NISTIR 7465

# MATHEMATICAL FOUNDATIONS OF MEASUREMENT SCIENCE FOR INFORMATION SYSTEMS

# REPORT OF A PLANNING WORKSHOP

NIST
National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

NISTIR 7465

# MATHEMATICAL FOUNDATIONS OF MEASUREMENT SCIENCE FOR INFORMATION SYSTEMS

# REPORT OF A PLANNING WORKSHOP

National Institute of Standards and Technology
Gaithersburg, MD 20899

**October 24, 2007**

**U.S. DEPARTMENT OF COMMERCE**
*Carlos M. Gutierrez, Secretary*

**NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY**
*James Turner, Acting Director*

# Mathematical Foundations of Measurement Science for Information Systems

*Report of a Planning Workshop*[*][†]

October 24, 2007

The Information Technology Laboratory (ITL) of the National Institute of Standards and Technology (NIST) is developing a new intramural research program on the mathematical foundations of measurement science for information systems. Among the long-term goals of this program is the understanding, and ultimately the measurement, of fundamental properties of information systems which relate to the reliability and security of our cyberinfrastructure.

As part of the startup of this program, NIST invited a group of subject-area experts for an informal one-day workshop to discuss the state of the art in this area and to consider the path forward. This workshop was held on May 25, 2007 in Gaithersburg, MD.

## 1 Attendees

The external panelists were

George Cybenko (GC)
Thayer School of Engineering
Dartmouth College

John Gilbert (JG)
Computer Science Department
University of California at Santa Barbara

[*] This report was compiled by Isabel Beichl and Ronald F. Boisvert of NIST. For further information, contact `boisvert@nist.gov`.

[†] Disclaimers: Certain commercial products are referenced in this report in order to faithfully document the discussions that took place at the workshop. The mention of such products does not imply any recommendation or endorsement by NIST. Opinions expressed by workshop participants are their own and do not represent the official views of NIST.

Brian Hunt (BH)
Department of Mathematics and Inst. for Physical Science and Technology
University of Maryland

Andrew Odlyzko
Director, Digital Technology Center
University of Minnesota

Edward Ott
Institute for Research in Electronics and Applied Physics
University of Maryland

Gregory Sorkin (GS)
Department of Mathematical Sciences
IBM T. J. Watson Research Center

Francis Sullivan (FS)
Director, IDA Center for Computing Sciences

Walter Willinger (WW)
AT&T Labs – Research

Attendees from within NIST included the following:

- Mathematical and Computational Sciences Division (ITL)

  Isabel Beichl
  Ronald Boisvert (RB), *Chief*
  Fern Hunt (FH)
  Manny Knill

- Computer Security Division (ITL)

  William Burr
  Donna Dodson
  Rene Peralta

- Advanced Networking Technology Division (ITL)

  Vladimir Marbukh (VM)
  Kevin Mills
  David Su, *Chief*

- Software Diagnostics and Conformance Testing Division (ITL)

  Paul E. Black

- Statistical Engineering Division (ITL)

  James Filliben
  Antonio Possolo, *Chief*

- Information Technology Laboratory (ITL)

  Sandy Ressler, *Manager, Complex Systems Program*

- Manufacturing Engineering Laboratory

  Albert Jones

# 2 Organization of the Workshop

Ronald Boisvert opened the workshop with an introduction to measurement science at NIST. He also provided a brief description of the NIST FY2007 Cyber Security initiative which is providing funding for the proposed program. His slides are included in the Appendix of this report. The external participants were invited to make brief presentations on their views of relevant mathematical research in this domain. Summaries of those presentations are provided in Section 3. The balance of the day was spent in directed discussions centered around questions distributed to the participants in advance. A summary of the main points from that discussion is provided in Section 4.

# 3 Contributed Talks

## 3.1 Quantitative Evaluation of Risk
*George Cybenko*

George Cybenko spoke about quantitative evaluation of risk to aid in decision making for investment in software protection strategies. For example, companies can invest in technologies to protect intellectual property marketed to the general public such as music or video games. Some strategies may be expensive to implement and deploy, but afford a great deal of protection, while other are less expensive, but may be more easily broken. Given that any protection scheme is likely to be broken at some point, the question asked is: what is the return on a given level of investment?

In this project, undertaken with Jeff Hughes at the Air Force Research Lab (Dayton), the aim was to develop a model that predicts how long it takes for a protection scheme to be broken. This, in effect, provides a measure of the security of such a protection scheme.

In particular, a probability distribution function for the time of first successful attack is sought. A typical way get at such information is to hire one or more

"Red teams" to attempt to defeat the security mechanism. Unfortunately, such a process obtains, at best, a single data point rather than a distribution. Instead, Cybenko and colleagues modeled the work of an attacker as a Markov process, i.e., the attacker moves from one state to another according to certain probabilities. (The process is a "partially observed" one, since as the attacker is unaware of the current state at any point in time.) Cybenko and colleagues used an "information market" approach with a collection of expert colleagues to develop the parameters needed by the model.

With the attacker model in place, Cybenko and colleagues use dynamic programming techniques to get an optimal policy given that the attackers have costs and benefits. This is, in effect, related to stock options pricing analyses.

*Key message:* The field of risk analysis in statistics is a mature discipline which holds promise as a means to provide metrics for cybersecurity. Security is a property of a system which is very difficult to precisely measure. Risk analysis has always had to deal with such uncertainties. Thus, it may be fruitful to consider a well-characterized measure of risk as a derived measure of "security".

## 3.2 Network Science in Need of Measurement Science *Walter Willinger*

Walter Willinger spoke on the need for measurement science for computer networks, relating some lessons learned from his own work in modeling of the Internet. His main message is that much of the existing work in the emerging area of network science is severely lacking in rigor, and that network science can only become a true "science" when accompanied by an appropriate measurement science.

As a concrete example, he cited studies of Internet router-level connectivity. Many of these studies have been based on data that map connectivity using large-scale trace-route experiments. From this data researchers have inferred power-law node-degree distribution and verified preferential attachment growth models. They have also observed hub-like cores which make the network vulnerable to planned attack. Such studies have been highly publicized.

However, Willinger argued that much of this work has been wrong at a fundamental level. He states that traceroute data are ambiguous, inaccurate, and incomplete, and hence cannot support any scientific conclusions about the real Internet. He also complained about a lack of statistical rigor in these studies, as well as the absence of any serious model validation. Bad models are dangerous because they can distort public opinion and cause bad policy to be made.

He believes that measurement science needs to step in to answer critical questions such as whether available Internet-related connectivity measures actually support the claims made in the complex networks literature. In other words, for what purpose can the measurements at hand be safely used? In particular, a good measurement science must play a role in

1. raising the level of data hygiene
2. matching statistical rigor to the quality of the data
3. promoting serious model validation

Finally, Willinger makes the point that researchers must be more specific about what aspects of the Internet that they are modeling, and then bring in much more detailed domain knowledge in order for the models to be useful. (For example, models of the router-level Internet that admit unbounded node degree are unrealistic, since real routers must have finite size.)

A copy of Willinger's slides are included in the Appendix of this report.

*Key message:* A rigorous measurement science for networks is critically needed. Improved data, more detailed models, and serious model validation are necessary to make network science a real science.

## 3.3 Economics of Security *Andrew Odlyzko*

Andrew Odlyzko spoke about the economics of information security. He stated that cryptography is mostly irrelevant to issues of day-to-day information security. People will always be involved, and so economics, psychology, and usability are critical to the understanding of practical security of systems.

Since absolute security is probably unattainable, it is important to understand the actual relationship between actual (imperfect) security controls and the level of risk they engender. He suggested by way of analogy the cat-and-mouse game played by original and after-market manufacturers. For example, printers are priced quite low, with the original manufacturer expecting to make most of its profits on consumables, like ink cartridges. Originally, it is the only vendor for the cartridges, but eventually after-market manufacturers will reverse engineer these and develop cheaper versions, thus stealing a good deal of the market. The original manufacturer can choose to invest alot initially to make the design complex so that reverse engineering is difficult. In computer security, code obfuscation and related technologies can be used to make software more difficult to understand and exploit for nefarious purposes. It would be useful to measure the value of such strategies, that is, to understand more clearly the tradeoff between investment in security technologies to the cost required to break them.

An interesting observation in this regard is that the efforts required to break into systems can vary, and that if the effort required is large then any attempt to do so is bound to leave traces. Can we measure these?

Other possible research directions he suggested were (a) to measure how long it takes information to disseminate around the web, and (b) to develop institutional mechanisms for collecting and distributing data that can be used by cyber security researchers.

Finally, Odlyzko made the point that even rough models can be useful in guiding decision making. (He quipped that economists have managed to make great careers in spite of the fact that the predictions of their models are rarely accurate.)

*Key message:* Since cyber attacks are often economically motivated, it may be useful to consider economic models. Since no system can be made absolutely impenetrable, it is of particular interest to measure the impact of imperfect security.

## 3.4 Some Issues of Network Topology *Edward Ott*

Ed Ott explained that simple graph models of information systems can be quite useful for defining questions, formulating solution techniques, and for gaining intuition. This is related to the principle of universality in physics, i.e., that solutions in simple cases can many times be applicable to more complex problems. He described a recent series of studies of the network models that he and his colleagues have undertaken, including the following.

- Characterization of the dynamical importance of network nodes and links using the largest eigenvalue of the associated adjacency matrix. See Juan G. Restrepo, Ed Ott, and Brian R. Hunt, *Physical Review Letters* **97**, 094102 (2006), as well as arXiv:0705.4503.

- Studies of the emergence of synchronization in complex networks of interacting dynamical systems. See Juan G. Restrepo, Ed Ott, and Brian R. Hunt, *Physica D: Nonlinear Phenomena* **224** pp. 114-122, as well as arXiv:0706.4454.

- Studies of percolation on large-scale networks. Here one considers how many nodes need to be broken in order to break a large graph into small disconnected components. This would have relevance to network degradation or attack, or to immunization and protection against epidemics. See arXiv:0704.0491.

- A similar problem is finding "communities" in networks, that is, a group of nodes with many connections to other nodes in the group. Hierarchical clustering and Laplacian spectral methods have been used to solve this problem.

Some discussion ensued about whether such simple models can actually be useful in predicting behavior of large-scale computer networks. Ron Boisvert made a comment that many other information systems have network structure, and such simple models might well be appropriate for their study. He cited the graphs associated with the static and dynamical structure of large computer programs as an example. Al Jones noted that manufacturing systems, health care systems, and web services also have such a structure. He explained that in some cases connections are not permanent. They change. But understanding the topology might nevertheless lead to useful performance metrics.

A copy of Ott's slides are provided in the Appendix of this report.

*Key message:* Searching for simple relationships between system topology and behavior is an important means for developing fundamental understanding of information systems.

## 3.5 Horizontal Integration *Francis Sullivan*

Francis Sullivan began his presentation noting that real networks do not seem to fit the statistical mechanical model. Although one can compute a power law exponent for a network model, it is not clear that this gives you any useful information. Nevertheless, he suggests that there may be other combinatorial quantities to measure which can provide real insight, such as distributions of cliques and independent sets, but these are much more challenging to compute.

He said that in a sense security is impossible because everybody is using the same technology. No matter how much Microsoft spends, attackers will eventually win if only because the ethos of attacking the most visible target draws in more people. He suggested that a greater diversity of operating systems would indirectly help security. Another approach to foil hackers would be technologies like virtualization which serve to disguise the real system running underneath.

He agreed with Andrew Odlyzko that cryptography is not the answer, although it still remains an important tool. If you have data it has to be unencrypted *sometime*, and since cryptography must be used by people, mistakes will be made. It is here where systems are the most vulnerable, and there may never be technological solutions that can overcome weaknesses in human behavior.

Sullivan related a theory of security called "horizontal integration", which was developed in a recent JASON study[1]. The study considers more agile mechanisms for managing classified data than the traditional hierarchical approach. The study focuses on measurement of risk rather than security. They turn risk into a commodity by tokenizing it and passing out the tokens to people who need it. Those with tokens expend them in the disclosure of information. There is a trade-off between convenience (e.g., expediency) and security that is made in an ad-hoc way at each transaction. Such a system would provide a more flexible means

[1] A "release" copy of this report was made available to us and is available to workshop participants on request.

of regulating information flow in battlefield situations, for example. It was suggested that computer operating systems could be the "battlefield" and a research direction could be to develop a theory and model of risk, like actuarial science.

### 3.6 Global Properties of Networks
*John Gilbert*

John Gilbert spoke about the need for measurement of global properties of networks and the tools that might be necessary to do this. He suggests going beyond measuring properties that have been the topic of most recent studies, such as density, diameter and degree distributions, and instead consider a more comprehensive set of graph-theoretic measures. The development of effective algorithms and software for computing (or estimating) such properties of graphs represents an important new research area.

Gilbert suggests work on developing the fundamentals of high-performance combinatorial computing would provide the underlying basis for a measurement science for information systems. Such techniques and tools are largely unavailable today. Such fundamental issues as what are the most efficient and effective computational primitives upon which to develop high performance software tools for graph-theortic computations remains unresolved. (Both he and Bruce Hendrickson of Sandia Labs have been studying this.) To support such work, he also suggested that NIST develop standard reference data sets and data generators for combinatorial computing.

Finally, he also suggested that more complex network models were needed to represent modern information systems. Such systems (e.g., the Internet) typically have a multi-level structure, and hence new abstract models of multi-level systems need to be constructed.

A copy of Gilbert's slides are provided in the Appendix of this report.

*Key message:* A measurement science for information systems needs the ability to perform non-trivial computations on large-scale graphs. To enable this, fundamentals of high-performance combinatorial computing must be developed.

## 4 Discussion

A series of general questions were posed to the panelists to elicit discussion related to potential goals and topics for NIST's program. A summary of the main threads of discussion is provided here. Speakers are identified by their initials; correspondence to full names is provided in Section 1.

**General topic***: Technical goals for NIST's program.*

- GC said that decision markets and information markets are very important and thinking of networks in terms of economics may be fruitful. He mentioned Michael Kearns' Penn-Lehman Automated Trading Project at the University of Pennsylvania as an example.

- Modeling unknown threats was posed as a very difficult challenge. One approach would be to characterize is the "normal" behavior of a computer network. By monitoring deviations from normality we could potentially detect threats of various kinds. Could this be scaled to network-wide measurements that could be used to characterize overall communications structure? The question of what measurements to make and where to make them is an interesting one. RB noted that DARPA[2] has put resources into anomaly detection for identifying intrusions, but that even in this simpler case the problem remains quite difficult.

- GS noted that IBM has had an effort in so-called autonomic computing for some time. One of the goals is to be able to automate the determination of dynamic control parameters for routers. One needs good measurements of network activity in order to design such controls.

- RB asked if insights from biological systems would help. WW stated that this has not gone beyond metaphor. BH brought up modeling of the spread of computer viruses. WW noted that there is nice mathematical work on the spread of viruses in scale-free networks, but that it would be more useful to consider the case of *real* networks. FS commented that the spread of a computer virus is not like a biological virus. The origins are different. Suppose you locate the origin of the computer virus, so what? RB suggested that understanding how the structure of the network might lend itself to controlling viruses would be interesting.

  GS explained that at IBM biological insights motivated both theoretical and practical work: biology was more than just metaphor. A computer virus can be characterized by a bit string, like DNA and real viruses, and typical commercial anti-virus packages include quarantine procedures. Computer virus bit strings, like organisms' DNA, can be used to create phylogenies; since computer viruses are often patched together from several parents, their phylogenies are not trees but directed acyclic graphs. These are interesting from a theoretical perspective, and also because they indicate common viral structures useful for efficient detection of many viruses. (If this means that a biological notion leads on to something non-biological, that's fine too, he said.) In the realm of metaphor, as hosts may first attack intruders with

[2] Defense Advanced Research Projects Agency

white blood cells and then develop specific immunity, machine-learning techniques can recognize likely computer viruses, which can then be subjected to automated techniques for recognizing them more efficiently, and also for "curing" infected programs. (Such analysis is currently done in a laboratory but could be done on end-user computers). Epidemiological notions then become even more relevant. Viruses spread by communication between machines. If a machine recognizes that it is infected and develops its own antidote, it can promptly communicate the antidote to machines it may have infected. This can lead to a much more favorable epidemiological model (and is another case of biology leading to something seemingly non-biological).

- GC thought that it is not the network itself but the information on the network that should be modeled. How much information fluidity is there? The network is interesting but it is only the dish; the actual food is what is really of interest.

