Maturità 2020

> Traccia:
> Le minacce dal web sono in continua evoluzione, descrivere le più comuni; e visto che l'integrità dei dati nell'ambito della sicurezza informatica ha un ruolo fondamentale, descrivere quali accorgimenti il programmatore può prendere in fase di definizione del DB.

Introduzione

La risorsa più importante di ogni organizzazione è l'**informazione**.
Essa è alla base dell'informatica, la cui definizione è "scienza che si occupa del trattamento automatico delle informazioni".
Ciò che viene trattato nell'informatica è il dato, ovvero la "descrizione elementare" di un'informazione. Il dato al giorno d'oggi controlla il mondo, è la "ricchezza" del nuovo millennio, soprattutto per le strategie aziendali. E' infatti nato all'incirca nel 2000 il **data-driven marketing**, cioè l'utilizzo dei dati concentrato al miglioramento del targeting e del personalizzare la comunicazione aumentando il coinvolgimento degli utenti.
I dati provengono praticamente da qualsiasi parte: dai social, dagli acquisti, dalla semplice navigazione sul web.
L'importanza viene sottolineata soprattutto da questa frase: "Il data-driven marketing oggi non è più una scelta, ma l'unica scelta possibile."
Data driven marketing

Il dato è quindi una delle risorse più importanti per l'economia; e come tutte le risorse che possono portare a del guadagno, la sua sicurezza viene messa a dura prova.

Sicurezza informatica

I dati devono essere quindi protetti da possibili attacchi, per poter garantire la **sicurezza informatica**.
La sicurezza informatica è l'insieme dei prodotti, dei servizi e dei mezzi per la protezione dei sistemi informatici per quanto riguarda i principi fondanti della CIA: data Confidentiality, data Integrity e system Availability.

Il principio che si può definire come più importante è il data Integrity, in italiano l'**integrità dei dati**.
Per integrità dei dati si intende la capacità di mantenere i dati e le risorse, garantendo la loro precisione, correttezza e inalterazione non autorizzata durante il loro intero ciclo di vita.
L'integrità è quindi uno degli aspetti più critici per la progettazione, realizzazione e utilizzo di qualsiasi sistema informatico.

L'integrità

Si possono distinguere due categorie principali di integrità: fisica e logica.

- Per integrità **fisica** si intende lo studio per la protezione dei dati al momento dell'archiviazione e del loro recupero dai supporti di memorizzazione.
  Questo tipo di integrità può quindi scontrarsi con difetti elettromeccanici, minacce naturali o altri tipi di problematiche che riguardano l'aspetto fisico della conservazione dei dati.
- L'integrità di cui andremo a parlare più approfonditamente è invece quella **logica**.
  L'integrità logica dei dati interessa in particolare la correttezza, la razionalità e la protezione da soggetti non autorizzati dei dati, andando quindi ad intersecarsi con l'argomento della **sicurezza dei dati**.
  La parte logica che interessa la correttezza e la razionalità include quindi le tematiche di integrità intra-relazionale e interazionale all'interno dei database relazionali.
  - L'integrità intra-relazionale concerne i vincoli all'interno di tuple della stessa relazione, quindi vincoli di dominio, di tupla, not null e di chiave.
  - L'integrità interazionale riguarda invece l'**integrità referenziale**, ovvero l'insieme di regole che garantiscono l'integrità dei dati quando si hanno relazioni associate tra di loro tramite chiave esterna. A questo tipo di integrità corrisponde quindi il vincolo di integrità relazionale.

La sicurezza dei dati, che si occupa di garantire la protezione dei dati verso le azioni provenienti da soggetti non autorizzati, tratta le problematiche relative agli **attacchi informatici**.

Integrità 1Integrità 2Integrità 3

Un attacco informatico è una qualsiasi azione, svolta da soggetti malintenzionati, che mira a sistemi informativi, dispositivi e reti. Un attacco, nel caso dei dispositivi, ha come obiettivo il renderli inoperativi, mentre gli attacchi relativi ai dati sono concentrati nel furto, nell'alterazione e nella cancellazione.

