Università degli Studi di Firenze

Scuola di Ingegneria - Dipartimento di Ingegneria dell'Informazione

Tesi di Laurea Triennale in Ingegneria Informatica

Progettazione di un servizio di Intrusion Detection System per Data Center tramite l'approccio Network Functions Virtualization

Candidato
Lorenzo Avenante

Relatore
Prof. Francesco Chiti

Anno Accademico 2022/23

Indice

Introduzione **i**

1 Teoria degli argomenti trattati **1**

1.1 La virtualizzazione 1

1.2 La Network Functions Virtualization 4

1.3 L'Intrusion Detection System 6

1.4 Cloud Computing 8

1.5 Data Center e Cloud Data Center 11

2 Progettazione di un IDS con l'approccio NFV **13**

2.1 Descrizione dell'architettura di un IDS sviluppato tramite l'approccio NFV e il suo funzionamento 13

2.2 Confronto tra un IDS tradizionale e l'IDS implementato con l'approccio NFV 18

3 Analisi del funzionamento dell'IDS di un caso di studio **20**

3.1 Definizione di attacco informatico e di attacco DoS 20

3.2 Caso di studio: attacco DoS a un Data Center avente IDS implementato con l'approccio NFV 23

3.3 Analisi qualitative e quantitative del caso di studio 25

3.4 Conclusioni 28

Bibliografia **30**

Introduzione

In questa tesi verrà proposto un modello di architettura per il servizio di sicurezza informatica chiamato Intrusion Detection System (IDS), attraverso l'approccio della virtualizzazione delle funzioni di rete (Network Functions Virtualization, NFV).
Nel primo capitolo verranno affrontati gli argomenti teorici utili allo sviluppo della tesi quali: la virtualizzazione, le Virtualized Network Functions (VNFs), la NFV, l'IDS e il Cloud Computing.
Nel secondo capitolo verrà descritta l'architettura dell'IDS attraverso l'approccio della NFV per un Data Center, poi verrà descritto il suo funzionamento e infine verrà effettuato un confronto con i modelli di IDS classici.
Nel terzo capitolo sarà definito che cos'è un attacco informatico, poi verrà descritto il comportamento di un Data Center, in cui è stato implementato l'IDS descritto nella tesi, sotto un attacco di tipo DoS e infine sarà valutato l'impatto delle risorse di rete utilizzate dall'IDS.

Capitolo 1

Teoria degli argomenti trattati

In questo capitolo verranno affrontati i concetti e le definizioni utili per lo sviluppo della tesi.

1.1 La virtualizzazione

La virtualizzazione, in informatica, è il processo che simula l'interfaccia dei dispositivi fisici attraverso suddivisione, aggregazione ed emulazione.
Con suddivisione si intende la creazione di diversi dispositivi virtuali da un unico dispositivo fisico, con aggregazione si intende la creazione di un unico dispositivo virtuale da tanti dispositivi fisici e con emulazione si intende la creazione di un dispositivo virtuale differente dal dispositivo fisico.
In altre parole, la virtualizzazione può essere definita come il livello di astrazione software posizionato tra il sistema operativo e le componenti hardware.
Si definisce l'Hypervisor come un piccolo sistema operativo specializzato, che gira su una macchina fisica, il quale permette la suddivisione e la allocazione delle risorse fisiche in risorse virtuali. Gli Hypervisor sono i responsabili della creazione e della gestione di macchine virtuali.

Definiamo le macchine virtuali (VM) come software che, attraverso un processo di virtualizzazione, creano un ambiente virtuale che emula tipicamente il comportamento di una macchina fisica. Dunque una macchina virtuale è composta da una propria CPU virtuale, da una propria memoria virtuale e da un proprio sistema operativo. [8]

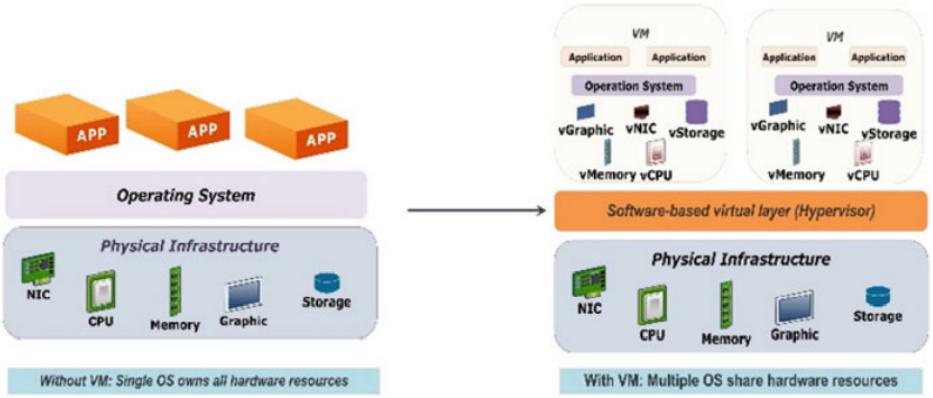

Figura 1.1: Confronto tra macchine standard e macchine virtuali (VM)

Una volta definito il concetto di virtualizzazione, è necessario rispondere alla domanda "perché virtualizzare?"
I principali aspetti positivi della virtualizzazione sono:

- Risparmio economico: risparmio di spazio, di manutenzione e di energia, poiché al posto di gestire tante macchine fisiche ne vengono gestite poche, perché sono virtualizzate.
- Efficienza: per un'unica macchina fisica risulta quasi impossibile sfruttare tutte le sue risorse computazionali, per questo motivo, se virtualizzata, più macchine virtuali avranno accesso alle sue risorse.
- Migrazione semplice: grazie alla versatilità della virtualizzazione, risulta semplice migrare una macchina virtuale da un dispositivo fisico a un altro.

Le macchine virtuali non rappresentano l'unica tecnologia per virtualizzare, anzi, ad oggi nel mondo del Cloud è preferibile l'utilizzo dei Container.
I Container si differenziano dalle macchine virtuali poiché all'interno della stessa infrastruttura fisica essi condividono lo stesso Kernel, infatti risultano più leggeri delle macchine virtuali perché non virtualizzano l'hardware ma utilizzano il Kernel della macchina fisica in modo efficiente attraverso l'isolamento dei processi. Inoltre, la gestione delle risorse virtuali non è più a carico dell'Hypervisor ma del Container Engine.