- FS suggested that monitoring for unexpected file changes would be important. Suppose we have a huge file: is it the same as it was 10 minutes ago? Are there sampling techniques that could be developed that would allow continuous monitoring of the state of files without significant degradation of system performance? For example, Michael Rabin considered the use of the first few coefficients of the Fourier transform as a rough measure of change.

**General topic***: Technical skills required to staff NIST's program.*

- GC mentioned Jon Kleinberg's graduate and undergraduate Network Science courses at Cornell as good background for the type of person that NIST might want to hire. Besides looking at information theorists and computer engineers familiar with networks, GC also suggested looking at other disciplines which might be quite relevant, such as the social sciences, economics, and statistics.

- GS suggests someone in random structures, discrete methods, probability, statistics, statistical mechanics, and computational combinatorics. FS concurred that probability and combinatorics will be very important. JG explained that these skills are also very applicable to biology and nanotechnology. He cited Berkeley, MIT, Georgia Tech, and the University of Maryland as centers for work in computational graph theory.

- GS added that machine learning and data mining are relevant to intrusion detection and probably in the general area of making sense of the behavior of a complex communications network.

- FH asked about the relevance of queuing theory. WW said that networks that can be described in closed form by queueing theory are too simplistic to be useful. He went on to explain that places like Bell Labs, Bellcore, and IBM used to have research groups that worked almost exclusively on queueing theory. However today if you look today at places like AT&T Labs-Research, Microsoft Research, or Google such groups don't exist any longer (which may say something about the relevance of that area). VM suggested that approximate queueing models can be used in much more complicated situations.

- WW suggested that stochastic control theory was a very highly relevant area for network modeling. Centers of excellence for such work include Cal Tech (Stephen Low, John Doyle), UIUC, and Cambridge. Operations research and optimization were also cited as related relevant skills. Ultimately, one is trying to steer a system toward optimality.

- Game theory was suggested as another tool useful in modeling network growth and dynamics. Tim Roughgarden of Stanford and Eva Tardos of Cornell are leaders in this area.

- FS suggests seeking people who are broad, flexible, and smart. We should value people who have the ability to look at real systems and form models.

**General topic**: *Potential unique contributions of NIST.*

- JG suggested that there would be real value for NIST to (a) compare simple mathematical models to real systems, (b) provide measurements of real systems for use by the research community, and (c) provide reference data on the properties of real systems.

- It was suggested that NIST might provide a center for the sharing of data on real networks. For example, there is no reliable information on how much data goes between different Internet service providers. It is unlikely to be able to persuade companies to release this information generally, but perhaps they would release to NIST provided the data could be properly anonymized. WW mentioned some relevant work by Matthew Roughan on privacy-preserving measurements[3].

- WW also mentioned that there is a special measurement component of the GENI effort, with a separate working group headed by Paul Barford of Wisconsin that is trying to ensure that measurements are not again an afterthought (as is the case with the current Internet), but are built in from the beginning[4].

[3] See http://internal.maths.adelaide.edu.au/people/mroughan/Papers/minenet06.pdf or browse his web page.

[4] For a recent working group document see http://www.geni.net/GDD/GDD-06-12.pdf.

- FS noted that one of the great things about NIST is that it is unbiased, and what NIST says will be taken well by industry.
- GS summarized his views as follows: no one knows how to solve the original problem; the proper scientific or engineering abstractions are not there yet. In this case the best approach may be to build the science from bottom up. Bring together a group of experts in the nitty-gritty details of security and practical systems, with a body of experience to draw on. Have theoreticians from the areas mentioned (graph theory, economics, queuing theory, etc.) talk to and look over the shoulders of experienced practitioners, and try to abstract and generalize.

# Appendix

## A Slides from Presentations

1. Foundations of Measurement Science for Information Systems (Ronald F. Boisvert)
2. Network Science in Need of Measurement Science (Walter Willinger)
3. Some Issues of Network Topology (Edward Ott)
4. Foundations of Measurement Science for Information Technology (John Gilbert)

# A-1 Foundations of Measurement Science for Information Systems

**Ronald F. Boisvert**

**Welcome!**

# Foundations of Measurement Science for Information Systems

**NIST**

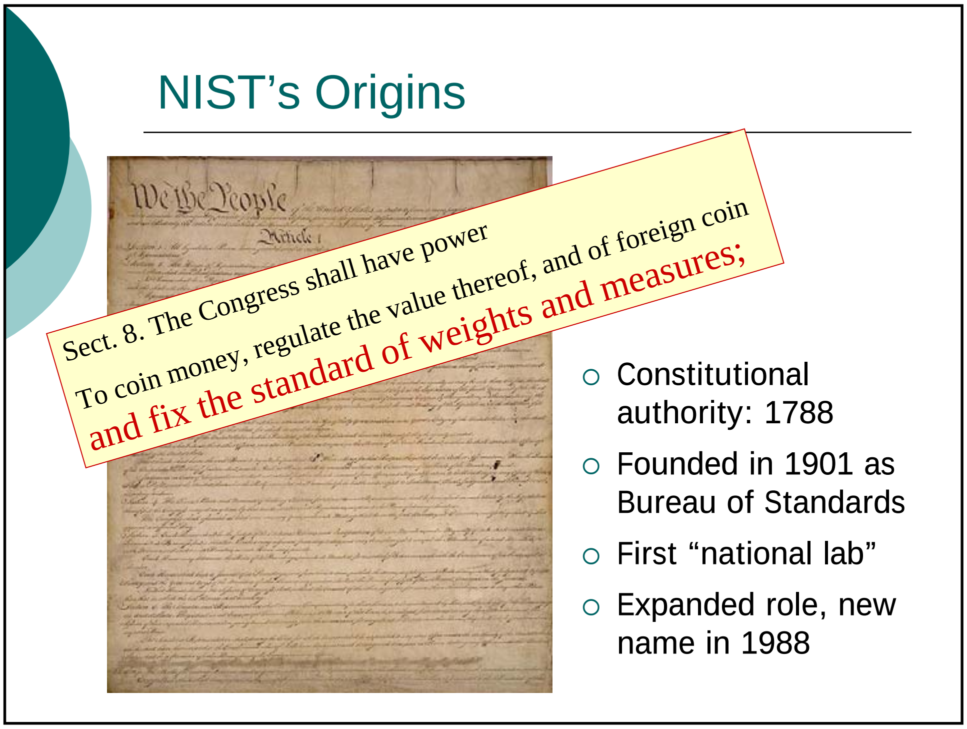
Information
Technology
Laboratory

May 25, 2007

*Ron Boisvert, Chief, Mathematical & Computational Sciences Division*

## This Talk

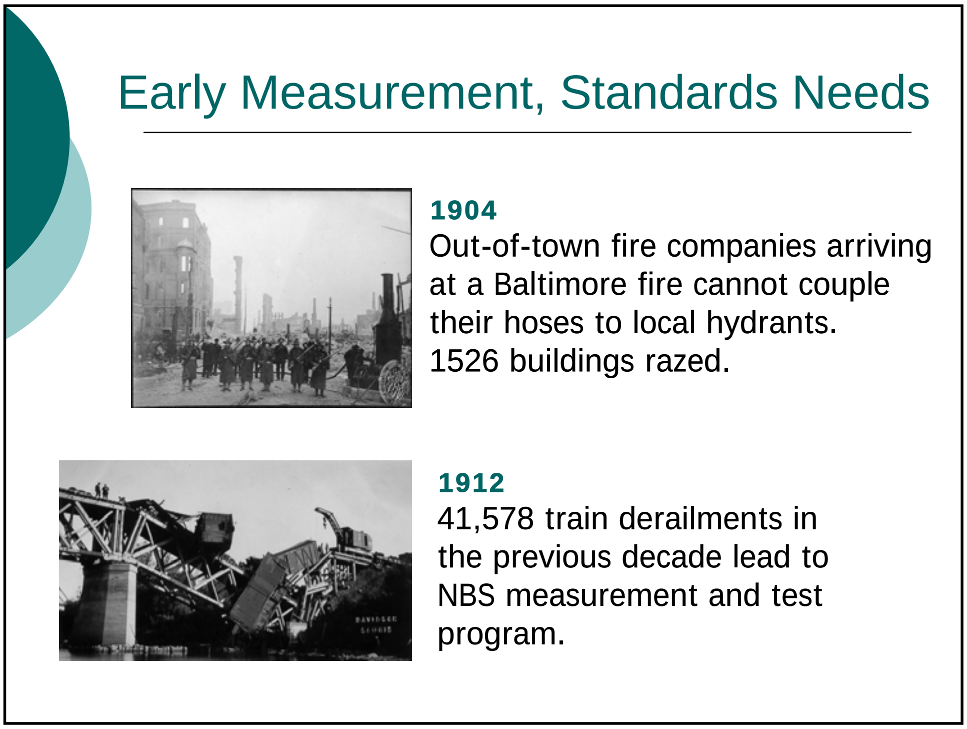
- Background
  - *NIST and Measurement Science*
  - *NIST Information Technology Lab (ITL)*
- FY07 Research Initiative
  - *Original Motivation*
  - *Proposed Directions*



## NIST's Origins

Sect. 8. The Congress shall have power
To coin money, regulate the value thereof, and of foreign coin
and fix the standard of weights and measures;

- Constitutional authority: 1788
- Founded in 1901 as Bureau of Standards
- First "national lab"
- Expanded role, new name in 1988



## Early Measurement, Standards Needs

**1904**
Out-of-town fire companies arriving at a Baltimore fire cannot couple their hoses to local hydrants.
1526 buildings razed.

**1912**
41,578 train derailments in the previous decade lead to NBS measurement and test program.

# NIST Assets and Mission

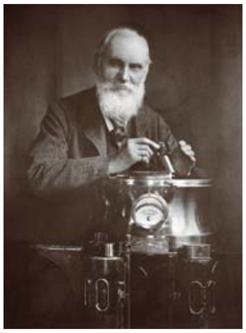
- **Laboratories**
  Gaithersburg, MD
  Boulder, Colorado
- $843 million FY 2007 budget
  ($677M Congressional appropriation)

2,800 employees
1,800 associates
850 users of facilities
1,500 affiliated field agents

**Mission**

*"To promote innovation and industrial competitiveness by advancing measurement science, standards, and technology in ways that enhance economic security and improve our quality of life."*

# Measurement



**Lord Kelvin**
**1824-1907**

*"I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of Science."*

*"If you can not measure it, you can not improve it."*

# Metrology

**Metrology:** science of measurement, embracing both experiment and theoretical determinations at any level of uncertainty in any field of science and technology*

- *Scientific or fundamental metrology* -- establishment of measurement units and systems, development of new measurement methods, realization of measurement standards, and traceability from these standards to users.
- *Applied or industrial metrology* -- application of measurement science to manufacturing and other processes, ensuring the suitability of measurement instruments, their calibration and quality control.
- *Legal metrology* -- regulatory requirements of measurements and measuring instruments for the protection of health, public safety, the environment, enabling taxation, protection of consumers and fair trade.

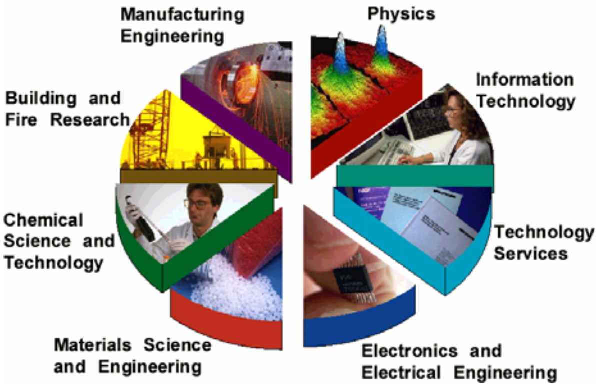
Key concepts

- traceability (through calibrations)
- characterization of uncertainty

* Bureau International des Poids et Measures (BIPM)

# NIST's Laboratories

Research to advance the nation's infrastructure for innovation:
measurements, test methods, evaluated data


Manufacturing Engineering
Physics
Information Technology
Building and Fire Research
Chemical Science and Technology
Technology Services
Materials Science and Engineering
Electronics and Electrical Engineering

Nanoscale Science & Technology

Neutron Research

## Output of the NIST Laboratories

**Measurement Science Research**
- 2,100 publications / year

**Calibrations and Accreditations**
- 3,200 items calibrated / year
- 826 Labs accredited

**National, International Standards**
- 450 committees

**Standard Reference Materials**
- 1,200 products available

Research Collaboratory for
STRUCTURAL BIOINFORMATICS

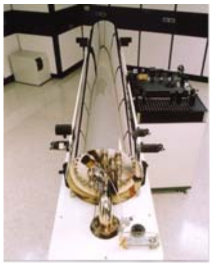
Welcome to the new home of the
PDB PROTEIN DATA BANK
Deposit | Search | Info | Download

**Standard Reference Data**
- 90 databases

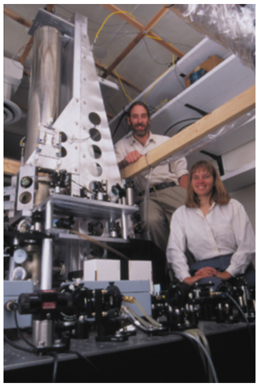
## Measurement Science: *The Second*

NBS
Pendulum
clock
(1904)
1s in 3 years

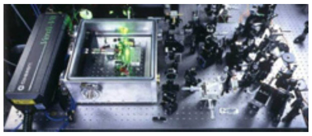
Ammonia
resonator
(1949)
1s in 300 years



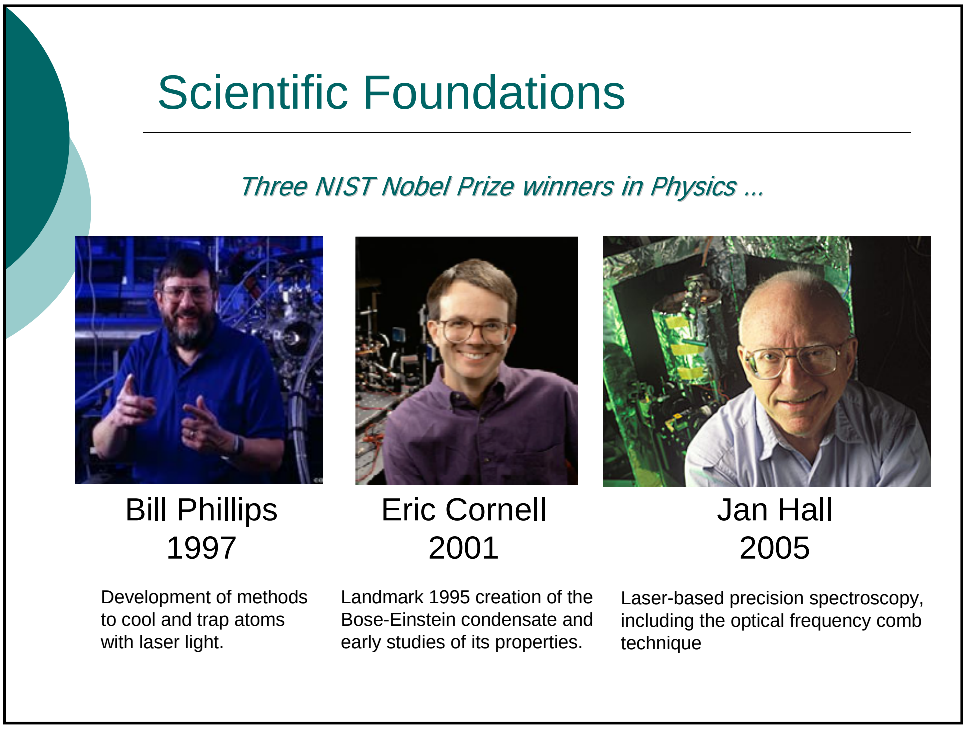
NIST 7
cesium beam
(1993)
1s in 6M years



NIST F1
cesium fountain
(1999)
1s in 30M years



Optical clock
(20xx)
1s in 30 billion years



## Scientific Foundations

*Three NIST Nobel Prize winners in Physics …*

Bill Phillips
1997

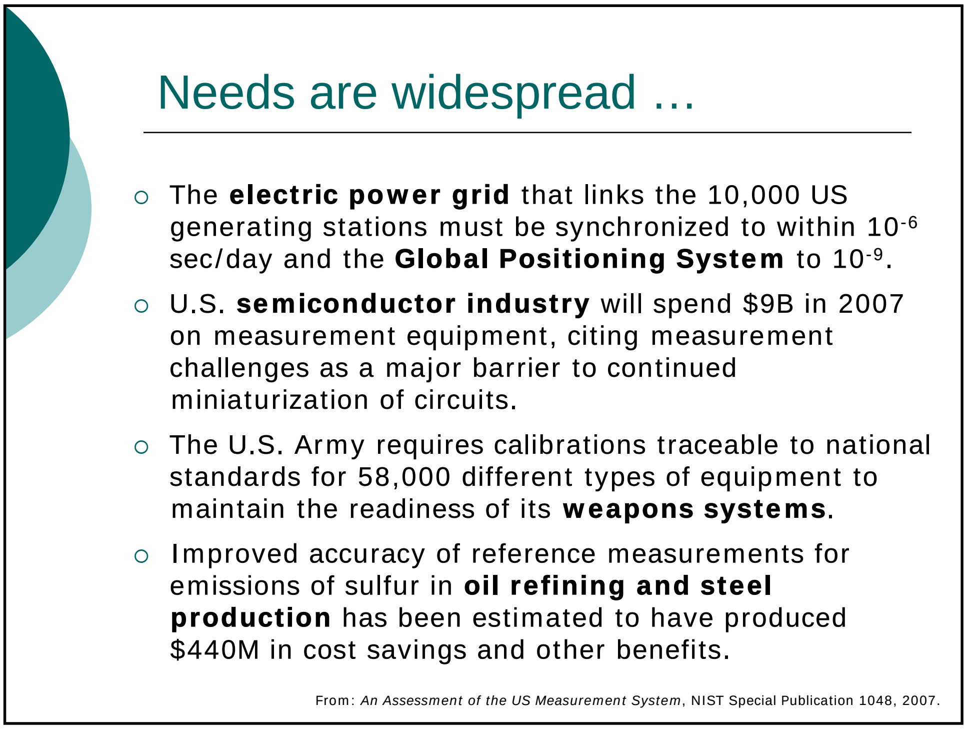
Development of methods to cool and trap atoms with laser light.

