



# Fingerprinting the Fingerprinters: Learning to Detect Browser Fingerprinting Behaviors

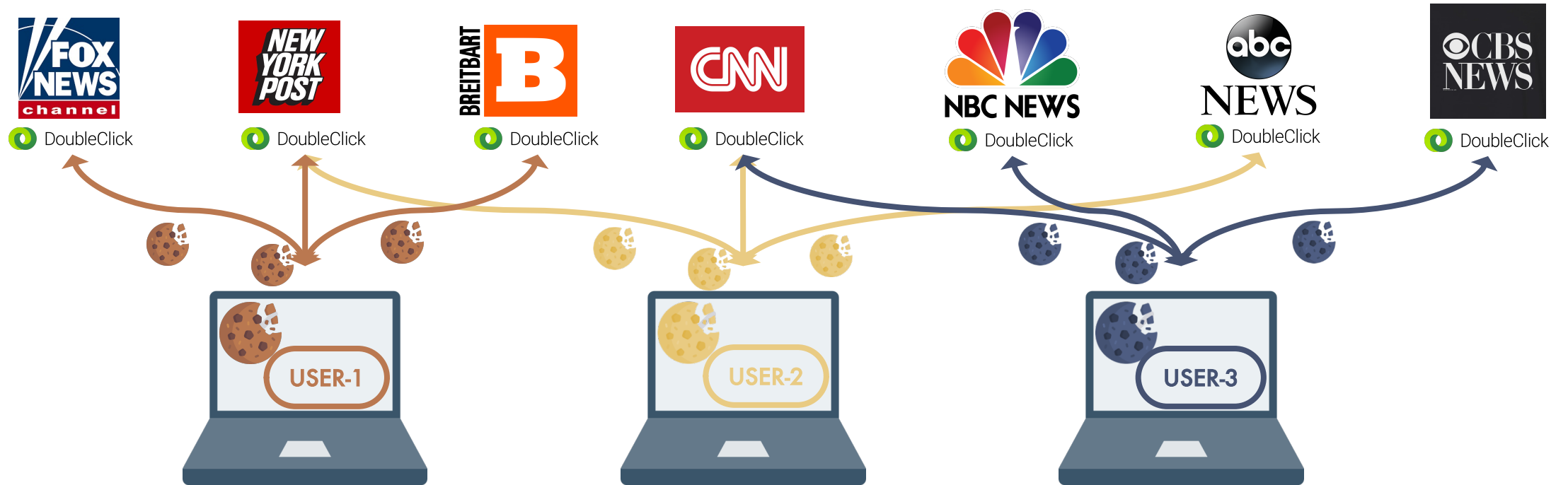
---

Umar Iqbal, Steven Englehardt, Zubair Shafiq

IEEE Symposium on Security & Privacy, 2021

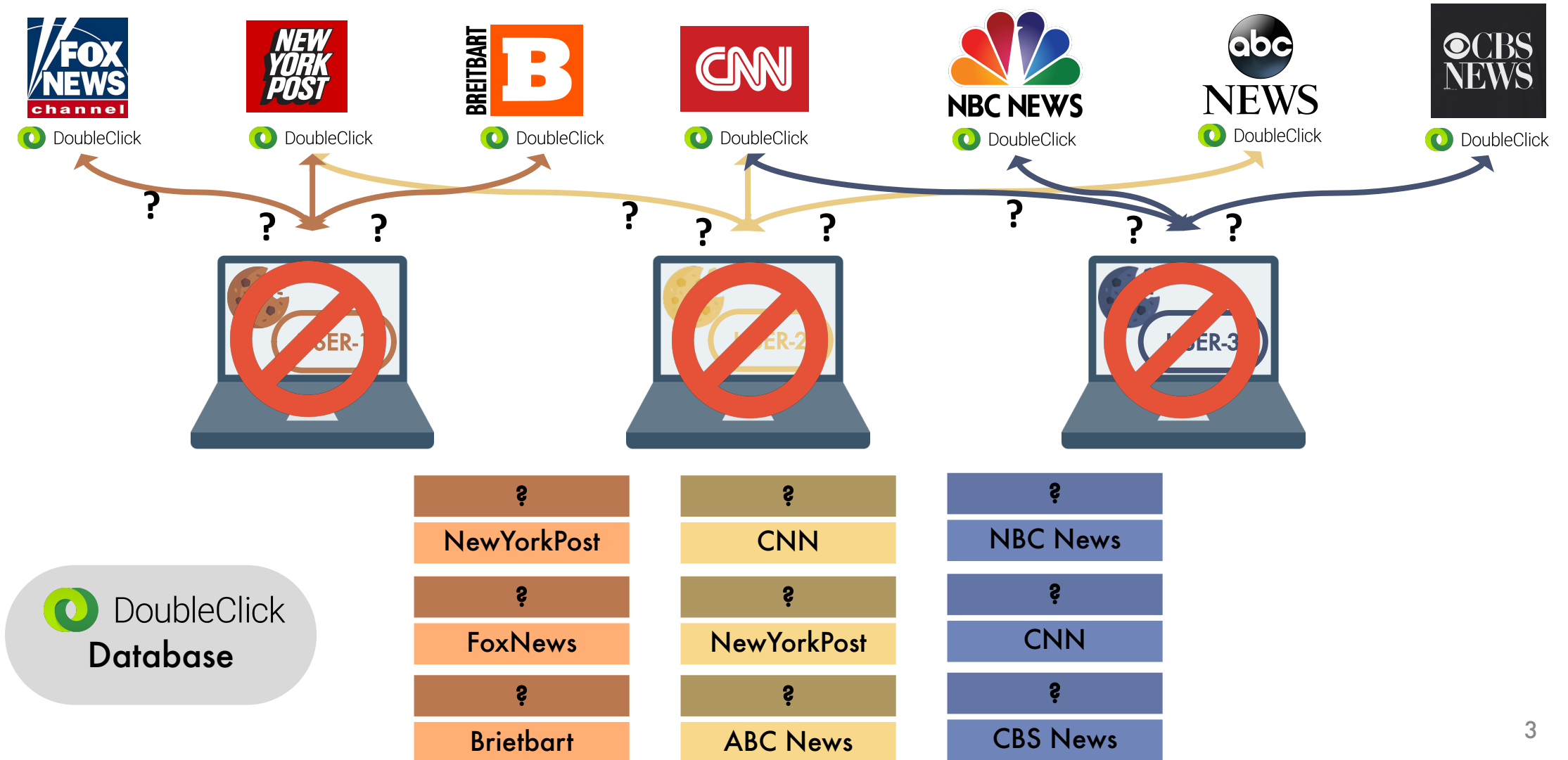


# Online Tracking



| USER-1      | USER-2      | USER-3   |
|-------------|-------------|----------|
| FoxNews     | NewYorkPost | CNN      |
| NewYorkPost | CNN         | NBC News |
| Brietbart   | ABC News    | CBS News |

# What if we block cookies?



# Cookie Blocking



## Full Third-Party Cookie Blocking and More

Mar 24, 2020 by John Wilander [@johnwilander](#)



