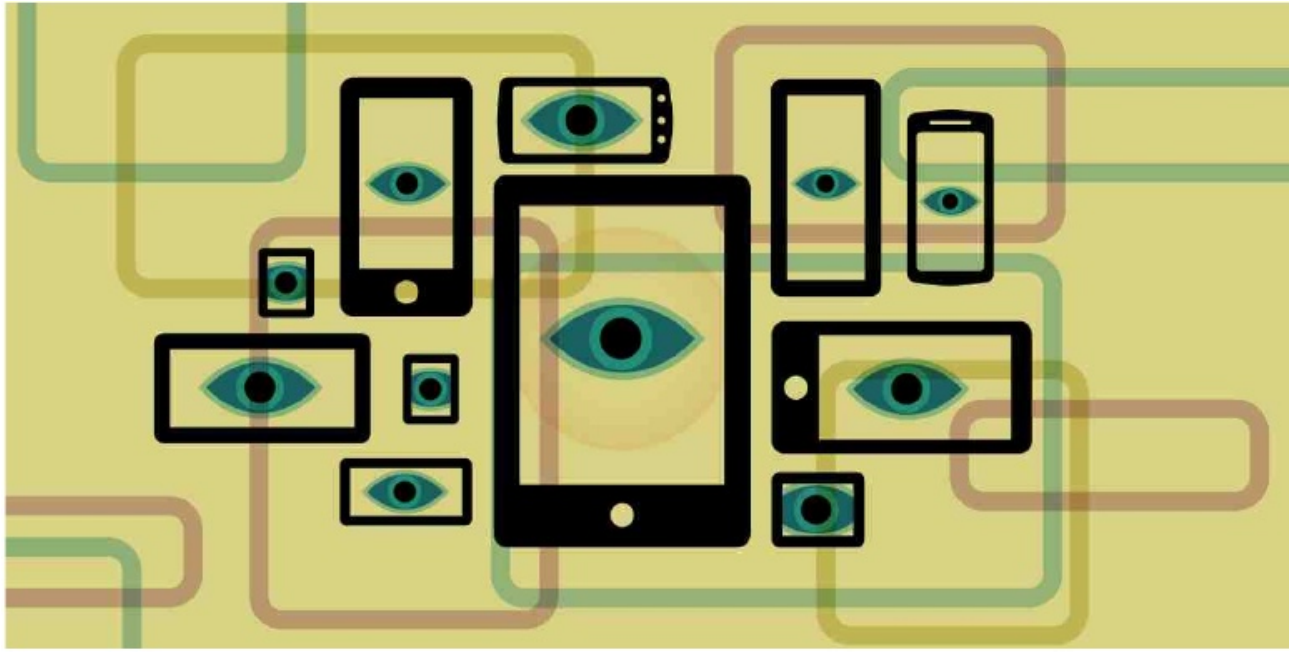
Az adatok „anonimitása” mítoszának leleplezése

PCLinuxOS Magazine – 2023. december

Írta:Paige Collings
Electronic Frontier Foundation



Az életünkről ma minden digitálisan rögzítenek és tárolnak valahol. Felvesznek minden bankkártyás vásárlást, az egészségi diagnózisokat, milyen zenét és könyveket kedvelünk és arra használják, hogy megjósolják, mi tetszik és mi nem, és – végső soron – kik vagyunk.

Gyakran ez a tudtunk, vagy belegyezésünk nélkül történik. A cégek által az online-viselkedésünkről a személyes adatokat megdöbbentő profitért adják el, ami arra ösztönzi az online-szereplőket, hogy minél többet gyűjtsenek össze. Minden egyes mozdulat az egérrel, vagy gesztus a képernyőn lekövethető és eladható hirdető cégeknek és adatbrókereknek, akik őket kiszolgálják.

A cégek igyekezetükben, hogy elfogadhatóvá tegyék a széles körű megfigyelést, gyakran hangoztatják, az adatokat személytelenné teszik. Ezzel feltételezhetően eltávolítják az összes személyes vonatkozást (mint a nevet) az adatból (így egy névtelen ember vett egy adott gyógyszert adott helyen és időben). A személyes adatok gyűjthetőek úgy is, hogy több ember adatait kombinálva eltávolítják a személyt azonosító információkat, így védve a felhasználó személyes terét.

A cégek néha azt hangoztatják, az adataink „anonimizáltak”, arra utalva, hogy ez egy egyirányú folyamat, az adathalmazból a személy visszabontása és azonosítása kizárt. De ez lehetetlen – az anonim adatok ritkán ilyenek. Ahogy Matt Blaze professzor, a kriptográfia és adatbiztonság szakértője röviden összegezte: „valami, ami anonimnak tűnik, inkább gyakran, mint ritkán, nem anonim, még ha legjobb szándékkal is úgy tervezték”.

Anonimizálás … és visszaazonosítás?

A személyes adatok az azonosítási skála különböző szintjén foglalnak helyet. A csúcson vannak azok, amikkel emberek közvetlenül beazonosíthatóak, mint a név, vagy a személyi szám, ezeket a „direkt azonosítóknak” nevezzük. A következő szinten az „indirekt azonosítók” jönnek, azok az információk, amik áttételesen kapcsolódnak személyekhez, mint a telefonszám és e-mail-cím. Ezt követik az adatok, amik több személyhez is kapcsolhatóak, mint a kedvenc étterem, vagy mozi. A spektrum másik végén vannak azok az adatok, amik egyetlen konkrét személyhez sem kapcsolhatóak – mint az összesített

népszámlálási adatok és az egyénekhez nem köthető adatok, mint az időjárás-jelentés.