Eric Cornell
2001

Landmark 1995 creation of the Bose-Einstein condensate and early studies of its properties.

Jan Hall
2005

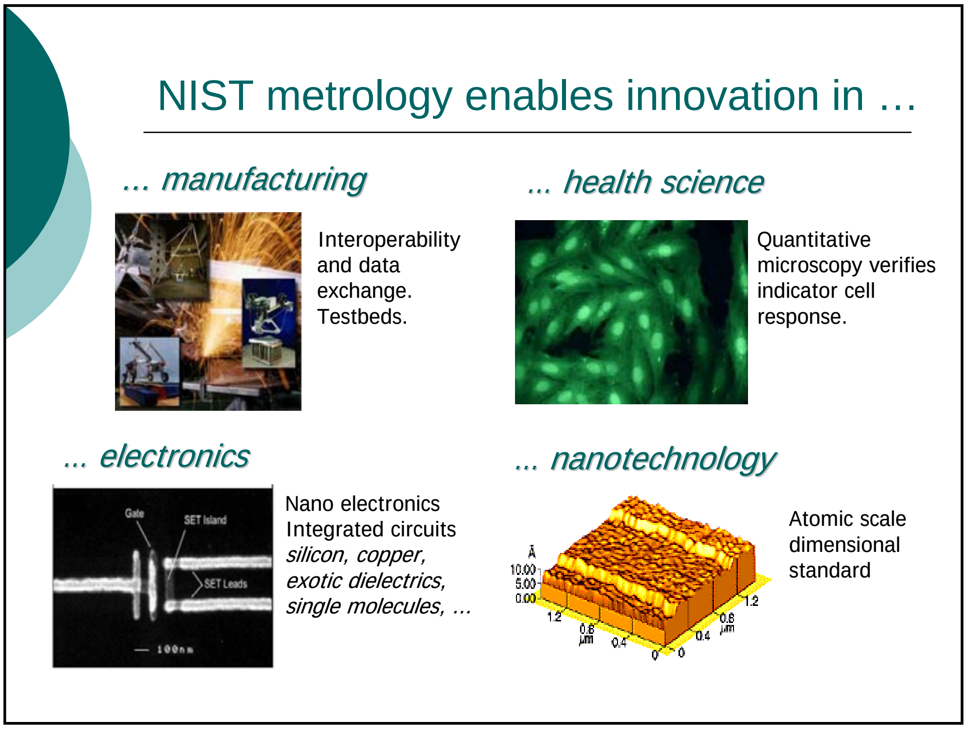
Laser-based precision spectroscopy, including the optical frequency comb technique



## Needs are widespread …

- The **electric power grid** that links the 10,000 US generating stations must be synchronized to within $10^{-6}$ sec/day and the **Global Positioning System** to $10^{-9}$.
- U.S. **semiconductor industry** will spend $9B in 2007 on measurement equipment, citing measurement challenges as a major barrier to continued miniaturization of circuits.
- The U.S. Army requires calibrations traceable to national standards for 58,000 different types of equipment to maintain the readiness of its **weapons systems**.
- Improved accuracy of reference measurements for emissions of sulfur in **oil refining and steel production** has been estimated to have produced $440M in cost savings and other benefits.

From: *An Assessment of the US Measurement System*, NIST Special Publication 1048, 2007.



## NIST metrology enables innovation in …

*… manufacturing*

Interoperability and data exchange. Testbeds.

*… health science*

Quantitative microscopy verifies indicator cell response.

*… electronics*

Nano electronics
Integrated circuits
*silicon, copper,*
*exotic dielectrics,*
*single molecules, …*

*… nanotechnology*

Atomic scale dimensional standard

## NIST metrology enables innovation in …

*…public safety and security*

**Measurements and standards infrastructure to ensure the accuracy, reliability, and security of systems critical to public safety**

Develop, compare, and test new technologies.
Enable safe and effective response to incidents.

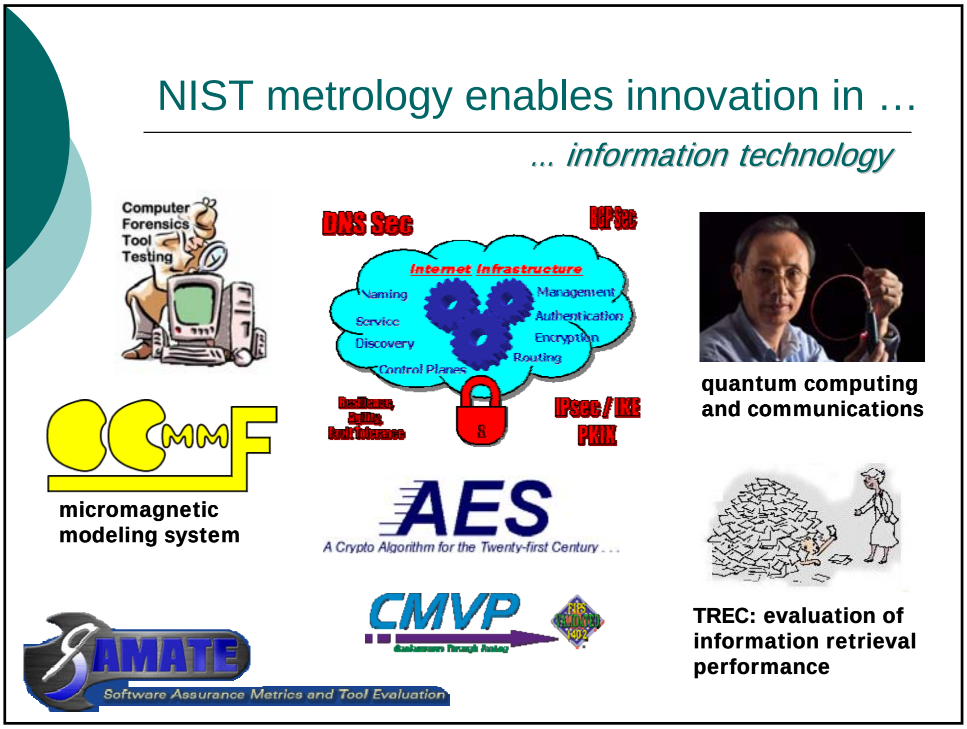
**World Trade Center Investigation**

**gas mask performance standards**

**biometrics**

**mail irradiation**

**DNA standards**



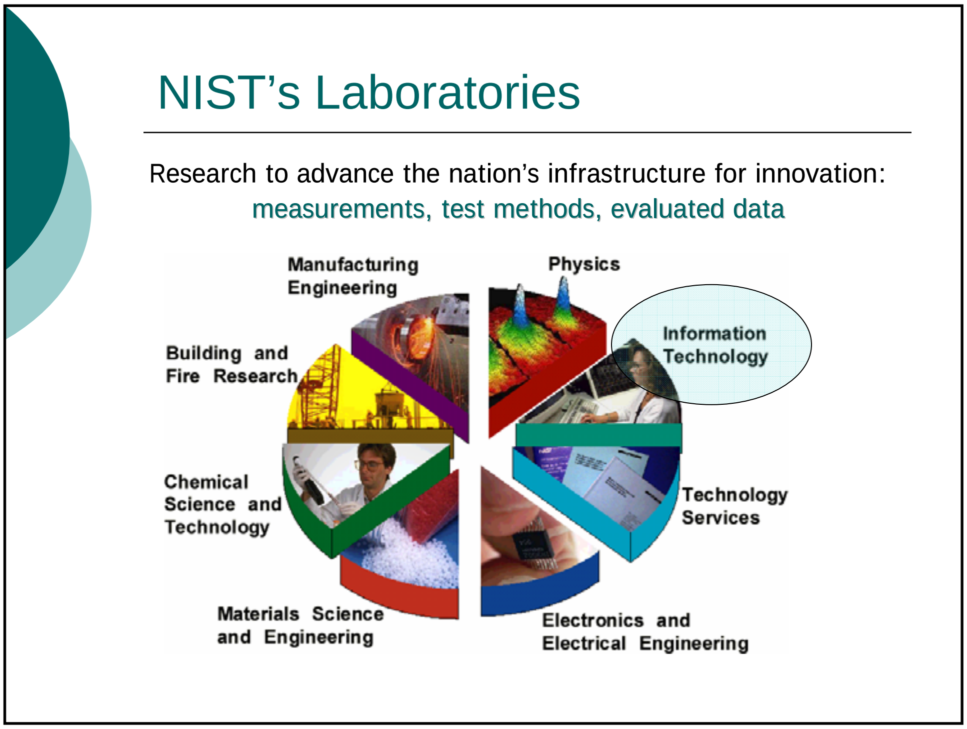
## NIST metrology enables innovation in …

*… information technology*

quantum computing and communications

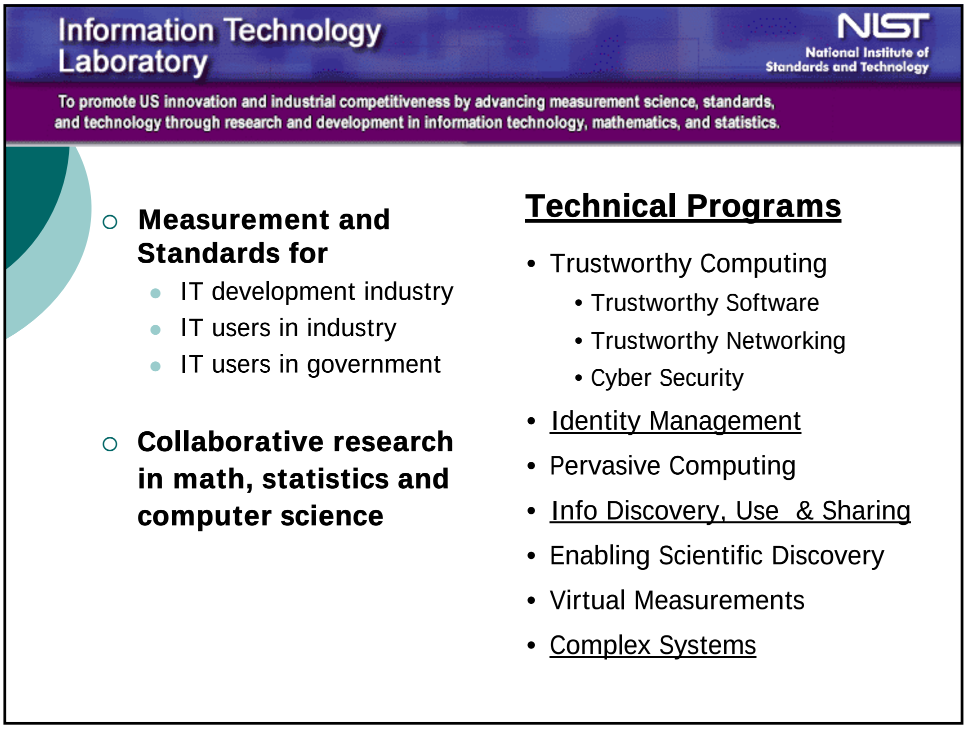
micromagnetic modeling system

TREC: evaluation of information retrieval performance



## NIST’s Laboratories

Research to advance the nation’s infrastructure for innovation:
measurements, test methods, evaluated data



Information Technology Laboratory

NIST National Institute of Standards and Technology

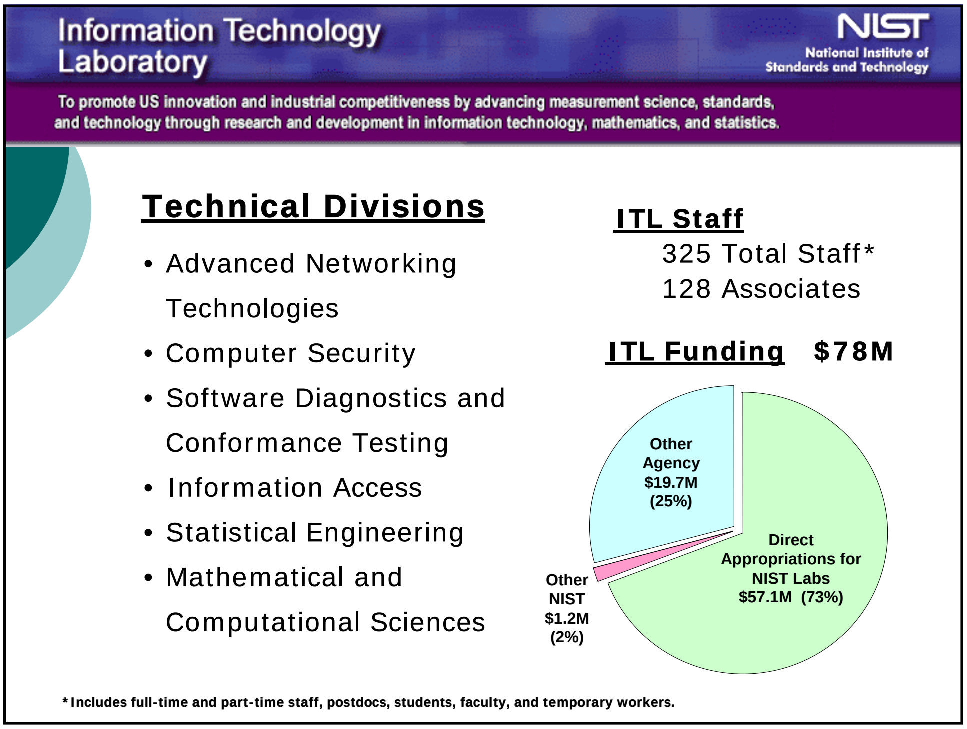
To promote US innovation and industrial competitiveness by advancing measurement science, standards, and technology through research and development in information technology, mathematics, and statistics.

○ **Measurement and Standards for**
  - IT development industry
  - IT users in industry
  - IT users in government

○ **Collaborative research in math, statistics and computer science**

## Technical Programs

- Trustworthy Computing
  - Trustworthy Software
  - Trustworthy Networking
  - Cyber Security
- Identity Management
- Pervasive Computing
- Info Discovery, Use & Sharing
- Enabling Scientific Discovery
- Virtual Measurements
- Complex Systems



Information Technology Laboratory

NIST National Institute of Standards and Technology

To promote US innovation and industrial competitiveness by advancing measurement science, standards, and technology through research and development in information technology, mathematics, and statistics.

## Technical Divisions

- Advanced Networking Technologies
- Computer Security
- Software Diagnostics and Conformance Testing
- Information Access
- Statistical Engineering
- Mathematical and Computational Sciences

**ITL Staff**

325 Total Staff*

128 Associates

**ITL Funding $78M**

Other Agency $19.7M (25%)
Direct Appropriations for NIST Labs $57.1M (73%)
Other NIST $1.2M (2%)

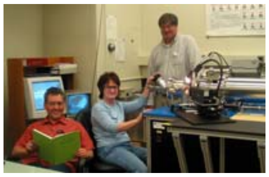
*Includes full-time and part-time staff, postdocs, students, faculty, and temporary workers.

