VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
FAKULTA INFORMAČNÍCH TECHNOLOGIÍ

WAP - Internetové aplikace
Projekt 2 – HotJar

Alexandra Slezáková (xsleza20)
xsleza20@stud.fit.vutbr.cz
22. marca 2023

1 Úvod

V projekte som sa zaoberala nástrojom HotJar, ktorý slúži na zaznamenávanie aktivity účastníkov webu. Tieto údaje môžu pomôcť obchodníkom lepšie pochopiť správanie používateľov a zlepšiť tak používateľskú skúsenosť na webe. Nástroj ponúka podobné, možno aj kvalitnejšie funkcie ako tomu je u Google Analytics.

Dáta, respektíve správanie návštevníkov je zaznamenávané pomocou kódu, ktorý je potrebné vložiť na stránku ako napríklad kód Google Analytics. HotJar ponúka rôzne typy výstupov ako napríklad heatmaps na zaznamenanie častí s najväčším pohybom alebo počtom klikov, nahrávky pohybu kurzora, úspešnosť vyplnenia formulára a získanie spätnej väzby.

V rámci projektu som sledovala, či nástroj dodržuje DNT popísané v sekcii 2. Sekcia 3 bližšie popisuje replikované pokusy zaznamenané v [1]. Prenášané dáta sú spomenuté v sekcii 4. Ďalej sekcia 5 popisuje zmeny, ktoré boli vykonané v HotJar skripte a sekcia 6 špecifikuje použitie vytvorenej webovej aplikácie pre účely projektu.

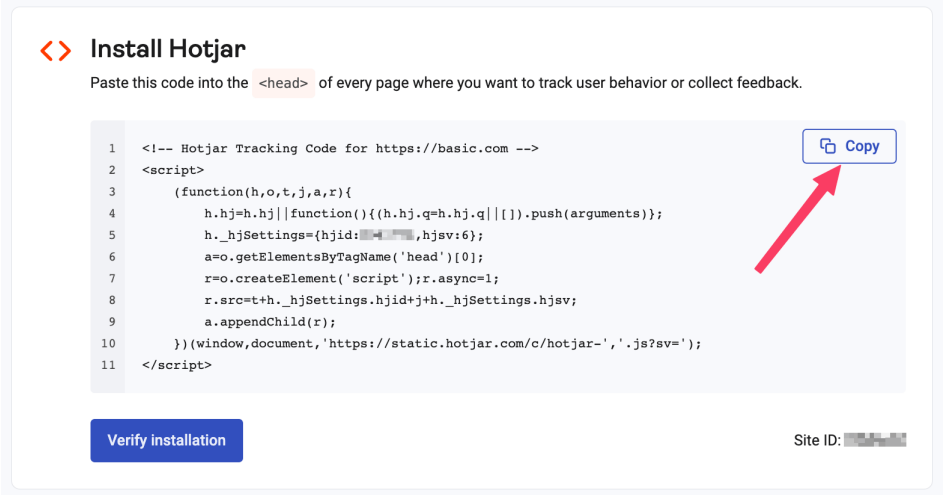
2 Do Not Track

„Do Not Track“ je možnosť nastaviteľná vo webovom prehliadači, ktorá pridáva pole do hlavičky HTTP žiadosti a informuje tým ostatné webové stránky, že daný užívateľ nechce byť sledovaný. Do roku 2011 bol DNT prijatý všetkými hlavnými prehliadačmi, ale mnohé webové stránky ho odmietli rešpektovať [2].

Táto sekcia popisuje, akým spôsobom HotJar dodržuje „Do Not Track“ v rôznych prehliadačoch. Ďalej sa sekcia zameriava na úpravu skriptu, ktorá umožní ignorovanie „Do Not Track“.

2.1 Dodržovanie „Do Not Track“

HotJar je možné nainštalovať jednoduchým spôsobom - vložením poskytnutého skriptu do hlavičky každej stránky, ktorá má byť sledovaná. Tento skript pomocou metódy `appendChild()` vloží ďalší skript do hlavičky webovej stránky (riadok 9 obrázku 1). Vo vloženom skripte sú inicializované vlastnosti objektu `window` využité ďalším vloženým skriptom, ktorý vykonáva samotné nahrávanie.



