Hybrid information flow monitoring against Web Tracking

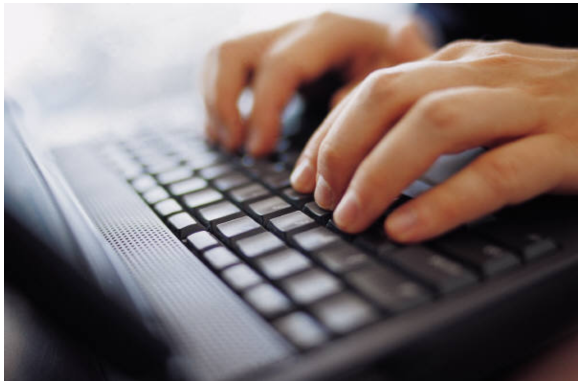
Nataliia Bielova (Inria INDES)

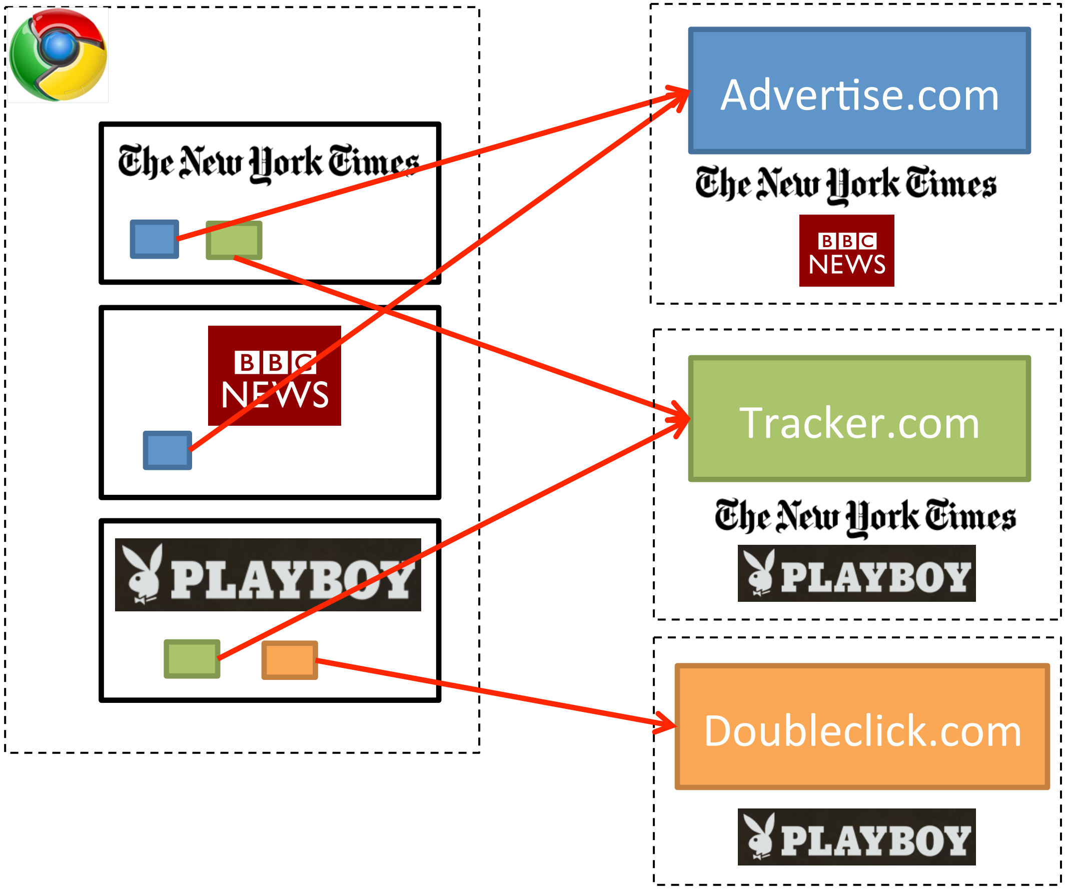
with Frederic Besson and Thomas Jensen (Inria CELTIQUE)

Security and Privacy Workshop – LabEx

18 December 2013

Web Tracking





Bigger browsing profiles
= increased value for trackers
= reduced privacy for users

