research funding stopped long before any hooks for this got added to the architecture.

Filling the architectural hole

- Some crafty marketing types announced that the reason IP TE *hadn't* been done was because it *couldn't* be done.

- Having thus "proved" that IP was incapable of traffic engineering, they pointed out that TelCos had been doing traffic engineering since the days of A.G.Bell so the obvious way to get it was to roll the clock back to the 1800s and discard IP for MPLS.

- The inescapable logic of this convinced all the equipment vendors (dazzled by the prospect of a forklift upgrade of the entire Internet), the trade press (eager for something new to talk about since no one wanted to hear any more about ATM) and half the tier-1 ISPs (the ones that had started life as TelCos).

“Optimal” Traffic Engineering (Linear Programming)

Given a topology of n nodes with c_{ij} as the capacity of the link from node i to node j and t_{sd} the traffic that enters at node s and departs at node d. If f_{sdij} is the fraction of demand t_{sd} that traverses the link from node i to node j then optimal TE is the set of f’s that minimizes:

$$\sum_{i,j} C\left(\sum_{s,d} \frac{f_{sdij} t_{sd}}{c_{ij}}\right)$$

such that:

$$\sum_{s,d} f_{sdij} t_{ij} \leq c_{ij}$$

$$\sum_{x} f_{sdxi} t_{xi} = \sum_{y} f_{sdiy} t_{iy}$$

This scales like n^4 in space and n^5 or n^6 in time, depending on the algorithm. I.e., for an 11 node network ~10,000 f values get computed.

IP routing

Every link has a “cost” (metric) so there are $O(n)$ costs. The lowest cost path from every node to every other is well defined and there are algorithms to compute the lowest cost paths in $n \log n$ steps.

A surprising result: Linear Programming solutions and routing solutions are *equivalent* in the sense that almost any linear programming solution can be turned into a set of link metrics that result in the same traffic flow and vice-versa.
(see “Internet Traffic Engineering without Full Mesh Overlaying” by Wang, Wang and Zhang in Infocom-01)

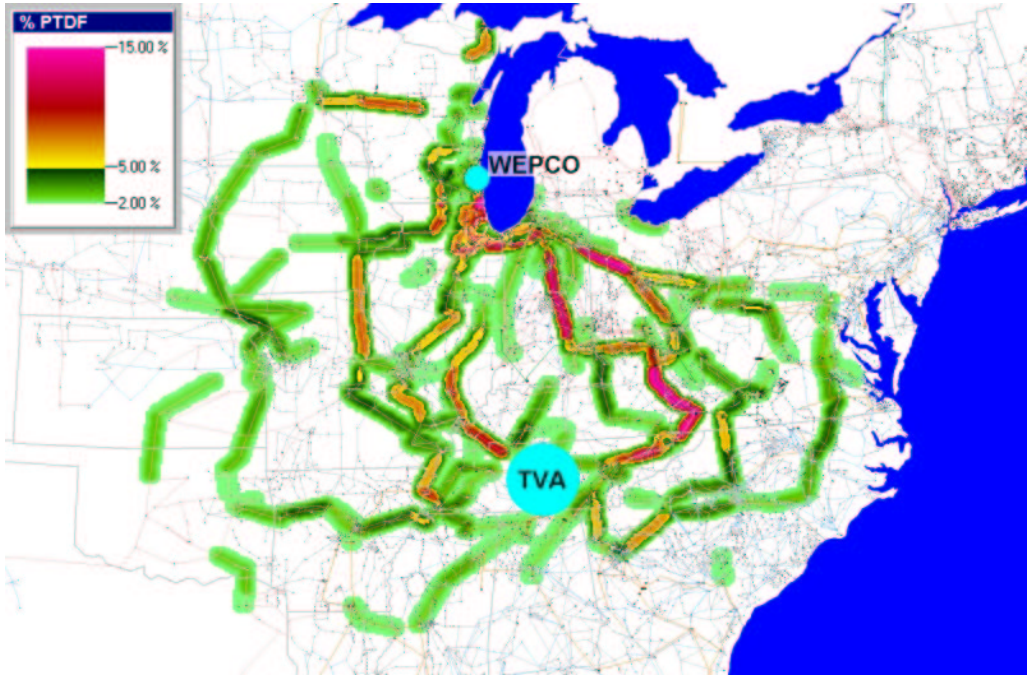
- I.e., you can do at least as much TE with IP routing as with MPLS and probably more.

