Electronic Frontier Foundation

Di
Problemi
Il nostro lavoro
Agire
Utensili
Donare
Ricerca

FLoC di Google è un'idea terribile

Di Bennett Cyphers
3 marzo 2021
Una donna guardata dietro uno specchio unidirezionale

Il cookie di terze parti sta morendo e Google sta cercando di crearne la sostituzione.

Nessuno dovrebbe piangere la morte del biscotto come lo conosciamo. Per più di due decenni, il cookie di terze parti è stato il fulcro di un'oscura, squallida multimiliardaria industria di sorveglianza pubblicitaria sul Web; l'eliminazione graduale dei cookie di tracciamento e di altri identificatori persistenti di terze parti è attesa da tempo. Tuttavia, poiché le fondamenta si spostano al di sotto dell'industria pubblicitaria, i suoi maggiori attori sono determinati a rimettersi in piedi.

Google sta guidando la carica per sostituire i cookie di terze parti con una nuova suite di tecnologie per indirizzare gli annunci sul Web. E alcune delle sue proposte mostrano che non ha imparato le giuste lezioni dal continuo contraccolpo al modello di business della sorveglianza. Questo post si concentrerà su una di queste proposte, Federated Learning of Cohorts (FLoC) , che è forse la più ambiziosa e potenzialmente la più dannosa.

FLoC vuole essere un nuovo modo per fare in modo che il tuo browser esegua la profilazione che i tracker di terze parti erano soliti fare da soli: in questo caso, ridurre la tua recente attività di navigazione in un'etichetta comportamentale, per poi condividerla con siti web e inserzionisti. La tecnologia eviterà i rischi per la privacy dei cookie di terze parti, ma ne creerà di nuovi nel processo. Può anche esacerbare molti dei peggiori problemi di non privacy con gli annunci comportamentali, tra cui la discriminazione e il targeting predatorio.

Il passo di Google per i sostenitori della privacy è che un mondo con FLoC (e altri elementi della " sandbox della privacy ") sarà migliore del mondo che abbiamo oggi, dove i data broker e i giganti dell'ad-tech tracciano e profilano impunemente. Ma quell'inquadratura si basa su una falsa premessa che dobbiamo scegliere tra "vecchio tracciamento" e "nuovo tracciamento". Non è né-o. Invece di reinventare la ruota di tracciamento, dovremmo immaginare un mondo migliore senza la miriade di problemi degli annunci mirati.

Siamo a un bivio. Dietro di noi c'è l'era dei cookie di terze parti, forse il più grande errore del Web. Davanti a noi ci sono due possibili futuri.

In uno, gli utenti possono decidere quali informazioni condividere con ogni sito con cui scelgono di interagire. Nessuno deve preoccuparsi che la loro navigazione passata venga ostacolata o sfruttata per manipolarli quando apriranno una scheda.

Nell'altro, il comportamento di ogni utente lo segue di sito in sito come un'etichetta, imperscrutabile a colpo d'occhio ma ricco di significato per chi sa. La loro storia recente, distillata in pochi pezzi, è “democratizzata” e condivisa con decine di attori anonimi che prendono parte al servizio di ogni pagina web. Gli utenti iniziano ogni interazione con una confessione: ecco cosa ho combinato questa settimana, per favore trattami di conseguenza.

Gli utenti e i sostenitori devono rifiutare FLoC e altri tentativi fuorvianti di reinventare il targeting comportamentale. Imploriamo Google di abbandonare FLoC e reindirizzare i suoi sforzi verso la creazione di un Web veramente user-friendly.
Cos'è FLoC?