Attacco informatico

Cenno storico

La storia narra che lo scienziato Guglielmo Marconi e il suo assistente Sir John Ambrose Fleming stessero per dimostrare il funzionamento di un rivoluzionario telegrafo senza fili, quando inaspettatamente la macchina riceve due messaggi: un insulto, e una filastrocca canzonatoria dell'italiano futuro premio nobel. L'esperimento risulterà quindi un disastro, dato che qualcuno è riuscito ad inserirsi nelle frequenze radio che Marconi dichiarava come sicure e private. Il colpevole si rivelerà essere l'inventore Nevil Maskelyne, che dal palazzo vicino riuscì ad interferire con le frequenze di Marconi.
Era il 4 giugno 1903, e questo è considerato il primo caso di "hacking" della storia.

Dai tempi di Marconi, le tecnologie e gli strumenti per la trasmissione, l'immagazzinamento, la gestione e la protezione di dati e informazioni hanno fatto passi da gigante, così come i potenziali attacchi, che si sono evoluti parallelamente.

Le tipologie di attacco si possono racchiudere in due categorie principali: attacco attivo e attacco passivo.
L'attacco attivo implica la manomissione di dati o il disturbo del loro funzionamento, danneggiandone l'integrità e la disponibilità. L'attacco passivo invece cerca di intercettare e utilizzare i dati senza danneggiare le risorse del sistema, quindi minacciandone solamente la riservatezza.

Attacchi passivi

L'unico tipo di attacco passivo esistente è l'analisi del traffico.
L'attacco di analisi, chiamato **sniffing**, è uno dei più semplici, in quanto utilizza strumenti preesistenti chiamati **packet sniffer**.
I packet sniffer sono nati con l'intento di essere programmi di diagnostica per le reti, riuscendo a catturare, analizzare e decodificare tutti i pacchetti in transito nelle comunicazioni.
Possono essere infatti molto utili agli amministratori di rete per monitorare lo stato della stessa, analizzando ed individuando eventuali problemi di comunicazione o tentativi di intrusione.
In breve tempo però sono diventati anche strumenti per malintenzionati.

Un attacco sniffing consiste nell'intercettare i singoli pacchetti transitanti nella comunicazione, per poterli analizzare e decodificare, al fine di riuscire ad estrapolare le informazioni contenutesi negli stessi pacchetti.
In questo modo l'individuo attaccante potrà servirsi dei dati ricavati dall'ascolto della trasmissione, senza aver avuto bisogno di introdursi nella rete o nei dispositivi.

Nella famiglia degli attacchi di analisi rientra anche il famoso attacco **Man In the Middle**.
Il Man In the Middle, abbreviato anche in MIM, è parte degli attacchi di analisi del traffico in quanto consiste nel posizionarsi all'interno della connessione tra due dispositivi, nel mezzo per l'appunto, per poter ricevere tutto il traffico che viene trasmesso.
In questo modo l'attaccante potrà sia ricevere i pacchetti trasmessi dai due utenti, ma anche trasmettere come se fosse parte della comunicazione, risultando come uno dei due partecipanti.
L'attacco MIM è però considerato più un attacco attivo, in quanto utilizza sistemi di falsificazione IP e intrusione nella rete, come vedremo in seguito con lo spoofing, che non coincide quindi con i dettami degli attacchi passivi.

Attacchi attivi

Negli attacchi attivi, al contrario degli attacchi passivi, esistono molte più tipologie, suddivise per obiettivo dell'attacco.

Intercettazione

Troviamo innanzitutto l'intercettazione, il cui metodo più utilizzato è lo ***spoofing***.

Lo spoofing è un tipo di attacco che consiste nel falsificare l'identità dell'host da cui viene eseguito l'attacco.
Esso può essere attuato in diverse maniere; il più utilizzato è l'IP spoofing, ma esistono il DNS Spoofing (chiamato anche shadow server), il MAC spoofing (chiamato anche MAC flooding), l'ARP spoofing o l'Email Spoofing (utilizzato principalmente come base per il phishing)).

