Federated Learning e Sicurezza Distribuita



Il *Federated Learning* (FL) si configura come un paradigma emergente e altamente promettente nell'ambito dell'apprendimento distribuito orientato alla privacy, consentendo a più entità di collaborare all'addestramento di modelli senza necessità di condividere i dati grezzi. Nonostante il FL integri meccanismi intrinseci di protezione della privacy, rimane suscettibile a una vasta gamma di minacce sia in termini di sicurezza sia di riservatezza.

Questo contributo esplora in dettaglio le sfide di sicurezza associate al FL, analizzando i vettori di attacco e i meccanismi di difesa lungo l'intero ciclo di vita del sistema. L'analisi si focalizza su tre fasi critiche: *auditing* dei dati e dei comportamenti, addestramento del modello e predizione. L'obiettivo è delineare come un sistema FL affidabile necessiti dell'integrazione di misure di sicurezza appropriate in ogni fase, garantendo un equilibrio tra la protezione della privacy, l'utilità del modello e le prestazioni complessive.

Introduzione

In un contesto in cui la protezione della privacy rappresenta un elemento cruciale nello sviluppo dell'intelligenza artificiale, il *Federated Learning* si è imposto come una soluzione innovativa nel panorama dell'apprendimento distribuito. Questo framework consente a organizzazioni e individui di collaborare nell'addestramento di modelli mantenendo i dati localmente, riducendo il rischio di esposizione di informazioni sensibili. Tuttavia, la natura distribuita e collaborativa di questa architettura introduce nuove superfici di attacco sfruttabili da avversari interni ed esterni.

Le sfide di sicurezza nel contesto del *Federated Learning* sono intrinsecamente complesse e richiedono approcci sistematici e multidisciplinari. I recenti sviluppi nella ricerca hanno evidenziato una serie di vulnerabilità specifiche, accompagnate dalla proposta di soluzioni avanzate per mitigare tali rischi. Le tendenze attuali si concentrano sull'implementazione di meccanismi robusti volti a proteggere non solo la privacy dei partecipanti, ma anche l'integrità e l'affidabilità del processo di apprendimento.

I concetti fondamentali della sicurezza nel *Federated Learning* comprendono la definizione di modelli di minaccia specifici e l'applicazione di principi di protezione mirati. I modelli di minaccia, in questo contesto, includono scenari che vanno da attacchi mirati da parte di partecipanti compromessi a intrusioni esterne su scala sistemica.

Le vulnerabilità di sicurezza e privacy nelle architetture FL rappresentano una delle principali aree di indagine. La natura distribuita di questi sistemi introduce complessità uniche che richiedono soluzioni personalizzate e specificamente adattate. Le soluzioni di difesa avanzate attualmente in sviluppo combinano tecniche di crittografia omomorfica, metodi di preservazione della privacy come il *differential privacy* e strumenti per il rilevamento di anomalie nel comportamento dei partecipanti.

Le future direzioni di ricerca in questo ambito mirano a un'evoluzione delle contromisure, considerando la crescita e la diversificazione delle minacce. L'obiettivo principale rimane quello di potenziare la resilienza dei sistemi FL, preservandone al contempo l'efficienza computazionale e l'usabilità pratica.

Il presente studio adotta un approccio multi-fase per esaminare le sfide di sicurezza lungo l'intero ciclo di vita di un sistema *Federated Learning*, partendo dalla fase di inizializzazione fino alla sua operatività continua. Questo

approccio consente una valutazione completa e integrata, tenendo conto delle interazioni tra le diverse componenti del sistema e dell'impatto complessivo delle misure di sicurezza implementate.

Architettura di Sicurezza e Threat Models nel Federated Learning

Architettura di Base

Il *Federated Learning* (FL) si configura come un framework avanzato per l'apprendimento distribuito, caratterizzato da una struttura architetturale complessa che mira a bilanciare efficienza computazionale e sicurezza dei dati. Alla base di questa architettura emergono componenti chiave, ciascuno dei quali ricopre un ruolo cruciale per il funzionamento e la resilienza del sistema.