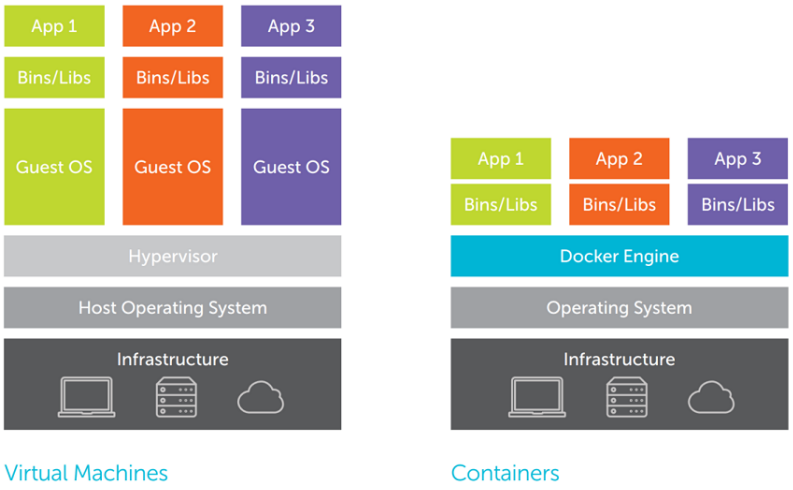

Figura 1.2: Confronto tra macchine virtuali e Container di tipo Docker

In generale si può dire che i Container risultano avere performance migliori dal punto di vista delle prestazioni e della scalabilità. Esistono comunque delle situazioni nelle quali non risulta opportuno utilizzare i Container, per esempio nel caso di applicazioni che fanno uso di dati critici, per questioni di sicurezza, è consigliato utilizzare le VM, poiché i Container utilizzano sempre i privilegi "root". [3]

1.2 La Network Functions Virtualization

La Network Functions Virtualization (NFV) può essere tradotta come "virtualizzazione di funzioni di rete" ed è una tecnica che ha come scopo quello di virtualizzare un'intera classe di funzioni dei componenti di rete, disaccoppiando le funzioni di rete dalle apparecchiature della stessa.
Una funzione di rete diventa un'istanza virtualizzata di un programma software personalizzato chiamato Virtual Function Network (VNF). Questo oggetto può essere creato su richiesta, messo in funzione ovunque sia necessario, senza la necessità di installare nuove apparecchiature e, inoltre, può essere spostato a piacimento nella rete e terminato quando la sua funzione non è più necessaria.
L'NFV consente di eseguire le funzioni di rete come istanze software nelle macchine virtuali o nei Container.
L'European Telecommunications Standards Institute (ETSI) ha definito l'architettura di riferimento per le NFV la quale è caratterizzata da tre blocchi: Network Functions Virtualization Infrastructure (NFVI), Virtual Network Functions (VNFs) e Network Function Virtualization management and orchestration (NFV-MANO) .
La NFVI è il blocco composto dalla combinazione dell'infrastruttura fisica e delle risorse virtualizzate, solitamente un Hypervisor è responsabile della gestione dell'ambiente virtuale, nel quale le VNFs vengono implementate, gestite, eseguite e terminate.
Il blocco delle VNFs è composto da una moltitudine di VNF e di Element Management System (EMS), i quali sono elementi che agiscono per la gestione delle VNFs. Una VNF può essere implementata in un'unica macchina virtuale oppure distribuita in più macchine, quando viene scelta la seconda opzione, è necessario che le funzioni vengano elaborate secondo un certo or-

dine poiché è possibile che alcune delle funzioni abbiano dipendenza da altre. Il NFV-MANO è il blocco responsabile per la gestione delle VNF. Gestisce dunque l'implementazione e il funzionamento delle VNF, inoltre, dispone di un database che memorizza informazioni utili per determinare le proprietà del ciclo di vita di servizi e risorse.

In alcuni casi viene definito anche un quarto blocco denominato Operation Support System (OSS) and Business Support System (BSS), il quale è responsabile del coordinamento con il sistema di rete tradizionale, infatti supervisiona il regolare funzionamento dei servizi di NFVI, VNFs e NFV-MANO. Si notifica che l'architettura ETSI per le NFV è in continua evoluzione e che alcune componenti di essa potrebbero essere aggiunte, sostituite o eliminate. [8] [5]

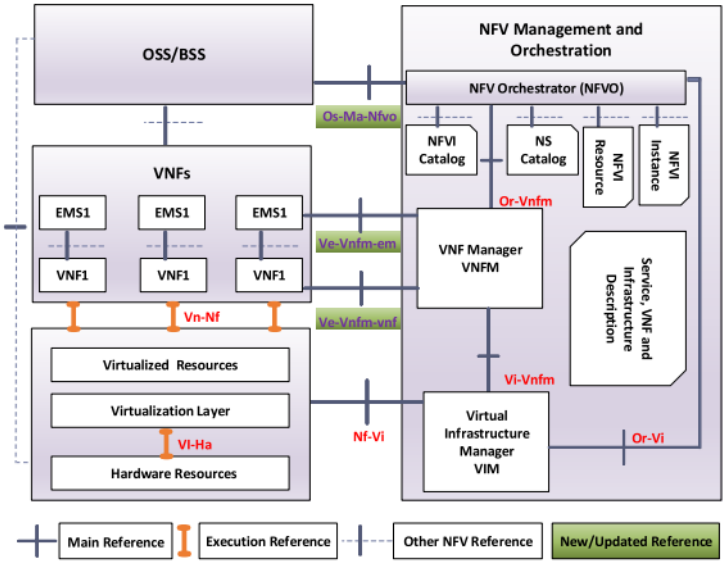

Figura 1.3: Architettura ETSI per le NFV

1.3 L'Intrusion Detection System

L'Intrusion Detection System (IDS) è un dispositivo hardware o un'applicazione software che monitora la rete da attività malevole o violazioni di policy. Un IDS è caratterizzato da due operazioni principali: monitoraggio e processamento dei dati.
Una volta installato un IDS in una rete, esso procede al monitoraggio del traffico della rete interna sugli end-point (Client e Server) o sui dispositivi di rete (Router, Switch, Proxy, ecc..) oppure in modalità ibrida monitorando entrambi.
Il monitoraggio viene svolto da dei moduli chiamati Agenti, i quali sono responsabili di raccogliere il traffico rete sul singolo dispositivo a cui sono affidati, essi sono gestiti da un Master Agent che sarà responsabile dell'invio dei dati raccolti al Modulo Centrale IDS, il quale sarà resposabile del processamento dei dati e della classificazione del traffico in normale o anomalo.
Le modalità con le quali vengono processati i flussi di traffico sono tre: analisi basata sulla firma, analisi basata sulle anomalie statistiche e analisi basata sui protocolli.