Nel 2019, Google ha presentato Privacy Sandbox , la sua visione per il futuro della privacy sul Web. Al centro del progetto c'è una suite di protocolli senza cookie progettati per soddisfare la miriade di casi d'uso che i cookie di terze parti attualmente forniscono agli inserzionisti. Google ha portato le sue proposte al W3C, l'ente di elaborazione degli standard per il Web, dove sono state discusse principalmente nel Web Advertising Business Group , un organismo composto principalmente da fornitori di tecnologia pubblicitaria. Nei mesi successivi, Google e altri inserzionisti hanno proposto decine di norme tecniche di uccelli a tema: PIGIN , TORTORA , SPARROW undefined undefined undefined undefined undefined undefined undefined undefined undefined undefined

FLoC è progettato per aiutare gli inserzionisti a eseguire il targeting comportamentale senza cookie di terze parti. Un browser con FLoC abilitato raccoglie informazioni sulle abitudini di navigazione dell'utente, quindi le utilizza per assegnare l'utente a una "coorte" o gruppo. Gli utenti con abitudini di navigazione simili, per qualche definizione di "simile", verrebbero raggruppati nella stessa coorte. Il browser di ogni utente condividerà un ID di coorte, indicando a quale gruppo appartengono, con siti web e inserzionisti. Secondo la proposta, almeno qualche migliaio di utenti dovrebbero appartenere a ciascuna coorte (anche se non è una garanzia).

Se suona denso, pensaci in questo modo: il tuo FLoC ID sarà come un breve riepilogo della tua attività recente sul Web.

La Google ha prova di concetto di utilizzato i domini dei siti visitati da ciascun utente come base per raggruppare le persone. Ha quindi utilizzato un algoritmo chiamato SimHash per creare i gruppi. SimHash può essere calcolato localmente sulla macchina di ogni utente, quindi non è necessario un server centrale per raccogliere dati comportamentali. Tuttavia, un amministratore centrale potrebbe avere un ruolo nell'applicazione delle garanzie sulla privacy. Per evitare che qualsiasi coorte sia troppo piccola (cioè troppo identificativa), Google propone che un attore centrale possa contare il numero di utenti assegnati a ciascuna coorte. Se alcuni sono troppo piccoli, possono essere combinati con altre coorti simili fino a quando non sono rappresentati abbastanza utenti in ciascuno di essi.

Affinché FLoC sia utile agli inserzionisti, la coorte di un utente rivelerà necessariamente informazioni sul suo comportamento.

Secondo la proposta, la maggior parte delle specifiche sono ancora nell'aria. La bozza delle specifiche afferma che l'ID di coorte di un utente sarà disponibile tramite Javascript, ma non è chiaro se ci saranno restrizioni su chi può accedervi o se l'ID sarà condiviso in altri modi. FLoC potrebbe eseguire il clustering in base agli URL o al contenuto della pagina invece che ai domini; potrebbe anche utilizzare un sistema basato sull'apprendimento federato (come suggerisce il nome FLoC) per generare i gruppi invece di SimHash. Inoltre, non è chiaro esattamente quante possibili coorti ci saranno. L'esperimento di Google ha utilizzato identificatori di coorte a 8 bit, il che significa che c'erano solo 256 coorti possibili. In pratica quel numero potrebbe essere molto più

alto; la documentazione suggerisce un ID di coorte a 16 bit composto da 4 caratteri esadecimali. Più coorti ci sono, più specifiche saranno; ID di coorte più lunghi significano che gli inserzionisti impareranno di più sugli interessi di ogni utente e avranno un tempo più facile per rilevare le impronte digitali.

Una cosa che viene specificata è la durata. Le coorti FLoC verranno ricalcolate su base settimanale, ogni volta utilizzando i dati della navigazione della settimana precedente. Ciò rende le coorti FLoC meno utili come identificatori a lungo termine, ma li rende anche misure più potenti di come gli utenti si comportano nel tempo.

Nuovi problemi di privacy

FLoC fa parte di una suite destinata a portare annunci mirati in un futuro che tutela la privacy. Ma il design principale prevede la condivisione di nuove informazioni con gli inserzionisti. Non sorprende che questo crei anche nuovi rischi per la privacy.

Impronte digitali

Il primo problema è il rilevamento delle impronte digitali. Il fingerprinting del browser è la pratica di raccogliere molte informazioni discrete dal browser di un utente per creare un identificatore univoco e stabile per quel browser. di EFF Il progetto Cover Your Tracks dimostra come funziona il processo: in poche parole, più il tuo browser appare o si comporta in modo diverso da quello degli altri, più è facile rilevare le impronte digitali.

