Technological Implications for Privacy

David Kotz
Department of Computer Science
Dartmouth College
dfk@cs.dartmouth.edu

© Copyright 2004 by the authors

The Web Eases Access

- It is easier for you to access information
- But, as more of life becomes digital & networked
  - Commerce...
  - Communication...
  - Entertainment...
- It is easier for *them* to track you

What Makes the Web Different?

“The Web is simply another medium of information distribution and gathering; it differs quantitatively because the volumes are so high and the costs so low"

↳ Junkbusters.com, submission to FTC, 1997

Informational Privacy

Privacy ...is the claim of individuals... to determine for themselves when, how, and to what extent information about them is communicated to others...”

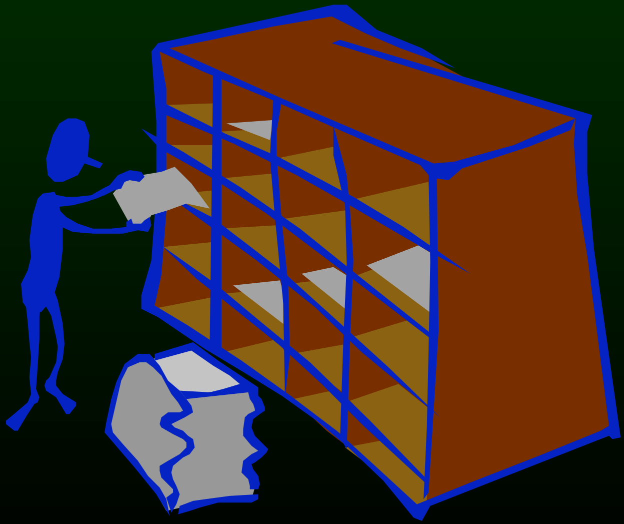
↪ Alan F. Westin

Information Protection on the Web

- People eavesdropping in your communications
- People hacking into personal or proprietary information (in databases, caches, or logs)
- People collecting, exchanging, and using personal information

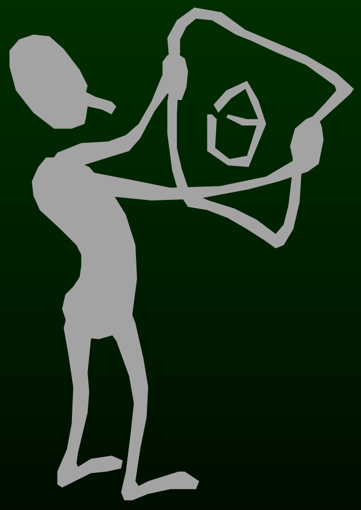
Federal Trade Commission

- FTC survey of 1400 web sites
  - 14% provide privacy notice of some kind
  - 2% list a complete policy
- FTC survey of 212 children's sites
  - 89% collect personal data
  - 54% disclose this fact
  - < 10% provide parental control



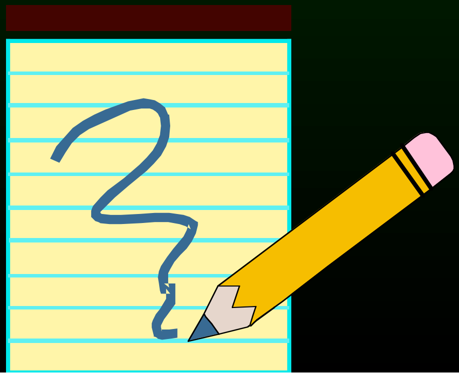
Outline

- Data-collection mechanisms
- Correlating data from multiple sources
- Possible solutions
  - Technical solutions
  - Legal solutions and regulations
- Summary



Data-Collection Mechanisms

- Direct
  - Information you enter into forms on Web pages
- Indirect
  - Monitoring and recording your surfing activity
- Derived
  - From the correlation of multiple direct and indirect

