**SISTEMI E SERVIZI DI TELECOMUNICAZIONE**

**INDICE:**

1. **INTRODUZIONE E FONDAMENTI**
   1. TRASMISSIONE E MODULAZIONE
   2. MEZZI TRASMISSIVI
   3. MULTIPLAZIONE
2. **RETI PUBBLICHE DI TELECOMUNICAZIONE**
   1. RETI BB WIRELINE
   2. ACCESSO WIRELESS
   3. ALTRE RETI DI ACCESSO
   4. CONNETIVITÀ 1
   5. MPLS
   6. CONNETIVITÀ 2
3. **VOCE E MULTIMEDIA SU IP**
   1. CODIFICA DELLA VOCE
   2. CODIFICA AUDIO E VIDEO
   3. SIP
4. **ELEMENTI DI ANALISI DEI SISTEMI**
   1. CODE 1.1
   2. CODE 1.2
   3. CODE 2.1
   4. CODE 2.2
   5. DISPONIBILITÀ
   6. ESEMPIO
5. **APPARATI DEI NETWORKING, DATA CENTER ED SDN**
   1. ROUTER
   2. SWITCH
   3. MIDDELBOXES
   4. DATA CENTER
   5. SDN
6. **GESTIONE DELLA RETE**
   1. GESTIONE
   2. SLA
   3. SNMP
   4. NETCONF
   5. SERVICE MANAGEMENT
7. **RETI RADIOMOBILI**
   1. INTRODUZIONE ALLE RETI RADIOMOBILI E CONCETTI BASE
   2. PRINCIPI DI SEGNALAZIONE 2G (GSM)
   3. 2G (GPRS), 3G E 4G
   4. 4G (VIRTUALIZZAZIONE) E 5G

**INTRODUZIONE E FONDAMENTI**

**TRASMISSIONE E MODULAZIONE**

**CANALI DI TRASMISSIONE:**

La trasmissione è un meccanismo che permette di trasferire a distanza informazioni attraverso un mezzo che può essere un collegamento radio, un filo di rame, ecc.. Per poter trasferire l’informazione è necessario per prima cosa trasformare i bit informativi in un segnale fisico trasmissibile; questa operazione si chiama operazione di modulazione. La trasmissione di informazioni avviene attraverso un canale di comunicazione che è una astrazione che rappresenta tutto ciò che sta tra il punto di partenza del segnale che deve essere trasmesso e il punto in cui il segnale viene ricevuto. La caratteristica principale dei canali di comunicazione è che sono introdotti due tipi di distorsione dell’informazione ovvero il rumore e la limitazione di banda che, qualsiasi tipo di canale di comunicazione si consideri, comporteranno delle modifiche alle informazioni trasmesse.

**BITRATE:**

Il primo risultato importante nella storia della trasmissione dei segnali è il teorema di campionamento di Nyquist sul bitrate che dice che se abbiamo un segnale di banda limitata B, questo può essere ricostruito esattamente se si possiede un numero di campioni che vengono presi ad una frequenza R che è maggiore o uguale del doppio della frequenza della banda (R>= 2B). La banda si misura in Hertz. Se si avesse un canale perfetto senza rumore con una banda B allora il canale sarebbe in grado di trasferire i dati fino a 2B simboli al secondo. I simboli sono dei segnali distinti che fanno parte di un alfabeto che possono avere V simboli, quindi il canale trasferisce con un bit rate di:

2B log V bit/s

Dato che V può essere arbitrariamente grande un canale perfetto non avrebbe limitazioni di capacità.

**FORMULA DELLA CAPACITÀ:**

Shannon ha formulato un teorema che stabiliva che il numero di bit al secondo che un canale può trasmettere è limitato dal rapporto segnale/rumore di questo canale, quindi se S è la potenza del segnale ed N è la potenza del rumore, allora vale questa formula: C = B log (1+S/N) dove C è la capacità massima di un canale con rumore. Questo significa che non è più possibile utilizzare un numero arbitrariamente grande di simboli da trasmettere in quanto esiste un limite imposto dal rapporto S/N.

**MODULAZIONE:**

Esistono due tipi di modulazione possibili:

* Modulazione analogica: La tecnica che permette di convertire un segnale in banda base in un segnale che può essere trasmesso in un canale situato su una certa banda di frequenze
* Modulazione numerica: assegnare a dei gruppi di bit delle forme d’onda o simboli differenti in modo tale da ottenere un segnale robusto tenendo conto della presenza del rumore e della distorsione

**CODIFICA:**

La codifica si occupa di trasformare gruppi di bit in altri gruppi di bit e può essere utilizzato in due punti di un sistema di comunicazione:

* Source coding: codifica di sorgente che esegue una compressione delle informazioni riducendo la ridondanza del segnale originale. La compressione può essere senza perdita (Huffman) o con perdita di dati (MPEG)
* Channel coding: proteggere l’informazione aggiungendo dei bit di ridondanza che permettono di correggere o di rivelare degli errori in fase di trasmissione

La modulazione numerica e la canne coding sono strettamente correlate fra loro, infatti si può dire che la modulazione è una forma di codifica.

La catena di trasmissione numerica dei dati è composta dalle seguenti fasi:

* Channel coding
* Digital modulation
* Frequency conversion
* Channel
* Frequency conversion
* De-modulation
* Decoding

**MEZZI TRASMISSIVI**

Vi sono due tipi di mezzi di trasmissione, quelli che utilizzano un mezzo fisico per la trasmissione del segnale detti mezzi di trasmissione guidata e quelli che utilizzano le onde radio per propagare il segnale nell’aria.

**DOPPINI:**

I Doppini sono dei mezzi di trasmissione guidata che è costituito da una coppia di collegamenti in rame intrecciati per evitare possibili interferenze di tipo crosstalk, se fossero paralleli funzionerebbero come delle antenne trasmettendo il loro segnale e ricevendo disturbi elettromagnetici. Esiste una classificazione dei doppini utilizzati nelle LAN basandosi sulla larghezza di banda e la relativa qualità generando delle categorie di doppini. Le più recenti ed ancora in uso sono le categorie che vanno dal 6 alla 8 che trasmettono i dati a velocità elevata rispettando delle condizioni riguardanti la massima distanza e la connettorizzazione (ovvero se l’attacco del cavo al connettore è fatta bene). I Doppini telefonici invece sono molto vecchi e non sono classificabili all’interno delle categorie dei doppini LAN.

**FIBRA OTTICA:**

Un collegamento in fibra ottica è composto da una fibra molto sottile in vetro che trasferiscono dei segnali ottici di qualità molto elevata con un tasso di errore quasi nullo e una capacità di banda molto alta. Il principio che permette la propagazione del segnale attraverso la fibra è quello della rifrazione totale. Vi sono due tipologie di fibre: le single mode e le multimode. La differenza tra le due è che nel caso della multimode è possibile utilizzare diversi modi di propagazione dell’informazione a differenti velocità mentre la single mode consente un’unica modalità di trasmissione.

**MULTIPLAZIONE**

La multiplazione è la tecnica che consente di condividere la capacità dei mezzi trasmissivi tra più utenti, servizi o collegamenti. Le tecniche di multiplazione si suddividono a seconda della risorsa che viene divisa: frequenza, tempo, codice o spazio.

La tecnica di multiplazione più semplice è la multiplazione a divisione di frequenza. Questa permette di prendere tanti segnali limitati in frequenza e posizionali su diverse frequenze di riferimento in un unico mezzo trasmissivo a banda larga.

La multiplazione a divisione di tempo sincrona suddivide l’asse temporale in una sequenza di slot dove periodicamente viene ripetuta la trasmissione dello stesso canale. La multiplazione a divisione di tempo sincrona o di pacchetto prevede che l’informazione sia inserita in dei pacchetti di dimensione variabile che vengono inseriti negli slot disponibili di tempo e a differenza della multiplazione sincrona la posizione di un pacchetto non è ripetuta periodicamente ma viene identificata con una etichetta.

