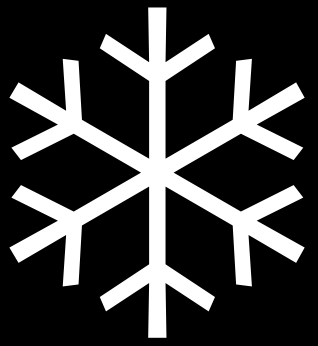
How Unique is Your Browser?

a report on the Panopticlick experiment



Peter Eckersley
Senior Staff Technologist
Electronic Frontier Foundation
pde@eff.org

What is “identifying information”?

Name & address!

But also...

Latanya Sweeney:

ZIP + DOB + gender

identifies almost all US residents

How?

7 billion people on earth

→ typically only ~20,000 per ZIP
→ divide by 365 for birthday
→ divide by ~70 for birth year
→ divide by 2 for gender

(on average works for ZIPs up to 50,000)

Bits of Information

We can measure information in bits:

Each bit of information required doubles the number of possibilities

Each bit of information obtained halves it

For instance

To identify a human, we need
$\log_2$ 7 billion = 33 bits

Learning someone's birthdate
$\log_2$ 365.25 = 8.51 bits

Surprisal and Entropy

Information from a particular value for a variable gives us *surprisal* or *self-information*:

Birthdate = 1st of March: 8.51 bits
Birthdate = 29th of February: 10.51 bits

The weighted average for that variable is the *entropy* of the variable

Surprisal of an event

$$I = - \log_2 \Pr(\text{event})$$

Entropy

$$H = \sum_{\text{events}} \Pr(\text{event}) \; . \; I$$

Adding surprisals

If variables are *independent*, surprisals add linearly (birthdate + gender are independent)

Starsign and birthdate are the opposite

Use joint distributions / conditional probability to model this

Now for an application...

Browser Tracking

“Track” → associate the browser's activities:

- at different times
- with different websites

What ways exist to track browsers?

Cookies

IP addresses

Supercookies

And Fingerprints

Browser has some combination of characteristics which, like DOB + ZIP + gender, are enough to distinguish it from all others

Fingerprint Privacy threats

Globally unique?

Fingerprint + IP → unique?

Occasional cookie undelete?

Auto linked cookie?

Fingerprinting rumours

“Analytics companies are using this method”

“DRM systems are using this method”

“Financial systems are using this method”

How good is it?

(Also: how bad is the logging of User Agent strings?)

Let's do an experiment to find out!

https://panopticlick.eff.org

Fingerprint information we collected

User Agent strings
Other browser headers
Cookie blocking?
Timezone (js)
Screen size (js)
Browser plugins + versions (js)
Supercookie blocking? (js)
System fonts (flash/java)

(Things Panopticlick didn't collect)

Quartz crystal clock skew
TCP/IP characteristics
Screen DPI
HTTP header ordering
Most ActiveX / Silverlight stuff
JavaScript quirks
CSS history
CSS font list (flippingtypical.com !)
More supercookies
lots more!

Data quality control

Use 3-month cookies and encrypted IP addresses

Can correct double counting if people return / reload

(Except: interleaved cookies)

(NOTE: the live data only uses the cookies!)

Dataset

Slightly over a million different browser-instances have visited Panopticlick.eff.org

