Internet and Data Centers

Appunti delle Lezioni di Internet and Data Centers

Anno Accademico: 2025/26

Giacomo Sturm

Dipartimento di Ingegneria Civile, Informatica e delle Tecnologie Aeronautiche
Università degli Studi "Roma Tre"

Sorgente del file LaTeX ed ultima versione del testo disponibile al link:
https://github.com/00Darxk/Internet-and-Data-Centers/

Indice

1 Introduzione: ISPs ed IXPs

Gli *Internet Service Provider* ISP o *Autonomous System* sono gestori autonomi della rete che hanno una responsabilità su un certo territorio. Su queste reti indipendenti viaggiano i pacchetti, gli ISP sono infatti responsabili di inoltrare i pacchetti che passano al loro interno per raggiungere la destinazione. In totale sono presenti circa 170000 ISP al mondo, alcuni di questi sono pubblicizzate e pubbliche per offrire servizi ad utenti comuni, altri offrono servizi a grandi compagni o altri ISP. Questi ISP formano una gerarchia, non dichiarata, che si può inferire, non essendo questi ISP governati da una struttura o organizzazione sovrastante. Alcuni di questi ISP sono privati e non si mostrano, per cui è possibile solamente inferire la gerarchia considerando questi ISP privati. Ogni ISP si cura solamente dei suoi interessi specifici, attraverso i loro clienti, permettono di connettersi ad un prezzo economico con altri servizi o utenti nel loro territorio. Questi ISP pagano altri ISP più grandi per raggiungere obiettivi sempre più remoti, in questo modo si può determinare una gerarchia che parte dagli ISP più piccoli ai più grandi.

Queste regioni geografiche che identificano i territori di un certo ISP non sono separate, ma sono fortemente sovrapposte, è solamente una divisione logica e parzialmente può essere interpretata come una suddivisione geografica.

Questi ISP comunicano tra di loro attraverso una connessione privata di rete *Private Network Interconnects* PNI, queste possono essere dei cavi fisici di proprietà degli ISP oppure possono essere affittati, che collegano due router tra questi due ISP. Il costo per questa PNI viene pagato in base ad accordi privati tra i vari ISP ed in base al traffico. Questo è un rapporto commerciale a varie livelli, ci può essere un cliente che paga per utilizzare i servizi di un altro ISP, oppure bilaterale o peer-to-peer, dove i due ISP si scambiano pacchetti destinati all'altro ISP, ed pagano entrambi insieme il costo del PNI.

Generalmente si paga sul picco di traffico giornaliero.

Questo processo tuttavia non è più efficiente, con il concetto di economia di scala, invece di stendere fili singoli tra tutti gli ISP, si creano di punti di scambio comuni dove tutti i provider portano una loro macchina ed un filo che la collega alla loro rete, a questo punto per comunicare con altri ISP si crea una connessione logica con uno degli ISP presenti nel punto di scambio. Questo si chiama *Internet eXchange Points* IXP. Questi sono governati da associazioni che mettono a servizio dei propri consorziati questi punti di scambio in modo sicuro. Quando si offre un servizio si cominciano ad offrirne di ulteriori, quindi oltre alla comunicazione offrono servizi di housing, di computing, etc. formando i tradizionali *Data Centers*. In questi punti di scambio ci sono interconnessioni di diverso tipo, commerciali, peer-to-peer, gratuiti, ma sempre offerti nello stesso punto di scambio. Al mondo sono presenti circa un migliaio di questi IXP, in Europa il più grande è ad Amsterdam: AMS-IX.

La rete di reti non è più una rete di macchine connesse tra di loro, ma coagula considerando questi punti di scambio comuni dove sono presenti delle macchine. Un *Data Center* è uno spazio dedicato contenente computer, risorse di massa e sistemi di telecomunicazione. Questi si trovano generalmente in scantinati, ed i più grandi in zone dove l'energia costa poco o è facilmente procurable, oppure in zone dove la temperatura è pressoché bassa, per diminuire il costo di raffreddamento. In questi data center si brucia una quantità molto elevata di energia. Per l'importanza di questi data center, sono costantemente sorvegliati e protetti. Ogni macchina ha una procedura di recupero

e sono pubbliche e quindi si assume che chi è in grado di avvicinarsi ad un macchina può effettuare queste procedure di recovery ed entrare in possesso della macchina. Quindi la sicurezza deve essere anche fisica per impedire alle persone di avvicinarsi a questi macchinari.

