

# Získávání informací o uživateli na webových stránkách

Ing. Tomáš Vondráček, Ing. Libor Polčák, Ph.D.

tomasvondrac@gmail.com, polcak@fit.vut.cz

Fakulta informačních technologií Vysokého učení technického v Brně

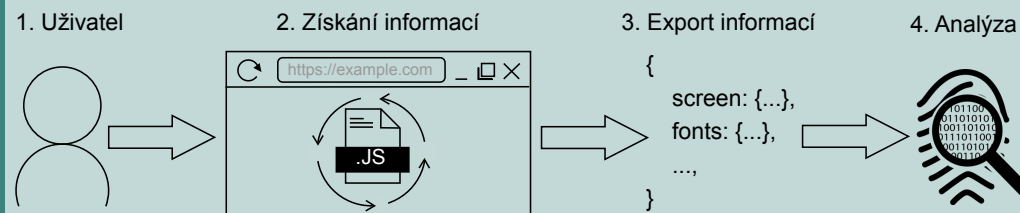
## Motivace

K síti Internet je v roce 2021 připojeno již 5 miliard uživatelů a existuje více než 1.8 miliardy webových stránek. Přitom každá webová stránka může o svých uživateli získat mnoho informací, kde jejich kombinací lze vytvořit potenciálně unikátní digitální otisk uživatele. Otisk, resp. informace obsažené v otisku, mohou v praxi například využít:

- ▶ reklamní agentury pro přesnější cílení reklamy,
- ▶ protokoly zajišťující autentizaci a autorizaci,
- ▶ společnosti určující rizikové skóre.

## Získávání informací

Proces získávání informací je realizován specifickým skriptem, který se nachází na libovolné webové stránce. Skript je spuštěn ihned, jakmile uživatel navštíví danou webovou stránku.



Navržená a implementovaná knihovna je schopna získat **127 informací** o uživateli. Informace se týkají konfigurace uživatelského zařízení, obrazovky nebo prohlížeče.

## Detekce webových rozšíření

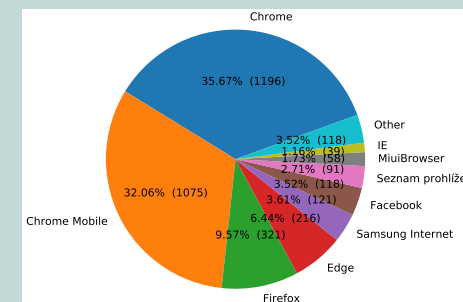
Uživatelé mají v dnešní době na výběr desítky webových rozšíření, kterými mohou maskovat své informace za účelem zvýšení anonymity. V rámci práce se podařilo detekovat **117 vedlejších efektů** v 19 testovaných webových rozšířeních. Tyto vedlejší efekty lze s výhodou použít pro přesnější identifikaci uživatele.

Vedlejší efekty se dále dělí na anomální a nekonzistentní dle typu získané informace. Metody pro získání anomálních informací jsou detekce: falešné nativní funkce, rozšiřitelnosti objektů, konfigurovatelnosti objektů a rozdílných výsledků. Nekonzistentní informace se detekují v rodině prohlížečů, operačním systému, zobrazovacím zařízení, preferovaných jazycích, úložiscích a HTTP hlavičkách.

## Datová sada

Celý vytvořený systém skládající se z knihovny, serveru a databáze byl nasazen na 4 komerčních webových stránkách. Dále bylo vytvořeno webové rozšíření pro účely analýzy stability informací v čase.

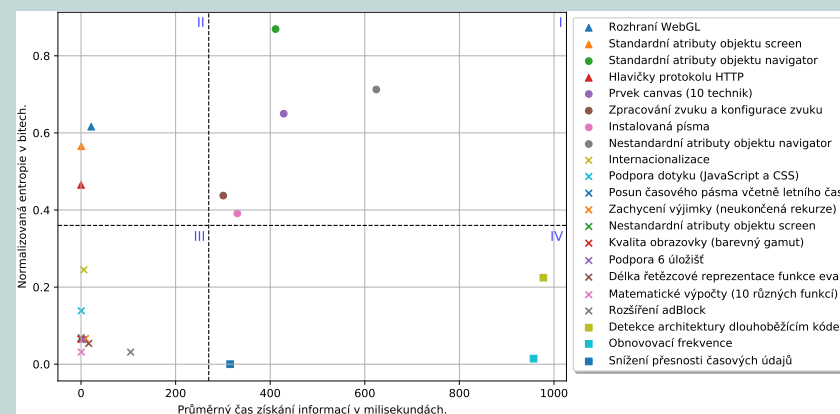
Za více než měsíc bylo z těchto 4 webů získáno **7 632** otisků. Po provedení přípravy dat zbylo potenciálně **3 353** unikátních otisků, na kterých je provedena analýza.



## Normalizovaná entropie a rychlost získání informací

Rychlost získání informací hraje velmi důležitou roli v získávání informací zejména z hlediska snížení zátěže webové stránky a rychlosti následné identifikace uživatele.

Nejlepší poměr entropie vůči času získání dosahuje skupina informací z **rozhraní WebGL**.



## Závěr

- ▶ Optimalizace nebo představení několika nových metod získávání informací
- ▶ Detekce 117 vedlejších efektů z 19 webových rozšíření
- ▶ Analýza včetně experimentů provedena na datové sadě obsahující 3 353 otisků
- ▶ 94 % otisků v datové sadě je zcela unikátních