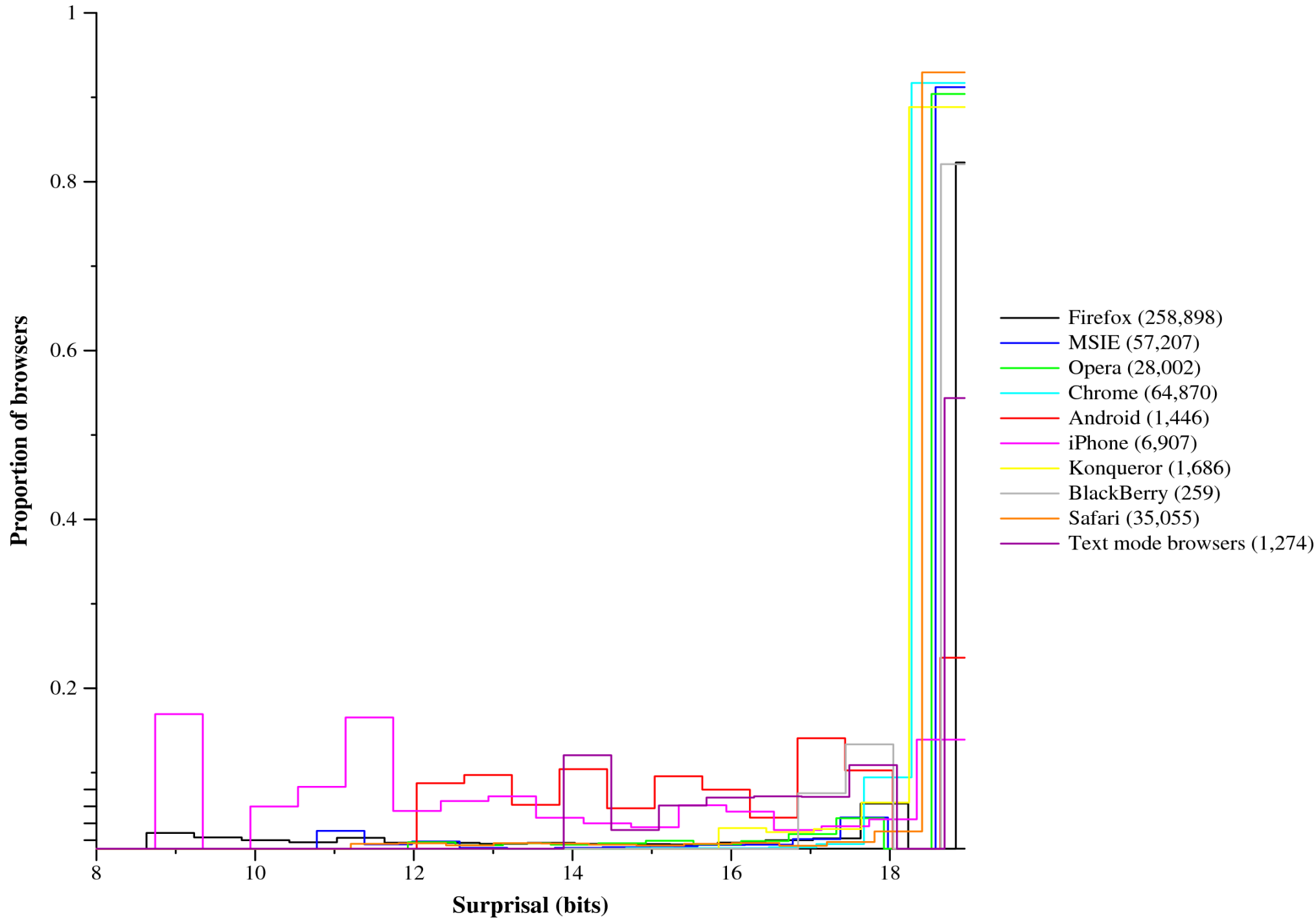
Privacy conscious users:
→ not representative of the wider Web userbase
→ the relevant population for some privacy questions

(analysed the first 500,000 or so)

83.6% had completely unique fingerprints
(entropy: 18.1 bits, or more)

94.2% of “typical desktop browsers” were unique
(entropy: 18.8 bits, or more)