I ***local participants***, noti anche come client, costituiscono il fulcro dell'architettura distribuita. Questi nodi mantengono i propri dataset privati localmente, riducendo drasticamente i rischi associati alla condivisione dei dati sensibili. Questo approccio decentralizzato non solo garantisce la privacy dei dati originali, ma promuove anche un paradigma di collaborazione efficiente nell'addestramento del modello globale, riducendo al minimo le esposizioni potenziali.

Il ***coordination mechanism*** per l'aggregazione dei modelli è il secondo elemento essenziale. Questo modulo gestisce la fusione degli aggiornamenti dei parametri forniti dai partecipanti, garantendo che il modello globale benefici del contributo distribuito mantenendo al contempo integrità e coerenza. Tecniche avanzate come il *Federated Averaging* (FedAvg) e le sue varianti sono frequentemente utilizzate per ottimizzare questa fase.

I ***communication protocols*** giocano un ruolo determinante nel

garantire la sicurezza e l’efficienza delle interazioni tra i nodi. Questi protocolli devono essere progettati per supportare un flusso bidirezionale sicuro, riducendo la latenza e proteggendo la riservatezza degli aggiornamenti mediante tecniche di crittografia *end-to-end* e meccanismi di autenticazione robusti.

Infine, i ***privacy-preserving computation methods***, come la crittografia omomorfica, il *secure multi-party computation* (SMPC) e il *differential privacy*, completano l’architettura di base. Questi strumenti sono indispensabili per proteggere la confidenzialità dei dati durante le fasi di elaborazione e trasmissione, fornendo un ulteriore strato di sicurezza.

Threat Models nel contesto del Federated Learning

I *threat models* nel contesto del FL comprendono una varietà di attori malevoli e vulnerabilità che possono compromettere la sicurezza e la robustezza del sistema. Tali minacce possono essere classificate in base al comportamento e alla posizione degli attaccanti.

Gli *honest-but-curious participants* rappresentano una minaccia sottile ma significativa: questi nodi seguono correttamente il protocollo, ma tentano di dedurre informazioni sensibili analizzando i gradienti o gli aggiornamenti ricevuti. Questa categoria richiede contromisure specifiche, come l’applicazione di tecniche di *differenziale privacy* per offuscare i dati sensibili.

I *malicious participants*, invece, perseguono attivamente l’obiettivo di compromettere il sistema. Essi possono manipolare i propri aggiornamenti al fine di alterare il modello globale (ad esempio, tramite attacchi di *poisoning*) o compromettere l’integrità del processo di apprendimento.

Gli *external adversaries* operano al di fuori del sistema,

cercando di sfruttare vulnerabilità nei protocolli di comunicazione per intercettare o manipolare i dati trasmessi. Gli attacchi più comuni in questa categoria includono attacchi di tipo *man-in-the-middle* e violazioni della confidenzialità tramite analisi del traffico.

I *compromised aggregators* costituiscono un'ulteriore minaccia critica. Un aggregatore compromesso può alterare il processo di aggregazione o esporre i dati sensibili dei partecipanti. Poiché l'aggregatore agisce come punto centrale di controllo, garantire la sua sicurezza è fondamentale per preservare l'integrità del sistema.

Gli obiettivi degli adversa*ries* possono essere diversificati e sofisticati, il furto di informazioni private è spesso la motivazione primaria, con gli attacchi mirati a estrarre informazioni dai parametri condivisi. Il *poisoning* del *training process* rappresenta un altro rischio significativo, in cui gli attaccanti iniettano dati o aggiornamenti malevoli per corrompere il modello.

La *compromise model integrity* è un obiettivo più ampio e strategico, finalizzato a degradare le prestazioni complessive o introdurre comportamenti indesiderati nel modello. Infine, gli *inference attacks*, in cui gli attaccanti cercano di dedurre informazioni sui dati di training analizzando il comportamento del modello, costituiscono una sfida particolarmente complessa, richiedendo contromisure avanzate per garantire una resilienza effettiva.

