PRIVACY OP HET WEB EN DE MOORD OP HET COOKIEMONSTER

Wanneer je aan het surfen bent op het internet kan je er niet meer aan ontsnappen: pop-ups, waarschuwingen, bijna dreigementen: "*Opgelet, deze website bevat cookies".* Dit is een gevolg van de Europese cookiewet, bedoeld om de privacy van mensen te beschermen. Cookies worden immers al jaren gebruikt om mensen te "tracken" op het internet. Op deze manier weten advertentiebedrijven welke websites je bezoekt. Zo kunnen ze reclame sturen die gebaseerd is op de webpagina's waar je naartoe surft.
Deze thesis zoekt een alternatieve methode om mensen te volgen op het internet, een methode die geen inbreuk doet tegen de nieuwe cookiewet, zodat advertentiebedrijven toch door kunnen gaan met het sturen van reclame die is afgestemd op jouw interesses.

Het is misschien al opgevallen dat wanneer je aan het surfen bent op het internet de advertenties lijken te weten wat je aan het doen bent: na het bezoeken van de website van Samsung krijg je plots reclame over de nieuwste smartphone, en als je op zoek bent naar een recensie van de laatste Tarantino-film word je overstelpt met advertenties voor Blu-rays.

De oorzaak hiervan is een techniek die *Behavioural Profiling* wordt genoemd: bedrijven proberen te volgen welke websites je bezoekt en gaan aan de hand daarvan advertenties tonen. Hiervoor wordt er typisch gebruik gemaakt van cookies.

Cookies zijn kleine bestandjes die worden opgeslagen op je webbrowser. Ze zijn ontworpen om informatie bij te houden die nuttig is voor het bezoeken van een webpagina: ben je ingelogd, in welke taal wil je de pagina bekijken, welke producten zitten er in je winkelmandje,…
Ze kunnen echter ook gebruikt worden om je te herkennen wanneer je een webpagina bezoekt: bij een eerste bezoek wordt er een bestandje opgeslagen met een code in waaraan je herkend kan worden (zoals een soort barcode). Bij een volgend bezoek kan de webpagina je herkennen aan de hand van de code in de cookie.
Advertentiebedrijven kunnen internetgebruikers herkennen over een heel netwerk van websites, zodat ze kunnen zien welke webpagina's je bezocht hebt.

Het volgen van bezoekers over verschillende websites wordt door velen beschouwd als een inbreuk op de privacy, en daarom keurde de Europese Unie in 2009 de Europese Cookiewet goed. Deze wet verplicht websites om toestemming te vragen aan hun bezoekers om cookies te mogen gebruiken. Hierbij moeten de bezoekers ook ingelicht worden over de functie van de cookies. Op deze manier wil de Europese Unie de privacy van de internetgebruikers beschermen.

Een gevolg van deze cookiewet is dat webpagina's hun bezoekers moeten inlichten indien ze cookies gebruiken voor advertentiedoeleinden, en vervolgens toestemming moeten krijgen om deze cookies te mogen plaatsen. Hierdoor zullen advertentiebedrijven veel minder in staat zijn om internetgebruikers te volgen over verschillende websites, met als gevolg dat ze minder accuraat kunnen voorspellen in welke advertenties ze geïnteresseerd zijn. Als resultaat hiervan zal er minder op advertenties geklikt worden, en dit betekent een verlies van inkomsten.

In deze thesis wordt er een alternatieve methode gezocht voor het herkennen van internetgebruikers zodat ze toch nog gevolgd kunnen worden over het internet zonder de cookiewet te schenden. In eerste instantie werd er op zoek gegaan naar bestaande technieken die hiervoor gebruikt kunnen worden. Na onderzoek van het domein werd er een taxonomie opgesteld. De taxonomie bestaat uit 4 types van technieken: lokale opslag, hardware, locatie en de webbrowser.

Methoden die gebruik maken van lokale opslag proberen een bezoeker te herkennen door die bezoeker een eigenschap te geven waardoor deze herkend kan worden; door een soort van naamkaartje op de bezoeker te kleven. Cookies zijn het bekendste voorbeeld van deze techniek. Lokale opslag mag sinds de cookiewet dus niet meer gebruikt worden zonder de toestemming van de bezoeker.

De hardware-methoden proberen informatie op te vragen over het apparaat waarmee de webpagina bezocht wordt. Indien er nuttige informatie wordt verkregen kan deze opgeslagen worden. Bij een volgend bezoek wordt de apparatuur vergeleken met die van vorige bezoekers, zodat een terugkerende bezoeker herkend kan worden.

Een derde methode maakt gebruik van de locatie vanaf waar iemand een webpagina bezoekt. Bij een bezoek wordt de locatie van de gebruiker opgevraagd en opgeslagen. Deze locatie kan vergeleken worden met de locaties van vorige bezoekers om in te schatten wie de huidige bezoeker is.

De laatste methode maakt gebruik van je webbrowser (bijvoorbeeld Internet Explorer of Firefox). Deze methode probeert zoveel mogelijk informatie over je browser te weten te komen: bijvoorbeeld welke lettertypes geïnstalleerd zijn, of welke versie van browser je gebruikt. Indien er voldoende informatie beschikbaar is kan een bezoeker op deze manier uniek herkend worden. Er moet echter rekening gehouden worden met het feit dat deze informatie kan veranderen, bijvoorbeeld na de installatie van een nieuw lettertype.

Om een keuze te maken tussen deze technieken werden ze met elkaar vergeleken door gebruik te maken van criteria die speciaal voor deze vergelijking zijn opgesteld: bijvoorbeeld ‘Hoe uniek kan iemand herkend worden?’ en ‘Hoe lang duurt het om een overeenkomstig profiel te vinden’.

Uit deze vergelijking bleek dat de webbrowser techniek de meest geschikte techniek is. Om te verifiëren in welke mate bezoekers nu écht herkend kunnen worden, werd er een webpagina gemaakt die bezoekers probeerde te herkennen aan de hand van de webbrowser-techniek. Hierbij werd er ook rekening gehouden met de inhoud van de pagina’s die bezocht werden (bijvoorbeeld sport, actualiteit,…) om een bezoeker te herkennen. Uit de test bleek dat 77% procent van de bezoekers correct werden herkend.

De cookiewet probeert het volgen van mensen op het internet aan banden te leggen. Deze thesis bewijst dat het verbieden van cookies onvoldoende is om dit te stoppen. Er zijn verscheidene alternatieve technieken beschikbaar. Deze worden in deze thesis uitvoerig met elkaar vergeleken, en hieruit blijkt dat browser fingerprinting de meest geschikte methode is. Deze methode werd ook effectief getest, en hieruit blijkt dat ondanks de cookiewet een groot deel van de bezoekers van een webpagina nog steeds herkend kan worden, en dus ook gevolgd kan worden over meerdere websites.