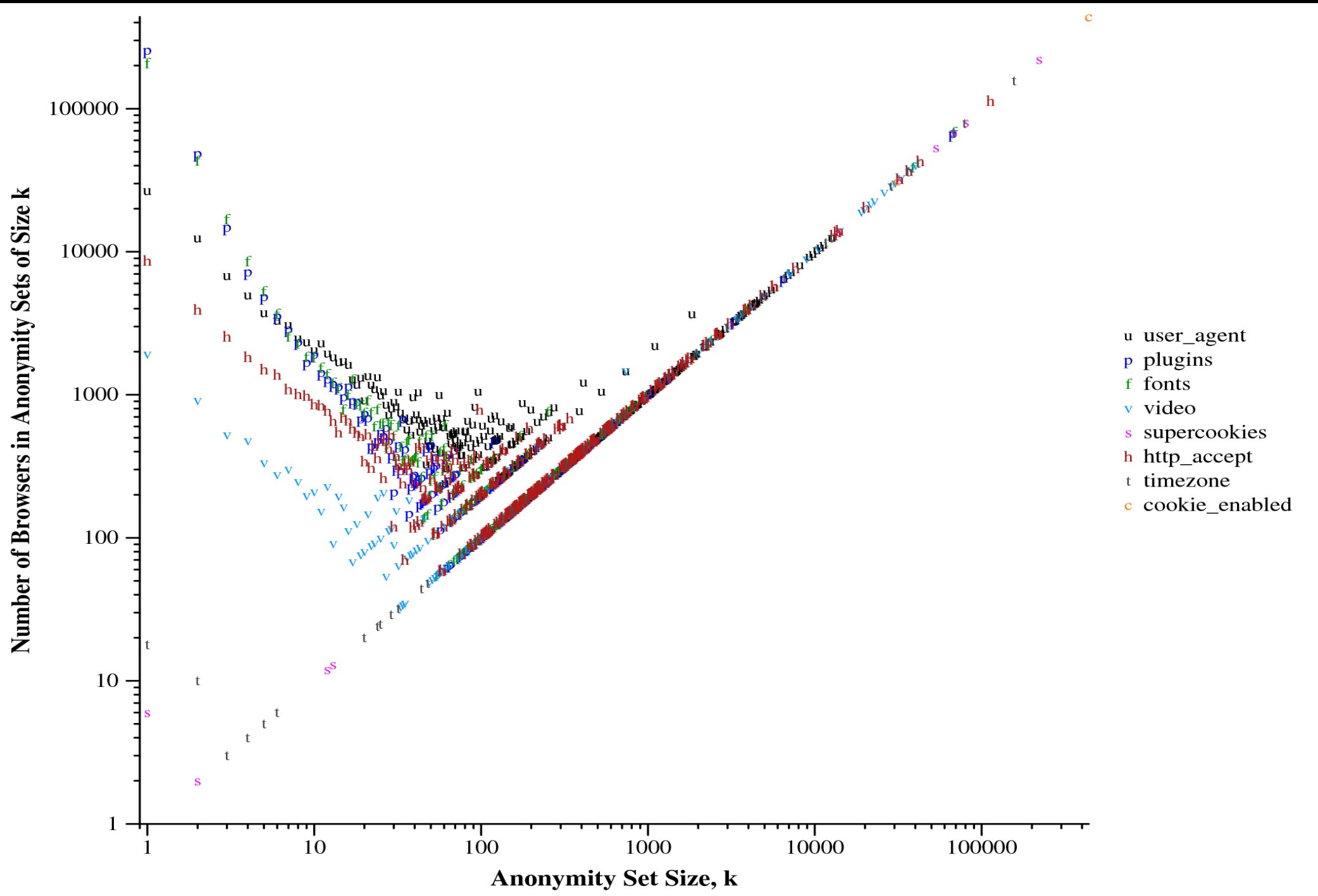
Which browsers did best?



