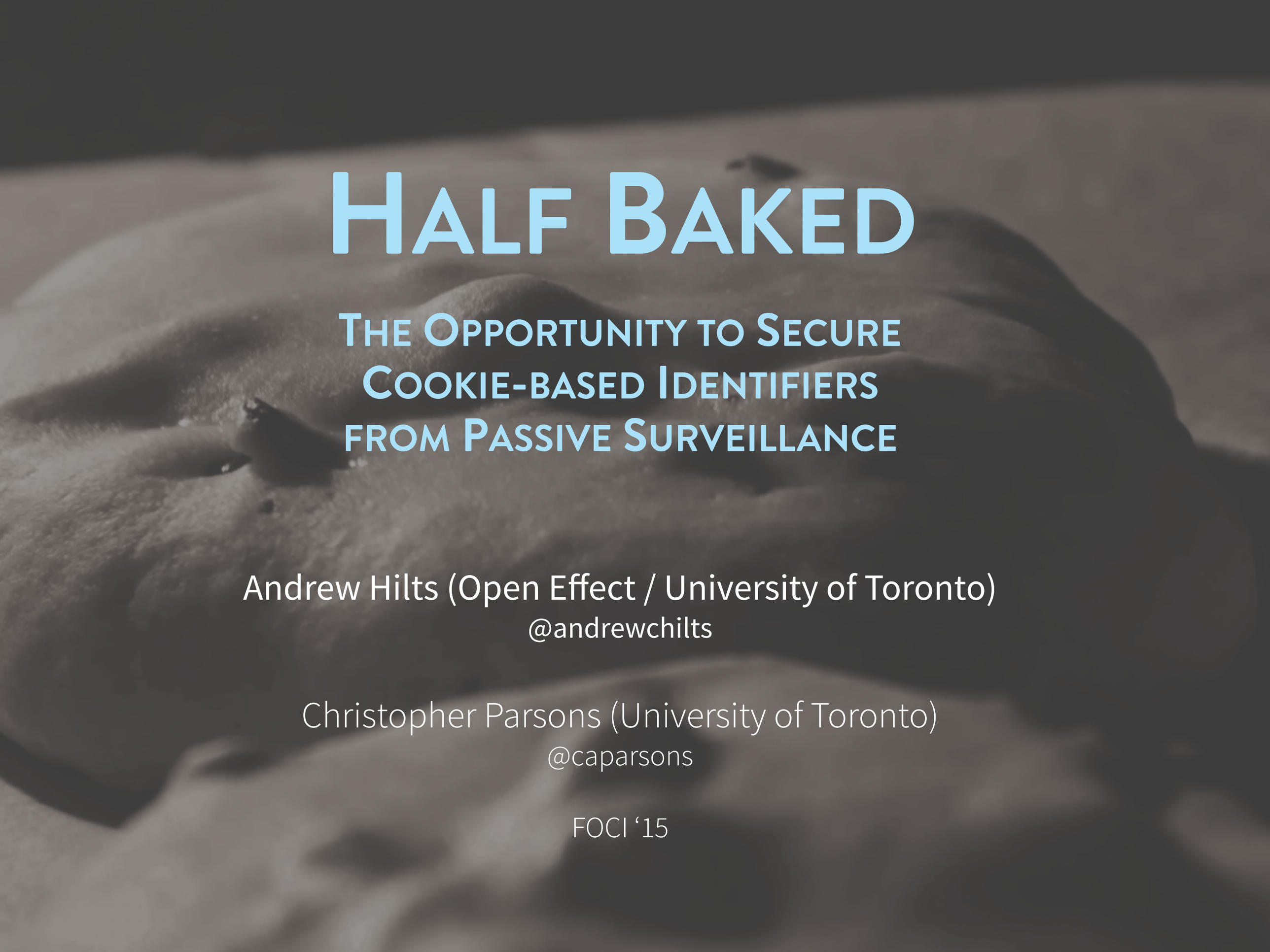


# Half Baked

## The Opportunity to Secure Cookie-based Identifiers from Passive Surveillance

Andrew Hilts (Open Effect / University of Toronto)
@andrewchilts

Christopher Parsons (University of Toronto)
@caparsons

FOCI ‘15

# Overview

Internet Metadata Surveillance

The role of cookie-based identifiers

HTTPS security and third party dependencies

Methodology

Results

Discussion

Conclusions

# Internet Metadata Surveillance

TOP SECRET//SI//REL USA, AUS, CAN, GBR, NZL

# Cookie-based identifiers

Same tracking identifier transmitted in the clear by multiple websites

nytimes.com

```
▶ Internet Protocol Version 4, Src: 192.168.1.104 (192.168.1.104), Dst: 96.17.10.80 (96.17.10.80)
▶ Transmission Control Protocol, Src Port: 61573 (61573), Dst Port: 80 (80), Seq: 1, Ack: 1, Len: 589
▼ Hypertext Transfer Protocol
  ▶ GET /b?c1=2&c2=3005403&ns__t=1438873671904&ns_c=UTF-8&c8=The%20New%20York%20Times%20-%20Breaking%
    Host: b.scorecardresearch.com\r\n
    Connection: keep-alive\r\n
    Accept: image/webp,*/*;q=0.8\r\n
    User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_10_4) AppleWebKit/537.36 (KHTML, like Gecko)
    DNT: 1\r\n
    Referer: http://www.nytimes.com/\r\n
    Accept-Encoding: gzip, deflate, sdch\r\n
    Accept-Language: en-US,en;q=0.8\r\n
  ▼ Cookie: UID=a8098d4-184.84.243.200-1415029960; UIDR=1415029960\r\n
      Cookie pair: UID=a8098d4-184.84.243.200-1415029960
      Cookie pair: UIDR=1415029960
    \r\n
    [Full request URI: http://b.scorecardresearch.com/b?c1=2&c2=3005403&ns__t=1438873671904&ns_c=UTF-
    [HTTP request 1/1]
    [Response in frame: 6425]
```

# HTTPS AND 3RD PARTY DEPENDENCIES

Progress!

