



SAPIENZA
UNIVERSITÀ DI ROMA

Generazione e visualizzazione grafica di traffico di reti

Facoltà di Ingegneria dell'Informazione, Informatica e Statistica
Laurea Triennale in Informatica

Francesco Pannozzo

Matricola 699427

Relatore

Prof. Daniele De Sensi

Anno Accademico 2023/2024

Tesi non ancora discussa

Generazione e visualizzazione grafica di traffico di reti

Relazione di tirocinio. Sapienza Università di Roma

© 2024 Francesco Pannozzo. Tutti i diritti riservati

Questa tesi è stata composta con L^AT_EX e la classe Sapthesis.

Email dell'autore: francesco.pannozzo@libero.it

*Dedicato alla
mia famiglia*

Sommario

Questa relazione descrive il lavoro di tirocinio interno svolto presso l'università La Sapienza, concretizzato nella realizzazione di un progetto volto a realizzare un software per poter visualizzare in forma grafica l'andamento del traffico di una rete. Il progetto ha come obiettivo di mostrare il traffico di rete al variare del tempo e ciò viene raggiunto tramite grafiche e animazioni generate programmaticamente. L'idea dell'ambito di tirocinio nasce dalla volontà di sperimentare una realizzazione front-end tramite la libreria Manim, un motore di animazioni per video matematici esplicativi. Nel capitolo dedicato all'introduzione descriveremo l'ambito in cui si è tenuto il tirocinio, le motivazioni alla base in cui è stato idealizzato l'argomento, seguito da un'analisi dello stato dell'arte attuale, descrivendo diverse realtà affini al progetto e studiandone caratteristiche, i loro pregi e mancanze. A seguire ci sarà una descrizione dei contributi che vuole dare il progetto, scaturiti dall'analisi dello stato dell'arte e infine una breve panoramica sulla base di partenza del progetto, descrivendo da cosa si è partiti, come è nata l'argomentazione che ha portato alla scelta del progetto e la modalità di svolgimento. Nel capitolo 2 parleremo del contesto di sviluppo, ovvero le tecnologie impiegate. Seguirà quindi una descrizione approfondita della scelta di certe tecnologie ma soprattutto del perchè le si è scelte. Il capitolo 3 descriverà il generatore di traffico e configurazione di rete, descrivendone funzionalità, condizioni iniziali di setup di questa parte di applicazione e infine descrivendo come il tutto viene generato, passando dalla distribuzione spaziale dei nodi e creazione dei links all'ottenimento dei files sui quali sono memorizzati i dati ottenuti. Il capitolo 4 descriverà l'analizzatore..

Indice

1	Introduzione	1
1.1	Ambito del tirocinio	1
1.2	Motivazioni	2
1.3	Stato dell'arte	2
1.4	Contributi	3
1.5	Base di partenza del progetto	4
2	Contesto di sviluppo: le tecnologie impiegate	5
2.1	Python	5
2.2	La libreria Manim	5
2.3	I formati Json e Yaml	6
3	Generatore di traffico e configurazione di rete	9
3.1	Funzionalità e setup	9
3.2	Esecuzione del generatore	17
3.3	Creazione dei links e nodi	18
3.4	Distribuzione dei nodi: aspect ratio	20
3.5	Generazione del traffico	21
3.6	Files prodotti	21
3.7	Aspetti realizzativi scartati: yaml flow e block style	25
4	Analizzatore di traffico	27
5	Visualizzatore grafico	29
6	Prestazioni e complessità computazionale	31
7	Conclusioni	33
8	Test	35

Capitolo 1

Introduzione

Nel mondo le reti informatiche sono oramai un concetto ben istanziato nella collettività, la loro presenza è soverchiante e si dirama nei più disparati settori. Basti pensare già alle reti PAN (Personal Area Network) le quali connettono dispositivi personali entro pochi metri e che ognuno di noi usa abitualmente nella propria casa, alle reti LAN (Local Area Network), anch'esse presenti nelle nostre case così come in uffici o edifici scolastici, le reti dei datacenter fino a giungere alla rete globale internet, la quale è creatrice a sua volta di paradigmi come può essere l'internet of things. Le reti informatiche sono impiegate nei più vari settori come l'istruzione, in cui le reti sono cruciali nelle scuole e nelle università per avere accesso a risorse educative o sfruttare l'e-learning, i servizi pubblici governativi e sanitari, nel settore ludico e multimediale come il gioco online e l'attuale streaming di contenuti multimediali: insomma, le reti informatiche sono di fatto una presenza piena e diffusissima ed è estremamente difficile riuscire a immaginare il mondo come lo vediamo oggi senza questa tecnologia. Con l'aumentare delle funzionalità legate alle reti, così come i dispositivi collegati a esse, capire cosa succede al loro interno, come si muovono i dati, è quindi di cruciale importanza, tramite l'analisi dei dati che vi fruiscono è possibile fare diagnostica, per quanto riguarda un discorso di monitoraggio, ma anche è possibile applicare le analisi in un ambito didattico e accademico. Capire cosa sta succedendo in una rete in modo immediato e visivo è lo scopo di questo progetto, il quale punta a mostrare, in modo grafico, l'andamento del traffico di una rete.