# Math & Computational Science

***Applied Mathematics***
***High Performance Computing***
***Scientific Visualization***
***Mathematical Software***

Ron Boisvert
*Division Chief*
boisvert@nist.gov

- Collaborative Research
  - within NIST: interdisciplinary, peer to peer
  - bring expertise, facilities / high local payoff

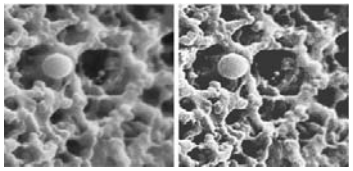


- Underlying R&D
  - research in math, CS anticipating NIST needs
  - tools, facilities to make us, customers more efficient

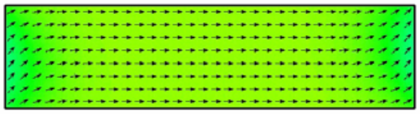
- Work with community
  - community-based measurement, standards
  - Web-based information services
  - wide distribution of tools

# Examples

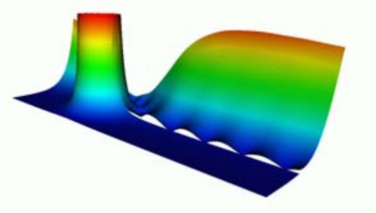
- Deblurring of images from scanning electron microscopes
  - Deconvolution
  - Inverse and ill-posed problems



- Software for modeling in nanomagnetics
  - Applied PDEs, Numerical analysis
  - Problem-solving environments



- Online handbook of special functions of applied math
  - Real and complex analysis
  - Mathematics on the Web



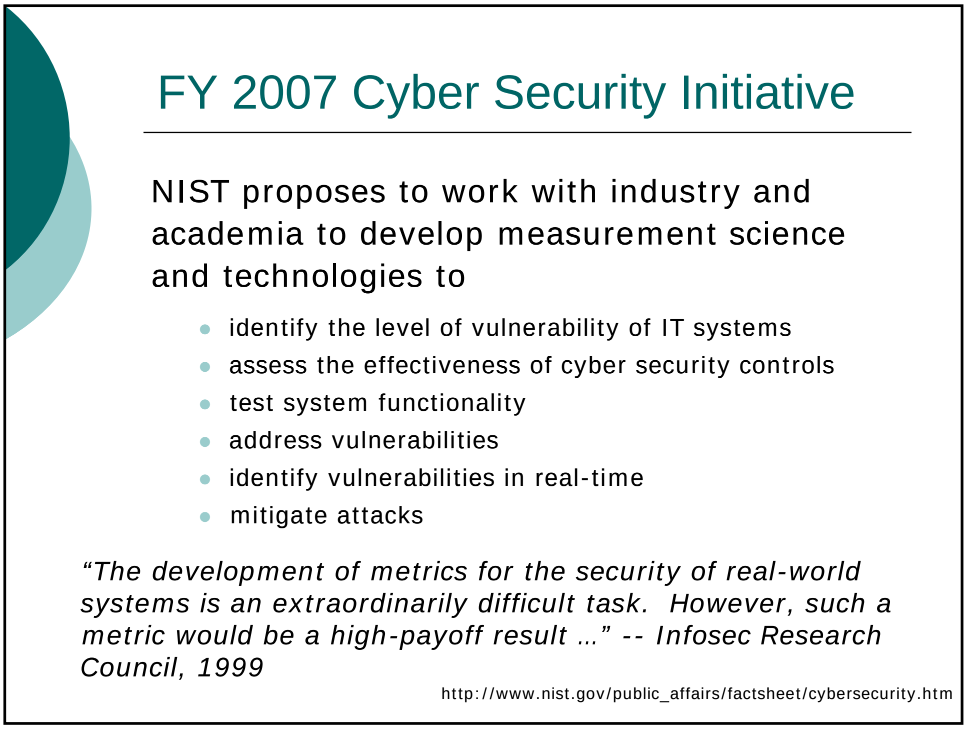
Today's subject

# NIST FY07 Cyber Security Initiative

*Innovative Technologies for National Security*

- The nation's IT infrastructure has grown phenomenally. Critical infrastructures—transportation, financial, power grids, military, intelligence systems, and health and safety—rely on computer, communication networks.
- In spite of efforts to secure, these systems remain vulnerable.
- Today's cyber security efforts are aimed at identifying particular vulnerabilities and determining whether well-known security controls are in place.
- There is no known way to measure the absolute security of a given system. Without metrics and measurement technologies, we can't determine the overall effectiveness of our controls.

http://www.nist.gov/public_affairs/factsheet/cybersecurity.htm



# FY 2007 Cyber Security Initiative

NIST proposes to work with industry and academia to develop measurement science and technologies to

- identify the level of vulnerability of IT systems
- assess the effectiveness of cyber security controls
- test system functionality
- address vulnerabilities
- identify vulnerabilities in real-time
- mitigate attacks

*"The development of metrics for the security of real-world systems is an extraordinarily difficult task. However, such a metric would be a high-payoff result ..." -- Infosec Research Council, 1999*

http://www.nist.gov/public_affairs/factsheet/cybersecurity.htm

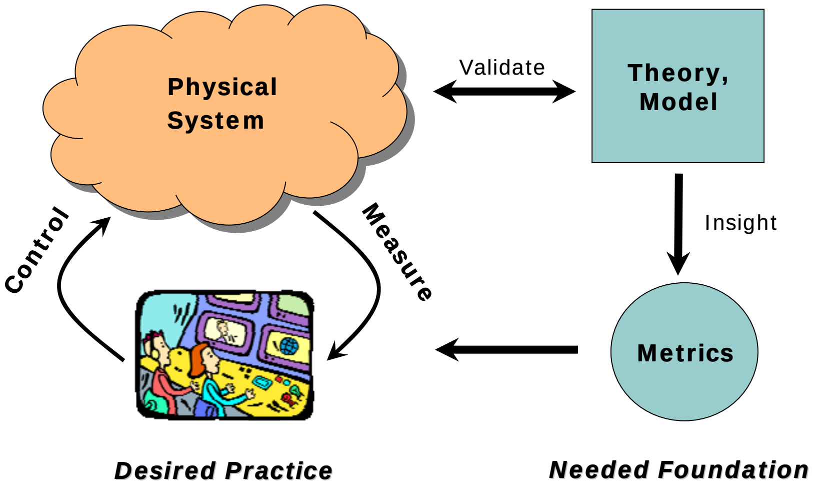
## Problem is More Fundamental

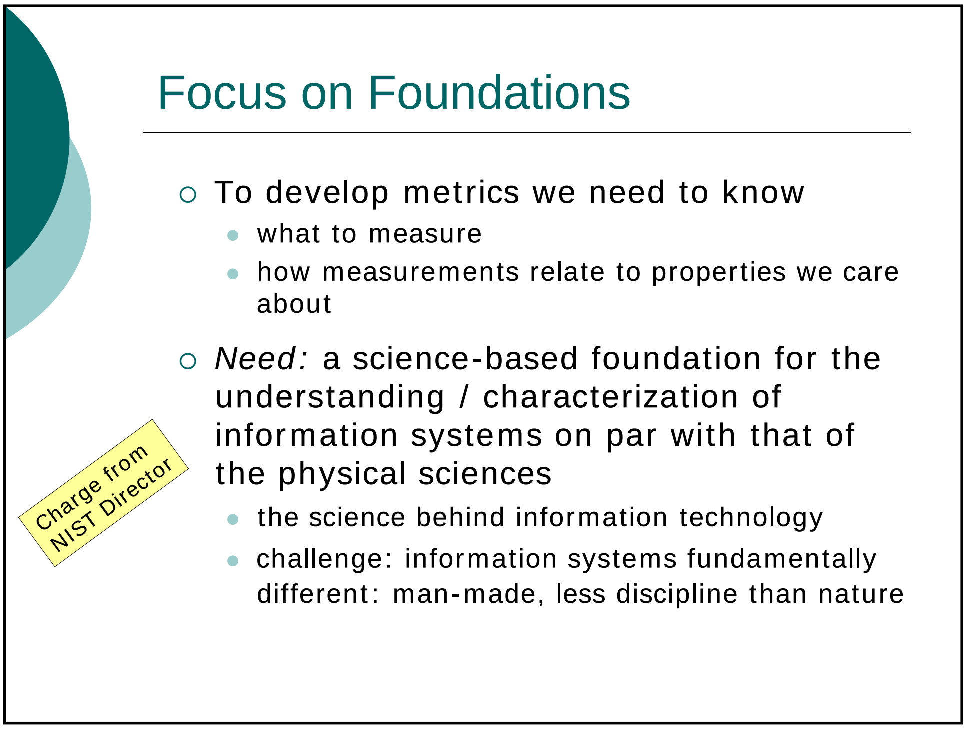
**We build, deploy large-scale information systems without complete understanding of their range of behaviors.**

*"[Despite] society's profound dependence on networks, fundamental knowledge about them is primitive. [G]lobal communication … networks have quite advanced technological implementations but their behavior under stress still cannot be predicted reliably.… There is no science today that offers the fundamental knowledge necessary to design large complex networks [so] that their behaviors can be predicted prior to building them."* — Network Science, National Research Council, 2006

## Science Foundation is Prerequisite

*Experiment – Model – Understand – Measure – Improve*

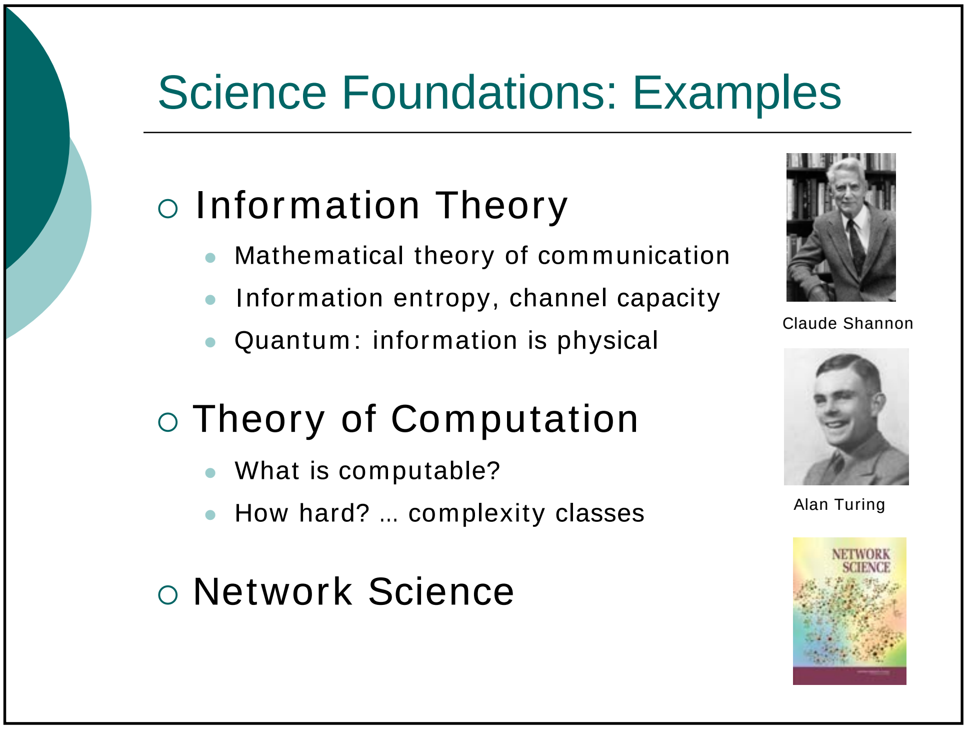




## Focus on Foundations

- To develop metrics we need to know
  - what to measure
  - how measurements relate to properties we care about
- *Need:* a science-based foundation for the understanding / characterization of information systems on par with that of the physical sciences
  - the science behind information technology
  - challenge: information systems fundamentally different: man-made, less discipline than nature

Charge from NIST Director

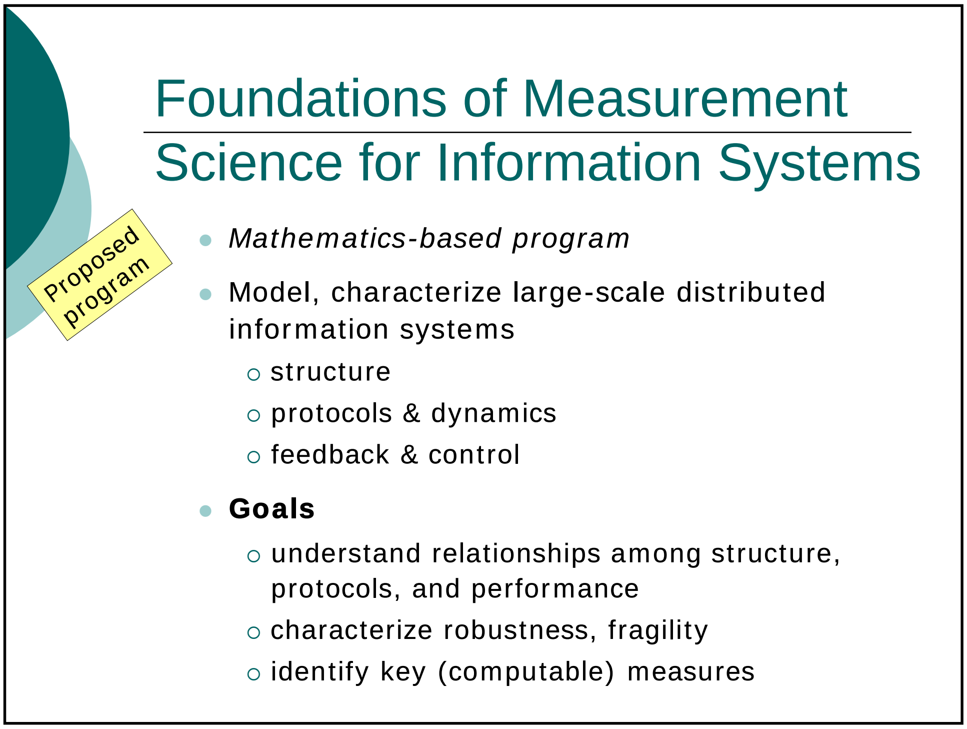


## Science Foundations: Examples

- Information Theory
  - Mathematical theory of communication
  - Information entropy, channel capacity
  - Quantum: information is physical
- Theory of Computation
  - What is computable?
  - How hard? … complexity classes
- Network Science