## Today's Firefox Blocks Third-Party Tracking Cookies and

September 3, 2019

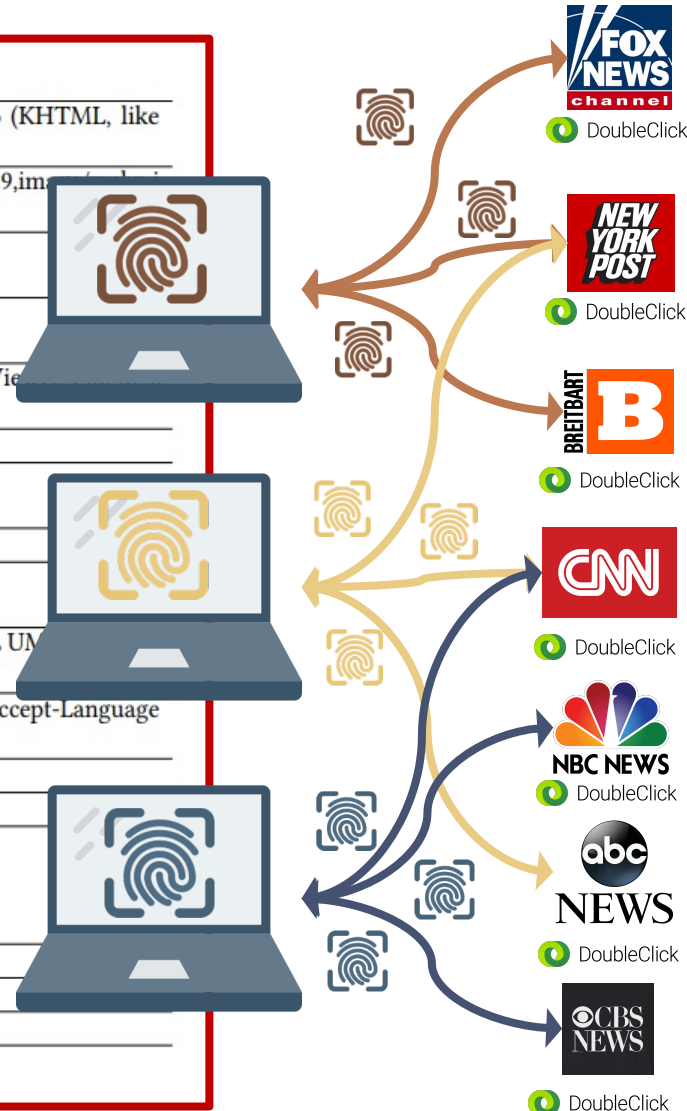


Building a more private web: A path towards making third party cookies obsolete

Tuesday, January 14, 2020

# Browser Fingerprinting

| Attribute                         | Source       | Example   |
|-----------------------------------|--------------|---|
| User agent                        | HTTP header  | Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/64.0.3282.119 Safari/537.36 |
| Accept                            | HTTP header  | text/html,application/xhtml+xml,application/xml;q=0.9,image/apng,*/*;q=0.8                                |
| Content encoding                  | HTTP header  | gzip, deflate, br   |
| Content language                  | HTTP header  | en-US,en;q=0.9  |
| List of plugins                   | JavaScript   | Plugin 1: Chrome PDF Plugin. Plugin 2: Chrome PDF Viewer Native Client. Plugin 4: Shockwave Flash...      |
| Cookies enabled                   | JavaScript   | yes   |
| Use of local/session storage      | JavaScript   | yes   |
| Timezone                          | JavaScript   | -60 (UTC+1)   |
| Screen resolution and color depth | JavaScript   | 1920x1200x24  |
| List of fonts                     | Flash or JS  | Abyssinica SIL,Aharoni CLM,AR PL UMing CN,AR PL UMing TW...   |
| List of HTTP headers              | HTTP headers | Referer X-Forwarded-For Connection Accept Cookie Accept-Language Accept-Encoding User-Agent Host          |
| Platform                          | JavaScript   | Linux x86_64  |
| Do Not Track                      | JavaScript   | yes   |
| Canvas                            | JavaScript   | Cwm fjordbank glyphs vext quiz, ☺<br>Cwm fjordbank glyphs vext quiz, ☺                                    |
| WebGL Vendor                      | JavaScript   | NVIDIA Corporation  |
| WebGL Renderer                    | JavaScript   | GeForce GTX 650 Ti/PCIe/SSE2  |
| Use of an ad blocker              | JavaScript   | yes   |



| USER-1      | USER-2      | USER-3   |
|-------------|-------------|----------|
| FoxNews     | NewYorkPost | CNN      |
| NewYorkPost | CNN         | NBC News |
| Brietbart   | ABC News    | CBS News |

# Outline

Background

State-of-the-Practice

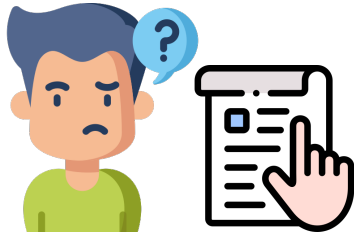
FP-Inspector

Syntactic-Semantic Representation

ML Based Detection

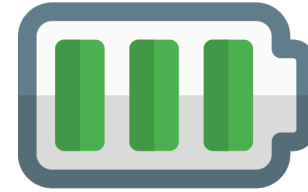
Measurements

# Countering Fingerprinting is Hard!



Sophisticated JS analysis is  
difficult & time consuming

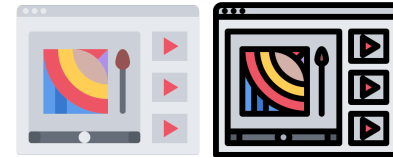
Remove  
APIs



Normalize  
APIs



Randomize  
APIs



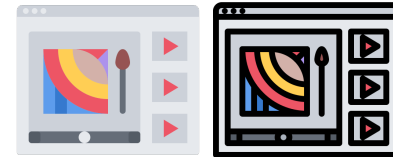
# Countering Fingerprinting is Hard!

