

SUPERVISOR:

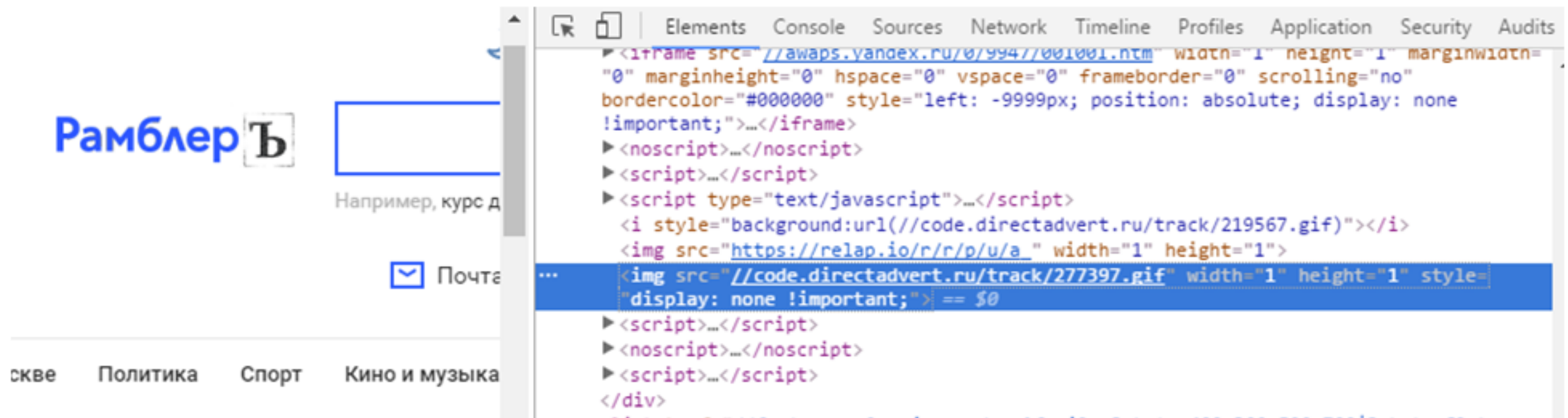
NATALIIA BIELOVA

STUDENT:

LEILA KUNTAR

USER DISCRIMINATION IN WEB

INTRODUCTION



- ▶ Tracking pixel (known as "Web beacon" or 1x1 pixel image) is used for collecting stats of users' behaviour in Web that can be used for:
 - ▶ selling aggregated information to third parties
 - ▶ targeting advertisement
 - ▶ building social graphs

EXISTING METHODS FOR PREVENTING WEB TRACKING

- ▶ Third-party cookie blocking
 - ▶ but may browser block the setting such cookies but at the same time allow their sending?
- ▶ Do not track (DNT) field in HTTP header
 - ▶ but there is no guarantee that websites and advertisers do not ignore this field
- ▶ Clearing client-side state
 - ▶ but there are fingerprinting techniques; also cookies can be duplicated in HTML Local Storage
- ▶ Miscellaneous (blocking pop-ups, private browsing mode, disabling JavaScript)

CONTRIBUTIONS

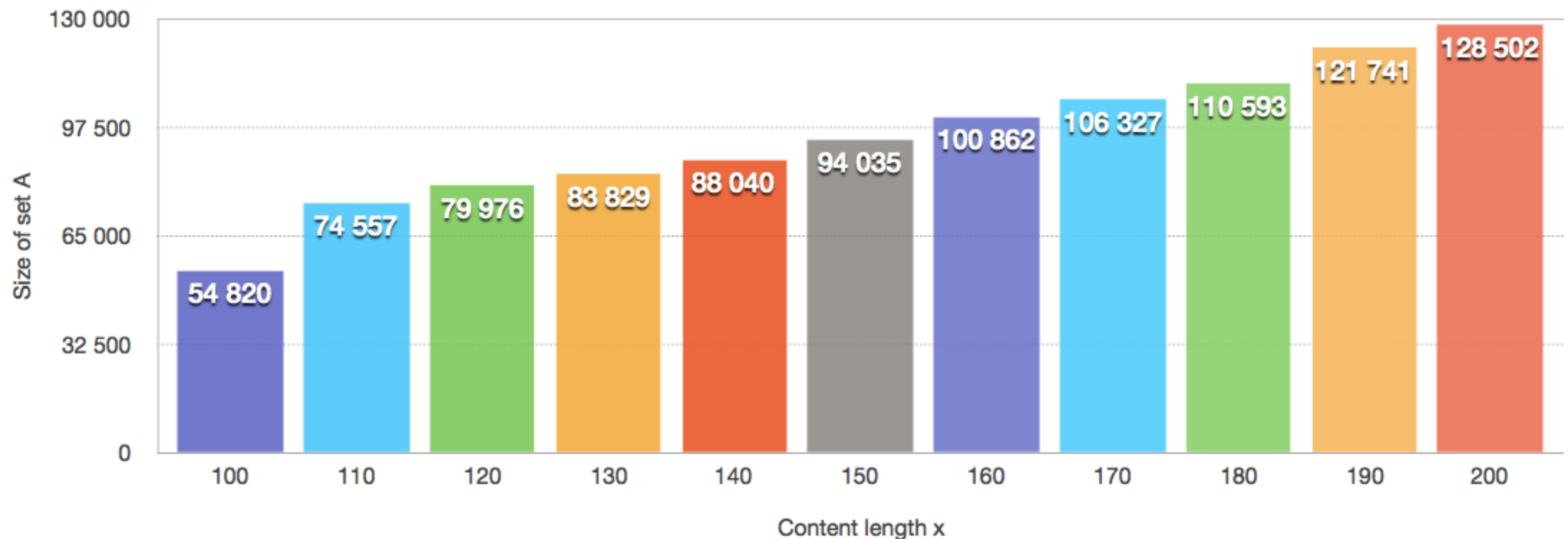
- ▶ Developing Python program that
 - ▶ parse JSON files (gathered by web crawler) with image links from top 800 domains
 - ▶ distinguish third-party URLs and fetch images
 - ▶ analyse if image is a beacon with two methods and put results in different databases
- ▶ Comparison of two methods that detect wherever image is a beacon
- ▶ Determining top providers of beacons
- ▶ Analyse the way beacons work

