SICUREZZA INFORMATICA DEI DATI GENOMICI

Riassunto tratto dal NIST IR 8432 Cybersecurity of Genomic Data

Autore: Aldo Pedico – Enterprise Cybersecurity

Contatto: pedicoaldo@gmail.com

PREMESSA

I rischi nell'ambito del trattamento dei dati genomici*: apparentemente sembra un tema molto lontano dalla quotidianità di molti di noi e riservato a pochi ma vi posso assicurare che affrontando nel dettaglio le implicazioni che possono nascere in seguito ad un attacco di un malintenzionato i danni possono essere devastanti.*

Tali danni possono avere conseguenze sia per la cibersicurezza sia per la privacy nei confronti di un singolo individuo o per una nazione.

Vi dico solo questo: i dati genomici possono essere usati per capire il livello di vulnerabilità degli individui e produrre armi chimiche e batteriologiche per un attacco fisico alle persone.

NIST propone soluzioni attraverso l'applicazione del nuovo paradigma Zero Trust *e per gli approfondimenti cita i manuali all'interno del documento.*

Per chi fosse interessato ad approfondire le minacce e le relative contromisure su questo tema, può scaricarsi il manuale pubblicato da NIST, nel mese di marzo 2023, gratuitamente disponibile dal sito https://nvlpubs.nist.gov/nistpubs/ir/2023/NIST.IR.8432.ipd.pdf

INDICE DEGLI ARGOMENTI

ASTRATTO

I dati genomici hanno permesso la rapida crescita della bioeconomia e sono preziosi per l'individuo, per l'industria e per la nazione perché hanno molteplici proprietà intrinseche che, correlati tra loro, li rendono diversi da altri tipi di dati.

Le caratteristiche dei dati genomici rispetto ad altri set di dati sollevano alcune sfide di sicurezza informatica e privacy corrispondentemente uniche che non sono adeguatamente affrontate con le attuali politiche, documenti di orientamento e controlli tecnici.

Sono state individuate lacune nelle pratiche di protezione durante tutto il ciclo di vita dei dati genomici per quanto riguarda:

1) *la loro generazione,*
2) *la loro condivisione sicura e responsabile,*
3) *il monitoraggio dei sistemi che elaborano tali dati,*
4) *la mancanza di documenti di orientamento specifici che affrontino le esigenze specifiche dei processori dei dati*
5) *le lacune normative/politiche rispetto alle minacce alla sicurezza nazionale e alla privacy nella raccolta, archiviazione, condivisione e aggregazione.*

Il rapporto propone una serie di idee di soluzione per affrontare i casi d'uso reali che si verificano in varie fasi del ciclo di vita dei dati genomici insieme alle strategie di mitigazione candidate e ai benefici attesi delle soluzioni.

1 - INTRODUZIONE

Il mondo è entrato in un'era di innovazione biologica accelerata basata principalmente sui molti usi dei dati genomici che includono lo sviluppo e la produzione di vaccini, lo sviluppo e la produzione farmaceutica, la diagnosi delle malattie, le innovazioni agricole che consentono una maggiore produzione alimentare, lo sviluppo di biocarburanti, la ricerca scientifica di base e traslazionale, i test sui consumatori, la genealogia e le forze dell'ordine, tra gli altri.

La tecnologia di sequenziamento genetico si è evoluta al punto tale che il sequenziamento di interi genomi è fattibile e conveniente.

Con il progredire di questa era, c'è una nuova consapevolezza dei rischi per la sicurezza nazionale, la sua economia, la sua industria biotecnologica e i suoi cittadini a causa degli attacchi alla sicurezza informatica mirati ai dati genomici, come evidenziato nell'ordine esecutivo sull'avanzamento della biotecnologia e dell'innovazione biomanifatturiera per una bioeconomia americana sostenibile, sicura e protetta.

Inoltre, le informazioni genetiche umane richiedono il rispetto delle politiche, delle leggi e dell'etica relative alla privacy.

Tuttavia, il valore intrinseco di alcuni dati genomici risiede nella capacità di condividere informazioni con la comunità più ampia, creando la necessità di bilanciare le restrizioni di accesso con le capacità di condivisione dei dati.