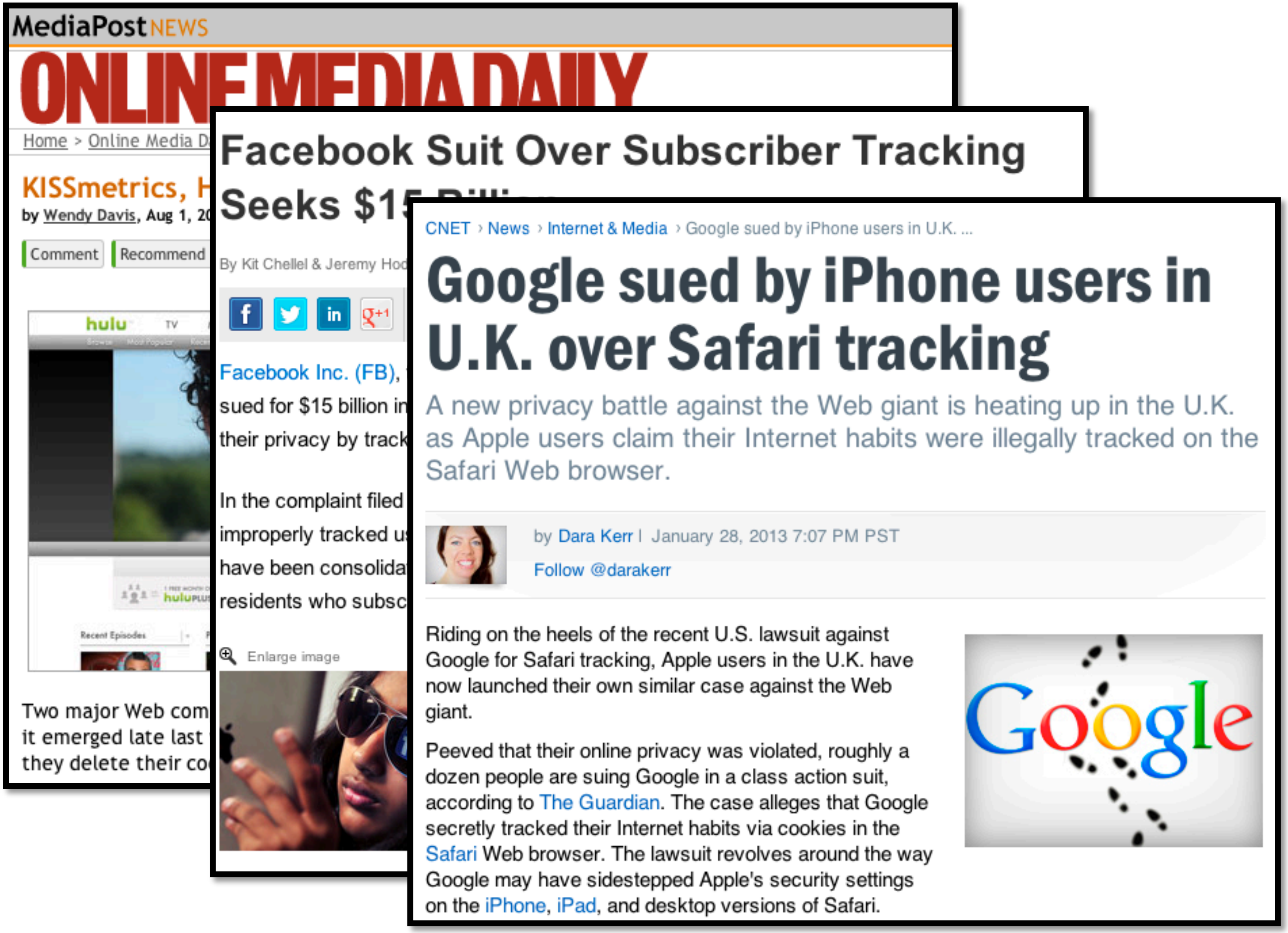
(Hypothetical tracking relationships only.)

Courtesy of Franziska Roesner

Doesn't cookie blocking already solve it?

- Blocking cookies prevents tracking
  - **only by browser-initiated HTTP requests**

- It doesn't protect from tracking
  - by using scripts
  - by other storage mechanisms
  - by browser fingerprinting



MediaPost NEWS

ONLINE MEDIA DAILY

Home > Online Media D

KISSmetrics, H

by Wendy Davis, Aug 1, 20

Comment Recommend

Two major Web com
it emerged late last
they delete their co

Facebook Suit Over Subscriber Tracking
Seeks $15

By Kit Chellel & Jeremy Hod

Facebook Inc. (FB),
sued for $15 billion in
their privacy by track

In the complaint filed
improperly tracked u
have been consolida
residents who subsc

Enlarge image

CNET › News › Internet & Media › Google sued by iPhone users in U.K. ...

Google sued by iPhone users in U.K. over Safari tracking

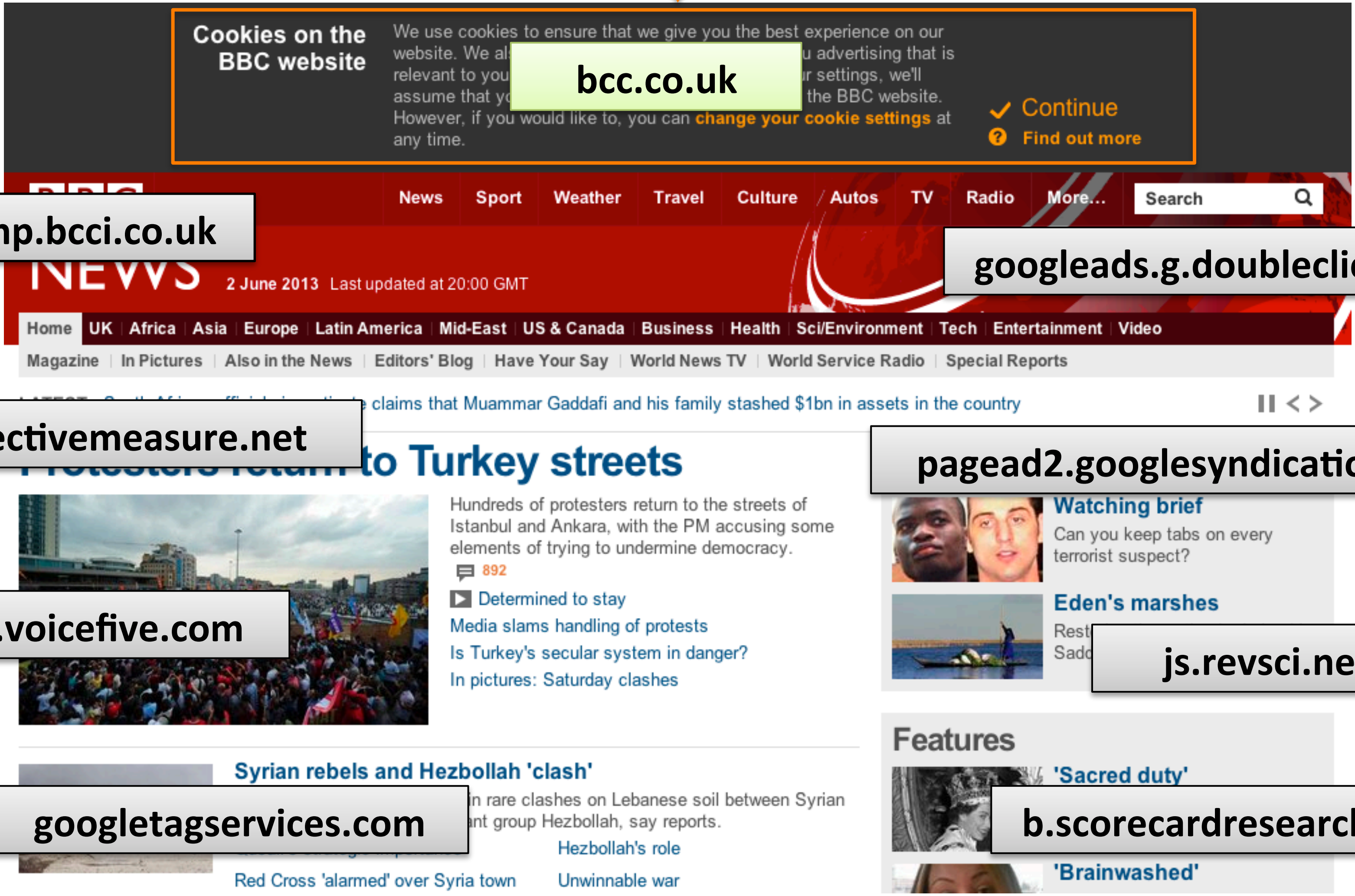
A new privacy battle against the Web giant is heating up in the U.K. as Apple users claim their Internet habits were illegally tracked on the Safari Web browser.