Az adatok anonimizálása általában két módon történhet. Egyik, hogy törölhetik a személyi azonosítókat, mint a nevet, vagy a TB-számot. Másik, hogy egyéb személyes adatcsoportokat módosítanak – kitakarják a bankszámlaszámot. Például, az USA törvény a Betegbiztosítás hordozhatóságáról és érvényességéről (HIPAA) előírja, hogy a kitakart adatoknál csak a zip-kód első három számjegyét lehet jelenteni.

Ugyanakkor, a gyakorlatban a személytelenítéshez nem csupán az azonosítási adatokat kell eltávolítani, hanem azokat is amikkel, ha ismert más információkkal kombinálják, be lehet téged azonosítani téged. Íme egy példa:

gondolj át, hány embernek azonos a ZIP-kódja veled, vagy a postai irányítószáma;
ezután gondold meg, hogy ezek közül hánynak van veled egy napon a születésnapja
most gondolj bele, hogy egyszerre hánynak ugyanaz a születésnapja, a ZIP-kódja és neme, mint neked.

Egy alapvető tanulmány szerint, ez a három jellemző elegendő az amerikai lakosság 87%-ának azonosítására. Egy másik tanulmány szerint is az USA lakosságának 63 %-a azonosítható ezzel a három ténnyel.

Nem bízhatunk a cégek önmegtartóztatásában. A személyes adatokból nyerhető pénzügyi nyereség és üzleti haszon gyakran felülírja a személyiség védelmét és anonimitást. A tényleges személyt visszaazonosítva (direkt azonosítók) és a személyes preferenciákkal (indirekt azonosítók) a cégek további profitot nyerhetnek az érzékeny adatainkat felhasználva. Például, egy weblap, ami „anonim” felhasználótól kér triviális információkat, alkalmas arra, hogy az információkból egyedi profilt alkossonbárkiről.

Tartózkodási hely megfigyelése

A rendszer gyakorlati működésének megértéséhez nézzük a helyadatokat. Ide tartoznak a mobil telefonos applikációk által a gyűjtött tartózkodási hely adatai: kezdve a bevásárlásodról a helyi üzletben, egészen egy egészségközpontban, egy bevándorlási klinikán, vagy egy tüntetéstelőkészítőtalálkozón való részvételedig. Az eszközeinken megtalálható helyadatok elegendően pontosak ahhoz, hogy a rendőrség egy bűntény helyszínéhez kössön valakit és a bíróságok, hogy ezen bizonyítékok alapján elítéljék. Sőt mi több, a kormány által gyűjtött adatokat hivatali dolgozók jogtalanul felhasználhatják, bűnözők és külföldi kormányok ellophatják és előre megjósolhatatlan módon új aljas célokból visszaélhetnek vele. Túlságoan gyakran történik, hogy az ilyen kifinomult megfigyelés kiemelten célozza a színesbőrüeket.

Gyakorlatban ez azt jelenti, hogy nincs mód a helyinformációk függetlenítésére a személyektől, mivel azok egymaguk is azonosítják a személyt. És még ha azt is mondják, hogy az adatokat anonimizálták, a személy visszakövetése lehetséges azok összevetésével más, nyíltan elérhető adatokkal, mint a választói névjegyzékek, vagy adatbrókerektől vásárolt információk. Egy 2013-as tanulmány szerint az személyek 50%-át egyedileg azonosítani lehet két, véletlenszerűen kiválasztott idő és helyadat alapján.

Jól csinálva lehet úgy helyadatokat gyűjteni, hogy közben megőrzik a személyek magánélethez való jogát, ha személytelenül gyűjtik a viselkedési adatokat ahelyett, hogy részletes idővonalat állítanának fel a helyadatokból. Például, egy ilyen gyűjtés, ami megmondja, hány telefon jelentette a helyzetét egy adott városban a múlt hónapban, de telefonszámok és más adatok nélkül, amivel esetleg közvetlenül személyhez köthetnék. Ugyanakkor, gyakran elvárják a szakértőktől, hogy az adatgyűjtés során finomított adatcsomagokat adjanak, amik sokkal hasznosabbak lehetnek bizonyos döntéshozók számára, de amelyek egyben sokkal inkább kockáztatják az emberek személyes adatainak kiszivárgását.

Mi több, a harmadik féltől származó helykövetők zömét úgy tervezték, hogy valós személyeket profilozzanak. Vagyis, amikor a nyomkövető adatot nyer, kell valami, amivel egy adott személyhez kapcsolhatja. Ez indirekt módon történhet úgy, hogy a begyűjtött adatot összekapcsolja egy adott eszközzel, vagy böngészővel, ami később adott

személyhez, vagy csoporthoz kapcsolható, pl. háztartáshoz. A nyomkövetők olyan egyedi azonosítókat is használhatnak, mint mobil hirdetés-azonosítók, vagy sütik, célzott üzenetek céljából. És az „anonim” személyi adatokból képzett profilok majdnem mindig visszakövethetők valós személyekhez – ebben a lakhelyükkel, olvasmányaikkal és vásárlásaikkal.

Az személyi adatainkkal foglalkozó adatbókerek szempontjából az adataink lehetnek hasznosak a profit szempontjából, vagy tényleg anonimak, de a mindkettő nem lehet egyszerre. Az EFF régóta ellenzi a helyadat-felderítő programokat, amik az életünket nyitott könyvvé tehetik a rendőrség, a célzott hirdetők, a személyiségtolvajok és a zavarosban halászók számára. Régóta hangoztatjuk, hogy a telefonokat anonimizálni kellene.

A közrend szempontjából kritikus, hogy a személyi adatok biztonságát ne áldozzuk fel a cégek zsebeinek megtöltése érdekében. Minden adatmegosztási rendszer kialakítása szempontjából a