Bezpečnost webových aplikací

Klára Pešková, Klara.Peskova@mff.cuni.cz
Katedra softwaru a výuky informatiky, MFF UK
Základy tvorby webu, ZS 2021/22

Digitální stopa - co to je?

- digital footprint, digital shadow
- různorodé záznamy o činnosti uživatele ve virtuálním prostředí
- informace, které po sobě zanechává uživatel při pohybu na internetu
- např. vyhledávání, nakupování, příspěvky na sociálních sítích, diskuse pod články
- bez souhlasu / se souhlasem
- nejen aktivity na internetu (př. kartička Tesco)

Proč je to důležité?

- cenné (drahé) informace
- o digitální stopu bychom se měli zajímat, je to naše "digitální já"
- měli bychom zveřejňovat jenom to, co není možné snadno zneužít

Různá rozdělení

- Vlastní - vzniká vlastní činností uživatele
  - Vědomá (aktivní)
    - interakce na sociálních sítích, přispívání do diskusních fór, vkládáním fotografií do fotobank, e-mailová komunikace, ...
  - Nevědomá (pasivní)
    - IP adresa, vyhledávané výrazy, údaje o stráveném času a činnosti na určité webové stránce (cookies), poskytovatel připojení, lokace, ...
- zanechaná přáteli (nepřáteli)
  - např. označení na fotografii + uvedení polohy + sdílení
- informace z internetu prakticky není možné odstranit

K čemu se (ostatním) digitální stopa hodí?

- reklamy
  - retargeting (opakované zobrazení reklamy na stejné zboží)
    - navštívené stránky
    - doba, kterou na nich uživatel strávil
    - kliknutí na odkazy, ...
    - ... s těmito údaji se obchoduje
- kriminalita
  - např. krádež osobních údajů, digitální identity, profilů; kyberšikana ...
- pracovní pohovor
- rizika: ztráta soukromí, možné zneužití digitální stopy

Beze stopy?

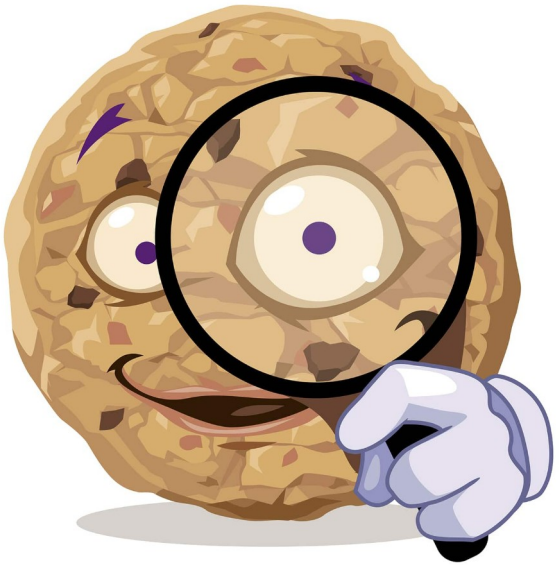
- nevytváření - téměř nemožné - proxy servery, šifrování, providelné mazání cookies...
- "rozmazání stopy" - několik účtů, které nic nespojuje
  - rozlišení osobních a pracovních aktivit
- "zametení" - prakticky nemožné
  - odstranění příspěvku - někdo ho už mohl sdílet, vyhledávače ho mají zaindexovaný, ...
  - odstranění profilu - např. Facebook - pouze deaktivace

Cookies

- co jsou cookies:
  - cookie jsou malé informace (páry klíč-hodnota), které si na náš počítač ukládá prohlížeč
  - cookies nastavuje stránka, kterou prohlížíme
  - mohou to být např. sesstion token, personalizované nastavení, … obecně cokoliv, co stránka potřebuje uchovat přes více requestů
  - při příští náštěvě stejného serveru prohlížeč pošle uložené cookies zpátky (stránce, která je nastavila)
- proč cookies existují? bezestavovost HTTP protokolu
  - jednotlivé requesty jsou od sebe oddělené, server potřebuje rozlišit uživatele