by Dara Kerr | January 28, 2013 7:07 PM PST
Follow @darakerr

Riding on the heels of the recent U.S. lawsuit against Google for Safari tracking, Apple users in the U.K. have now launched their own similar case against the Web giant.

Peeved that their online privacy was violated, roughly a dozen people are suing Google in a class action suit, according to The Guardian. The case alleges that Google secretly tracked their Internet habits via cookies in the Safari Web browser. The lawsuit revolves around the way Google may have sidestepped Apple's security settings on the iPhone, iPad, and desktop versions of Safari.

Thanks to EU ePrivacy Directive



Cookies on the BBC website

We use cookies to ensure that we give you the best experience on our website. We al u advertising that is relevant to you r settings, we'll assume that y the BBC website. However, if you would like to, you can **change your cookie settings** at any time.

✓ Continue

Find out more

News
Sport
Weather
Travel
Culture
Autos
TV
Radio
More...
Search

NEWS

2 June 2013 Last updated at 20:00 GMT

Home | UK | Africa | Asia | Europe | Latin America | Mid-East | US & Canada | Business | Health | Sci/Environment | Tech | Entertainment | Video
Magazine | In Pictures | Also in the News | Editors' Blog | Have Your Say | World News TV | World Service Radio | Special Reports

claims that Muammar Gaddafi and his family stashed $1bn in assets in the country

to Turkey streets

Hundreds of protesters return to the streets of Istanbul and Ankara, with the PM accusing some elements of trying to undermine democracy.

892

Determined to stay

Media slams handling of protests

Is Turkey's secular system in danger?

In pictures: Saturday clashes

Watching brief

Can you keep tabs on every terrorist suspect?

Eden's marshes

Rest

Sadd

Syrian rebels and Hezbollah 'clash'

in rare clashes on Lebanese soil between Syrian

ant group Hezbollah, say reports.

Hezbollah's role

Red Cross 'alarmed' over Syria town

Unwinnable war

Features

'Sacred duty'

'Brainwashed'

bcc.co.uk

emp.bcci.co.uk

googleads.g.doubleclick.net

effectivemeasure.net

pagead2.googlesyndication.com

b.voicefive.com

js.revsci.net

googletagservices.com

b.scorecardresearch.com

Don’t browser extensions solve it?

AdBlockPlus: blocks scripts/requests **only from known advertisement companies**

Ghostery: blocks scripts/requests **only from known tracking companies**

- They don’t protect from tracking by other companies
- They don’t protect form tracking by the main (first-party) website

Tracking is complicated

- Much discussion on tracking, but limited knowledge about concrete technologies

- In this talk:
  - How tracking works
    - Cookies and browser fingerprinting
  - Address gaps with new analysis
    - Quantitative information flow