PROBLEMI DI SICUREZZA INFORMATICA E PRIVACY

Gli attacchi informatici mirati ai dati genomici includono attacchi contro la:

1. RISERVATEZZA,
2. INTEGRITÀ *e*
3. DISPONIBILITÀ.

1. *Contro la* RISERVATEZZA

   Possono minacciare la nostra economia attraverso il furto della proprietà intellettuale di proprietà dell'industria biotecnologica, consentendo ai concorrenti di ottenere un vantaggio economico sleale accedendo ai dati genomici detenuti negli Stati Uniti.

2. *Contro la* INTEGRITÀ

   Possono interrompere la produzione biofarmaceutica, la produzione agroalimentare e l'attività di bioproduzione.

3. *Contro la* DISPONIBILITÀ

   Includono la CRITTOGRAFIA *a scopo di riscatto, la* CANCELLAZIONE *dei dati e la* DISABILITAZIONE *delle apparecchiature automatizzate critiche utilizzate nella ricerca, nello sviluppo e nella produzione.*

I potenziali danni degli attacchi informatici ai dati genomici minacciano anche la sicurezza nazionale, tra cui consentire lo sviluppo di armi biologiche e la sorveglianza, l'oppressione e l'estorsione dei cittadini, militari e personale di intelligence sulla base dei loro dati genomici.

Gli attacchi informatici mirati ai dati genomici possono anche danneggiare gli individui consentendo il ricatto, la discriminazione basata sul rischio di malattia e la perdita di privacy dalla rivelazione di consanguineità o fenotipi nascosti tra cui salute, stabilità emotiva, capacità mentale, aspetto e abilità fisiche.

Oltre ai rischi per la privacy che possono sorgere a causa di un attacco informatico, possono sorgere anche rischi per la privacy non correlati alla sicurezza informatica durante l'elaborazione dei dati genomici: possono sorgere quando nell'elaborazione dei dati genomici c'è insufficiente: 1) prevedibilità*, 2)* gestibilità *e 3)* dissociabilità*.*

1. *Insufficiente* prevedibilità*: può causare problemi di* PRIVACY *se le persone sono sorprese da ciò che sta accadendo con i loro dati genomici.*
2. *Insufficiente* gestibilità*: può verificarsi quando non sono in atto le capacità per consentire un'amministrazione adeguatamente granulare dei dati genomici, ad esempio, gli individui potrebbero dover essere in grado di eliminare alcuni o tutti i loro dati genomici da un set di dati.*
3. *Insufficiente* dissociabilità*: introduce rischi per la privacy quando è consentito l'accesso ai dati genomici grezzi, invece di utilizzare adeguate tecnologie di miglioramento della privacy per estrarre solo le informazioni necessarie (senza rivelare i dati grezzi).*

AMBITO E OBIETTIVI DEL DOCUMENTO

Questo documento identifica le caratteristiche dei dati genomici rispetto ad altri tipi di dati di alto valore e fornisce una panoramica introduttiva delle risorse di gestione del rischio di sicurezza informatica e privacy e delle classificazioni dei rischi durante il ciclo di vita dei dati genomici.

Identifica inoltre le sfide più comuni per la protezione dei dati genomici, le attuali pratiche di sicurezza informatica e privacy all'avanguardia per i dati genomici e le lacune associate a queste pratiche.

Infine, viene presentata una serie di casi d'uso che simulano le sfide della vita reale con strategie di mitigazione candidate per affrontare ciascuna sfida, insieme ai benefici attesi forniti dalle soluzioni proposte.

2 - CONTESTO

Un'intera sequenza del genoma può essere utilizzata per identificare un individuo in alcuni casi e i dati grezzi possono superare i 100 gigabyte di dimensione.

La PORTABILITÀ, la PRIVACY, la CATENA DI CUSTODIA, l'INTEROPERABILITÀ, la GESTIONE DEL CONSENSO e la REINTERPRETAZIONE dei dati genomici sono tutti punti chiave per il loro uso efficace nel settore sanitario.