- L'analisi basata sulla firma riesce ad individuare il traffico anomalo generato da dei tipi di attacchi riconosciuti, la loro "firma" può essere un certo tipo di pacchetto inviato, una certa sequenza di pacchetti o altre modalità riconosciute come malevole. Questa modalità di analisi risulta poco efficacie nel caso di attacchi non conosciuti.

- L'analisi basata sulle anomalie statistiche utilizza la Machine Learning. Per essere efficacie l'IDS ha bisogno di un periodo di apprendimento, nel quale impara il comportamento della rete e ne definisce uno "normale" basandosi sulla banda usata, i tipi di pacchetti inviati e le connessioni

fatte tra i dispositivi. Una volta completato l'apprendimento, l'IDS sarà in grado di riconoscere il comportamento non "normale". Questa tecnica può produrre molti falsi positivi (traffico ritenuto anomalo, il quale risulta legittimo), soprattutto quando l'IDS non ha ancora appreso in maniera capillare il comportamento della rete.

- L'analisi basata sui protocolli funziona attraverso delle tabelle di flusso di traffico predeterminate, se il traffico non rientra nelle regole viene definito anomalo.

L'IDS non effettua nessuna azione di "remediation", ovvero, se notifica che il traffico è anomalo, non provvede in alcun modo a mettere in sicurezza la rete.
Per questo motivo, si definisce un nuovo dispositivo chiamato Intrusion Prevention System (IPS), il quale è un'estensione dell'IDS con l'abilità di bloccare i flussi di traffico ritenuti malevoli. Esso dunque, aggiunge alle funzioni di monitoraggio e processamento dei dati, quella di risanamento della rete. Questa operazione verrà svolta da dei moduli dedicati, i quali bloccheranno o limiteranno i flussi di traffico di rete ritenuti malevoli.
In ambito di sicurezza informatica esiste un altro dispositivo specializzato per il monitoraggio del traffico di rete chiamato Firewall. Questo si differenzia dall'IDS perché è responsabile di bloccare o lasciar passare il traffico che proviene dall'esterno verso la rete interna e quello dalla rete interna verso l'esterno, mentre un IDS monitora i pacchetti che vengono scambiati nella rete interna.
Nel caso in cui un attaccante riuscisse a compromettere una macchina interna a una rete, questa, se fosse monitorata solo da un Firewall, non riuscirebbe ad accorgersi che al suo interno vi è un intruso. Se invece nella stessa rete

viene implementato anche un IDS, sarà possibile accorgersi di questo tipo di attacchi. [6] [7]

1.4 Cloud Computing

Il Cloud Computing è un modello per abilitare, tramite la rete, l'accesso diffuso, agevole e su richiesta, ad un insieme condiviso e configurabile di risorse di elaborazione (ad esempio reti, server, memoria, applicazioni e servizi) che possono essere acquisite e rilasciate rapidamente e con minimo sforzo di gestione o di interazione con il fornitore di servizi [**?**].
Questo modello Cloud è composto da cinque caratteristiche essenziali, tre modalità di servizio e quattro modelli di distribuzione.
Caratteristiche essenziali:

- Self-service su richiesta. Un consumatore può acquisire unilateralmente e automaticamente le necessarie capacità di calcolo, come tempo macchina e memoria, senza richiedere interazione umana con i fornitori di servizi.

- Ampio accesso in rete. Le capacità sono disponibili in rete e accessibili attraverso meccanismi standard che promuovono l'uso attraverso piattaforme eterogenee come client leggeri o pesanti (come ad esempio telefoni mobili, tablet, laptops e workstations).

- Condivisione delle risorse. Le risorse di calcolo del fornitore sono messe in comune per servire molteplici consumatori utilizzando un modello condiviso (multi-tenant), con le diverse risorse fisiche e virtuali assegnate e riassegnate dinamicamente in base alla domanda. Dato il senso di indipendenza dalla locazione fisica, l'utente generalmente non ha

controllo o conoscenza dell'esatta ubicazione delle risorse fornite, ma può essere in grado di specificare la posizione ad un livello superiore di astrazione (ad esempio, paese, stato o data center).

- Elasticità rapida. Le risorse possono essere acquisite e rilasciate elasticamente, in alcuni casi anche automaticamente, per scalare rapidamente verso l'esterno o l'interno in relazione alla domanda. Al consumatore, le risorse disponibili spesso appaiono illimitate e disponibili in qualsiasi quantità, in qualsiasi momento.

- Servizio misurato. I sistemi Cloud controllano automaticamente e ottimizzano l'uso delle risorse, facendo leva sulla capacità di misurazione ad un livello di astrazione appropriato per il tipo di servizio (ad esempio memoria, elaborazione, larghezza di banda e utenti attivi). L'utilizzo delle risorse può essere monitorato, controllato e segnalato, fornendo trasparenza sia per il fornitore che per l'utilizzatore del servizio.

Modelli di servizio:

- Software come Servizio (SaaS: Software as a Service). La facoltà fornita al consumatore è quella di utilizzare le applicazioni del fornitore funzionanti su un'infrastruttura Cloud. Le applicazioni sono accessibili da diversi dispositivi attraverso un'interfaccia leggera (thin client), come ad esempio un'applicazione email su browser, oppure da programmi dotati di apposita interfaccia. Il consumatore non gestisce o controlla l'infrastruttura cloud sottostante, compresi rete, server, sistemi operativi, memoria, e nemmeno le capacità delle singole applicazioni.

- Piattaforma come Servizio (PaaS: Platform as a Service). La facoltà fornita al consumatore è quella di distribuire sull'infrastruttura Cloud

applicazioni create in proprio oppure acquisite da terzi, utilizzando linguaggi di programmazione, librerie, servizi e strumenti supportati dal fornitore. Anche in questo caso il consumatore non gestisce né controlla l'infrastruttura cloud sottostante, ma ha il controllo sulle applicazioni ed eventualmente sulle configurazioni dell'ambiente che le ospita.

