Third-party Tracking on the Web: A Swedish Perspective

Joel Purra and Niklas Carlsson

Linköping University, Sweden

@ IEEE LCN, Dubai, Nov. 2016

LINKÖPINGS UNIVERSITET

The New York Times - Bre

www.nytimes.com

SECTIONS SEARCH

SUBSCRIBE NOW LOG IN

U.S. INTERNATIONAL 中文

The New York Times

Thursday, February 19, 2015 | Today's Paper | Video | CAC 40 +0.20% †

World U.S. Politics New York Business Opinion Technology Science Health Sports Arts Style Food Home Travel Magazine Real Estate ALL

Obama Tells of Programs to Fight the Draw of Extremists

By JULIE HIRSCHFELD DAVIS

The president outlined his administration's efforts to counter what he called "violent extremism" in a speech to law enforcement, community and religious leaders.

311 Comments

From a Private School in Cairo to ISIS Killing Fields

Kim Gordon, pictured in Los Angeles, said her new memoir is "the most conventional thing I've done." Sam Comen for The New York Times

Sonic Youth's Rock Recluse, Opening Up

By JOE COSCARELLI

In her memoir, "Girl in a Band," Kim Gordon, Sonic Youth's

The Opinion Pages

OP-ED CONTRIBUTOR

My Own Life

By OLIVER SACKS

I am now face to face with dying. But I am not finished with living.

- Blow: The Obama Years
- Collins: A Gun on Every Corner
- Kristof: The Cost of a Decline in Unions
- Greenhouse: Groundhog Day at the Supreme Court

When One Penalty Is Enough

By THE EDITORIAL BOARD

The fines for not signing up for health insurance in 2014 under the Affordable Care Act are coming due at tax time. Fines for 2015 will be worse.

MORE IN OPINION

- Editorial: Regulating the Drone Economy
- Op-Ed: How to Hold Colleges Accountable

We are all tracked …

- When browsing, information is recorded by the servers you communicate directly with
  - Resources from other services might be requested as well, with or without being visible.
- Information can be passively recorded during transmission; some of which can't be avoided
- Specialized tracking code can actively extract extended information

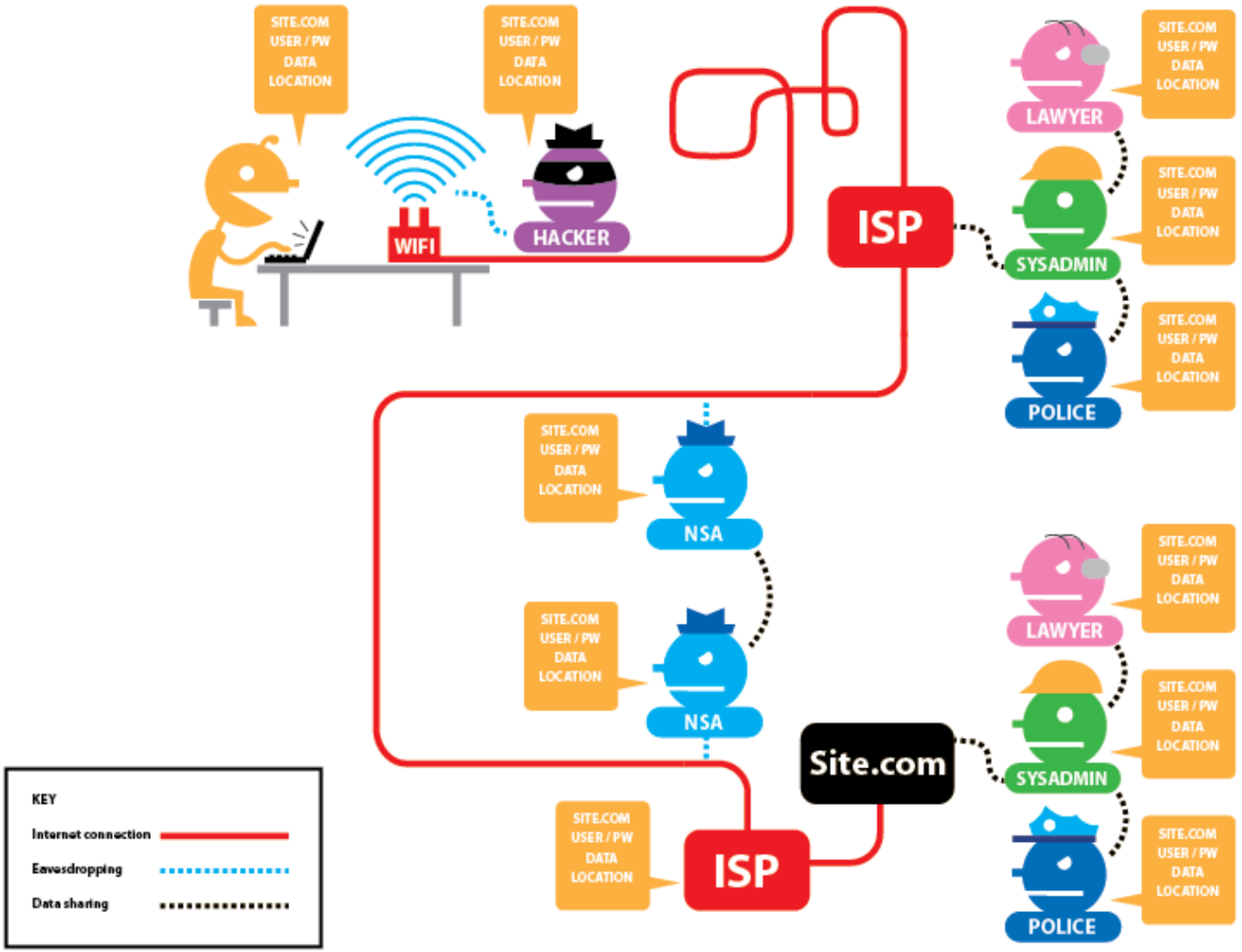
Why is tracking used?

- Information is collected and stored to gain knowledge about the visitors a website has.
  - Website owners: to improve/personalize content
  - Advertisement firms: to sell targeted ads
  - Media analytics firms: to verify statistics (for ads).
  - Data brokers: to package and sell (inferred) user data

The downside

- Users lose control over who they share information with. This can be considered an invasion of privacy.
- Information is easily stored and easily retrieved,
  - Anything done online in the past can haunt you for ever.
  - Self-censorship, effectively limiting freedom of speech.
- What is illegal for governments, companies are allowed to do, through user agreements. Governments still have control over companies within their jurisdiction.
- The full scope of the tracking is still unknown
  - Could become a historical thought police.
  - Could mean online companies have a grip on all current and future politicians, company leaders and celebrities.

Passive tracking and HTTPS



Passive vs active tracking

- Passive tracking: Anyone can listen in anywhere along the network path ...
  - People are becoming increasingly aware of monitoring by ISPs and nation state ...
  - HTTPS prevents passive tracking of some information (e.g., exact page, browser model, OS, language settings, cookies, etc.)

Passive vs active tracking

- Passive tracking: Anyone can listen in anywhere along the network path ...
  - People are becoming increasingly aware of monitoring by ISPs and nation state ...
  - HTTPS prevents passive tracking of some information (e.g., exact page, browser model, OS, language settings, cookies, etc.)
- Active tracking: A script or plugin executed in the browser to extract and collect extended information.
  - HTTPS does **not** prevent this.
  - Example info include time spent on each page, window size, screen resolution, color depth, mouse movements, scrollbar location, installed fonts, plugins and extensions.

Passive vs active tracking

- Passive tracking: Anyone can listen in anywhere along the network path ...
  - People are becoming increasingly aware of monitoring by ISPs and nation state ...
  - HTTPS prevents passive tracking of some information (e.g., exact page, browser model, OS, language settings, cookies, etc.)
- Active tracking: A script or plugin executed in the browser to extract and collect extended information.
  - HTTPS does **not** prevent this.
  - Example info include time spent on each page, window size, screen resolution, color depth, mouse movements, scrollbar location, installed fonts, plugins and extensions.
- We focus on third-party tracking, but ask if sites implementing HTTPS use less tracking themselves

This paper …

- … presents measurement methodology and characterization of the current third-party tracking landscape

This paper …

- … presents measurement methodology and characterization of the current third-party tracking landscape
  - Third-party usage across a number of website classes and breakdown the coverage of different tracker types
  - Aggregate analysis that combines the tracker services based on the organizations operating them so to gain insights into the big players aggregate coverage
  - Try to answer if websites that have adopted HTTPS in fact are more privacy conscious (on behalf of their users) and use less third-party tracking.

This paper …

- … presents measurement methodology and characterization of the current third-party tracking landscape
  - Third-party usage across a number of website classes and breakdown the coverage of different tracker types
  - Aggregate analysis that combines the tracker services based on the organizations operating them so to gain insights into the big players aggregate coverage
  - Try to answer if websites that have adopted HTTPS in fact are more privacy conscious (on behalf of their users) and use less third-party tracking.