1.1 Ambito del tirocinio

Il progetto fa parte del percorso di tirocinio interno intrapreso presso l'Università La Sapienza di Roma. L'argomento su cui verte il progetto è la realizzazione di un visualizzatore grafico dell'andamento del traffico di una rete, basato su animazioni programmatiche. Il tool permette di visualizzare gli switch rappresentanti i vari endpoints e i link che li collegano i quali vengono colorati tramite animazioni nel tempo in base al traffico di rete precedentemente analizzato. Nel tool è presente anche una parte generativa di traffico di rete, una creazione di traffico fittizia di vitale importanza ai fini di testing. Il progetto è suddiviso quindi in tre parti; una parte che si occupa della generazione di una configurazione network e relativo traffico

di rete, una parte che analizza il traffico di rete calcolandone le medie percentuali di intervalli di tempo e aggiornate ripetutamente e infine una parte che visualizza il traffico analizzato dall'analizzatore producendo una rappresentazione video.

1.2 Motivazioni

L'idea di sviluppare un visualizzatore grafico di traffico di rete è nata, in sede di proposta, dal Professore Daniele De Sensi, relatore del tirocinio, e dalla mia volontà di sviluppare un'applicazione avente il front-end come focus dell'esperienza. Nel mio personale corso di studi presso il Dipartimento di Informatica non ho avuto modo di studiare e approfondire un discorso legato al front-end, per cui la volontà di intraprendere questo percorso nasce in primis da un forte interesse verso questo aspetto dell'informatica e in secondo luogo per un completamento di formazione professionale personale.

1.3 Stato dell'arte

L'esigenza di analisi di reti informatiche ha portato alla luce svariati tool che permettono appunto di analizzare cosa avviene in una rete, di studiarne i dati statistici e di visualizzare graficamente determinati scenari. L'università americana Johns Hopkins[8] ha stilato una lista di software per la visualizzazione e analisi di reti[9]:

- **Gephi[4]:** Gephi è il software leader di visualizzazione ed esplorazione per tutti i tipi di grafici e reti ed è open source. Le sue caratteristiche includono l'analisi esplorativa dei dati mediante manipolazioni di reti in tempo reale, analisi dei collegamenti per rivelare le strutture sottostanti delle associazioni tra oggetti, analisi dei social network per la creazione di connettori di dati sociali per mappare le organizzazioni della comunità e le reti di piccoli mondi, analisi della rete biologica per rappresentazione di modelli di dati biologici ed esportazione e creazione poster per promuovere e divulgare il lavoro scientifico con mappe stampabili di alta qualità.
- **Cytoscape[3]:** è una piattaforma software open source per visualizzare reti complesse e integrarle con qualsiasi tipo di dati. Consiste in una piattaforma per visualizzare reti di interazioni molecolari e percorsi biologici, potendo integrare queste reti con annotazioni, profili di espressione genica e altri dati. Originariamente progettato per la ricerca biologica, ora è una piattaforma generale per l'analisi e la visualizzazione di reti complesse.
- **GraphVis[5]:** è un software di visualizzazione di grafici open source. Caratteristiche, I programmi di layout Graphviz accettano descrizioni di grafici in un semplice linguaggio di testo e creano diagrammi in formati utili, come immagini e SVG per pagine web; PDF o Postscript per l'inclusione in altri documenti; o visualizzare in un browser grafico interattivo. Graphviz ha molte funzionalità utili per diagrammi concreti, come opzioni per colori, caratteri, layout di nodi tabulari, stili di linea, collegamenti ipertestuali e forme personalizzate.