METHODS COMPARISON

	Method “Check the size”	Method “Check the content length”
Count of beacons	2 431 277	2 316 179
Count of 3 rd party images	5 873 372	5 830 542

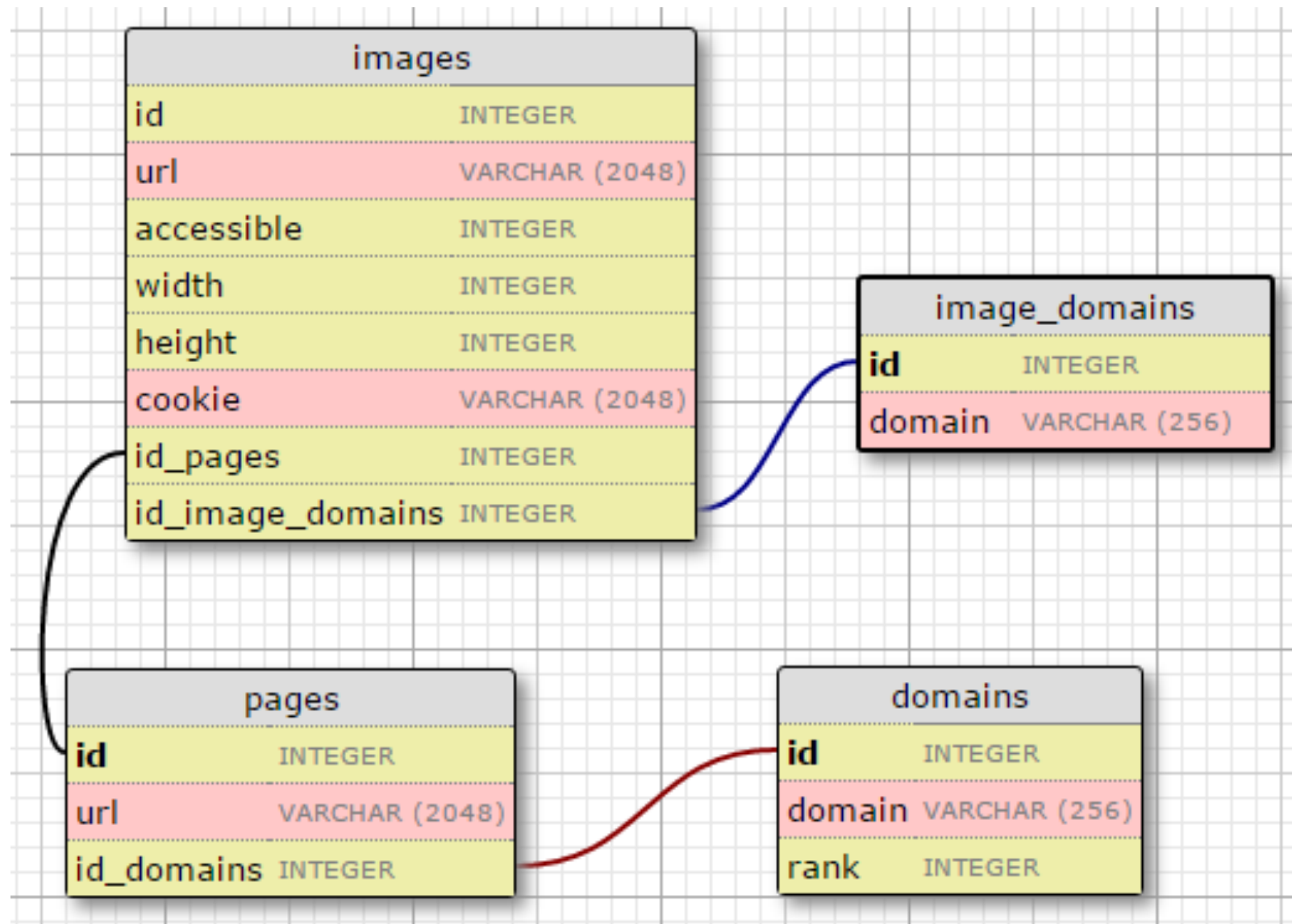
- ▶ Method “Check the content length”
 - ▶ content length is less or equal 100 bytes and content type is “IMG” (but these fields are not necessary fields in HTTP header)
- ▶ There are beacons in PNG format so content length of response is even more than 1000 bytes
- ▶ Method “Check the size” checks that image has size 1x1 px or 0x0 px (in case when server returns 204 status and content type is “IMG”)

