Allegato Tecnico

Indice

Il presente allegato tecnico descrive i trasferimenti sistematici che sono avviati automaticamente da siti web che incorporano "risorse di e a richiesta" verso alcune aziende di Stati Uniti e Russia, nonché le problematiche per la privacy e la sicurezza dei visitatori che questo determina.

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

Content Delivery Network

1. Sebbene le risorse di cui è composta una pagina web siano spesso distribuite con licenze open source, permettendone una libera installazione sul server di chi ne volesse fare uso, alcuni web master utilizzano dei Content Delivery Network (CDN) forniti da terze parti per trasferire su di essi il costo della banda necessaria al trasferimento e talvolta anche la responsabilità degli aggiornamenti.

2. Alcuni dei CDN oggetto della presente segnalazione non richiedono ai webmaster alcun pagamento. In cambio del servizio ottengono sistematicamente dati personali da tutti i visitatori dei siti su cui vengono incorporate le risorse da loro fornite.

3. I servizi oggetto della presente segnalazione che costituiscono dei Content Delivery Network o fornitori di questi e che vengono "interrogati" come tali dal Titolare del trattamento sono:

   - AWS (fornito da Amazon Web Services, Inc. - Infrastruttura Iaas, PaaS, SaaS)
   - CloudFront (fornito da Amazon, Inc.)
   - Azure (fornito da Microsoft Corporation - Infrastruttura PaaS, SaaS)
   - le diverse CDN offerte da Adobe, Inc. [1]
   - le diverse CDN offerte da Microsoft Corporation [2]
   - FontAwesome (fornito da Cloudflare, Inc.)
   - JSDeliver (fornito da Cloudflare, Inc.)
   - Unpkg (fornito da Cloudflare, Inc.)
   - CDNJS (fornito da Cloudflare, Inc.)
   - CloudflareCDN (fornito da Cloudflare, Inc.)
   - Google Hosted Libraries (fornito da Google LLC)
   - Akamai CDN (fornito da Akamai Technologies, Inc.)
   - JQuery (fornito da StackPath LLC)
   - Fastly
   - Yandex CDN (fornito da Yandex LLC)

Sistemi di Web Analytics

4. Un servizio di Web Analytics traccia e riporta le attività dei visitatori di un sito web al Titolare dello stesso, fornendo statistiche aggregate rispetto a innumerevoli dimensioni: demografiche, tecniche, economiche e comportamentali.

[1] l'elenco dei domini di Adobe, Inc. oggetto delle verifiche del nostro osservatorio è disponibile all'indirizzo https://github.com/MonitoraPA/monitorapa/blob/main/cli/check/browsing/hosts/500-adobe.hosts

[2] l'elenco dei domini di Microsoft Corporation oggetto delle verifiche del nostro osservatorio è disponibile all'indirizzo https://github.com/MonitoraPA/monitorapa/blob/main/cli/check/browsing/hosts/508-microsoft.hosts

5. Tramite tali software gli amministratori di un sito web possono sfruttare sofisticate analisi sui dati dei diversi visitatori senza però possedere la capacità di accesso a quelli "non pre-elaborati". Questi sono quelli effettivamente registrati dal fornitore del servizio a fronte di ogni singola azione effettuata dall'utente.

6. L'introduzione di un software di Web Analytics su un sito web comporta l'inserimento all'interno del codice HTML che sarà inviato al browser del visitatore, di un riferimento a un file o un frammento di codice di programma JavaScript.

7. Esistono molti modi di effettuare tale inclusione nelle pagine HTML poi inviate al browser dei visitatori, fra cui:

   - l'inclusione diretta in un tag `<script>` che determini lo scaricamento e l'esecuzione sul browser del codice di tracciamento.
   - l'introduzione di piccoli "snippet" (frammenti) di codice JavaScript indicati nella documentazione tecnica del servizio
   - l'abilitazione di un Web Analytics in altri servizi forniti da un diverso responsabile del trattamento.