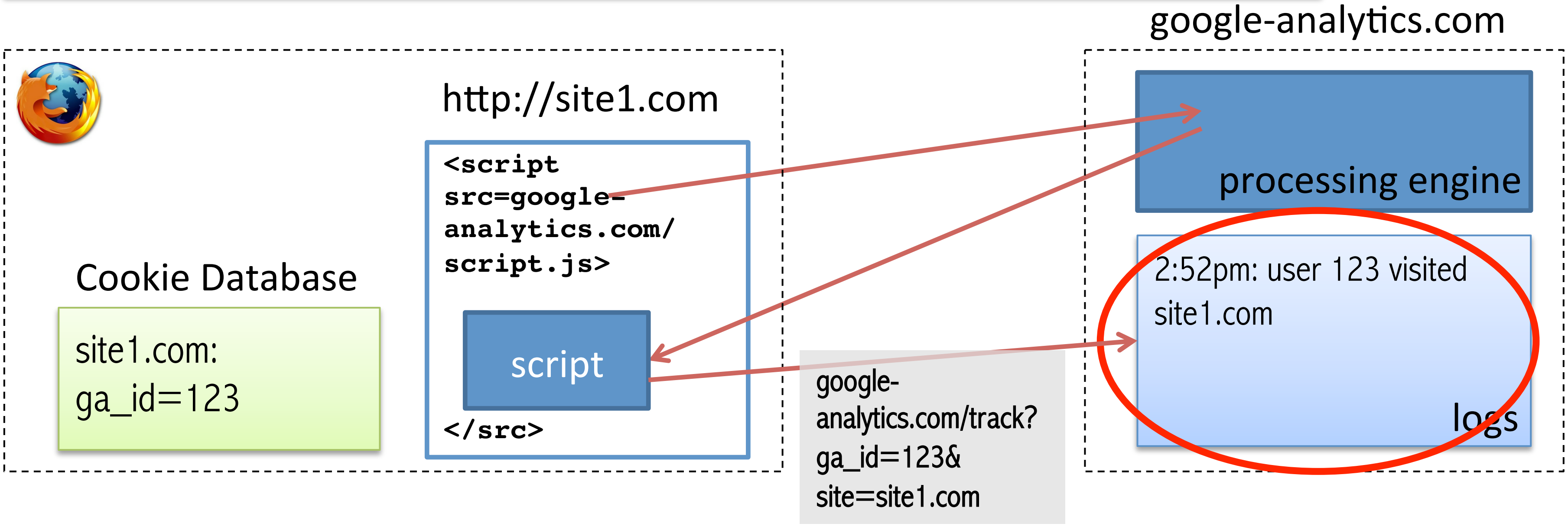
Mechanisms Required By Trackers

- Ability to store/create user identity in the browser
  - Store: cookies + other browser storages
  - Create: fingerprinting browser and OS properties

- Ability to communicate user identity back to tracker
  - Browser: cookies + other HTTP headers
  - JavaScript: embed information in URLs ⬅

Tracking by storing identity

```
// google-analytics.com/script.js
var url = “http://google-analytics.com/track?ga_id= “
     + encodeURI(document.cookie)
     + “&site= “ + encodeURI(document.location);
new Image().src = url;
```



