

Sigurnosne prijetnje na Internetu

Povijest *web trackera*

Nikola Bukvić, 18.12.2024.

Pregled predavanja

- Motivacija
- Osnovne definicije
- Vrste *web trackera*
- Razlozi teškom praćenju povijesti *web trackera*
- Početak *web trackera*
- Razvoj *web trackera*
 - Različite statistike te razlozi
- Regulative koje su danas implementirane

Pitanja za ispite

- Koja je razlika između *first-party* i *third-party* web *trackera*?
- Navedi 3 vrste web *trackera* po ulozi i ukratko objasni što su.
- Zašto ima malo informacija o povijesti web *trackera*?
- Zašto je *Google Analytics* bio toliko dominantan među ostalim web *trackerima*?
- Zašto web *trackerima* raste međusobna povezanost kroz godine?

Motivacija

- Važnost osobnih podataka i privatnosti
- *Web trackeri* mogu biti zloupotrijebljeni, zbog čega je važno znati što su *web trackeri* i u koje se svrhe mogu upotrebljavati
- Potrebna diskusija o regulativama koje kontroliraju kako se *web trackeri* smiju koristiti
- Najvažnije: manjak arhiviranja povijesti *web trackera*, mali broj izvora o povijesti *web trackera* i Interneta

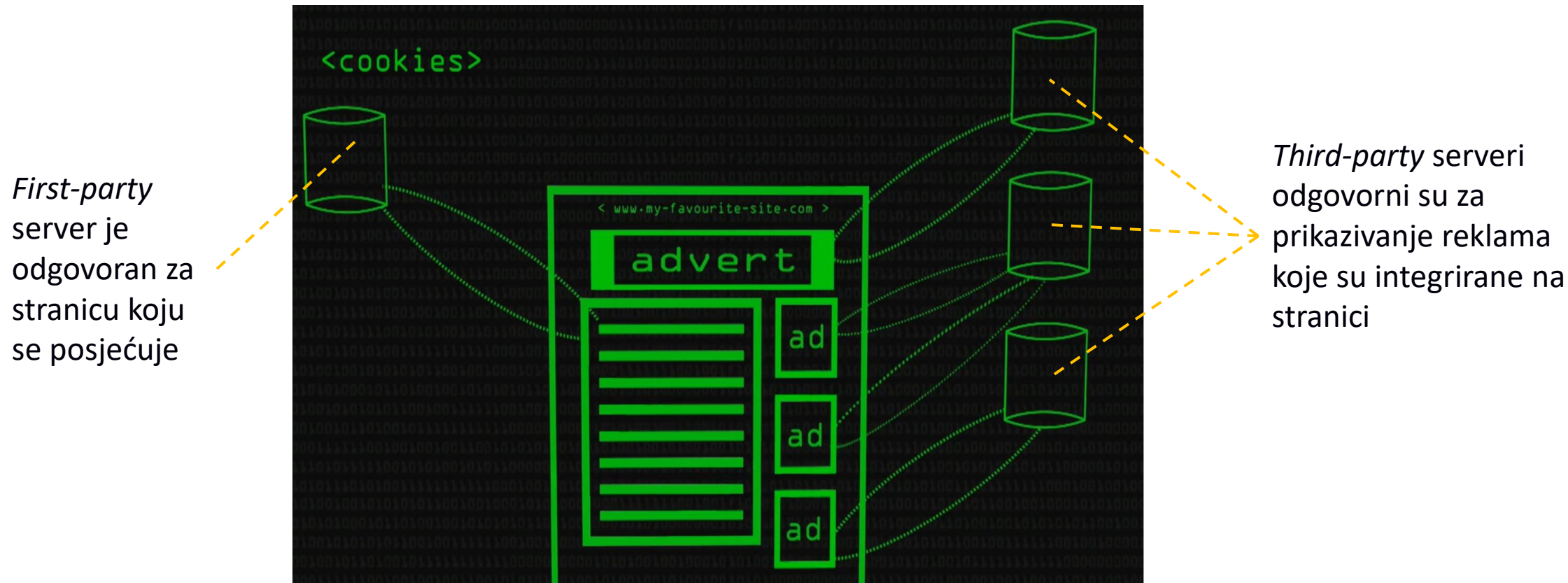
Glavne definicije

- *Web tracker*:
 - Alat ili skripta koju web stranice koriste za praćenje aktivnosti korisnika kako bi prikupili podatke o njihovom ponašanju
- Vrste:
 - HTTP kolačići (engl. *cookies*) treće strane, super kolačići, ugrađene skripte, digitalni otisak (engl. *fingerprinting*), praćenje kroz lokalnu pohranu

Web trackers (tehnike)