Security Challenges nel *Federated Learning* per Sistemi IoT

IoT-Specific Considerations: Sfide Fondamentali nell'Integrazione

La complessità intrinseca dei sistemi IoT introduce sfide

particolari, derivanti principalmente dalla loro eterogeneità strutturale. Le ***heterogeneous device capabilities*** rappresentano uno dei principali ostacoli: i dispositivi, dotati di risorse computazionali, memoria e caratteristiche hardware variabili, devono collaborare in maniera efficace all'interno del medesimo ecosistema di *Federated Learning* (*FL*). Questa varietà introduce difficoltà non solo tecniche ma anche potenziali vulnerabilità sfruttabili dagli attaccanti.

I *resource constraints*, come limitazioni in termini di *computational power*, memoria disponibile e capacità energetica, costituiscono un ulteriore elemento critico. La progettazione di soluzioni di sicurezza per sistemi IoT deve essere ottimizzata per garantire un'adeguata protezione, mantenendo un basso impatto in termini di utilizzo delle risorse computazionali ed energetiche.

Sfide di Connettività e Distribuzione

Le *bandwidth limitations* rappresentano una problematica significativa per la natura distribuita del *Federated Learning*, che richiede frequenti scambi di dati tra dispositivi partecipanti e server centrale. Nei sistemi IoT, spesso caratterizzati da connessioni instabili e velocità di trasmissione ridotte, l'efficienza e la sicurezza del sistema possono risultare compromesse. Protocolli di sicurezza tradizionali, sebbene robusti, potrebbero rivelarsi troppo onerosi in termini di consumo di banda.

La distribuzione geografica dei dispositivi IoT introduce ulteriori complessità. Dispositivi collocati in aree geografiche remote o in ambienti eterogenei sono soggetti a condizioni di rete e ambientali differenti, complicando l'implementazione di misure di sicurezza uniformi. Inoltre, le *dynamic network conditions*, caratterizzate da connettività intermittente e latenza variabile, richiedono soluzioni di sicurezza adattive per garantire efficacia operativa anche in situazioni di rete sfavorevoli.

Scalabilità e Domini Applicativi

La scalabilità rappresenta una sfida cruciale: gestire la sicurezza di miliardi di dispositivi connessi richiede approcci innovativi e altamente efficienti. Questa problematica assume particolare rilevanza considerando i molteplici domini applicativi in cui operano i sistemi IoT, tra cui il settore sanitario, l'industria manifatturiera e le *smart grids*. Ciascun dominio presenta requisiti di sicurezza specifici, che richiedono soluzioni personalizzate e interoperabili.

IoT-Specific Threat Models: Vulnerabilità e Minacce

All'interno dei modelli di minaccia specifici per l'IoT, emergono vulnerabilità che richiedono un'attenzione mirata. Le vulnerabilità a livello di dispositivo rappresentano un punto di ingresso critico: la compromissione di un singolo dispositivo potrebbe avere ripercussioni su larga scala, compromettendo l'intero sistema. I problemi di connettività di rete, oltre a degradare le prestazioni, possono essere sfruttati come vettori di attacco da parte di attori malevoli.

Attacchi e Minacce alla Sicurezza

Gli attacchi di esaurimento delle risorse (*resource exhaustion attacks*) costituiscono una minaccia particolarmente insidiosa per i dispositivi IoT, spesso caratterizzati da risorse computazionali e energetiche limitate. Inoltre, le minacce alla sicurezza fisica, data l'esposizione di molti dispositivi IoT in ambienti non controllati, aumentano il rischio di manomissioni hardware.

Propagazione degli Attacchi e Privacy

La propagazione degli attacchi tra domini (*cross-domain attack propagation*) rappresenta un rischio significativo in ecosistemi IoT connessi tramite *FL*. Un attacco inizialmente

confinato a un dominio applicativo potrebbe propagarsi, compromettendo l'integrità di altri domini correlati. Inoltre, il leakage di dati sensibili, in particolare in applicazioni critiche come quelle sanitarie, costituisce una preoccupazione fondamentale. Proteggere la privacy degli utenti in tali contesti richiede l'adozione di tecniche avanzate, come la *homomorphic encryption* e i metodi di anonimizzazione dei dati.