google-analytics.com

http://site1.com

Cookie Database

site1.com:
ga_id=123

```
<script
src=google-
analytics.com/
script.js>
```

script

```
</src>
```

google-
analytics.com/track?
ga_id=123&
site=site1.com

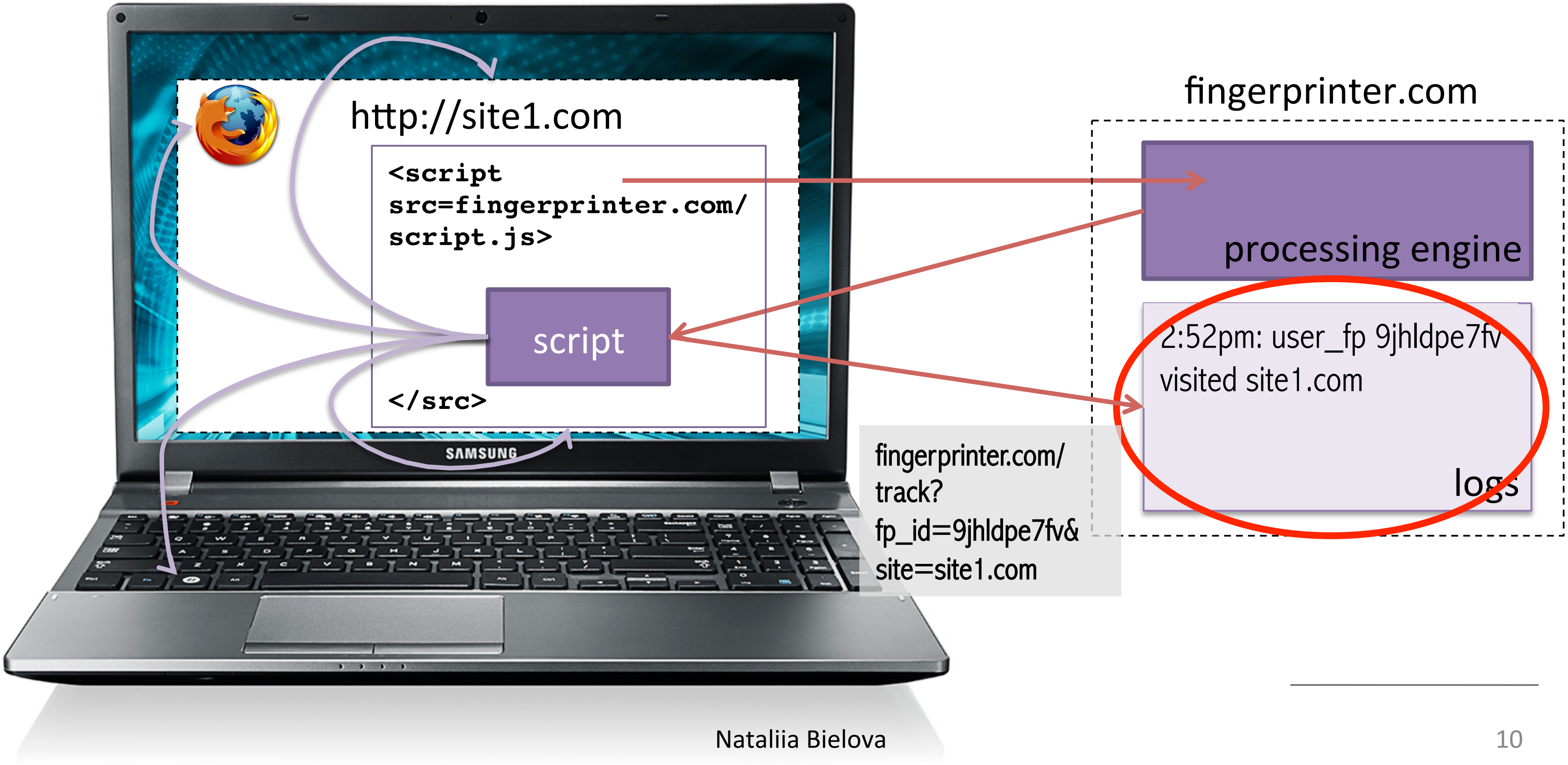
processing engine

2:52pm: user 123 visited
site1.com

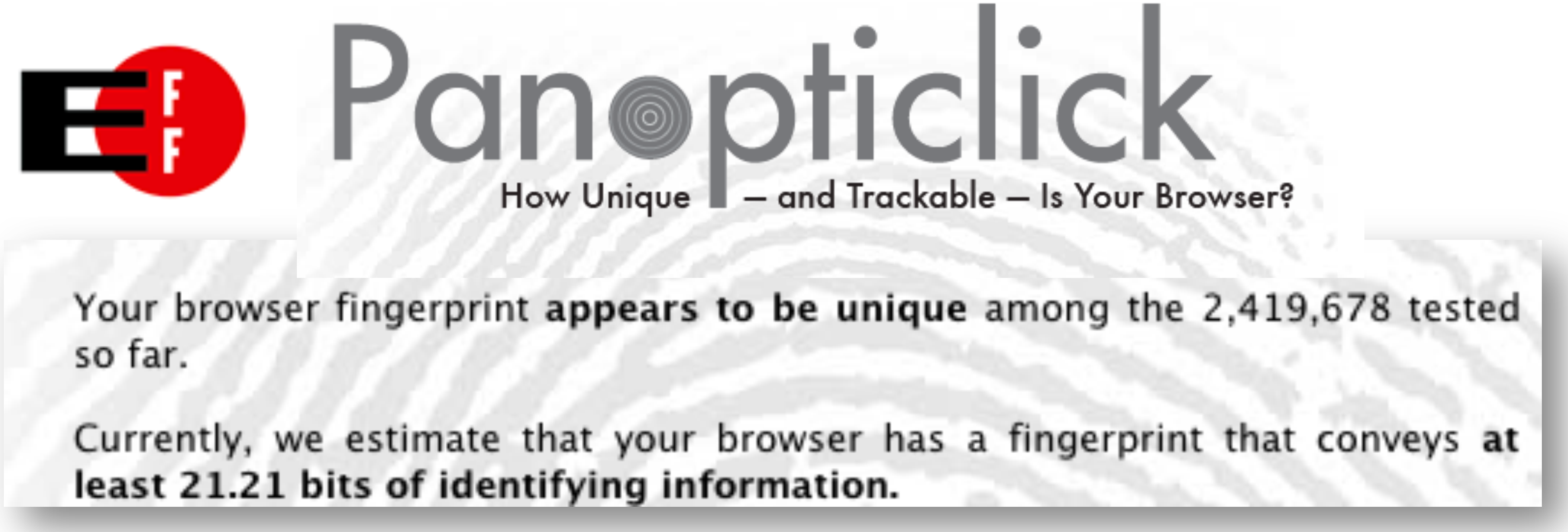
logs

Tracking by creating identity

Browser and operating system properties are used to track repeated visits to a site.



Tracking by creating identity



Your browser fingerprint **appears to be unique** among the 2,419,678 tested so far.

Currently, we estimate that your browser has a fingerprint that conveys **at least 21.21 bits of identifying information.**

83.6% of browser fingerprints are unique among all observed (500 000 browsers) [Eckersley, PETS’2010]

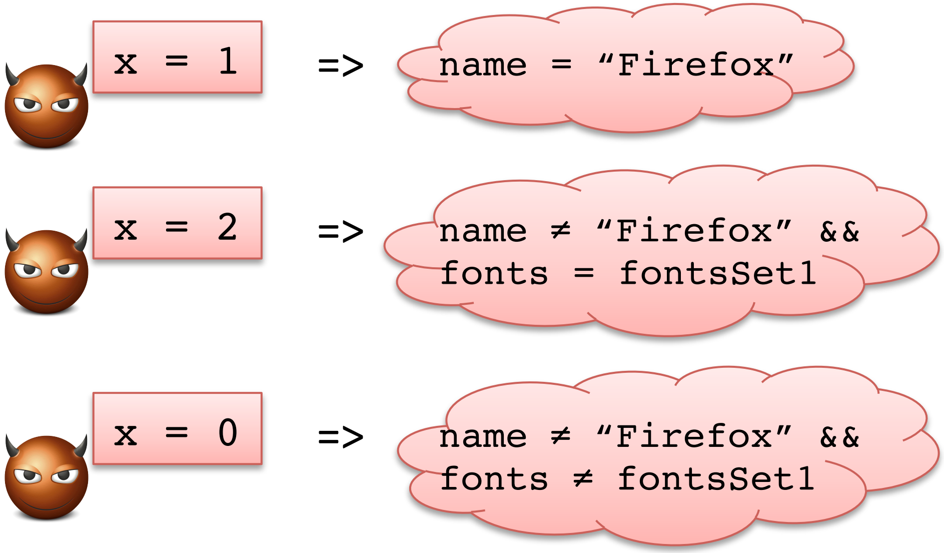
Which browser properties create a fingerprint?

Browser property	Source
Browser name and version, Operating system name and version	HTTP JavaScript
File types accepted, language used	HTTP
Plugins installed in the browser	JavaScript
Time zone	JavaScript
Screen size and color depth	JavaScript
Fonts installed	Flash
Some of browser preferences	HTTP JavaScript
Support for new technologies	JavaScript

Give the most identifying Information [Eckersley'2010]

What does tracker learn?

```
var x = 0;
if (name == “Firefox”) {
     x = 1;
}
else {
     if (fonts == fontsSet1) {
          x = 2;
     }
}
output x;
```



x = 1 => name = “Firefox”

x = 2 => name ≠ “Firefox” && fonts = fontsSet1

x = 0 => name ≠ “Firefox” && fonts ≠ fontsSet1

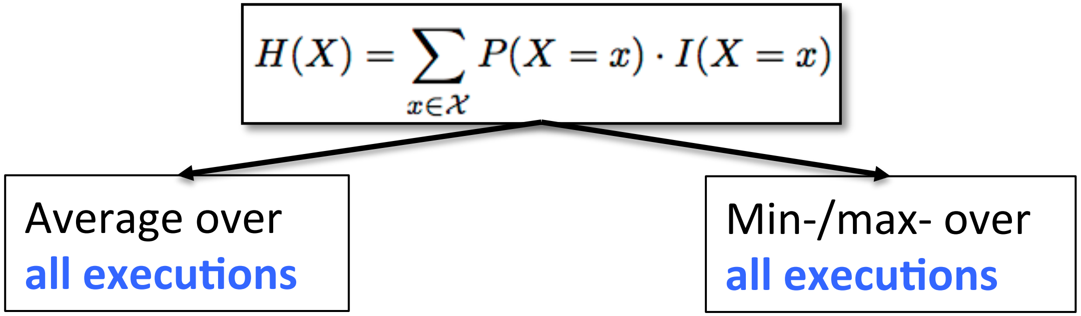
Depending on user’s browser, **different executions** of the same script **leak different quantity** of information!