Claude Shannon

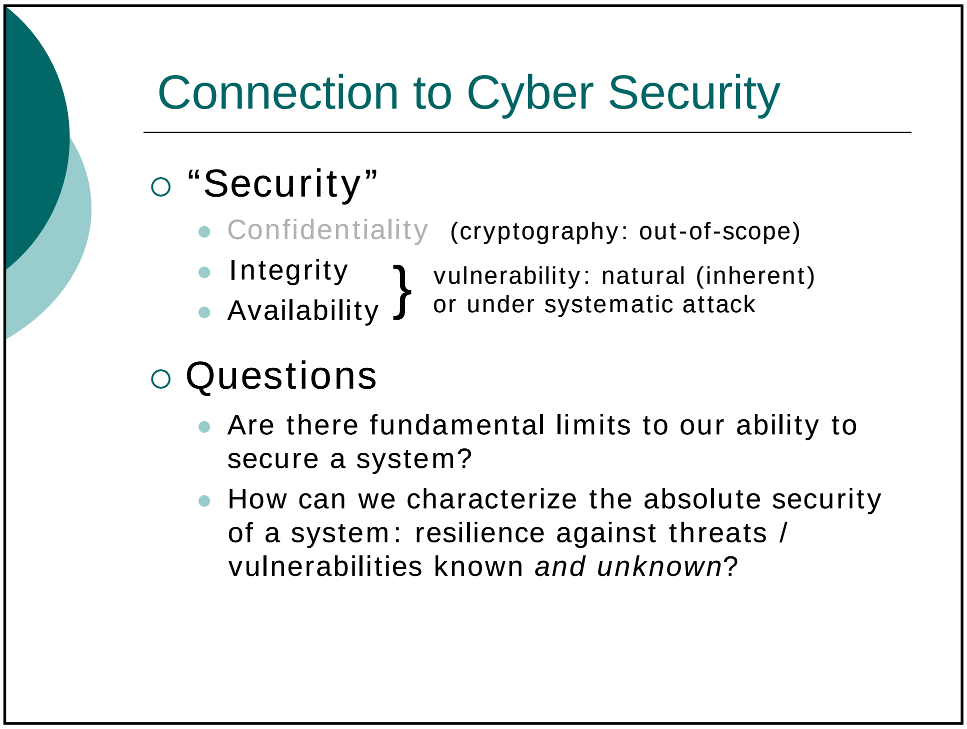
Alan Turing



## Foundations of Measurement Science for Information Systems

Proposed program

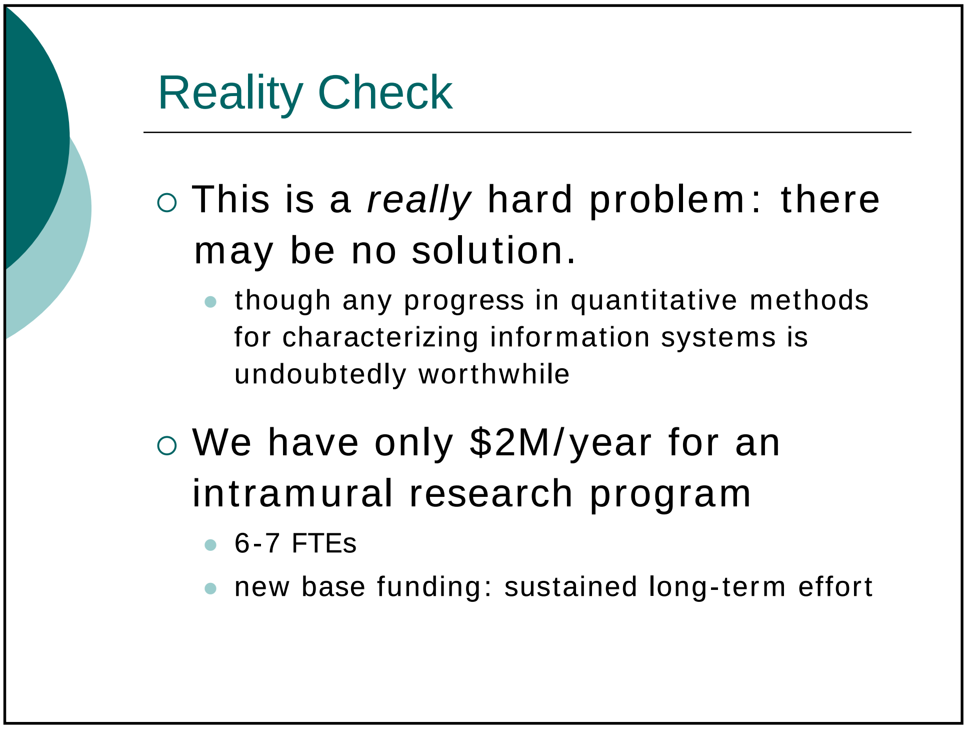
- *Mathematics-based program*
- Model, characterize large-scale distributed information systems
  - structure
  - protocols & dynamics
  - feedback & control
- **Goals**
  - understand relationships among structure, protocols, and performance
  - characterize robustness, fragility
  - identify key (computable) measures



## Connection to Cyber Security

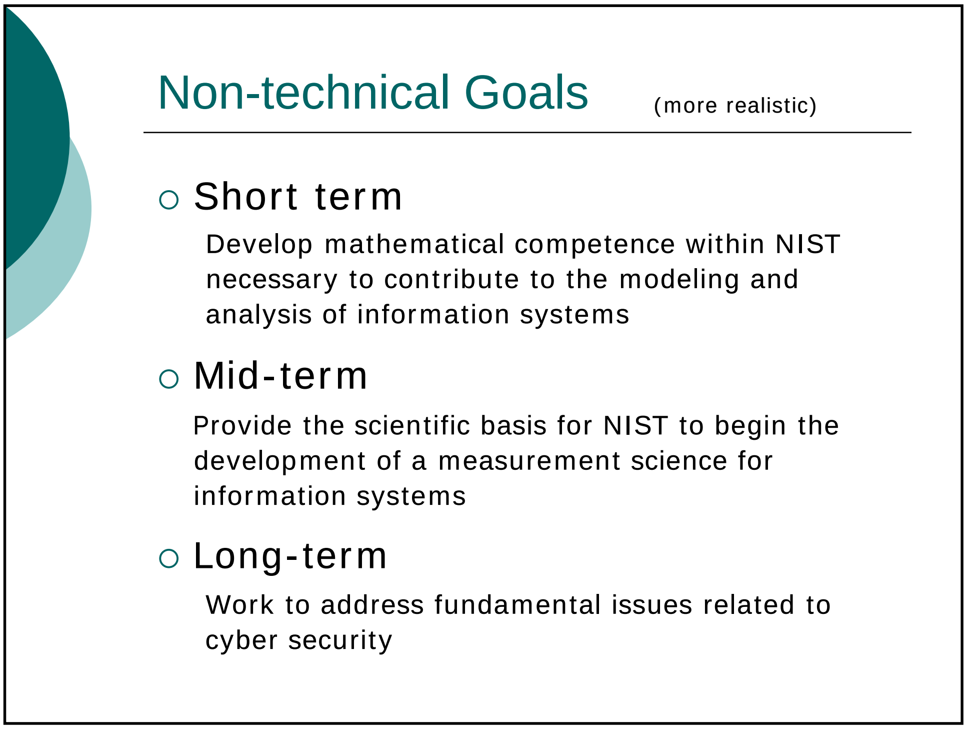
- "Security"
  - Confidentiality (cryptography: out-of-scope)
  - Integrity
  - Availability

  } vulnerability: natural (inherent) or under systematic attack
- Questions
  - Are there fundamental limits to our ability to secure a system?
  - How can we characterize the absolute security of a system: resilience against threats / vulnerabilities known *and unknown*?



## Reality Check

- This is a *really* hard problem: there may be no solution.
  - though any progress in quantitative methods for characterizing information systems is undoubtedly worthwhile
- We have only $2M/year for an intramural research program
  - 6-7 FTEs
  - new base funding: sustained long-term effort



## Non-technical Goals (more realistic)

- Short term

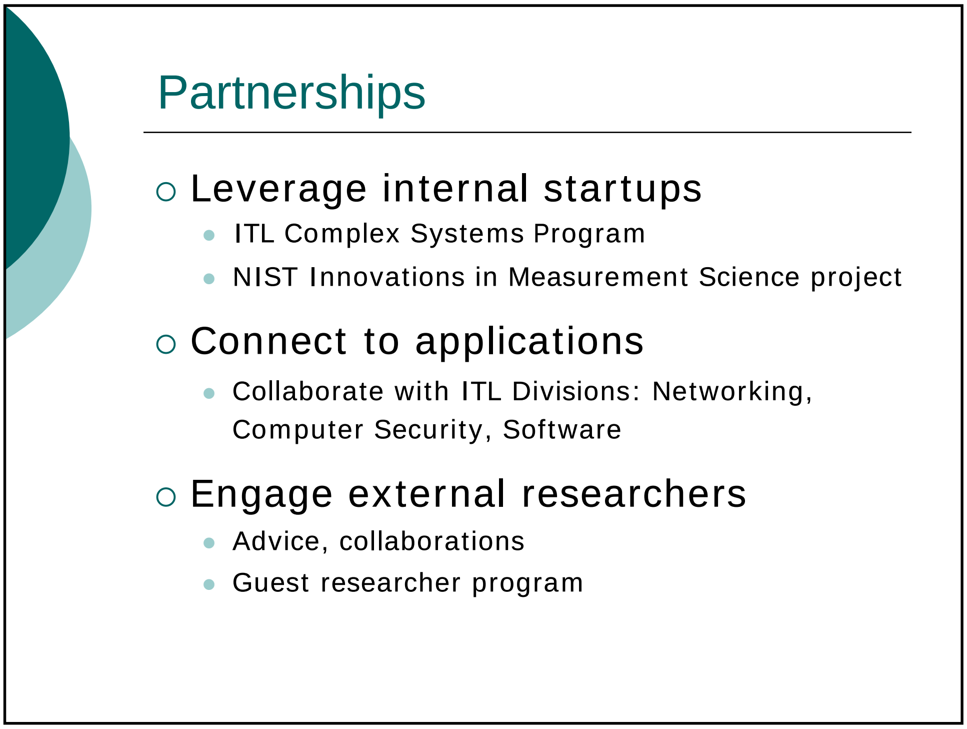
  Develop mathematical competence within NIST necessary to contribute to the modeling and analysis of information systems

- Mid-term

  Provide the scientific basis for NIST to begin the development of a measurement science for information systems

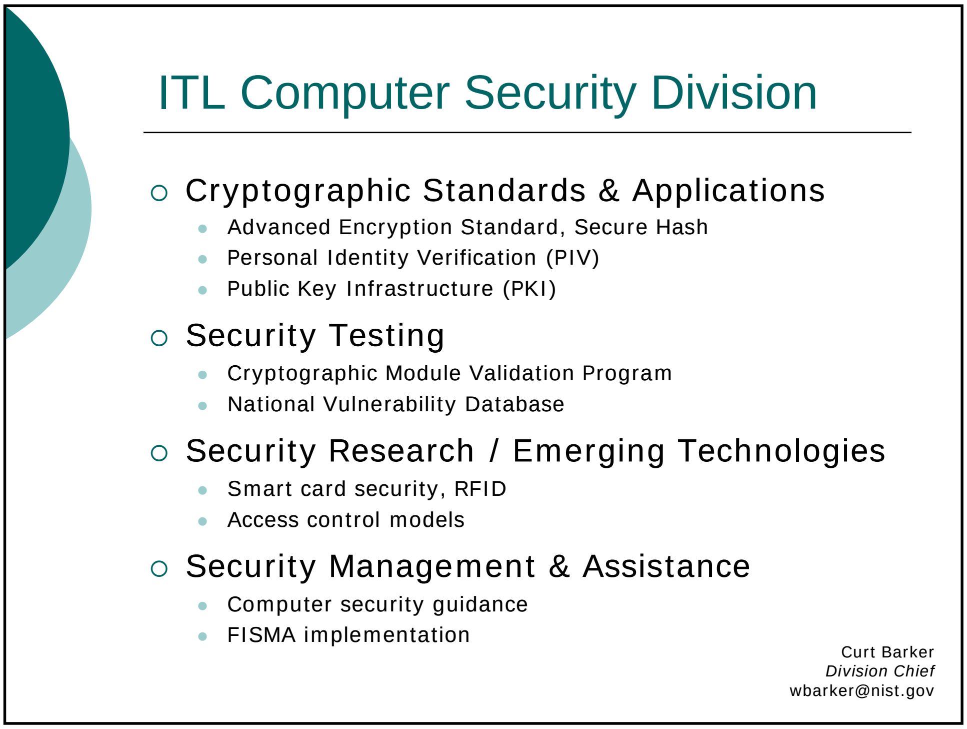
- Long-term

  Work to address fundamental issues related to cyber security



## Partnerships

- Leverage internal startups
  - ITL Complex Systems Program
  - NIST Innovations in Measurement Science project
- Connect to applications
  - Collaborate with ITL Divisions: Networking, Computer Security, Software
- Engage external researchers
  - Advice, collaborations
  - Guest researcher program



## ITL Computer Security Division

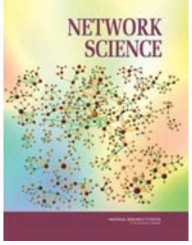
- Cryptographic Standards & Applications
  - Advanced Encryption Standard, Secure Hash
  - Personal Identity Verification (PIV)
  - Public Key Infrastructure (PKI)
- Security Testing
  - Cryptographic Module Validation Program
  - National Vulnerability Database
- Security Research / Emerging Technologies
  - Smart card security, RFID
  - Access control models
- Security Management & Assistance
  - Computer security guidance
  - FISMA implementation

Curt Barker
*Division Chief*
wbarker@nist.gov

# ITL Complex Systems Program

- Develop metrics for description, prediction and control of complex systems.
- Identify and fund (internally) projects
- Work with other NIST laboratories towards interdisciplinary efforts & seek out external partners.

New in FY2007



"In spite of society's profound dependence on networks, fundamental knowledge about them is primitive. [G]lobal communication networks have quite advanced technological implementations, but their behavior under stress still cannot be predicted reliably."

"There is no science today that offers the fundamental knowledge necessary to design large, complex networks in such a way that their behaviors can be predicted prior to building them."

complex systems

Information Technology Laboratory

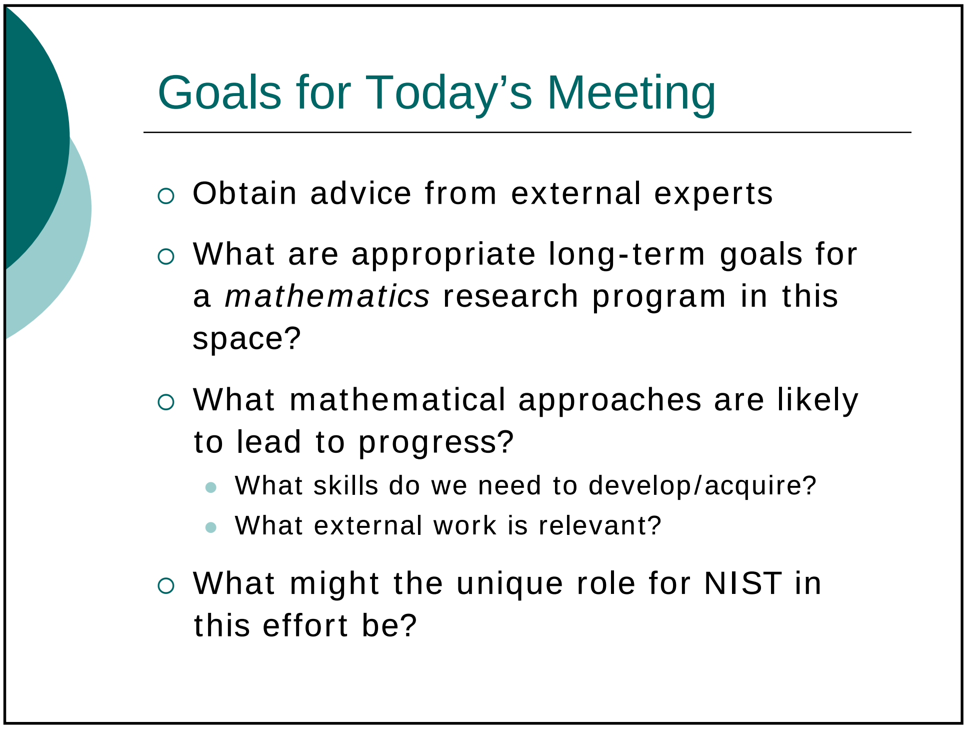
Sandy Ressler
*Program Manager*
sressler@nist.gov

# NIST Innovations in Measurement Science Program

New in FY2007

- Project: Measurement Science for Complex Information Systems
- Lead: ITL Advanced Networking Technologies Division
- **Goal:** measure, predict, control macroscopic behavior in complex information systems (e.g., Internet and distributed systems like the Grid)
  - Establish models and analysis methods that (1) are computationally tractable, (2) reveal macroscopic behavior, and (3) establish causality.
  - Characterize distributed control techniques, including: (1) economic mechanisms to elicit desired behaviors and (2) biological mechanisms to organize components

Kevin Mills
*Project Lead*
kmills@nist.gov



# Goals for Today’s Meeting

- Obtain advice from external experts
- What are appropriate long-term goals for a *mathematics* research program in this space?
- What mathematical approaches are likely to lead to progress?
  - What skills do we need to develop/acquire?
  - What external work is relevant?
- What might the unique role for NIST in this effort be?

# A-2 Network Science in Need of Measurement Science

Walter Willinger

# Network Science in Need of Measurement Science:

# Lessons Learned from Modeling the Internet

Walter Willinger
AT&T Labs-Research
walter@research.att.com

## Recap: What Network Science says about the Internet

- Concrete example: Router-level connectivity
  - Data: Large-scale traceroute experiments
  - Inference: Power-law node degree distribution
  - Modeling: Preferential attachment-type growth model
  - Model validation: “fits” the data (i.e., node degree distribution)
  - Highly publicized claims
    - High-degree nodes form a hub-like core
    - Fragile/vulnerable to targeted node removal
    - Achilles’ heel
    - Zero epidemic threshold
- Similar examples
  - Autonomous System or AS-level connectivity
  - Overlay networks (e.g., P2P, WWW)

## Fact: Network Science got it all wrong!

The Internet is exactly the opposite of what the "theory" of Network Science claims in essentially every meaningful aspect

These claims are not controversial, they are simply wrong!

