New Challenges in IT

Peter Lee, Department Head

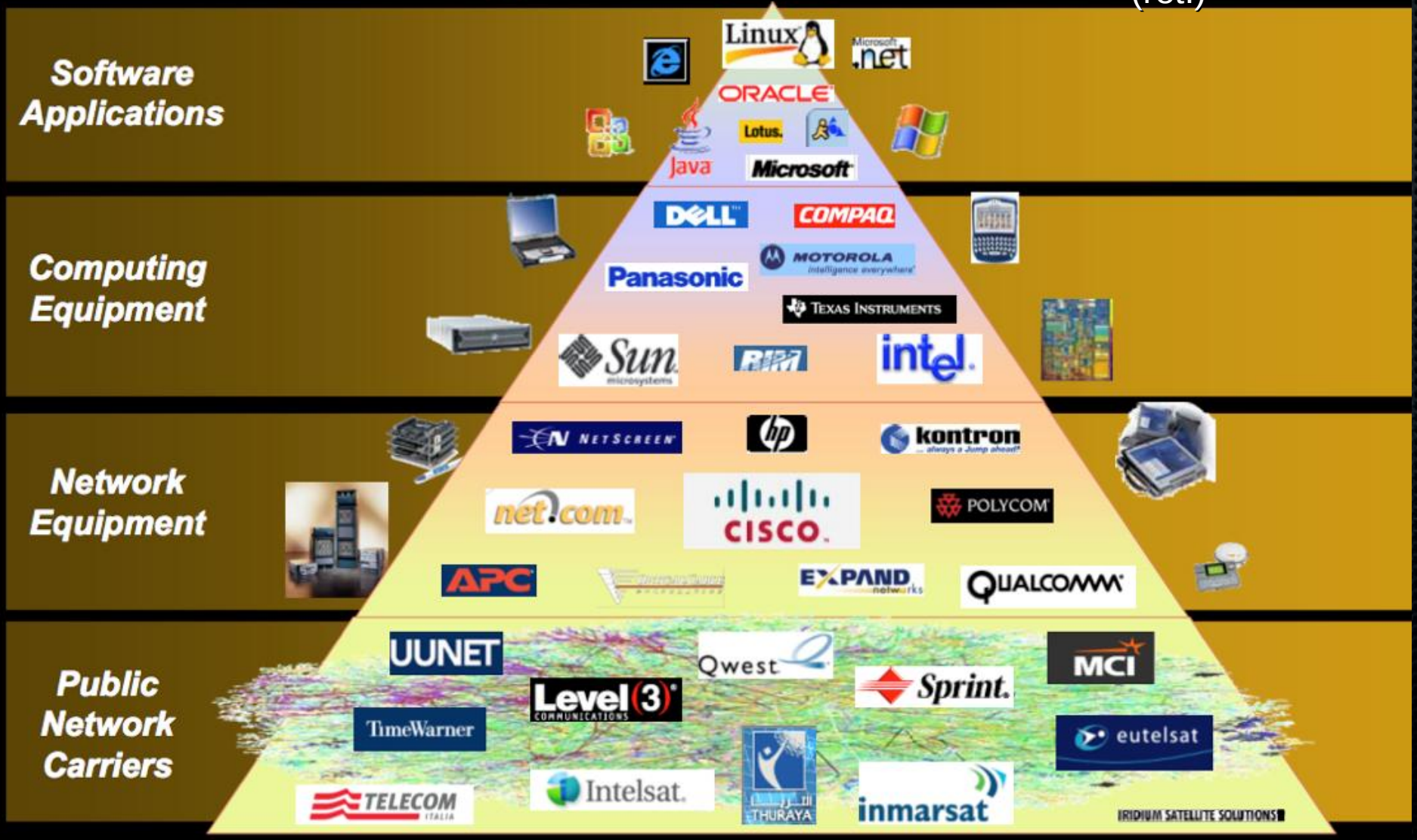
Computer Science Department
Carnegie Mellon University
April 24, 2008

IT research landscape

- Drivers of change in critical national IT needs:
  - globalization
  - data-intensive science, engineering, and commerce
  - multicore computing and the end of Moore’s law
  - Internet as a social and computing mechanism

DOD Networks
Commercial Technology Inside

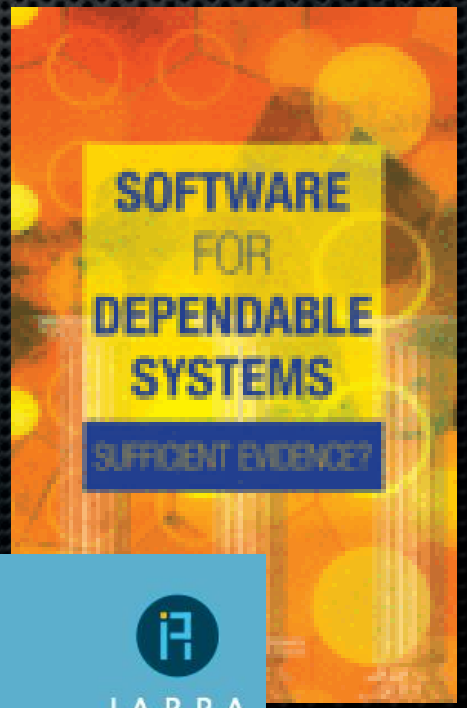
Courtesy, Gen. Steven Boutelle (ret.)