8. In tutti i casi, l'inclusione di un sistema di Web Analytics nel codice HTML delle pagine inviate al browser dell'utente determina automaticamente l'invio di una richiesta HTTP ai server del fornitore di tale servizio.

9. I servizi di Web Analytics oggetto della presente segnalazione sono:

   - CloudflareInsights (fornito da Cloudflare, Inc.)
   - Google Analytics (fornito da Google LLC)

Social Network e Advertising

10. In ambito commerciale viene definito *Social Network* una piattaforma online che permette a diverse persone di instaurare e mantenere relazioni personali e/o commerciali, talune sono veicolate dalla piattaforma stessa, con altre persone che condividano interessi simili.

11. I Social Network, tipicamente gratuiti per coloro che vi inseriscono contenuti, basano il proprio modello di business direttamente sulla profilazione degli utenti attraverso le loro abitudini e le loro relazioni. I profili cognitivo-comportamentali vengono così usati per prevedere comportamenti a valle di determinati stimoli. Tali previsioni vengono rivendute, con modalità automatizzata, a chiunque sia interessato ad analizzare o alterare comportamenti e/o opinioni, sfruttandole per finalità che possono andare da quelle meramente commerciali sino a quelle politiche.

12. È importante osservare come ciò non riguardi solo i contenuti visualizzati dall'utente direttamente sul Social Network utilizzato in un dato

momento. Avvengono su tutti i siti web che ne incorporano direttamente o indirettamente i servizi, come quelli oggetto della presente segnalazione.

13. Tali sistemi possono essere integrati in vari modi all'interno di un sito web, ma tipicamente comportano il trasferimento di diversi dati personali e in varie modalità tramite le richieste HTTP necessarie al loro funzionamento. Ai dati personali inviati direttamente durante l'interazione con il servizio, si aggiungono tutti quelli attivamente inviati dai vari JavaScript innescati ed eseguiti dal browser per conto di tali servizi. A mero titolo esemplificativo, quando una pagina web incorpora video forniti tramite il dominio `www.youtube-nocookie.com`, sebbene non vengano installati nuovi cookie sul browser del visitatore, è possibile osservare delle chiamate di tipo POST HTTP verso l'indirizzo `https://www.youtube-nocookie.com/youtubei/v1/log_event?alt=json&key=<CHIAVE_UNIVOCA>` contenente diverse informazioni relative all'attività dell'utente.

14. I Social Network e i sistemi di Advertising oggetto della presente segnalazione sono:

    - Facebook (fornito da Meta Platforms, Inc.)
    - YouTube (fornito da Google LLC)
    - Google Double Click (fornito da Google LLC)
    - Moat Ads (fornito da Oracle Corporation)
    - Twitter, Inc.

Software as a Service **e altri servizi** ***in Cloud***[3]

15. Costituiscono Software as a Service (SaaS), ovvero software fornito come servizio, tutti quei sistemi che permettono accesso "online" all'utilizzo di un applicativo informatico, senza cioè prevederne l'installazione sul proprio hardware.

16. Analogamente alle piattaforme di Social Network e di Advertising precedentemente indicate, i SaaS possono effettuare ulteriori trasferimenti di dati personali raccogliendoli e/o inviandoli in varie modalità per finalità di profilazione. A mero titolo esemplificativo, la Privacy Policy accessibile dal link presente sull'interfaccia introdotta nelle pagine da parte di Google reCAPTCHA [4], dichiara esplicitamente l'uso dei dati raccolti per la profilazione dei visitatori a fini di pubblicità personalizzata.