Anche lo spoofing fa parte, come detto in precedenza, della tipologia di attacchi **Man in the middle**.
Esistono altre tipologie di spoofing che derivano da quelle descritti precedentemente, agendo su tutti i livelli dello modello ISO/OSI.

Le tipologie di spoofing più utilizzate, tra quelle citate, sono quelle che agiscono su **IP** e **DNS**.

- L'IP spoofing consiste nel falsificare l'indirizzo IP dell'host da cui viene eseguito l'attacco. In questo modo l'host viene considerato attendibile dalla rete, cercando di ottenere dati sensibili.

- Il DNS spoofing, chiamato anche **shadow server**, consiste nel creare record DNS fasulli, inserendoli nella cache del server, attraverso una tecnica chiamata ***cache poisoning***, per poter deviare le richieste DNS ad un server secondario, che si posizionerà quindi davanti al server originario, coprendolo (da qui **shadow** server).
  L'utente quindi verrà reinderizzato ad un sito fasullo, molto probabilmente del tutto simile a quello originario, dove inserirà i dati sensibili che verranno poi rubati dall'individuo attaccante.

Nella tipologia spoofing rientra anche l'attacco chiamato **SQL Injection** che riguarda in maniera particolare l'attacco dei database.

- Con questa tecnica è possibile inserire delle stringhe di SQL all'interno di campi di input, che possono essere ad esempio i campi di inserimento utente e password, oppure di ricerca, in maniera tale da fare eseguire le query all'interno del DB, cercando di causare danni o per ottenere dati. Questo tipo di attacco viene utilizzato in particolare in casi in cui l'applicazione che si sta attaccando presenti vulnerabilità piuttosto evidenti.

Produzione

Un altro tipo di attacco è la produzione, che consiste nell'introdurre nuovi elementi nel sistema, al fine di condurre degli attacchi di sabotaggio che hanno l'obiettivo di ridurre l'integrità e la disponibilità delle risorse del sistema.

Le principali tecniche di disturbo sono gli attacchi malware e gli attacchi di tipo DoS.

- I **malware** sono software "malevoli" che mettono a rischio un sistema. Essi cercano di invadere, danneggiare o disattivare computer, sistemi e reti. Il loro scopo è quello di lucrare illecitamente a spese degli utenti.
  Le due categorie più utilizzate all'interno dei malware sono **virus** e **worm**.
  - I **virus** sono software che hanno il compito di moltiplicarsi ***"infettando"*** altri file o altri host presenti nella rete, senza farsi rilevare dall'utente. Solitamente sfruttano falle o vulnerabilità presenti in un sistema operativo o in un software, cercando di rendere inutilizzabile il dispositivo attraverso attività distruttive o di ostruzionismo.
  - Gli attacchi **worm** sono simili ai virus, ma si differenziano per la capacità di riuscire a moltiplicarsi senza doversi "legare" ad altri file o software, utilizzando invece direttamente gli host presenti nella rete.

Come il virus punta a rendere inutilizzabile il dispositivo, occupando una grande quantità di risorse computazionali.
Spesso un attacco worm o un attacco virus coincide con l'installazione di altre tipologie di malware (famiglia di cui worm e virus fanno parte), come ad esempio **backdoor** o **keylogger**, sfruttati dall'hacker per ottenere dati sensibili.
Il worm più potente mai creato è ***MyDoom*** (in italiano "Il mio destino"). In circolazione dal 2004 e imbattuto sino al 2016, è riuscito a causare 38 miliardi di danni in tutto il mondo, diffondendosi attaccando tutti i file presenti sul PC.
WormVirus

- Un ulteriore attacco di produzione è il **Denial of Service** (DoS) o la sua evoluzione, il **Distribuited Denial of Service** (DDoS).
  Questo attacco mira a "tenere occupato" un host, solitamente server, con operazioni sostanzialmente inutili, in maniera tale da potergli impedire di offrire i propri servizi alla rete.