Obr. 1: HotJar inštalácia.

Skript na zaznamenanie udalostí vykonaných na stránke obsahuje podmienku, ktorá kontroluje, či príznak `doNotTrack` je nastavený na 1 alebo nie. A teda HotJar dodržuje toto nastavenie.

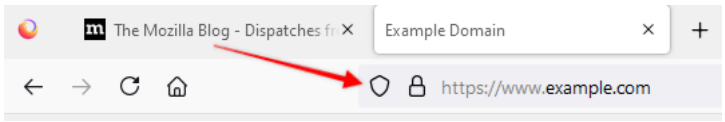
2.2 Odstránenie obmedzenia na prenos dát

Aj keď HotJar dodržuje použitie „Do Not Track“ príznaku, skript je možné upraviť odstránením tejto podmienky, ktorá príznak kontroluje. Ďalšou možnosťou je natvrdo nastaviť tento príznak už v `html` súbore. Aj keď príznak má vlastnosť `readonly`, je možné ho nastaviť definovaním novej vlastnosti objektu `navigator` zobrazenej na obrázku 2.

```
Object.defineProperty(window.navigator, 'doNotTrack', {
    value: null
});
```

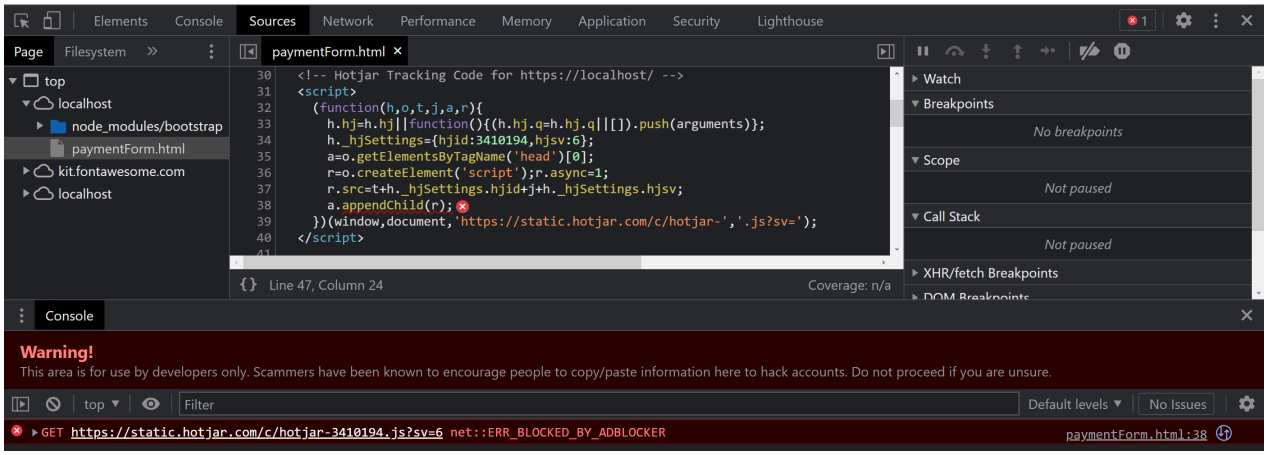
Obr. 2: Zmena príznaku `doNotTrack`.

Spomenuté zmeny fungovali v prípade použitia prehliadača *Google Chrome*, *Microsoft Edge*, *Firefox* alebo *Opera*, kde interakcia so stránkou bola nahraná, aj keď možnosť „Do Not Track“ bola v prehliadači zapnutá. V prehliadači *Firefox* je však možné zapnutie rozšírenej ochrany pred sledovaním, zobrazené na obrázku 3, ktorá podľa výpisu v konzole zabraňuje načítaniu určitých skriptov a blokuje aj prístup ku cookies.



Obr. 3: Rozšírená ochrana vo Firefox.

Webový prehliadač *Opera*, podobne ako *Firefox*, ponúka ochranu súkromia, nazývanú Tracker Blocker, ktorá priamo blokuje metódu `appendChild()`, zobrazené na obrázku 4, bez ohľadu na zapnutie „Do Not Track“. V tomto prípade teda nedochádza k vloženiu HotJar skriptov do hlavičky `html` dokumentu.