80% of DoD IT is COTS (including the classified bits!)

The certification challenge

- Today, **certification** (esp. in defense and intelligence agencies) **does not focus on the actual code**
  - based on standards and conformance tests
  - a significant national security risk
- Is there a new certification technology?

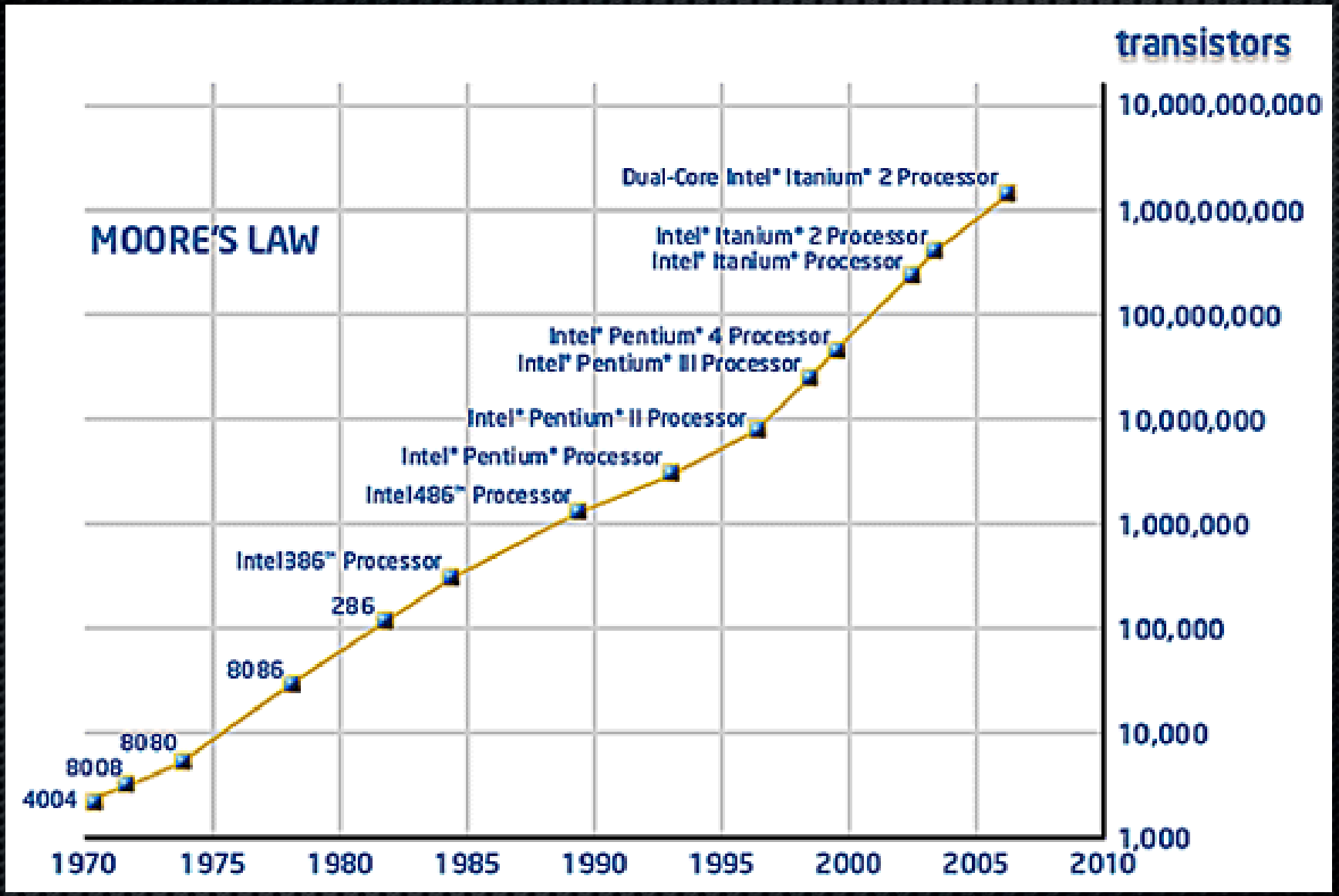


D. Jackson, et al

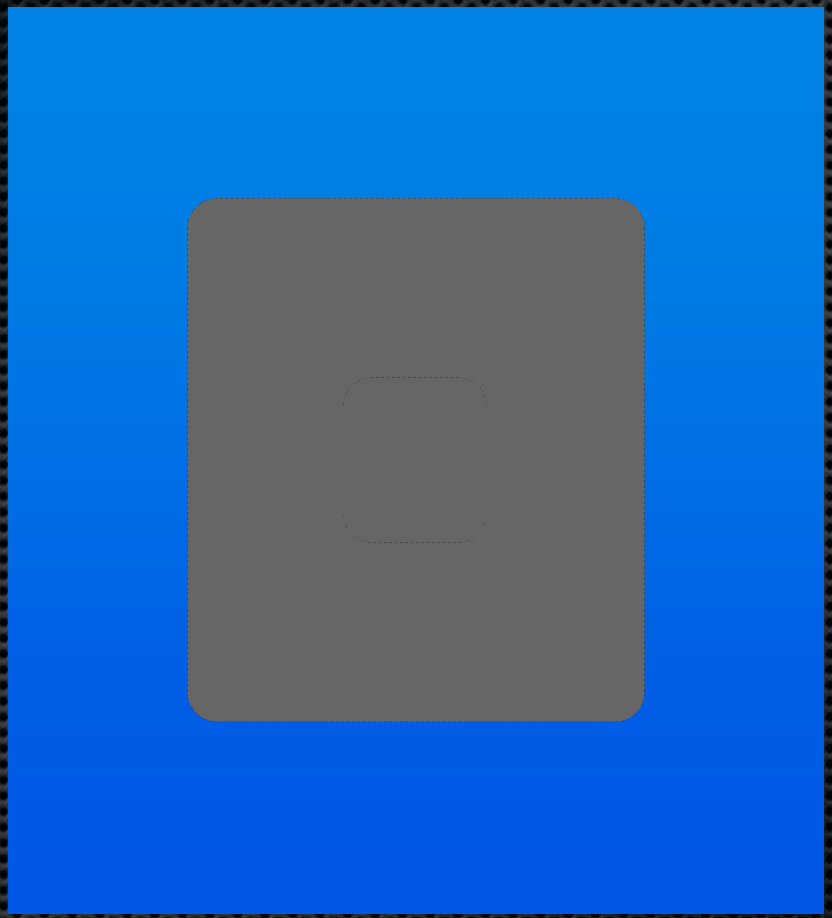
A. Myers, et al

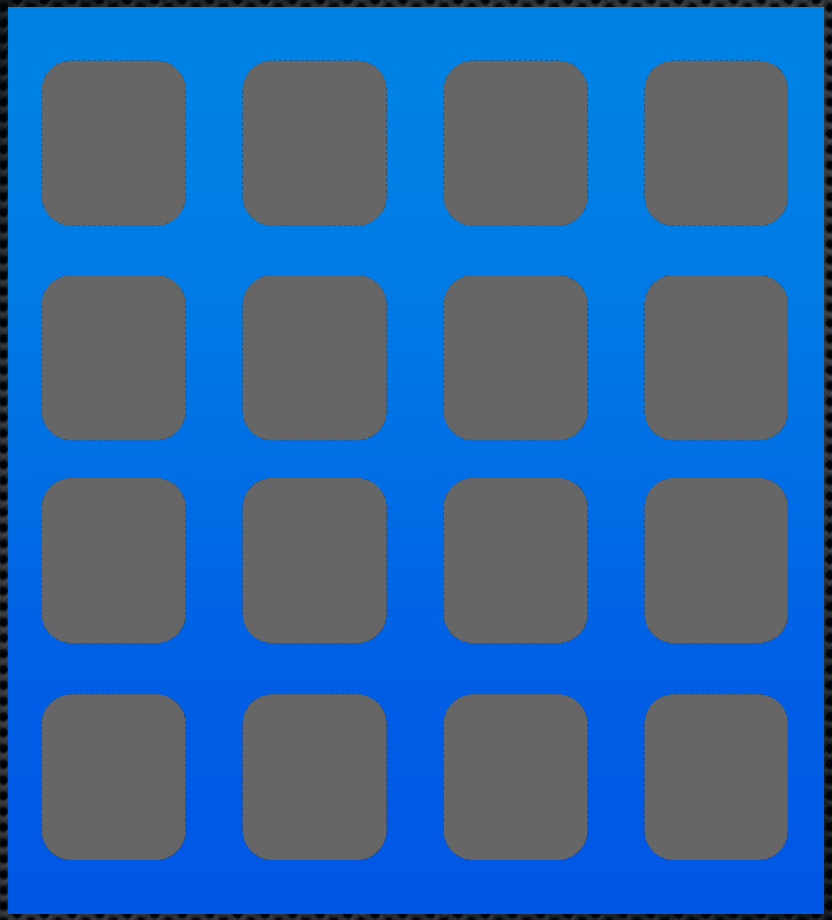
STONESOUP
Safely Taking on Executables Of Unknown Provenance

IARPA



Courtesy, Intel Corp.

