Allegato Tecnico

Indice

Il presente allegato tecnico descrive i trasferimenti sistematici avviati automaticamente dai siti web che utilizzano le API di Google Fonts, nonché le problematiche per la privacy e la sicurezza dei visitatori che tale strumento determina.

L'allegato descrive inoltre l'output dell'osservatorio automatico distribuito di Monitora PA, che include le evidenze raccolte dei trasferimenti in corso.

Google Fonts

1. Google Fonts è un servizio per webmaster fornito gratuitamente da Google e composto di:
   - una libreria di font con licenza libera
   - un sito web interattivo per navigare tale libreria
   - un servizio HTTP che fornisce un'API di tipo ReST [1] per incorporare tali font direttamente dai server di Google .
2. Sebbene le licenze con cui i font sono distribuiti ne permettano l'installazione sul server del sito che ne fa uso, l'utilizzo dell'API di Google viene spesso preferito per scaricare su Google il costo della banda necessaria alla loro distribuzione, nonché la responsabilità del loro aggiornamento.
3. Esistono molti modi di effettuare tale inclusione nelle pagine HTML inviate al browser dei visitatori, fra cui:
   - l'inclusione diretta in un tag `<link rel="stylesheet">` di un riferimento a https://fonts.googleapis.com/css specificando in query string i font, gli effetti e i parametri desiderati [2]
   - l'introduzione all'interno di CSS locali del sito di direttive `@import` [3] che puntino a indirizzi web costruiti secondo la stessa specifica
   - l'introduzione all'interno di CSS locali del sito di direttive `@font-face` [4] che facciano riferimento a file TTF, WOFF o WOFF2 serviti dal server https://fonts.gstatic.com
4. Durante lo scaricamento di suddette risorse, Google istruisce il browser affinché mantenga la risposta in una cache locale, dedicata allo specifico sito. [5] Nel caso dei CSS, la direttiva `Cache-Control` impone al browser di mantenere il dato in cache per ventiquattro ore (`max-age=86400`) dopo le quali la direttiva `stale-while-revalidate=604800` impone al browser di rivalidare la cache, richiedendo nuovamente la risorsa in una richiesta asincrona nei sei giorni successivi.

[1] https://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm
[2] https://developers.google.com/fonts/docs/getting_started
[3] https://developer.mozilla.org/en-US/docs/Web/CSS/@import
[4] https://developer.mozilla.org/en-US/docs/Web/CSS/@font-face
[5] https://developer.chrome.com/blog/http-cache-partitioning/#how-will-cache-partitioning-affect-chromes-http-cache

L'identificazione del visitatore

5. La prima richiesta inviata ai server di Google per ottenere i CSS o i font da utilizzare durante l'elaborazione della pagina, nonché tutte le richieste inviate per convalidare o aggiornare la versione in cache delle stesse risorse una volta al giorno, causano il trasferimento di diversi dati personali del visitatore a favore di Google LLC, fra cui:
    - indirizzo IP
    - data, ora e timezone della navigazione sul sito
    - stringa identificativa dello User Agent
    - lingue note all'utente, attraverso l'header HTTP Accept-Language
    - data dell'ultima visita, attraverso l'header HTTP If-Modified-Since
    - l'indirizzo della pagina che incorpora la risorsa, attraverso l'header HTTP Referer
    - eventuali cookie precedentemente impostati
6. Questi dati iniziali, il cui trasferimento automatico è determinato dalle specifiche dei protocolli TCP/IP e HTTP, sono sufficienti per Google a identificare il visitatore nella maggioranza dei casi.
7. Per esempio, ogni qualvolta l'utente utilizzi, in contemporanea al sito web su cui è incorporata una risorsa fornita da Google, uno dei servizi o delle App gratuite offerte da Google che prevedono l'autenticazione, sarà sufficiente confrontare l'IP o il cookie identificativo inviati alla prima richiesta con quelli degli utenti correntemente autenticati per stabilirne l'identità.
8. Inoltre diversi altri servizi di Google LLC come YouTube, reCAPTCHA o Google Maps, una volta incorporati in una pagina web, possono determinare via JavaScript l'inclusione dei Font da server di Google, avviando sistematicamente suddetti trasferimenti.
9. La straordinaria quantità di informazioni raccolte all'insaputa degli utenti su milioni di applicazioni mobili e siti web rende estremamente facile, per Google, l'utilizzo di un qualsiasi insieme di dati personali anche non identificativi, con funzione di identificativi di fatto[6], deanonimizzando i dati ad essi associati[7] per svelare l'identità del soggetto da cui sono stati emessi e perfezionarne la profilazione, in particolare attraverso l'header HTTP Referer, che permette di associare all'utente i dati deducibili dal contenuto della pagina visitata. Si pensi ad esempio ad una persona che visiti la pagina di un ospedale dedicata ad una determinata visita specialistica.
10. Dunque per Google è sempre possibile identificare i cittadini europei e tracciarne l'attività, le opinioni e gli interessi su tutti i siti web (e tutte

[6] https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf
[7] https://www.csee.umbc.edu/~kunliu1/p3dm08/proceedings/2.pdf

le App) che utilizzino i suoi servizi, per ricondurne poi le registrazioni all'identità personale. In altri termini, anche in totale assenza dell'IP, tutti i dati raccolti attraverso le risorse incorporate da Google rimangono riconducibili all'interessato, restando dati personali soggetti al GDPR.