This paper …

- … presents measurement methodology and characterization of the current third-party tracking landscape
  - Third-party usage across a number of website classes and breakdown the coverage of different tracker types
  - Aggregate analysis that combines the tracker services based on the organizations operating them so to gain insights into the big players aggregate coverage
  - Try to answer if websites that have adopted HTTPS in fact are more privacy conscious (on behalf of their users) and use less third-party tracking

Methodology

- Developed data collection tool
  - Headless phantom.js browser
- Visit front page of large number of sites
  - HTTP vs HTTPS (with and without www)
  - Measure redirects etc.
  - Process/execute scripts to build pages
  - No blocking
  - Extract URL, domain, and other info
- Classify resources
  - Internal vs. external
  - Known trackers (using Disconnect.me)
  - Type of resource; e.g., advertising, analytics, content

Swedish perspective

SUMMARY OF DOMAIN LISTS.

Domain lists			Selection		
List(s)	Date	Total size	Type	Size	Unique
.SE Health Report	27/3/14	980	curated (9 categories)	915	915
.se zone	10/7/14	1,318,000	random	100,000	100,000
.dk zone	23/7/14	1,260,000	random	10,000	10,000
.com zone	27/8/14	114,178,000	random	10,000	10,000
.net zone	27/8/14	15,096,000	random	10,000	10,000
Reach 50	1/9/14	50	top	50	50
Alexa top-1M	1/9/14	1,000,000	top	10,000	9,986
–"–	–"–	–"–	random	10,000	9,959
–"–	–"–	–"–	all .se	3,364	3 364
–"–	–"–	–"–	all .dk	2,637	2,637
Total	–	132,852,050	–	156,907	156,045

- Measurements performed from Sweden
- Important and popular Swedish domains
- Global baseline

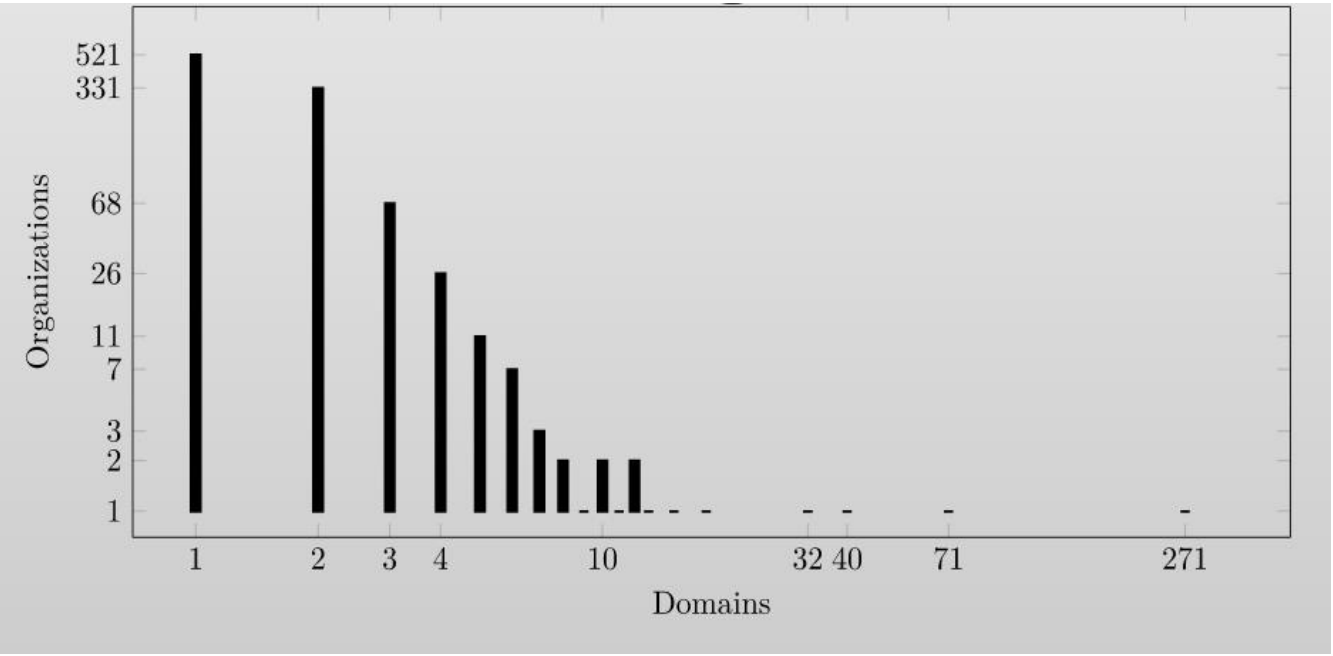
What are third-party resources?

- A resource belonging to the origin's primary domain is called internal. Otherwise it's an external resource.
- Assumption: Any external resource is a third-party resource.

Domain examples
example.se (primary domain)
www.example.se (subdomain)
example.org (third-party domain)
doubleclick.net (known tracker domain)

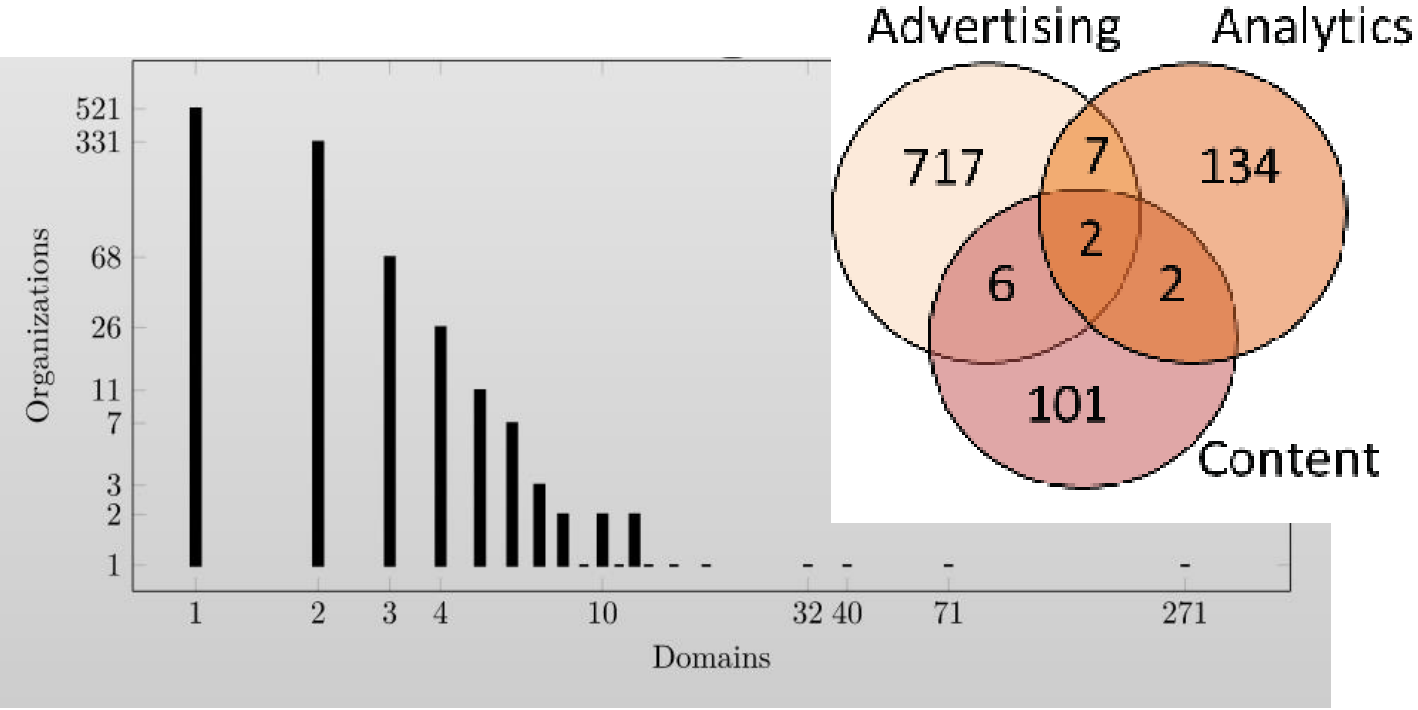
Resource examples
Branded (videos, services, images)
Unbranded (fonts, useful scripts, images)
Ads (scripts, images, flash)
Web beacons (hidden images, analytics scripts)