C'è una differenza tra l'*hosting* e l'*housing*, nel primo l'hardware è posseduto dal data center, mentre nel secondo l'hardware non è fornito dal data center me viene procurato dal cliente. Altri centri invece hanno un solo cliente che è il proprietario, questi generalmente vengono realizzati solo dai più grandi come Google, Microsoft, etc.

Questa tendenza è nata negli ultimi decenni, quando oltre ad offrire connettività gli ISP cominciarono ad offrire hosting ed housing, creando propri data center. La connettività è ormai diventata un bene di consumo a basso costo, quasi dato per scontato. Questi data center possono essere interni ad una singola rete o condivisi tra varie reti.

Dentro un IXP si può facilmente inserire un data center, avendo già la struttura e l'organizzazione per ospitarli, unendo i router e macchine dei vari ISP a risorse di calcolo offerte dal data center.

La definizione di *cloud* fornita dal NIST, *National Institute of Standard and Technology*, è caratterizzata dalla possibilità di scalare su richiesta, un accesso a banda capiente, la possibilità di attivare risorse su richiesta, risposta rapida e misurazione accurata.

I servizi offerti da piattaforme cloud sono IaaS, *Infrastructure as a Service*, macchine virtuali, risorse di massa, firewall, tutti realizzati virtualmente; PaaS *Platform as a Service*, la gestione di macchine virtuali database, risorse; Saas *Software as a Service*, il software che viene eseguito su queste macchine virtuali può essere comprato o affittato direttamente da queste piattaforme cloud. Altri servizi specializzati possono offrire calcolo massivi. Alcuni servizi di cloud sono reti pubbliche, altri possono essere privati o ibridi. I leader del settore sono AWS, *Amazon Web Services*, Microsoft Azure e Google Cloud. Questi cloud possono popolare diversi data center, AWS si trova su 25 regioni, in questo caso macro-regioni geografiche. Microsoft si trova su più di 60 regioni mentre Google principalmente in America del Nord.

Quando si compra un servizio su una piattaforma cloud, questo è sparo non necessariamente su uno stesso data center, poiché per trasferire una macchina virtuale è estremamente facile. Le risorse acquistate su piattaforme cloud sono distribuite ed in base alla necessità del gestore possono essere trasferite facilmente. Questo vale sia per macchine virtuali che per la memoria di massa.

In Europa è stato creato un sistema federato di ISP europei, GAIA-X, per realizzare hub nazionali indipendenti dall'hardware per garantire servizi specificando un modello con garanzia di controllo sulla locazione delle macchine e la strada che percorrono i dati.

In Italia l'Agenzia per la Cybersicurezza Nazionale ha fondato un cloud marketplace per offrire alla pubblica amministrazione di acquistare risorse cloud da parte di provider qualificati. La CONSIP ha un servizio di supporto per i servizi cloud dedicati alla pubblica amministrazione. Inoltre c'è un programma per portare ed abilitare la pubblica amministrazione all'uso dei servizi cloud. Questo è previsto dal PNRR, nel primo obiettivo, attraverso una migrazione ad un cloud nazionale PSN, oppure ad un provider certificato. Anche la stessa migrazione ad un altro cloud viene trattato come un servizio senza dover necessariamente conoscere il suo funzionamento tecnico.

1.1 PSN: Polo Strategico Nazionale

Questo progetto ha l'obiettivo di creare un grande data center nazionale che fornisca vari servizi cloud.

I principali servizi offerti dal PSN sono housing ed hosting, cloud IaaS Private e Shared, può offrire cloud pubblica gestita dalla PSN garantendone la sicurezza, utilizzando sistemi offerti da produttori certificati, si può realizzare un sistema ibrido con elementi interni alla rete.

I public cloud esistenti che parteciperebbero a questo progetto sono Azure, Google Cloud ed Oracle.

Sulla rete privata del PSN, sul suo hardware, vengono offerti servizi IaaS dedicati e condivisi, CaaS e *Disaster Recovery* (DR) e PaaS, fornendo elementi applicativi e middleware come servizio.