- Infrastruttura come Servizio (IaaS: Infrastructure as a Service). La facoltà fornita al consumatore è quella di acquisire elaborazione, memoria, rete e altre risorse fondamentali di calcolo, inclusi sistemi operativi e applicazioni. Il consumatore non gestisce né controlla l'infrastruttura cloud sottostante, ma controlla sistemi operativi, memoria, applicazioni ed eventualmente, in modo limitato, alcuni componenti di rete (esempio firewalls).

Modelli di Distribuzione:

- Cloud privato. L'infrastruttura Cloud è fornita per uso esclusivo da parte di una singola organizzazione comprendente molteplici consumatori (ad esempio filiali). Può essere posseduta, diretta e gestita dall'organizzazione stessa, da una società terza o da una combinazione delle due, e può esistere dentro o fuori le proprie sedi.

- Cloud comunitario. L'infrastruttura Cloud è fornita per uso esclusivo da parte di una comunità di consumatori di organizzazioni con interessi comuni (ad esempio missione, requisiti di sicurezza, vincoli di condotta e di conformità). Può essere posseduta, diretta e gestita da una o più delle organizzazioni della comunità, da una società terza o una combinazione delle due e può esistere dentro o fuori le proprie sedi.

- Cloud pubblico. L'infrastruttura Cloud è fornita per un uso aperto a qualsiasi consumatore. Può essere posseduta, diretta e gestita da

un'azienda, da un'organizzazione accademica o governativa oppure da una combinazione delle precedenti. Esiste dentro le sedi del fornitore Cloud.

- Cloud ibrido. L'infrastruttura è una composizione di due o più infrastrutture Cloud (privata, comunitaria o pubblica) che rimangono entità distinte, ma unite attraverso tecnologie standard o proprietarie, che abilitano la portabilità di dati e applicazioni (ad esempio per bilanciare il carico di lavoro tra Cloud).

[4]

1.5 Data Center e Cloud Data Center

Un data center è una struttura che centralizza le operazioni e le apparecchiature IT condivise da un'organizzazione allo scopo di archiviare dati ed eseguire applicazioni. Poiché ospitano le risorse più critiche di un'organizzazione, i data center sono fondamentali per la continuità delle operazioni quotidiane. Di conseguenza, la sicurezza e l'affidabilità dei data center e delle loro informazioni sono tra le massime priorità di ogni organizzazione.
Le principali operazioni svolte grazie ad un Data Center sono: archiviazione, backup e ripristino dei dati, applicazioni di produttività (per esempio e-mail) transazioni di e-commerce, Big Data, apprendimento automatico e intelligenza artificiale.
In passato, i data center erano infrastrutture fisiche altamente controllate, con l'avvento del Cloud questo modello è cambiato. Ad eccezione dei casi in cui le restrizioni normative richiedono un Data Center on-premise (ovvero con l'infrastruttura fisica di proprietà dell'azienda stessa) senza connessioni a Internet, la maggior parte delle moderne infrastrutture di data center si è

evoluta da server fisici on-premise a infrastrutture virtualizzate che supportano applicazioni e carichi di lavoro in ambienti multi-cloud.
Dunque una azienda che intende sviluppare il proprio Data Center in Cloud, acquista da un provider un servizio di tipo IaaS, Paas o SaaS, dipendentemente dalle proprie necessità, e sviluppa il progetto. L'azienda non dovrà preoccuparsi dei costi di mantenimento e della affidabilità dell'infrastruttura che saranno totalmente a carico del provider.

Capitolo 2

Progettazione di un IDS con l'approccio NFV

In questo capitolo verrà prima proposto un modello di architettura di un IDS con l'utilizzo della NFV e successivamente verrà confrontato questo modello con un IDS standard.
Il contesto teorico nel quale si sviluppa l'architettura e si definisce il comportamento dell'IDS è quello di un Data Center. In particolare ci immedesimiamo in una azienda che fornisce il servizio di IDS attraverso moduli software e non i classici dispositivi hardware.

2.1 Descrizione dell'architettura di un IDS sviluppato tramite l'approccio NFV e il suo funzionamento

In questo capitolo verrà proposta una possibile architettura per un IDS con NFV basata sul modello ETSI e il suo funzionamento.

L'architettura è caratterizzata da tre livelli:

- Il primo livello verrà chiamato "Infrastructure": fanno parte di questo livello tutte le risorse fisiche e virtuali, oltre a un Hypervisor che sarà responsabile della gestione delle risorse virtuali.

- Il secondo livello verrà chiamato "Virtual Network Functions": fanno parte di questo livello tutte le diverse VNFs e gli elementi di esecuzione di quest'ultime.
  Definiamo dunque i tre moduli che eseguono le funzioni di monitoraggio, salvataggio e processamento.
  Il modulo "Agent" è responsabile dell'attività di monitoraggio, vengono istanziati il numero di Agent nella rete in rapporto 1:1 con i dispositivi all'interno di essa (client, server, switch, router, ecc..).
  Il modulo "Storage" è responsabile dell'attività di salvataggio delle informazioni raccolte nella rete da parte degli Agent. Si noti che è possibile che esista più di un'istanza Storage all'interno della rete.
  Il modulo "Processor" è responsabile del processamento dei dati raccolti dagli Agent e salvati nel modulo Storage. Si noti, come per il modulo Storage, che è possibile che esista più di un'istanza Processor all'interno della rete.

- Il terzo livello verrà chiamato "Management and Orchestration": fanno parte di questo livello i moduli chiamati: Controller, Orchestrator, Machine Learning e IDS Central Module.
  Il modulo "Controller" ha il compito di allocare e de-allocare le risorse fisiche e virtuali, in particolare è responsabile della creazione e distruzione dei moduli Agent, Storage e Processor, questo è reso possibile grazie alla conoscenza totale della rete attraverso una tabella che asso-

cia le risorse computazionali agli elementi creati.

Il modulo "Orchestrator" è il responsabile del corretto funzionamento delle funzioni virtualizzate di rete e dei moduli a loro legate (Agent, Storage, Processor), inoltre, possiede il ruolo decisionale su dove installare i vari moduli per minimizzare la distanza tra questi e il dispositivo a loro associati.

Il modulo "Machine Learning" eseguirà algoritmi di apprendimento del comportamento della rete sui dati raccolti, affinché la metodologia di identificazione di un intruso nella rete diventi sempre più efficace.