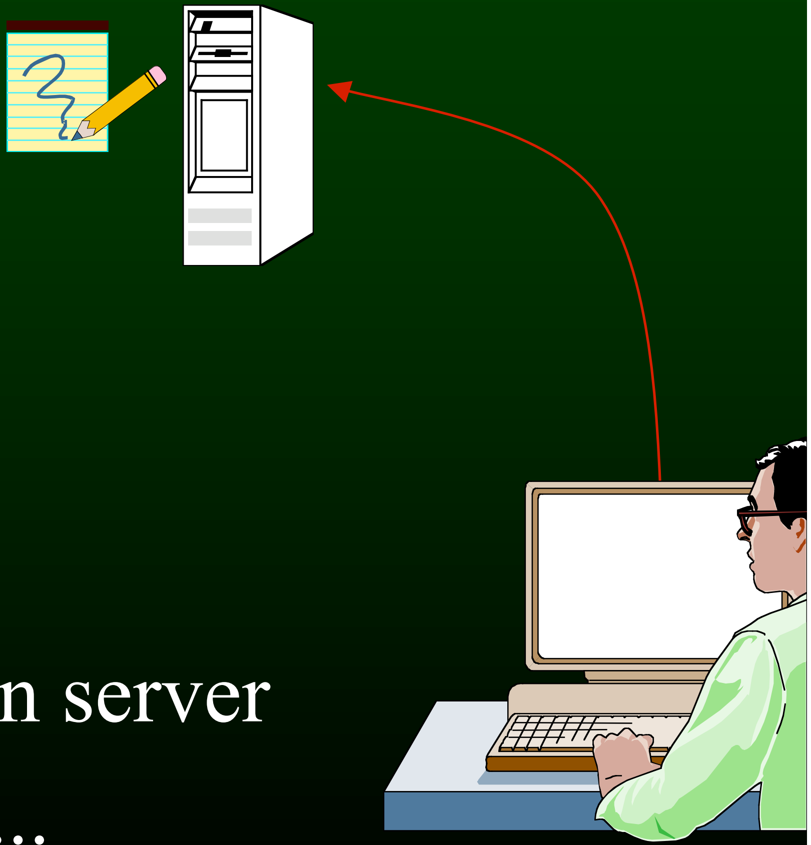


Web forms

- Entry boxes on Web page
- You might enter *identifying* information
  - Name
  - Address
  - Social Security number
  - E-mail address
  - Credit-card number
  - password
  - …

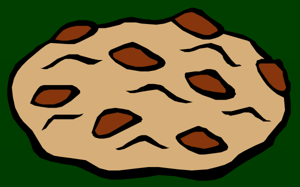
Server Logs

- Every HTTP request from client to server
  - URL for page you want
  - URL for page you're leaving
  - Time
  - IP address of browser
  - Browser brand & version
  - OS brand & version
- All recorded in a "log" file on server
- This web page demonstrates...



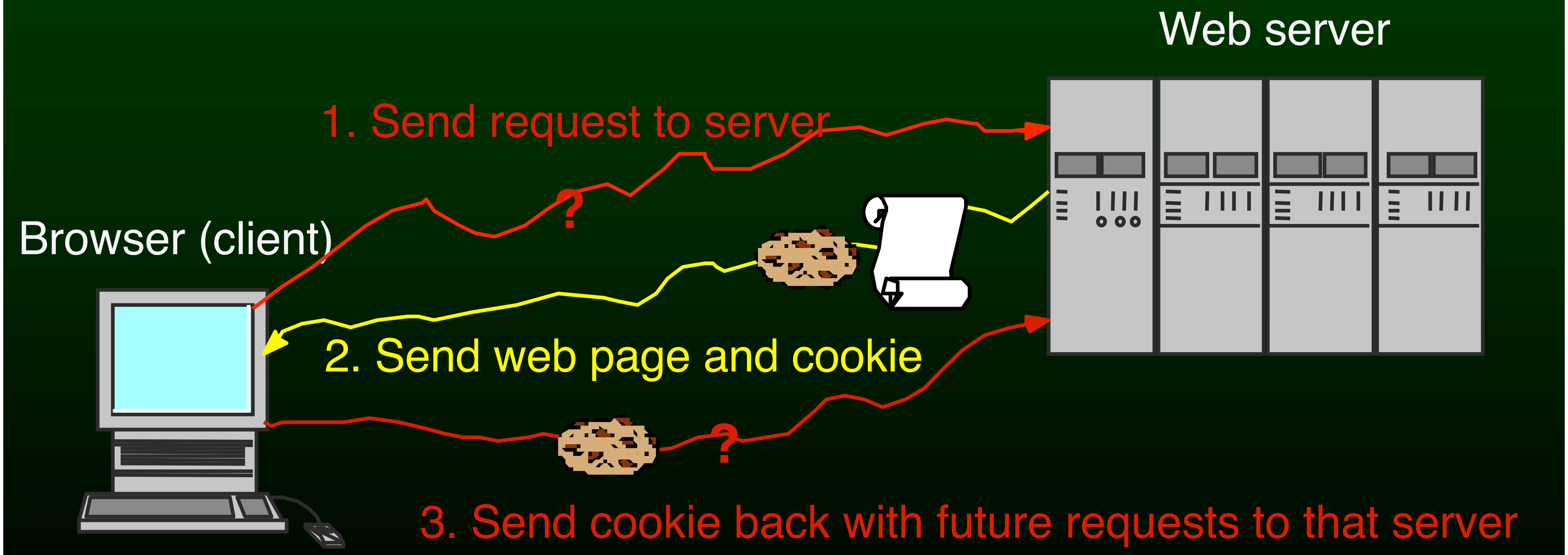
Using Server Logs

- Server’s administrator can analyze the logs
    - Common analysis produces summary stats
- But you can go further…
    - Try to extract picture of each browsing user
        - Sequence and timing of accesses
    - But identified only by IP address



Cookies

- Data, not food
- Data, not software



Web server

1. Send request to server

Browser (client)

2. Send web page and cookie

3. Send cookie back with future requests to that server

Why Cookies?