Google ha promesso che la stragrande maggioranza delle coorti FLoC comprenderà migliaia di utenti ciascuna, quindi un ID di coorte da solo non dovrebbe distinguerti da poche migliaia di altre persone come te. Tuttavia, ciò offre ancora alle impronte digitali un enorme vantaggio. Se un tracker inizia con la tua coorte FLoC, deve solo distinguere il tuo browser da poche migliaia di altri (piuttosto che da poche centinaia di milioni). In termini di teoria dell'informazione, le coorti FLoC conterranno diversi bit di entropia, fino a 8 bit, nella prova di Google per la prova del concetto. Questa informazione è ancora più potente dato che è improbabile che sia correlata con altre informazioni che il browser espone. Ciò renderà molto più facile per i tracker mettere insieme un'impronta digitale unica per gli utenti FLoC.

Google ha riconosciuto che questa è una sfida, ma si è impegnata a risolverla come parte del più ampio piano di "budget per la privacy" che deve affrontare il rilevamento delle impronte digitali a lungo termine. Risolvere il rilevamento delle impronte digitali è un obiettivo ammirevole e la sua proposta è una strada promettente da perseguire. Ma secondo le domande frequenti , quel piano è "una proposta in una fase iniziale e non ha ancora un'implementazione del browser". Nel frattempo, Google inizierà a testare FLoC già questo mese .

Il rilevamento delle impronte digitali è notoriamente difficile da fermare. Browser come Safari e Tor si sono impegnati in guerre di logoramento di anni contro i tracker, sacrificando ampie parti dei propri set di funzionalità al fine di ridurre le superfici di attacco delle impronte digitali. La mitigazione delle impronte digitali generalmente comporta l'eliminazione o la limitazione di fonti di entropia non necessarie, che è ciò che è FLoC. Google non dovrebbe creare nuovi rischi relativi alle impronte digitali fino a quando non avrà capito come affrontare quelli esistenti.

Esposizione cross-contesto

Il secondo problema è meno facilmente spiegabile: la tecnologia condividerà nuovi dati personali con tracker che possono già identificare gli utenti. Affinché FLoC sia utile agli inserzionisti, la coorte di un utente rivelerà necessariamente informazioni sul suo comportamento.

La pagina Github del progetto affronta questo problema in anticipo:

Questa API democratizza l'accesso ad alcune informazioni sulla cronologia di navigazione generale di un individuo (e quindi, gli interessi generali) a qualsiasi sito che opta per essa. … I siti che conoscono le PII di una persona (ad esempio, quando le persone accedono utilizzando il proprio indirizzo e-mail) potrebbero registrare e rivelare la loro coorte. Ciò significa che le informazioni sugli interessi di un individuo possono eventualmente diventare pubbliche.

Come descritto sopra, le coorti FLoC non dovrebbero funzionare da sole come identificatori. Tuttavia, qualsiasi azienda in grado di identificare un utente in altri modi, ad esempio offrendo servizi di "accesso con Google" ai siti su Internet, sarà in grado di collegare le informazioni apprese da FLoC al profilo dell'utente.

Due categorie di informazioni possono essere esposte in questo modo:

Informazioni specifiche sulla cronologia di navigazione. I tracker possono essere in grado di decodificare l'algoritmo di assegnazione della coorte per determinare che qualsiasi utente che appartiene a una coorte specifica probabilmente o sicuramente abbia visitato siti specifici.

Informazioni generali su dati demografici o interessi. Gli osservatori possono apprendere che, che i in generale , è molto probabile membri di una specifica coorte siano un specifico tipo di persona. Ad esempio, una particolare coorte può rappresentare in modo eccessivo gli utenti giovani, donne e neri; un'altra coorte, elettori repubblicani di mezza età; un terzo, giovani LGBTQ +.