- **igraph[6]**: è una collezione di librerie per creare, manipolare grafici e analizzare ponendo l'enfasi nell'efficienza, portabilità e facilità d'uso. Igraph è open source e gratuito e può essere programmato in R, Python, Mathematica e C/C++
- **UCINET6[17]**: è un pacchetto software per l'analisi dei dati dei social network. UCINET viene fornito con il tool di visualizzazione di rete NetDraw. Può leggere e scrivere una moltitudine di file di testo diversamente formattati, nonché file Excel. I metodi di analisi dei social network includono misure di centralità, identificazione di sottogruppi, analisi di ruolo, teoria dei grafi elementari e analisi statistica basata sulla permutazione. Inoltre, il pacchetto dispone di potenti routine di analisi delle matrici, come l'algebra delle matrici e la statistica multivariata.
- **SocNetV[15]**: è un'applicazione software gratuita multiplatforma per l'analisi e la visualizzazione dei social network. Tra le caratteristiche principali troviamo il poter disegnare i social network, caricare i campi da un file supportato (GraphML, GraphViz, Adjacency, EdgeList, GML, Pajek, UCINET, ecc.), personalizzare attori e collegamenti tramite sistema punta e clicca, analizzare le proprietà dei grafici e dei social network, produrre report HTML e incorporare layout di visualizzazione di rete
- **Pajek[10]**: è un software per la visualizzazione e l'analisi delle reti. La sua forza risiede nel poter analizzare reti complesse potendo arrivare fino a un miliardo di vertici. L'analisi e la visualizzazione vengono eseguite utilizzando sei tipi di dati: rete (grafico), partizione, vettore, cluster (sottoinsieme di vertici), permutazione (riordinamento dei vertici, proprietà ordinali); e gerarchia (struttura generale ad albero sui vertici).

1.4 Contributi

Da un punto di vista grafico e quindi di visualizzazione, la maggior parte degli strumenti sopra elencati, permette una certa forma di personalizzazione nella disposizione dei nodi, sia automaticamente attraverso algoritmi di layout sia manualmente, permettendo agli utenti di spostare i nodi per ottimizzare la visualizzazione o per enfatizzare certi aspetti della rete. Tuttavia, la possibilità di avere animazioni dinamiche che mostrino l'andamento del traffico nel tempo, mostrando la variazione del colore in base alla quantità dello stesso, risulta essere una caratteristica meno comune nei software di analisi di rete, nello specifico:

- Gephi: Non supporta nativamente animazioni dinamiche basate su traffico in tempo reale. Tuttavia, la sua flessibilità e la capacità di aggiungere plugin potrebbero permettere implementazioni personalizzate.
- Cytoscape: Anche se fortemente orientato all'analisi statica, plugin o estensioni potrebbero aggiungere capacità simili.
- GraphVis (Graphviz): Principalmente orientato verso la visualizzazione statica; non supporta direttamente animazioni dinamiche dei link basate sul traffico.

- **igraph**: Come libreria di analisi, non è orientato verso la visualizzazione in tempo reale o animazioni dei link basate su traffico nel suo utilizzo standard.
- **UCINET** (con **NetDraw**): Focalizzato sull'analisi statica di reti sociali; non supporta animazioni dinamiche in base al traffico.
- **SocNetV**: Orientato all'analisi statica e alla visualizzazione; non è progettato per visualizzare animazioni dinamiche basate sul traffico.
- **Pajek**: Simile agli altri, è più un tool per l'analisi statica e la visualizzazione di grandi reti, senza un supporto diretto per animazioni dei link basate su traffico.

In questo contesto, l'inserimento di una caratteristica che permetta di fare quanto premesso come base del progetto di tirocinio, risulta particolarmente indicata nel contribuire a fornire una soluzione visiva come strumento aggiuntivo di analisi di una rete, di debugging e anche come strumento didattico. La possibilità di avere un riscontro visivo istantaneo di cosa avviene nel tempo in una rete, a livello di traffico, può dare immediato feedback nel caso ci fosse un problema di congestione in un punto nevralgico, oppure mostrare parti di rete libere dove poter studiare un reindirizzamento dello stesso, volto a ottimizzare le prestazioni. A livello didattico ciò si potrebbe mostrare per presentazioni così come per didattica tramite banalmente spiegazioni. Insomma i benefici derivanti da una rappresentazione del generale sono evidenti e ciò può essere di grosso aiuto nell'analisi così anche solo come semplice rappresentazione del traffico di rete, nonchè uno strumento complementare a quanto già presente in circolazione.

1.5 Base di partenza del progetto

Il progetto è partito da zero, si basa sullo sviluppo totalmente nuovo dell'applicazione ed è stato tutto idealizzato e pianificato in sede di proposta. Come approfondirò in seguito, nella sezione dedicata alla tecnologia impiegata, il progetto non è l'unica cosa a essere partita da zero, poichè il linguaggio scelto per sviluppare l'applicazione è Python[13], linguaggio non incluso nel mio personale percorso di studi e che ho dovuto necessariamente studiare da zero per poter affrontare il percorso di tirocinio. Il lavoro è stato svolto individualmente.

Capitolo 2

Contesto di sviluppo: le tecnologie impiegate

Il progetto, a livello di tecnologia impiegata, pone le fondamenta su tre aspetti che andremo a elencare di seguito, descrivendo le varie motivazioni che hanno spinto a sceglierli.