Vulnerabilità di Security e Privacy nei Sistemi *Federated Learning* (FL) Distribuiti

Criticità Specifiche dell'Architettura Distribuita

Vulnerabilità della Distribuzione delle Risorse

Nei sistemi *Federated Learning* (FL) distribuiti, la gestione delle risorse computazionali presenta vulnerabilità peculiari intrinseche alla natura collaborativa del processo di apprendimento. La distribuzione asimmetrica delle capacità computazionali tra i nodi partecipanti rappresenta potenziali punti di debolezza durante il processo di *training* federato. I nodi con risorse limitate possono trasformarsi in *bottleneck* per l'intera architettura distribuita, compromettendo non solo le performance ma anche la sicurezza del sistema nel suo complesso.

L'eterogeneità delle risorse disponibili tra i nodi introduce ulteriori vulnerabilità nei meccanismi di sincronizzazione del *training* distribuito. Nodi con capacità computazionali significativamente divergenti possono generare disallineamenti temporali nel processo di apprendimento, creando aperture sfruttabili per attacchi mirati a compromettere la convergenza del modello.

Vulnerabilità nella Coordinazione Distribuita

La natura intrinsecamente distribuita del *Federated Learning* introduce fragilità critiche nei meccanismi di coordinazione. Il processo di orchestrazione del *training* federato risulta particolarmente suscettibile a interruzioni selettive delle comunicazioni tra i nodi. Un attore malevolo potrebbe sfruttare queste lacune per alterare la sincronizzazione del processo di *training*, influenzando la qualità e l'affidabilità del modello finale.

La gestione distribuita del consenso durante l'aggregazione dei modelli locali presenta specifici punti di debolezza. Attacchi mirati ai meccanismi di consenso possono indurre divergenze significative nell'apprendimento, portando alla generazione di modelli inconsistenti o compromessi. La complessità intrinseca nel coordinare molteplici nodi distribuiti espande ulteriormente la superficie di attacco, facilitando manipolazioni intenzionali del processo di aggregazione.

Vulnerabilità nell'Aggregazione Federata

L'architettura distribuita presenta vulnerabilità rilevanti nella fase di aggregazione dei modelli locali. I nodi intermedi incaricati dell'aggregazione parziale costituiscono potenziali punti critici di compromissione, dove un attore malevolo potrebbe manipolare gli aggiornamenti del modello prima che raggiungano l'aggregatore finale.

La struttura gerarchica tipica dell'aggregazione distribuita introduce fragilità nella propagazione degli aggiornamenti. Un compromesso nei livelli intermedi della gerarchia potrebbe propagarsi attraverso molteplici rami dell'architettura distribuita, amplificando gli effetti dell'attacco e compromettendo l'integrità del modello globale.

Vulnerabilità nella Sincronizzazione Temporale

La sincronizzazione temporale in un sistema *Federated Learning*

distribuito rappresenta un ulteriore vettore di vulnerabilità. Attacchi basati sul *timing* possono sfruttare le differenze nei tempi di *training* e aggregazione tra i nodi per inferire informazioni sensibili sui dati di *training* o manipolare deliberatamente il processo di apprendimento.

I ritardi nella propagazione degli aggiornamenti attraverso l'architettura distribuita generano vulnerabilità specifiche relative alla consistenza del modello. Tali asincronicità possono essere sfruttate per orchestrare attacchi mirati a compromettere la convergenza del *training* federato, minando la robustezza e l'affidabilità del sistema nel suo complesso.

Soluzioni di Difesa Avanzate nel *Federated Learning*

IoT-Optimized Defense Mechanisms: Strategie per la Protezione dei Sistemi FL Distribuiti

Nel contesto del *Federated Learning* su sistemi IoT, le *resource-aware solutions* rappresentano un elemento cardine per garantire un addestramento distribuito sicuro ed efficiente. La *lightweight cryptography* emerge come un paradigma essenziale durante le fasi di aggregazione e scambio dei parametri del modello, fornendo meccanismi di cifratura ottimizzati in grado di salvaguardare la riservatezza degli aggiornamenti del modello, pur rispettando i severi vincoli computazionali dei dispositivi IoT partecipanti.