**RETI PUBBLICHE DI TELECOMUNICAZIONE**

**RETI BB WIRELINE (RETI DI ACCESSO CON CAVO)**

**RETI DI ACCESSO A BANDA LARGA IN RAME:**

Il doppino in rame ha la seguente architettura: parte dalla casa e tramite un cablaggio verticale e raggiunge un punto di interconnessione alla base dell’edificio, da qui raggiunge vari ripartitori situati negli armadi stradali che raggiungeranno il MDF (Main Distribution Frame) ovvero il permutatore di centrale.

Le tecnologie xDSL sfruttano al massimo la banda disponibile sui doppini telefonici già installati e alcune di queste tecnologie sono: ADSL2+, VDSL2, G.FAST. Queste tecnologie sono influenzate dalla lunghezza del doppino in quanto minore è la distanza, maggiore è la capacità della banda, infatti per essere efficienti le tecnologie ADSL necessitano di una distanza massima di pochi KM, le VDSL di poche centinaia di metri, le G.FAST sono pensate per avere in rame solamente il tratto che collega la casa al primo punto di interconnessione. L’allocazione delle frequenze varia per le diverse tecnologie e ancora una volta maggiore è la larghezza della banda minore sarà la distanza coperta. La banda è divisa in porzioni che variano in base al tipo di architettura:

* ADSL2+: porzione di UP stream, DOWN stream e la porzione per la linea telefonica
* VDSL: porzione per la linea telefonica, porzione dedicata all’UP stream e infine una larga porzione dedicata a vari possibili schemi di distribuzione sia in UP che in DOWN che utilizzano una Frequency Division Duplex (FDD)
* G.FAST: una porzione di frequenza sovrapposta alla VDSL per le comunicazioni in cui è sconsigliato l’utilizzo del G.Fast e una porzione solamente dedicata al Time Division Duplex (TDD) per trasmettere i dati in intervalli di tempo dedicati all’UP e altri intervalli dedicati al DOWN

**DISTURBI:**

Le reti in doppiono presentano dei problemi di disturbi nella trasmissione dell’informazione che generalmente sono generati dai doppini adiacenti. Se supponiamo vi siano due simboli S1 e S2 che devono essere trasmessi rispettivamente ad R1 e R2 avremo la possibilità che alcune informazioni di S1 indirizzate ad R1 raggiungano in realtà R2 e viceversa.

Una soluzione che permette di ottenere dei miglioramenti significativi delle prestazioni in caso di questi problemi è il Vectoring. Questa tecnologia considera l’insieme dei simboli che vengono trasmessi in un certo istante (S1,S2; ..) come un vettore di simboli s^ e l’insieme dei simboli ricevuti (R1, R2, ..) come un altro vettore di simboli r^. Dati questi due vettori è possibile dire che: r^= T s^ dove T è una matrice che introduce la componente di distorsione k1 e k2 nella comunicazione. Quindi per trasmettere correttamente i segnali si dovrà fare partire: (1- k1 k2)^-1 (s1 –k1 s2) e (1- k1 k2)^-1 (s2 –k2 s1).

Se si vuole trasmettere utilizzando più di un doppino, nel caso in cui un singolo doppino non fornisca sufficiente banda, è possibile utilizzarli entrambi ed ottenere una capacità di banda pari a circa tre doppini. Questo perché dati N doppini è possibile ricavare N-1 canali phantom, utilizzando la differenza tra i modi comuni degli altri doppini. Questa tecnica tuttavia genera dei doppini fantasma che sono estremamente sensibili alla distorsione quindi per evitarla sarebbe necessario che anche questi fossero intrecciati fra loro.

**STRUTTURA DELLE RETI DI ACCESSO A LARGA BANDA:**

La maggior parte delle reti utilizzano un misto di fibra ottica e di rame, in particolare più ci si avvicina alla casa dell’utente più è probabile che ci sia una parte di rete in rame in quanto i costi aumentano notevolmente più ci si allontana dalla centrale.

Una delle soluzioni più usate attualmente per le reti domestiche è la rete FTTCab/VDSL che prevede una parte di rete ottica che va dalla centrale al ripetitore e una parte in rame che porta la connessione nelle case degli utenti, questa rete fornisce la maggiore velocità di trasmissione possibile al costo più basso.

La soluzione FTTB/VDSL porta la fibra fino all’edificio dell’utente (l’unica parte in rame è quella verticale nell’edificio) ed è configurata come un albero ottico passivo. In questo caso alla base dell’edificio vi è un Optical Network Termination che ha la stessa funzione del ripetitore MSAN, ma con costi maggiori in quanto per ogni edificio deve essere presente un ONT a differenza di un singolo MSAN.

Un esempio di rete basata su fibra ottica è la Open Fiber. Questa rete possiede due punti di flessibilità da cui partono delle diramazioni in fibra che splittano la capacità di trasmissione; il punto di flessibilità primario ha uno split 1:4, mentre lo split secondario ha uno split 1:16. A partire dallo split secondario vi sono dei collegamenti diretti per ogni singolo edificio, mentre le linee di fibra che collegano lo split primario con il secondario sono assegnate ai diversi operatori (vodafone,wind).

**ACCESSO WIRELESS**

Le reti di accesso trattate sono reti di accesso wireless da postazione fissa (Fixed Wireless Access) come le reti fornite da Eolo. La propagazione del segnale avviene tramite onde radio e questa metodologia porta con se una serie di problemi:

* Distanza: la distanza di propagazione del segnale è limitata ed è pari a 1/d^2
* Assorbimento: ci sono dei fattori che assorbono il segnale come fattori atmosferici o fisici
* Multipath self-interference: ogni antenna ha un lobo principale che dirige il segnale verso i terminali, se ci sono degli oggetti fisici ad esempio tra l’antenna e il terminale il segnale rimbalzerà su questo ostacolo allungando il suo percorso ma raggiungendo comunque il terminale, questo però causa dei problemi ai simboli successivamente inviati in quanto si sovrappongono con quelli che hanno rimbalzato e sono arrivati più tardi del previsto
* Interferenze: interferenze classiche da altre antenne presenti nella zona o di dispositivi che utilizzano le onde radio come mezzo trasmissivo o che disturbano la frequenza in qualche modo

Queste reti tuttavia portano diversi vantaggi per la trasmissione del segnale che sono:

* Basse frequenze: utilizzano frequenze basse che garantiscono una grande propagazione
* Frequenze licenziate: consentono di gestire più facilmente le interferenze da parte di altri dispositivi
* Tecnologie di antenna sofisticate: utilizzando alcune tecnologie come ad esempio la MIMO si hanno delle prestazioni eccellenti in termini di qualità del segnale ricevuto e insensibilità sulle interferenze ricevute

Riguardo alle basse frequenze è possibile dire che c’è un rapporto inversamente proporzionale tra area di copertura e frequenza utilizzata, di fatto se si utilizza una frequenza di 3,5GHz si avrà una copertura estremamente minore rispetto a quella che si avrebbe utilizzando 700MHz.

Vi sono due tipi di possibili trasmissioni del segnale tramite antenne in base alla loro configurazione. Una antenna può essere strutturata in maniera tale da trasmettere e ricevere le informazioni con la stessa efficienza in ogni direzione, questo tipo di antenna viene definita omnidirezionale. Esistono altre antenne che sono strutturate in modo da trasmettere e ricevere il segnale verso specifici settori con un angolo di azione che va dai 45° ai 120°, queste antenne sono chiamate antenne direzionali.

**ANTENNA MULTIPLE INPUT MULTIPLE OUTPUT (MIMO):**

Le antenne MIMO sono costituite da più elementi radianti il cui segnale può essere sincronizzato in maniera tale che l’insieme degli elementi radianti produce una antenna che può essere configurata per essere più o meno sensibile verso determinate direzioni. Questo implica che possono essere trasmessi più segnali in contemporanea ad un ricevitore che riceve una combinazione dei segnali inviati e tramite un operazione matematica simile a quella del vectoring ricostruisce in maniera affidabile i simboli trasmessi. I vantaggi principali sono che è possibile trasmettere più simboli in parallelo, ma soprattutto, tramettendo in parallelo, si ottiene una grande resistenza al rumore e alle interferenze varie.

I modelli più recenti di MIMO sono chiamati Massive MIMO che possiedono un alto numero di elementi radianti che può svolgere la funzione di diverse antenne in base al servizio che deve essere fornito. Infatti tramite software è possibile modulare e aggiornare nel tempo i lobi di radiazione in base ad esigenze che vanno da un alto traffico mobile ad un alto numero di edifici ecc..