- HTTP kolačići:
 - Mala tekstualna datoteka koja sprema web stranica na korisnikov uređaj (najčešće preglednik)
 - Kolačići prve strane (engl. *first-party*)
 - Postavljaju ih web stranice koje korisnik posjećuje, tj. one na čiju je adresu korisnik otišao u pregledniku
 - Kolačići treće strane (engl. *third party*)
 - Postavljaju ih druge web stranice ili usluge koje su integrirane u stranicu koju korisnik pregledava, koje korisnik ne posjećuje izravno

Web trackers (tehnike)



Web trackers (tehnike)

- Ugrađene skripte:
 - Skripte koje se učitavaju unutar web stranica, najčešće pomoću `<script>` oznake u HTML datoteci
 - Skripte mogu koristiti Web API kao alat za dohvat podataka
 - Primjer: Skripta čita *navigator.userAgent* pomoću JavaScript API-a kako bi se dohvatila vrsta preglednika
- Digitalni otisak:
 - Tehnika koja prikuplja podatke o korisnikovom uređaju kroz različite web stranice da se uspije identificirati korisnika

Web trackers (vrste po ulozi)

- Analitičko praćenje:
 - Skripta za web analitiku koja se izvršava u kontekstu prve strane
 - Postavlja *first-party* kolačiće, ali kasnije ih prosljeđuje trećoj strani
- Osnovno praćenje:
 - Praćenje je uključeno kao treća strana (npr. *iframe*) na glavnoj stranici i koristi kolačiće treće strane za praćenje korisnika na različitim web stranicama
- Prisilno praćenje:
 - Korisnik je prisiljen posjetiti trackerovu domenu direktno (npr. putem pop-up prozora ili redirekcije), čime se dopušta postavljanje *first-party* kolačiće

Web trackers (vrste po ulozi)

- Preneseno praćenje:
 - *Tracker* ne postavlja vlastite kolačiće, već se oslanja na drugi *tracker* koji mu prosljeđuje identifikatore
- Osobno praćenje:
 - Slično osnovno praćenju, ali korisnici direktno posjećuju stranicu *trackera* u drugim kontekstima
 - Često uključuje društvene *widgete* (npr. „*Like*” i „*Tweet*” gumbi)
- Preneseno analitičko praćenje:
 - Kombinacija analitičkog i prenesenog praćenja
 - Podaci se postavljaju na jednoj domeni, ali se prosljeđuju na drugoj domeni

Poteškoće istraživanja *web trackinga*

- Nepotpuni podaci:
 - Rane web stranice nisu sačuvane ili *third-party* skripte nedostaju
- Dinamički sadržaj:
 - *Trackeri* se mijenjaju svaki put kad korisnik posjeti stranicu
- Blokiranje arhiviranja:
 - Mnoge stranice sprječavaju *crawlanje* pomoću *robots.txt* datoteke
- Evolucija tehnika:
 - Napredne metode poput *fingerprintinga* teže je detektirati

Poteškoće istraživanja *web trackinga*

- Najbolji pristup problemu – *Wayback Machine*:
 - Alat unutar *Internet Archivea* koji omogućava da se pregledaju povijesne verzije web stranice od kada je prvi put pokrenut, 1996. godine

