

Allegato Tecnico

Indice

Outlook 365	2
L'identificazione del mittente	3
Microsoft 365 e altri servizi	3
L'identificazione del visitatore	4
La collocazione geografica dei data center	5
Le evidenze raccolte dall'osservatorio di Monitora PA	5
Firma	6

Il presente allegato tecnico descrive i trasferimenti sistematici che sono avviati automaticamente dalle Pubbliche Amministrazioni che utilizzano Outlook 365 e gli altri servizi connessi ai servizi cloud della piattaforma Microsoft 365, nonché le problematiche per la privacy e la sicurezza dei cittadini che tale utilizzo determina.

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

Outlook 365

1. Le pubbliche amministrazioni oggetto della presente segnalazione utilizzano per la posta elettronica ordinaria (PEO) il noto servizio fornito da Microsoft Corporation denominato Outlook 365, evidenziato dalla presenza dei server dell'azienda nei record MX dei domini istituzionali delle stesse.
2. Ogni email inviata a/o da dette pubbliche amministrazioni comporta un trasferimento di molti dati personali verso Microsoft, ottenuti tramite l'analisi:
 - dei diversi header SMTP¹ inviati dal client di posta elettronica in uso da parte di ciascun mittente (nonché dai server SMTP attraversati) e dei metadati da essi deducibili²
 - del contenuto di ciascuna mail non cifrata
 - dalle relazioni che intercorrono fra ciascun mittente e i destinatari
 - dell'identificativo del device e della posizione geografica del destinatario qualora si utilizzi l'App Outlook 365³
 - dei dati elencati al punto 8, qualora il destinatario utilizzi l'interfaccia Web del servizio

Tale trasferimento sistematico di dati personali verso Microsoft rimane attivo anche nel caso in cui la casella di destinazione non esista o sia stata disattivata. Fintanto che il record MX⁴ del DNS continui a puntare ai server di Microsoft, tutte le email inviate alle caselle di quel dominio verranno ricevute e trattate da Microsoft. Solo a valle del trattamento sarà segnalata al mittente l'impossibilità di consegna del messaggio.

3. In molti casi i dati ricevuti e trattati da Microsoft includono categorie speciali di dati personali protetti dall'articolo 9 del GDPR: per esempio uno studente potrebbe inviare alla segreteria di un istituto che utilizzi Outlook 365 un certificato medico oppure un medico di famiglia potrebbe menzionare un paziente durante un consulto con uno specialista di un ASL che utilizzi il medesimo servizio.

¹Oltre agli header **From:**, **Sender:**, **To:**, **cc:** e **bcc:** che permettono di identificare mittenti e destinatari della comunicazione, risultano di particolare interesse, fra gli header previsti dalla *sezione 3 del RFC 2076*, gli header **Received** (che permettono di ricostruire il percorso del messaggio attraverso i diversi server SMTP, siano essi in reti pubbliche o private), l'header **Date:** che permette di conoscere l'ora cui il messaggio è stato inviato, gli header **Reply-To:**, **References:** e **In-Reply-To:** (nonché gli header non-standard **Thread-Index:** e **Thread-Topic:**) che permettono di ricostruire l'intero scambio di email, nonché alcuni header non-standard inclusi da alcuni client di posta come **Fax:**, **Telefax:**, **Phone:** e **X-Mailer** che permettono di acquisire ulteriori informazioni sui recapiti mittente e su software e sistemi operativi che questi utilizza.

²<https://labs.rs/en/metadata/>

³<https://play.google.com/store/apps/datasafety?id=com.google.android.gm>

⁴https://en.wikipedia.org/wiki/MX_record

L'identificazione del mittente

4. Il mittente di ogni email ricevuta dai server di Microsoft è identificabile tramite l'indirizzo di posta elettronica indicato negli header **From:** e **Sender:**⁵.
5. Anche quando tale indirizzo non riporti associato il nome e il cognome della persona fisica che lo sta utilizzando, è probabile che Microsoft possa comunque identificarla. Si pensi come esempio a uno studente che utilizzi uno pseudonimo per accedere ai servizi forniti al proprio istituto da Microsoft 365 Education: anche se le email inviate ai docenti non riportassero nel contenuto e/o nel mittente il nome dello studente, il solo accesso dalla propria rete domestica o dal proprio smartphone a Outlook 365 (o a qualsiasi altro servizio o App della suite), rivelerebbe a Microsoft la propria identità⁶ permettendogli di associarla permanentemente allo pseudonimo utilizzato.

Microsoft 365 e altri servizi

6. Outlook 365 è spesso utilizzato dalle Pubbliche Amministrazioni in associazione con diversi altri servizi di Microsoft come parte di una piattaforma software completa e coerente denominata Microsoft 365.
7. Oltre a Outlook 365, Microsoft 365 include:
 - OneDrive (storage condiviso)
 - Office online
 - Microsoft Teams
 - Microsoft To Do
 - ClipChamp (video editor online)
 - Skype
 - Microsoft Family Safety (parental control)
8. In tutte le versioni disponibili, l'utilizzo di questi servizi comporta il trasferimento sistematico verso Microsoft di diversi dati personali fra cui:
 - indirizzo IP
 - data, ora e "timezone" dell'accesso a ciascun servizio
 - stringa identificativa dello User Agent
 - "impronta digitale" ("fingerprint" dedotto con varie tecniche) dello User Agent utilizzato⁷
 - lingue preferite dall'utente, attraverso l'header **HTTP Accept-Language**
 - eventuali cookie precedentemente impostati.