The Washington Post

The Switch

**News sites could protect your privacy with encryption. Here's why they probably won't.**

By Andrea Peterson December 11, 2013 Follow @kansasalps

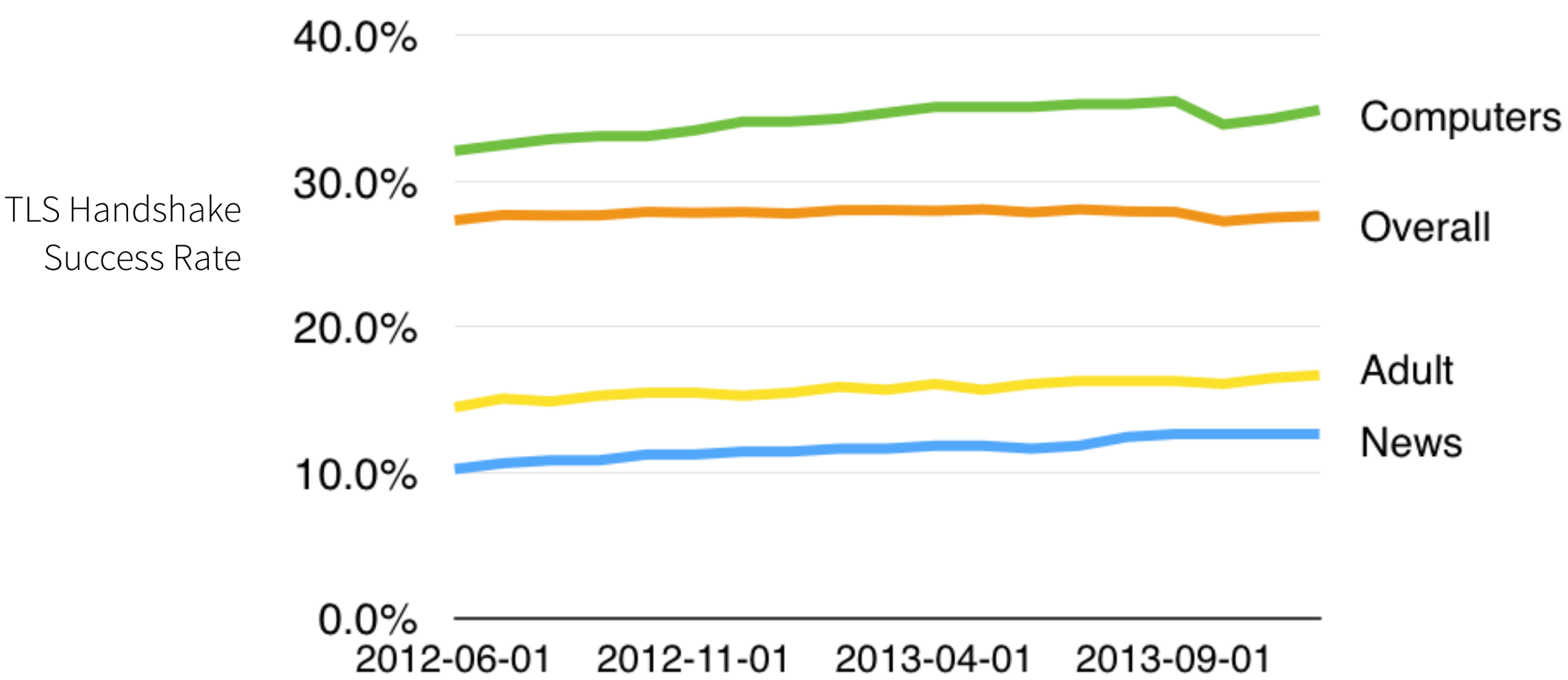
# Research Questions

To what extent are cookie-based identifiers actually encrypted in transit?

Does this differ across different Alexa categories?

Does the level of HTTPS support a website offers affect the number and security of transmitted identifiers?

# TLS Adoption Over Time




Longitudinal data from Zmap dumps: https://scans.io/study/umich-https

# Primary Data Collection

Alexa Web Information Service

Get top 500 sites per category

Categorized URL List

Selenium Chrome Engine

Load each URL

HTTP(S) Requests

mitmproxy

Intercept (and decrypt) each HTTP(S) Request initiated by loaded URL

Request Headers

Results db

Parse and save headers for analysis
Run tests

# Flagging Identifier Cookies

key matches `id`

key matches `pref`

key matches regexes:

```
.*id$

ident

uuid

user

_id

fingerprint
```

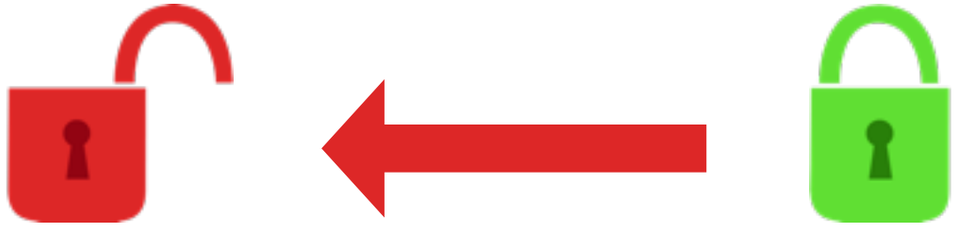
value matches regexes:

```
.*id$

ident

uuid

user

_id

fingerprint

id=

[0-9a-f]{8}-[0-9a-f]{4}-[1-5][0-9a-f]{3}-[89ab][0-9a-f]{3}-[0-9a-f]{12}
```