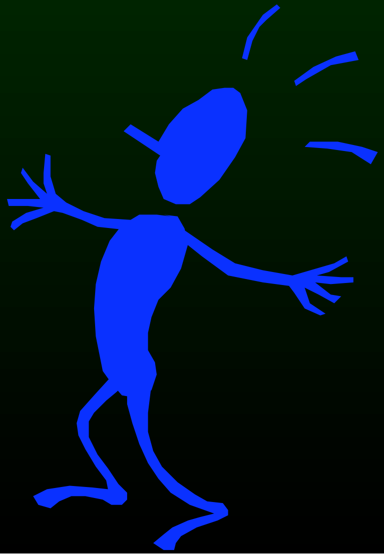
- Web servers are *stateless*
  - They process a sequence of *independent* requests
  - No way to link new request with earlier requests
- Cookies provide state
  - Server sends info it wants to remember, as a cookie
  - Browser stores cookie, possibly for months
  - Browser sends cookie with new requests

The Sweet Side of Cookies

- Cookie state allows interactive applications
  - “Shopping basket”
  - Automatic log-in
  - Customized home pages
  - Personalized “what’s new” listings

The Dark (Burnt?) Side of Cookies

- Cookies identify your browser to server
  - From click to click, or session to session
- Server can thus track *your* movements
  - What pages you fetch, what images you see
  - How long between fetches
  - What advertisements you've seen
  - What preferences you chose



Other sources

- Traditional databases
  - Direct Marketers
  - Credit agencies
  - Motor vehicle records
  - Magazine subscriptions
  - …
- DejaNews
  - Profile of an author (which newsgroups)

THE
DMA
Direct Marketing Association

Correlating Information

- The real danger comes from correlation
  - Merging information collected in one way
  - With information collected in another way
- Biggest concern:
  - Data collected about “anonymous” surfer
  - … is linked with an identified surfer

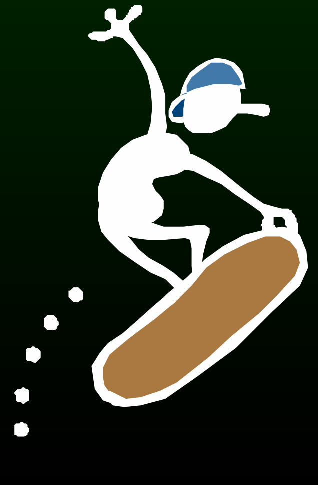
Cookies For Advertising

- Consider on-line advertising agency
  - E.g., doubleclick.net
  - <IMG> tag leads your browser to fetch ad from agency
  - Agency sends cookie along with image
  - Cookies help them to avoid repeating ads

CLICK
HERE
Just $9.99
Per Trade
Free Charts, News and Research
Free Unlimited Real-Time Quotes
TRADE
NOW

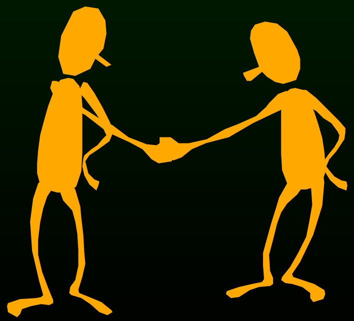
Cookies Across Web Sites

- Many Web sites reference that agency's ads
- Agency sees every page you reference on every server that sells to Agency
  - Recall the "referring URL" is sent with every request
- Agency can correlate surfing patterns across servers



Matching a Name to that Surfer

- ❑ If one web site learns your identity
  - ❑ Sell the information to agency
  - ❑ Agency uses cookie to locate your surfing pattern
- ❑ Agency then
  - ❑ uses (or sells) your name and surfing pattern, or
  - ❑ sells your name to other web servers, in real time



P3P

- Platform for Privacy Preferences
  - Proposed standard
- User specifies privacy preferences
  - Browser records them
- Server specifies privacy policy
  - Sent to browser on request
- Browser matches preferences with policy
  - Ask user whenever policy weaker than preferences

Regulation: Industry

online privacy alliance

- Microsoft
- AOL
- Apple
- Gateway
- Netscape
- Yahoo!
- …

- American Advertising Federation
- Direct Marketing Association
- Disney
- MCI
- Time Warner
- Nationsbank
- …

Regulation: Congress

- *Personal Information Privacy Act of 1997*
  - No traffic in SS#s without consent
- *Childrens' Privacy Act of 1997*
  - Must provide some parental control
- But… *WIPO Treaties Implementation Act*
  - Passed the House on August 4
  - May disallow the removal or blocking of cookies

Regulation: White House

- Al Gore’s July 31 announcement
  - Protect personal info and medical records
  - Stop identity theft
    - Approved by Senate on 7/30
  - Protect children’s privacy on-line
  - Challenge the private sector
    - *Online Privacy Alliance*
    - Warned them that government will step in

Summary

- Privacy is hard to find on the Web
- Technologies developed for one purpose...
  - e.g., cookies
- Can be used for other purposes
  - e.g., tracking users' surfing habits
- Correlation mechanisms
  - You may not be anonymous when you think you are
- Regulation likely

Web Resources

- Collected during this research
  - http://www.cs.dartmouth.edu/~dfk/tangled-web.html