**TECNICHE DI MODULAZIONE E MULTIPLAZIONE DEI COLLEGAMENTI RADIO:**

Le tecniche di modulazione e multiplazione dei collegamenti radio adottano tecniche sofisticate di multiplexing e accesso multiplo e alcune di queste sono:

* CDMA: Code Division Multiple Access è un meccanismo che divide il codice in parti e le assegna ai canali
* OFDM: Ortogonal Frequency Division Multiplexing è una tecnica utilizzata per la multiplazione che utilizza un numero elevato di carriers a piccola banda
* OFDMA: Ortogonal Frequency Division Multiple Access funziona come OFDM solo che è possibile avere un accesso multiplo da diversi canali assegnando diversi carrier nel tempo alle diverse comunicazioni

Il Code division multiplexing è una tecnica che utilizza dei codici ortogonali, ovvero un insieme di parole di codice la cui caratteristica è che il prodotto scalare vale 0 se e solo se ogni codice ortogonale è diverso dagli altri. L’utilizzo di questi codici è molto utile in fase di ricezione in quanto se i segnali trasmessi dovessero unirsi nel tragitto, il ricevitore effettuando una moltiplicazione otterrebbe il dato originale corretto. Ad esempio se una stazione base è in comunicazione con 4 terminali e si assegna ad ogni flusso che collega la stazione ai terminali un codice binario, supponiamo che vengano trasmesse delle sequenze di dati a 3 terminali ma una è errata. Il CDMA prevede che le sequenze trasmesse vengano sommate tra di loro quando ricevute e successivamente moltiplicate per ognuno dei codici assegnati per verificare se è stato ricevuto del segnale trasmesso o meno in maniera corretta.

Per quanto riguarda OFDM l’idea è di trasmettere il segnale suddividendo la banda in un numero molto grande di sotto bande di capacità estremamente ridotta. Ognuna di queste bande viene utilizzata come carrier con cui si tramette un segnale. Il vantaggi di questo schema sono che i carrier sono tra loro indipendenti quindi è possibile utilizzare degli schemi di modulazione numerica diversi per ogni carrier e, a causa della frazione molto piccola di banda utilzzata dai carrier, i dati permangono per molto tempo e questo porta una insensibilità al problema del multipath perché i segnali si sovrappongono complessivamente meno rispetto all’utilizzo di un single carrier.

**FIXED WIRELESS ACCESS ARCHITETTURE:**

Le reti Fixed Wireless Access sono composte da una stazione base e tante subscriber station che ricevono il sengale. A seconda della distanza ci sono 3 modalità di propagazione differenti:

* Line on Site (LOS): propagazione per larghe distanze senza nulla in mezzo
* Near Line on Site (NLOS): propagazione per medie distanze con degli ostacoli in mezzo
* Indoor: propagazione per distanze al di sotto del Km

**ALTRE RETI DI ACCESSO**

Vi sono molte altre reti di accesso oltre a quelle in fibra e rame menzionate prima. Alcune degne di nota sono:

* Reti cellulari (3G/4G/5G): le reti 5G sono ormai equiparabili alle reti di accesso da postazione fissa
* Reti di accesso satellitare: hanno un enorme vantaggio nel dispiegamento della rete che fornisce il servizio anche in aree molto difficili da coprire dalle altre possibili soluzioni
* Reti ibride fibra/coassiale (HFC): fibre utili in paesi che hanno la TV via cavo della quale viene sfruttata la struttura della rete per offrire servizi di accesso ad internet
* Powerline: reti di accesso tramite rete elettrica utili in ambiti domestici

**RETI SATELLITARI:**

Le reti satellitari hanno un’architettura piuttosto semplice da descrivere in quanto c’è un terminale presso l’utente chiamato VSAT che punta verso il satellite e quest’ultimo ritrasmette ad una stazione di terra che dà accesso alla rete fissa. Dato il lungo tragitto che devono compiere i segnali in questa architettura vi sono dei problemi per quanto riguarda i servizi di comunicazione vocale ma non per le reti di accesso ad Internet. I satelliti si dividono in tre fasce:

* Satelliti geostazionari: stanno a 35800km di quota e vengono utilizzati per TV ed internet
* Satelliti di orbita media: stanno tra i 5000km e 15000km di quota e sono utilizzati per i servizi di localizzazione come il GPS
* Satelliti di orbita bassa: stanno a 1000km di quota e vengono utilizzati per le comunicazioni vocali e internet

I satelliti hanno due tipi di copertura la single beam e la multi beam, dove il primo e più datato illumina una area terrestre con un fascio unico, mentre il secondo copre con segnali differenti tante aree.

**CONNETTIVITÀ**

I sevizi di connettività si suddividono in due grandi categorie: la connettività generalizzata e la connettività dedicata. La prima si riferisce a connettività come quella ad internet fornita da un provider, mentre la seconda offre un servizio pensato per una azienda e interconnettere i punti della azienda tra loro.

Lo schema della connettività generalizzata è organizzato con una rete composto da un insieme di entità autonome che utilizzano un piano di instradamento gerarchico. La condivisione delle risorse è gestita dall’ISP.

Lo schema della connettività dedicata è uno schema di indirizzamento privato con segregazione del traffico che garantisce la disponibilità di risorse dedicate. Vi è un punto di accesso alla rete generale tramite un’interfaccia.

**RETI PRIVATE FISICHE E VIRTUALI:**Realizzare un servizio di connettività dedicata significa quindi realizzare una rete privata e può essere implementata in vari modi in base al livello di virtualizzazione. I principali modelli che si possono considerare sono:

1. Il modello più semplice fisico è quello di una rete costituita da N siti che possiedono una propria rete interna e delle interfacce che collegano tramite collegamenti fisici i siti tra di loro (collegamenti punto a punto in fibra ottica)
2. Una versione migliorata del primo modello prevede che i collegamenti tra i siti non siano dei collegamenti diretti fisici ma dei circuiti che forniscono delle frequenze dedicate. I circuiti rendono più facili le operazioni di riorganizzazione del traffico della rete o risolvere eventuali guasti
3. Rete di tipo virtuale di livello 2 realizzate con un nucleo di VLAN su cui si attivano i collegamenti tra i vari siti utilizzando dei router
4. Reti di tipo virtuale di livello 3 che utilizzano un modello network based che utilizza una rete privata virtuale costituita da dei tunnel IP.

Nelle reti private virtuali network-based di solito vi è un provider di servizi di connettività che fornisce questo tipo di reti a diversi clienti. Ogni cliente per ogni sede ha dei router che si affacciano su un nucleo di rete in cui ci sono delle funzioni di routing. È possibile che due clienti distinti si affaccino al nucleo di rete tramite lo stesso nodo, tuttavia a livello logico la rete del provider funziona come se avesse N piani separati e indipendenti dedicati ai singoli clienti.

Per le reti private che utilizzano uno schema di reti di livello 2 sono sostanzialmente delle VLAN. Le VLAN sono nate per avere una maggiore flessibilità e riconfigurabilità nelle reti di campus/edificio; ad esempio se vi sono due edifici con N terminali distinti tramite la VLAN è possibile collegarli per poi visualizzarli come N insiemi da un unico router per effettuare l’instradamento. Dato che su alcuni collegamenti possono esistere più VLAN in contemporanea (collegamenti trunk), per distinguere il traffico delle varie VLAN si utilizza un formato specifico che prevede di inserire 4 byte aggiuntivi all’interno della trama ethernet in modo tale da identificare la VLAN e quando raggiungono gli switch che possiedono delle porte VLAN-aware in grado di identificare quale tipo di VLAN sta passando e la gestiscono come devono.