Challenge:

How to **automatically** evaluate **how much information** a tracker **learns through one execution** of the script?

Static analysis for Quantitative Information Flow

- **Traditionally**, static analysis compute expected leakage
  - using **Information Entropy**



$$H(X) = \sum_{x \in \mathcal{X}} P(X = x) \cdot I(X = x)$$

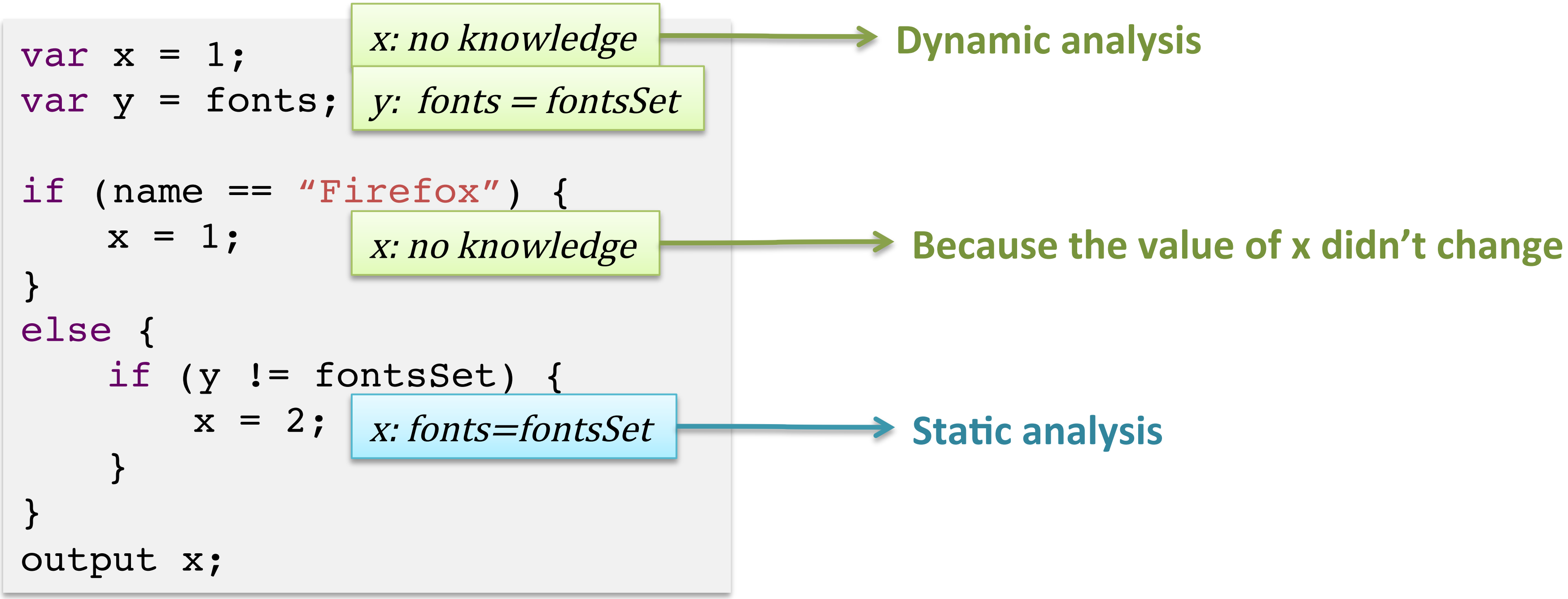
Average over **all executions**

Min-/max- over **all executions**

- **In reality, we only have one execution of a script!**
  - in **one execution** → tracker uniquely **identifies the user**
  - in another execution → tracker just learns FireFox is used

[Besson, Bielova, Jensen CSF’2013]

Hybrid monitoring



```
var x = 1;
var y = fonts;

if (name == “Firefox”) {
    x = 1;
}
else {
    if (y != fontsSet) {
        x = 2;
    }
}
output x;
```