Remove

Limit website functionality  
& cause breakage

difficult & time consuming

Randomize  
APIs



# Detecting Fingerprinting Scripts

## [canvas fingerprinting]

canvas width, height  $\geq 16$  px  
calls to save, restore = None  
[Blocked]

Heuristics as ground truth<sup>[1]</sup>

```
e.width = 780, e.height = 150;
var n = "UNICODE STRING",
i = e.getContext("2d");
i.save(); i.rect(0,0,10,10); i.rect(2,2,6,6);
t.push(!1 === i.isPointInPath(5, 5, "evenodd")
? "yes" : "no"), i.restore(), i.save();
var r = i.createLinearGradient(0, 0, 200, 0);
...
i.shadowColor="rgb(85,85,85)",i.shadowBlur=3,
i.arc(500,15,10,0,2*Math.PI,!0),i.stroke(),
i.closePath(),i.restore(),t.push(e.toDataURL())
}
```

Accuracy

```
}
d.textBaseline = "top";
d.font = "14px 'Arial'";
...
d.fillText("http://valve.github.io", 4, 17);
return b.toDataURL()
}
};
"object" === typeof module &&
  "undefined" !== typeof exports && (
    module.exports = n);
g.ClientJS = n
})(window);
```

Coverage

# Detecting Fingerprinting Scripts

```
e.width = 780, e.height = 150;  
var n = "UNICODE STRING",  
i = e.getContext("2d");  
i.save(); i.rect(0,0,10,10); i.rect(2,2,6,6);  
t.push(!1 === i.isPointInPath(5, 5, "evenodd"  
? "yes" : "no"), i.restore(), i.save());
```

## Accuracy & Coverage Issues

Heuristics as ground truth<sup>[1]</sup>

```
...  
d.fillText("http://valve.github.io", 4, 17);  
return b.toDataURL()  
}  
};  
"object" === typeof module &&  
  "undefined" !== typeof exports && (  
    module.exports = n);  
g.ClientJS = n  
})(window);
```

Coverage

# Outline

Background

State-of-the-Practice

FP-Inspector

Syntactic-Semantic Representation

ML Based Detection

Measurements

# Detecting Fingerprinting Behavior

```
<script>
  // Canvas font fingerprinting script.
  Fonts = ["monospace", ... , "sans-serif"];
  CanvasElem = document.createElement("canvas");
  CanvasElem.width = "100";
  CanvasElem.height = "100";
  context = CanvasElem.getContext('2d');
  FPDict = {};

  for (i = 0; i < Fonts.length; i++) {
    CanvasElem.font = Fonts[i];
    FPDict[Fonts[i]] = context.measureText("example").width;
  }

  var img = document.createElement("img");
  img.src = "tracker.com/track_user/?userId={FPDict}";
</script>
```

