Tricks of the Programming Trade

Jon Bentley

Avaya Labs Research

IP Telephony | Contact Centers | Unified Communication | Services

© 2006 Avaya Inc.

AVAYA

Three Kinds of Knowledge

General

1. Science
2. Project Management
3. "Tricks of the Trade"

Medicine

1. Chemistry, Biology
2. Operating teams, Medical records
3. Blood samples

Software

1. Algorithm design, database theory
2. Team organization, project scheduling
3. This talk

About This Talk

- **Goals**
  - For the listeners: a few amusing stories, and a prod to think about tricks
  - For the talker: more stories and tricks
- **Requests**
  - Please do: think, talk
  - Please don't: answer from table-lookup, take notes
    - Google "programming pearls"

Outline

- Introduction
- Problem Definition
- The Back of the Envelope
- Debugging
- Other Tricks
- Teaching the Tricks

Problem Definition

A public opinion polling firm wishes to draw a random sample from a hardcopy list of precincts

The Current Status

- An hour with a table of random numbers

The User's Idea

- Input
  - The user types *n* precinct names (10-character strings) and an integer *m* < *n*
  - Typically, *m* ~ 20 and *n* ~ 200
- Output
  - Random selection of *m* names

A Better Problem

Input

- Integers *m* and *n*

Output

- A sorted list of *m* integers in the range 1..*n*

Example

- If *m* = 3 and *n* = 8, the output might be 2, 3, 5

Code: A dozen lines of Basic

Conceptual blocks

West Point Cadet: Input is a hardcopy list of precincts

Classroom Assignments

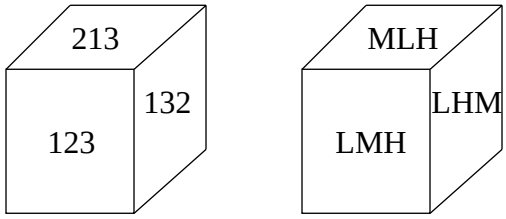
- Read a file of *n* lines and the parameter *m*, then print *m* lines at random
  - Prove that your program produces a random output
- Read the parameters *m* and *n*, then print a sorted list of *m* random integers in the range 1..*n* without duplicates
  - Suppose that *m* and *n* are small; make the program as small as possible
  - Suppose that *m* and *n* are large; make the program as fast as possible
- From a file of unknown size, print *m* lines at random
  - Must they be printed in order?
  - Can one read the file twice?
- Given a vague description, formulate a precise problem
  - Woolsey: Problem solvers vs. problem definers

Another Randomizing Problem

- For *n* = 72 psychological subjects, randomly permute
  - order of experimenters (1, 2, 3)
  - stress conditions (High, Medium, Low)
- Output of the Program

```
1 3L 2M 1H
2 3H 1M 2L
3 1L 2H 3M
4 1M 2L 3H
   ...
```

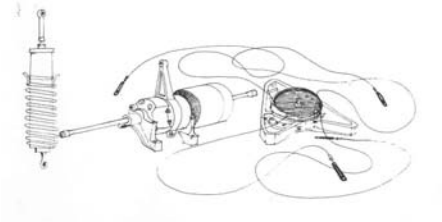
A More Elegant Solution



Why it failed

Problems to Be Defined

Adams: Decelerators for Mariner IV's panels



What is the simplest digital camera?

Dallen's bunkers: 5" logs and 3 sand bags?

How much {security, reliability} to build into a new member of a product family?

How should an individual ascend from a submarine sunk in 100 feet of warm water? (1929: Momsen, 200 feet, DSM; 1944: *USS Tang*)

Summarizing Your Solution

Moore's "Elevator Statement"

From *Crossing the Chasm*, Revised 1999

For <target customer>
who <statement of need or opportunity>
the <product name> is a <product category>
that <key benefit, compelling reason to buy>.
Unlike <primary competitive alternative>
our product <statement of primary differentiation>.

An Example

For computer programmers
who need an informative break from work,
"Tricks of the Programming Trade"
is a technical talk
that is fun and immediately useful.
Unlike specialized talks on recent research,
this talk covers old and fundamental topics
that are often used but rarely discussed.

A Quiz

A conversation about software engineering

How much water flows out of the Mississippi River per day?

One Possible Answer: The river is about a mile wide, twenty feet deep, and flows at five miles per hour

1 mile × 1/250 mile × 120 miles / day

~ 1/2 mile^3 / day

It discharges roughly one-half cubic mile per day

But what does this have to do with software?

Two Answers Are Better Than One

How much water flows out?