So much for "Network Science" as a "science" …

## Main Question:

## What went wrong when applying Network Science to the Internet?

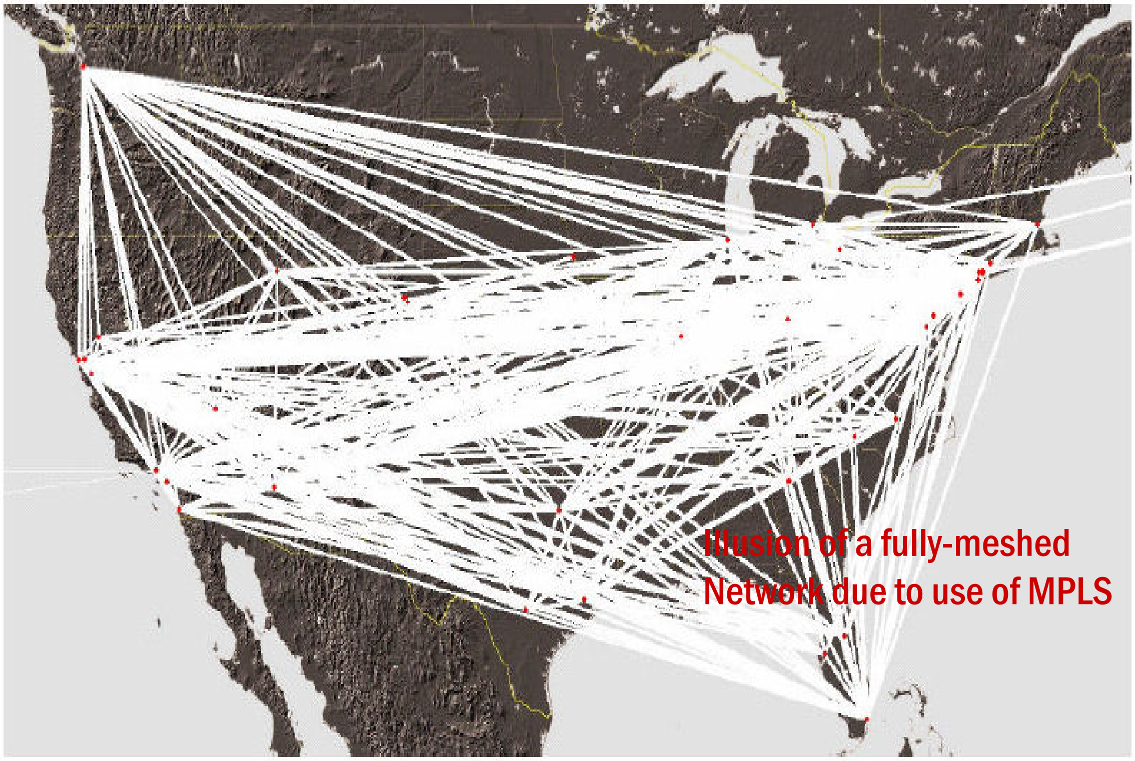
Network Science can only become a "science" when accompanied by an appropriate Measurement Science!

## Measurement Science

- Provide answers to the following type of questions
  - "Do the available Internet-related connectivity measurements support the sort of claims that can be found in the  existing complex networks literature?"
  - "For what purpose can the measurements at hand be safely used?
- Basic requirements (among others)
  - Insist on high level of data hygiene
  - Insist on a level of statistical rigor that matches the quality of the available data
  - Insist on taking model validation serious
- Illustration: ISP router-level topology

## MISTAKE #1: Lack of Data Hygiene

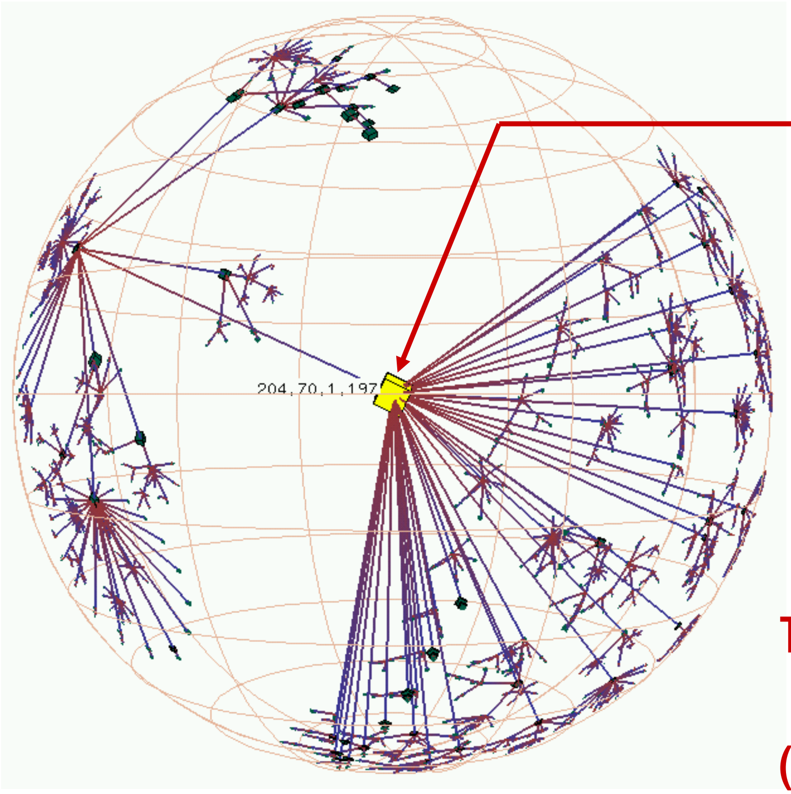
- traceroute-based measurements are ambiguous
  - traceroute is strictly about IP-level connectivity
  - traceroute cannot distinguish between high connectivity nodes that are for real and that are fake and due to underlying Layer 2 (e.g., Ethernet, ATM) or Layer 2.5 technologies (e.g., MPLS)
- traceroute-based measurements are inaccurate
  - Requires some guesswork in deciding which IP addresses/interface cards refer to the same router ("alias resolution" problem)
- traceroute-based measurements are incomplete/biased
  - IP-level connectivity is more easily/accurately inferred the closer the routers are to the traceroute source(s)
  - Node degree distribution is inferred to be of the power-law type even when the actual distribution is not



Illusion of a fully-meshed Network due to use of MPLS

*Background image courtesy JHU, applied physics labs*

http://www.cs.washington.edu/research/networking/rocketfuel/



- www.savvis.net
- managed IP and hosting company
- founded 1995
- offering “private IP with ATM at core”

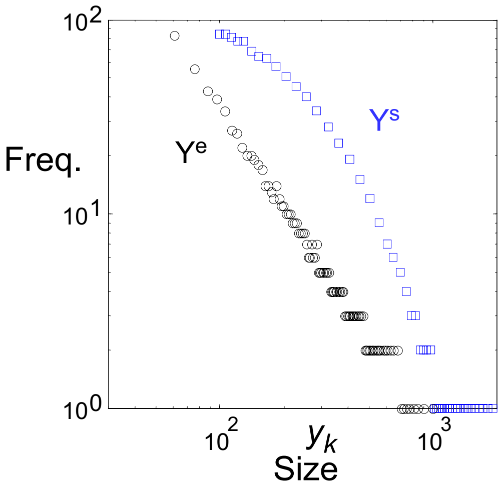
This “node” is an entire network! (not just a router)

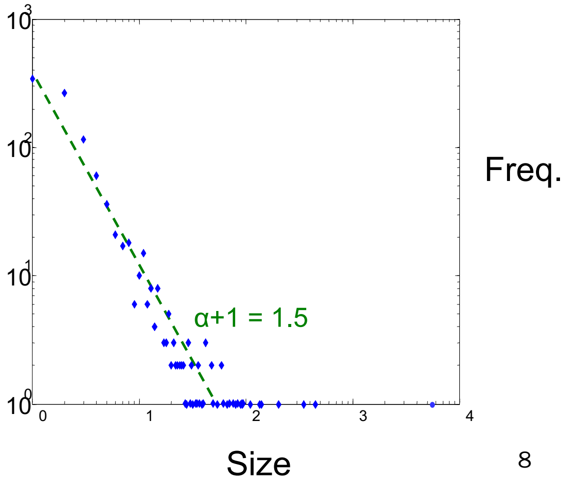
http://www.caida.org/tools/measurement/skitter/

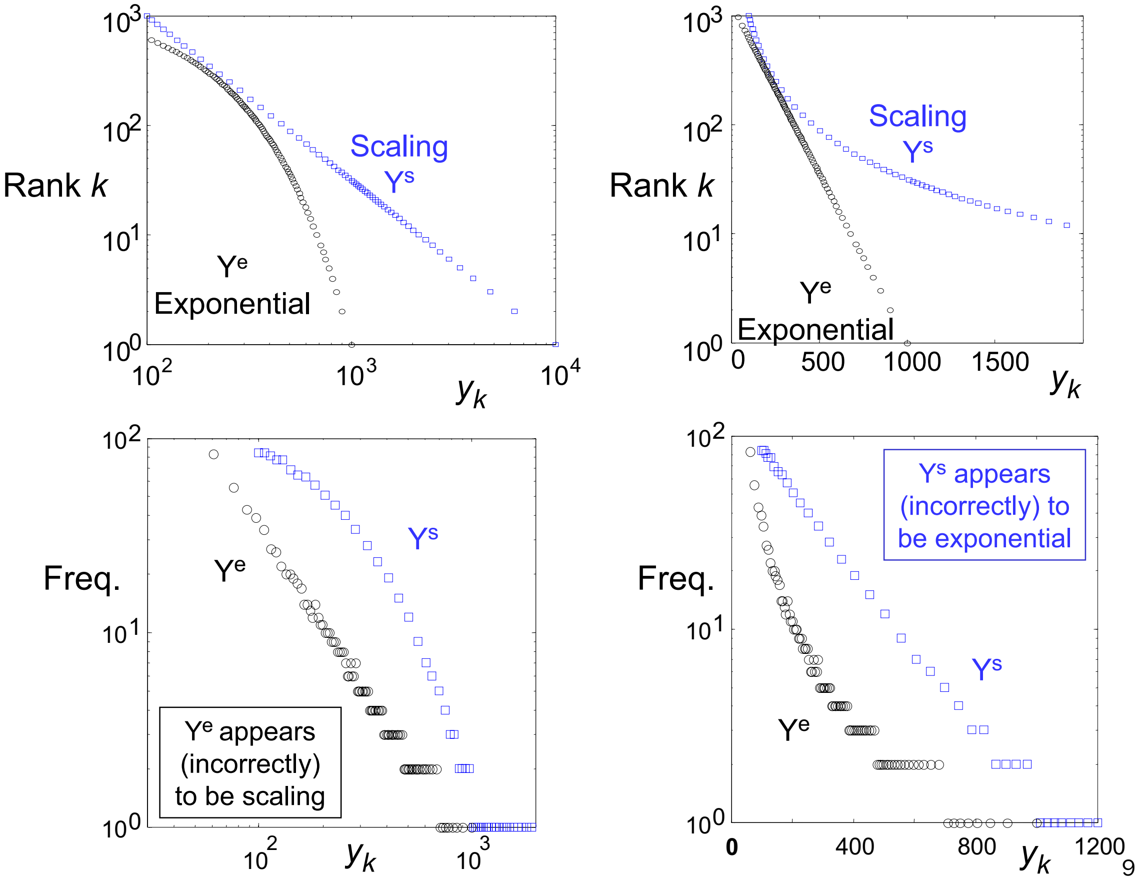
## MISTAKE #2: Lack of Statistical Rigor

Given: Samples from an exponential distribution
Want: Claim power law behavior
Recipe: Use size-frequency plots!

Given: Samples from a Pareto distribution with $\alpha$=1.0
Want: Claim power law with $\alpha$=1.5
Recipe: Use size-frequency plots!

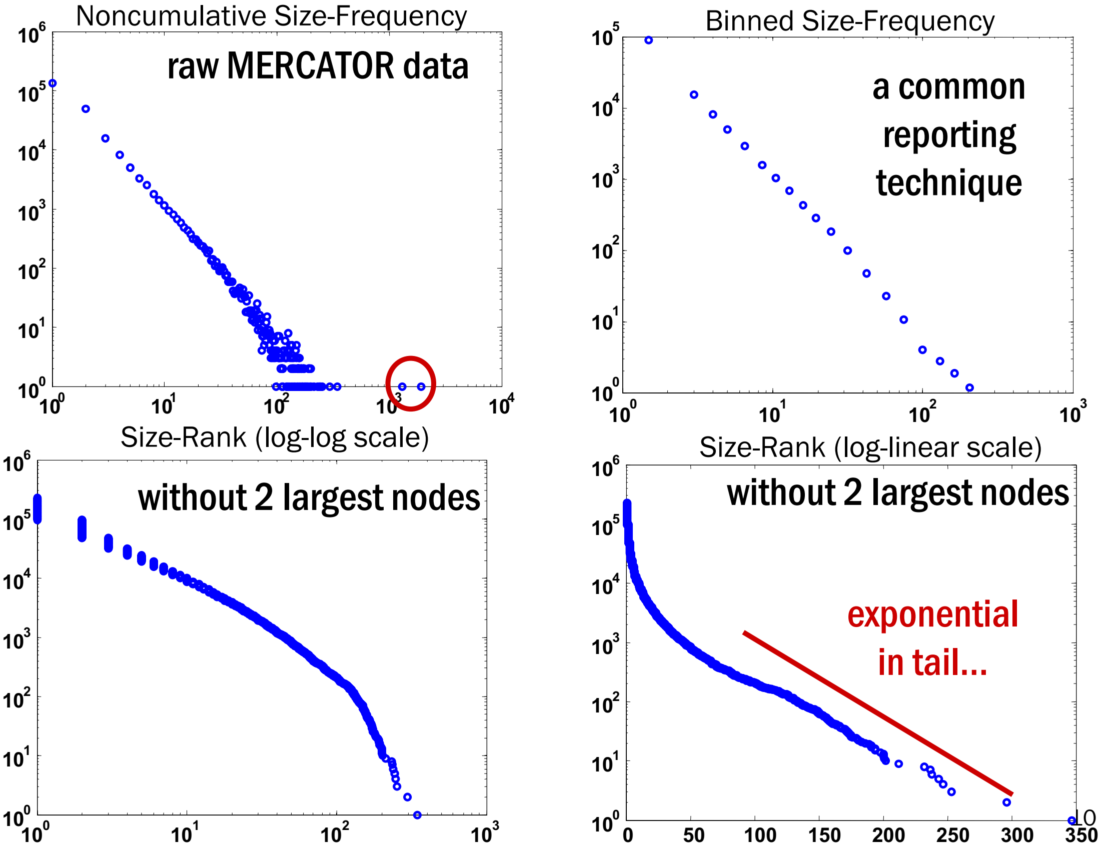






$Y^e$ appears (incorrectly) to be scaling

$Y^s$ appears (incorrectly) to be exponential



Noncumulative Size-Frequency

**raw MERCATOR data**

Binned Size-Frequency

**a common reporting technique**

Size-Rank (log-log scale)

**without 2 largest nodes**

Size-Rank (log-linear scale)

**without 2 largest nodes**

**exponential in tail…**

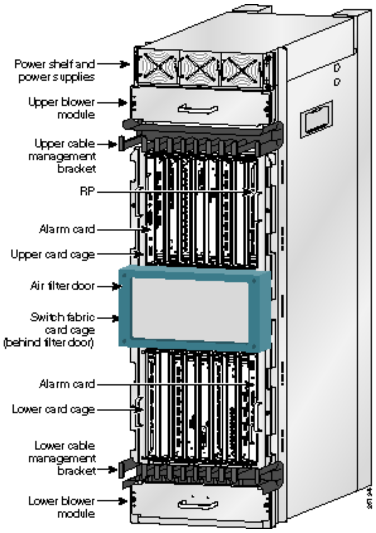
## MISTAKE #3: Lack of serious Model Validation

- Mathematical Modeling 101
  - For one and the same observed phenomenon, there are usually many different explanations/models
  - All models are wrong, but some are “damned lies”
- Model validation ≠ data fitting
  - The ability to reproduce a few graph statistics does not constitute “serious” model validation
  - Which of the observed properties does a proposed model have to satisfy before it is deemed “valid”?
- What constitutes “serious” model validation?
  - What new kinds of measurements does the proposed model suggest for the purpose of model validation

## Cisco 12000 Series Routers

- Modular in design, creating flexibility in configuration.
- Router capacity is constrained by the number and speed of line cards inserted in each slot.