Nella versione VLAN 802.1ad vi sono due livelli di VLAN, una VLAN relativa a gruppi di utenti e un’altra VLAN gestita dal provider che gestisce le prime VLAN. Questo consente di avere un grande numero di VLAN

**MPLS**

Il Multi Protocol Label Switching (MPLS) è una tecnologia importantissima nelle reti pubbliche di comunicazione. Prima di vedere di cosa tratta MPLS è necessario ricordare le modalità con cui vengono fatti transitare i pacchetti all’interno di una rete di dati. I meccanismi di Forwarding sono principalmente due:

1. Forwarding by destination address: utilizzato nelle reti IP comuni che prevede l’instradamento tramite l’analisi dell’indirizzo di destinazione. Esistono delle tabelle di instradamento contenenti i prefissi degli indirizzi di destinazione organizzati gerarchicamente. Se ci sono più match nella tabella dei prefissi viene utilizzato match con il prefisso più lungo.
2. Forwarding by label swapping: utilizza delle etichette sulla intestazione del pacchetto che ne caratterizzano il percorso. Quando un pacchetto con una etichetta viene instradato tramite la consultazione delle tabelle di forwarding, questo cambierà il valore dell’etichetta basandosi sempre su un valore presente nella tabella di forwarding. Questo tipo di tecnica è conection oriented e il percorso permane finché resta attiva la connessione, poi viene cancellato. Inoltre c’è un exact matching in quanto esiste una sola etichetta di pacchetto associata ad una specifica riga della tabella di forwarding.

MINUTO 16

**CONNETTIVITÀ 2**

**VOCE E MULTIMEDIA SU IP**

**Codifica della voce**

I segnali voce identificano dei segnali di comunicazione della voce umana e tipicamente in un contesto di una sessione bidimensionale, in cui i tempi di ritardo sono di fondamentale importanza affinché la conversazione avvenga nel migliore modo possibile.

La voce umana occupa una banda che sta tra i 300 e i 3400 HZ. Secondo il teorema del campionamento, che dice che la frequenza di campionamento deve essere maggiore di due volte della frequenza massima del segnale che noi vogliamo trasmettere, avremo Fc > 6800 HZ. Gli standard utilizzano frequenze da 8KHz per il campionamento della voce. Ogni campione viene codificato tramite codifica PCM (Pulse Code Modulation) che prevede l’utilizzo di 8 bit per campione con quantizzazione non lineare in quanto i segnali di intensità bassa sono più difficili da riprodurre e quindi necessitano di più livelli rispetto ai segnali di alta intensità che possono essere compresi in meno livelli.

**ALGORITMI DI CODIFICA (CODEC):**

Gli algoritmi di codifica della voce si sono migliorati col tempo fino a definire due tipologie di codec:

* Waveform Codec: cerca di riprodurre nel tempo l’andamento della forma d’onda sonora. Questo codec quindi non si occupa solo della voce, ma si occupa di riprodurre qualsiasi segnale sonoro seguendo l’andamento della banda. Il tipo di Waveform codec più importante è il G726 ADPCM (Adaptive Post Code Modulation) che riduce il bitrate tramite la codifica delle differenze tra campioni successivi.
* Codec Predittivi: si basano sulla conformazione del sistema vocale dell’essere umano per prevedere che tipo di suono sarà prodotto immediatamente dopo aver prodotto un suono. Il funzionamento generale di questi tipi di codec si basa sull’utilizzo di un codebook che è in grado di generare un insieme ampio di segnali tramite sintesi dei segnali, questi segnali vengono confrontati con il segnale originale e tramite il calcolo della differenza tra i due ed una analisi del risultato si ottiene un segnale che presumibilmente sarà simile al segnale successivo in arrivo. Questo tipo di codifica funziona bene per i segnali vocali ma malissimo per gli audio.

**STANDARD IETF PER VOCE SU IP:**

L’informazione audio e video vengono passate con un meccanismo di codifica Audio/Video Coding che produce dei dati. Questi dati vengono inseriti in dei pacchetti di trasporto RTP (Real-Time Transport Protocol), successivamente in pacchetti UDP, poi pacchetti IP. Per quanto riguarda il controllo della sessione, viene utilizzato un protocollo SIP che producono dei dati che vengono immagazzinati in pacchetti UDP o TCP che passeranno al livello IP.

I livelli RTP, UDP e IP generano un overhead di grande dimensione, infatti il minimo valore di byte prodotti dai vari livelli sono: RTP 12 bytes, UDP 8 bytes, IP 20 bytes. Questo comporta un conseguente consumo di banda non indifferente. Per ridurre l’overhead si potrebbero utilizzare pacchetti di dimensione maggiore che ridurrebbero l’overhead percentuale ma introdurrebbero dei ritardi di pacchettizzazione che non sempre sono tollerabili.

Il livello di ritardo in una conversazione vocale è definito dalla somma del ritardo di codifica e pacchettizzazione e del ritardo di rete. Dopo diversi studi è stato determinato che il one-way delay deve essere inferiore ai 150ms. Se il ritardo è inferiore di 25ms non è necessario eseguire alcuna operazione per migliorare la comunicazione, ma nel caso in cui il ritardo sia compreso tra 25ms e 150s è necessario utilizzare i cancellatori d’eco.

**Codifica audio e video**

**SIP**

Il SIP (Session Initiation Protocol) è un protocollo IETF per la gestione delle sessioni di voce e multimediali. Questo protocollo è integrato con SDP (Session Description Protocol) che consente la descrizione dei media coinvolti nella sessione. Il protocollo venne inizialmente utilizzato per le applicazioni PC-to-PC e PC-to-Phone e, solo in un secondo momento, venne adottato anche per la telefonia fissa e mobile. Il SIP dispone di molte estensioni che forniscono servizi supplementari, di messaggistica e di notifiche di eventi.

SIP è un protocollo di tipo testuale dove gli agenti sono identificati da degli URL e il trasporto dei dati avviene tramite protocollo UDP (User Datagram Protocol) non affidabile e non orientato alla connessione o TCP (Transmission Control Protocol) affidabile e orientato alla connessione; le primitive utilizzate da questo protocollo si chiamano come per http “metodi” e alcune di queste sono:

* INVITE: metodo che si occupa di invitare un agente a partecipare ad una sessione
* ACK: metodo per verificare l’attivazione corretta di una sessione
* BYE: metodo per concludere una sessione

**SESSIONI SIP:**

In un ambito ideale dove tutti gli agenti sono noti a tutti la sessione SIP viene stabilita in modalità peer-to-peer. Per stabilire un contato un agente invia un messaggio di INVITE all’agente con il quale si vuole collegare, questo accetta la connessione e restituisce un messaggio di conferma all’agente che lo ha invitato, una volta ricevuto questo messaggio l’agente invitante invia un messaggio di ACK di conferma di avvenuta connessione. Una volta stabilita la connessione può avvenire la comunicazione tramite RTP (Real-Time Transport Protocol) e quando uno degli agenti desidera interrompere la connessione lo comunicherà tramite un messaggio di BYE.

In un ambito di interlavoro fra più reti dove non vi è una conoscenza completa degli interlocutori è necessario introdurre un SIP server. Il SIP server si occupa quindi di impostare una sessione tra due o più interlocutori appartenenti a reti differenti che in assenza dello stesso non potrebbero comunicare, una volta stabilita una connessione tra gli interlocutori la comunicazione avviene tramite RTP che però non passa dal SIP server e la conclusione della connessione avviene direttamente tra gli interlocutori come se fossero in modalità peer-to-peer.

**VOIP:**

Con l’invio di ulteriori messaggi come ad esempio il messaggio RINGING e TRYING è possibile anche stabilire connessioni adatte anche alla rete telefonica. Questo tipo di connessione PC-Telefono viene identificata con VoIP e consente di collegare terminali IP in rete con altri terminali in rete e anche con terminali che invece non lo sono.

Uno dei problemi principali riscontrati nelle comunicazioni VoIP riguarda il NAT (Network Address Translation). Questa componente consente ai dispositivi appartenenti ad una rete locale e dunque con IP privati di condividere un unico indirizzo IP pubblico. Dato che gli indirizzi IP privati non possono essere utilizzati per comunicare al di fuori della loro rete locale è necessario che questi vengano tradotti affinché la connessione con l’esterno sia eseguibile. Altro problema delle connessioni VoIP riguarda il fatto che è necessario tradurre il proprio payload in uscita.