L'*efficient privacy preservation* nel contesto del *Federated Learning* richiede l'adozione di metodologie avanzate, quali il *differential privacy* e il *secure aggregation*, implementate in maniera ottimizzata per dispositivi caratterizzati da risorse computazionali e di memoria limitate. Le *adaptive security measures*, opportunamente calibrate, proteggono sia il processo

di *training* locale che la fase di aggregazione globale, modulando dinamicamente i livelli di protezione in funzione delle risorse disponibili e della sensibilità dei dati di *training*.

Gli *energy-efficient protocols*, progettati ad hoc per il paradigma del *Federated Learning*, ottimizzano il consumo energetico durante le fasi di comunicazione degli aggiornamenti del modello, garantendo che i dispositivi IoT possano partecipare al processo di addestramento senza comprometterne la longevità operativa.

Protezione a Livello Rete (*Network-Level Protection*) nel *Federated Learning*

La sicurezza della comunicazione è un aspetto critico nel *Federated Learning*, dove la trasmissione sicura degli aggiornamenti del modello è imprescindibile. I *secure communication protocols* vengono adattati per affrontare in modo efficace le esigenze specifiche del traffico generato dal *Federated Learning*, proteggendo gli aggiornamenti del modello durante il transito tra i dispositivi *edge* e il server di aggregazione.

La *bandwidth-efficient encryption* viene ottimizzata per ridurre al minimo l'overhead di comunicazione derivante dallo scambio degli aggiornamenti del modello, mantenendo al contempo un elevato livello di protezione. Le *dynamic security policies* sono implementate per adattarsi dinamicamente alle diverse fasi del processo di *Federated Learning*, includendo regole granulari per la protezione durante l'addestramento locale, l'aggregazione e la distribuzione del modello globale.

I *distributed trust mechanisms* giocano un ruolo chiave nel contesto del *Federated Learning*, in cui la fiducia reciproca tra i partecipanti è critica. Questi meccanismi decentralizzati consentono di verificare l'integrità degli aggiornamenti del modello, prevenendo attacchi sofisticati

come il *model poisoning*.

Sicurezza dei Dispositivi (*Device Security*) nei Sistemi FL

La protezione dei dispositivi coinvolti nel *Federated Learning* richiede misure altamente specializzate per salvaguardare sia i dati di *training* locali che il processo stesso di addestramento. I *secure boot mechanisms* garantiscono l'integrità dell'ambiente di *training* locale, verificando l'autenticità del software FL e prevenendo manipolazioni del processo attraverso tecniche avanzate di autenticazione.

Gli *hardware security modules* forniscono uno strato di protezione dedicato alle operazioni critiche, inclusa la gestione sicura delle chiavi crittografiche necessarie per proteggere gli aggiornamenti del modello e il *secure storage* dei parametri del modello locale. Questi moduli rappresentano una componente essenziale per assicurare la confidenzialità del processo di *training* distribuito.

La *firmware protection* si concentra sulla salvaguardia dei componenti software responsabili dell'addestramento locale e delle comunicazioni con il server di aggregazione. Parallelamente, il *resource monitoring* svolge una funzione cruciale nel rilevamento tempestivo di anomalie nel processo di *training*, identificando possibili compromissioni del modello o attacchi come il *type inference*.

Questa architettura di sicurezza multilivello, progettata specificamente per il paradigma del *Federated Learning* in ambienti IoT, garantisce la protezione delle diverse fasi operative, dal *training* locale fino all'aggregazione globale. L'obiettivo è preservare la privacy dei dati, garantire l'integrità del modello risultante e mitigare i rischi legati a minacce emergenti nel panorama della sicurezza cibernetica.

Linee Guida di Implementazione e Sviluppi Futuri nei Sistemi FL Distribuiti

Ottimizzazione dell'Architettura Distribuita

Nel contesto dei sistemi FL distribuiti, l'ottimizzazione dell'architettura richiede un approccio metodico e avanzato, capace di sfruttare la natura gerarchica e dinamica del sistema. Il *distributed computation offloading* riveste un ruolo cruciale, poiché consente di implementare strategie di redistribuzione computazionale basate su algoritmi adattivi in grado di valutare, in tempo reale, le capacità e lo stato operativo di ciascun nodo. Questo processo ottimizza l'allocazione del carico, migliorando la scalabilità e la resilienza del sistema.