x: no knowledge → Dynamic analysis

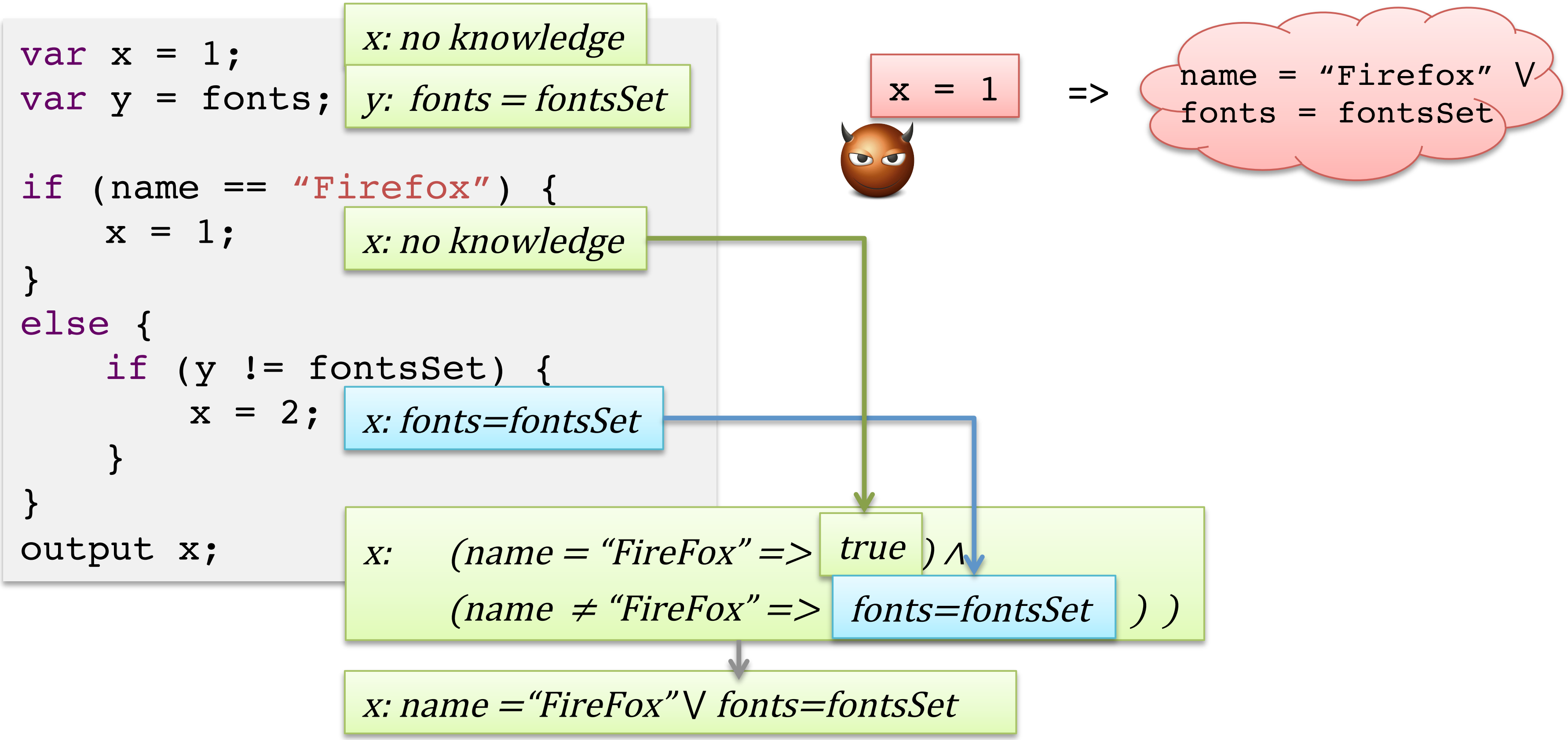
y: fonts = fontsSet

x: no knowledge → Because the value of x didn’t change

x: fonts=fontsSet → Static analysis

[Besson, Bielova, Jensen CSF'2013]

Hybrid monitoring



```
var x = 1;
var y = fonts;

if (name == "Firefox") {
    x = 1;
}
else {
    if (y != fontsSet) {
        x = 2;
    }
}
output x;
```

x: no knowledge

y: fonts = fontsSet

x: no knowledge

x: fonts=fontsSet

x: (name = "FireFox" => true) ∧
(name ≠ "FireFox" => fonts=fontsSet))

x: name ="FireFox" ∨ fonts=fontsSet

```
x = 1
```

=>

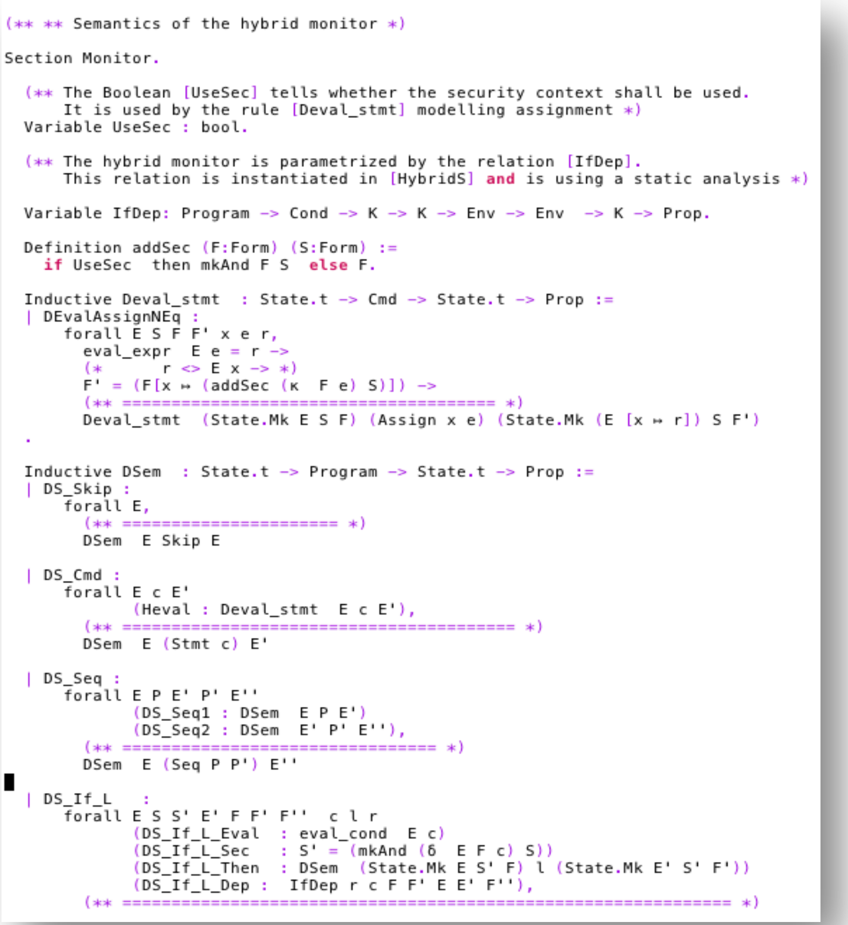
```
name = "Firefox" ∨
fonts = fontsSet
```

Hybrid monitor **precisely models the knowledge** of the tracker!