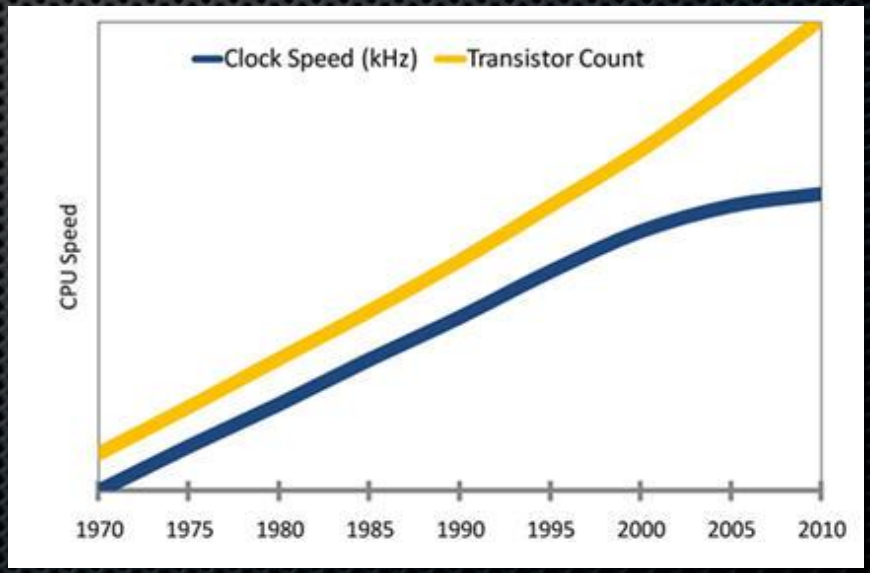




Chip manufacturers are betting that multicore processor technology will keep US industry on the Moore’s Law curve

The multicore challenge

- **Performance is not scaling with transistor count**
- We don’t know how to write software for multicore!
  - one of the holy grails of computer science, particularly urgent today
  - Not just algorithms, but also languages, architeture



A Story: Machine Translation Research

ج ميعوروس ص ينيون مسئولون حث /شينخوا /ينايرر 17بكين
ال قضيةبشان "النفس ضبط وممارسة الهدوءالتزام "علي المعنية الاطراف
الشعبية الديمقراطية كوريا بجمهورية الخاصة النووية.

و زيرونائب تشانغ ون يانغ الصيني الخارجية وزير نائب التقي وقد
الكسندرالروسي الخارجية
من السعي مواصلة الي المعنية الاطراف دعيا حيث غداء مادبة علي لوسيوكوف
السلميالحل اجل
الحالي المعقد الوضع ظل في الحوار خلال من.

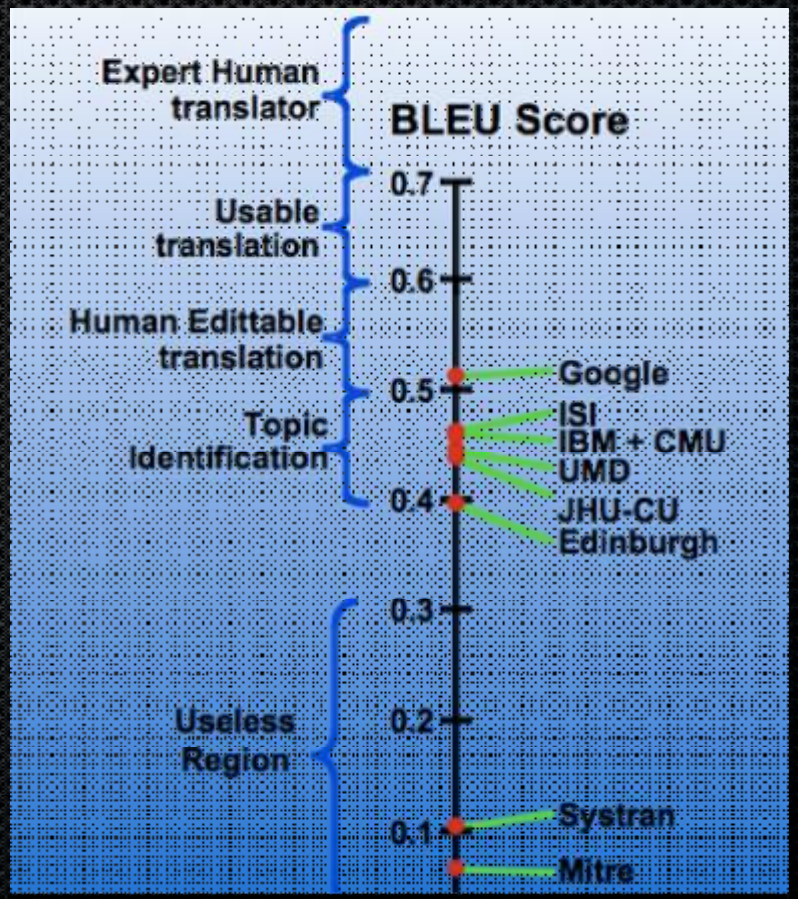
Beijing January 17 / Shinhua / the Chinese and Russian officials urged all parties concerned to " remain calm and exercise restraint " over the nuclear issue of the Democratic People's Republic of Korea.