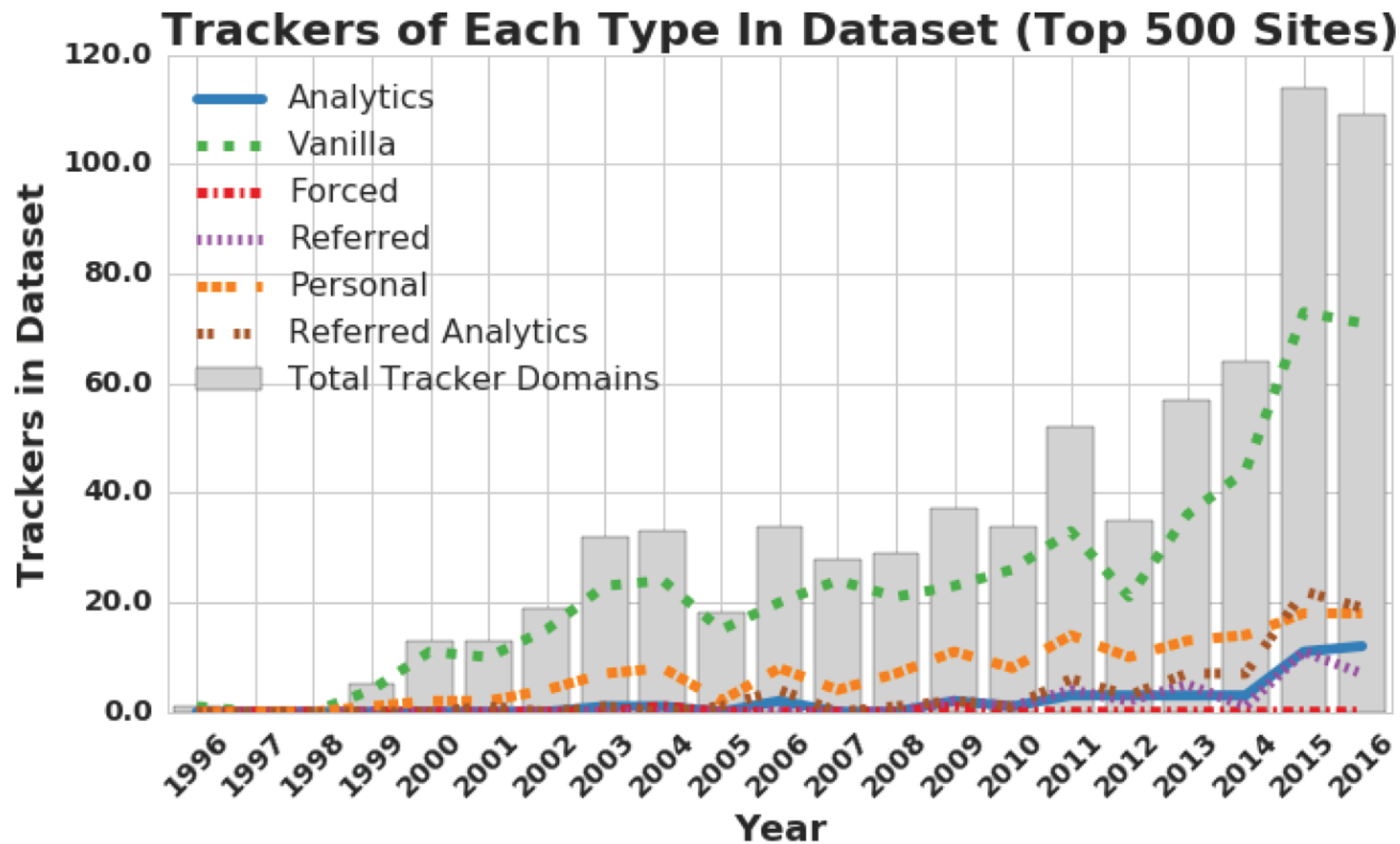


Početak *web trackera*

- 1994:
 - Pohrana podataka o korisnicima te manjak tehnike spremanja informacija za buduće posjete istih lokacija
 - Lou Montulli iz *Netscapea* razvija HTTP kolačiće kao način za omogućavanje trajnih veza između web stranica i korisnika no nije se pazilo kako će se tehnologija koristiti u budućnosti
- 1995:
 - Osnovan DoubleClick (jedna od prvih kompanija koja se specijalizirala za ciljano oglašavanje putem *third-party* kolačića)

Razvoj *web trackinga* (1996-2016)

- 1996: Prvi osnovni *trackeri* (osnovno praćenje) iz domene *microsoft.com* na domeni *digital.net* kao kolačići
- 1999: Prvi osobni *trackeri* na *go.com* od *Disneya*
- Rani 2000-ti:
 - Povećanje broja *trackera*, pojava *third-party trackera* i skripti
 - Najveći broj prisilnih *trackera* prije nego što su se još pojavili popup blokeri (npr. U 2004. za *Internet Explorer*)
- 2008: *Google* akvizirao *DoubleClick*