HOW MANY BEACONS WERE FOUND BY METHOD “CHECK THE CONTENT LENGTH” BUT NOT BY METHOD “CHECK THE IMAGE SIZE”
DEPENDING ON CONTENT LENGTH UPPER BOUND



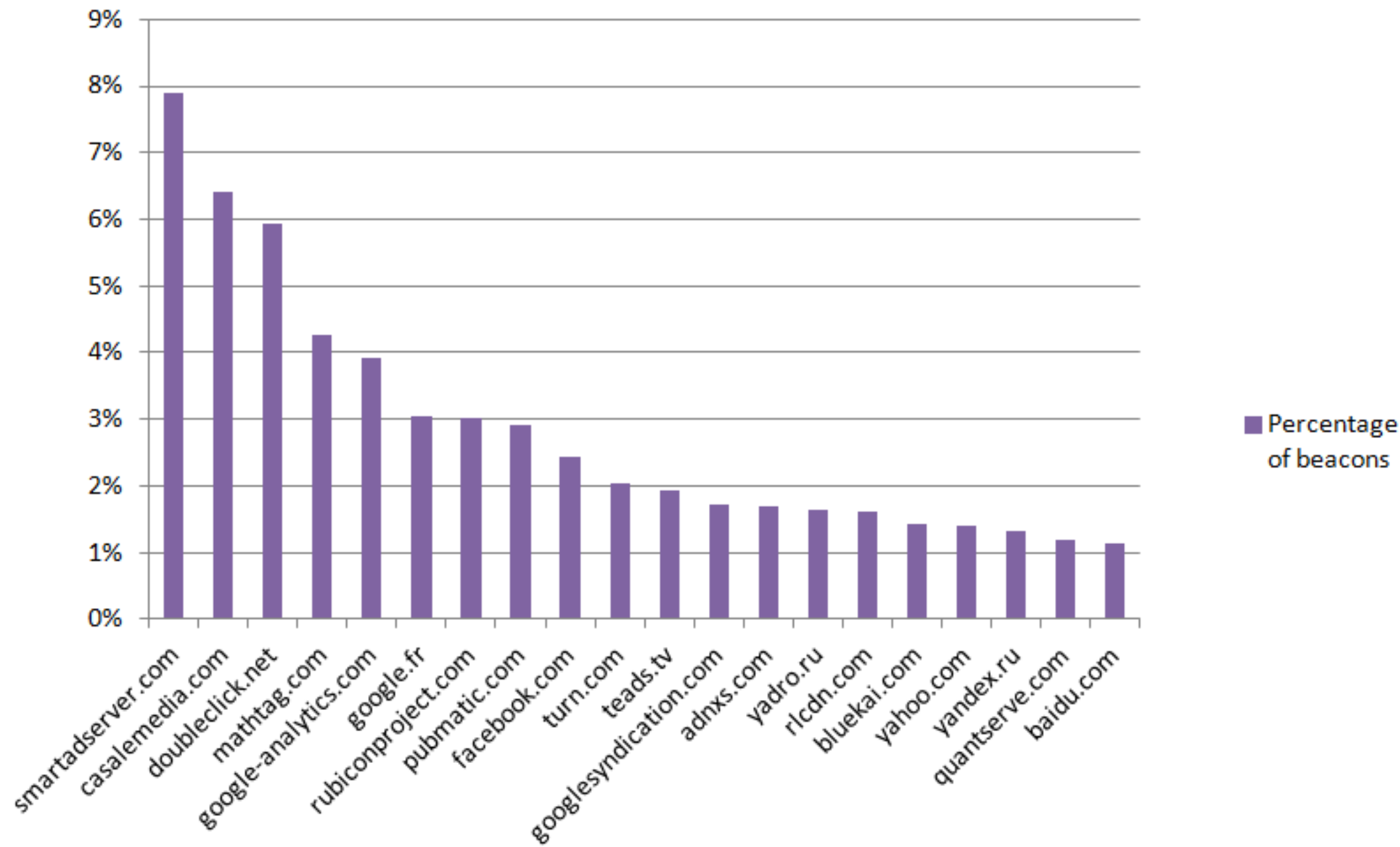
- ▶ Compare the sets of 1 million beacons in order to find beacons that are presented in one database and are absent in second database
- ▶ With x increasing the symmetric difference of two methods also increases