He met with vice Chinese foreign minister Yang Chang won the deputy of the Russian foreign minister Alexander Losyukov at a lunch with invited interested parties to continue the search for a peaceful solution through dialogue under the current complicated situation.

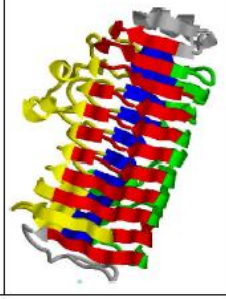
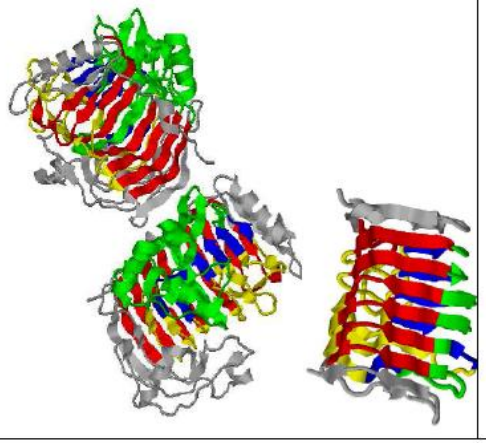
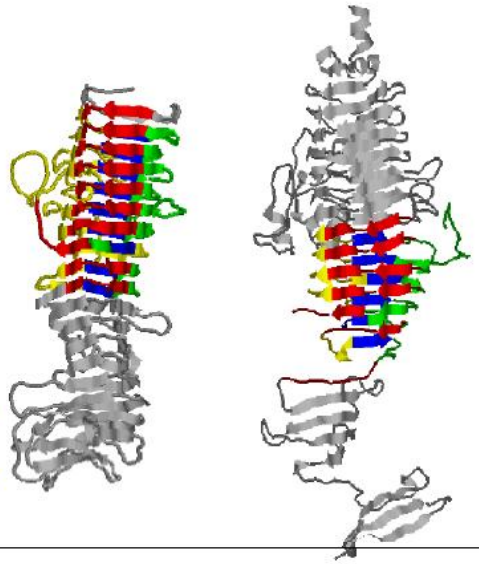
Courtesy of J. Carbonell

2005 NIST MT Challenge

- Google “comes out of nowhere”, employing massive data-intensive computing to win
- Confirms the scalability of a range of algorithms



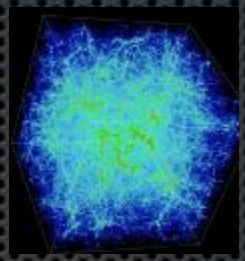
The underlying algorithms have impact beyond machine translation!

Group	Perfect match	Good match	OK match
Missing rungs	0	1-2	3 or more
PDB-ID	**1czf**	1air, 1bhe, 1bn8, 1dbg, **1ee6**(right), 1idj, **1ktw**(left), 1qcx, 1qjv, 1rmg	**1dab**(left), 1ea0, **1tyu**(right)
			

Predicted alignment for known B-helices on cross-family validation, obtained via large-scale, data-intensive computing

Courtesy of J. Carbonell

A New Era in Scientific Discovery?