(I think of this as analogous to physics looking at collisions in center of mass frame vs. lab frame. In lab frame problem is complex and quadratic. In center of mass frame the same problem is obvious and linear.)

If not circuits, what?

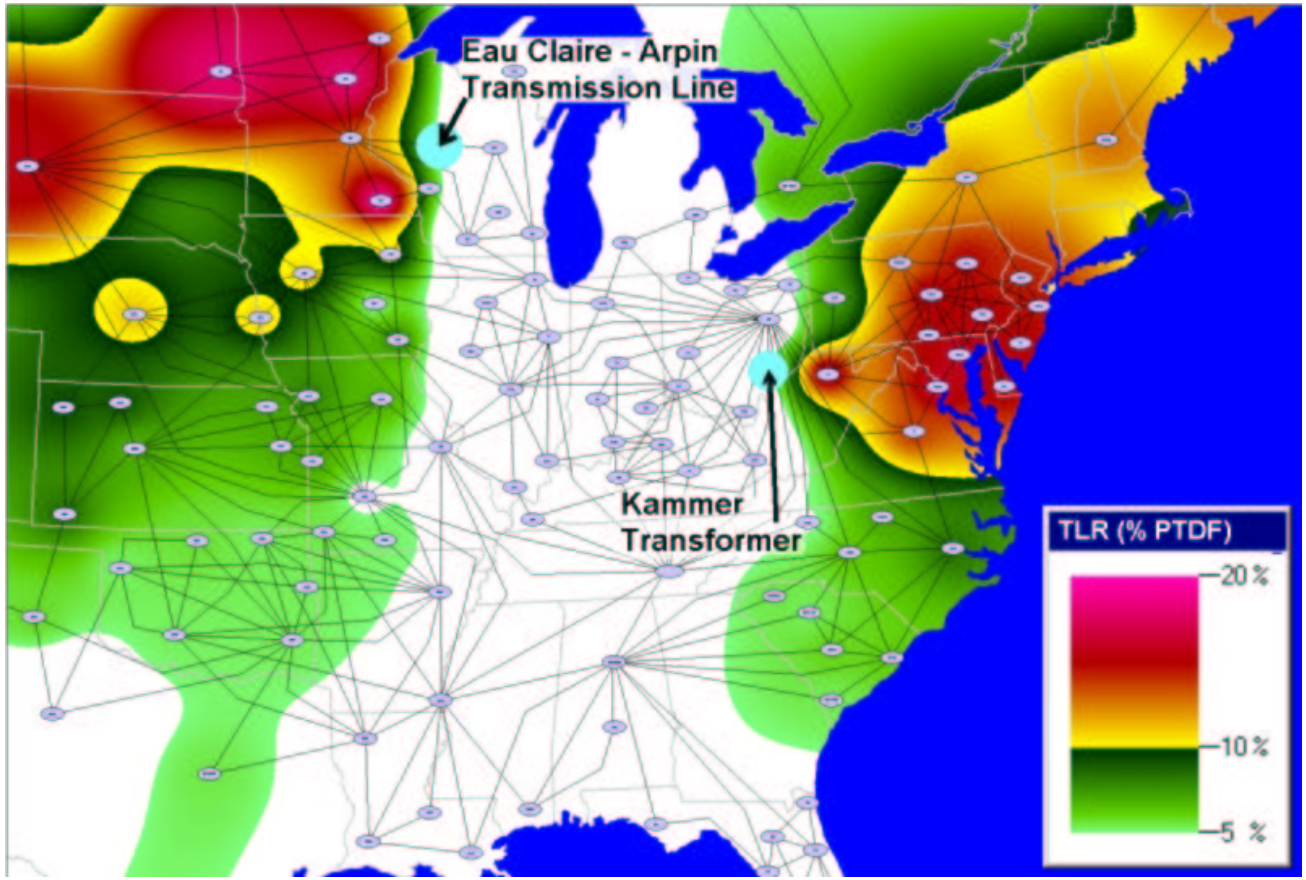
Almost *any* other communication model (e.g., post office, FedEx, freight trains, cargo ships) is a better fit to modern networking than circuits.