Infine, il "IDS Central Module" è responsabile dell'azione di rilevamento di un intruso nella rete. La sua funzionalità principale è quella di decidere quali dispositivi analizzare.

La caratteristica principale di questa architettura è che non prevede l'implementazione di specifiche risorse hardware nella rete, infatti, tutti i moduli presenti in essa (Agent, Controller, Orchestrator, ecc..) sono risorse software che possono essere implementate in un qualsiasi ambiente virtuale. Disaccoppiando i moduli software dai corrispettivi hardware, si rende questa architettura molto elastica ad aumenti o diminuzioni di dispositivi da monitorare, inoltre, eventuali aggiornamenti del parco macchine potrebbero non dare disservizi, in quanto sarebbe possibile migrare i moduli software all'interno della rete per garantire la continuità del servizio.

Andiamo ora a descrivere in modo più dettagliato il funzionamento dell'IDS. Per installare l'IDS nella rete è necessario installare in una macchina sicura l'IDS Central Module, dopodiché verrà eseguito uno scansionamento di tutti gli elementi della rete, per questo procedimento sarà necessario che l'IDS abbia accesso a un dispositivo con visibilità massima, per esempio un Domain Controller. Dopo aver effettuato lo scansionamento verranno creati i

moduli Agent con un rapporto 1:1 con i dispositivi nella rete (si identificano come dispositivi tutti gli elementi con un proprio Ip, dunque anche istanze di macchina virtuale), questi moduli saranno caratterizzati dall'esecuzione delle funzioni di monitoraggio.
Definiamo due diverse funzioni di monitoraggio:

- Monitoraggio in modalità Standard: prevede un utilizzo di risorse computazionali e di rete al massimo del 10%. La funzione di monitoraggio scansionerà il traffico in uscita e in entrata del dispositivo, inviando i pacchetti monitorati al modulo Storage più vicino.

- Monitoraggio in modalità Hunting: non prevede un limite massimo all'utilizzo di risorse computazionali e di rete. Anche in questo caso il risultato della funzione di monitoraggio del traffico del dispositivo sarà inviata al modulo Storage più vicino. Questa modalità viene attivata in due circostanze: quando l'IDS Central Module notifica un intruso, oppure quando un dispositivo nuovo si connette alla rete.

In ogni caso la procedura del monitoraggio dei dispositivi prevede che tutti i dati in uscita o in entrata in un dispositivo vengano inoltrati anche all'Agent a lui associato, quest'ultimo esegue una minima operazione di processamento dei dati ricavando le seguenti informazioni: la tipologia, il volume, la provenienza o la destinazione dei pacchetti. Dopodiché l'Agent provvede all'invio delle informazioni ottenute al modulo Storage più vicino, nel caso in cui tutti i moduli Storage siano saturi, tramite il lavoro del Controller e dell'Orchestrator è possibile istanziare un nuovo modulo Storage.
Il lavoro di monitoraggio non produrrebbe alcun risultato se non venisse processato. Andiamo ad analizzare come vengono processati i dati raccolti.
L'IDS Central Module possiede una lista di tutti gli elementi di rete e defini-

sce di quale dispositivo vanno processati i dati raccolti e con quale priorità. Una volta determinato il dispositivo da analizzare, il modulo Orchestrator invia un messaggio a un modulo Processor affinché possa svolgere le sue funzioni di processamento.
La metodologia secondo la quale viene scelto quale Processor deve eseguire l'analisi è composta da due fattori: la distanza tra il modulo Processor e il modulo Storage che contiene i dati da processare e la quantità di elementi in coda da processare. Con questo metodo si esaltano i vantaggi dell'utilizzo della NFV, infatti, nel caso in cui siano stati allocati troppi moduli Processor, sarà possibile liberare le risorse computazionali, in caso contrario sarà possibile istanziare un nuovo modulo Processor cosicché la funzione di processamento possa essere svolta senza troppi rallentamenti, come per il modulo Storage.
Nel caso in cui vengano richieste al modulo Processor due o più analisi in modo concorrenziale, esso eseguirà prima tutti i processamenti con priorità alta e successivamente quelli con priorità bassa.
Definiamo due modalità di processamento dei dati:

- Processamento singolo: si analizzano i dati raccolti da un singolo dispositivo.
- Processamento concatenato: si analizzano i dati raccolti da diversi dispositivi concatenati tra loro.

I due metodi di processamento, come suggerisce il loro nome, differiscono solamente per il numero di dispositivi di cui si devono analizzare i dati, quindi, nel caso di processamento concatenato, il modulo Processor potrà dare il risultato solo dopo aver analizzato i dati di tutti i dispositivi coinvolti.
In entrambi i casi la procedura del processamento dei dati è la seguente: per

prima cosa si esegue un'analisi dei dati basata sulla firma, se questa dà esito positivo il dispositivo o tutti i dispositivi concatenati verranno identificati come intrusi e sarà necessaria una attività di risanamento della rete da parte o di un nuovo modulo a cui saranno affidate le attività di "remediation" oppure all'intervento dell'uomo.
Nel caso in cui l'esito della prima analisi fosse negativo, si procede all'analisi basata sulle anomalie statistiche (analisi comportamentale basata su Machine Learning), se questa dà esito positivo il dispositivo o tutti i dispositivi concatenati verranno identificati come "sospetti" ed entreranno in modalità di monitoraggio Hunting. In questo caso saranno applicate le tecniche di risanamento solo dopo una seconda conferma, poiché l'utilizzo del metodo di analisi comportamentale prevede un considerevole tasso di falsi positivi, ovvero di dispositivi identificati dal sistema come intrusi, che però risultano legittimi.
Nel caso in cui l'analisi basata sulle anomalie statistiche dovesse portare a un falso positivo, il modulo Processor invierà i dati al modulo di Machine Learning cosicché possa utilizzare i dati per migliorare l'analisi comportamentale.

2.2 Confronto tra un IDS tradizionale e l'IDS implementato con l'approccio NFV

Un IDS tradizionale è caratterizzato dall'utilizzo di dispositivi hardware in corrispondenza dei nodi di rete cruciali, quali gli hub e gli switch. Questi dispositivi eseguono le operazioni di monitoraggio, salvataggio e processamento dei pacchetti che vengono inoltrati dai dispositivi di rete e sono gestiti da un dispositivo IDS centrale. Si noti come sia necessario che, durante la realizzazione della rete, debba essere tenuto conto delle risorse di banda che

il dispositivo IDS consuma.