I meccanismi principali utilizzati per attraversare il NAT sono:

* STUN (Session Trasversal Utilities for NAT): questo meccanismo esterno fa da tramite tra le componenti che cercano di comunicare e, osservando esternamente la comunicazione riesce ad individuare i veri indirizzi degli agenti consentendo la comunicazione
* TURN (Trasversal Using Relay NAT): un nodo che si occupa di tutto anche della traduzione dei media in aggiunta alla traduzione degli indirizzi che effettua anche lo STUN.

Per le reti VoIP pubbliche esiste una componente chiamata SBC (Session Border Controller) che si occupa della gestione della comunicazione tra dispositivi interni alla rete privata e dispositivi esterni. Alcune delle sue mansioni sono:

* Controllo delle risorse: verifica che le risorse siano coerenti con le risorse della rete interna
* Accounting: si occupa di intercettare i messaggi di INVITE, ACK e BYE
* Topology hiding
* Stateful firewall e NAT

I sistemi VoIP sono anche in grado di fornire i servizi tipici dei centralini tra cui:

* Identificazione del chiamante (CLIP)
* Inoltro e trasferimento delle chiamate
* Attesa di chiamata con messaggi e/o musica

**CODE 1**

**SISTEMI DI SERVIZIO E CODE M/M/1:**

Un sistema di servizio è un’astrazione utile per studiare le prestazioni di sistemi reali come ad esempio la coda di attesa alla porta di uscita di un server. I valori più importanti di cui si tiene considerazione in un sistema di servizio sono:

* λ = frequenza di arrivo
* B = dimensione del buffer
* N = numero di server
* L = processi in coda
* W = tempo speso in coda
* Y = processi in servizio
* H = tempo di servizio
* = tempo medio di servizio
* T = W+H (tempo totale speso nel sistema)
* X = L+Y (processi totali nel sistema)

Un fenomeno tipico che si può verificare nei sistemi reali è quello della perdita di una o più richieste in arrivo ad un sistema con il buffer pieno. Quindi se B= 0 avremo un sistema a pura perdita, mentre se B = ∞ avremo un sistema a pura attesa.

**FORMULA DI LITTLE:**

La formula di Little è valida solo sotto condizioni generiche riguardanti i sistemi di servizio a pura attesa che sono:

* Avere un sistema stabile (non in transitorio)
* Avere un sistema work-conserving, ovvero che il lavoro effettuato dal server non venga perso

La formula enuncia che il numero medio di clienti in un sistema è uguale al tasso medio di arrivo moltiplicato per il tempo medio nel sistema: **X= λT**

La formula vale anche per i sottoinsiemi L e Y: **L= λW** , **Y= λH = =p**

Dove **p** è il carico del sistema, ossia la probabilità che il sistema sia occupato (almeno un job presente).

Affinché p sia una probabilità è dunque necessario che p<=1 e che quindi <= µ

Nel caso in cui p>1 il sistema è considerato instabile in quanto la coda cresce indefinitamente e non si possono più applicare le formule precedenti.

Anche nel caso in cui p=1 il sistema è considerato instabile salvo casi particolari.

**DISTRIBUZIONI DEGLI ARRIVI E DEI SERVIZI:**

Supponiamo di osservare il flusso di auto su un’autostrada e la coda ad un casello. Se il flusso è di 10 auto al minuto e la coda è stabile, e lunga in media 100 auto, cosa possiamo affermare sulla base della formula di Little?

Il tempo medio di soggiorno nel sistema è: T= 100/10 = 10 min. Quindi µ>10 auto/min.

Esistono diversi sistemi in grado di soddisfare queste condizioni ad esempio:

* µ=12 auto/min quindi p= 10/12 = 0.83. In questo caso il sistema è spesso occupato ma non sovraccarico
* µ= 10 auto/min quindi p=0.25. In questo caso il sistema è raramente occupato e la variabilità della coda è elevata

**MODELLI MARKOVIANI:**

In un modello Markoviano la distribuzione dei tempi di interarrivo e di servizio è esponenziale. In sostanza se A(t) è il numero di arrivi in un periodo t , allora A(t) = λ t e quindi il processo di arrivo è senza memoria :

**P{tj > t}=**

Il comportamento dei sistemi senza memoria può essere contro intuitivo in quanto un sistema a partire da un momento in cui viene osservato, non dipende da quanto successo prima.

**CODA M/M/1/K:**

Una coda che dispone di un servente, interarrivo e servizio esponenziali e buffer con K posizioni si chiama coda M/M/1/K. Il sistema in questo caso si risolve con una catena di Markov in cui ogni stato rappresenta il numero di elementi nel sistema e gli archi che collegano gli stati fra di loro rappresentano le transazioni e sono etichettati con la loro frequenza. Le equazioni risolutive sono quindi:

P0= e gli stati successivi Pi= \*

Nel caso in cui K avesse come valore ∞ le formule si semplificano e si ottiene:

* **X = p/1-p**
* **T= 1/ µ-**
* **L= / 1-p**
* **Y= p**
* **W = 1/ µ \* p/1-p**
* **H = 1/µ**

Possiamo dire dunque che essendo la distribuzione dei tempi esponenziale all’aumentare del carico **p** il ritardo aumenta **esponenzialmente** rispetto al suo aumento.

GUARDARE ESEMPIO FINALE DA VIDEOLEZIONE

**CODE 2**

**SISTEMA A PERDITA E PRINCIPIO PASTA:**

Se un sistema a coda ha un buffer finito e il processo degli arrivi permette che giungano al sistema più job di quanti ne può contenere, si ha il fenomeno della perdita. Il calcolo della probabilità di perdita è generalmente complesso, ma nel caso in cui il processo di arrivo è di Poisson, quindi con interarrivi esponenziali, vale il principio PASTA (Poisson Arrivals See Time Averages) anche conosciuto come Random Observer Property. Gli arrivi quindi “campionano” lo stato del sistema e la distribuzione di probabilità dello stato che trovano corrisponde alla media temporale (time average) in cui si trovano gli stati del sistema. Di conseguenza in una coda M/M/1/K la probabilità di perdita è:

**= \***

**ESEMPIO SU PASTA DI SISTEMI SENZA MEMORIA:**

Consideriamo un sistema a coda M/M/1 con λ=20 richieste al secondo e tempo di servizio medio 40ms. Supponendo che ad un certo istante gli arrivi vengano dirottati su un buffer secondario B, dopo quanto tempo, in media, la coda sarà vuota? In quel momento, quante richieste in media saranno parcheggiate in B?

Essendoci un dirottamento degli arrivi ad un istante causale abbiamo un random observer quindi la coda sarà vuota nel tempo:

**T = 1/µ- λ**

In quel momento nel buffer (che non viene scaricato) saranno mediamente presenti:

**X= p/1-p = λT job**

**CODA M/D/1:**

Una coda con un servente, tempo di interarrivo esponenzizale e tempo di servizio deterministico (tutti i job ci mettono lo sesso tempo ad essere serviti) si chiama coda M/D/1. L’analisi del sistema per buffer infinito ha i seguenti risultati:

* **X= (1-p/2)\*(p/1-p)**
* **Y= p**
* **L=**
* **T=**
* **W=**
* **H= 1/**

PER ESEMPIO GUARDA LEZIONE

**CODE CON PIU’ SERVENTI:**

In una coda M/M/N vi sono N serventi con una capacità che estraggono job da un’unica coda e li elaborano in parallelo; il job in testa alla coda viene assegnato al primo servente libero e in questo modo la capacità complessiva del sistema è pari a Nµ. Il carico complessivo normalizzato si indica con A= λ/µ e la condizione di non saturazione è A<N. I sistemi di questo tipo possono essere risolti con una catena di Markov. Il comportamento di una catena di questo tipo approssima quello di una coda con un singolo servente di capacità complessiva pari a Nµ per valori di A/N elevati (vicini a 1) mentre per i valori piccoli la capacità utilizzata è di µ.

**CODA M/M/N/0:**

Una coda M/M/N/0 ha N serventi e buffer nullo ed è un esempio tipico di sistema a pura perdita. Questo tipo di coda ha una grande importanza nella modellazione di situazioni in cui un numero molto grande di sorgenti casuali con un tasso di richieste molto basso competono per un numero finito di risorse come ad esempio nelle sessioni VoIP su trunk. Questo sistema si analizza con una catena di Markov e il parametro di prestazione significativo è la probabilita di perdita Pn (PASTA). Il carico sul sistema si indica con A e si misura in Erlang. La formula per il calcolo di Pn si chiama formula B di Erlang e si indica con B(N;A). Non esiste una espressione chiusa per tale formula, ma si può ricordare che il valore di B per un rapporto A/N costante decresce al crescere di N.