Esistono diverse tipologie di attacchi DoS, alcune cercano di impedire l'accesso di un individuo specifico ad un network o una risorsa, mentre altre cercano di rendere la risorsa inaccessibile per qualsiasi utente.
Queste tipologie di interruzioni, nel caso di attacchi ad aziende, causano perdite finanziarie che possono essere anche molto pesanti.

- La metodologia più utilizzata è il ***buffer overflow***, che consiste nell'invio di un traffico maggiore di quanto il sistema a cui si mira sia in grado di gestire. Questo attacco permette quindi di riuscire a rendere inutilizzabile il sistema, facendolo collassare.
- Un altra strategia appartenente alla famiglia DoS è il ***SYN flood***, che attacca il server autenticandosi solamente in maniera parziale. In questo modo verrà lasciata in attesa la connessione sulla porta per il ***three way handshake***, procedendo quindi a fare la stessa operazione su tutte le porte fino a mettere fuori uso il server.

Il DDos, che può essere visto come "l'evoluzione" del DoS, dato quest'ultimo che avviene principalmente da una sola macchina, consiste nell'utilizzo di diversi dispositivi che prendono di mira una singola risorsa.
Per questo motivo ha infatti molta più probabilità di successo rispetto al DoS normale per la maggior quantità di traffico prodotto. Inoltre questo metodo viene preferito per l'enorme difficoltà nel rintracciare la fonte da dove avviene l'attacco, dato che questa proviene da più punti e non da un singolo host.

- I computer che sono controllati dall'attaccante vengono chiamati ***zombie*** o ***bot***. Dopo essere stati infettati entrano a far parte di una botnet, ovvero una rete di dispositivi infettati, a libera disposizione dell'attaccante.
  Quando il numero di bot è ritenuto esiguo dall'attaccante, essi vengono attivati, sommergendo il server bersaglio di richieste di connessione. Il flusso enorme di risposte provocherà nel sistema una tale "inondazione" di traffico, rendendolo inadeguato alla gestione delle abituali funzioni on-line.

DoS e DDoSDoSLibro di sistemi

Phishing

Il phishing è una tipologia di attacco che concerne anche con il mondo delle truffe; ha infatti l'obiettivo di rubare le informazioni e i dati personali delle vittime per poterli utilizzare a scopo di lucro.
E' anche considerato come una forma di **adescamento**, la molteplicità degli attacchi avviene infatti attraverso l'inganno psicologico dell'utente, sfruttando le meccaniche dell'ingegneria sociale.

- Il metodo utilizzato solitamente è l'email, data la larga probabilità di trovare utenti "deboli". Il messaggio viene costruito in maniera tale da sembrare proveniente da un'organizzazione attendibile, come potrebbero essere una banca o la posta.
  Generalmente il contenuto indica problemi relativi alla sicurezza dell'account della vittima, invitandolo a cliccare su un link, che a prima vista può sembrare corretto. In realtà collega ad un sito fittizio controllato da chi ci sta attaccando. E' quasi impossibile riconoscere un sito fittizio da un sito "originale", dato che spesso sono totalmente identici, tranne ovviamente che per l'indirizzo URL, che può essere il primo campanello d'allarme, unito al fatto di non essere in una connessione sicura. Solitamente gli URL fittizzi hanno al loro interno punti o underscore che ad un primo sguardo non risaltano. (Es. https://face.book.com)
  Dopodichè viene richiesto all'utente di inserire le proprie credenziali o i propri dati, come se si trattasse di un login o di un cambio password, cosa che ovviamente non accade. I dati arrivano quindi in mano al soggetto attaccante, che avrà quindi libero arbitrio sul loro utilizzo.

Phishing
Phishing Youtube

Difesa dagli attacchi

Per poter garantire l'integrità dei dati all'interno di una base di dati è necessario che essa sia garantita a livello di DBMS e livello di dato stesso.

Anche se piuttosto utopico, spesso per la protezione di un database nelle imprese vengono usate password di default o facilmente indovinabili.

- Questo in realtà è un problema facilmente risolvibile, attraverso l'attuazione di una policy che preveda il ricambio regolare di password, con una lunghezza e con un numero di simboli e caratteri alfanumerici.