DATABASE STATISTICS

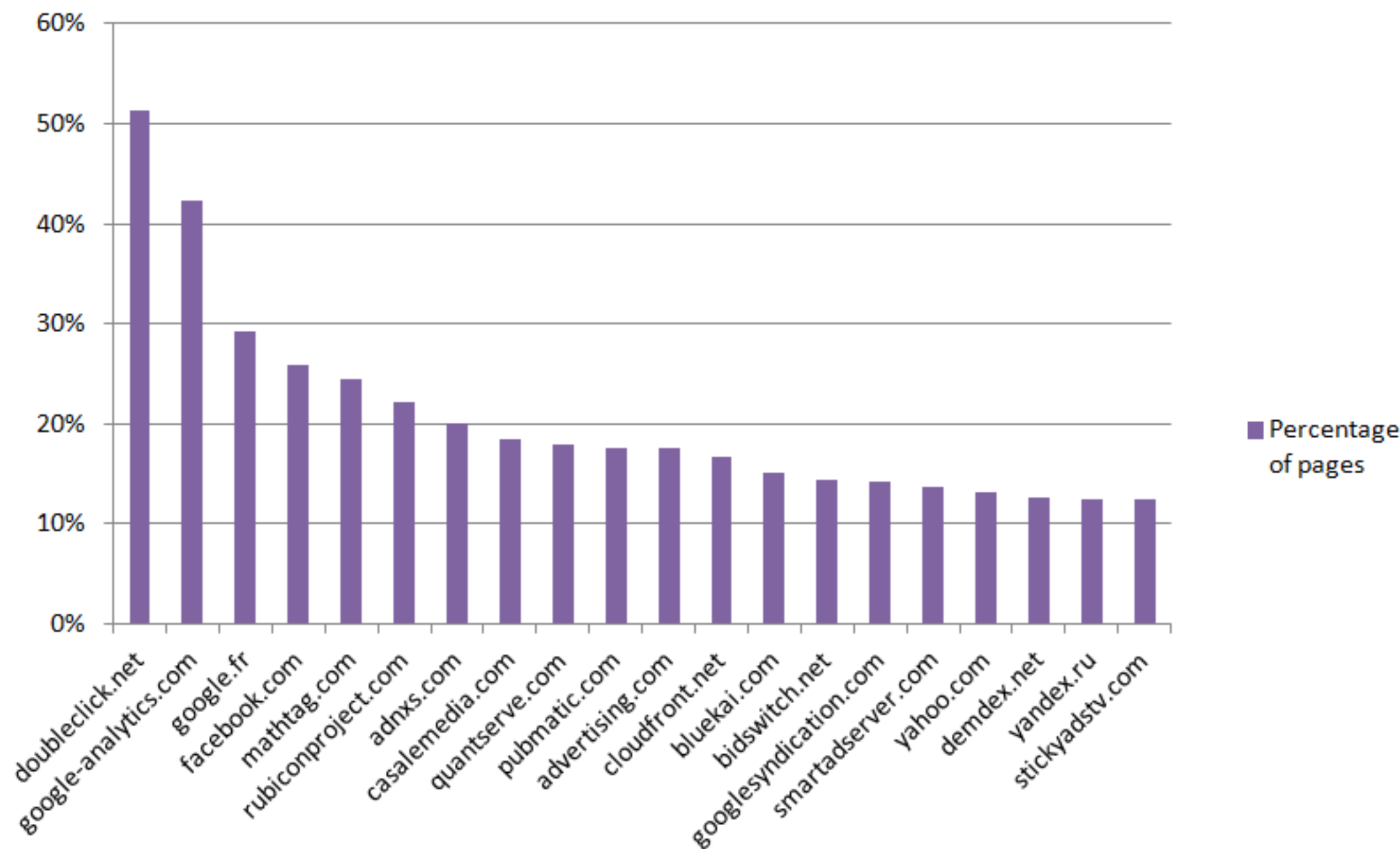


- ▶ Count of domains is 800 (only 5% without beacons)
- ▶ Count of pages is 124 214 (89.7% contain beacons)
- ▶ Count of third-party images in database is 5 873 372
- ▶ Count of beacons is 2 431 277 (41%)
- ▶ Count of unique image domains is 4 348 (40.3% provide beacons)