HTML canvas element

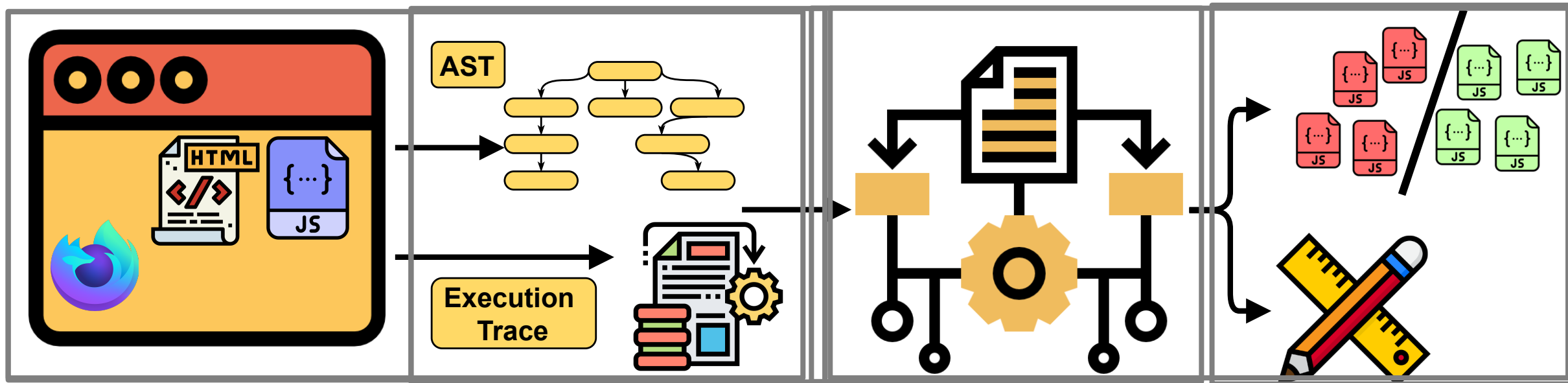
APIs with fingerprinting potential

Network request with fingerprint

# FP-Inspector

Syntactic-Semantic representation  
of fingerprinting scripts

ML to automatically learn fingerprinting  
behavior



OpenWPM  
instrumented Firefox

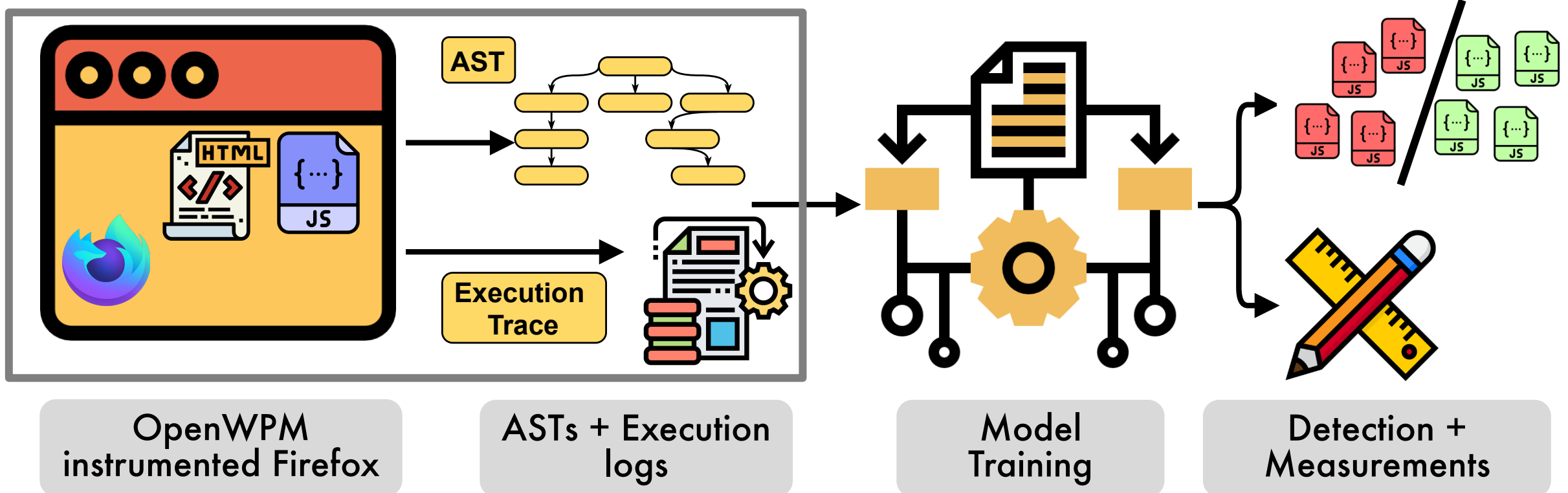
ASTs + Execution  
logs

Model  
Training

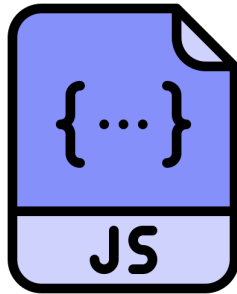
Detection +  
Measurements

# FP-Inspector

Syntactic-Semantic representation  
of fingerprinting scripts



# Syntactic Representation

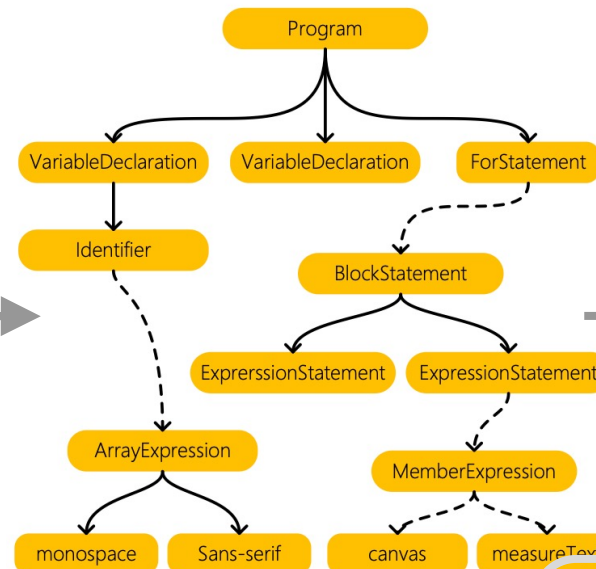


Source as a signal of functionality

People write code differently

Code normalization with  
Abstract Syntax Trees  
(ASTs)

```
// Canvas font fingerprinting script.  
Fonts = ["monospace", ... , "sans-serif"];  
CanvasElem = document.createElement("canvas");  
CanvasElem.width = "100";  
CanvasElem.height = "100";  
context = CanvasElem.getContext('2d');  
FPDict = {};  
  
for (i = 0; i < Fonts.length; i++) {  
  CanvasElem.font = Fonts[i];  
  FPDict[Fonts[i]] = context.measureText("example").width;  
}
```



ArrayExpression:monospace

MemberExpression:getContext

ForStatement:var

CallExpression:canvas

MemberExpression:measureText

MemberExpression:width

MemberExpression:font

MemberExpression:length

# Semantic Representation



Execution as a signal of functionality

Frequency + Higher-level semantic of APIs