⁵<https://www.rfc-editor.org/rfc/rfc2822#section-3.6.2>

⁶<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0279942>

⁷https://www.amiunique.org/fp#table_js_wrapper

9. In molti casi i dati personali ricevuti e trattati da Microsoft, includono le categorie speciali di dati personali protetti dall'articolo 9 del GDPR. Uno studente, per esempio, potrebbe accedere sistematicamente ai materiali caricati su OneDrive dal docente di religione cattolica (la cui fruizione dell'insegnamento è facoltativa in Italia), rivelando a Microsoft le proprie convinzioni religiose. Come ulteriore esempio si pensi ad alcuni dati biometrici: l'identificazione facciale, l'impronta vocale (ceduta durante una video conferenza su Microsoft Teams), quelli contenuti in un referto caricato su OneDrive da un medico ospedaliero per condividerlo con un collega.
10. L'uso dei servizi in questione può comportare anche la cessione di innumerevoli dati personali di soggetti terzi non direttamente coinvolti e quindi inconsapevoli, menzionati su documenti redatti dal personale o caricati sullo storage di OneDrive.

L'identificazione del visitatore

11. Tutte le richieste inviate dal browser verso i server dei servizi sopra indicati per il recupero delle risorse (HTML, script, CSS, media vari), sia che siano per il primo "caricamento" che per la convalida della "cache", causano un 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 preferite dall'utente, attraverso l'header HTTP **Accept-Language**
 - data dell'ultima visita, attraverso l'header HTTP **If-Modified-Since**
 - eventuali cookie precedentemente impostati.
12. Questi dati iniziali, il cui trasferimento automatico è determinato dalle specifiche dei protocolli HTTP e TCP/IP, sono sufficienti per l'identificazione dell'utente in modo univoco nella maggioranza dei casi, anche quando ci si limiti alla sola visualizzazione di un contenuto pubblicamente disponibile (come un documento condiviso su OneDrive, etc.) senza preventiva autenticazione.
13. Per stabilire l'identità di un utente sarebbe sufficiente che questo utilizzasse un sito web con una risorsa fornita da Microsoft in contemporanea con una sua App gratuita che preveda un'autenticazione. Per una precisa identificazione, basterà confrontare l'attuale IP, con o senza il cookie identificativo, con quello di eventuali utenti attualmente connessi.

14. La straordinaria quantità di informazioni raccolte tramite telemetria e all'insaputa degli utenti su milioni di sistemi operativi e applicazioni desktop e mobili, rende estremamente facile l'utilizzo di un qualsiasi insieme di dati personali, con funzione di identificativi di fatto⁸. I dati deanonimizzati associati⁹ a questi, sveleranno l'identità del soggetto emittente e perfezioneranno la profilazione.
15. Per Microsoft è dunque sempre possibile identificare i cittadini europei tracciandone l'attività e deducendone opinioni e interessi, per ricondurne poi queste informazioni all'identità personale. In altri termini anche in totale assenza dell'IP e nonostante l'utilizzo di eventuali pseudonimi, tutti i dati raccolti attraverso i servizi in questione rimarranno riconducibili all'interessato, rimanendo per cui dati personali, quindi soggetti al GDPR.
16. L'utilizzo di tali servizi comporta perciò un trasferimento sistematico di dati personali degli utenti verso Microsoft.

La collocazione geografica dei data center

17. Microsoft dichiara di elaborare e/o conservare presso i propri "data center" situati in territorio europeo i dati dei cittadini europei.
18. Tuttavia il software utilizzato per tale elaborazione in questi "data center" è controllato direttamente dall'azienda madre e dunque sarà sottoposto alla normativa vigente nel Paese di appartenenza.
19. Il controllo sul software che elabora i dati dei cittadini europei implica ovviamente la possibilità di accedere ai dati stessi. Microsoft potrebbe, per ottemperare alle leggi USA a cui sottostà, essere costretta a inviare presso uno o più dei suoi "data center" europei, degli "aggiornamenti" in grado di prelevare qualsiasi dato che possa essere di interesse di una agenzia governativa e successivamente, rimuovere ogni evidenza di tale "data breach" prima del riavvio dei servizi.

Le evidenze raccolte dall'osservatorio di Monitora PA

20. 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.
21. L'allegato `MonitoraPA_2023-05-30.zip` contiene copia compressa dell'output relativo ai servizi qui analizzati, riportando i valori dei record

⁸https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf

⁹<https://www.csee.umbc.edu/~kunliu1/p3dm08/proceedings/2.pdf>

MX presenti nella configurazione DNS dei domini istituzionali di ciascun ente analizzato.

22. Il formato del file TSV, attentamente studiato per massimizzare la possibilità dei cittadini di utilizzare personalmente l'osservatorio di Monitora PA ¹⁰, 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 ¹¹)
 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")

Alba, 03 giugno 2023

Firma

Giacomo Tesio

¹⁰<https://github.com/MonitoraPA/monitorapa/blob/main/ARCHITETTURA.md>

¹¹<https://indicepa.gov.it/ipa-dati/dataset/enti>