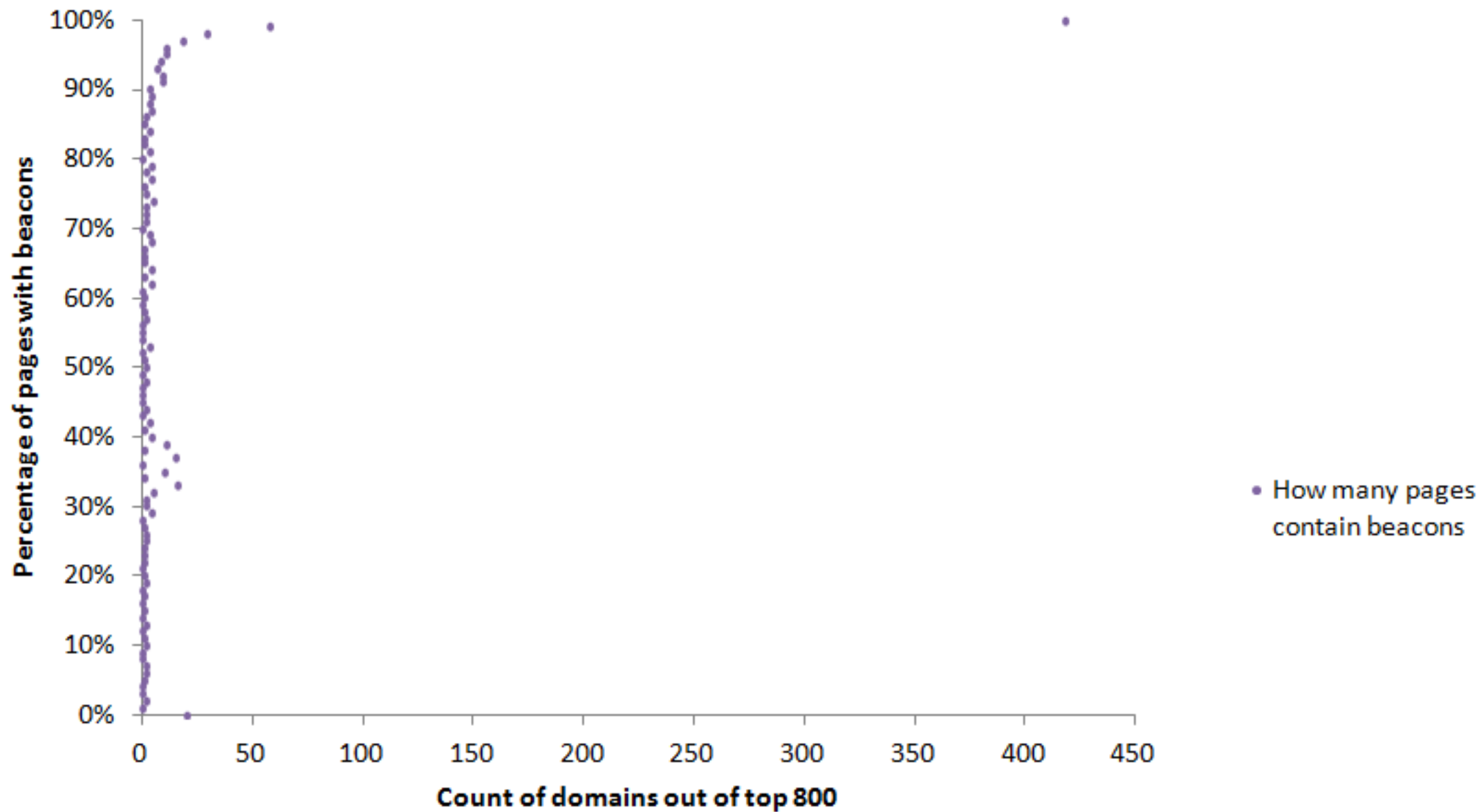
HOW MANY BEACONS OUT OF 2 431 277 ARE DELIVERED BY PROVIDERS OF BEACONS



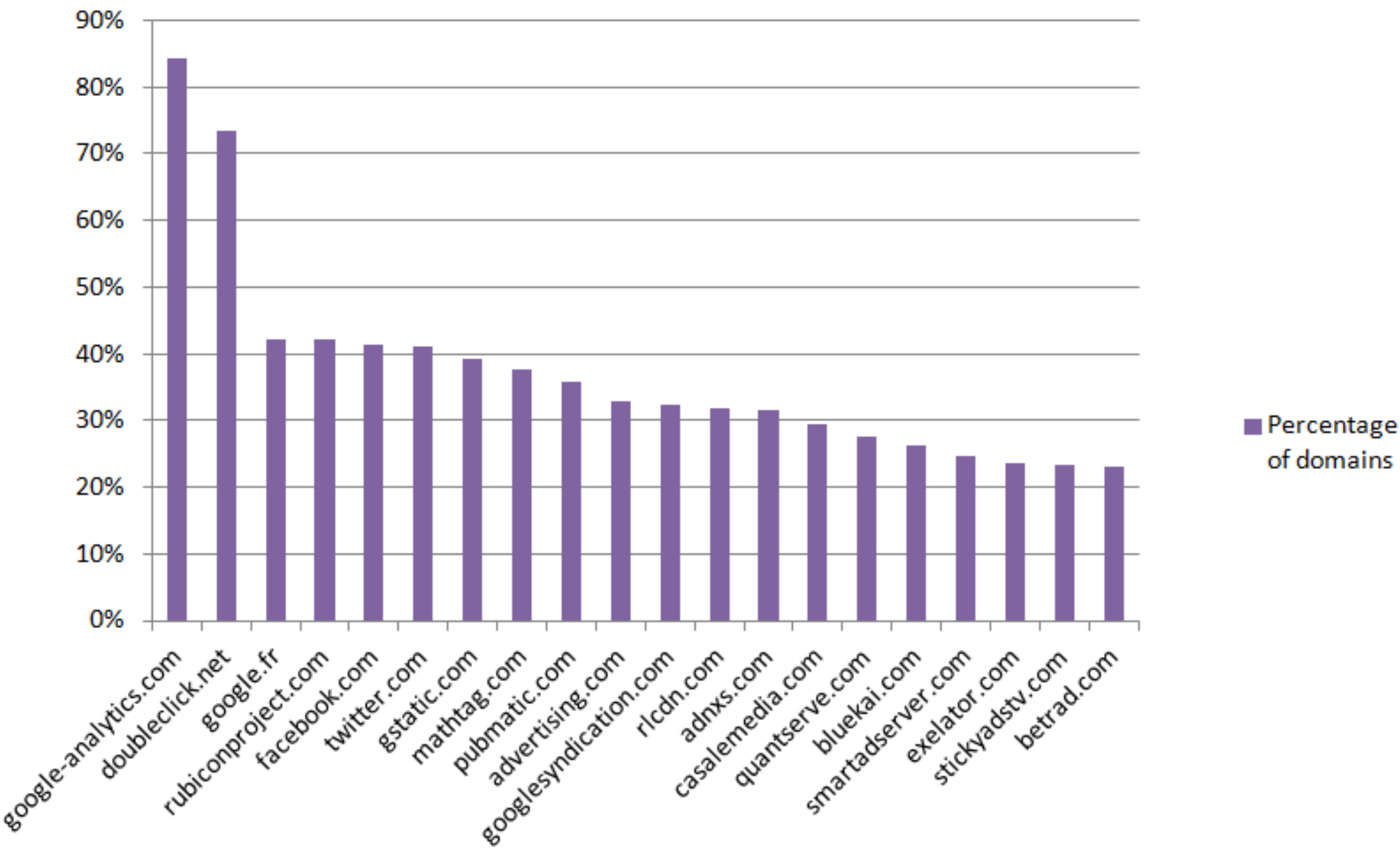
HOW MANY UNIQUE PAGES ARE SERVED BY PROVIDERS OF TRACKING PIXELS (OUT OF 124 214 PAGES)