Blocked domains on Disconnect.me



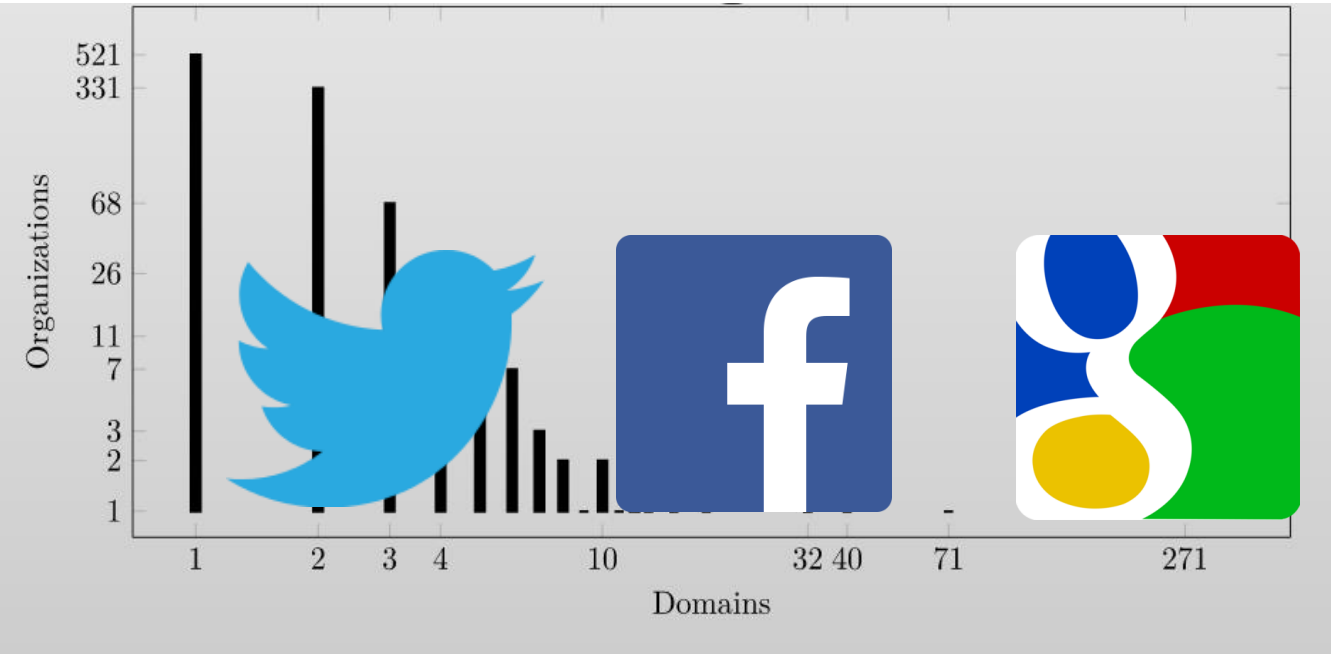
- Many have few: 521 out of 980 organization have 1 domain; 331 have 2 domain.
- Some have many: Google has 271, Yahoo 71, AOL 40, Microsoft 32.

Blocked domains on Disconnect me



- Many have few: 521 out of 980 organization have 1 domain; 331 have 2 domain.
- Some have many: Google has 271, Yahoo 71, AOL 40, Microsoft 32.
- Spread over advertising, analytics, content

Blocked domains on Disconnect.me



- Many have few: 521 out of 980 organization have 1 domain; 331 have 2 domain.
- Some have many: Google has 271, Yahoo 71, AOL 40, Microsoft 32.
- Spread over advertising, analytics, content
- “Disconnect category”: Google, Facebook, Twitter

The New York Times - Bre

www.nytimes.com

SECTIONS SEARCH

SUBSCRIBE NOW LOG IN

U.S. INTERNATIONAL 中文

The New York Times

Thursday, February 19, 2015 | Today's Paper | Video | CAC 40 +0.20% †

World U.S. Politics New York Business Opinion Technology Science Health Sports Arts Style Food Home Travel Magazine Real Estate ALL

Obama Tells of Programs to Fight the Draw of Extremists

By JULIE HIRSCHFELD DAVIS

The president outlined his administration's efforts to counter what he called "violent extremism" in a speech to law enforcement, community and religious leaders.

311 Comments

From a Private School in Cairo to ISIS Killing Fields

Kim Gordon, pictured in Los Angeles, said her new memoir is "the most conventional thing I've done." Sam Comen for The New York Times

Sonic Youth's Rock Recluse, Opening Up

By JOE COSCARELLI

In her memoir, "Girl in a Band," Kim Gordon, Sonic Youth's

The Opinion Pages

OP-ED CONTRIBUTOR

My Own Life

By OLIVER SACKS

I am now face to face with dying. But I am not finished with living.

- Blow: The Obama Years
- Collins: A Gun on Every Corner
- Kristof: The Cost of a Decline in Unions
- Greenhouse: Groundhog Day at the Supreme Court

When One Penalty Is Enough

By THE EDITORIAL BOARD

The fines for not signing up for health insurance in 2014 under the Affordable Care Act are coming due at tax time. Fines for 2015 will be worse.

MORE IN OPINION

- Editorial: Regulating the Drone Economy
- Op-Ed: How to Hold Colleges Accountable

The New York Times - Bre

www.nytimes.com

SECTIONS SEARCH

U.S. INTERNATIONAL 中文

The New York Times

Thursday, February 19, 2015 | Today's Paper | Video | CAC 40 +0.20% †

World U.S. Politics New York Business Opinion Technology Science Health Sports Arts Style Food Home Travel

Obama Tells of Programs to Fight the Draw of Extremists

By JULIE HIRSCHFELD DAVIS

The president outlined his administration's efforts to counter what he called "violent extremism" in a speech to law enforcement, community and religious leaders.

311 Comments

From a Private School in Cairo to ISIS Killing Fields

Kim Gordon, pictured in Los Angeles, said her new memoir is "the most conventional thing I've done." Sam Comen for The New York Times

Sonic Youth's Rock Recluse, Opening Up

By JOE COSCARELLI

In her memoir, "Girl in a Band," Kim Gordon, Sonic Youth's

The Opi

OP-ED CONTRIBUTOR

My Own Life

By OLIVER SACKS

I am now face to face with dying. But I am not finished with living.

- Blow: The Obama Years
- Collins: A Gun on Every Corner
- Kristof: The Cost of a Decline in Unions
- Greenhouse: Groundhog Day at the Supreme Court

MORE IN OPINION

- Editorial: Regulating the Drone Economy
- Op-Ed: How to Hold Colleges Accountable

DISCONNECT Help Share

f 0 g 1 t 7

Advertising 5 requests

Analytics 4 requests

Social 0 requests

Content 0 requests

Whitelist site Visualize page

Show counter Cap counter

Time saved Bandwidth saved

Upgrade to Premium

The New York Times

The New York Times - Brea

www.nytimes.com

SECTIONS SEARCH

Times JOURNEYS
Changing the way you think about travel.
DISCOVER & BOOK

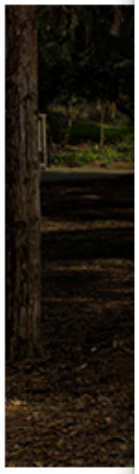
World U.S. Politics New York Busine

Obama Tells of Programs to Fight the Draw of Extremists

By JULIE HIRSCHFELD DAVIS

The president outlined his administration's efforts to counter what he called "violent extremism" in a speech to law enforcement, community and religious leaders.

311 Comments



Kim Gordon, p
conventional t

Sonic Youth's Rock Recluse, Opening Up

By JOE COSCARELLI

In her memoir, "Girl in a Band," Kim Gordon, Sonic Youth's

From a Private School in Cairo to ISIS Killing Fields

DISCONNECT

Show list view

scorecardresearch.com

This site is informed when you visit the following sites:

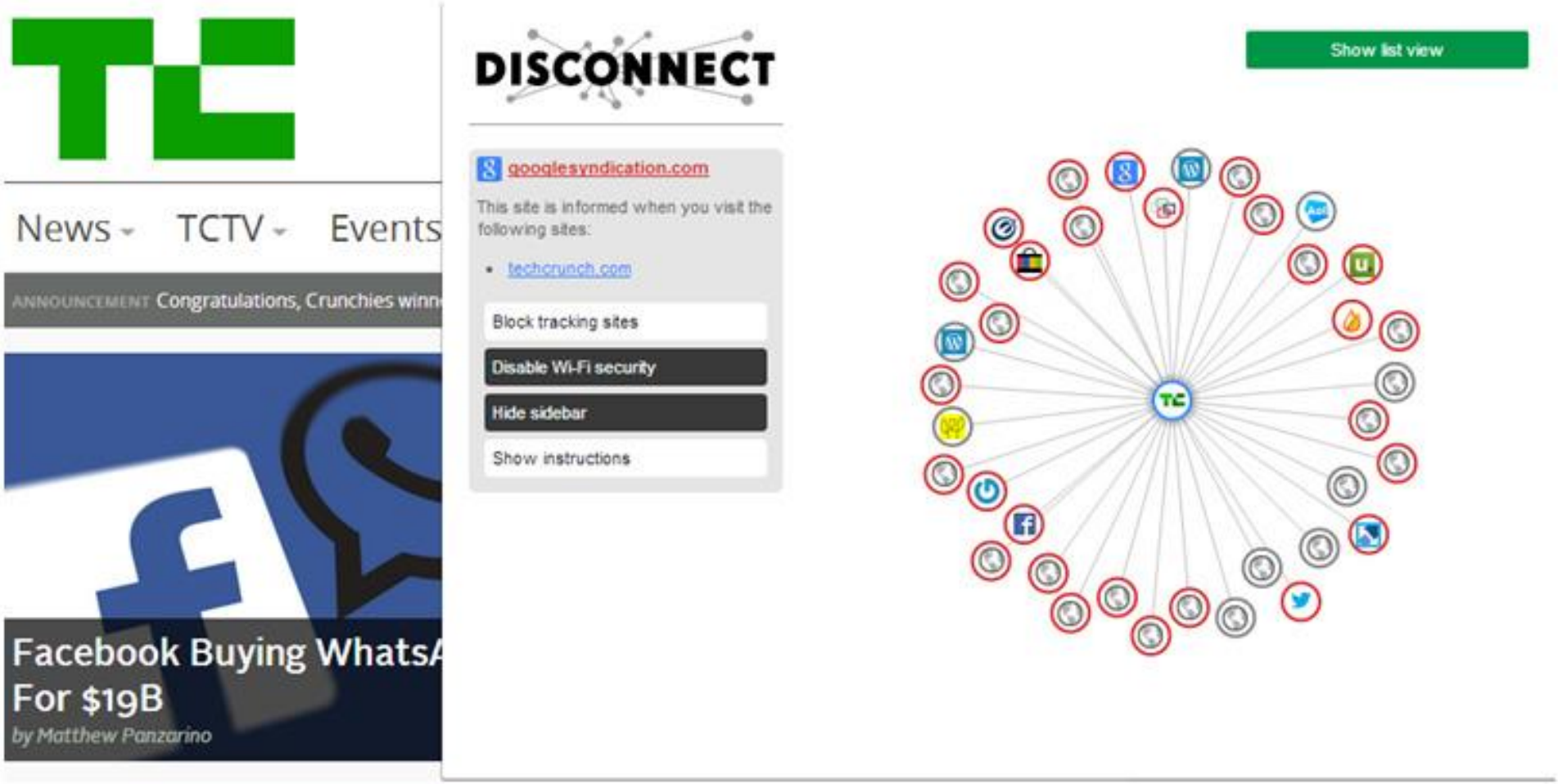
- nytimes.com

Unblock tracking sites

Hide sidebar

Show instructions

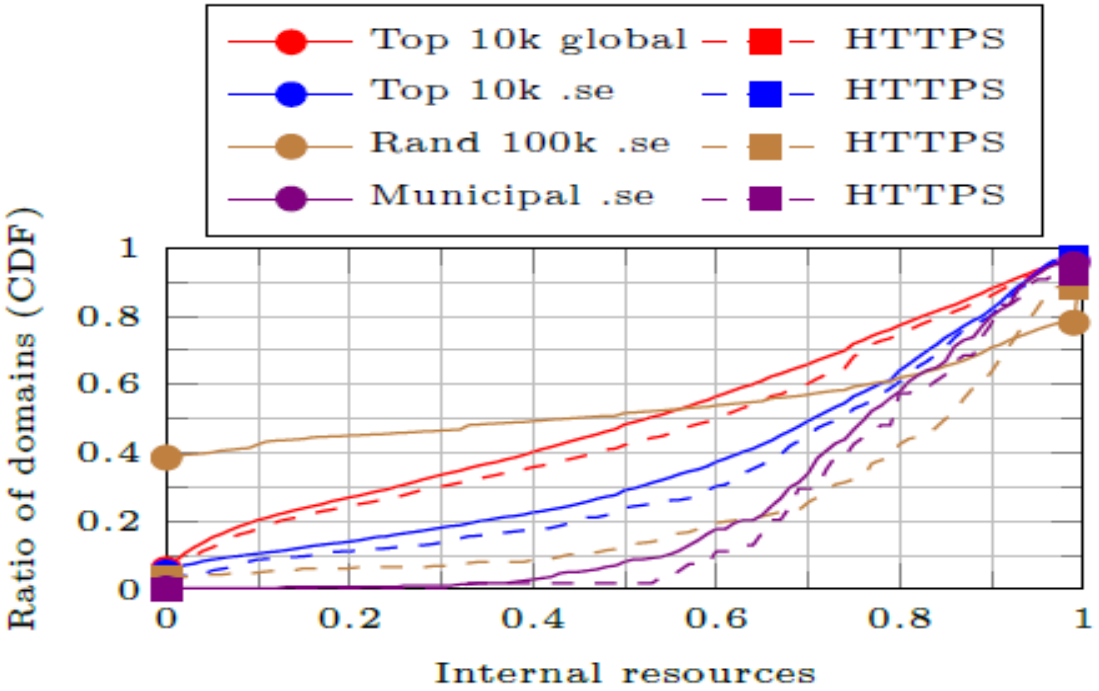
The New York Times


News
TCTV
Events
ANNOUNCEMENT Congratulations, Crunchies winn
Facebook Buying WhatsA
For $19B
by Matthew Panzarino
DISCONNECT
googlesyndication.com
This site is informed when you visit the following sites:
techcrunch.com
Block tracking sites
Disable Wi-Fi security
Hide sidebar
Show instructions
Show list view

External third-party resources

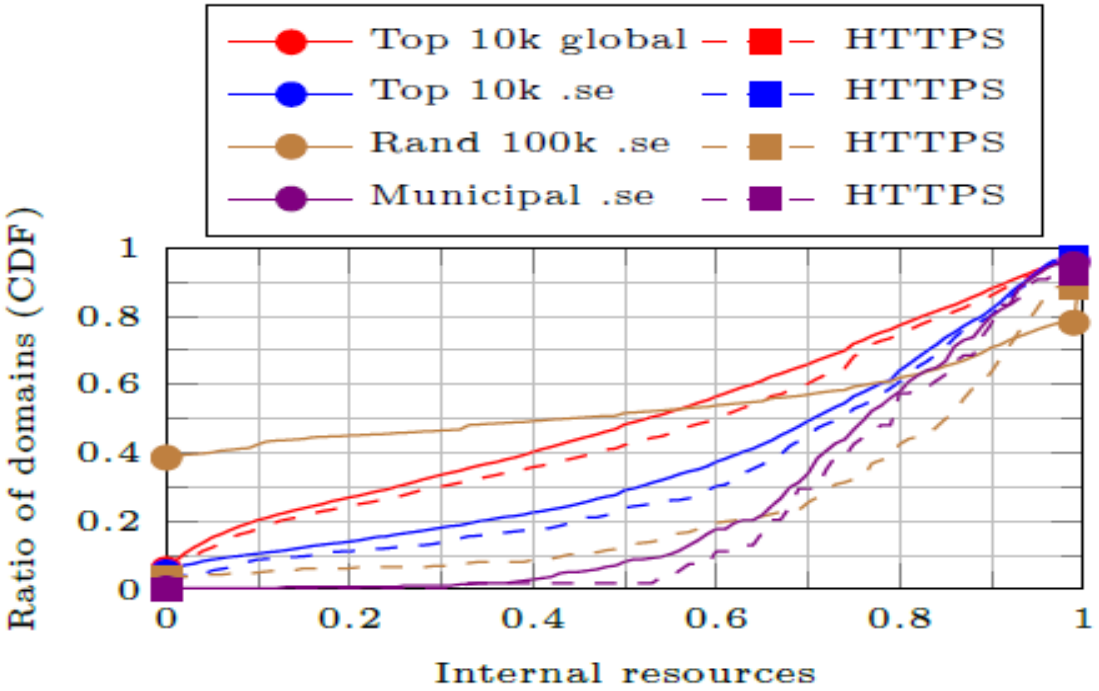
- Upper bound: Third-parties typically have server logs and/or analytics software to record your online habits
  - Each third-party (external) resource leaks at least some info