Servizi di Cloud ibridi vengono gestiti nel territorio nazionale dal PSN, utilizzando un'infrastruttura ibrida utilizzando cloud pubblici di partner commerciali e privati sulla rete del PSN, installati sull'infrastruttura locale. Questo come il Public Cloud PSN Managed ed il Secure Public Cloud garantiscono la presenza dei dati sul territorio nazionale. Per il resto dei servizi cloud, i dati sono localizzati presso il *Cloud Service Provider* (CSP) e non garantisce la *data sovereignty*

L'infrastruttura del PSN è un data center a doppia regione, interconnessa via *Virtual Data Center Network* (VDCN), simulando una stessa LAN, duplicando i dati tra le due regioni.

Servizi PaaS messi a disposizione da parte del PSN su una piattaforma per erogare elementi applicativi e middleware, astraendo l'infrastruttura sottostante. Questi servizi sono *Database as a Service* (DaaS), verifica dell'identità e gestione degli accessi, big data e servizi AI.

Analogamente fornisce applicazioni basate su container con servizi CaaS.

L'accesso a questi servizi avviene solamente tramite la rete TIM, uno dei consorzianti che ha vinto il contratto, questa implementa dei sistemi di sicurezza per garantire che eventuali attacchi non riescano a penetrare la rete interna del PSN, utilizzando dei tunnel per inviare il flusso di dati all'infrastruttura del PSN.

I quattro data centers sono collegati su due regioni, tra queste regioni il traffico viene gestito tramite il protocollo *Multi Protocol Label Switching* MPLS, sul backbone IP di TIM.

Tra DC della stessa regione le VLAN sono trasportate con VXLAN, *Virtual eXtensive LAN*, queste utilizzano espedienti tecnologici per duplicare infrastrutture fisiche utilizzando dei tag per dividere i pacchetti destinati alle varie copie. Il protocollo VXLAN permette di connettere reti diverse condividendo pacchetti di livello due tra le reti, trattandola come una singola LAN, trasparente dal punto di vista delle macchine nelle varie LAN.

1.2 SPC: Sistema Pubblico di Connettività

Questo è un bando multi-fornitore, il vincitore del bando ha preso la parte più grande del mercato della pubblica amministrazione, mentre i restanti fornitori hanno ottenuto il rimanente.

Questa rappresenta una rete di grandi dimensioni dedicata a connettere le sedi di ogni singola Pubblica Amministrazione tra di loro ed all'internet. Offre diversi servizi di trasporto wired e wireless, su rete elettrica, ottica o da parte di dati satellitari.

Per questi servizi viene definito un *Service Level Agreement* (SLA). Ad ogni servizio il fornitore assegna una misura di qualità, il jitter è la distanza tra i vari pacchetti, se è pari a zero questi arrivano tutti allineati. Se la misura della qualità dovesse scendere al di sotto di terminate soglie,

sono previsti rimborsi, chiamati "penali" all'amministrazione. Gli SLA contemplano anche guasti o anomalie. Vari ISP realizzano questo SPC, a ciascuno è assegnato un gruppo di PA, l'accesso ad internet è gestito dai singoli fornitori, in base all'esito dell'asta multi-fornitore.

Gli ISp realizzano le reti delle PA usando MPLS, per comunicare tra di loro su una rete QXN, *Qualified Exchange Network*, attualmente collocato presso gli IXP di Roma (Namex) e Milano (MIX).

2 Algoritmi di Instradamento

Si considerano algoritmi di instradamento solo per infrastrutture fisiche, connessi da reti fisse.

Esistono due tipi principali di algoritmi di instradamento, *distance vector* e *state packet*. In TCP si utilizza una variante del distance vector. Queste sono due filosofie opposte.