As much as flows in

The Mississippi basin is about 1000 by 1000 miles; annual runoff from rain is about one foot. Yearly drainage is

1000 miles × 1000 miles × 1/5000 mile/year

~ 200 miles^3 / year

Daily outflow is

200 miles^3 / year / 400 days / year

~ 1/2 mile^3 / day

A Cheating Triple Check

Long ago: a telephone call to a canoeist colleague

An almanac reports that the river's discharge is 640,000 cubic feet per second, or 0.4 cubic miles per day

A crucial engineering technique

sheer dumb luck

A modern answer

- **Google water flow mississippi river**
  - **"At New Orleans, the average flow rate is 600,000 cubic feet per second."**
- **Google 600000 cubic feet per second in cubic miles per day**
  - **600 000 ((cubic feet) per second) = 0.352178813 (cubic miles) per day**

Tabular Computation

How much water flows in?

$$\frac{1000\ \text{miles} \mid 1000\ \text{miles} \mid 1\ \text{mile}}{5000\ \text{year}}$$

Cancel terms

$$\frac{\cancel{1000\ \text{miles}} \mid \cancel{1000\ \text{miles}} \mid \cancel{1\ \text{mile}} \mid 200\ \text{mile}^3}{\cancel{5000}\ \text{year}}$$

Multiply by 1 year = 400 days (almost)

$$\frac{\cancel{1000\ \text{miles}} \mid \cancel{1000\ \text{miles}} \mid \cancel{1\ \text{mile}} \mid 200\ \text{mile}^3 \mid \text{year}}{\cancel{5000}\ \text{year} \mid 400\ \text{days}}$$

Cancel again

$$\frac{\cancel{1000\ \text{miles}} \mid \cancel{1000\ \text{miles}} \mid \cancel{1\ \text{mile}} \mid \cancel{200}\ \text{mile}^3 \mid \cancel{\text{year}} \mid 1}{\cancel{5000\ \text{year}} \mid \cancel{400}\ \text{days} \mid 2}$$

A Quiz

How much water flows out of the Mississippi River per day?

One Possible Answer: The river is about a kilometer wide, 10m deep, and flows at 8km per hour

$1 \text{ km} \times 1/100 \text{ km} \times 200\text{km / day}$

$\sim 2 \text{ km}^3 \text{ / day}$

It discharges roughly two km^3 per day

Two Answers Are Better Than One

How much water flows out?

As much as flows in

The Mississippi basin is about 1500 by 1500 km; annual runoff from rain is about 1/3 m. Yearly drainage is

$1500 \text{ km} \times 1500 \text{ km} \times 1/3000 \text{ km/year}$

$\sim 750 \text{ km}^3 \text{ / year}$

Daily outflow is

$750 \text{ km}^3 \text{ / year} \;/\; 375 \text{ days / year}$

$\sim 2 \text{ km}^3 \text{ / day}$

A Cheating Triple Check

A telephone call to a canoeist colleague

An almanac reports that the river's discharge is 640,000 cubic feet per second, or 1.6 km^3 per day

A crucial engineering technique:

sheer dumb luck

A modern answer

- **Google water flow mississippi river**
  - **"At New Orleans, the average flow rate is 600,000 cubic feet per second."**
- **Google 600000 cubic feet per second in cubic kilometers per day**
  - **600 000 ((cubic feet) per second) = 1.46794533 (cubic kilometers) per day**

Tabular Computation

How much water flows in?

$$\frac{1500\ \text{km} \mid 1500\ \text{km} \mid 1\ \text{km}}{\mid \mid 3000\ \text{year}}$$

Cancel terms

$$\frac{\cancel{1500\ \text{km}} \mid \cancel{1500\ \text{km}} \mid \cancel{1\ \text{km}} \mid 750\ \text{km}^3}{\mid \mid \cancel{3000}\ \text{year} \mid}$$

Multiply by 1 year = 375 days (almost)

$$\frac{\cancel{1500\ \text{km}} \mid \cancel{1500\ \text{km}} \mid \cancel{1\ \text{km}} \mid 750\ \text{km}^3 \mid \text{year}}{\mid \mid \cancel{3000}\ \text{year} \mid \mid 375\ \text{days}}$$

Cancel again

$$\frac{\cancel{1500\ \text{km}} \mid \cancel{1500\ \text{km}} \mid \cancel{1\ \text{km}} \mid \cancel{750}\ \text{km}^3 \mid \cancel{\text{year}} \mid 2}{\mid \mid \cancel{3000\ \text{year}} \mid \mid \cancel{375}\ \text{days} \mid}$$