External third-party resources



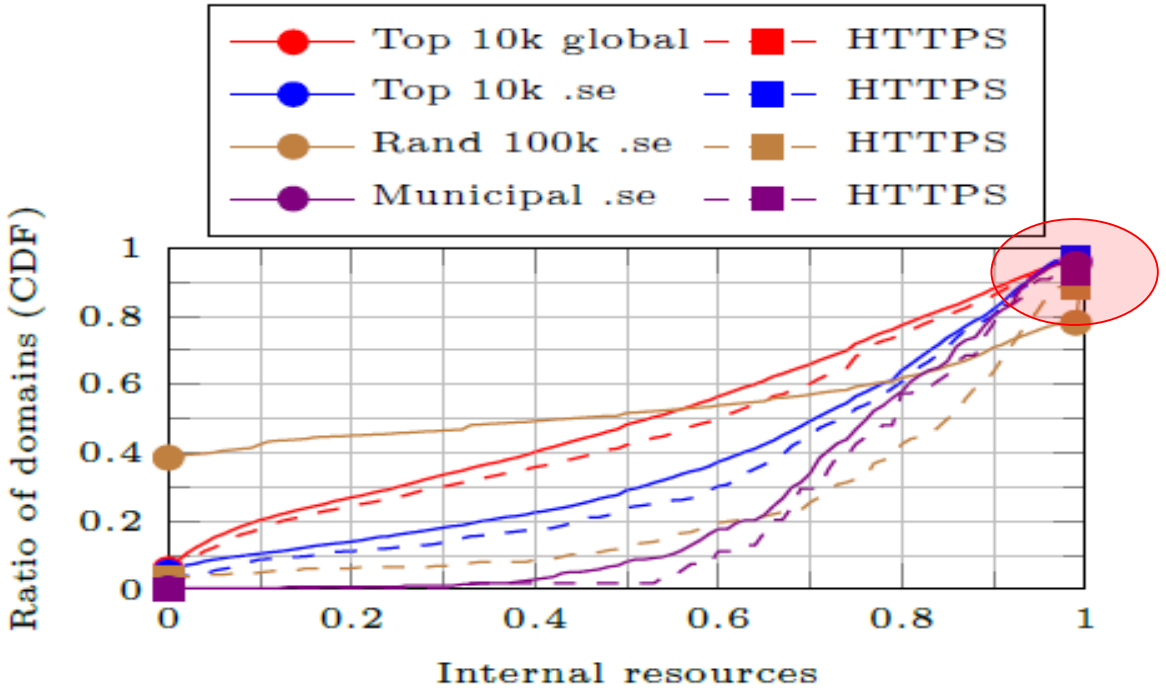
- Upper bound: Third-parties typically have server logs and/or analytics software to record your online habits
  - Each third-party (external) resource leaks at least some info

External third-party resources



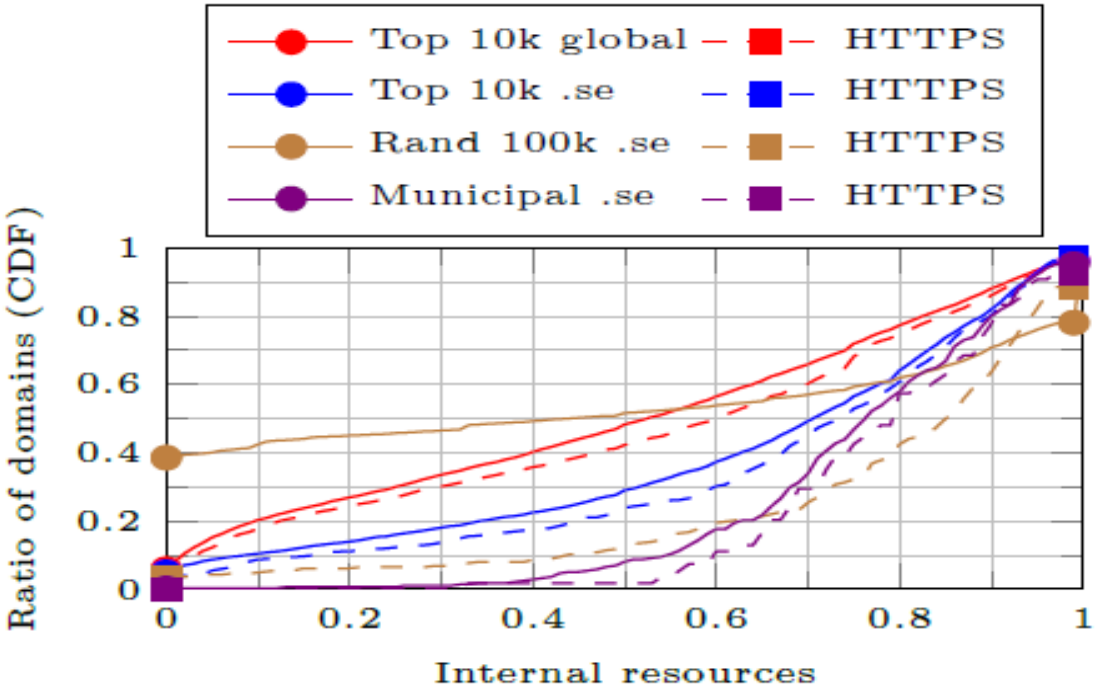
- Upper bound: Third-parties typically have server logs and/or analytics software to record your online habits
  - Each third-party (external) resource leaks at least some info
- External resource usage high
  - Especially among most popular domains (e.g., 93% at least some)

External third-party resources



- Upper bound: Third-parties typically have server logs and/or analytics software to record your online habits
  - Each third-party (external) resource leaks at least some info
- External resource usage high
  - Especially among most popular domains (e.g., 93% at least some)

External third-party resources

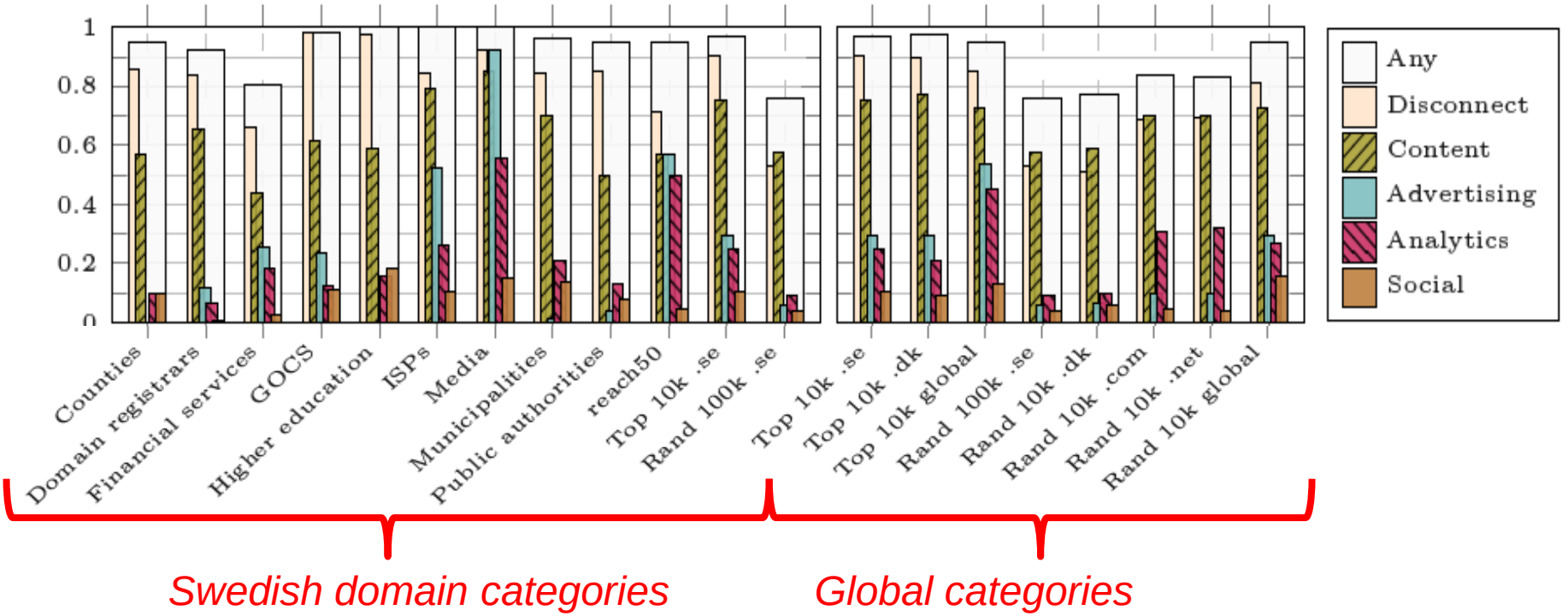


- Upper bound: Third-parties typically have server logs and/or analytics software to record your online habits
  - Each third-party (external) resource leaks at least some info
- External resource usage high
  - Especially among most popular domains (e.g., 93% at least some)
- HTTP and HTTPS results similar (except for rand 100k .se)