Broj trackera na popularnim web stranicama kroz povijest

- Kompleksnost web trackera raste kao i njihova uloga u prikupljanju informaciju

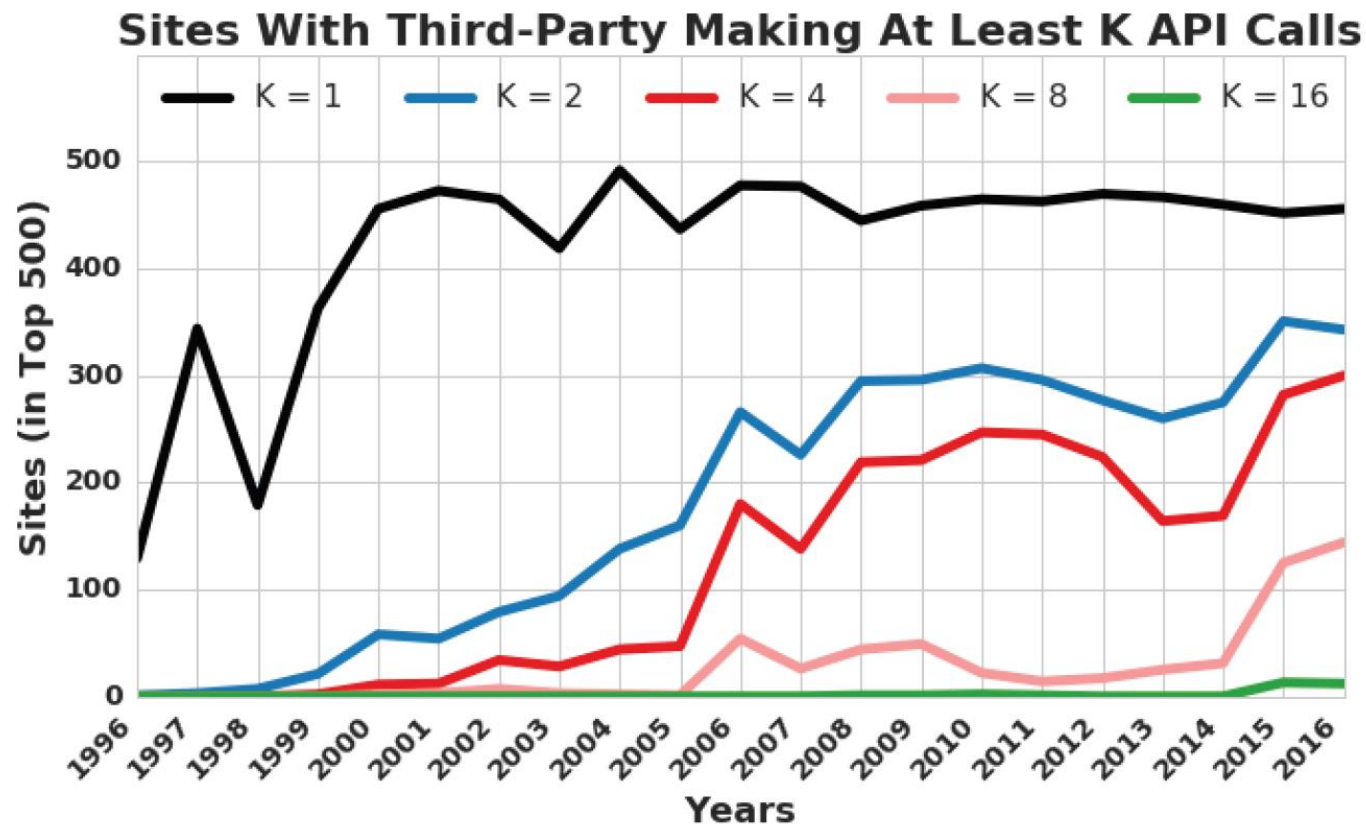
Year	1Type	2Type	3Type	4Type
1996	100.00% (1)	0	0	0
1998	0	0	0	0
2000	100.00% (13)	0	0	0
2002	100.00% (19)	0	0	0
2004	96.97% (32)	3.03% (1)	0	0
2006	100.00% (34)	0	0	0
2008	100.00% (29)	0	0	0
2010	94.12% (32)	2.94% (1)	2.94% (1)	0
2012	88.57% (31)	11.43% (4)	0	0
2014	93.75% (60)	4.69% (3)	1.56% (1)	0
2016	86.24% (94)	11.01% (12)	2.75% (3)	0

Složenost *web trackera* kroz godine

- 1Type – postotak nađenih *web trackera* koji samo jednoj vrsti po ulozi
- 2Type – postotak nađenih *web trackera* koji pripadaju dvama vrstama po ulogama istovremeno
- 3Type – postotak nađenih *web trackera* koji pripadaju trima vrstama ulogama istovremeno
- 4Type – postotak nađenih *web trackera* koji pripadaju četirima vrstama po ulogama istovremeno

Razvoj *web trackinga* (1996-2016)