Reminders About Quick Calculations

- How To Do Them
  - Two answers are better than one
  - Tabular computations
  - Dimension checks; quick checks
  - Common sense
  - Rules of thumb
  - Safety factors
- When To Do Them
  - After design, before implementation
  - Before any efficiency improvements
  - Cost-benefits analysis

Exercises

- Advertisement: a salesperson drives 100,000 miles a year
  - Taxi driver: I drive 30,000 miles a month
- How much blood does a human heart pump in an hour?
- Newspaper: a quarter has “an average life of 30 years”
- How can you weigh a car using only a dollar bill?
- Dozen web sites: Zeroth Iditarod dogsled went “674 miles in 27.5 hours”
- Catton on McClellan at the Chickahominy: “I wish I knew how deep it is.” A young Lieutenant’s answer?
- When is a cyclist with removable media faster than a high-speed data line?

How much does the snow on my tent weigh?

How Much Does the Snow on my Tent Weigh?

Facts

- The tent is a square, 9 feet on a side
- The snow is 30 inches or 2.5 feet high
- Weight of snow
  - Light snow is about one-tenth as dense as water
  - A cubic foot of water weighs about 62 pounds

Calculation

- A cubic foot of snow weighs about 6 pounds
- A square foot of snow 2.5 feet deep weighs about 15 pounds
- The tent is about 81 square feet
- The snow above me weighs about 1200 pounds

Quality of Estimates

Albert Michelson's estimates of the speed of light

Physics, 1907

1879: 299,910 ± 50 km/s

1926: 299,796 ± 4

1935: 299,774 ± 11

Google "speed of light": 299 792 458 m / s

How good are you at estimating parameters?

Try an "estimation quiz"

netlib.bell-labs.com/cm/cs/pearls/quiz.html

Or, Google estimation quiz

Father's Day

"My father beat [back-of-the-envelope calculations] into me. I come from the coast of Maine, and as a small child I was privy to a conversation between my father and his friend Homer Potter. Homer maintained that two ladies from Connecticut were pulling 200 pounds of lobsters a day. My father said, "Let's see. If you pull a pot every fifteen minutes, and say you get three legal per pot, that's 12 an hour or about 100 per day. I don't believe it!"

"Well it is true!" swore Homer. "You never believe anything!"

Father wouldn't believe it, and that was that. Two weeks later Homer said, "You know those two ladies, Fred? They were only pulling 20 pounds a day."

Gracious to a fault, father grunted, "Now that I believe."

– Roger Pinkham

Rule of 72

If $X \times Y = 72$, then

- Finance: X% interest for Y years doubles the money
- Math: $(1 + X/100)^Y \sim 2$

Examples

- $1.08^9 = 1.9990$ $1.09^8 = 1.9926$
- $1.24^3 = 1.9066$ $1.03^{24} = 2.0328$

Applications

- A virus spreads at 6% per hour
- Moore's Law
  - If computers get twice as good every 18 months, they get one percent better each week.
  - $1.009^{78} \sim 2.011$

Little's Law

The average number of things in the system is the product of the average rate at which they leave the system and the average time each one spends in the system

- Math: $N = \lambda \times T$

Weide's night club

- N = 60 people, T = 3 hours for the average Joe, so $\lambda = N / T$ gives a rate of 20 people/hour. The line is 20 people long, so the wait is an hour.

Denning's wine cellar

- N = 150 cases, λ = 25 cases/year, therefore T = 6 years

In a stable population, what is the birth rate?

Rules of Thumb

- Pi seconds is a nanocentury
- 90-10 (80-20) Rules
  - 10% of X accounts for 90% of Y
    - (population, beer)
    - (code, run time)
    - (code, function)
- The cheapest, fastest and most reliable components of a computer system are those that aren't there.
- Whenever possible, steal code.
- Luis Prieto-Portar: How to estimate the cost of a building?

Rules for Warriors

- How many inches of armor do I need?
- Will I collide with that vessel?
- How much can a warrior carry how far in a day's walk?
- How accurate is that weapon?
- How quickly must it be accurate?
- How to avoid singularities in (theta, phi) gun mounts?

More Exercises

- How much water is in the Mississippi River and its tributaries right now?
- How to perform a non-decomposable task that takes 100 CPU years?
- How many times does your heart beat in a year?