**ESERCIZI CODE**

SE SERVE GUARDALO

**AFFIDABILITA’ E DISPONIBILITA’**

L’affidabilità R(t) è una funzione che fornisce la probabilità che un sistema funzioni per un certo tempo senza guasti; F(t)= 1 – R(t) è la guastabilità del sistema (ovvero la probabilità che il sistema si guasta prima di t). Sia λ(t) il tasso di guasto di un sistema, allora questa funzione è una funzione del tempo che assume la forma di una vasca da bagno (ovvero alta mortalità precoce che crolla in un periodo di vita normale fino a risalire nella fase di invecchiamento). Quindi per noi λ(t)= λ (ovvero è costante dato che consideriamo solo il periodo di vita normale) e il processo dei guasti è senza memoria:

* **R(t)=**
* **F(t) =1 -**

**DISPONIBILITA’:**

Dato che i guasti si riparano, nella maggioranza dei casi ciò che importa non è l’affidabilità di un sistema, ma bensì la sua disponibilità, ovvero la probabilità che sia funzionante nel momento in cui viene utilizzato. La disponibilità D di un sistema dipende sia dalla sua guastabilità sia dal tempo di riparazione, quindi:

D=

Dal punto di vista della misurazione D si valuta come 1 – la sommatoria delle durate dei guasti fatte un periodo di osservazione T.

Supponiamo di conoscere il MTBF e MTTR di un certo componente o sottosistema: possiamo dunque calcolare il valore per quel componente. Se conoscessimo tutti i valori D di tutti i componenti di un sistema allora potremmo calcolare il valore complessivo di D; questo calcolo ha distinte formule in base alle configurazioni dei sistemi:

* Se le componenti sono in serie (una dietro l’altra) il sistema funziona solo se tutti i suoi componenti funzionano. **D= d1\*d2\*…**
* Se le componenti sono parallele il sistema funziona se almeno una componente funziona. **p=1-d, D=1-p1\*p2\*…**
* Se le componenti sono in un sistema più complesso, questo può essere sempre ricondotto ad una combinazione tra componenti in serie e/o in parallelo

**SISTEMI IN RISERVA CALDA:**

Alcuni sistemi per configurazione dispongono di diverse parti guastabili all’interno di un’unica componente e se anche solo una di queste si guasta è necessario riparare l’intera componente. In questo caso si ha che: R(t)= 2 - , da cui è possibile ricavare il valore di MTBF= 3/2

**SISTEMI NON IN SERIE O PARALLELO:**

Se il sistema preso in considerazione non è riconducibile in alcun modo ad una combinazione di sistemi in serie o parallelo è necessario adottare un approccio differente, ovvero quello dei fault trees che si basa sul teorema della probabilità totale: P(A) = P(A|B) + P(A|¬B).

Per risolvere questi sistemi si costruisce un albero dove alla radice il sistema è completamente funzionante e successivamente sui rami vengono identificate delle situazioni in cui il sistema può essere guasto o meno, fino all’ultimo ramo che identifica una situazione di guasto. La disponibilità totale sarà dunque calcolata come 1 – la somma delle probabilità di guasto dei vari percorsi ottenuti dalla generazione dell’albero. Questo sistema è efficace anche per risolvere i sistemi in serie e parallelo.

**TRE MICROTEOREMI SULLA DISPONIBILITA’:**

1. Nessun sistema costituito da componenti con d<1 può avere D=1
2. La disponibilità di un sistema in parallelo è sempre maggiore della massima disponibilità delle sue componenti
3. La disponibilità di un sistema a serie è sempre più piccola della minima disponibilità delle sue componenti

**ESERCIZI AFFIDABILITA**

SE SERVE GUARDALO

**ESEMPIO**

SE SERVE GUARDALO

[**Apparati di Networking, Data Center ed SDN**](about:blank#section-6)

[**Router, switch e middleboxes**](about:blank)

**ROUTER:**

Il router è il nodo fondamentale della rete che si occupa della trasmissione dei dati e del loro indirizzamento. Gestisce il protocollo di routing (OSPF, BGP, ecc..) ottenendo così il percorso migliore di trasmissione dei dati. Questi percorsi vengono scritti all’interno di una tabella di forwarding. Il meccanismo di funzionamento del router ideale è piuttosto semplice: il pacchetto arriva con un indirizzo ip di destinazione, se questo indirizzo sta in un certo range all’interno della tabella di forwarding allora il router saprà da quale porta inviare il pacchetto.

In realtà il router è molto più complicato di quanto descritto, in quanto ad esempio non esiste un unico protocollo di instradamento ma ce ne sono vari che collaborano per inserire gli indirizzi all’interno della tabella di forwarding. Inoltre non vi è una sola tabella di instradamento ma vi è anche la ILM che serve per gestire le etichette MBLS. Questa è solo una delle componenti che sono presenti nei router più complessi ve ne sono altre come lo schedule dell’uscita del router.

Il processo di forwarding è più complicato di quanto sembri e si articola nelle seguenti fasi:

1. Matching: si occupa di identificare quali righe della tabella hanno un match sul prefisso significativo dell’indirizzo IP
2. Longest prefix: una volta eseguito il match questa funzione va a vedere quale tra i prefissi che hanno ottenuto il match sono i più lunghi (quindi meno generici). Questa operazione non necessariamente da come risultato una sola riga.
3. Preference: spesso ci sono delle preferenze sulla selezione della riga come ad esempio una out static.
4. Metric: processo che utilizza dei valori di metrica per selezionare le righe da mantenere (ad esempio la metrica che valuta il peso minore di costo tra le righe)
5. Equal cost multipath: nel caso di OSPF è possibile che vi siano più righe nella tabella con le stesse identiche caratteristiche e questa funzione si occupa di distribuire equamente su ognuna di queste il traffico di dati

**GESTIONE DELLE CODE DI USCITA:**

Nei router le code di uscita spesso sono gestite tramite FIFO (First In First Out), quindi il primo che arriva viene gestito e non si implementa nulla che dà delle priorità al traffico. Se invece si desidera assegnare un qualche livello di priorità si può utilizzare la cosiddetta priorità assoluta, ovvero assegnare un livello di priorità ad una classe di traffico per poter servire tale classe finche ci sono dei pacchetti appratenti ad essa. La priorità assoluta può portare ad un problema di saturazione della rete da parte dei processi ad alta priorità quindi per far fronte a questo problema sono utilizzati degli algoritmi di partizione del link che implementano il fair queuing. Questo algoritmo approssima un meccanismo di scheduling chiamato Generalized Processor Sharing che assegna ad ogni classe di traffico una quota R in modo tale da distribuire la capacità del link equamente in quote suddivise tra le classi attive che stanno generando del traffico.

**SWITCH:**

Gli switch ideali hanno la semplice funzione di accettare in ingresso un pacchetto che possiede un indirizzo MAC, lo switch osserva se tale indirizzo MAC è presente in una tabella di forwarding che contiene esplicitamente tale indirizzo o meno, in caso positivo il pacchetto va alla porta indicata dalla tabella. Nel caso in cui l’indirizzo non è presente nella tabella viene eseguita una operazione di flooding, ovvero mandare a tutte le porte il pacchetto. Le funzionalità offerte dallo switch sono complementate da due algoritmi: l’algoritmo di learning che si basa sull’osservazione del traffico e utilizzando gli indirizzi MAC sorgente identifica l’origine del traffico e la sua destinazione e l’algoritmo RSTP (Rapid Spanning Tree Protocol) che evita la creazione di loop che causano la presenza di una stessa destinazione su più porte.

Il protocollo RSTP tuttavia è poco ottimizzato nei percorsi dunque è possibile utilizzare altri protocolli più efficaci come ad esempio il TRILL. Lo switching moderno proposto dal TRILL (Transparent Interconnection of Lots of Links) non utilizza lo spanning tree, ma bensì utilizzando un protocollo di incapsulamento e di instradamento IS-IS (ogniuno degli switch conosce come è formata la topologia) che utilizza come indirizzo di destinazione gli switch finchè non si raggiunge il livello precedente all’uscita dal sistema interno.