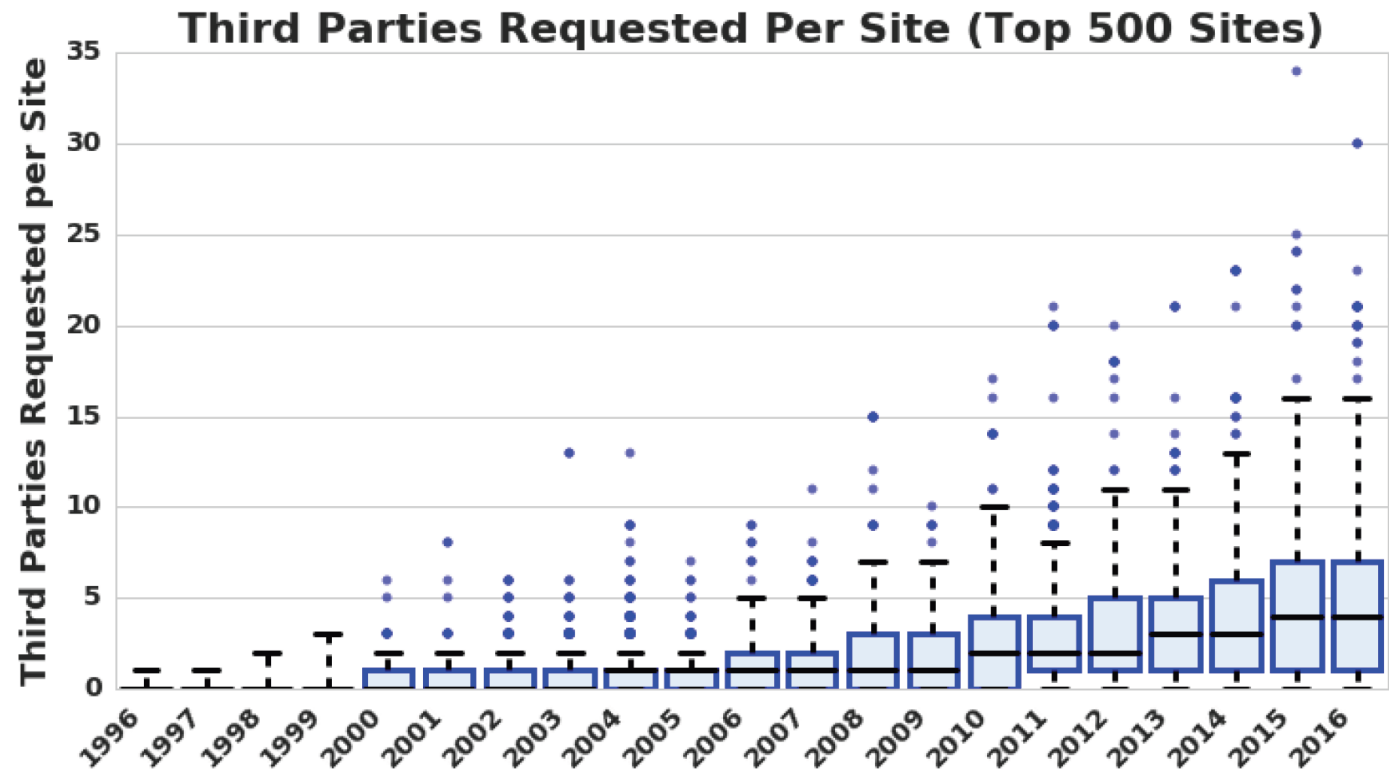
- Najčešće korišteni *fingerprinting* API-jevi:
 - *navigator.plugins*
 - *screen.width*



Broj stranica sa *web trackerom* koji izvršava barem K API poziva s digitalnim otiscima

Razvoj *web trackinga* (1996-2016)

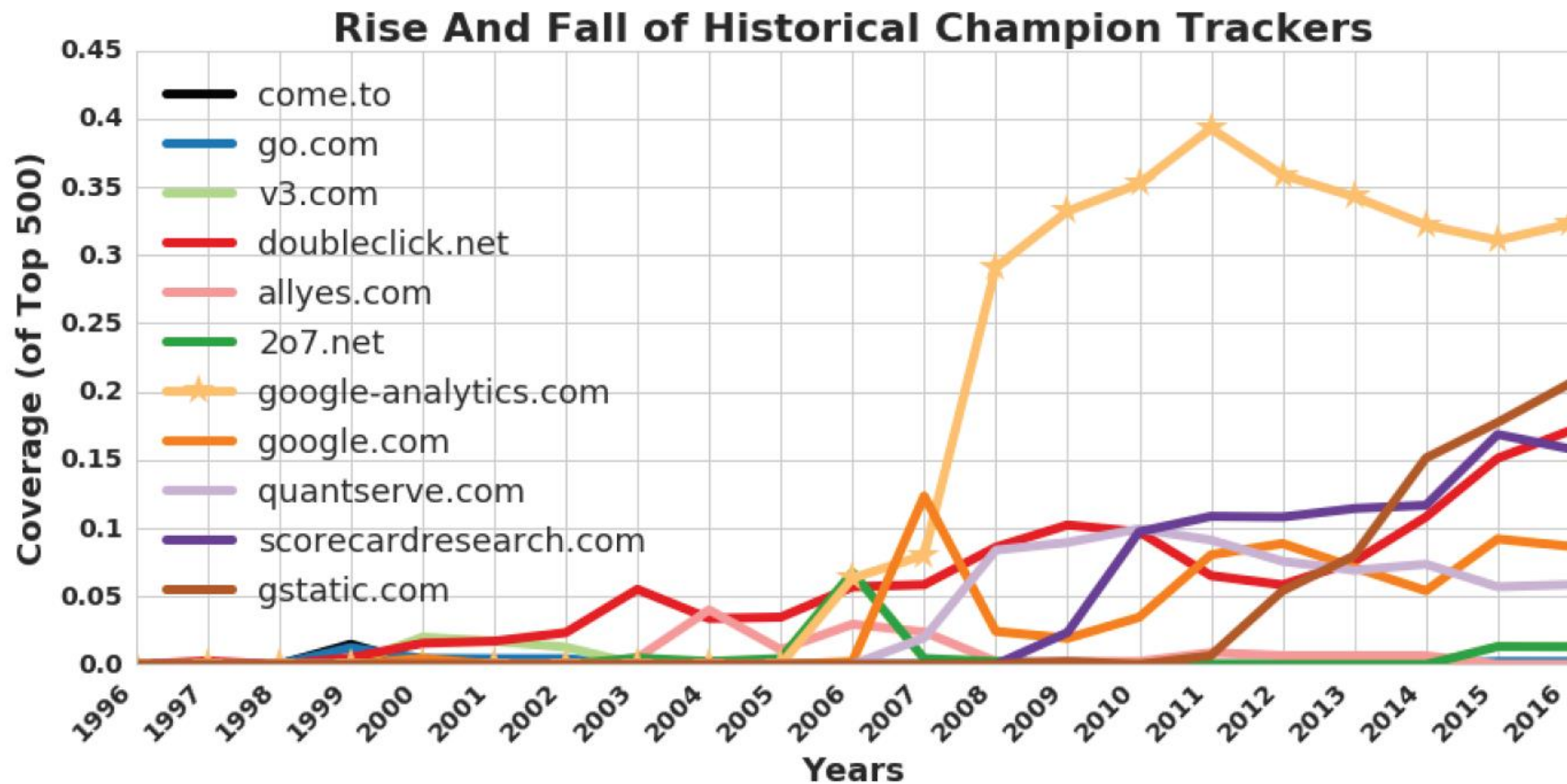
- Rane 2000:
 - Samo 5% stranica kontaktira više od 5 *third-party* domene
- 2016:
 - Gotovo 40% stranica kontaktira barem 5 *third-party* domena