COUNT OF DOMAINS (OUT OF TOP 800) AND THE PERCENTAGE OF PAGES THAT HAVE AT LEAST ONE BEACON (RATIO TO THE TOTAL COUNT OF PAGES THAT HAVE THIRD-PARTY IMAGES)

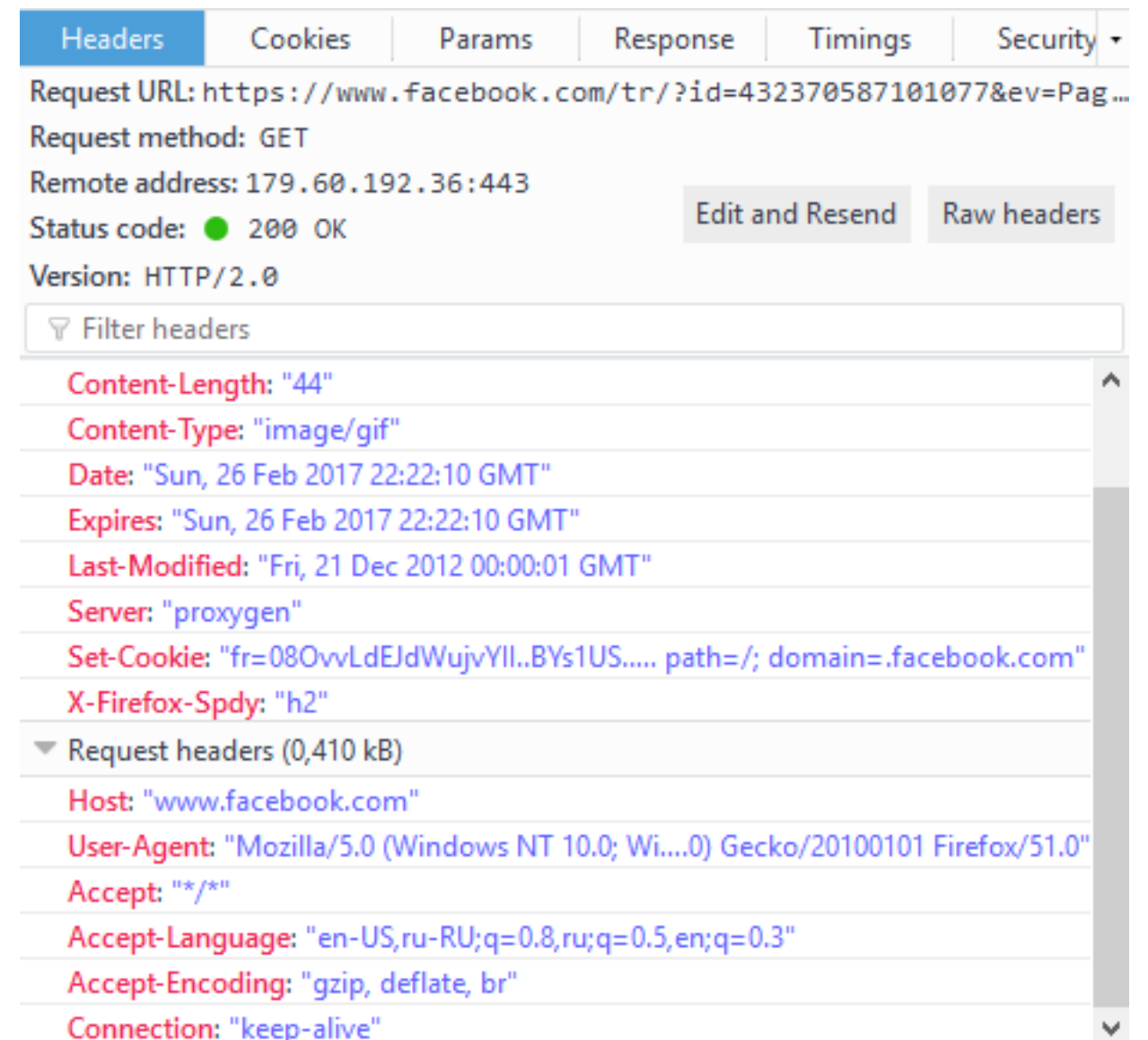


HOW MANY UNIQUE DOMAINS (OUT OF 800) ARE SERVED BY PROVIDERS OF TRACKING PIXELS



FACEBOOK PIXEL

- ▶ Everyone can create his own Facebook pixel in Adverts Manager
- ▶ A small piece of code embedded into the header code of page
- ▶ 41.25% of domains contains on their pages such beacons
- ▶ 751 unique Facebook pixels that were found on 59 023 pages in database