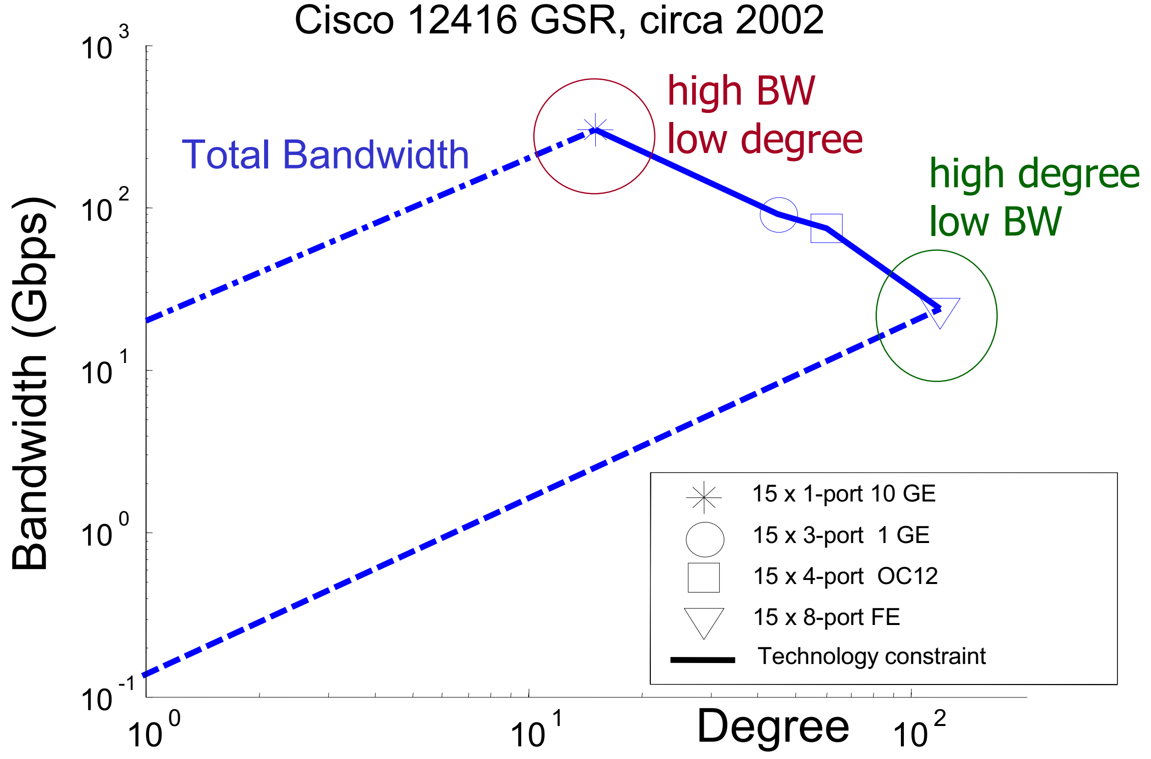
| Chassis | Rack size | Slots | Switching Capacity |
|---|---|---|---|
| 12416 | Full | 16 | 320 Gbps |
| 12410 | 1/2 | 10 | 200 Gbps |
| 12406 | 1/4 | 6 | 120 Gbps |
| 12404 | 1/8 | 4 | 80 Gbps |


Power shelf and power supplies
Upper blower module
Upper cable management bracket
RP
Alarm card
Upper card cage
Air filter door
Switch fabric card cage (behind filter door)
Alarm card
Lower card cage
Lower cable management bracket
Lower blower module

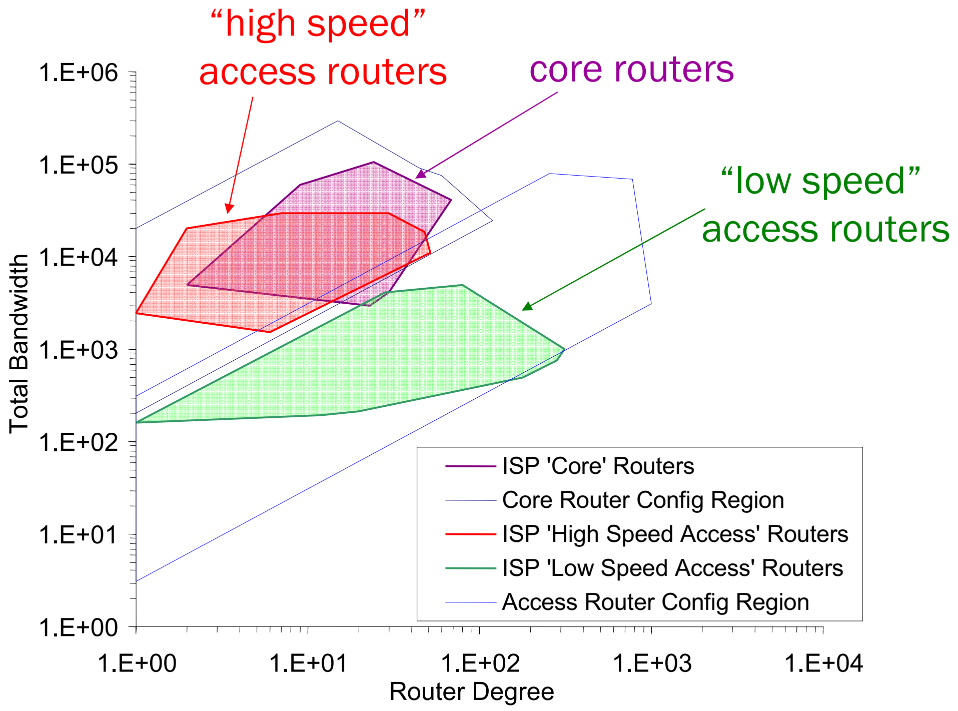
Source: www.cisco.com

## Router Technology Constraint

Cisco 12416 GSR, circa 2002



## AT&T Router Deployment (c.2003)



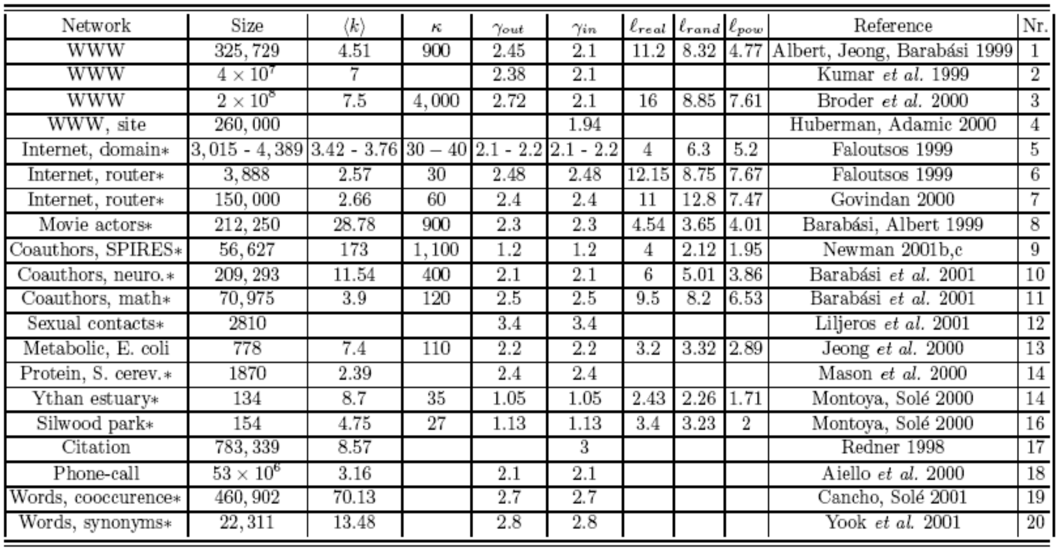
## Network Science and the Internet: “Lies, damned lies and statistics”

- How to lie with statistics …
  - Power-law (scale-free) node degree distribution
- (White) lies …
  - Preferential attachment-type models
- Damned lies …
  - Achilles’ heel
  - Fragile/vulnerable to targeted node removal
  - Zero epidemic threshold
- Bad analysis of bad data = bad models (“damned lies”)
  - ***“Bad [models] are potentially important: they can be used to stir up public outrage or fear; they can distort our understanding of our world; and they can lead us to make poor policy choices.” (J. Best)***

## How to avoid such fallacies: A case for Measurement Science

- Make node degree distribution a non-issue
  - Good reasons
    - High-quality data but low variability (e.g., exponential)
    - Low-quality data
    - High-quality data and high variability (e.g., power-laws)
  - Preferential attachment-type models
    - dead on arrival
  - Only reasonable alternative
    - Bring in and rely on domain knowledge
- What new kinds of measurements does the proposed model suggest for the purpose of model validation
  - Preferential attachment-type models: None
  - HOT models: Check router configs against existing router technology

# What about other applications of Network Science?  Same story!



| Network | Size | $\langle k \rangle$ | $\kappa$ | $\gamma_{out}$ | $\gamma_{in}$ | $\ell_{real}$ | $\ell_{rand}$ | $\ell_{pow}$ | Reference | Nr. |
|---|---|---|---|---|---|---|---|---|---|---|
| WWW | 325, 729 | 4.51 | 900 | 2.45 | 2.1 | 11.2 | 8.32 | 4.77 | Albert, Jeong, Barabási 1999 | 1 |
| WWW | $4 \times 10^7$ | 7 | | 2.38 | 2.1 | | | | Kumar *et al.* 1999 | 2 |
| WWW | $2 \times 10^8$ | 7.5 | 4, 000 | 2.72 | 2.1 | 16 | 8.85 | 7.61 | Broder *et al.* 2000 | 3 |
| WWW, site | 260, 000 | | | | 1.94 | | | | Huberman, Adamic 2000 | 4 |
| Internet, domain* | 3, 015 - 4, 389 | 3.42 - 3.76 | 30 – 40 | 2.1 - 2.2 | 2.1 - 2.2 | 4 | 6.3 | 5.2 | Faloutsos 1999 | 5 |
| Internet, router* | 3, 888 | 2.57 | 30 | 2.48 | 2.48 | 12.15 | 8.75 | 7.67 | Faloutsos 1999 | 6 |
| Internet, router* | 150, 000 | 2.66 | 60 | 2.4 | 2.4 | 11 | 12.8 | 7.47 | Govindan 2000 | 7 |
| Movie actors* | 212, 250 | 28.78 | 900 | 2.3 | 2.3 | 4.54 | 3.65 | 4.01 | Barabási, Albert 1999 | 8 |
| Coauthors, SPIRES* | 56, 627 | 173 | 1, 100 | 1.2 | 1.2 | 4 | 2.12 | 1.95 | Newman 2001b,c | 9 |
| Coauthors, neuro.* | 209, 293 | 11.54 | 400 | 2.1 | 2.1 | 6 | 5.01 | 3.86 | Barabási *et al.* 2001 | 10 |
| Coauthors, math* | 70, 975 | 3.9 | 120 | 2.5 | 2.5 | 9.5 | 8.2 | 6.53 | Barabási *et al.* 2001 | 11 |
| Sexual contacts* | 2810 | | | 3.4 | 3.4 | | | | Liljeros *et al.* 2001 | 12 |
| Metabolic, E. coli | 778 | 7.4 | 110 | 2.2 | 2.2 | 3.2 | 3.32 | 2.89 | Jeong *et al.* 2000 | 13 |
| Protein, S. cerev.* | 1870 | 2.39 | | 2.4 | 2.4 | | | | Mason *et al.* 2000 | 14 |
| Ythan estuary* | 134 | 8.7 | 35 | 1.05 | 1.05 | 2.43 | 2.26 | 1.71 | Montoya, Solé 2000 | 14 |
| Silwood park* | 154 | 4.75 | 27 | 1.13 | 1.13 | 3.4 | 3.23 | 2 | Montoya, Solé 2000 | 16 |
| Citation | 783, 339 | 8.57 | | | 3 | | | | Redner 1998 | 17 |
| Phone-call | $53 \times 10^6$ | 3.16 | | 2.1 | 2.1 | | | | Aiello *et al.* 2000 | 18 |
| Words, cooccurence* | 460, 902 | 70.13 | | 2.7 | 2.7 | | | | Cancho, Solé 2001 | 19 |
| Words, synonyms* | 22, 311 | 13.48 | | 2.8 | 2.8 | | | | Yook *et al.* 2001 | 20 |

# A-3 Some Issues of Network Topology

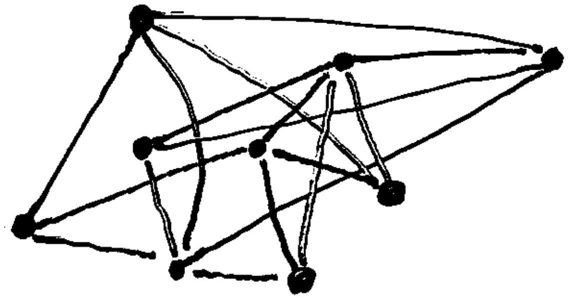
**Edward Ott**

# SOME ISSUES OF NETWORK TOPOLOGY

Edward Ott
University of Maryland

Topics:

- Node & link ranking
- Percolation
- Community structure



# NODE AND LINK RANKING

- Web page ranking

  Google ~~[illegible]~~ etc.

  Ref. 'The $25,000,000,000 Eigenvector,' SIAM Rev. (2006)

- Dynamics

  $$\overset{\text{Adj. matrix}}{A}u = \lambda u$$

  $\lambda_{max}$ is 'determining' for Sync., Perc., etc.

  ⇒ Rank nodes or links by the size of the effect of their removal on $\lambda_{max}$

  Ref. Restrepo, Ott, Hunt, Phys. Rev. Lett. 97, 094102 (2006).

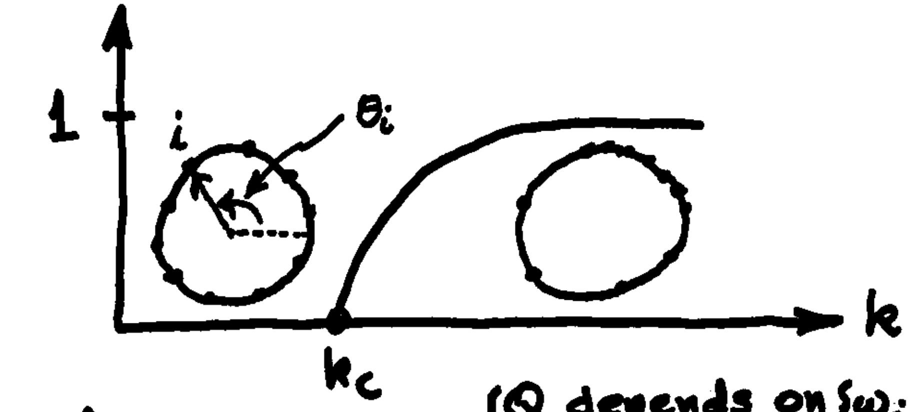
  'Dynamical importance'

- Issue: Ranking network component importance depends on the particular process of interest.

# SYNC

$$d\theta_i/dt = \omega_i + k \sum_j A_{ij} \sin(\theta_j - \theta_i)$$

$r =$ 'Coherence' $= |\langle \exp i\theta_i \rangle|$



$$k_c = Q/\lambda_{max} \quad \begin{cases} Q \text{ depends on } \{\omega_i\} \\ \lambda_{max} \text{ reflects topol.} \end{cases}$$

- ↑ Separates the effects of node properties and topology.

- Addresses 'relation of ntk. structure and function'

Ref. Restreppo, O., Hunt, Physica D (2006).