Known trackers

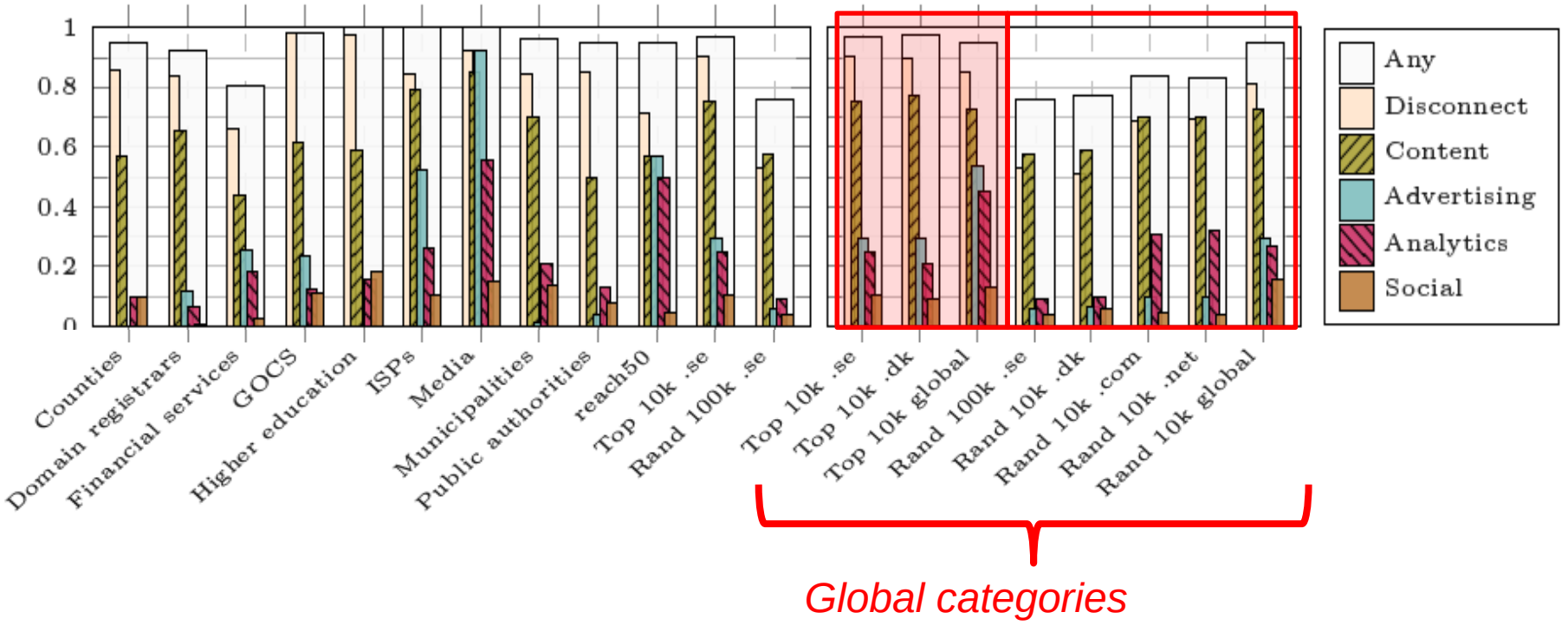
- Lower bound
  - Use Disconnect’s tracker list (2,149 known domains: resp. for <10% external)
  - Only front page

Known trackers



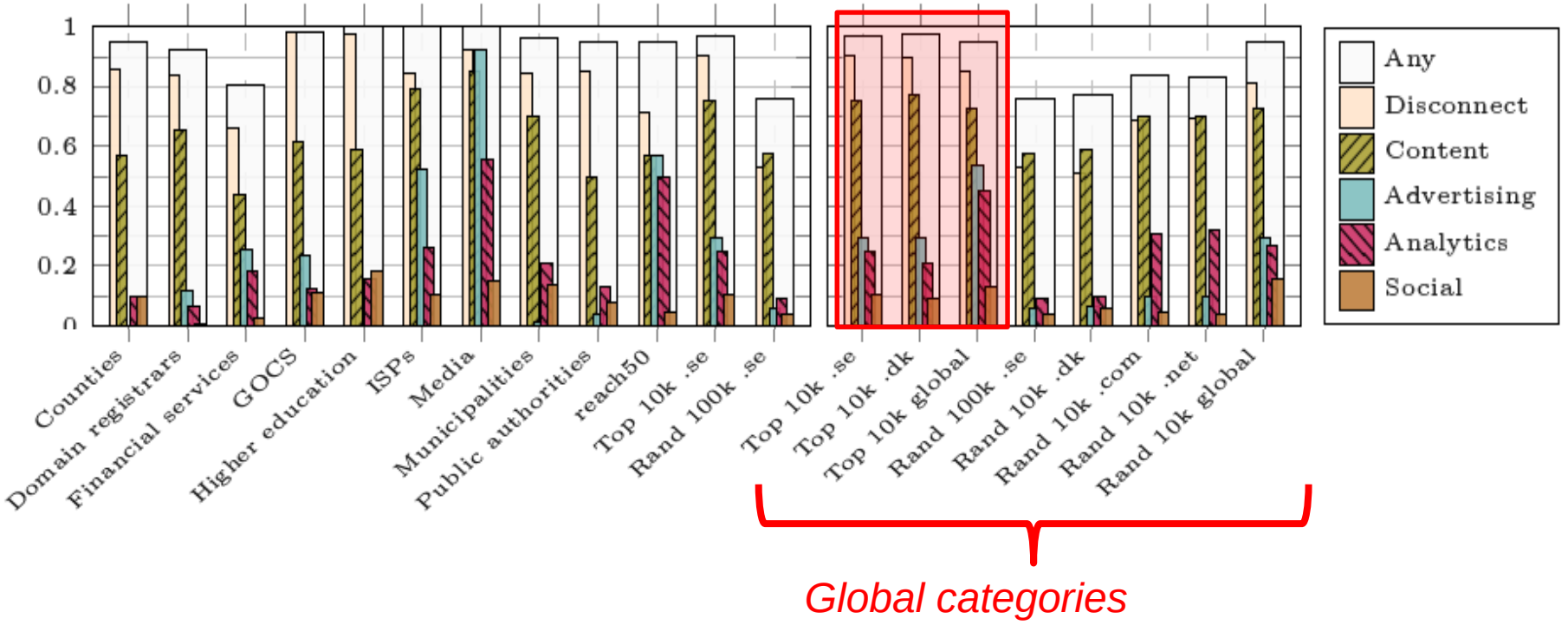
- Lower bound
  - Use Disconnect's tracker list (2,149 known domains: resp. for <10% external)
  - Only front page

Known trackers



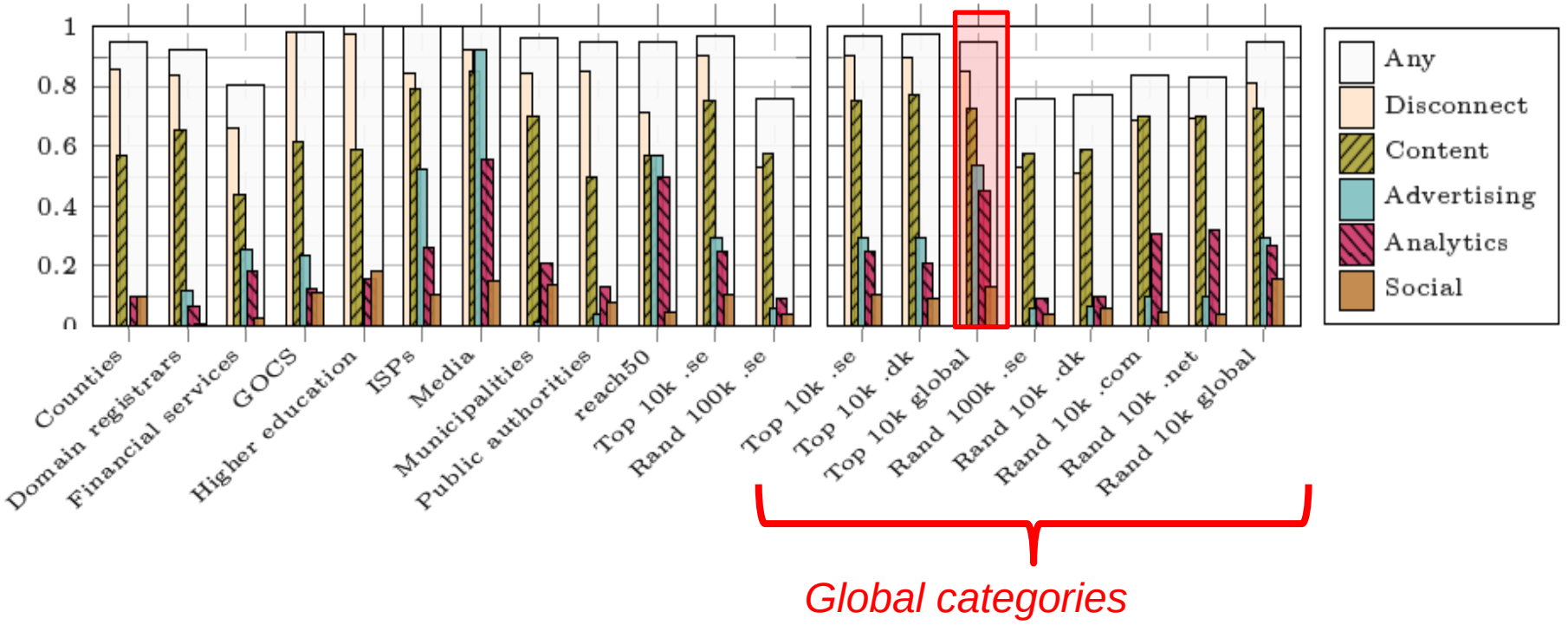
- Lower bound
  - Use Disconnect's tracker list (2,149 known domains: resp. for <10% external)
  - Only front page
- Biggest differences: popular vs. less popular (e.g., advertising)