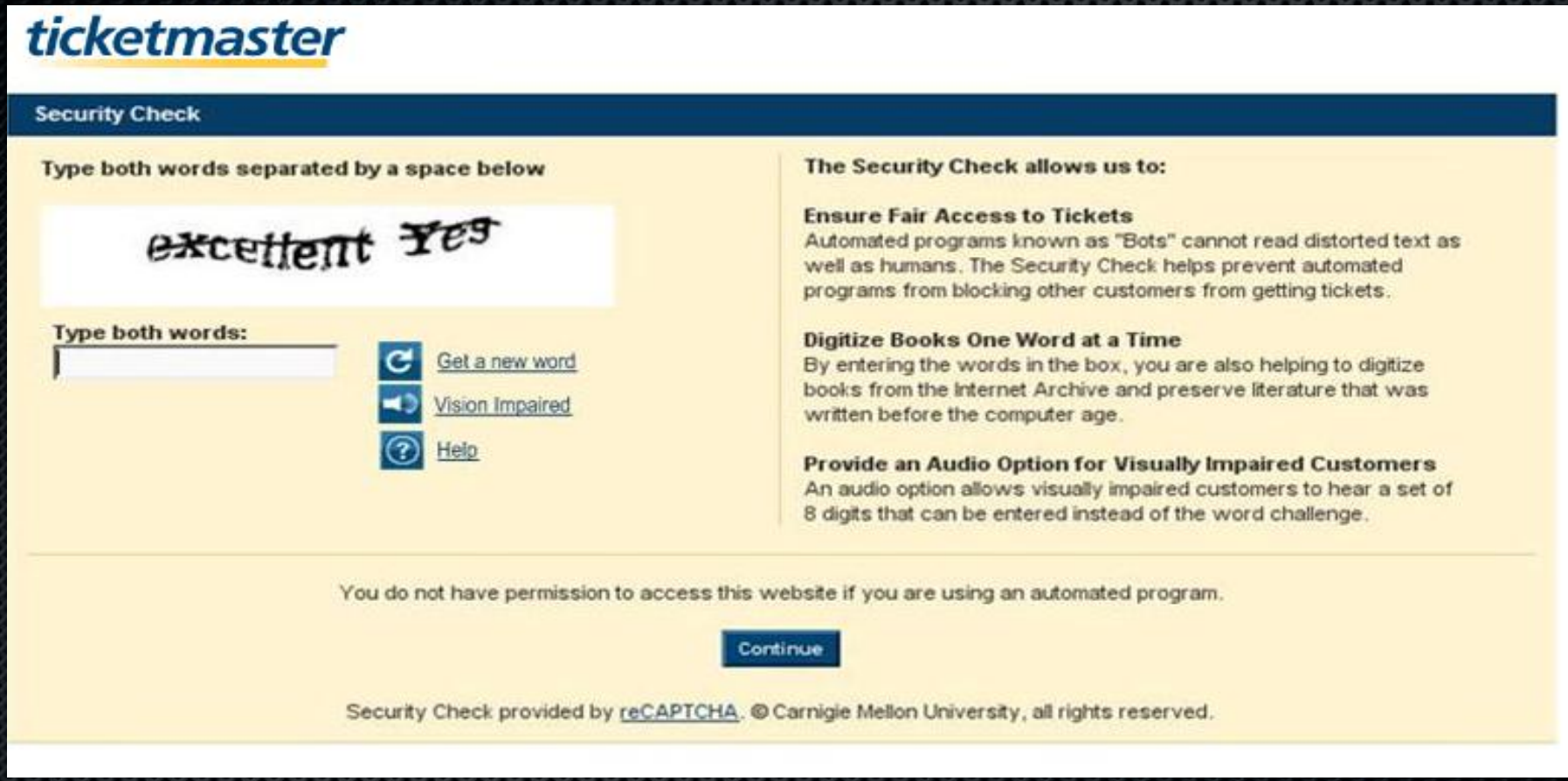
Data-Intensive eScience

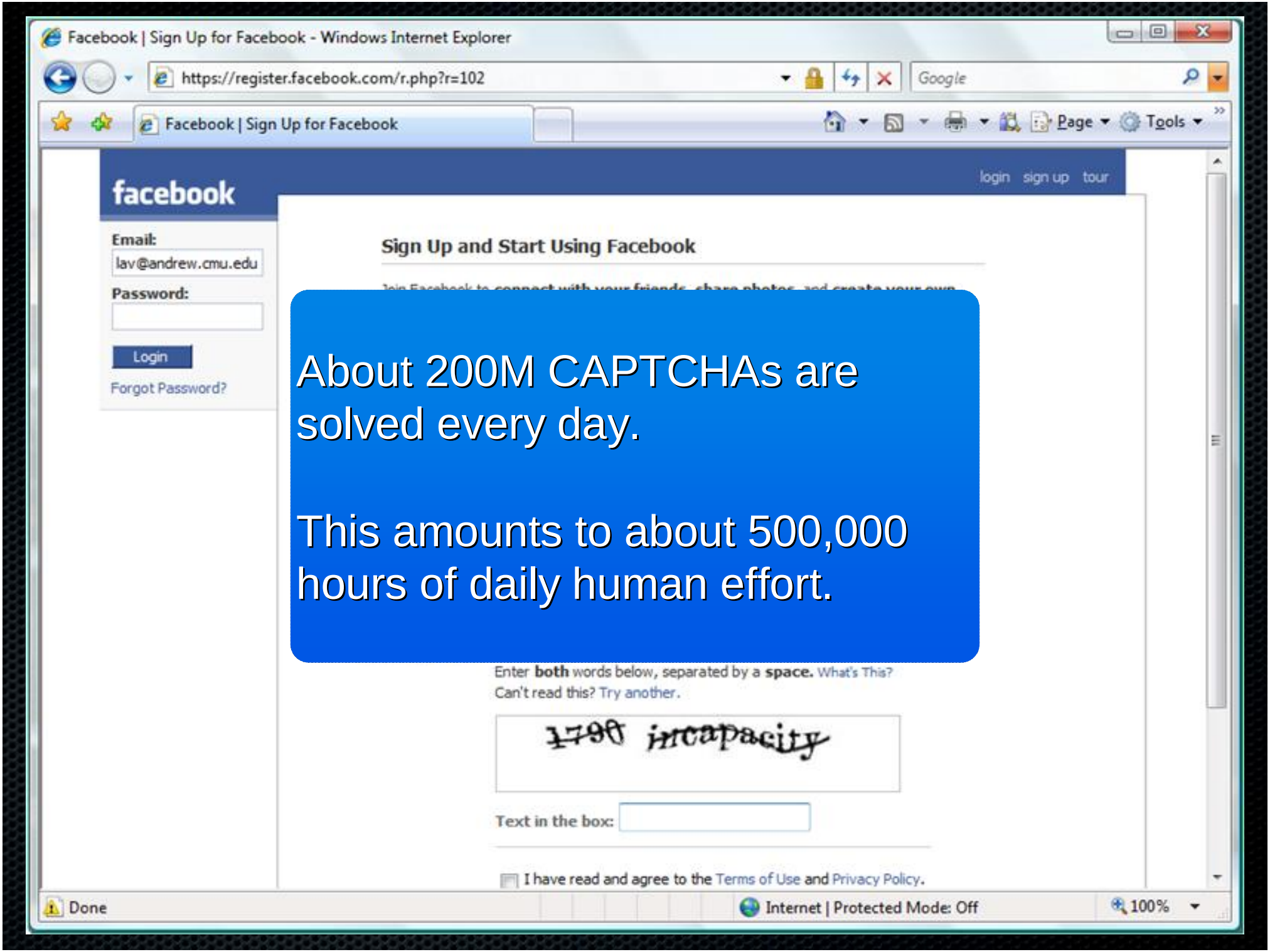
Theory & Experimentation

Modeling & Simulation

The DISC challenge

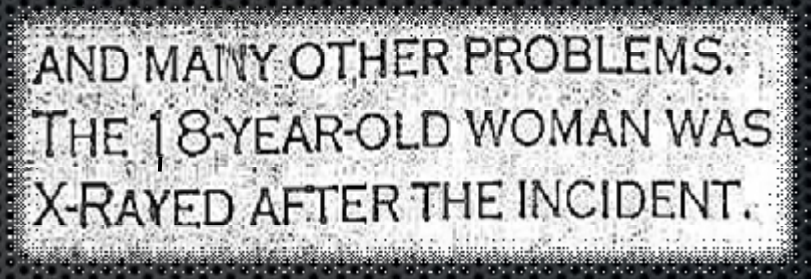
- DISC: Data-Intensive Scalable Computing
- Is there a scalable, reliable, sharable computing architecture for dealing with data-intensive problems?
- Can “cloud computing” provide a basis for eScience?
  - can we make this as accessible and interactive as, say, Google.com?


ticketmaster
Security Check
Type both words separated by a space below
excellent Yes
Type both words:
Get a new word
Vision Impaired
Help
The Security Check allows us to:
Ensure Fair Access to Tickets
Automated programs known as "Bots" cannot read distorted text as well as humans. The Security Check helps prevent automated programs from blocking other customers from getting tickets.
Digitize Books One Word at a Time
By entering the words in the box, you are also helping to digitize books from the Internet Archive and preserve literature that was written before the computer age.