Jestion: Are there other simple relationships between topol. & dynamical behavior appropriate to other behaviors/functions

# PERCOLATION

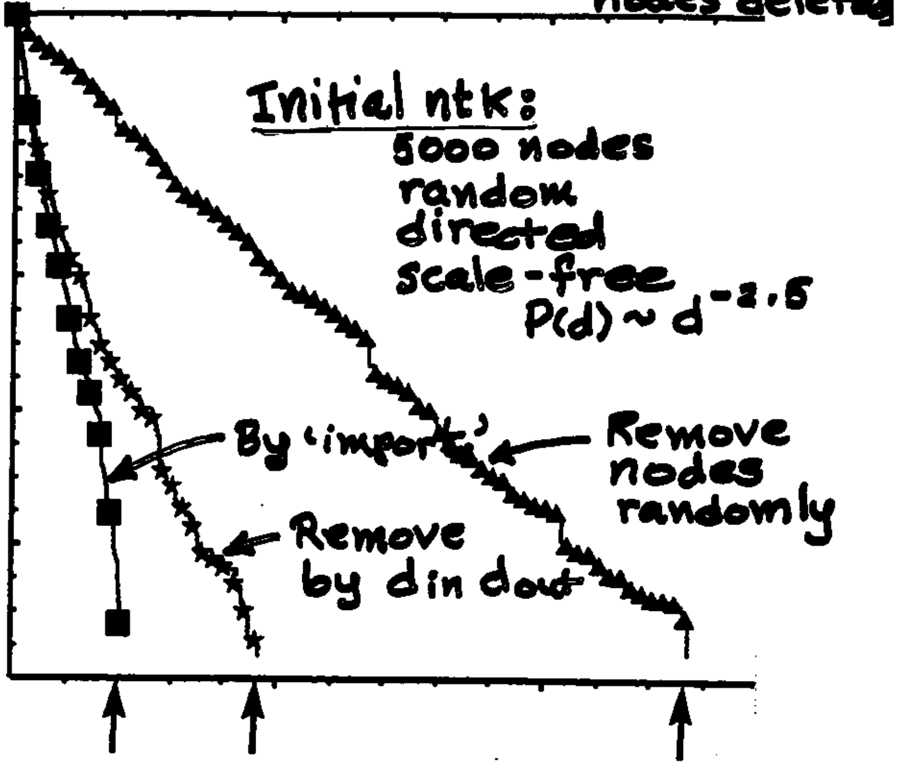
- Q. How many nodes (or links) does one have to remove to break a large graph of N nodes into many small ($\ll N$) disconnected pieces?

- The answer depends on which nodes (or links are removed. E.g.,
  - Remove by degree
  - Remove randomly
  - Remove by importance

- Relevance
  - Degradation.
  - Attack.
  - Immunization to prevent epidemic spread.

•• Q. What can be done for more complicated situations where assessing system breakdown is not a purely topological issue?

Size of largest component] ÷ [# of remaing nodes] vrs [fraction of nodes deleted]

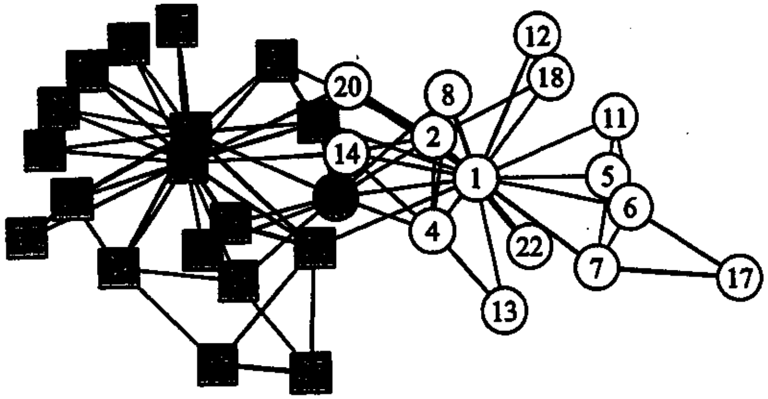


- Arrows from theory based on $\lambda_{max} = 1$ at Perc. threshold
- Ref.: Restreppo, Ott, Hunt, "Weighted Perc. on Directed Ntks.," arXiv: 0704.0491 V1

DISCOVERING COMMUNITIES IN NTKS.

- Given a ntk. graph, break it into communities.
- What is a 'community'?
- One answer: A group of nodes with many connections to other nodes in the group, but with relatively few to nodes in other groups.
- Methods to find communities:
  Laplacian spectoral methods (CS)
  Hierarchal clustering (Sociology)
  Modularity (Girvan, Newman, 2002)
  Resistor ntk (Wu, Huberman, 2003)
  Potts model (Bornholtz & Riechart '04)
- Review article: M.E.J. Newman
  European Physical Journal B (2004).
- Question: Other defs. of 'community' may be appropriate depending on the relevant ntk. function. Directed and weighted ntks.?

## EXAMPLE: Zachary's Karate Club Network



- Friendship ntk. constructed by Zachary (1977) over a period of 2 yrs.
- Club split in two due to a dispute between two factions.
- Test of a community discovery algorithm: does it predict correct split from the friendship graph?

## CONCLUSIONS

- Ranking
- Sync.
- Perc.
- Community structure

- •• Simple graph models are useful for defining questions, formulating solution techniques, gaining intuition, etc.
- •• Real cases may require analysis of more complex, application specific models. Are simple models useful? How general are phenomena and solution techniques found in simple problems ('universality')?

# A-4 Foundations of Measurement Science for Information Technology

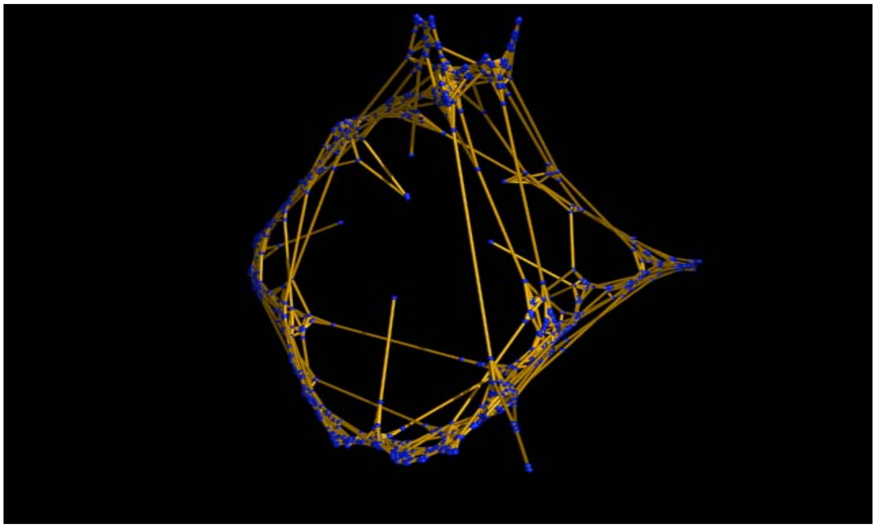
John Gilbert

## Foundations of Measurement Science for Information Technology

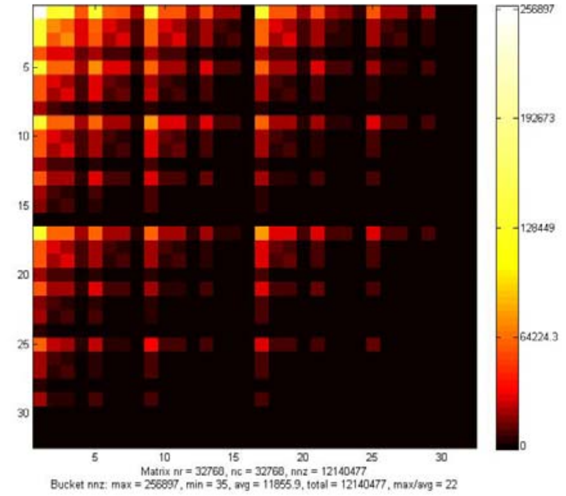
**A few research areas in measurement science for complex networks:**

- Measurement of global properties of networks:
  - Not just density, diameter, degree distribution, etc.
  - Connectivity, robustness
  - Spectral properties: Laplacian eigenvectors, Cheeger bounds, …
  - Other global measures of complexity?
  - Sensitivity analysis of all of the above
  - Stochastic settings for all of the above
- Multiscale modeling of complex networks
- Building useful reference data sets and generators
- Fundamentals of high-performance combinatorial computing
- Tools: How will results be used by nonspecialists?

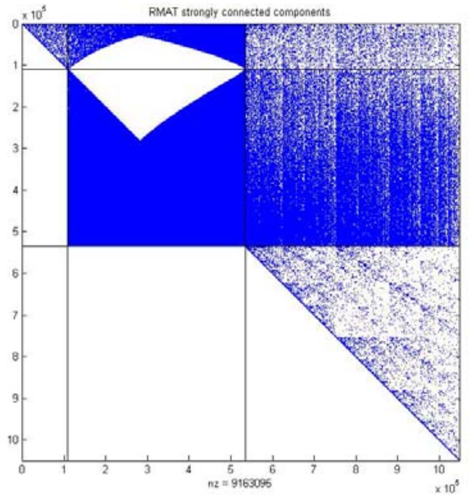
## SSCA Benchmark Graph (scale 10)

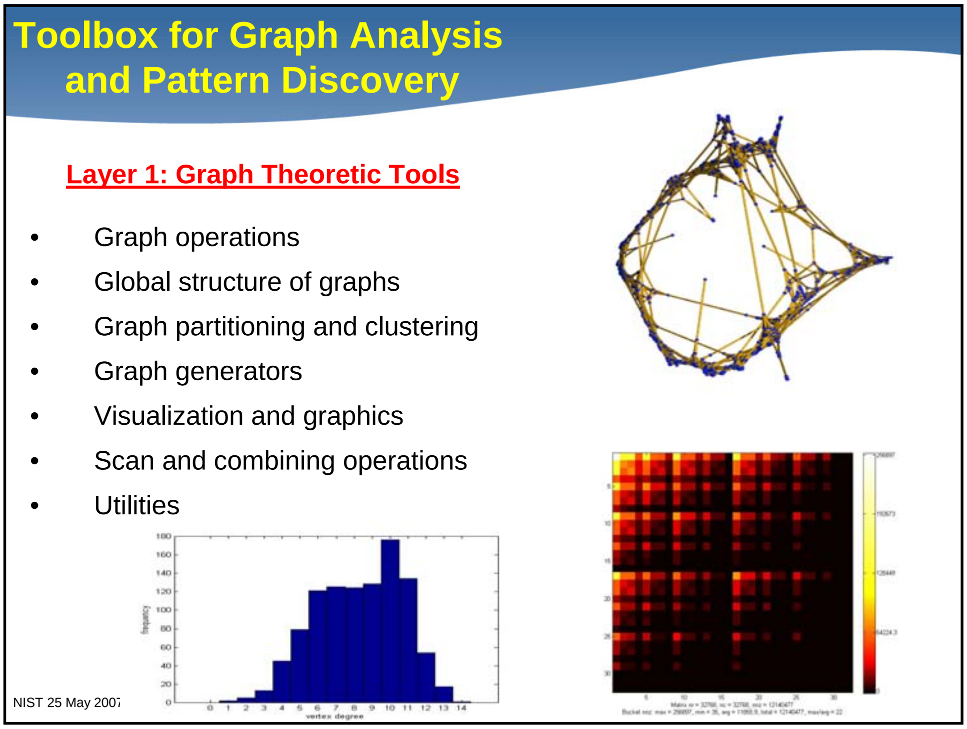


## RMAT Approximate Power-Law Graph



## Strongly Connected Components

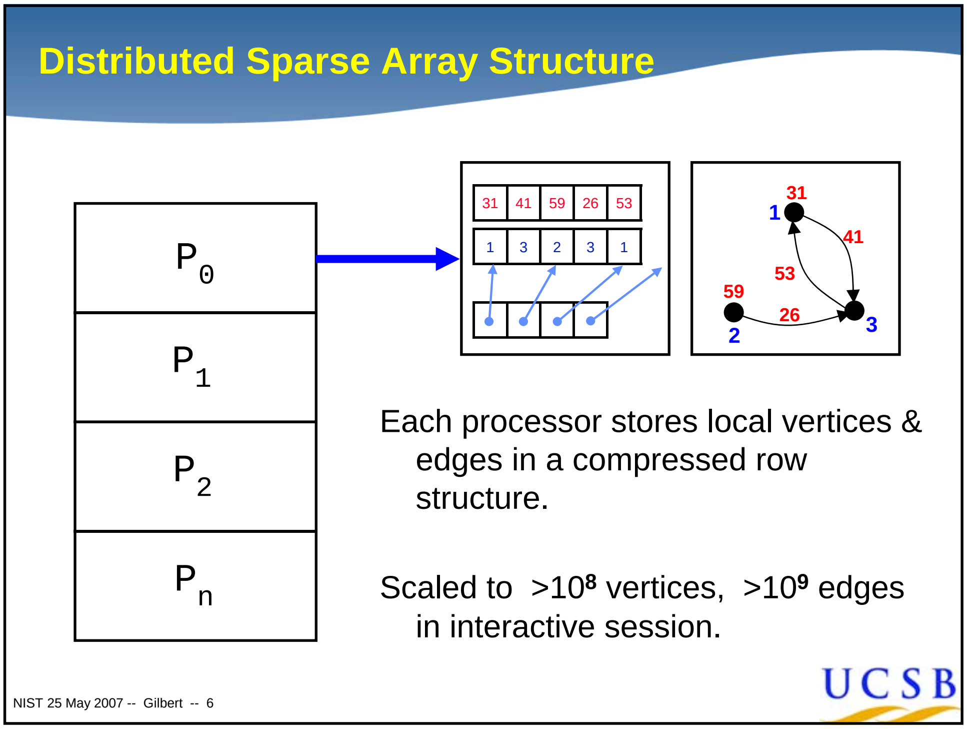




## Toolbox for Graph Analysis and Pattern Discovery

**Layer 1: Graph Theoretic Tools**

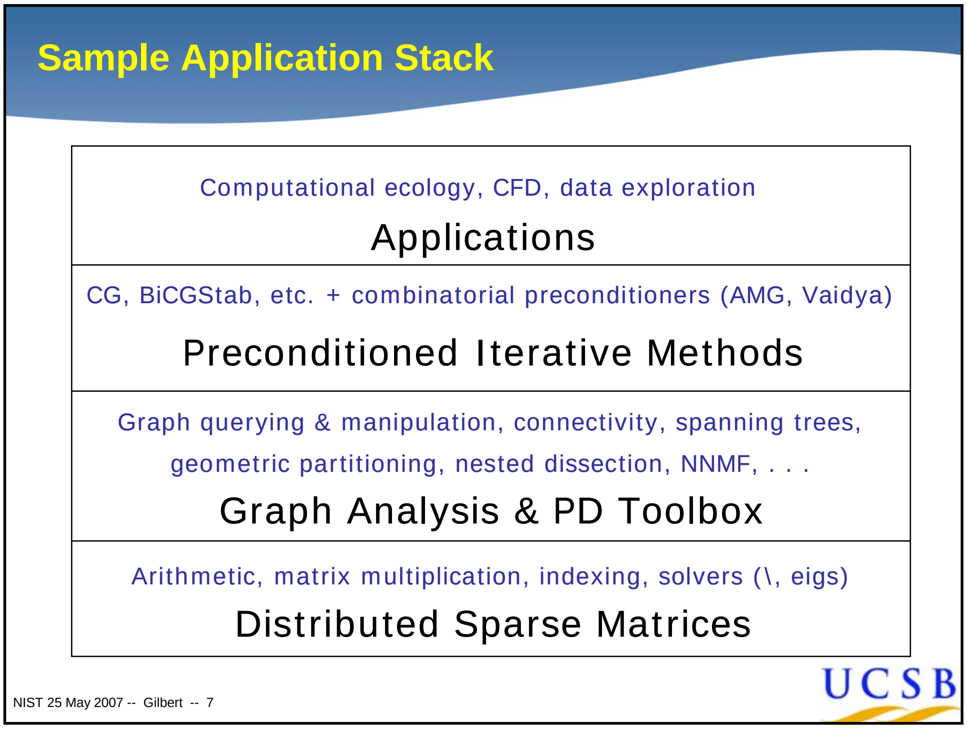
- Graph operations
- Global structure of graphs
- Graph partitioning and clustering
- Graph generators
- Visualization and graphics
- Scan and combining operations
- Utilities



## Distributed Sparse Array Structure

Each processor stores local vertices & edges in a compressed row structure.

Scaled to $>10^8$ vertices, $>10^9$ edges in interactive session.



## Sample Application Stack

Computational ecology, CFD, data exploration

**Applications**

CG, BiCGStab, etc. + combinatorial preconditioners (AMG, Vaidya)

**Preconditioned Iterative Methods**

Graph querying & manipulation, connectivity, spanning trees, geometric partitioning, nested dissection, NNMF, . . .

**Graph Analysis & PD Toolbox**

Arithmetic, matrix multiplication, indexing, solvers (\, eigs)

**Distributed Sparse Matrices**

UCSB