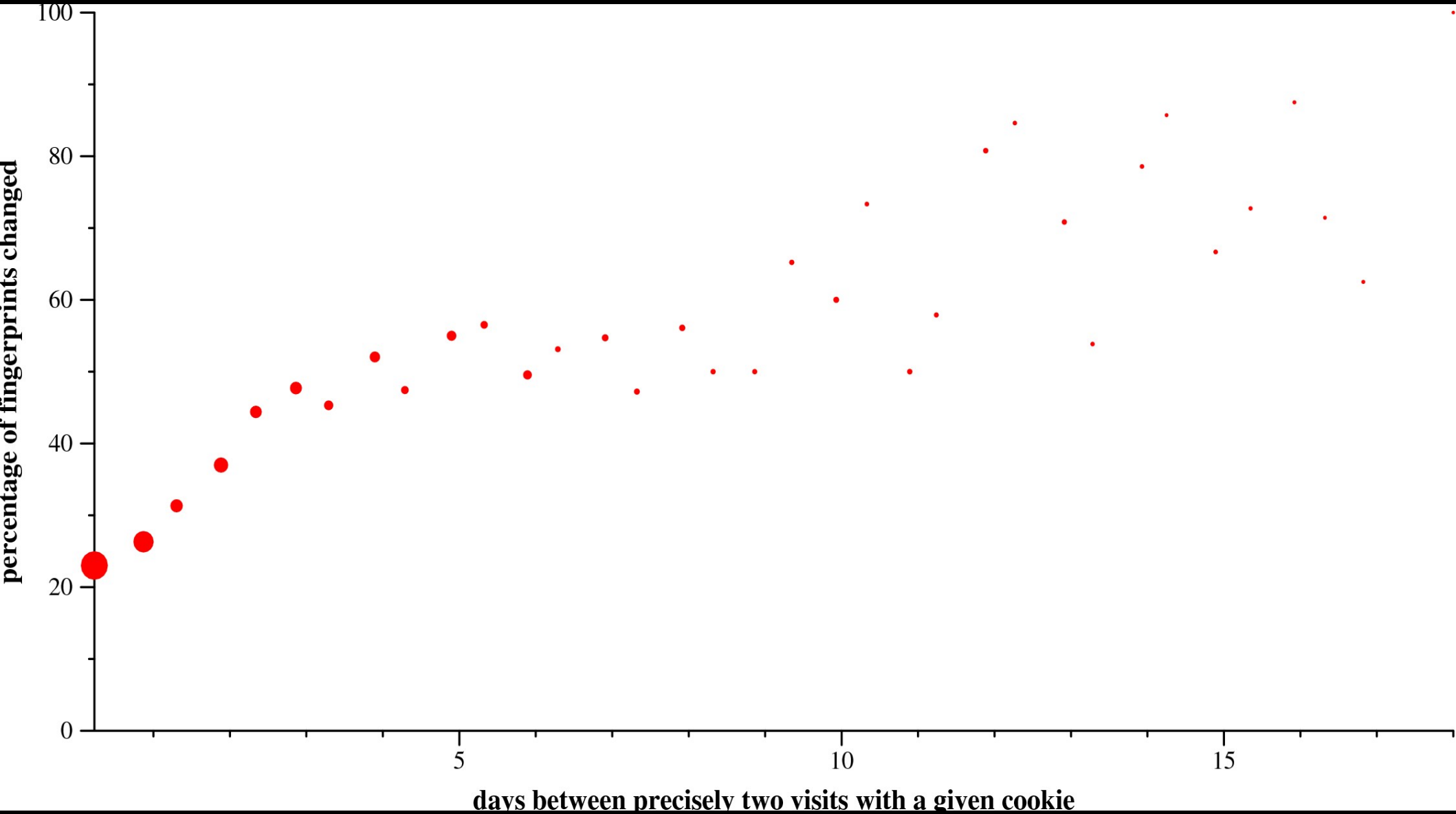
Which variables mattered?

Variable	Entropy
User Agent	10.0 bits
Other headers	6.09 bits
Cookies enabled?	0.353 bits
Timezone	3.04 bits
Screen size	4.83 bits
Plugins	15.4 bits
Supercookies	2.12 bits
Fonts	13.9 bits

Or in more detail...



Are fingerprints constant?



Rate of change of fingerprints

Very high!

Looks like good protection

(but it isn't)

Fuzzy Fingerprint Matching

- Test for Flash/Java

- If yes, and only only one of the 8 components has changed [much], we match

Guessed 66% of the time

99.1 % correct; 0.9% false-positive

so…

Which browsers did well?

Those without JavaScript

Those with Torbutton enabled

iPhones and Androids [*]

Cloned systems behind firewalls

Paradox: some “privacy enhancing” technologies are fingerprintable

- Flash blockers
- Some forged User Agents
- “Privoxy” or “Browzar” in your User Agent!

Noteworthy exceptions:

- NoScript
- TorButton

Test vs. Enumerate

Plugins and fonts → long lists of facts about a computer are very identifying!

Possible solution: testing rather than enumeration

("Does this browser have the FRANKENSTEIN font installed?")

Other solution: browsers do not supply this stuff to websites at all...

Fingerprintability vs Debuggability

Do we need all this for a browser?

Mozilla/5.0 (X11; U; Linux i686; en-AU; rv:1.9.1.9) Gecko/20100502 Seamonkey/2.0.4

All this for each plugin?

Shockwave Flash 10.1 r53

How much of a problem is this?

Many fingerprints are globally unique

Defensive measures

Power users:
- Block JavaScript with NoScript
- Use Torbutton (possibly without Tor)

Everyone else needs to wait for the browsers to fix it

Some of the browsers have started!