Ciò significa che ogni sito che visiti avrà una buona idea del tipo di persona con cui sei al primo contatto, senza dover fare il lavoro di rintracciarti sul web. Inoltre, poiché la tua coorte FLoC si aggiornerà nel tempo, i siti che possono identificarti in altri modi saranno anche in grado di monitorare come cambia la tua navigazione. Ricorda, una coorte FLoC non è altro e niente di meno che un riepilogo della tua recente attività di navigazione.

Dovresti avere il diritto di presentare diversi aspetti della tua identità in diversi contesti. Se visiti un sito per informazioni mediche, potresti fidarti di esso con informazioni sulla tua salute, ma non c'è motivo per cui abbia bisogno di sapere qual è la tua politica. Allo stesso modo, se visiti un sito web di vendita al dettaglio, non dovrebbe essere necessario sapere se hai letto di recente sul trattamento per la depressione. FLoC erode questa separazione di contesti e presenta invece lo stesso riassunto comportamentale a tutti coloro con cui interagisci.

Oltre la privacy

FLoC è progettato per prevenire una minaccia molto specifica: il tipo di profilazione individualizzata abilitata dagli identificatori cross-context oggi. L'obiettivo di FLoC e di altre proposte è evitare di consentire ai tracker di accedere a informazioni specifiche che possono legare a persone specifiche. Come abbiamo mostrato, FLoC può effettivamente aiutare i tracker in molti contesti. Ma anche se Google è in grado di iterare il suo design e prevenire questi rischi, i danni della pubblicità mirata non si limitano alle violazioni della privacy. L'obiettivo principale di FLoC è in contrasto con altre libertà civili.

Il potere di mirare è il potere di discriminare. Per definizione, gli annunci mirati consentono agli inserzionisti di raggiungere alcuni tipi di persone escludendone altri. Un sistema di targeting può essere utilizzato per decidere chi può vedere annunci di lavoro o offerte di prestito con la stessa facilità con cui pubblicizza scarpe.

Nel corso degli anni, il meccanismo della pubblicità mirata è stato spesso utilizzato per lo sfruttamento , la discriminazione e il danno . La possibilità di indirizzare gli annunci alle persone in base all'etnia, alla religione, al sesso, all'età o alle capacità consente annunci discriminatori per

lavoro, alloggio e credito. Il targeting basato sulla storia del credito, o sulle caratteristiche sistematicamente associate ad esso, consente annunci predatori per prestiti ad alto interesse. Il targeting basato su dati demografici, posizione e affiliazione politica aiuta i fornitori di disinformazione politicamente motivata e soppressione degli elettori. Tutti i tipi di targeting comportamentale aumentano il rischio di truffe convincenti .

Invece di reinventare la ruota di tracciamento, dovremmo immaginare un mondo migliore senza la miriade di problemi degli annunci mirati.

Google, Facebook e molte altre piattaforme pubblicitarie cercano già di frenare determinati usi delle loro piattaforme di targeting. Google, ad esempio, limita la capacità degli inserzionisti di indirizzare gli annunci a persone in " categorie di interessi sensibili ". Tuttavia, questi sforzi spesso falliscono; attori determinati possono solitamente trovare soluzioni alternative alle restrizioni a livello di piattaforma su determinati tipi di targeting o determinati tipi di annunci .

Anche con il potere assoluto su quali informazioni possono essere utilizzate per indirizzare chi, le piattaforme sono troppo spesso incapaci di prevenire l'abuso della loro tecnologia. Ma FLoC utilizzerà un algoritmo non supervisionato per creare i suoi cluster. Ciò significa che nessuno avrà il controllo diretto sul modo in cui le persone sono raggruppate. Idealmente (per gli inserzionisti), FLoC creerà gruppi che hanno comportamenti e interessi significativi in comune. Ma il comportamento online è collegato a tutti i tipi di caratteristiche sensibili: dati demografici come sesso, etnia, età e reddito; Tratti della personalità "big 5" ; anche la salute mentale undefined

Google ha proposto di poter monitorare gli output del sistema per verificare eventuali correlazioni con le sue categorie sensibili. Se rileva che una particolare coorte è troppo strettamente correlata a un particolare gruppo protetto, il server amministrativo può scegliere nuovi parametri per l'algoritmo e dire ai browser degli utenti di raggrupparsi di nuovo.