```
// Canvas font fingerprinting script.
Fonts = ["monospace", ... , "sans-serif"];
CanvasElem = document.createElement("canvas");
CanvasElem.width = "100";
CanvasElem.height = "100";
context = CanvasElem.getContext('2d');
FPDict = {};

for (i = 0; i < Fonts.length; i++) {
    CanvasElem.font = Fonts[i];
    FPDict[Fonts[i]] = context.measureText("example").width;
}
```

Document.createElement True

HTMLCanvasElement.height True

HTMLCanvasElement.width True

HTMLCanvasElement.getContext True

CanvasRenderingContext2D.measureText True

Element Tag Name canvas

HTMLCanvasElement.width 100

HTMLCanvasElement.height 100

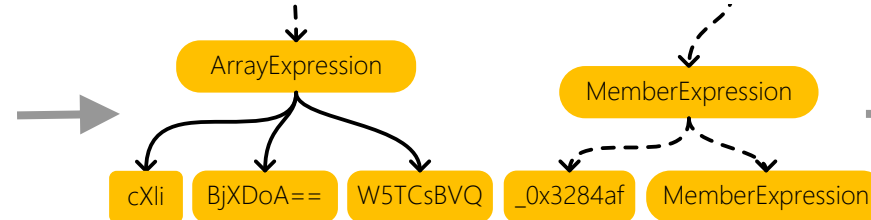
CanvasRenderingContext2D.measureText 50 (no. of calls)

CanvasRenderingContext2D.measureText 7 (no. of chars)

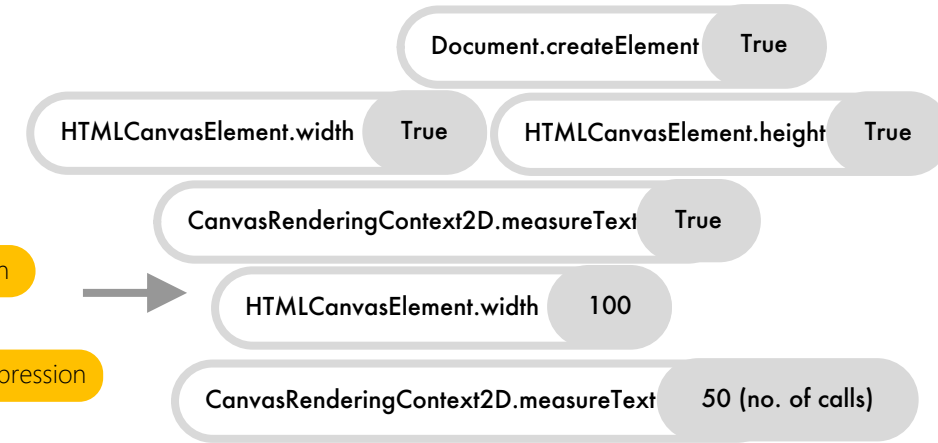
# Syntactic vs Semantic

```
var _0x2c4a=['\x63\x58\x49\x69','\x42\x6a\x58\x44\x6f\x41\x3d\x3d','\x55\x54\x72\x43\x69\x73\x4f\x77\x4f\x38\x4f\x6c\x50\x45\x6e\x43\x6d\x77\x30\x3d','\x49\x38\x4f\x38\x49\x4d\x4f\x42\x77\x70\x72\x44\x6e\x41\x3d\x3d','\x77\x35\x54\x43\x73\x42\x56\x51','\x77\x37\x62\x43\x69\x4d\x4f\x38\x77\x35\x54\x43\x73\x42\x56\x51\x3284af={};
for(i=0x0;i<_0x1b2b65[_0x5d52('0x7','\x28\x6d\x68\x26')];i++){_0x1d1d56[_0x5d52('0x8','\x67\x33\x48\x21')]=_0x1b2b65[i];_0x3284af[_0x1b2b65[i]]=_0x4d24cc[_0x5d52('0x9','\x35\x70\x64\x4c')]( _0x5d52('0xa','\x28\x6d\x68\x26'))['_\x77\x69\x64\x74\x68'];}
```