Distribucija broja third-party poziva po stranici

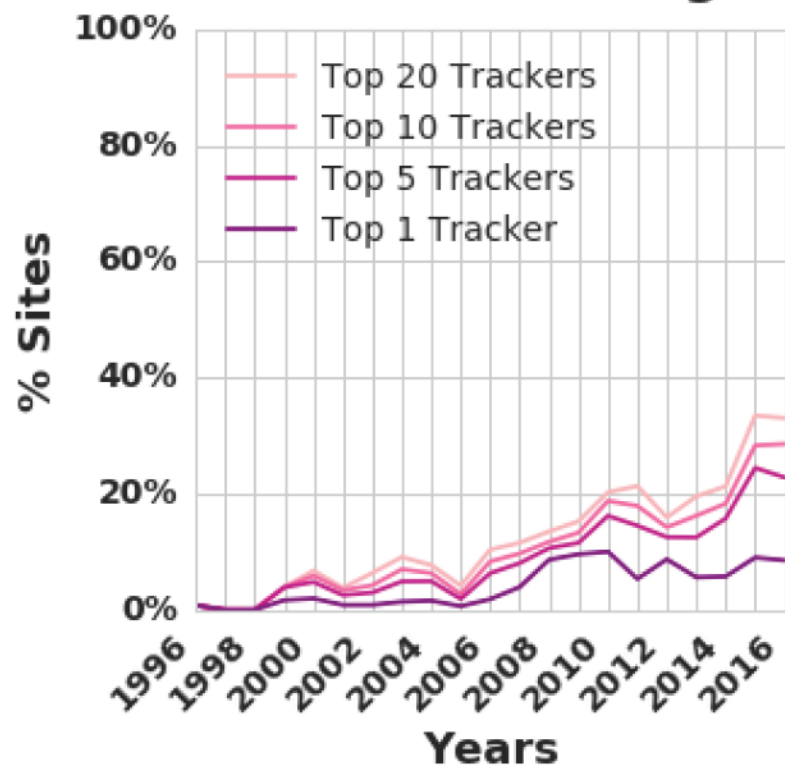
Razvoj *web trackinga* (1996-2016)

- Najpopularniji trackeri:
 - *Google Analytics, DoubleClick, ScorecardResearch*
 - Top 5 trackera pokriva 70% najpopularnijih stranica
 - *Google Analytics* u 2011. prisutan na 35% top stranica
 - *Google* ima najveći pristup podacima zahvaljujući akvizicijama kompanija za online oglašavanje i praćenje, poput *DoubleClicka*
 - Alati poput *Google Analyticsa* popularni su zbog svoje besplatne dostupnosti i jednostavnosti korištenja, što ih čini ključnim izborom za web stranice
 - *Google* koristi *Google Analytics* da stekne uvid u ogromne količine podataka o prometu na webu.