Come altri dati sensibili, in ogni fase del ciclo di vita dei dati genomici, dalla creazione all'archiviazione, dall'analisi alla diffusione, i dati possono essere a rischio di intercettazione, danneggiamento, sovrascrittura o cancellazione.

Si stima che il mercato dei test genetici DTC valga oltre 1,3 miliardi di dollari e si prevede che crescerà fino a circa 3,5 miliardi di dollari entro la fine del 2026.

I dati genomici trasferiti e condivisi rappresentano decine di milioni di individui che forniscono le loro informazioni.

La perdita di controllo dei dati genomici può causare rischi per la PRIVACY, la SICUREZZA PERSONALE e la SICUREZZA NAZIONALE, poiché gli avversari possono utilizzare i dati genomici per motivi nefasti come la sorveglianza, l'oppressione e l'estorsione.

Come riportato dagli esperti di sicurezza nazionale, le minacce alla sicurezza possono sorgere attraverso la creazione di armi biologiche specifiche per la popolazione o identità compromesse di agenti della sicurezza nazionale.

3 - SFIDE E PREOCCUPAZIONI ASSOCIATE ALLA GESTIONE DELLE INFORMAZIONI GENOMICHE

Le preoccupazioni per la sicurezza identificate includono:

1. VIOLAZIONE*: i dati genomici possono essere utilizzati per la sorveglianza della popolazione e l'oppressione, nonché per l'estorsione dei nostri cittadini, militari e personale di intelligence.*
2. ARMI BIOLOGICHE*: i potenziali rischi connessi alle armi biologiche dirette contro la sequenza di DNA di un individuo o alla clonazione di un individuo sono stati ritenuti poco pratici e non preoccupazioni che dovrebbero guidare il lavoro di cibersicurezza genomica.*

3. PRODUZIONE DI PRODOTTI TOSSICI O AGENTI INFETTIVI*: si ipotizza che a causa della mancanza di integrità e dei controlli di manomissione che tipicamente esistono, i dati genomici potrebbero essere corrotti alterando sequenze o annotazioni e che "Questi cambiamenti potrebbero ritardare i programmi di ricerca o provocare la produzione incontrollata di prodotti tossici o agenti infettivi" causando perdita di vite umane e conseguenze economiche.*

SFIDE IN MATERIA DI PRIVACY

Le sfide alla PRIVACY *derivanti dall'uso di dati genomici umani includono problemi per gli individui come:*

1. *l'abilitazione al ricatto,*
2. *la discriminazione basata sul rischio di malattia e*
3. *la rivelazione di consanguineità nascosta o fenotipi tra cui salute, stabilità emotiva, capacità mentale, aspetto e abilità fisiche.*

I dati genomici umani dovrebbero essere classificati come INFORMAZIONI *di* IDENTIFICAZIONE PERSONALE (PERSONALLY IDENTIFIABLE INFORMATION PII).

I potenziali problemi di privacy identificati includono:

1. REIDENTIFICAZIONE DI DATI GENOMICI DE-IDENTIFICATI (ANONIMIZZATI?)*: i dati genomici umani, anche piccoli frammenti dell'intero genoma di una persona, di solito possono essere reidentificati per alcune popolazioni se combinati con set di dati disponibili, come dati di ascendenza, dati genomici identificati auto-condivisi di parenti lontani, inferenza del cognome, età, ecc.*
2. LA RIVELAZIONE IMPREVISTA DEI PARENTI DI SANGUE DEGLI INDIVIDUI PUÒ PORTARE ALLA PERDITA DELLA DIGNITÀ QUANDO TALI RELAZIONI VENGONO IDENTIFICATE*: possono essere rivelati legami consanguinei che possono essere imbarazzanti o incriminanti, con conseguenti danni psicologici o di reputazione.*

DISCRIMINAZIONE E PROBLEMI DI REPUTAZIONE

Le discriminazioni e le preoccupazioni reputazionali identificate includono:

1. FALSA IDENTIFICAZIONE*: la cattiva gestione del campione, le deviazioni delle procedure di laboratorio del crimine o la modifica intenzionale dei dati genomici digitali producono un rischio*

per le persone di essere incastrati per crimini che non hanno commesso o estorto con le informazioni genomiche falsificate.