Gli switch moderni utilizzano una combinazione del protocollo di learning e SPB/TRILL in base al tipo di pacchetto che ricevono in ingresso. La struttura degli switch è molto simile a quella dei router; infatti abbiamo un primo livello di instradamento che utilizza delle tabelle, un secondo livello fortemente hardwareizzato utilizzata per il trasporto dei pacchetti ed infine un meccanismo di accodamento intelligente.

**MIDDLEBOXES E SERVER:**

Altri apparati necessari per realizzare una struttura di rete sono i server e i middleboxes. Questi ultimi fanno riferimento a degli apparati che implementano delle funzioni che non possono essere integrate all’interno dei router. Alcune funzioni necessarie alla rete che possono essere implementate da queste componenti sono:

* Firewall/ IDS (Intrusion Detection System)/ anti DDoS
* Virtualizzazione di reti
* Load balancer per le prestazioni

Alcuni tipi di server necessari per implementare una rete svolgono le seguenti funzioni:

* Autenticazione
* DHCP
* DNS
* CDN

La prima funzione che andiamo ad analizzare è quella del firewall. Il firewall è un dispositivo che serve per realizzare alcune funzioni legate alla sicurezza che regolano il traffico di una rete. Le regole imposte dal firewall vengono eseguite in sequenza e si occupano di imporre regole per accettare o non accettare delle determinate connessioni. I firewall moderni hanno un comportamento di tipo stateful, ovvero sono dei firewall che tengono conto dei fatti avvenuti in passato all’interno della rete.

L’Intrusion Detection System (IDS) tipicamente viene utilizzato non-in-line per implementare delle regole per il traffico in maniera similare al firewall. La differenza sostanziale è che una copia del traffico proveniente dall’esterno viene indirizzata al IDS che lo esamina senza però intralciare la comunicazione. Le regole imposte in questo tipo di sistema sono Signature-based oppure anomaly-based. La prima suppone che si conosce la caratteristica tipica di un particolare tipo di attacco, mentre il secondo identifica nuovi attacchi mai visti basandosi su comunicazioni nuove non identificate. Normalmente si usano per la maggiore i sistemi Signature-based in quanto gli anomaly-based incappano in dei falsi positivi per la maggior parte delle volte.

I sistemi anti-DDos tipicamente non risiedono all’interno della rete aziendale. Gli attacchi DDos sono attacchi che hanno come obiettivo quello di esaurire le risorse all’interno dei server della rete (es. esaurire il numero massimo di connessioni TCP/IP) oppure di esaurire la banda a disposizione. Esistono tipicamente due architetture di difesa dagli attacchi DDos spesso adottate in simultanea e sono:

1. Fidarsi del proprio ISP (Internet Service Provider) che dispone di alcuni filtri di traffico in grado di bloccare eventuali attacchi di banda prima che questa venga raggiunta
2. Utilizzare un provider che dispone di una grande intelligenza all’interno del suo cloud in grado di deviare il traffico prima del suo ingresso nella rete a rischio all’interno del suo stesso cloud, per analizzarlo e determinare la presenza di rischi, pulire il traffico e infine inviare alla rete il traffico pulito

**ARCHITETTURA DI UN FRONT-END PROTETTO**

Una architettura molto semplice e sicura consiste nell’avere un firewall all’ingresso della nostra rete che implementa anche le funzioni di NAT, un edge router che si occupa di ricevere tutti dati dall’esterno e infine un IPS/IDS che sta tra le due componenti appena menzionate per l’analisi dei flussi. Una soluzione più sofisticata prevede l’utilizzo di due livelli di firewall e IPS/IDS.

La sicurezza tuttavia non è l’unica caratteristica di cui tener conto nella realizzazione di un front-end, una tra le altre caratteristiche fondamentali è la capacità di distribuire il traffico in maniera efficiente (Load Balancing). Il Load Balancing è un problema molto diffuso nei siti con ampi bacini di utenza e si può affrontare in base a dei livelli in cui si agisce:

* Soluzioni a livello di rete: dove si possono utilizzare il caching, la CDN (Contend Delivery Network) oppure il DNS Load Balancing
* Soluzioni a livello applicativo: dove si utilizzano dei reverse proxy per distribuire il traffico in maniera intelligente sui vari server
* Soluzioni a livello di rete del data center: dove si utilizzano dei Load Balancer che ricevendo degli IP virtuali in ingresso distribuiscono il carico agli IP fisici tramite algoritmi di balancing, mantenendo la persistenza delle sessioni tramite l’utilizzo di cookie

**Data Center**

I data center sono degli spazi contenenti un insieme di server e una grande quantità di storage che vengono distribuiti. Questi data center hanno due tipi di interfacce verso l’esterno: uno verso i client che richiedono un servizio (ovvero front-end verso internet) e un secondo front-end che va verso altri data center dello stesso tipo (interfaccia di backoffice). Oltre alle componenti fisiche appena menzionate che necessitano di una gestione fisica delle infrastrutture, ovviamente il data center dispone di funzionalità di tipo informatico riguardanti la sua gestione come la gestione logica gestita dal Fault Configuration Accounting Performace Security (FACPS). Il data center normalmente fa parte di una infrastruttura che contiene per lo meno un data center principale e un data center di disaster recovery, oppure nelle soluzioni più moderne vi sono le infrastrutture cloud dove vi sono più data center geograficamente distanti collegati fra loro.

**EVOLUZIONE DEI SERVER E STORAGE:**

Una volta il data center era una grande stanza con tanti server dedicati ad un applicativo con un sistema di storage dedicato. Il vantaggio principale di questo tipo di soluzione è la sua semplicità di realizzazione, ma fondamentalmente non esiste alcuna condivisione a livello di storage e CPU tra i vari armadi server.

In un secondo momento si fecero dei passi avanti verso la condivisione dei dati tramite il consolidamento. Il consolidamento dello storage consiste nel creare una SAN (Storage Area Network) alla quale tutti i server sono collegati in modo tale che le risorse possano essere condivise tra questi.

L’infrastruttura generale di un data center consolidato si articola in 3 parti:

1. Provider access: accesso verso le reti esterne
2. Router, switch, load balancer: insieme di armadi che si occupano delle funzionalità della sicurezza e distribuzione delle risorse
3. Struttura fisica: assetto RAC che collega fra di loro una grande quantità di armadi server

Tutta l’infrastruttura deve essere alimentata in maniera estremamente precisa in quanto costi, distribuzione della potenza e surriscaldamento sono dei fattori fondamentali da tenere sempre in considerazione. La soluzione più tradizionale utilizza la corrente alternata proveniente dalla rete elettrica, vi sono dei motori diesel nel caso in cui salti la corrente, successivamente vi è un’altra batteria che si carica con corrente continua nel caso in cui nulla funzioni, infine la corrente viene trasformata in alternata fino a raggiungere i server e poi diventa nuovamente continua per essere distribuita a questi ultimi. Per gestire il calore le file di server sono posizionate fra un corridoio freddo da cui prendono aria e uno caldo dove la espellono.

**HIGH AVAILABILITY E DISASTER RECOVERY:**

High availability e disaster recovery si occupano entrambe di fornire delle soluzioni a dei problemi che possono presentarsi all’interno della rete. In termini di rapporto frequenza dell’evento/costo dell’evento è evidente che la high availability si occupa di problemi frequenti ma di basso impatto economico, mentre la disaster recovery ha frequenza molto bassa (può anche non succedere mai) ma conseguenze economiche disastrose.