Tracking cookies

- dvě "kategorie" cookies: (technicky stejné)
- first-party cookies
  - patří stránce, kterou uživatel otevřel
- third-party cookies (tracking cookies)
  - jsou do stránky vloženy pomocí javascriptu
  - typicky patří nějaké stránce, která zobrazuje reklamu



Retracking

- jak to funguje:
  - autor stránky (first-party) umístí na svojí stránku nenápadný (neviditelný) kousek kódu (pixel), který získá od reklamních stránek (third-party)
  - když si uživatel (second-party) zobrazí stránku, pixel vytvoří third-party cookie
  - pokud uživatel při prohlížení webu narazí na stránku, která používá stejnou reklamní third-party, uživateli se zobrazí reklama na tu samou ledničku, kterou nedávno hledal

Tracking cookies - uklidnění

- If this type of tracking keeps you up at night, consider that an advertiser can already track the sites you visit based a combination of your IP address, browser version, location, and any number of other factors—so getting rid of the tracking cookies only eliminates a small piece of the puzzle when it comes to tracking your behavior online. There are also only a few advertisers big enough to really track you across the majority of web sites—and one has to assume Google already knows everything else you're doing online.

  zdroj:
  https://lifehacker.com/fact-and-fiction-the-truth-about-browser-cookies-5461114

Úkoly - digitální stopa

- Najděte moji adresu

- Clickclickclick.click

Same-origin policy (SOP)

- důležitý nástroj webové bezpečnosti
- prohlížeč povolí skriptům na jedné stránce přístup k datům na druhé stránce pouze pokud mají same-origin
- cíl - ochránit informace z jedněch stránek před okolním světem (zabránit krádeži cookies)
- Co je origin?
  Posuzuje se odkud pochází request

Co je origin?

- protocol://xxx.domain.cz:[port]

Compared URL	Outcome	Reason
http://www.example.com/dir/page2.html	Success	Same scheme, host and port
http://www.example.com/dir2/other.html	Success	Same scheme, host and port
http://username:password@**www.example.com**/dir2/other.html	Success	Same scheme, host and port
http://www.example.com:**81**/dir/other.html	Failure	Same scheme and host but different port
https://www.example.com/dir/other.html	Failure	Different scheme
http://**en.example.com**/dir/other.html	Failure	Different host
http://**example.com**/dir/other.html	Failure	Different host (exact match required)
http://**v2.www.example.com**/dir/other.html	Failure	Different host (exact match required)
http://www.example.com:**80**/dir/other.html	Depends	Port explicit. Depends on implementation in browser.

zdroj: https://en.wikipedia.org/wiki/Same-origin_policy

- pokud dojde k porušení, prohlížeč zastaví request

Same-origin policy

- důležité hlavně pro moderní aplikace, které používají cookies - např. k identifikaci uživatele (jeho session), k uložení citlivých informací
- SOP implementuje prohlížeč

SOP - čemu zabrání

- modelová situace:
  - uživatel je přihlášený na Facebooku
  - v jiném tabu si otevře zlou stránku
  - (bez SOP) JavaScript na zlé stránce si může s Facebookovou stránkou dělat, co chce - vše, co může uživatel
    - číst zprávy, posílat věci na zeď, přečíst si údaje v přihlašovacím formuláři pomocí HTML DOM po jejich vyplnění (před odesláním formuláře)
  - Facebook ale CHCE používat Javasctipt, ale jenom "svůj" => same-origin policy

Rozvolnění

- embedding:
  - multimédia <img>, <video>, <audio> ...
  - vkládání javascriptu <script>, css <link rel="stylesheet">
  - fonty (@font-face) - některé prohlížeče povolují, některé ne
  - <iframe> - externí HTML

=> SOP platí pouze pro skripty

Jak povolit?

- Cross-Origin Resource Sharing (CORS)

  = Header v HTTP response

- nastavení v PHP:

```
<?php
$server = $_SERVER['HTTP_ORIGIN'];
Header('Access-Control-Allow-Origin: ' . $server);
?>
```