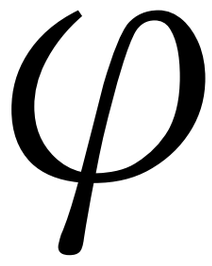


Public Health for the Internet

Toward a New Grand Challenge for Data Management

Joseph M. Hellerstein, UC Berkeley

Tyson Condie, Minos Garofalakis, Boon Thau Loo, Petros Maniatis,
Timothy Roscoe, Nina A. Taft

Our Last Grand Challenge

- “We recommend a ten-year goal for the database research community: . . . Make it easy for everyone to store, organize, access, and analyze the majority of human information online.”

  - The Asilomar Report, 1998

- “Google’s mission is to organize the world’s information and make it universally accessible and useful.”

  - Google Corporation

- Declare victory and *move on.*

Why a Grand Challenge?

We choose to go to the moon. We choose to go to the moon in this decade and do the other things, not only because they are easy, but because they are hard, because that goal will serve to organize and measure the best of our energies and skills, because that challenge is one that we are willing to accept, one we are unwilling to postpone, and one which we intend to win.

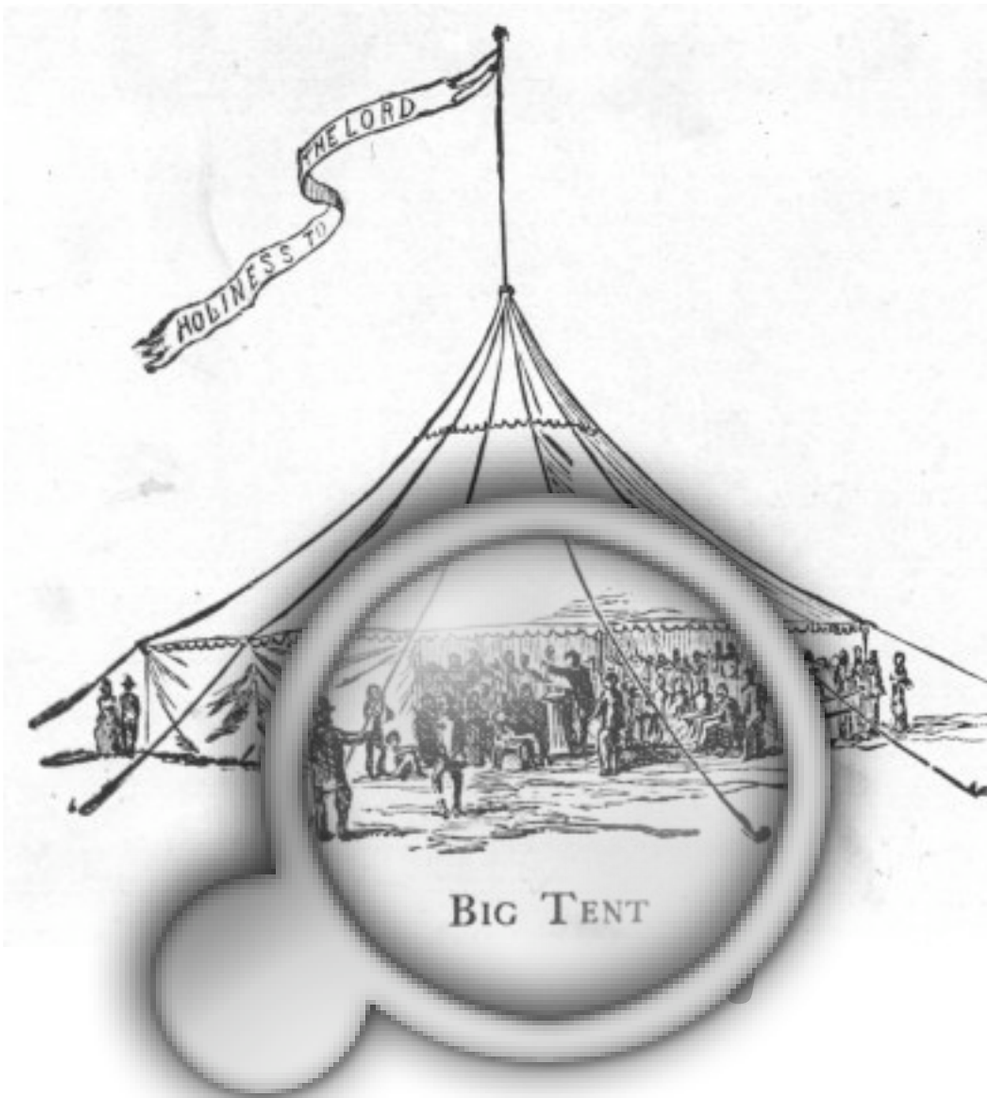
– John F. Kennedy

Every generation needs a new revolution.

– Thomas Jefferson

This Generation is Different

- More diverse geographically
- More diverse technically
- More diverse in application interests
- Query: Can we harness this new breadth to a shared purpose?



A Big, Focused Tent

Motivation for a New Grand Challenge: The Common Good

But history will judge you, and as the years pass, you will ultimately judge yourself, in the extent to which you have used your gifts and talents to lighten and enrich the lives of your fellow men. In your hands lie the future of your world and the fulfillment of the best qualities of your own spirit.