[Besson, Bielova, Jensen CSF'2013]

Hybrid monitor for quantitative information flow

- Monitoring **one execution**
  - Dynamic + static
- Automatic **quantification of information leakage**:
  - Symbolic representation of tracker's knowledge at runtime
- Strong **formal guarantees**
  - Provably correct approximation of actual tracker's knowledge



```
(** ** Semantics of the hybrid monitor *)

Section Monitor.

  (** The Boolean [UseSec] tells whether the security context shall be used.
      It is used by the rule [Deval_stmt] modelling assignment *)
  Variable UseSec : bool.

  (** The hybrid monitor is parametrized by the relation [IfDep].
      This relation is instantiated in [HybridS] and is using a static analysis *)

  Variable IfDep: Program -> Cond -> K -> K -> Env -> Env  -> K -> Prop.

  Definition addSec (F:Form) (S:Form) :=
    if UseSec  then mkAnd F S  else F.

  Inductive Deval_stmt  : State.t -> Cmd -> State.t -> Prop :=
  | DEvalAssignNEq :
      forall E S F F' x e r,
        eval_expr  E e = r ->
        (*       r <> E x -> *)
        F' = (F[x ↦ (addSec (κ  F e) S)]) ->
        (** ======================================= *)
        Deval_stmt  (State.Mk E S F) (Assign x e) (State.Mk (E [x ↦ r]) S F')
  .

  Inductive DSem  : State.t -> Program -> State.t -> Prop :=
  | DS_Skip :
      forall E,
        (** ====================== *)
        DSem  E Skip E

  | DS_Cmd :
      forall E c E'
             (Heval : Deval_stmt  E c E'),
        (** ========================================= *)
        DSem  E (Stmt c) E'

  | DS_Seq :
      forall E P E' P' E''
             (DS_Seq1 : DSem  E P E')
             (DS_Seq2 : DSem  E' P' E''),
        (** ================================= *)
        DSem  E (Seq P P') E''

  | DS_If_L   :
      forall E S S' E' F F' F''  c l r
             (DS_If_L_Eval  : eval_cond  E c)
             (DS_If_L_Sec   : S' = (mkAnd (б  E F c) S))
             (DS_If_L_Then  : DSem  (State.Mk E S' F) l (State.Mk E' S' F'))
             (DS_If_L_Dep :  IfDep r c F F' E E' F''),
        (** ================================================================ *)
```

All the theorems are proven in Coq: http://www.irisa.fr/celtique/ext/QIF/

Towards guaranteed protection from Web Tracking (ongoing)

- Our hybrid monitor [Besson, Bielova, Jensen CSF'2013] evaluates how much tracker learns

Challenge:

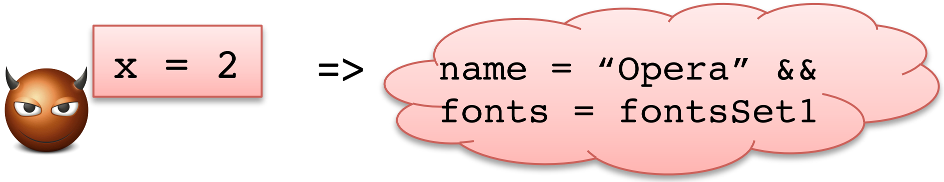
Which mechanism can **provably guarantee** that **every user is protected** from being tracked?

Towards guaranteed protection from Web Tracking (ongoing)

```
var x = 0;
if (name == “Opera”) {
     x = 1;
     if (fonts == fontsSet1) {
          x = 2;
     }
}
output x;
```

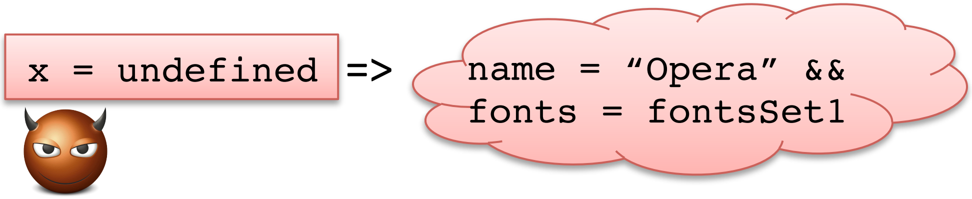
Program instrumentation

```
var x = 0;
if (name == “Opera”) {
     x = 1;
     if (fonts == fontsSet1) {
          x = undefined;
     }
}
output x;
```



Opera browser (very rare) + fontsSet1

=> the user is easily identifiable



Modifying/halting one program execution does not improve user’s protection!

Towards guaranteed protection from Web Tracking (ongoing)

```
var x = 0;
if (name == “Opera”) {
     x = 1;
     if (fonts == fontsSet1) {
```

x = 2 => name = “Opera” &&

<span>Our idea</span>:

Several users (i.e. several executions) **have to be made undistinguishable** for the tracker!

instrumentation

```
var x = 0;
if (name == “Opera”) {
     x = 1;
     if (fonts == fontsSet1) {
          x = undefined;
     }
}
output x;
```

x = undefined => name = “Opera” && fonts = fontsSet1

Modifying/halting one program execution does not improve user’s protection!

Summary

- **Web tracking** is done by different technologies(see Inria ConfLunch*)
  - cookies, other browser storages, fingerprinting
- **Hybrid information flow monitoring** [Besson, Bielova, Jensen CSF'2013]
  - monitors one execution
  - provably correctly approximates tracker's knowledge
- **Towards guaranteed protection** against Web tracking (ongoing)
  - Program instrumentation
  - Systematic lying about browser properties provably improves privacy
- Analyzing **stability of browser fingerprints** (ongoing)
  - https://stopfingerprinting.inria.fr

*http://videos.rennes.inria.fr/confLunch/NataliiaBielova/indexConfLunchBielova.html