**PROTEZIONE DEI DATI:**

La protezione dei dati è un elemento chiave per poter fare ripartire il sistema dopo un guasto/disastro. La base di dati per essere protetta deve essere periodicamente duplicata e vi sono due strategie tipicamente utilizzate:

* DB Shadowing asincrono: si tiene conto di tutte le transazioni che il DB server principale effettua sullo storage principale e periodicamente viene inviato il file di log delle transazioni al Backup DB server che esegue le medesime transazioni sullo storage secondario
* DB Shadowing sincrono: ogni qualvolta viene effettuata una operazione sullo storage principale, viene inviato un log al Backup DB server il quale effettua un aggiornamento dello storage secondario

Per migliorare le prestazioni vi sono delle soluzioni diverse rispetto alle precedenti che si basano sul mirroring in cui concettualmente sarebbe possibile non perdere alcuna transazione in caso di guasto/disastro. Le strategie di mirroring utilizzate sono:

* Logical volume mirroring: si basa sul software del sistema operativo del DB Server principale che implementa delle copie di transazioni sincrone, ovvero quando viene effettuata una operazione di scrittura sullo storage principale in contemporanea la stessa operazione viene eseguita sullo storage secondario e il commit della transazione viene inviato solo quando l’operazione di scrittura è avvenuta con successo in entrambi gli storage. Queste soluzioni sono completamente sincrone, ma possono creare dei problemi di ritardo.
* Physical volume mirroring: in questo caso è lo storage principale che si occupa di eseguire una copia della transazione sullo storage secondario e, dopo aver ricevuto l’ack da parte del server secondario, invia il commit all DB server principale. Anche qui abbiamo un problema di ritardo per soluzioni geograficamente distanti tra loro

**PROTEZIONE DELL’ACCESSO ALLA RETE:**

È necessario implementare un accesso protetto alla rete in grado di tollerare possibili guasti dei link di collegamento e degli edge router.

Una prima soluzione è quella di avere una duplicazione dei router che funzionano come se fossero un singolo router logico; il più importante protocollo utilizzato per implementare tale soluzione è il VRRP (Virtual Router Redundancy Protocol). Il funzionamento di questa soluzione consiste nell’utilizzare i VIP (Virtual IP) per identificare un unico indirizzo IP per entrambe le copie di router che di base possiedono due IP diversi; in questo modo è possibile implementare una architettura di master tra i router duplicati e solo il master risponderà alle richieste in entrata al VIP e, nel caso in master dovesse guastarsi, tramite delle elezioni subentrerebbe un nuovo master in grado di rispondere alle richieste in uscita. Il processo di elezione non consente al sistema di funzionare e in quel momento potrebbe esserci una perdita di informazioni. Per quanto riguarda le operazioni in uscita si opta per delle soluzioni di multipath OSPF oppure per avere più efficienza delle soluzioni a livello più basso ovvero a livello MPLS o addirittura fisico.

**DISASTER RECOVERY:**

La disaster recovery sostanzialmente si occupa di costruire due copie del data center allineate con un collegamento. Le due copie generalmente sono distanti tra loro. Implementare soluzioni di disaster recovery è molto costoso e il costo aumenta in base alle distanze e al tipo di replica che si vuole avere (sincrona o asincrona).

La disaster recovery prevede tre requisiti fondamentali per la sua implementazione e che ne definiscono la sua qualità:

* RPO (Recovery Point Objective): identifica la quantità di dati che possono essere persi in seguito all’evento. Generalmente vengono persi gli ultimi secondi/minuti antecedenti al disastro.
* RTO (Recovery Time Objective): identifica il tempo necessario affinchè sia possibile fare ripartire il sistema con i database allineati in seguito al disastro. Questo tempo è generalmente il massimo fra RPO e NRO
* NRO (Network Recovery Objective): tempo massimo di attesa affinchè la rete esterna sia in grado di vedere la rete di backup operativa

Le soluzioni disponibili per la disaster recovery sono raggruppate in delle tier che vanno da tier1 a tier7; in base alla tier di appartenenza cambiano i valori di RTO e il livello di RPO. Nella tier1 abbiamo dei backup su nastro, una soluzione semplice con costi bassi ma che comporta un RTO di giorni in quanto è necessario effettuare un restore del database. Nelle tier3 e 4 iniziano ad essere presenti le repliche tipicamente asincrone che possono essere parziali o complete; in questo caso i tempi di recupero sono nell’ordine delle diverse ore di lavoro e i costi sono una via di mezzo. Nelle tier5, 6 e 7 abbiamo degli allineamenti di tipo sincrono transazionale o realizzati da infrastruttura; queste soluzioni estremamente efficaci sono molto costose che comportano un RPO=0 e a livello di tempo si va dai pochi minuti ad un massimo di 1h.

**VIRTUALIZZAZIONE:**

Il processo di virtualizzazione consiste nell’avere diversi utenti che fanno uso di una quantità di memoria e CPU virtuali allocate al loro dispositivo che condividono le stesse componenti hardware. Nei data center un esempio di virtualizzazzione consiste nella migrazione delle Virtual Machines che possono avvenire nel caso di balancing, ovvero quando si distribuisce il carico tra due VM una con tanto carico di lavoro e una con poco, oppure possono avvenire nel caso di ottimizzazione, ovvero quando abbiamo 3 VM accese con un basso carico di lavoro e per risparmiare corrente ne spengo 2 e passo tutto il carico di lavoro sulla restante VM accesa.

Gli insiemi dei data center di una azienda sono sottoposti a due tipi di traffico:

* Traffico nord-sud: traffico che prevede che dei terminali effettuino delle richieste sui server della nostra infrastruttura
* Traffico est-ovest: traffico di backoffice, ovvero il traffico generato dalle operazioni di allineamento, replica e migrazione dei dati tra i siti dove ci sono i data center

**RETI RADIOMOBILI**

**INTRODUZIONE ALLE RETI RADIOMOBILI E CONCETTI BASE**

Multiplexing: multiple stazioni che trasmettono e un ricevitore usati per gli accessi da parte di piu clienti ad una antenna. I tipi di multiplexing sono FDMA, TDMA, CDMA.

Duplexing: separazione dei canali di downstream e upstream utilizzando lo stesso mezzo di comunicazione. I tipi di duplexing sono TDD e FDD.

Una rete mobile è una rete che consente a dei termnali di accedere ai sevizi di telefonia e rete. La rete mobile è composta da RAN (Radio Access Network) e CN (core network). Il primo gestisce la connettivita radio con i terminali il secondo connette le RAN alle infrastrutture telefoniche e di rete. I CN possono essere a commutazione di pacchetto per internet e a commutazione di circuito per la telefonia.

L’area di coprtura cellulare è offerta da delle base station che coprono un area generalmente di forma esagonale detta cella. Le operazioni legate alla mobilità che possono essere eseguite in un cella sono: Cell Selection, location update e handover. Le prime due si usano quando ci si muove e non si fanno chiamate e handover si usa mentre si è in chiamata e ci si muove.

Cell selection: le antenne mandano un messaggio in broadcast, i cellulari che ricevono questo messaggio in base alla potenza scelgono a quale antenna accoppiarsi.

Location Update: se non si è in chiamata un terminale è identificato all’interno di una location area che è composta da più celle. Se si sposta di LA deve essere notificato al DB che ne salva la posizione.

Handover: se si passa da una cella allaltra in chiamata si attiva l’handover. Questa procedura consente di cambiare BS di riferimento senza far cadere la chiamata prima che il segnale della prima BS sia troppo scarso e interferisca con lacomunicazione.

Paging: procedura utilizzata quando si vuole identificare la cella in cui risiede un terminale per iniziare una comunicazione. Si sa la LA di un terminale manon la cella, quindi tutte le BS di una LA mandano un messaggio in broadcast, quando il terminale riceve il messagio risponde e dunque fa capire dove sta.

Il processo che serve per decidere luogo configurazione e impostazioni di accesso ai noti si chiama Radio Planning. Il radio planning è composto da due fasi:

* Coverage planning: seleziona dove installare le BS e decide come configurarle . tiene in considerazione il traffico stimato che dipende da popolazione urbanizzazione ecc. capisce come posizionare le antenne in base alle esigenze di traffico o morfologiche del terreno
* Frequency planning: decide come assegnare le risorse rafdio (frequenze) alle diverse BS. Tiene conto del possibile riuso delle frequenze assegnando a delle celle delle frequenze. Le celle sono disposte in modo tale da non avere adiacenti delle altre celle con la stessa frequenza.

**PRINCIPI DI SEGNALAZIONE 2G (GSM)**

**2G (GPRS), 3G E 4G**