– John F. Kennedy

Technical Parameters

- DISTRIBUTED: The Revolution Will Not Be Centralized

  – with apologies to Gil Scott Heron

- MULTIPARTY: The People, United, Will Never Be Defeated

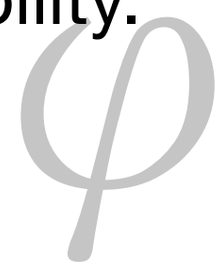
  – Sergio Ortega

- SENSITIVE: Information is Knowledge. – Albert Einstein

  Knowledge is Power. – Sir Francis Bacon

  With great power comes great responsibility.

  – Uncle Ben



A Grand Challenge:

Public Health for the Internet (PHI / φ)

Computers on the Internet should **organize themselves** into a **worldwide community watch**,

tracking and containing the spread of viruses, worms, spyware and other **harmful traffic**.

This should be done without sacrificing end-user privacy or autonomy, and without placing undue responsibility or control into the hands of any one party.

φ

Who Cares?

The New York Times

Attack of the Zombie Computers Is Growing Threat

"It's a huge scientific, policy, and ultimately social crisis, and no one is taking any responsibility for addressing it," said K. C. Claffy, a veteran Internet researcher at the San Diego Supercomputer Center.

— Last Sunday's Front Page (01/07/2007)

- Malware cost global businesses $169-204B in 2004
- DB market revenues that year were ~ $7B

From Medicine to Public Health

- Security tools focused on “medicine”
  - Vaccines for Viruses
  - Improving the world one patient at a time
- Opportunity in the “Public Health” arena
  - Public Health: population-focused, community-oriented
  - Epidemiology: incidence, distribution, and control in a population

φ : A New Approach

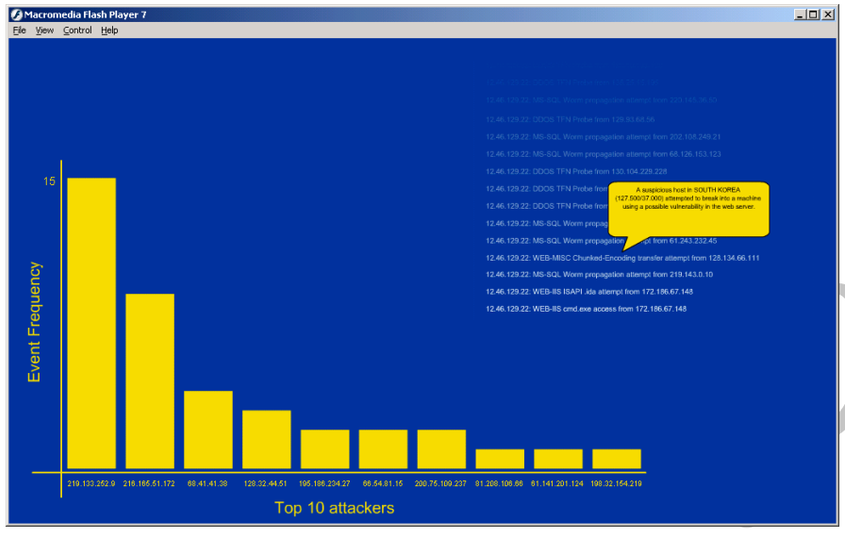
- A Peer-2-Peer community watch for the Internet
  - The endpoints are the sensors
  - Knit them together directly

- Population-wide monitoring
  - Not just at the chokepoints

- Engage end users: education and prevention
  - Encouraging healthy behavior

Gives and Gets for Joining φ

- Gives: sensor data
  - Firewall logs (Windows, OS X, Snort)
  - Spam filter logs (Outlook, Thunderbird), Phishing URLs
  - Modest compute/BW overhead
- Gets: insight, better security, karma, geeky fun



What is to be done?

- Distributed data streams
- Data security: *verifiability*, distributed resource governors, privacy
- Codesign of networks and data processing
- Architecture via protocol design
- Uncertainty, data reduction, heterogeneity
- And a lot of networking, security and ML work

Kicking off a Grand Challenge

- It’s hard to articulate a good one
- But the metaquery is harder
  - How do we build community momentum?
- Some recent models:
  - PlanetLab
  - Project IRIS
  - Hadoop

http://www.openphi.net

A Center for Disease Control?

- Who owns the Center? What do they Control?

- This will be unpopular at best
  - Electronic privacy for individuals
  - The Internet as “a broadly surveilled police state”?

- Provider disincentives
  - Transparency = embarrassment

- And hard to deliver
  - Can monitor the chokepoints (ISPs)

Why Us?

- The bad guys don’t target SIGs, they target computers

- Data management is a key component
  - Yes, the other folks could develop it too
  - Do you want to let them? Do they want to wait?

- Lots for us to learn
  - Networking, Security, ML
  - We play here anyhow, so focus the collaborations

Energizing the End-Users

- Firewalls: Fiction
  - mobility
  - tunneling

- Endpoints: Everywhere
  - Internet, intranet, hotspot
  - Toward a uniform architecture

- End-users will help
  - Populist appeal to home users is timely
  - Imagine the cool net-security download

φ