Si vogliono delle certe qualità da questi algoritmi, si vuole avere algoritmi efficienti, per evitare che il calcolo dei cammini abbia un peso eccessivo rispetto all'instradamento dei pacchetti. La tabella di instradamento dentro ai router non viene realizzata con un tabella, ma con strutture ad albero specializzate per avere un accesso il più veloce possibile. SI vuole mantenere la maggior quantità possibile ci computazione sull'hardware, inoltre la dimensione dei pacchetti è piccola, e dipende dalla dimensione dei pacchetti ethernet di 1500 byte, definita dal primo consorzio NIX. Inoltre le risorse computazionali dei router non sono necessariamente sufficienti per poter gestire la complessità della rete. Questi pacchetti sono piccoli, poiché essendo in competizione dovevano realizzare schede di rete più economiche dei loro competitori. Utilizzando meno memoria si hanno schede di rete più economiche, e questo rappresenta un errore da parte del consorzio NIX, anche se si è affermato come lo standard per le comunicazioni via filo, è necessario avere operazioni di inoltro estremamente veloci poiché i singoli pacchetti sono estremamente piccoli.

Un router è diviso nel *data plane* e nel *control plane*, gli algoritmi di routing sono presenti nel control plane, e la tabella di instradamento composta da questo control plane viene inoltrata al data plane. Parlando con le altre macchine i protocolli di routing devono determinare la tabella di instradamento. Si vuole che questi protocolli siano efficienti, poiché su un router sono presenti molti altri servizi aggiuntivi, quindi il tempo del processore è limitato.

Inoltre si vuole individuare un cammino ottimo, un cammino che impieghi il minor tempo possibile per raggiungere la sua destinazione. Per determinare l'ottimalità del cammino si utilizzano criteri come il numero di hop o il costo delle linee, talvolta assunto inversamente proporzionale alla velocità.

Questo cammino se minimizza il numero di hop, minimizza il numero di immissioni da parte dei router che attraversa. Un'altra risorse è l'utilizzo delle linee, avendo una banda limitata, non può mandare i pacchetti su una stessa linea, avendo anche un buffer limitato per le singole linee. Il carico corrente della rete è tuttavia difficile da calcolare. Se si considerasse solamente il carico della rete, si creerebbe un fenomeno di retroazione causando un oscillazione della rete incontrollabile. Altrimenti si potrebbero scegliere linee con un packet loss minimo.

In genere si scelgono algoritmi che si basano sul numero di hop. Questi algoritmi devono essere robusti e dinamici, poiché alcune linee riscontrare malfunzionamenti, errori di configurazione da parte di amministrazioni di rete, etc. Nello stesso tempo, si vuole essere stabile, non si vuole cambiare il routing durante la trasmissione. Tutti i pacchetti devono essere inviati e ricevuti nello stesso ordine, anche se esiste il livello TPC per riordinarli, nessuna macchina deliberatamente invia pacchetti fuori sequenza. Per questo un'oscillazione del routing non è accettabile, poiché grava sul livello TPC, aumentando il tempo necessario per sistemare questi pacchetti.

Si è realizzato un protocollo inter-dominio che oscilla in continuazione, dato che il suo effetto non era completamente compreso. Per cui è possibile realizzare esperimenti dove la rete diventa instabile.

Il traffico può essere classificato, quando entra in una rete di un ISP, per determinare se si tratta di traffico utente generico o mission critical o privilegiato o VoIP. Inoltre questi algoritmi devono essere equi, nessun nodo deve essere privilegiato o danneggiato.

L'ultimo criterio è l'economicità, si vuole ridurre ic osti di configurazione e manutenzione dei protocolli di routing.

Questi criteri sono talvolta contrastanti, e bisogna scegliere l'algoritmo migliore per un dato caso d'uso.

Si possono classificare questi algoritmi in statici e dinamici. GLi algoritmi dinamici applicano un instradamento in funzione della topologia e del carico della rete, mentre algoritmi statici hanno applicazione ristretta poiché prendono decisioni indipendenti dallo stato della topologia della rete. Questi algoritmi statici vengono utilizzati in casi semplici. La configurazione manuale delle macchine è una configurazione statica, questa è sempre presente anche in piccole parti in ogni rete. Se in una rete sono presenti topologie ad albero, queste non hanno bisogno di una configurazione dinamica. Invece per topologie magliate è opportuno utilizzare un algoritmo dinamico, poiché al taglio di un link o allo spegnimento di un nodo, bisogna ridistribuire il carico e riorganizzare la rete in modo veloce, senza compromettere l'operabilità delle altre macchine.