I limiti principali dell'implementazione di un IDS tradizionale in una rete sono: l'ottimizzazione delle risorse e la poca elasticità a cambi di topologia, quali l'aumento o la diminuzione del parco macchine da scansionare.

Il modello di IDS implementato con NFV risulta superare i limiti dell'IDS tradizionale, poiché i moduli che eseguono le funzioni di monitoraggio, salvataggio e processamento (Agent, Storage e Processor) possono essere istanziati quando sono necessari, possono essere de-allocati quando non servono più e possono essere migrati nella rete per ottimizzare l'utilizzo delle risorse.

Si sottolinea che sia gli IDS tradizionali che l'IDS proposto in questa tesi sono *resource-hungry*, ovvero, che necessitano di molte risorse di rete. In particolare gli IDS tradizionali hanno bisogno di molta banda, mentre l'IDS con NFV richiede determinate risorse computazionali, di memoria e di banda.

Capitolo 3

Analisi del funzionamento dell'IDS di un caso di studio

In questo capitolo verrà descritto in primis che cos'è un attacco informatico e il tipo di attacco DoS, successivamente sarà descritto il comportamento dell'IDS con NFV durante un attacco di tipo DoS e infine verranno riassunti i miglioramenti apportati dall'approccio NFV all'IDS rispetto ai modelli classici.

3.1 Definizione di attacco informatico e di attacco DoS

Un attacco informatico è un tentativo malevolo e intenzionale da parte di un individuo o di un'organizzazione di violare il sistema informativo di un altro individuo o azienda. Di solito, l'attaccante viola la rete della vittima per ottenere qualche tipo di vantaggio.
Le motivazioni di un attacco possono essere di svariato tipo, in queste motivazioni ci sono tre categorie principali: criminale, politica e personale.

Gli aggressori a sfondo criminale cercano guadagni finanziari attraverso il furto di denaro, il furto di dati o l'interruzione dell'attività. Allo stesso modo, le persone con motivazioni personali, come i dipendenti attuali o gli ex dipendenti scontenti, rubano denaro, dati o provano a distruggere il sistema informatico della società a cui sono legati. Gli attaccanti con motivazioni socio-politiche cercano l'attenzione per le loro cause. Di conseguenza, rendono noti al pubblico gli attacchi, questo fenomeno è conosciuto anche come *hacktivismo*, Anonymous ne è il gruppo più famoso.

Un'altra motivazione di attacchi informatici è la cyberwarfare, attacchi effettuati per danneggiare una nazione con cui si è in guerra da un punto di vista cibernetico e non.

I tipi più comuni di attacchi informatici sono: i malware (virus, worm, ransomware, ecc..), phishing, Man in the Middle (MitM), Denial of Service (DoS) e Distributed Denial of Service (DDoS).

Andiamo adesso a descrivere il tipo di attacco DoS.

L'attacco informatico di tipo DoS (Denial of Service) si prefigge come obiettivo quello di aumentare il traffico dati di un dispositivo o di una rete in modo esponenziale affinché le risorse computazionali e di banda non siano sufficienti per erogare i servizi presenti in essa.

Solitamente questo tipo di attacco viene utilizzato per due motivi principali: per creare un danno reputazionale oppure come diversivo per il lancio di un altro attacco.

Un'azienda che viene colpita da un attacco DoS, non potrà erogare tutti i servizi informatizzati durante l'attacco e durante la procedura di risanamento della rete. Questo comporta, oltre che a un disservizio per i propri clienti, la possibilità di incorrere in sanzioni amministrative, in contenziosi legali e alla svalutazione dell'azienda stessa da un punto di vista economico e repu-

tazionale.

Oltre al danno reputazionale, è possibile che, durante un attacco di tipo DoS, gli autori lancino un nuovo attacco, per esempio un ransomware, poiché quando la rete si trova in stato di sovraccarico, oltre al malfunzionamento dei servizi erogati dall'azienda attaccata, potrebbero subire delle interruzioni tutte quelle attività legate alla sicurezza informatica, cosicché sia più facile per gli attaccanti lanciare un attacco senza che questo venga rilevato.

Per effettuare un attacco di tipo DoS è necessario che l'hacker invii una grandissima mole di pacchetti da un dispositivo a una macchina bersaglio, in cui è attivo il servizio che vogliamo rendere inaccessibile, utilizzando una determinata tecnica, per esempio "Syn-Flood".

Se l'attacco viene effettuato da più di un dispositivo verso una o più macchine bersaglio, questo viene chiamato DDoS (Distributed Denial of Service).

Solitamente questi attacchi, DoS e DDoS, vengono effettuati contro l'infrastruttura informatica di una azienda. Il flusso di traffico dei pacchetti inviati con intenzioni malevole sarà dall'esterno verso l'interno della rete aziendale. Di norma il dispositivo di sicurezza che filtra i pacchetti che entrano ed escono dentro una rete aziendale è il firewall, nel caso in cui questo dispositivo non dovesse rilevare l'attacco, se l'azienda ha il servizio di IDS attivo, avrà comunque alte chance di rilevare l'attacco.

Nel caso in cui un dispositivo interno alla rete dovesse essere compromesso, se questo lanciasse un attacco di tipo DoS a un altro dispositivo appartenente alla stessa rete, il firewall non sarebbe in grado di rilevare l'attacco in alcun modo, mentre l'IDS, anche in questo caso, avrebbe alte chance di rilevare l'attacco. [2] [1]