2. DISCRIMINAZIONE BASATA SUL RISCHIO DI MALATTIA IDENTIFICATO DAI DATI GENOMICI DI UN INDIVIDUO*: mentre alcune forme di discriminazione negli Stati Uniti sono proibite dalla legge, le leggi federali sono scritte in modo restrittivo e non proibiscono la discriminazione basata sui dati genomici di un individuo per cose come l'assicurazione sulla vita, l'accettazione nell'esercito o nelle comunità residenziali per anziani (ad esempio, in base al rischio di Alzheimer).*

3. CONSEGUENZE NON INTENZIONALI DERIVANTI DALLA DISTORSIONE DEL CAMPIONE*: l'IA e le tecniche di analisi statistica analizzano grandi insiemi di dati genomici per trovare malattie, fattori di rischio per malattie, prendere decisioni terapeutiche e prevedere la prognosi del paziente.*

   Questa distorsione dei dati campione può essere amplificata, in particolare nelle tecniche di IA, e influire sui risultati di queste tecniche di analisi e previsione che possono comportare discriminazioni, con conseguenti potenziali danni a coloro i cui dati genomici non sono rappresentati nel set di campioni.

   Questo pregiudizio degli algoritmi di IA è stato studiato dal NIST nel campo del riconoscimento facciale.

PREOCCUPAZIONI ECONOMICHE

Le preoccupazioni economiche identificate includono:

1. VIOLAZIONE DELLA PROPRIETÀ INTELLETTUALE*: l'esfiltrazione di dati genomici può comportare la perdita della proprietà intellettuale per le istituzioni, con conseguenti perdite economiche in futuro.*

2. INTERRUZIONE OPERATIVA*: le interruzioni della generazione, dell'archiviazione o dell'utilizzo dei dati genomici possono influire negativamente sulle operazioni di produzione, influenzando la produzione di alimenti agricoli, la produzione biofarmaceutica e la bioproduzione.*

3. ESTORSIONE*: i cattivi attori o gli stati nazionali possono estorcere individui sulla base dei loro dati genomici, minacciando di rivelare informazioni sensibili sulla salute o sui parenti codificate nel genoma dell'individuo, con conseguente danno finanziario, psicologico e perdita di reputazione per l'individuo.*

4. SANZIONI E RESPONSABILITÀ*: la negligenza nei controlli di sicurezza informatica dei dati genomici da parte dell'azienda o dell'istituzione potrebbe violare i loro obblighi normativi, compromettere il*

loro accesso ai dati in futuro e comportare perdite finanziarie significative attraverso sanzioni imposte.

5. REVOCARE L'ACCESSO AI DATI*: la negligenza in un'adeguata protezione della sicurezza o della privacy può indurre le persone a non acconsentire più all'utilizzo dei propri dati genomici in studi scientifici, il che potrebbe ridurre l'efficacia della ricerca e minacciare la bioeconomia.*

ALTRE POTENZIALI PREOCCUPAZIONI FUTURE

Altre potenziali preoccupazioni future identificate includono:

1. PERDITA DI INDIVIDUALITÀ*: sebbene non sia una minaccia realistica al momento o nel prossimo futuro, la clonazione di un individuo potrebbe teoricamente essere realizzata con le sue informazioni genomiche.*
   Ciò potrebbe causare danni psicologici o finanziari alla persona e al suo clone.
2. INGEGNERIA UMANA*: le informazioni genomiche potrebbero essere utilizzate per l'eugenetica, mirando più finemente alla parentela o ad altri tratti umani.*
   Ciò potrebbe causare danni sociali o danni psicologici all'individuo.