Uno di questi algoritmi statici è il *flooding* che consiste nell'inoltra ogni pacchetto a tutte le interfacce connesse al router.

Gli algoritmi dinamici possono essere successivamente divisi in routing isolato, dove ogni router decide senza comunicare con altri router; il routing centralizzato, dove un router centrale determina la scelta migliore, questa strategia è rimasta per molto tempo sottovalutata; il routing distribuito, questa è la strategia corrente della rete, dove ogni router informa i propri vicini le informazioni note. Nel distance vector i router informano i propri vicini rispetto alla condizione globale, mentre nel link state packet i router inviano alcuni pacchetti verso tutta la rete che raccontano della topologia locale.

Algoritmi di routing isolato sono *hot potato* e *backward learning*, l'algoritmo utilizzato dai switch o bridge, che determinano in base alla provenienza dei pacchetti le destinazioni possibili rispetto alle varie interfacce.

Per comunicare sulla rete è richiesto un indirizzo di rete associato ad una scheda di rete, ed una netmask associata. Una netmask indica da un indirizzo qual è la parte di rete e quale di host. Per mandare il pacchetto apparentemente non serve, ma è necessaria per determinare la rete dentro cui la macchina è presente e può raggiungere direttamente.

Si controlla prima se la macchina destinazione è direttamente raggiungibile, inviando un pacchetto MAC di livello due, poiché non è necessario un pacchetto di livello 3.

Ogni macchina almeno in questo senso ha meccanismi di routing. Inoltre per tutti i destinatari che non sono locali è presente un default gateway, il primo indirizzo disponibile nella rete. Inoltre è necessario conoscere l'IP del DNS per risolvere gli indirizzi. Una macchina avendo queste quattro informazioni può navigare in rete.

Anche un router ha interfacce e netmask differenti per ogni rete su cui è connesso. Queste vengono inserite nella tabella di instradamento e rappresentano le reti direttamente connesse, queste non possono essere rimosse dalla tabella di instradamento, per configurazione.

In ogni router è presente una parte di routing statico, che rappresenta questa configurazione di interfacce.

Se il routing è completamente statico questa tabella contiene per ogni nodo da raggiungere le linee da usare, compilata dall'amministratore di rete, chiamato ad intervenire in presenza di guasti. Una variante quasi-statica consiste l'amministratore fornisce più alternative in ordine di priorità.

Se il destinatario è direttamente connesso, bisogna determinare l'indirizzo MAC della scheda di destinazione con una richiesta ARP, oppure verso il next hop. Questi pacchetti di livello due vengono creati localmente nelle varie reti locali che attraversa, mentre rimane invariato il pacchetto di livello tre, eccetto per il campo ttl ed il checksum che viene ricalcolato.

Una versione del flooding chiamato selective flooding viene utilizzato come sotto-protocollo di altri protocolli per inoltrare solo su un insieme di linee selezionato, scartando pacchetti troppo vecchi, scartando un pacchetto al suo secondo passaggio per un nodo.

Nel routing isolato ogni *intermediate system* calcola in modo indipendente le proprie tabelle. L'hot potato invia il pacchetto alla linea con coda più breve, di interesse solo teorico.

Nel *backward learning* dai pacchetti in ingresso vengono determinate le macchine raggiungibili su quella linea, IEEE 802.1D a livello due, questi protocolli non aprono pacchetti di livello superiore. Per effettuare il backward learning è importante che non siano presenti maglie, altrimenti un pacchetto potrebbe entrare da più interfacce, accoppiandolo ad un algoritmo per il calcolo dello spanning tree. Può essere raffinato aggiungendo un campo che specifica il costo di un cammino, incrementandolo ad ogni hop. Si possono mantenere alternative ordinate. Quando la destinazione è ignota si effettua flooding. Per la sua necessità di avere una struttura ad albero non è presente in internet, dato che non è possibile tagliare arbitrariamente connessioni.