Questa soluzione suona sia orwelliana che sisifea. Per monitorare il modo in cui i gruppi FLoC sono correlati alle categorie sensibili, Google dovrà eseguire controlli massicci utilizzando dati su razza, sesso, religione, età, salute e situazione finanziaria degli utenti. Ogni volta che trova una coorte che si correla troppo fortemente lungo uno qualsiasi di questi assi, dovrà riconfigurare l'intero algoritmo e riprovare, sperando che nella nuova versione non siano implicate altre “categorie sensibili”. Questa è una versione molto più difficile del problema che sta già tentando, e spesso non riesce , di risolvere.

In un mondo con FLoC, potrebbe essere più difficile indirizzare gli utenti direttamente in base all'età, al sesso o al reddito. Ma non sarà impossibile. I tracker con accesso a informazioni ausiliarie sugli utenti saranno in grado di apprendere cosa "significano" i raggruppamenti FLoC, che tipo di persone contengono, attraverso l'osservazione e la sperimentazione. Coloro che sono determinati a farlo saranno comunque in grado di discriminare. Inoltre, questo tipo di comportamento sarà più difficile da controllare per le piattaforme di quanto non lo sia già. Gli inserzionisti con cattive intenzioni avranno una plausibile negabilità: dopo tutto, non prendono direttamente di mira le categorie protette, ma raggiungono le persone in base al comportamento. E l'intero sistema sarà più opaco per utenti e regolatori.
Google, per favore non farlo

Abbiamo scritto di FLoC e dell'altra serie iniziale di proposte quando sono state introdotte per la prima volta , definendo FLoC "l'opposto della tecnologia per la tutela della privacy". Speravamo che il processo di standardizzazione avrebbe fatto luce sui difetti fondamentali di FLoC, inducendo Google a riconsiderare la possibilità di portarlo avanti. In effetti, diversi problemi sulla pagina ufficiale Github sollevano le esatte stesse preoccupazioni che si evidenziano qui. Tuttavia, Google

ha continuato a sviluppare il sistema, lasciando i fondamentali quasi invariati. Ha iniziato a proporre FLoC agli inserzionisti, vantando che FLoC è un sostituto " efficace al 95% " per il targeting basato sui cookie. E a partire da Chrome 89, rilasciato il 2 marzo undefined undefined undefined

Non commettere errori, se Google porta a termine il suo piano di implementare FLoC in Chrome, probabilmente darà a tutte le "opzioni" coinvolte. Il sistema sarà probabilmente opt-in per gli inserzionisti che ne trarranno vantaggio e opt-out per gli utenti che rischiano di essere feriti. Google lo pubblicizzerà sicuramente come un passo avanti per la "trasparenza e il controllo dell'utente", sapendo benissimo che la stragrande maggioranza dei suoi utenti non capirà come funziona FLoC e che pochissimi faranno di tutto per disattivarlo. Si darà una pacca sulla spalla per aver inaugurato una nuova era privata sul Web, libera dal malvagio cookie di terze parti, la tecnologia che Google ha aiutato a estendere ben oltre la sua durata di conservazione, guadagnando miliardi di dollari nel processo .

Non deve essere così. Le parti più importanti della sandbox per la privacy, come l'eliminazione di identificatori di terze parti e la lotta al rilevamento delle impronte digitali, cambieranno davvero il Web in meglio. Google può scegliere di smantellare la vecchia impalcatura per la sorveglianza senza sostituirla con qualcosa di nuovo e unicamente dannoso.

Rifiutiamo decisamente il futuro di FLoC. Questo non è il mondo che vogliamo, né quello che gli utenti meritano. Google ha bisogno di imparare le lezioni corrette dall'era del monitoraggio di terze parti e progettare il suo browser in modo che funzioni per gli utenti, non per gli inserzionisti.

Nota: abbiamo contattato Google per verificare alcuni fatti presentati in questo post, nonché per richiedere ulteriori informazioni sulla prossima versione di prova di Origin. Non abbiamo ricevuto risposta al momento della pubblicazione.