Headers Cookies Params Response Timings Security ▾

Request URL: `https://www.facebook.com/tr/?id=432370587101077&ev=Pag...`

Request method: GET

Remote address: 179.60.192.36:443

Status code: ● 200 OK Edit and Resend Raw headers

Version: HTTP/2.0

Filter headers

- Content-Length: "44"
- Content-Type: "image/gif"
- Date: "Sun, 26 Feb 2017 22:22:10 GMT"
- Expires: "Sun, 26 Feb 2017 22:22:10 GMT"
- Last-Modified: "Fri, 21 Dec 2012 00:00:01 GMT"
- Server: "proxygen"
- Set-Cookie: "fr=08OvvLdEJdWujvYII..BYs1US..... path=/; domain=.facebook.com"
- X-Firefox-Spdy: "h2"

▼ Request headers (0,410 kB)

- Host: "www.facebook.com"
- User-Agent: "Mozilla/5.0 (Windows NT 10.0; Wi....0) Gecko/20100101 Firefox/51.0"
- Accept: "*/*"
- Accept-Language: "en-US,ru-RU;q=0.8,ru;q=0.5,en;q=0.3"
- Accept-Encoding: "gzip, deflate, br"
- Connection: "keep-alive"

Status	Method	File	Domain	Cause	Type	Trans	Headers	Cookies	Params
▲ 304	GET	fbevents.js	connect.facebook.net	JS script	js	3,9	Filter cookies		
▲ 304	GET	fbevents.plugins.identity.js	connect.facebook.net	JS script	js	3,5	Request cookies		
● 200	GET	/tr/?id=432370587101077&ev=PageView&dl=fil...	www.facebook.com	JS img	gif	4	fr: "08OvvLdEJdWujvYII..BYs1US...1.0.BYs		
● 200	GET	/tr/?id=432370587101077&ev=post&dl=file:///...	www.facebook.com	JS img	gif	4			

Site

Cookie Name

▼ facebook.com

facebook.com

fr

Name: fr

Content: 08OvvLdEJdWujvYII..BYs1US...1.0.BYs1US.

Domain: .facebook.com

Path: /

Send For: Any type of connection

Surfing a page with Facebook pixel before authorisation in Facebook

Method	File	Domain	Cause	Type	Trans	Headers
GET	/tr/?id=432370587101077&ev=PageView&dl=fil...	www.facebook.com	JS img	gif	4	Filter cookies
GET	/tr/?id=432370587101077&ev=post&dl=file:///...	www.facebook.com	JS img	gif	4	Request cookies
GET	fbevents.js	connect.facebook.net	JS script	js	3,9	c_user: "10000209
GET	fbevents.plugins.identity.js	connect.facebook.net	JS script	js	3,5	csm: "2"
POST	/tr/	www.facebook.com	JS subdocument	gif	4	datr: "tFWzWF3L

Site	Cookie Name
facebook.com	sb
facebook.com	c_user
facebook.com	xs
facebook.com	csm
facebook.com	pl
facebook.com	lu
facebook.com	wd
facebook.com	presence
facebook.com	p

Name: csm

Content: 2

Domain: .facebook.com

Path: /

Send For: Any type of connection

Surfing a page with Facebook pixel after authorisation in Facebook

GOOGLE BEACON

Surfing www.elevenia.co.id

Before authorisation in Gmail

Status	Method	File	Domain	Cause	Type	Trans	Headers	Cookies
200	GET	collective=1&v=j46&a=2111443680&t=event&...	www.google-analytics.com	img	gif			
200	GET	tap.php?v=4894&nid=1986&put=154964915612...	pixel.rubiconproject.com	img	gif			
200	GET	tap.php?v=7751&nid=2249&expires=30&put=...	pixel.rubiconproject.com	img	gif			
200	GET	/ads/user-lists/960475416/?label=MCC remarke...	www.google.fr	img	gif			

After authorisation in Gmail

Status	Method	File	Domain	Cause	Type	Trans	Headers	Cookies
200	GET	collective=1&v=j46&a=2111443680&t=event&...	www.google-analytics.com	img	gif			
200	GET	xuid?id=1&mid=3335&xuid=1549649156127290...	eb2.3lift.com	img	gif			
200	GET	/ads/user-lists/960475416/?fmt=3&num=1&cv...	www.google.fr	img	gif			
200	GET	/ads/user-lists/960475416/?label=MCC remarke...	www.google.fr	img	gif			
200	GET	/ads/user-lists/937736793/?fmt=3&num=1&cv...	www.google.fr	img	gif			
200	GET	/ads/user-lists/960195341/?fmt=3&num=1&cv...	www.google.fr	img	gif			
200	GET	tap.php?v=7751&nid=2249&expires=30&put=...	pixel.rubiconproject.com	img	gif			