Il routing centralizzato è un meccanismo di routing gestito da un'entità centrale, supponendo l'esistenza di un *Routing Control Center* (RCC) che conosce la topologia della rete, questo spesso non è realistico. È rimasta per molto tempo teorica, ma recentemente sono apparse esigenze di trasferire un grade volume di traffico tra due macchine. Quindi un'autorità centrale tramite dei protocolli, *Software Defined Networking* (SDN), determina la topologia della rete in quel momento e stabilisce la strada migliore per poter trasferire quella grande quantità di dati. Si può scegliere di effettuare routing guidati da protocolli noti o imporre flussi di traffico.

Nel routing distribuito non esiste un RCC, ma tutte le funzionalità sono da tutti gli is, seguendo lo stesso paradigma, comunicando con i propri vicini. Il primo distance vector invia ai propri vicini una parte della tabella di routing, una proiezione rimuovendo colonne relative a macchine locali. Mentre il link state packet invia informazioni sui suoi vicini verso il resto della rete.

2.1 Algoritmi Distance Vector

Ogni is invia una tabella detta distance vector, ad ogni is adiacente. Questa tabella contiene la parte essenziale della sua tabella di instradamento, da cui vengono omessi dettagli locali. Ogni is ricevendo queste tabelle dei propri vicini ricalcolano la tabella di instradamento integrando queste informazioni.

Le destinazioni vengono apprese con un meccanismo di passa-parola. Inizialmente la tabella di instradamento non può essere vuota, altrimenti rimarrebbe vuota ad ogni iterazione dell'algoritmo. La prima informazione da inserire sono gli indirizzi direttamente connessi, a distanza di un singolo hop, nel data plane. Queste informazioni vengono prese nel control plane ed inviate ai suoi vicini. Questi effettuano lo stesso inviando i propri indirizzi direttamente connessi.

Ogni destinazione nella tabella è corredata dal costo del cammino, dalla destinazione all'is stesso. Questo vengono calcolati dal distance vector ottenuto dai vicini, sommando i costo della linea di ingresso da cui è stato ricevuto.

Molte aziende e produttori cercando di vendere di più aggiungono caratteristiche e modificano questi protocolli. Il protocollo di distance vector originario appartiene a Cisco.

Per passare l'informazione bisogna inviare i vari distance vector, questi possono essere inviati secondo vari criteri. Possono essere inviati periodicamente oppure ad ogni modifica della tabella di instradamento.

I nodi adiacenti da aggiornare vengono appresi in base a protocolli di servizio appositi, oppure sono ignoti, inviando i pacchetti di distance vector in multicast.

Il tempo necessario affinché tutte le macchine conoscano tutte le altre è abbastanza lungo, fino a decine di secondi, un tempo estremamente lungo per ingegneria delle reti. Non si può intasare la rete inviando distance vector molto grandi, con più di 1000 nodi. Si preferiscono protocolli come SPF per reti più dinamiche.

Per ogni linea il distance vector è composto da due colonne, una prima che contiene la rete e la seconda il costo per raggiungerla. I costi delle linee vengono sommati, e vengono inseriti nella tabella di instradamento dell'is le connessioni alle reti di costo minore, per ogni LAN determinata.

Questi is non devono essere sincronizzati altrimenti periodicamente tutte le macchine smetterebbero di funzionare per mandare e ricevere i vari distance vector. Per questo il tempo da attenere per inviare i distance vector vengono sfasati di una certa quantità. Sono presenti inoltre meccanismi per garantire che questo processo non sia sincrono.

Il problema cruciale che si crea con i distance vector è il *count to infinity*, riescono a reagire rapidamente a buone notizie, e lentamente alle cattive notizie.

Una buona notizia può essere l'aggiunta di una nuova rete, questa propaga il suo distance vector e l'informazione viene propagata sull'intera in un tempo relativamente breve.

Se questa connessione viene tagliata, il vicino non può aggiornare la sua tabella di instradamento. Queste macchina avrebbero raggiunto la rete ora disconnessa passando per macchine altre macchine adiacenti. Poiché non possono eliminare questa entry nella tabella di instradamento, ad ogni nuovo distance vector la distanza alla rete disconnessa incrementa continuazione, tendendo all'infinito. Si crea una specie di eco.

Una buona notizia arriva a distanza k in k passi. Le cattive notizie invece si propagano in un tempo che è funzione del valore convenzionalmente attribuito ad infinito. Questo valore si attribuisce alla lunghezza del cammino più lungo più uno.