Obfuscated



Syntactic representation



Semantic representation

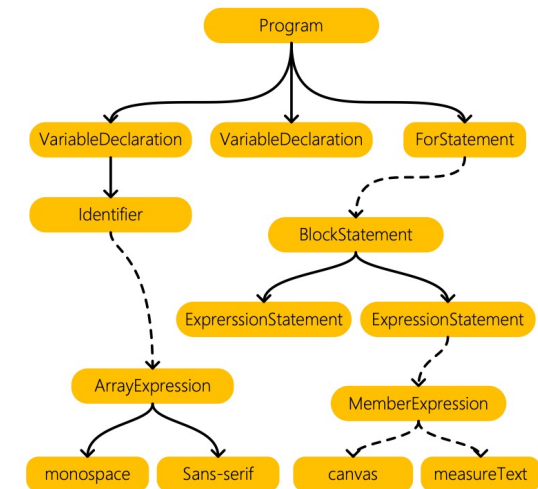
```
(function(g) {
  .....
  n.prototype = {
    getCanvasPrint: function() {
      var b = document.createElement("canvas"),d;
      try {
        d = b.getContext("2d")
      } catch (e) {
        return ""
      }
      d.textBaseline = "top";
      d.font = "14px 'Arial'";
      ...
      d.fillText("http://valve.github.io", 4, 17);
      return b.toDataURL()
    }
  };
  "object" === typeof module &&
  "undefined" !== typeof exports && (
    module.exports = n);
  g.ClientJS = n
})(window);
```