11. Lo scaricamento dei font o dei CSS da server di Google Fonts comporta il trasferimento sistematico dei dati personali degli utenti verso Google LLC, avviato automaticamente prima che il visitatore possa esprimere qualsivoglia forma di consenso ad ogni prima visita della giornata.

12. Google dichiara esplicitamente di registrare "i dettagli della richiesta HTTP, inclusi il timestamp, l'URL richiesto e tutte le intestazioni HTTP (incluse la stringa referrer e user agent) fornite in relazione all'utilizzo dell'API CSS" [8], ma non è possibile individuare nella Privacy Policy [9] o nella pagina dedicata alla Retention Policy [10] indicazioni esatte circa i tempi di mantenimento e gli utilizzi dei dati raccolti specificatamente tramite Google Fonts.

13. Inoltre Google riconosce esplicitamente di includere i dati personali raccolti attraverso l'utilizzo di Google Fonts nelle analisi effettuate tramite Google Analytics [11], rendendoli anche parzialmente accessibili al pubblico in una pagina dedicata[12].

La collocazione geografica dei data center

14. Google dichiara di elaborare presso i suoi data center europei i dati dei cittadini europei.

15. Tuttavia il software utilizzato per tale elaborazione è controllato centralmente dagli USA come dimostrato dal blackout globale del dicembre 2020 e dalla sua simultanea risoluzione in tutto il pianeta.

16. Il controllo sul software che elabora i dati dei cittadini europei implica ovviamente la possibilità di accedere ai dati stessi. Ad esempio Google potrebbe essere costretta, nei termini delle già citate normative USA, a inviare presso uno o più dei suoi data center europei aggiornamenti in grado di prelevare i dati di interesse di una agenzia governativa e rimuovere ogni evidenza di tale "data breach" prima del riavvio dei servizi.

[8] https://developers.google.com/fonts/faq?hl=it#what_does_using_the_google_fonts_api_mean_for_the_privacy_of_my_users

[9] https://policies.google.com/privacy#inforetaining

[10] https://policies.google.com/technologies/retention

[11] https://developers.google.com/fonts/faq#what_does_using_the_google_fonts_api_mean_for_the_privacy_of_my_users

[12] https://fonts.google.com/analytics

Le evidenze raccolte dall'osservatorio di Monitora PA

17. L'osservatorio automatico distribuito di Monitora PA analizza una serie di siti web per identificare problemi di non conformità alla normativa, producendo una serie di file TSV (tab-separated values) contenenti le evidenze delle problematiche raccolte durante l'analisi.

18. L'allegato `103-google-fonts.tsv` contiene copia dell'output relativo a Google Fonts, riportando le evidenze raccolte sull'utilizzo delle API di Google Fonts o sullo scaricamento di font dai suoi server (con il conseguente trasferimento sistematico di dati personali ad esso conseguente)

19. Il formato del file TSV, attentamente studiato per massimizzare la possibilità dei cittadini di utilizzare personalmente l'osservatorio di Monitora PA [13], è costituito da sei colonne dalla semantica posizionale. Nell'ordine da sinistra a destra le colonne contengono:

    1. Identificativo AgID-IPA del titolare del trasferimento (come presente nella seconda colonna del file CSV da AgID [14])
    2. Tipologia di automatismo verificato (nel caso in questione, Web)
    3. Indirizzo dell'automatismo verificato
    4. Data e ora della verifica
    5. Stato di completamento della verifica
       `"1"` significa che la verifica è stata completata con successo
       `"0"` significa che non è stato possibile completare la verifica
    6. Evidenze riscontrate della problematica oggetto di verifica:
       - se la colonna è vuota, il problema oggetto di analisi non è stato individuato sul sito dall'osservatorio automatico
       - se la colonna non è vuota, contiene le evidenze del problema individuate dall'osservatorio (o dati sull'errore che ha impedito la verifica, quando la colonna precedente contiene `"0"`)

[13] https://github.com/MonitoraPA/monitorapa/blob/main/ARCHITETTURA.md
[14] https://indicepa.gov.it/ipa-dati/dataset/enti

Alba, 26 settembre 2022

Firme

Giacomo Tesio

GIACOMO TESIO

Massimo Maria Ghisalberti

MASSIMO MARIA GHISALBERTI