La *hierarchical model compression* rappresenta un progresso chiave nell'ambito delle comunicazioni distribuite. Questa tecnica integra metodi di compressione differenziati per ciascun livello della gerarchia del sistema, bilanciando con precisione la riduzione del traffico di rete con il mantenimento delle prestazioni del modello globale. La compressione granulare permette inoltre una maggiore adattabilità alle caratteristiche specifiche dei nodi, massimizzando l'efficienza computazionale e di comunicazione.

Strategie di Comunicazione Multi-livello

L'efficientamento delle comunicazioni in sistemi FL distribuiti richiede un'implementazione stratificata delle strategie di scambio dei dati. Il *tiered communication protocol* introduce un modello gerarchico nel quale ciascun livello della rete applica protocolli ottimizzati e specifici per le proprie esigenze operative, consentendo l'esecuzione di *selective parameter updates*. Questa metodologia riduce drasticamente il volume complessivo di dati trasferiti,

limitando la latenza e ottimizzando la larghezza di banda disponibile.

L'*asynchronous aggregation strategy* si configura come una soluzione altamente innovativa per gestire la natura asincrona delle comunicazioni in sistemi distribuiti su larga scala. Questo approccio consente ai sottogruppi di nodi di effettuare l'aggregazione locale indipendentemente dalla sincronizzazione globale, migliorando la throughput complessiva e mitigando i colli di bottiglia legati alla sincronizzazione centralizzata.

Raccomandazioni per l'Implementazione Distribuita

Un'implementazione efficace dei sistemi FL distribuiti richiede un approccio strategico e rigoroso alla gestione delle risorse. La *distributed resource orchestration* dovrebbe essere supportata da framework dinamici, progettati per allocare e monitorare in tempo reale le risorse computazionali e di rete attraverso i diversi livelli della gerarchia.

Parallelamente, un *adaptive security framework* rappresenta un elemento imprescindibile per l'architettura distribuita. Questo framework deve includere meccanismi di rilevamento delle anomalie distribuiti (*distributed anomaly detection*) e strumenti di risposta coordinata agli incidenti di sicurezza. La capacità di adattarsi automaticamente alle condizioni operative variabili garantisce una protezione coerente e proattiva contro le minacce emergenti.

Direzioni di Ricerca Emergenti

I futuri orientamenti della ricerca nel dominio dei sistemi FL distribuiti si concentrano su paradigmi altamente innovativi che promettono di rivoluzionare le attuali pratiche. La *distributed privacy enhancement* è un'area prioritaria, focalizzata sullo sviluppo di metodologie avanzate di protezione della privacy che operino in modo sinergico con le architetture multi-livello.

Il *dynamic topology management* emerge come un campo di ricerca cruciale, mirato a progettare meccanismi auto-adattivi per la gestione della topologia di rete in sistemi FL distribuiti. Questo filone si basa su modelli predittivi e architetture auto-organizzanti, capaci di rispondere in modo proattivo alle variazioni nei requisiti di *training* e alle condizioni dinamiche della rete.

Prospettive Future

L’evoluzione dei sistemi FL distribuiti richiederà un approccio multidimensionale per bilanciare efficienza operativa, sicurezza e scalabilità. Le future implementazioni dovranno integrare soluzioni sempre più sinergiche, capaci di gestire la complessità delle architetture distribuite, mantenendo elevati standard di *privacy* e *security*. L’attenzione sarà rivolta allo sviluppo di meccanismi adattivi e auto-evolutivi che rispondano in modo dinamico alle mutevoli esigenze del sistema, consentendo una mitigazione proattiva delle sfide emergenti nel panorama cibernetico globale.

Fonti:

Condividi sui Social Network:
https://www.ictsecuritymagazine.com/articoli/federated-learning-fl/