Dormant

NULL

NULL

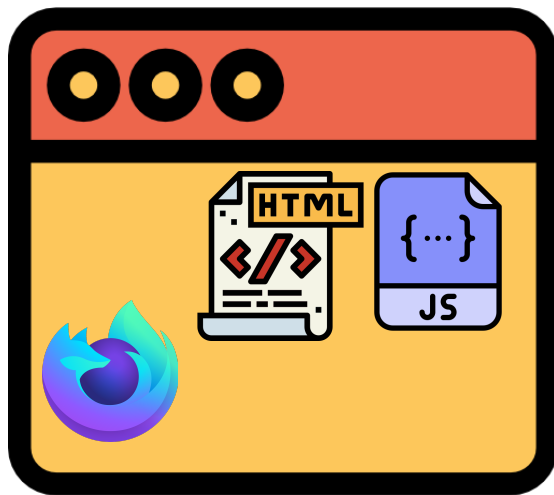
Semantic Representation



Syntactic representation

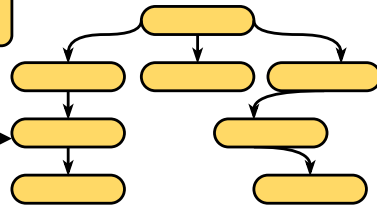
# FP-Inspector

ML to automatically learn fingerprinting behavior

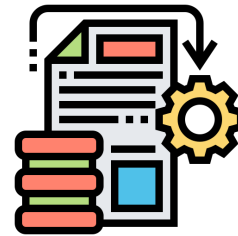


OpenWPM  
instrumented Firefox

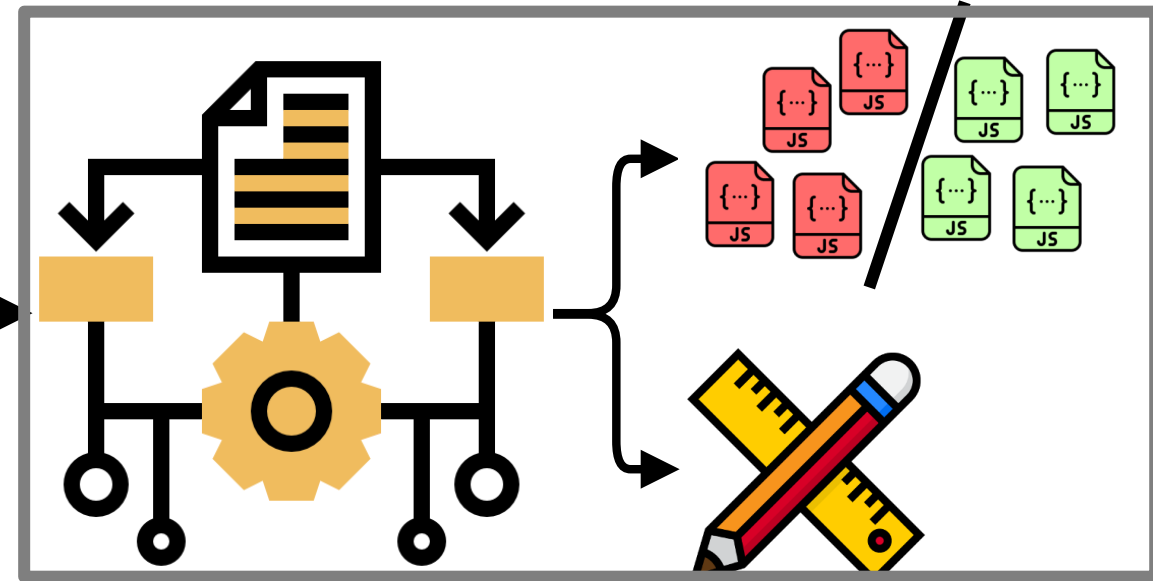
AST



Execution  
Trace



ASTs + Execution  
logs



Model  
Training

Detection +  
Measurements

# Classification



No ground truth

Heuristics as ground truth<sup>[1]</sup>

Decision tree-based ensemble

Iterative improvement

Incomplete ground truth



Iterative  
re-training

Enhanced ground truth



26% more scripts than heuristics



99.9% accuracy



10% net gain with syntactic + semantic

# State of Fingerprinting in the Wild



How much?

One-third of popular websites

519 / Million '16<sup>[1]</sup>



2,349 / 100K '19

Where?

Websites that use ads for monetization

14% News

Who?

Popular trackers & Ad fraud services

DoubleVerify

Integral Ad Service

# Key Takeaways



FP-Inspector improves the state-of-the-art



Syntactic-Semantic representation for improved accuracy



Machine learning for automated detection



Prevalence of browser fingerprinting

# Key Takeaways



Open-source implementation & detections

## FP-Inspector

Artifact release for our IEEE Symposium on Security and Privacy 2021 paper entitled Fingerprinting the Fingerprinters: Learning to Detect Browser Fingerprinting Behaviors



<https://uiowa-irl.github.io/FP-Inspector/>



Incorporated by filter lists

EasyList



DISCONNECT.



New fingerprinting category in EasyPrivacy

### Fingerprinting

```
breaktime.com.tw^$third-party  
brightedge.com^$third-party  
citrusad.net^$third-party  
clickguardian.app^$third-party  
clickyab.com^$third-party  
franecki.net^$third-party  
guoshipartners.com^$third-party  
hostpush.info^$third-party  
ismatlab.com^$third-party  
iwqzrm.com^$third-party  
klangoo.com^$third-party  
p30rank.ir^$third-party  
ppcprotect.com^$third-party  
push4site.com^$third-party  
ravelin.net^$third-party  
sabavision.com^$third-party  
sealine.pro^$third-party  
simility.com^$third-party  
spideraf.com^$third-party
```

# Question?



Paper Link

[umariqbal.com/papers/fpinspector-sp2021.pdf](http://umariqbal.com/papers/fpinspector-sp2021.pdf)

Source & Data

[uiowa-irl.github.io/FP-Inspector/](https://uiowa-irl.github.io/FP-Inspector/)

Umar Iqbal



@umaarr6



umariqbal.com

Contact details

# References

1. Online tracking: A 1-million-site measurement and analysis  
Steven Englehardt, Arvind Narayanan  
2016 ACM SIGSAC conference on computer and communications security (CCS)
2. Icons from flaticons