4 - STATO ATTUALE DELLE PRATICHE

LACUNE TECNICHE NELLE SOLUZIONI

1. ATTUALMENTE, I PRODUTTORI DI SEQUENCER NON FORNISCONO SOFTWARE BILL OF MATERIAL (SBOM) PER I LORO DISPOSITIVI

   Pertanto i professionisti della sicurezza non hanno visibilità sulle potenziali vulnerabilità del loro software e quindi non possono consigliare adeguatamente gli utenti dei sequencer *su come affrontare le vulnerabilità del software scoperto attraverso patch o altre misure di mitigazione.*

2. I SEQUENZIATORI SONO IN GENERE CONNESSI A UNA RETE E A INTERNET

   Ciò fornisce l'accesso a il produttore per gli aggiornamenti e il trasferimento dei file all'archiviazione sicura.

   Non ci sono linee guida sugli indirizzi di rete e sui protocolli di rete corrispondenti necessari ai sequenziatori *per operare in modo efficace.*

   Se questa guida esistesse, sarebbe possibile sequenziare micro segmenti sulla rete fornendo loro solo i necessari risorse per il loro corretto funzionamento in linea con i principi di sicurezza informatica ZTA.

   Questo mitigherebbe le possibilità che gli avversari sfruttino le vulnerabilità nel Hw *o* Sw *di* Sequencer *per l'esfiltrazione dei dati e l'utilizzo dei* sequencer *come punti di ingresso che consentono il movimento laterale in tutta la rete aziendale.*

3. *Il problema generale del confinamento dei dati (cioè gli utenti autorizzati e/o il loro software che condividono l'accesso non autorizzato ai dati), è un noto problema irrisolto nella sicurezza informatica.*

   A causa dei rischi per la privacy dei soggetti e dell'alto valore di molti tipi di dati genomici, il problema del confinamento è rilevante per questi dati.

   Questo problema è più comunemente affrontato da controlli contrattuali che possono essere particolarmente complessi quando i controlli sono tra più organizzazioni.

   Inoltre, i controlli contrattuali in genere non impediscono la condivisione non autorizzata dei dati ma prevedono sanzioni se viene eseguita.

   Queste sanzioni in genere non possono rimediare alla perdita di privacy dei pazienti e degli interessati.

4. *La maggior parte della condivisione e dell'elaborazione dei dati genetici avviene frequentemente in ambienti cloud, sfruttando i contenitori (ad esempio, Docker o Pod).*

   Molti scanner di vulnerabilità per la sicurezza informatica non sono ottimizzati per la scansione dei contenitori, con conseguente incapacità di identificare alcune vulnerabilità e un numero elevato di falsi positivi.

5 - Soluzioni Disponibili per Soddisfare le Esigenze Attuali

Microsegmentazione Automatica della Rete dei Sequenziatori con MUD (Manufacturer Usage Description)

NCCoE ha recentemente sviluppato implementazioni di modelli di una soluzione per aiutare a proteggere i dispositivi Internet of Things (IoT) *che possono avere applicazioni per sequenziatori genomici.*

NIST ha rilasciato NIST SP 1800-15, Securing Small-Business and Home Internet of Things (IoT) Devices*:* Mitigating Network-Based Attacks Using MUD *nel 2021.*

MUD limita un dispositivo gestito a comunicare solo con una "lista consentita" di indirizzi di rete consentiti (protocolli Internet e porte) specifici per ciascun tipo di dispositivo, **in linea con i principi di sicurezza informatica ZTA**.

Descrizione del Caso d'Uso

I sequencer *sono dispositivi costosi gestiti da software personalizzato che forniscono solo una visibilità limitata sulle connessioni e le configurazioni di rete.*

I sequenziatori devono essere connessi a Internet e all'interno della rete per l'assistenza del produttore, l'accesso degli utenti e l'archiviazione dei dati, ma il numero di indirizzi di comunicazione e protocolli necessari per il corretto funzionamento è limitato.

In questo modo, i sequencer possono essere paragonati ai dispositivi IoT.

A causa della connettività dei sequenziatori, possono essere un obiettivo interessante per l'esfiltrazione dei dati, l'implementazione di ransomware, l'impianto di malware e il movimento laterale all'interno della rete aziendale dell'utente.

IDEA DI SOLUZIONE

MUD potrebbe essere applicato ai sequencer.