A Great Bug Report

We [Wilks & Becker] found that `qsort` is unbearably slow on "organ-pipe" inputs like "0123443210":

```
main(int argc, char **argv)
{   int n=atoi(argv[1]), i, x[100000];
    for (i = 0; i < n; i++)
        x[i] = i;
    for ( ; i < 2*n ; i++)
        x[i] = 2*n-i-1;
    qsort(x, 2*n, sizeof(int), intcmp);
}
```

(Continued …)

Wilks and Becker, Cont.

Here are the timings on a Pentium:

```
$ time a.out 2000
real    5.85s
$ time a.out 4000
real    21.65s
$ time a.out 8000
real    85.11s
$
```

This is clearly quadratic behavior – each time we double the input size, the run time goes up by a factor of four.

A simple experiment to reveal functional form: quadratic when it should be $\Theta(n \log n)$

Approaches to Debugging

Program Verification

- Don't make any mistakes

Theory of Test Data Selection

- Uncover all mistakes

Advanced Debuggers

- Use them if you have them

The Right Attitude

Debugging Steubenville, Ohio

From Roueché's *Medical Detectives*

36 cases of salmonella from December 1980 through February 1981

A Good Debugger

- "I was prepared for hard work, for plenty of old-fashioned shoe-leather epidemiology."
- "In such outbreaks, there is always a common denominator -- an eating place, a wedding reception,a picnic, a dinner party, a church social. Something in which everybody is involved."
- Dominant age group: 20-29

Debugging Computer Systems

Some Puzzles

- The programmer can't log in standing up
- A banking system "quits" on international data
- The program worked once twice
- The puddle in my office, near where my predecessor had a leak from the ceiling

A Moral

- Debugging is usually unbelieving
- The key is skepticism: the magician

Perspective Shift: Tiny MSE

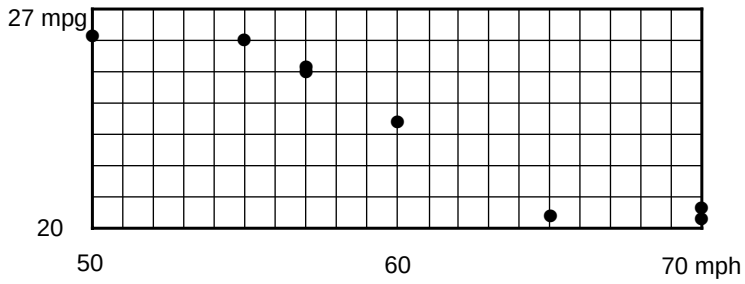
- **Math**
  - Back-of-the-envelope estimates
  - Rule of 72, Little's Law, Optimization
- **Science**
  - Experiments
    - Hypothesis testing: Significance of age 20-29
    - Functional form: Becker and Wilk's Qsort
    - Parameter estimation: Custer at the Chickahominy
- **Engineering**
  - Problem Definition, Heuristics, Debugging

Experiment: MPG as a function of MPH

A 1999 Jeep Grand Cherokee (EPA 21 mpg highway)

MPH	Order	Min								
		1	2	3	4	5	6	7	8	9
50	6	27.2	26.2	25.8	25.4	26.2	26.1	25.7	26.1	26.1
55	1	24.9		25.1	24.8	25.3	26.0	26.1		
57	5	25.4	24.3	25.7	25.6	25.7	25.2	25.2		
57	8	25.4	26.9	27.1	26.4	26.6	25.2	25.0		
60	2	26.0	24.8	24.3	23.3	23.4	23.4	23.4		
65	3	21.4	21.9	20.8	20.7	20.5	20.4	20.4		
70	4		20.6	20.6						
70	7	20.3								

An Optimization Search



Ten Design Hints from *Programming Pearls*

Work on the right problem.
Explore the design space of solutions.
Look at the data.
Use the back of the envelope.
Exploit symmetry.
Design with components.
Build prototypes.
Make tradeoffs when you have to.
Keep it simple.
Strive for elegance.

Teaching the Tricks

- Rarely (never?) as a one-hour lecture
- Woven through an education
  - Freshman
    - Computers for poets (Breadth CS)
    - Intro to IT
    - First course in programming
  - Junior
    - IT for the masses
    - Algorithms and data structures
  - Senior
    - Capstone design course

Summer Reading

- (Per Dave Reed)
- Three Classics
  - Polya, *How To Solve It*
  - Highet, *The Art of Teaching*
  - Strunk and White, *The Elements of Style*
- Yesterday's Find
  - R. E. D. Woolsey, *The Woolsey Papers*

Nurturing Trickiness

- **Observe**
- **Read**
- **Discuss**
- **Practice**
- **Reflect**