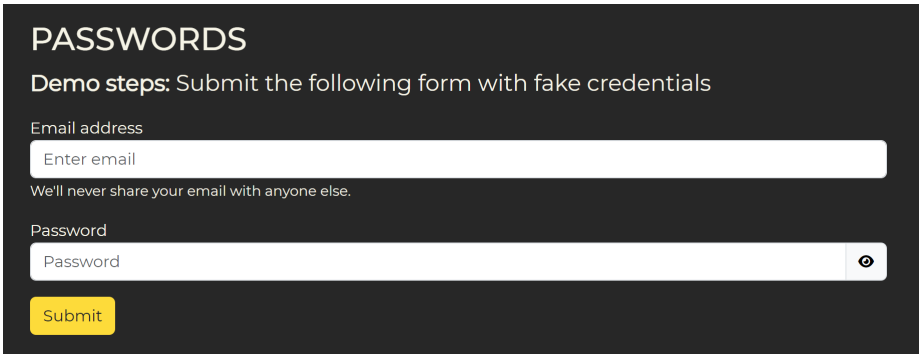
Obr. 4: Webový prehliadač *Opera*.

3 Expertimenty

Táto sekcia sa zaoberá experimentami uvedenými v [3].

3.1 Heslá sú zahrnuté v záznamoch relácie

Podľa [3] sú heslá v zázname vynechané. Toto tvrdenie bolo overené pomocou prihlasovacieho formulára na obrázku 5.



Obr. 5: Prihlasovací formulár.

Nápoveda HotJar[1] špecifikuje, že vstupné pole s ID atribútom `password` alebo `pass` je ignorované. Aj po nastavení spomenutého ID bolo heslo súčasťou nahrávky. Obsah tohto poľa bol maskovaný určitým počtom hviezdičiek rôznym od reálnej dĺžky hesla. Formulár ďalej obsahoval pole pre email, ktoré bolo taktiež maskované hviezdičkami.

HotJar ponúka možnosť použiť atribút `data-hj-allow`, ktorý umožňuje nepotláčanie obsahu vstupného poľa typu `text` a teda užívateľský vstup nie je nahradený hviezdičkami.

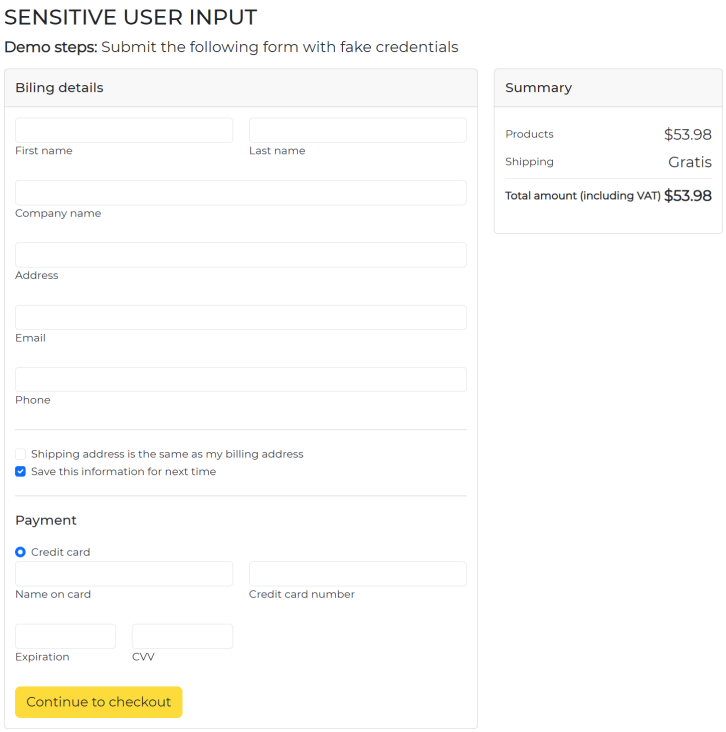
Únik dát bol odhalený pri použití atribútu `data-hj-allow` a nepoužití ID atribútu s hodnotou `pass` alebo `password` poľa pre heslo. Aj keď pole typu `password` je maskované, na každej webovej stránke je možné heslo zobraziť. Vtedy sa mení typ tohto poľa z `password` na `text` a celé heslo je tak zaznamenané v nahrávke[2], ak sa heslo počas zobrazenia upravuje. Avšak v prípade použitia špecifických ID atribútov bolo pole maskované bez ohľadu na jeho typ.