Provide an Audio Option for Visually Impaired Customers
An audio option allows visually impaired customers to hear a set of 8 digits that can be entered instead of the word challenge.
You do not have permission to access this website if you are using an automated program.
Continue
Security Check provided by reCAPTCHA. © Carnigie Mellon University, all rights reserved.


Facebook | Sign Up for Facebook - Windows Internet Explorer
https://register.facebook.com/r.php?r=102
Google
Facebook | Sign Up for Facebook
Page
Tools
facebook
login sign up tour
Email:
lav@andrew.cmu.edu
Password:
Login
Forgot Password?
Sign Up and Start Using Facebook
About 200M CAPTCHAs are solved every day.
This amounts to about 500,000 hours of daily human effort.
Enter both words below, separated by a space. What's This?
Can't read this? Try another.
1790 incapacity
Text in the box:
I have read and agree to the Terms of Use and Privacy Policy.
Done
Internet | Protected Mode: Off
100%

Digitizing Books

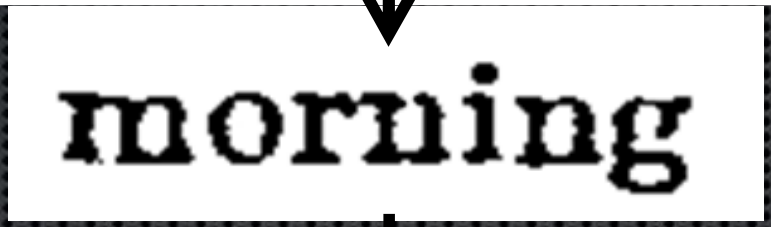
Scan



OCR

AND MARY OTHER PROBLEMS.
THE 1B-YEAR-OLD WOMAN WAS
X-RATED AFTER THE INCIDENT.