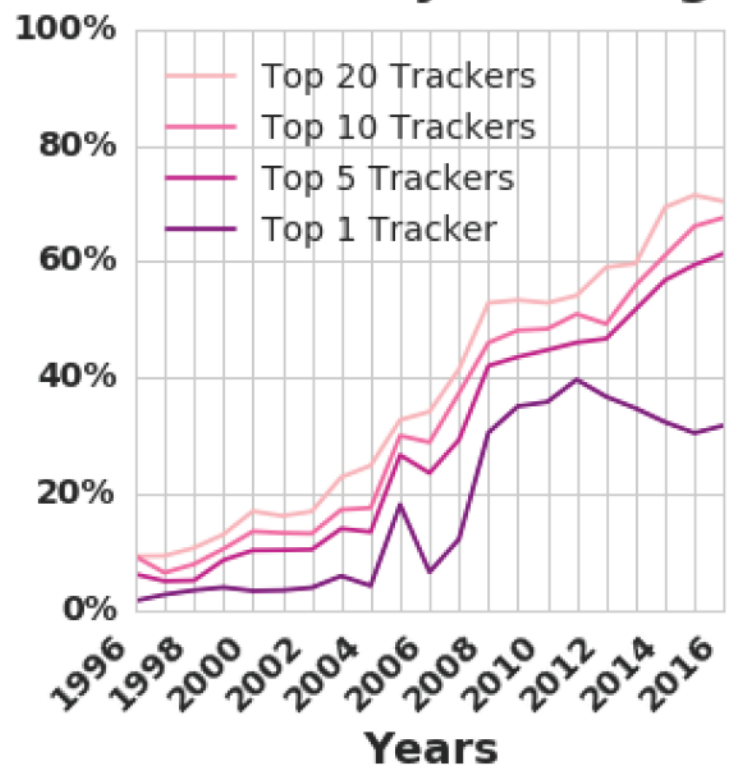


Broj trackera na popularnim web stranicama kroz povijest

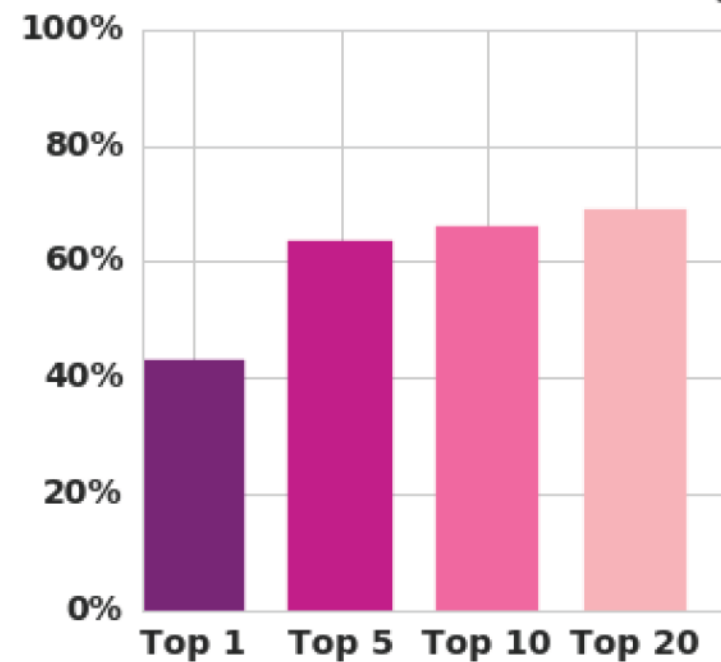
Tracker Coverage



Third Party Coverage



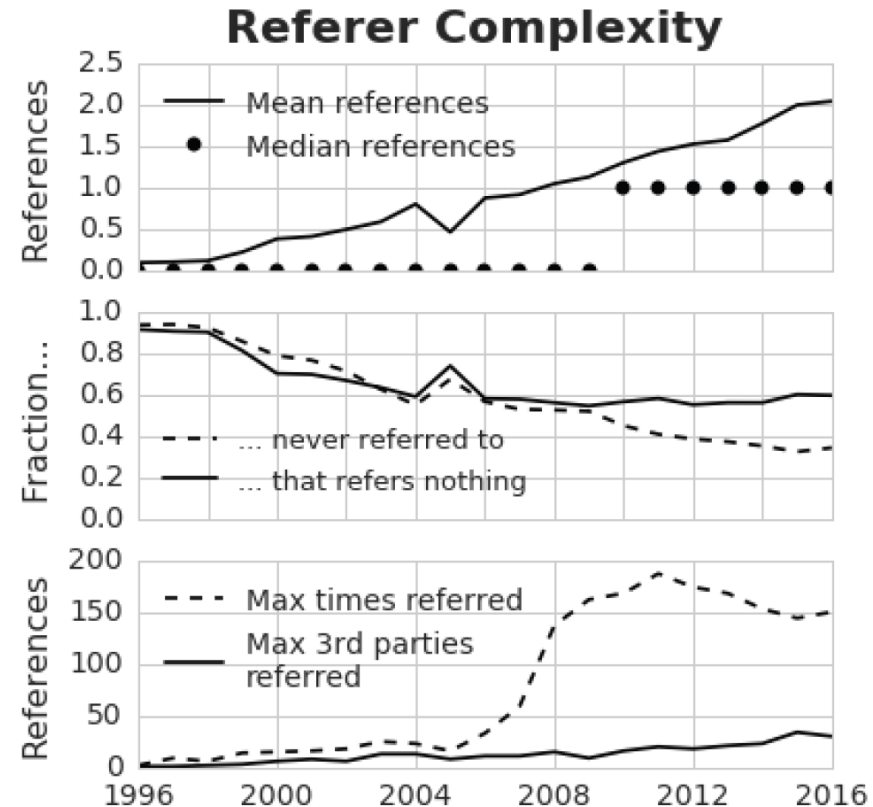
2016 Tracker Coverage



Rast top web trackera

Razvoj *web trackinga* (1996-2016)

- Povezanost među *trackerima* raste kroz godine – sve više *trackera* dijeli podatke s drugim *trackerima*
- Tako se lakše prate korisnici, profiliranje je preciznije te nastaju bolje ciljani oglasi



Promjene u učestalosti s kojom domene spominju ili upućuju na druge domene (na temelju HTTP Referer-a)

Trendovi nakon 2016.