No cookies for this request

Filter cookies

Request cookies

APISID: "j45wBlu

HSID: "AAMe3D

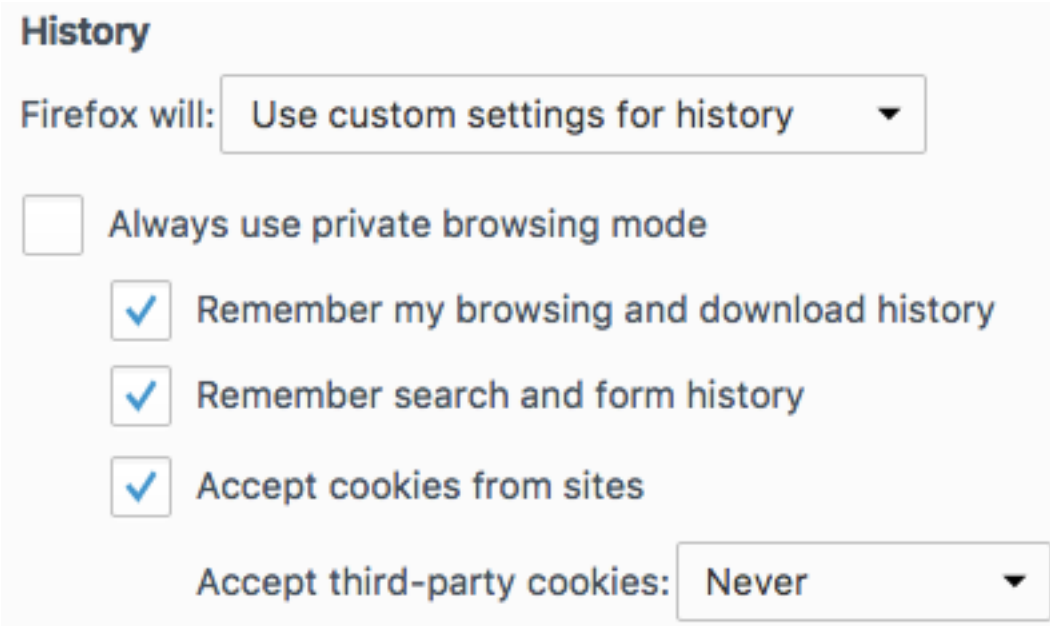
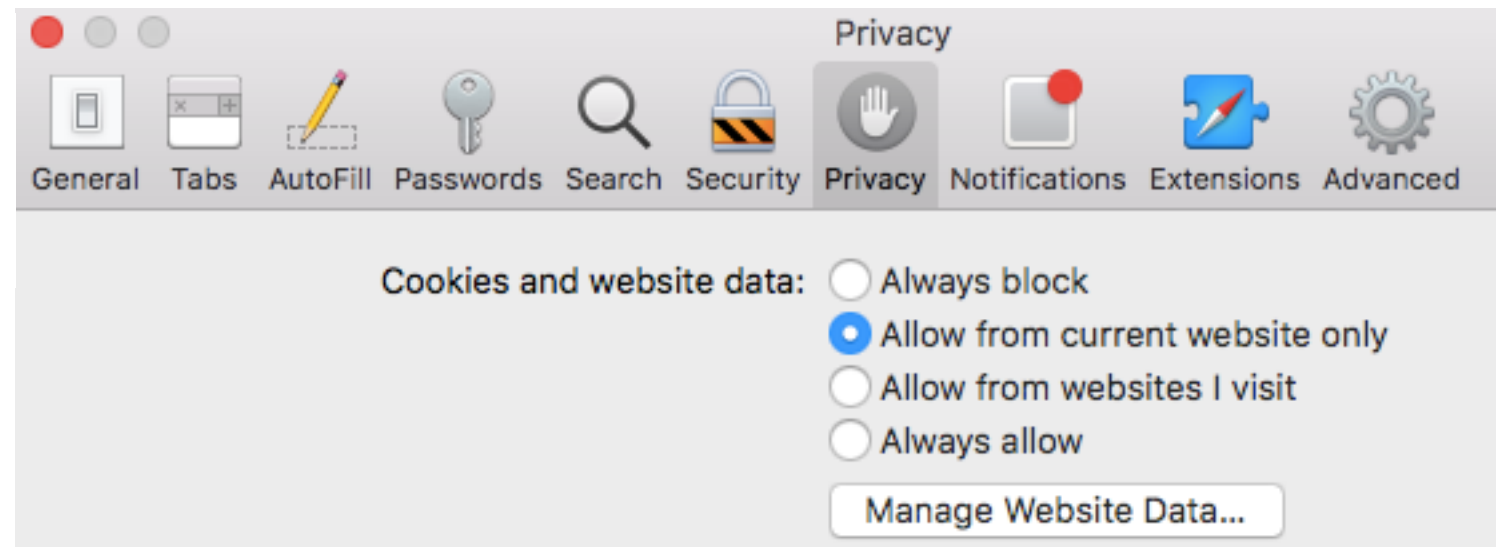
NID: "97=tXVIYR

SAPISID: "Z6s11

SID: "YgTZVKL0c

SSID: "AyG-KrtV

BLOCK THIRD-PARTY COOKIES



Content settings

Cookies



- ☐ Allow local data to be set (recommended)
- ☒ Keep local data only until you quit your browser
- ☐ Block sites from setting any data
- ☒ Block third-party cookies and site data

[Manage exceptions...](#)[All cookies and site data...](#)

CONCLUSION

- ▶ Web tracking via beacons is a widely spread form of tracking nowadays
- ▶ If such tracking is not anonymous it infringes users' privacy
- ▶ Beacons may have non standard sizes (e.g. 2x2, 1x2)
- ▶ Disabling JavaScript doesn't protect against tracking
- ▶ Modern browsers have option "Don't accept 3rd party cookies" and it work appropriately
 - ▶ but websites may keep users' identifiers in HTML Local Storage

STILL TRUE?



on the internet, no one knows you're a cat.

ROFLCAT.COM