[1] Povolenie obmedzení: Allowing restrictions

[2] Niekedy bolo heslo viditeľné v nahrávke, ktorá bola prehratá opakovane. Pri prvom pozretí nahrávky bolo heslo maskované.

3.2 Citlivé používateľské vstupy sú redigované čiastočným a nedokonalým spôsobom

[3] uvádza, že HotJar vynecháva špecifické polia `input` podľa ich typu. Ďalej popisuje, že pole pre meno, email, telefón, adresu a niektoré údaje z kreditnej karty neboli maskované. Od napísania článku toto HotJar pravdepodobne zmenil a všetky vstupné polia sú defaultne maskované. Pre účely zopakovania tohto experimentu bol vytvorený formulár zobrazený na obrázku 6.



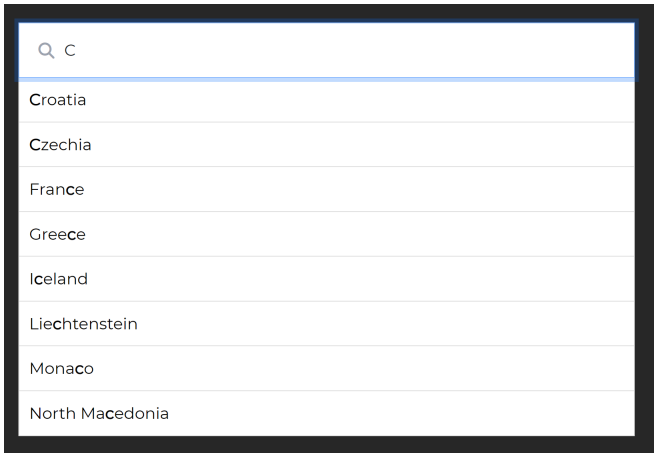
Obr. 6: Prihlasovací formulár.

Ako prvé boli vykonané experimenty, kde všetky vstupné polia mali špecifikované ID atribúty tak, ako je to uvedené v nápovede HotJar spomenutej v sekcii 3.1. Všetky polia boli maskované určitým počtom hviezdičiek a teda k úniku dát vôbec nedošlo.

Všetky údaje so vstupných polí typu `text` boli v nahrávke zaznamenané v prípade, keď bol použitý `data-hj-allow` atribút a ID atribúty boli volené ľubovoľne. K maskovaniu nedošlo ani v prípade použitia atribútu `autocomplete`. Ak ID atribút obsahoval slovo `ccnumber` obsah vstupného poľa bol nahradený hviezdičkami.

3.3 Vykreslený obsah stránky

Zdroj [1] uvádza, že spoločnosti zaznamenávajúce relácie zhromažďujú aj vykreslený obsah stránky. Tento experiment bol vykonaný na jednoduchom filtri krajín Európy, ktorý je zobrazený na obrázku 7.



Obr. 7: Filter krajín Európy.

Aj keď obsah vstupného poľa nebol zaznamenaný, podľa nájdených výsledkov, ktoré sú v zázname viditeľné, je možné predpokladať, čo sa užívateľ snažil hľadať.

3.4 Nahrávacie služby môžu zlyhať pri ochrane používateľských údajov

V [1] je popísaný spôsob, akým nahrávacie služby môžu zlyhať pri ochrane používateľských údajov. V dobe vzniku článku nástenka HotJar poskytovala prehrávanie nahrávok na stránke HTTP. To umožňovalo „man-in-the-middle“ vložiť skript na stránku prehrávania a extrahovať tak všetky údaje zo záznamu. Dnes už HotJar používa HTTPs.

4 Prenášané dáta

Dáta boli prenášané až po zatvorení tabu vo webovom prehliadači a konkrétne boli prenesené údaje o pozícií, kde došlo ku kliknutiu alebo pohybu myši, ďalej sa prenášali údaje so vstupných polí vrátane časovej značky, času, selektoru a typu. Medzi ďalšie prenášané dáta patrila poloha užívateľa (zaznamenávanie tejto informácie bolo možné vypnúť), operačný systém, veľkosť obrazovky.