Known trackers



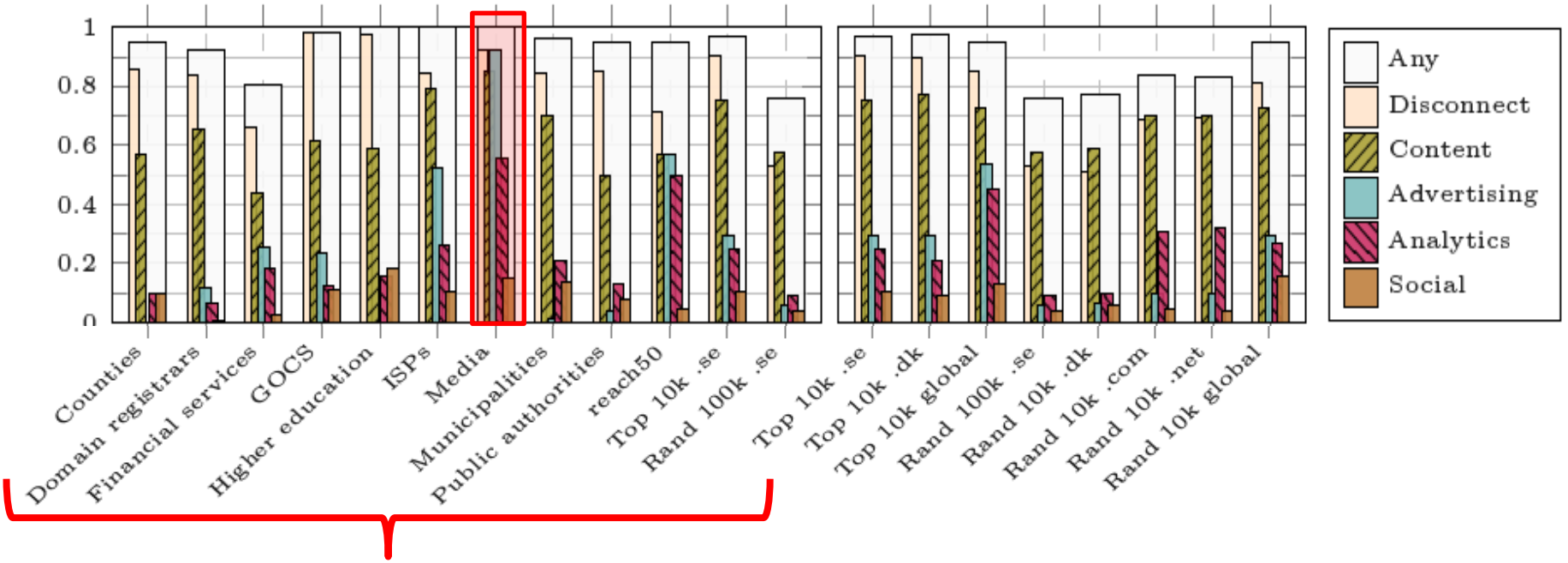
- Lower bound
  - Use Disconnect's tracker list (2,149 known domains: resp. for <10% external)
  - Only front page
- Biggest differences: popular vs. less popular (e.g., advertising)
- Popular has at least one known tracker in 95+ % of cases

Known trackers



- Lower bound
  - Use Disconnect's tracker list (2,149 known domains: resp. for <10% external)
  - Only front page
- Biggest differences: popular vs. less popular (e.g., advertising)
- Popular has at least one known tracker in 95+ % of cases
  - 70+ % use at least 2; 10% more than 12; 1% allow 48

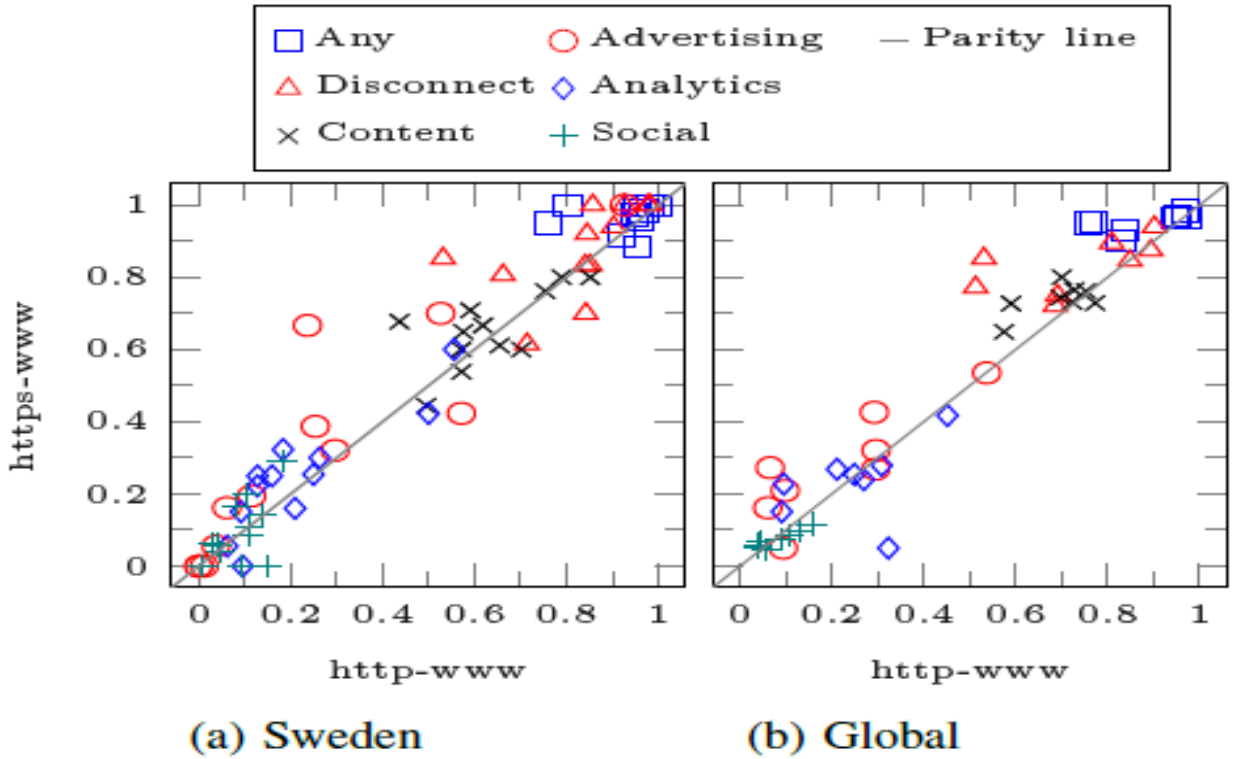
Known trackers



Swedish domain categories

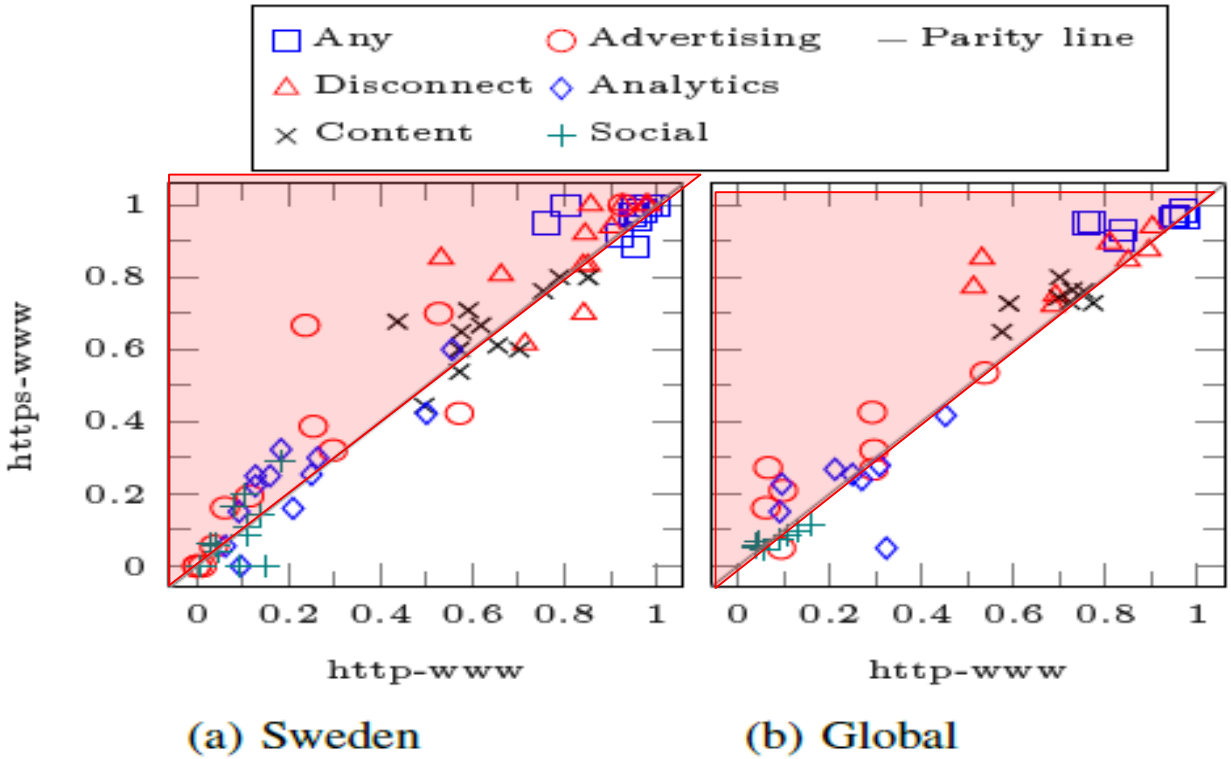
- Lower bound
  - Use Disconnect's tracker list (2,149 known domains: resp. for <10% external)
  - Only front page
- Biggest differences: popular vs. less popular (e.g., advertising)
- Popular has at least one known tracker in 95+ % of cases
  - 70+ % use at least 2; 10% more than 12; 1% allow 48
  - Other: Media worst (e.g., 50% >7 trackers); content typically not blocked …