One that shares a lot of the same problems and has an excellent mathematical foundation is the operation of a power distribution grid. The following shows how a transfer from a generator to a utility flows through the grid:
(from IEEE Spectrum Feb 01 and www.powerworld.com)

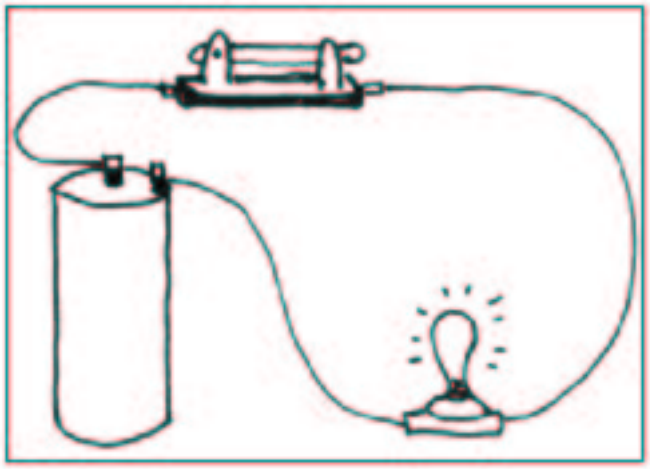


Power distribution (cont.)

This is a visualization of the state of the grid during a June 1998 power crisis in the Midwest that increased spot market prices by 3000%.
(from IEEE Spectrum Feb 01 and www.powerworld.com)



Why doesn't the Grid work like the phone system?



We get the same indoctrination for electrical circuits as for phones yet we didn't make the same mistakes. E.g.,

- the power company doesn't track individual power flows through each of your appliances.
- when a generator sells power to a utility they don't ask which customers it's going to.
- no one asks if it's "fair" to turn on a TV or a light.

Maxwell's Laws

A large part of the reason we didn't get confused is because in 1864 James Clerk Maxwell said everything there was to say about electricity:

$$\oint_S \mathbf{E} \cdot \mathbf{dA} = 0$$

$$\oint_S \mathbf{B} \cdot \mathbf{dA} = 0$$

$$\int_S \mathbf{E} \cdot \mathbf{dA} = \frac{\partial}{\partial t} \int_S \mathbf{B} \cdot \mathbf{dA}$$

$$\int_S \mathbf{B} \cdot \mathbf{dA} = \frac{\partial}{\partial t} \int_S \mathbf{E} \cdot \mathbf{dA}$$

There is ***no*** equivalent statement for communications.

What are we missing?

- The creation and operation of the power grid is made possible by the rigorous understanding of electricity embodied in Maxwell's laws.
- We have no such laws for network traffic and thus keep retreating to the Ohm's law world of circuits in a (futile) effort to stay on firm ground.
- We have tools like Ito calculus that might allow us to deal with stochastic packet flows in much the same way Maxwell dealt with deterministic electron flows. There are even mathematicians (e.g., Frank Kelly at Cambridge in the UK) applying these tools to the net, albeit working to solve different problems.
- Like the revolution that Maxwell started in 1864, supplying a little bit of missing theory might completely change the way we look at the world.