5 Úpravy skriptu

Po preskúmaní skriptu na zaznamenávanie relácie bola nájdená funkcia `getSuppressedText()`, na obrázku 8, ktorá prevádza text na hviezdičky.

```
1    getSuppressedText: function(e, t) {
2        return j[e] ? j[e](t) : a(t)
3    }
```

Obr. 8: Funkcia na potlačenie textu.

Uprávou funkcie tak, aby vracala len parameter `t` bolo možné získať všetky údaje, ktoré boli vyplnené vo vstupnom poli. V skripte bola ďalej upravená hodnota vlastnosti `suppression` z `full` na `none`. Všetky úpravy skriptu sú označené pomocou komentára `//CHANGES` pre jednoduchšie vyhľadávanie.

6 Použitie vytvorenej webovej aplikácie

Webová aplikácia sa nachádza na stránke https://wap-project.8u.cz/. Na úvodnej stránke je menu s tromi vykonanými pokusmi. Na zaznamenávanie relácie je použitý HotJar s plánom zadarmo. Záznamy sú dostupné po prihlásení sa pomocou **Google účtu**. Prístupové údaje sú:

- email: wap.project2@gmail.com
- heslo: Wap-project2*

V prípade, že HotJar zobrazuje varovanie, že HotJar skript nie je správne nainštalovaný, je to len z toho dôvodu, že dlhšiu dobu nebolo nič na stránke zaznamenané. Nie je teda potrebné overovať správne nainštalovanie HotJar. Po vzniknutí nahrávky upozornenie zmizne. Ďalej HotJar zobrazuje nahrávky niekoľko minút po zatvorení sledovanej stránky.

Ku experimentom popísaným v sekcii 3.1 je možné pristúpiť cez prvú položku v menu, **Passwords**. Ďalej je možné vybrať jeden z troch experimentov s heslami. Experiment *Changed replay script* používa na zaznamenávanie interakcie používateľa upravený skript a teda dochádza k úniku údajov. Formulár použitý v tomto experimente je totožný s formulárom v *No data leaks*. V *Data leaks* taktiež dochádza k úniku údajov vyplnených vo formulári kvôli nesprávne použitým ID atribútom vstupných polí a použitiu atribútu `data-hj-allow`. Na vytvorenie nahrávky je použitý defaultný HotJar skript. Formulár *No data leaks* bol vytvorený podľa predpisov z nápovedy HotJar a nedochádza tak k úniku žiadnych údajov.

Experiment zo sekcie 3.2 je dostupný z položky **Sensitive user input**. Rovnako ako aj pri experimentoch s heslami boli vykonané 3 pokusy - s upraveným skriptom, s nesprávnymi ID atribútmi a bez úniku údajov.

Posledná položka v menu zahrňuje pokus zo sekcie 3.3.

7 Záver

Aj keď HotJar rešpektuje príznak „Do Not Track“, úpravou HTML kódu je možné ho zmeniť, ak prehliadač priamo nepodporuje lepšiu ochranu ako to bolo v prehliadači Opera alebo Firefox.

Znovu vykonané experimenty podľa [1] ukazujú, že došlo k zmene zaznamenávania relácie. Všetky vstupné polia sú defaultne maskované, pokiaľ nemajú atribút `data-hj-allow`. Použitím tohto atribútu nedochádza k potláčaniu vstupného poľa typu `text`. Ostatné typy sú vždy maskované bez ohľadu na použitie spomenutého atribútu. Ďalej bolo ukázané, že úpravou skriptu veľké množstvo údajov unikne.

Literatúra

[1] Gunes Acar, Steven Englehardt, and Arvind Narayanan. No boundaries: data exfiltration by third parties embedded on web pages. *Proceedings on Privacy Enhancing Technologies*, 2020:220–238, 10 2020.

[2] What is „Do Not Track“. https://www.avast.com/c-what-is-do-not-track.

[3] Steven Englehardt. No boundaries: Exfiltration of personal data by session-replay scripts, 2017.