HTTP vs HTTPS



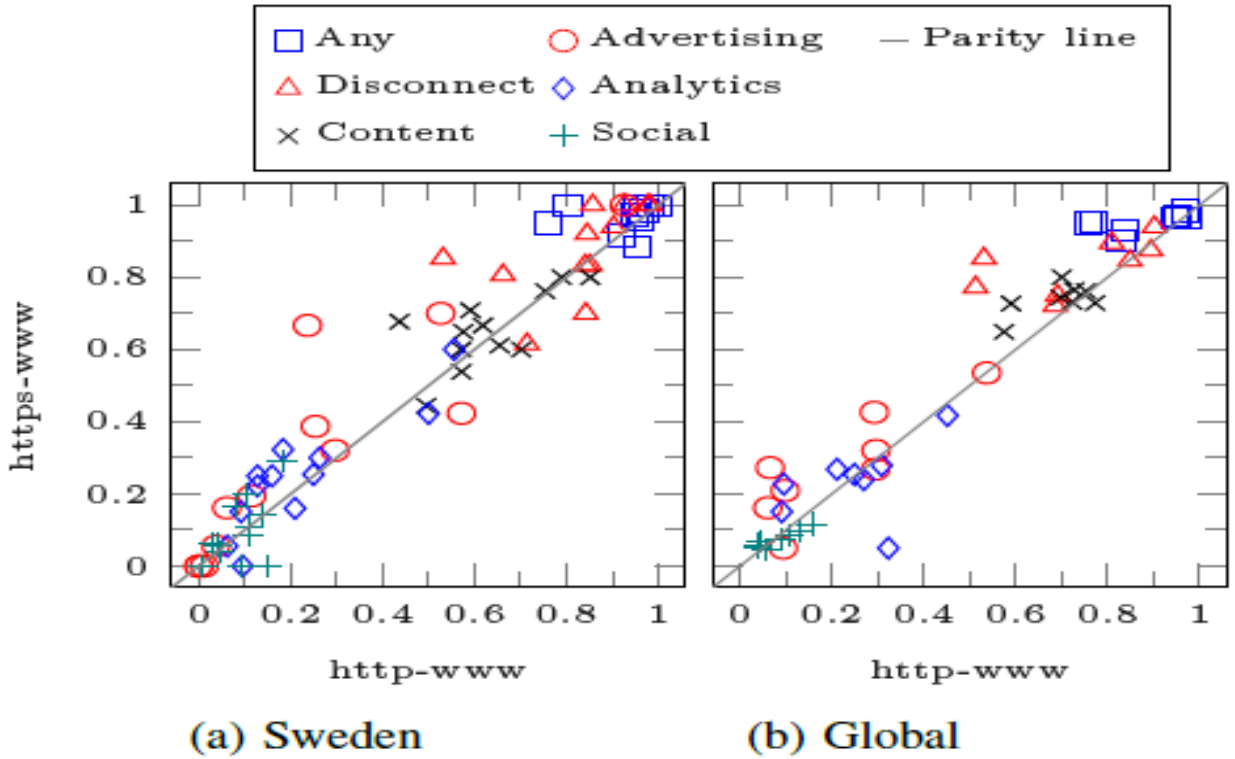
- Small differences between HTTP and HTTPS
  - If anything, slightly higher for HTTPS …

HTTP vs HTTPS



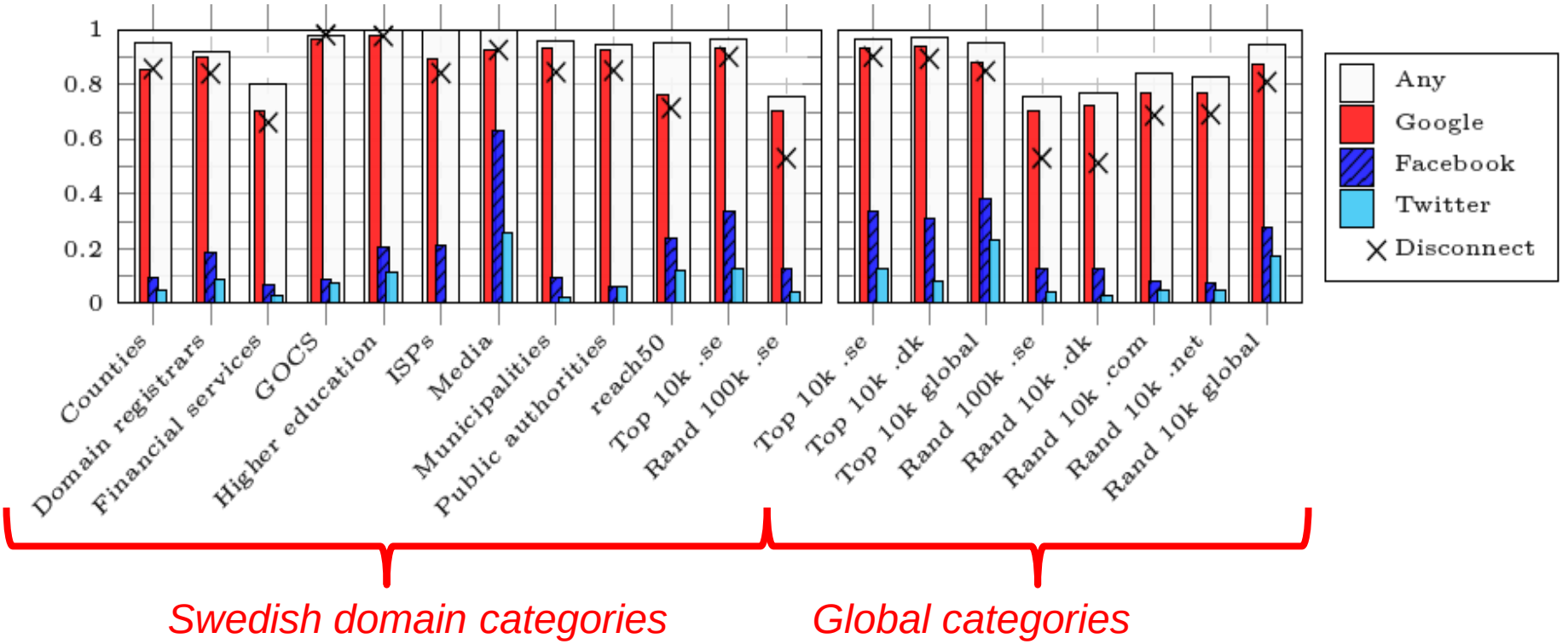
- Small differences between HTTP and HTTPS
  - If anything, slightly higher for HTTPS …

HTTP vs HTTPS



- Small differences between HTTP and HTTPS
  - If anything, slightly higher for HTTPS …
- As tracked by third-parties when using HTTPS as when using HTTP

The big players



Swedish domain categories *Global categories*

- Google has 90+ % coverage in popular domains
  - Even higher than disconnect (owns domains outside the Disconnect category)
- Facebook and Twitter far behind

Conclusions

- Measurement framework for automated, repeatable data collection of websites (tools made public)
- Analysis of the third-party tracking landscape
  - Swedish perspective vs global baseline
  - Across domain categories
  - Breakdown based on tracker types
  - HTTP and HTTPS
- HTTPS domains use at least as much (if not more) third-party tracking

Thanks for listening!

Third-party Tracking on the Web: A Swedish Perspective

LINKÖPINGS UNIVERSITET

Niklas Carlsson (niklas.carlsson@liu.se)
Research overview and pubs: www.ida.liu.se/~nikca/