Poiché i sequencer *hanno un numero limitato di modelli di comunicazione, MUD consente a una rete di limitare la comunicazione del sequencer all'interno della rete locale ed esternamente solo alle risorse necessarie per il corretto funzionamento.*

Esiste un'architettura MUD che implementa MUD e contiene un MUD Manager che fornisce la microsegmentazione del dispositivo in base all'elenco di indirizzi consentiti fornito dal produttore.

Questa microsegmentazione fornita dal MUD Manager inibisce notevolmente la capacità degli avversari di accedere a un dispositivo gestito; e se accedono a un dispositivo, la loro capacità di muoversi lateralmente attraverso il resto della rete è severamente limitata dal MUD Manager.

La Figura 3 illustra una potenziale architettura MUD per un sequencer.

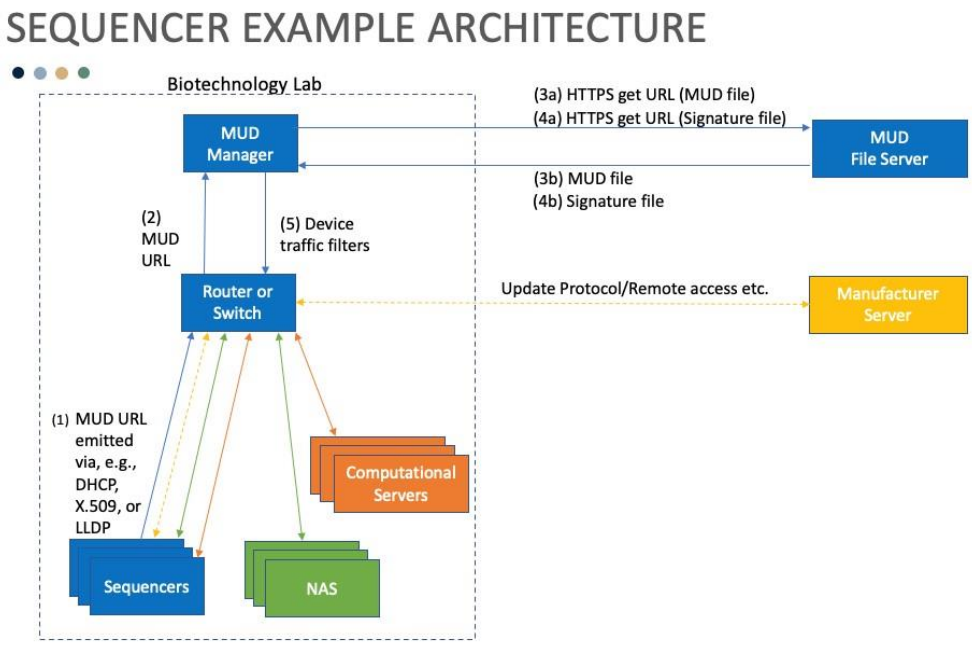

FIG. 3. ARCHITETTURA MUD NOZIONALE PER UN SEQUENCER

Un vantaggio significativo della soluzione MUD consiste nella possibilità d'implementare al suo interno anche dispositivi legacy (il cui software non può essere modificato) per fornire le informazioni MUD necessarie quando si è connessi alla rete.

La crittografia omomorfica multi-parte federata, è una soluzione che può virtualmente eliminare i rischi di riservatezza, d'integrità, di perdita ed anche risolvere il problema del confinamento.

Si tratta di una tecnica che consente il calcolo su dati crittografati aggregati su più set di dati.

L'esfiltrazione dei dati grezzi è impedita in quanto l'utente autorizzato è in grado di ottenere risultati solo dal calcolo che coinvolge più set di dati e gli utenti autorizzati non possono accedere ai dati grezzi in chiaro.

BENEFICI ATTESI

Questa soluzione può consentire progressi più rapidi nel trattamento oncologico e nella medicina di precisione consentendo la collaborazione tra organizzazioni senza complesse negoziazioni contrattuali o accordi di condivisione perché il rischio di esfiltrazione dei dati o della violazione della privacy è praticamente eliminato.