- GDPR (2018.): Europska regulativa koja prisiljava web stranice na transparentnije praćenje
- CCPA (2020.): Kalifornijski zakon o privatnosti, prvi u SAD-u koji slijedi sličan model kao GDPR

Trendovi nakon 2016.

- Od 2016. Google Chrome je popularniji više nego ikad prije
- Preglednici s ugrađenim zaštitama:
 - Firefox uveo Intellegent Tracking Prevention (ITP) za blokiranje neovlaštenih kolačića
 - Brave, Librewolf, Mullvad

Zaključak

- *Web tracking* je postao složeniji
- Samo dio tvrtki ima monopol nad tržištem informacija
- Rast brige za privatnost
- Najvažnije: Potreba za zapisivanjem i/ili arhiviranjem (na bolje načine) povijesti *web trackera* i interneta za bolju edukaciju

Literatura

- [1] - Lerner, Ada, Anna Kornfeld Simpson, Tadayoshi Kohno, and Franziska Roesner. "Internet jones and the raiders of the lost trackers: An archaeological study of web tracking from 1996 to 2016." In 25th USENIX Security Symposium (USENIX Security 16). 2016.
- [2] - Kaspersky; What are Cookies? - <https://www.kaspersky.com/resource-center/definitions/cookies>, pristupljeno 16.12.2024.
- [3] - W3Schools; Web APIs - Introduction - https://www.w3schools.com/js/js_api_intro.asp, pristupljeno 16.12.2024.
- [4] - MDN Web Docs; Introduction to web APIs - https://developer.mozilla.org/en-US/docs/Learn/JavaScript/Client-side_web_APIs/Introduction, pristupljeno 16.12.2024.
- [5] - web.dev; Fingerprinting - <https://web.dev/learn/privacy/fingerprinting>, pristupljeno 16.12.2024.
- [6] - YouTube; NBC News; The History Of Internet Tracking And The Battle For Privacy | NBC News NOW - <https://www.youtube.com/watch?v=rKmO1nPGrVc>, pristupljeno, 7.12.2024.
- [7] - YouTube; What Happens When You Click "Accept All?" - https://www.youtube.com/watch?v=kDJTTeL_E-4, pristupljeno 16.12.2024.
- [8] - TheViewPoint; Origins and Pioneers of AdTech: 1990s- 2010 - <https://www.theviewpoint.com/insights/origins-and-pioneers-of-adtech-1990s-2010>, pristupljeno 14.12.2024.
- [9] - GDPR.eu; What is GDPR, the EU's new data protection law? - <https://gdpr.eu/what-is-gdpr/>, pristupljeno 16.12.2024.

Dodatna literatura

- [1] - Englehardt, Steven, and Arvind Narayanan. "Online tracking: A 1-million-site measurement and analysis." In Proceedings of the 2016 ACM SIGSAC conference on computer and communications security, pp. 1388-1401. 2016.
- [2] - Ermakova, Tatiana, Benjamin Fabian, Benedict Bender, and Kerstin Klimek. "Web tracking-A literature review on the state of research." (2018).
- [3] - Englehardt, Steven, Dillon Reisman, Christian Eubank, Peter Zimmerman, Jonathan Mayer, Arvind Narayanan, and Edward W. Felten. "Cookies that give you away: The surveillance implications of web tracking." In Proceedings of the 24th International Conference on World Wide Web, pp. 289-299. 2015.
- [4] - MDN Web Docs; A typical HTTP session - <https://developer.mozilla.org/en-US/docs/Web/HTTP/Session>, pristupljeno 15.12.2024.
- [5] - Am I Unique ?; Learn how identifiable you are on the Internet - <https://amiunique.org/>, pristupljeno 14.12.2024.
- [6] - Electronic Frontier Foundation; Cover Your Tracks - <https://coveryourtracks.eff.org/>, pristupljeno 14.12.2024.
- [7] - PrivacyTests; What are the best browsers in 2024? - <https://privacytests.org/>, pristupljeno 16.12.2024.
- [8] - Ghostery; Introducing Who's In the Know: The Privacy Pulse Report - <https://www.ghostery.com/blog/privacy-report-advertisers-and-adblockers>, pristupljeno 10.12.2024.

Hvala!