Invece per potersi difendere da attacchi come l'SQL Injection è importante accertare una validazione dei dati forniti dall'utente.

- Come primo passo può essere utilizzata la funzione mysql real escape string, che andrà a rimuovere i caratteri speciali dalla stringa, eliminando quindi gli attacchi concernenti l'utilizzo di apici o uguali.
- Invece come secondo passo è consigliato l'utilizzo di una funzione **is numeric**, che andrà a controllare semplicemente se l'input inserito sia di tipo numerico, così da eludere, insieme alla real escape string, l'utilizzo di attacchi con numeri.

La sicurezza nei DBMS però non si ferma al semplice controllo della validazione dei dati, ma deve essere gestita attraverso una serie di passaggi, in modo tale da garantire l'integrità.

- Il primo passaggio che si effettua è il **partizionamento**.
  Il partizionamento consiste nella suddivisione del database in più parti, ognuna caratterizzata da un livello di sensibilità. In questo modo il database verrà diviso in componenti più piccoli, che risulteranno più sicuri, oltre che veloci, da gestire. Aumenterà quindi la ridondanza, garantendo l'integrità dato l'accesso da DB separati.
- Il secondo passaggio che andremo ad attuare è la **cifratura** dei dati.
  Cifrare i dati all'interno di un database permette di proteggere i dati dal furto o dalla manomissione, dal momento che non sono interpretabili se non tramite la chiave di decifratura.
  Molto comune è l'utilizzo di funzioni hash, che permettono di attuare una forma di crittografia a "senso unico", quindi rendendo l'operazione di decifrazione impossibile.
  Questo tipo di criptazione è utilizzato nel caso delle password, in modo tale che un hacker non possa interpretarle al momento dell'accesso al DB.
  Gli algoritmi più utilizzati sono l'MD5, lo SHA1 (anche se ormai obsoleto) e lo SHA2.
  Criptazione

- Successivamente il passaggio da compiere per poter salvaguardare i dati è il **lock di integrità**, attraverso l'utilizzo di transazioni.

  Una transazione è una sequenza di operazioni atomiche che devono essere eseguite senza concludersi con uno stato intermedio.

  I lock, letteralmente "blocchi", sono un meccanismo usato per disciplinare l'accesso a risorse condivise.
  Semplicemente i dati vengono marcati attraverso delle "etichette", che ne definiscono sensibilità e livelli di accesso.
  Per questo meccanismo viene implementato il modulo **Lock Manager** nel DBMS, che tiene traccia delle risorse in uso, delle transazioni che le stanno usando e delle transazioni che ne

hanno fatto richiesta.
In questo modo se dovesse venire riscontrata qualche anomalia verrà eseguita l'istruzione di **abort**, che può essere a runtime o di sistema, per concludere la transizione. L'abort a runtime si verifica al momento del riscontro di un'anomalia da parte del DBMS, mentre l'abort di sistema viene lanciato nel caso di un'interruzione brusca per intervento esterno, per bug o per spegnimento del computer.
L'entrata in funzione del meccanismo di abort prevede l'avvio di un altro meccanismo, il **rollback**, un'operazione che permette di riportare la base di dati a uno stato precedente.
Rollback
Lock di integrità
Transizione

- L'ultimo passo che si compie al momento della progettazione del database per proteggere l'integrità dei dati è l'utilizzo di **Finestre e viste**.
  Le viste sono fondate su una query di tipo SELECT, il cui risultato può essere utilizzato come se fosse una tabella.
  Le finestre invece sono semplicemente le aree dove è possibile svolgere determinate operazioni, in questo caso di visualizzazione di dati.
  L'utilizzo delle finestre e delle viste permette quindi di creare sottoinsiemi del DB dove possono essere inserite le informazioni in maniera parziale, in modo tale da non consentire all'utente di avere accesso a dati riservati.
  Viste

Sicurezza del DBMS

Attacco ai database (Wiki)
Protezione delle basi di dati (Wiki)