3.2 Caso di studio: attacco DoS a un Data Center avente IDS implementato con l'approccio NFV

Il contesto teorico in cui andiamo a sviluppare il caso di studio è un Data Center aziendale. I dipendenti dell'azienda si possono connettere alla rete interna tramite cablaggio o in modalità wireless, questi possono fare accesso alle risorse del database aziendale o ai servizi erogati.
Supponiamo che un computer aziendale sia stato compromesso attraverso un malware, e supponiamo che il funzionamento di questo software malevolo sia di far partire un attacco di tipo DoS al server che eroga il servizio di posta elettronica. Andiamo a simulare in che modo il modello di IDS che utilizza la NFV possa rilevare tale attacco.
Al tempo t_0 il sistema IDS non identifica alcun dispositivo connesso alla rete come intruso, inoltre, il dispositivo compromesso non si trova in modalità di monitoraggio "Hunting".
Al tempo t_1 il malware si attiva e fa paritre un attacco di tipo DoS verso il server mail aziendale.
Al tempo t_2 l'Agent incaricato di monitorare il dispositivo compromesso svolge la sua funzione e raccoglie i dati, i quali vengono inviati al modulo Storage più vicino. Nel caso in cui nessun modulo Storage avesse disponibilità per immagazzinare tali dati, sarà necessario che l'IDS, attraverso il modulo Controller e Orchestrator, vada a creare un nuovo modulo Storage su cui salvare i dati, questo caso è molto raro.
Al tempo t_3 l'IDS incarica un modulo Processor di eseguire il processamento dei dati raccolti sul dispositivo infetto, questo sarà messo in coda a una lista di processamenti con bassa priorità. Qualora tutte le code dei Processor fos-

sero piene, verrà creato un nuovo modulo Processor che eseguirà le funzioni di processamento dei dati raccolti dall'Agent, questo caso risulta più frequente rispetto a quello del modulo Storage, soprattutto in presenza di numerosi monitoraggi in modalità "Hunting".
Al tempo t_4 il modulo Processor invia i risultati delle proprie funzioni all'IDS Central Module.
Se il processamento dà esito positivo grazie all'analisi sulla firma, significa che il dispositivo scansionato risulta essere un intruso che ha eseguito un attacco noto. In questo caso verrà notificato al responsabile della sicurezza aziendale quale dispositivo è stato riconosciuto come intruso e quale tipologia di attacco è stato effettuato basandosi sulla firma rilevata. In questo caso al tempo t_{4bis} saranno effettuate le operazioni per il risanamento della rete, queste possono essere automatizzate o eseguite tramite intervento umano.
Se il processamento dà esito positivo grazie all'analisi basata sulle anomalie statistiche, significa che potrebbe essere presente nel sistema un intruso, dunque l'IDS attribuisce l'etichetta di "sospetto" al dispositivo monitorato, per cui l'agente eseguirà l'attività di monitoraggio in modalità "Hunting".
Al tempo t_5 il modulo Agent eseguirà le sua operazione di monitoraggio e invierà i dati raccolti a un modulo Storage.
Al tempo t_6 il modulo IDS assegnerà le operazioni di processamento dei dati a un Processor, questa operazione verrà inserita nella coda con priorità alta.
Al tempo t_7 il modulo Processor invia i risultati all'IDS Central Module. Se il processamento dà esito positivo una seconda volta, tale dispositivo verrà identificato come intruso e al tempo t_{7bis} saranno svolte le operazioni di risanamento della rete, in modalità automatica o manuale.
Nel caso in cui il secondo processamento non dia esito positivo, l'IDS Central Module rimuoverà al dispositivo monitorato l'etichetta di "sospetto" e comu-

nicherà al modulo di Machine Learning di allenarsi sui dati raccolti affinché la possibilità di falsi positivi diminuisca.
Invece, nel caso in cui al tempo t_4 il Processor dia esito negativo, l'IDS continuerà a svolgere le sue operazioni in modalità standard.
Nei due casi analizzati in cui il processamento dei dati ha dato esito negativo, affinché l'IDS possa accorgersi che nella propria rete vi è un dispositivo intruso, sarà necessario aspettare almeno un altro ciclo composto dalle operazioni descritte da t_2 a t_4.
Sarà compito degli sviluppatori delle funzioni di processamento quello di minimizzare la possibilità che questo accada, in quanto l'impatto che i falsi positivi hanno sulla rete determina un aumento dell'utilizzo delle risorse da parte dell'IDS per periodi molto limitati, mentre, l'impatto di un vero negativo, ovvero un dispositivo intruso che il sistema IDS non ha identificato come tale, può provocare danni devastanti a tutta la rete.

3.3 Analisi qualitative e quantitative del caso di studio

Eseguiamo adesso un'analisi qualitativa e quantitativa sul tempo di esecuzione, il costo computazionale e il traffico di rete generato dall'IDS nella rete.
Definiamo una topologia caratterizzata da 2 switch da 100 Mb/s in comunicazione tra loro, a cui sono collegati 20 dispositivi ciascuno tra client e server.
Supponiamo che all'interno di questa rete, i dispositivi vengano monitorati dagli Agent con un tempo di campionamento di 5 secondi.
Il tempo necessario affinché l'Agent esegua la funzione di monitoraggio su un dispositivo compromesso che ha iniziato l'attività malevola sarà dunque

inferiore o uguale a cinque secondi, ovvero, $t_2 - t_1 \leq 5$ sec.
In questo lasso di tempo, il costo computazionale sarà dovuto alle funzioni di monitoraggio svolte dall'Agent, il cui impatto non dovrebbe essere rilevante.
Per quanto riguarda il consumo di banda, in questo caso l'IDS necessita di inoltrare un numero prestabilito di pacchetti in uscita dal dispositivo infetto al modulo Agent e successivamente, dopo un piccolo processamento dei dati, ri-inoltrerà questi dati a un modulo Storage.
In questo caso il traffico generato dall'Agent sarà due volte il numero dei pacchetti monitorati.
Supponiamo che l'Agent monitori 1 Mb di pacchetti, l'impatto nella banda sarà di 2 Mb ogni 5 secondi. Moltiplicando il consumo di banda per i 20 dispositivi collegati allo switch, si ha che, durante una situazione di monitoraggio standard, l'IDS consuma 40 Mb ogni 5 secondi, ovvero, 8 Mb/s.
Il tempo necessario affinché il Processor esegua l'analisi sui dati raccolti è dato dal tempo di esecuzione di un singolo processamento per il numero di elementi in coda.
Supponiamo che, in una situazione standard, siano attivi 4 moduli Processor per switch, e che il tempo di esecuzione della funzione di processamento sia di un secondo. Nel caso peggiore, quando tutte le quattro code dei Processor sono piene, il tempo di processamento dei dati del dispositivo intruso sarà di 5 secondi, mentre nel caso medio di 2 secondi.
Il costo computazionale delle funzioni di processamento sarà elevato.
L'impatto del consumo di banda da parte di un modulo Processor sarà di 1 Mb al secondo, poiché 1 Mb sono i dati da analizzare nella funzione di processamento e 1 secondo è il tempo di esecuzione per una singola analisi.
Quindi, in situazioni standard, saranno attivi 4 Processor per switch, in media 1 Processor per 5 dispositivi, che consumeranno 4 Mb/s di banda.