The Norwich line steamboat train, from New-London for Boston, this morning ran off the track seven miles north of New-London.



Words Digitized Per Day

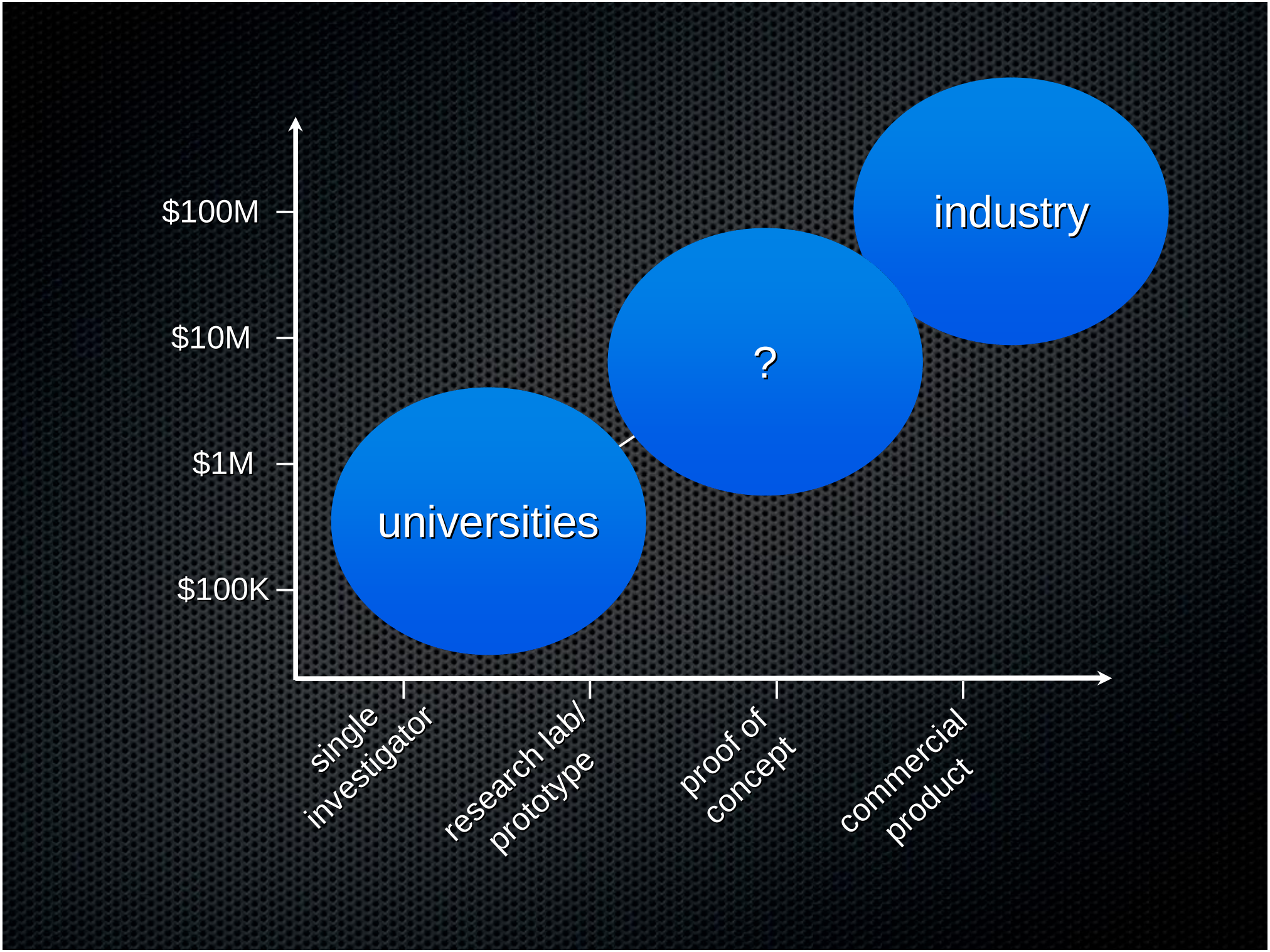
Courtesy of L. von Ahn

The society-scale computing challenge

- For the first time in history, we have the ability to marshall the abilities of hundreds of millions of people
- How can this be used to solve problems of critical national importance?

Summary

- Significant changes are driving IT research into new directions that are relevant to NIST areas
- Some examples
  - certification
  - multicore computing
  - DISC
  - society-scale computing


$100M
$10M
$1M
$100K
universities
?
industry
single investigator
research lab/ prototype
proof of concept
commercial product