17. I Software as a Service e i sistemi in Cloud oggetto della presente segnalazione sono:

    - Turnstile (fornito da Cloudflare, Inc.)
    - Google Maps (fornito da Google LLC)
    - Google Translate (fornito da Google LLC)

[3] https://www.iso.org/standard/60544.html
[4] https://www.google.com/intl/en/policies/privacy/

- Google Tag Manager (fornito da Google LLC)
- Google Search (fornito da Google LLC)
- Vimeo (fornito da Vimeo, Inc.)
- hCaptcha (fornito da Intuition Machines, Inc.)
- AddThis (fornito da Oracle Corporation)
- AddToAny

L'identificazione del visitatore

18. La prima richiesta inviata ai server dei servizi sopra indicati per ottenere gli script, i CSS, le immagini 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, causano il trasferimento di diversi dati personali del visitatore a favore dei fornitori del servizio, fra cui:

    - indirizzo IP
    - data, ora e timezone della navigazione sul sito
    - stringa identificativa dello User Agent
    - "impronta digitale" ("fingerprint" dedotto con varie tecniche) dello User Agent utilizzato
    - 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

19. Questi dati iniziali, il cui trasferimento automatico è determinato dalle specifiche dei protocolli HTTP e TCP/IP, sono sufficienti ai detti fornitori per identificare in modo univoco e nella maggioranza dei casi, il visitatore.

20. Per esempio, per stabilire l'identità di un utente sarebbe sufficiente che questo utilizzasse un sito web con una risorsa fornita da Google in contemporanea con una sua App gratuita che preveda una autenticazione. Basterà poi, il confronto dell'IP e/o del cookie identificativo inviati con quelli degli utenti attualmente connessi per una precisa identificazione.

21. La straordinaria quantità di informazioni raccolte all'insaputa degli utenti su milioni di applicazioni mobili e siti web, rende estremamente facile l'utilizzo di un qualsiasi insieme di dati personali anche non identificativi, con funzione di identificativi di fatto[5]. I dati deanonimizzati ad essi associati[6] sveleranno l'identità del soggetto da cui sono stati emessi e perfezioneranno la profilazione. Questo in particolare, attraverso l'header HTTP Referer che permette di associare all'utente i dati deducibili dal contenuto della pagina visitata. Si pensi per esempio a una persona

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

che visiti la pagina di un ospedale dedicata a una determinata visita specialistica.

22. Per le aziende fornitrici dei servizi sopra indicati, è dunque sempre possibile identificare i cittadini europei e tracciarne l'attività, le opinioni e gli interessi su tutti i siti web (comprese tutte le App) che utilizzino tali 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 rimangono riconducibili all'interessato, restando dati personali e quindi soggetti al GDPR.

23. Lo scaricamento e l'esecuzione di tali risorse comporta perciò il trasferimento sistematico dei dati personali degli utenti verso queste aziende fornitrici.

La collocazione geografica dei data center

24. Alcune delle società segnalate dichiarano di elaborare e/o conservare presso i propri data center situati in territorio europeo i dati dei cittadini europei.

25. Tuttavia il software utilizzato per tale elaborazione è controllato direttamente dall'azienda madre e dunque sottoposto alla normativa vigente nel rispettivo Paese di appartenenza.

26. Il controllo sul software che elabora i dati dei cittadini europei implica ovviamente la possibilità di accedere ai dati stessi. Come 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. Analoghi rischi si pongono anche per le aziende situate in altri Paesi.

Le evidenze raccolte dall'osservatorio di Monitora PA

27. L'osservatorio automatico distribuito di Monitora PA analizza una certa quantità 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.

28. L'allegato `MonitoraPA_2023-02-14.zip` contiene copia compressa dell'output relativo ai servizi qui analizzati, riportando l'elenco delle chiamate inviate presso i rispettivi server con il conseguente trasferimento sistematico dei dati personali dei visitatori.

29. Il formato del file TSV, attentamente studiato per massimizzare la possibilità dei cittadini di utilizzare personalmente l'osservatorio di Monitora PA [7], 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 [8])
    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"`)

15 febbraio 2023

Firme

Giacomo Tesio

Massimo Maria Ghisalberti

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