L'impatto delle risorse computazionali e di banda degli elementi gestionali dell'IDS dovrebbero essere non rilevanti.
Ricapitolando, in una situazione standard, l'IDS necessiterà di 12 Mb/s di banda per switch, ovvero del 12% delle risorse di rete e necessiterà delle risorse computazionali per l'esecuzione delle funzioni di 8 Processor e di 40 Agent, che possiamo stimare come 8 GB di RAM e almeno 12 CPU dedicate, che dovranno essere distribuite in modo proporzionale nella rete.
Infine, per quanto riguarda le risorse di memoria per l'immagazzinamento dei dati, si ha che i moduli Storage riceveranno 0.2 Mb/s di dati da salvare da ogni Agent, moltiplicando questo numero per il numero di elementi della rete si ottiene 8 Mb/s, ovvero 1 MB/s.
Si nota che la frequenza di dati da salvare nei moduli Storage sarà: 60 MB/-min, 3600 MB/h ovvero 3.52 GB/h, 84.38 GB/day e circa 590 GB/week.
L'IDS necessiterà di capacità di immagazzinamento dei dati di circa 1 TB e provvederà alla pulizia degli Storage una volta al giorno, eliminando i dati più vecchi di 7 giorni.
In caso in cui un dispositivo sia in modalità di monitoraggio "Hunting", l'Agent aumenterà la frequenza di campionamento di dieci volte, dunque la banda utilizzata dal singolo Agent passerà da 0.4 Mb/s a 4 Mb/s, dunque la banda utilizzata dall'IDS sul singolo switch passerà da 12 Mb/s a 15.6 Mb/s, ovvero per ogni dispositivo in modalità "Hunting" aumenterà di 3.6 Mb/s.
Si ricordi che la probabilità in cui più di un dispositivo entri contemporaneamente in modalità "Hunting" è molto bassa, poiché significherebbe che più dispositivi sono stati compromessi ed eseguono delle operazioni malevole nella rete.
L'impatto computazionale sarà invece minore, sarà necessario creare uno o due moduli Processor per aiutare a svolgere tutti i processamenti richiesti

dall'IDS.
L'ultima domanda a cui dobbiamo rispondere è: in quanto tempo il modello proposto di IDS può rilevare un intruso nella propria rete?
In almeno 5 secondi eseguirà la funzione di monitoraggio sul dispositivo compromesso e al più in 5 secondi verrà eseguita la funzione di processamento. Se questa dà esito positivo sull'analisi basata sulla firma, il sistema IDS avrà identificato l'intruso al più in 10 secondi, se questa darà esito positivo sull'analisi basata sulle anomalie statistiche sarà necessario processare almeno una seconda volta i dati raccolti di quel dispositivo specifico.
Questo processo sarà velocizzato poiché il monitoraggio entrerà in modalità "Hunting" e i Processor inseriranno l'analisi da fare nella coda ad alta priorità. Dunque il secondo processamento sarà eseguito in meno di 2 secondi e in totale l'IDS sarà in grado di rilevare un intruso tramite l'analisi sulle anomalie statistiche al più in 12 secondi.

3.4 Conclusioni

Il trend degli ultimi anni è quello di disaccoppiare la parte hardware da quella software, in modo tale che per un'azienda non sia più necessario comprare dispositivi specifici per avere determinati servizi.
Tramite la virtualizzazione, ci è possibile replicare l'ambiente nel quale sviluppare un servizio attraverso moduli software.
In questa tesi è stato proposto il modello di un IDS basato interamente su moduli software, il quale necessita che nella rete siano presenti determinate risorse computazionali, di immagazzinamento e di banda.
Al contrario, un modello classico basato su dispositivi hardware, necessiterà dell'acquisto dei moduli hardware, delle risorse di banda e la sua corretta

implementazione sarà vincolata alla topologia della rete in cui deve essere installato. Inoltre, come sottolineato in questo progetto di tesi, non risulterebbe elastico a: variazioni dei dispositivi da monitorare, variazioni della topologia della rete, aggiornamenti del parco macchine e aggiornamenti hardware e software dell'IDS stesso.
La realizzazione di software tramite NFV si sposa sia con il tema dell'elasticità progettuale che con il tema dell'ottimizzazione delle risorse nella rete. In questo caso, i moduli Agent, Storage e Processor possono essere allocati in una qualsiasi macchina della rete di cui non vengono utilizzate le risorse. Inoltre, qualora fosse necessario, sarebbe possibile migrare i moduli da una parte all'altra della rete per bilanciare l'utilizzo delle risorse nella rete.
Gli sviluppi futuri per questa tesi sono:

- descrivere un'architettura per un Intrusion Prevention System (IPS), tramite l'aggiunta di un modulo dedicato al risanamento della rete.
- descrivere la metodologia della migrazione di un modulo all'interno della rete. Questo servirà per eventuali aggiornamenti del parco macchine e per ribilanciare il carico computazionale all'interno della rete.

Bibliografia

[1] Cisco.com. Attacco informatico - quali sono le minacce informatiche comuni?, Mar 2023.

[2] IBM.com. Cos'è un attacco informatico?, 2023.

[3] Ann Mary Joy. Performance comparison between linux containers and virtual machines. In *2015 International Conference on Advances in Computer Engineering and Applications*, pages 342–346, 2015.

[4] Peter M. Mell and Timothy Grance. Sp 800-145. the nist definition of cloud computing. Technical report, Gaithersburg, MD, USA, 2011.

[5] A. U. Rehman, Rui L. Aguiar, and João Paulo Barraca. Network functions virtualization: The long road to commercial deployments. *IEEE Access*, 7:60439–60464, 2019.

[6] A.M. Riyad, M.S. Irfan Ahmed, and R.L. Raheemaa Khan. Multi agent based intrusion detection architecture for the ids adaptation over time. In *2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, pages 1–4, 2017.

[7] A.M. Riyad, M.S. Irfan Ahmed, and R.L. Raheemaa Khan. Multi agent based intrusion detection architecture for the ids adaptation over time.

In *2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, pages 1–4, 2017.

[8] Shao Ying Zhu, Sandra Scott-Hayward, Ludovic Jacquin, and Richard Hill. *Guide to Security in SDN and NFV: Challenges, Opportunities, and Applications*. Springer Publishing Company, Incorporated, 1st edition, 2017.