# HTTPS In Practice

No HTTPS

HTTPS downgrade



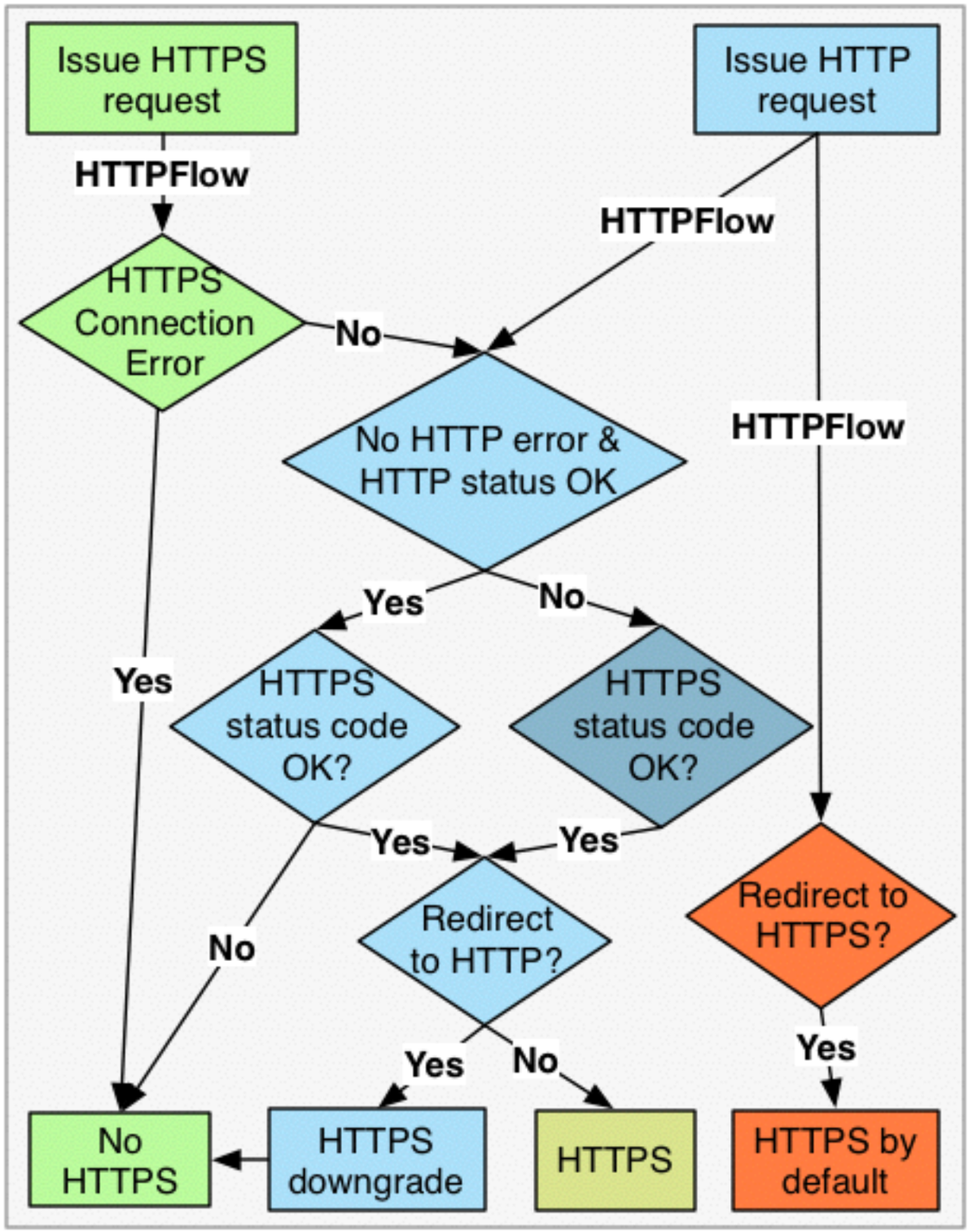
HTTPS support

OR

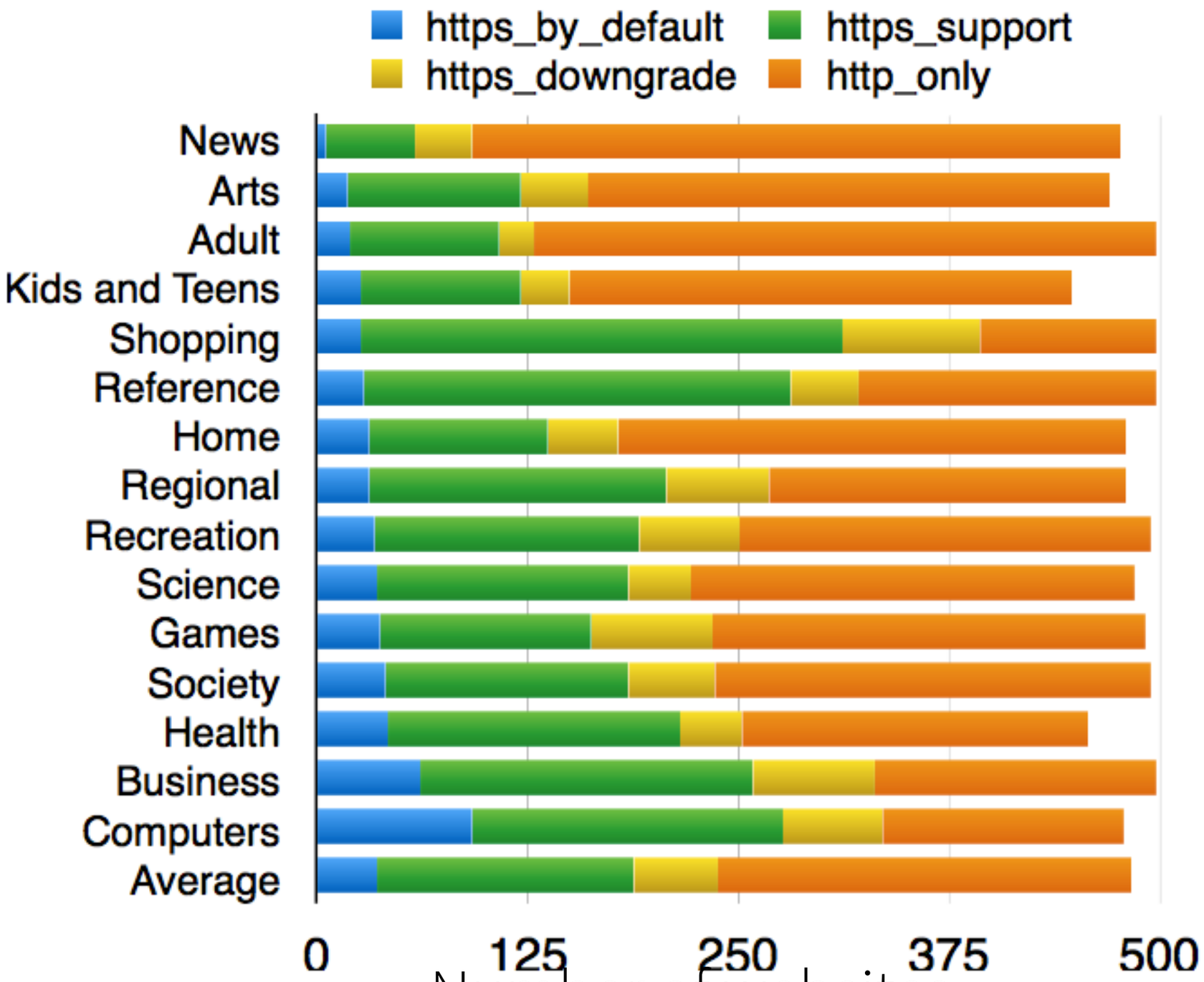
HTTPS by default

# Categorizing Practical HTTPS Support

Run this test on 5 random paths for each hostname and reconcile results.



# HTTPS Support in Practice by Category



# First and Third Party HTTPS

Websites that downgrade HTTPS connections communicate with the most third parties. (14.8 avg)

Website's HTTPS practice by % of third party comms secured

- No HTTPS: **68%**
- HTTPS Downgrade: **64%**
- HTTPS Support: **61%**
- HTTPS by Default: **93%**

31.4% of identifier transmissions to top ad trackers that support HTTPS were secured. This leaves 68.6%, or 9414 transmissions that could have been secured.

# Discussion

b.scorecardresearch.com is the only top tracker that does not support HTTPS. Unique ID sent insecurely from 1772 different websites.

News websites only encrypt identifier transmissions 23% of the time, leaving their reader's habits vulnerable to surveillance.

Given that 9/10 top trackers support HTTPS, it shouldn't be difficult for news and other sites to turn on TLS for those trackers for a big privacy benefit.

# Limitations

Single user-agent (Selenium Chrome webdriver)

Single network vantage point

Single page load

HTTPS in practice test error rate of 6%

Identifier cookie test doesn’t check for base64 or other obfuscation

UMich data only has one cert per IP address

# Conclusions

Not much has changed in terms of overall TLS adoptions post-Snowden.

Many websites can simply switch on major ad tracker security to better protect their users' privacy.

We plan to re-run tests every 6 months to assess changes in ecosystem.