2.1 Python

Uno dei primi aspetti di cui si è tenuto conto è stata la scelta del linguaggio di programmazione che, come accennato precedentemente, è Python. Ci sono diversi validi motivi per cui puntare su questa tecnologia; in primis è materia di insegnamento alla facoltà di Informatica de La Sapienza, ciò ha quindi una forte valenza accademica, in secondo luogo risulta essere il linguaggio più usato al mondo, ad affermarlo è l' Institute of Electrical and Electronics Engineers (IEEE)[7][1] un'associazione internazionale di scienziati professionisti con l'obiettivo della promozione delle scienze tecnologiche. Il linguaggio ha molte caratteristiche ottime, come una sintassi semplice e leggibile che lo rende facile da imparare e semplice da usare per gli sviluppatori esperti accorciando di gran lunga i tempi di sviluppo, una grande versatilità per poter essere usato in ambiti diversi come l'intelligenza artificiale, il web development, data analysis e molto altro, un ampio supporto delle librerie, due delle quali usate proprio nel progetto (di cui ne parlerò a breve), una grande comunità in cui trovare facilmente risorse, tutorial e supporto, una interoperabilità che permette un'ottima integrazione con altre tecnologie e altri linguaggi, orientato agli oggetti volto a facilitare la gestione del codice e migliora il riuso, scalabile e di facile integrazione.

2.2 La libreria Manim

Il secondo aspetto risiede nella scelta della libreria Manim[16], che viene definita come "Animation engine for explanatory math videos". Lo scopo di Manim è quindi quello di animare concetti tecnici legati alla matematica e si affida alla semplicità di Python per generare animazioni in modo programmatico. Manim può produrre anche immagini e gif, ma è nella produzione di video che splende, in questo modo è possibile progettare animazioni e renderle visibili in movimento, coprendo figure

algebriche, grafici cartesiani, grafi e molto altro[16]. La libreria offre molta libertà sui risultati che si vogliono ottenere, può renderizzare singole immagini, gif e da molte opzioni per quanto riguarda l'output video, fornendo la possibilità di renderizzarli nei formati più comuni con la possibilità di personalizzazione del framerate:

- 480p (SD): 854 x 480
- 720p (HD): 1280 x 720
- 1080p (HD): 1920 x 1080
- 1440p (2K): 2560 x 1440
- 2160p (4K): 3840 x 2160

Manim viene impiegato principalmente per presentazioni che implicino aspetti matematici, la sfida del progetto è stata quella di cercare di sfruttare le potenzialità della libreria e renderle al servizio di uno strumento di analisi sul traffico di reti, una sfida vinta come potremo vedere in seguito. Il terzo aspetto tecnologico riguarda l'aspetto di gestione dei dati. Un visualizzatore grafico di traffico di reti ha bisogno principalmente di due insiemi di informazioni importanti; uno riguarda tutte le informazioni che riguardano il come è costruita la rete, parliamo quindi degli endpoint quali sono gli switch e conseguenti informazioni annesse, pensiamo ad esempio all'indirizzo di rete, un nome identificativo e così via, ma parliamo anche dei link che collegano i vari endpoint, con la necessità di tenere traccia delle loro capacità trasmissive, la tipologia di rete, se ha una struttura a grafo completo, mesh, torus o disposizione libera e molte altre informazioni che discuteremo in seguito.

2.3 I formati Json e Yaml

L'altro insieme di informazioni deriva dal traffico vero e proprio, rendendo quindi necessario un sistema di mantenimento dei dati legati ai pacchetti trasmessi. L'analisi di questi due insiemi di informazioni ha portato alla valutazione di tre sistemi per la strutturazione di dati; json[11], yaml[19] e csv[14]. Dopo attenta analisi si è deciso di adottare il formato json per strutturare e memorizzare i dati legati al traffico, con la possibilità di scegliere anche il formato yaml, mentre per i dati relativi alla descrizione della rete il compito è stato affidato esclusivamente a yaml, csv è stato scartato. Perché csv è stato scartato? Sebbene csv rappresenti una valida alternativa tenendo conto di aspetti prestazionali, essendo un formato molto veloce da analizzare, da leggere e scrivere, tuttavia lo diventa meno quando c'è bisogno di strutturare maggiormente i dati con strutture più complesse dalla semplice forma tabellare, tipicamente usate nei database. L'idea di scartarla, sia per strutturare i dati della rete che per quelli del traffico, deriva principalmente dai seguenti motivi:

- **Leggibilità:** uno dei primi intenti del progetto era di rendere l'applicazione il più leggibile possibile, questo perché si è fortemente voluto attribuirne anche scopi di debugging e didattici, laddove avere una certa leggibilità è più che ragionevole.

- **Strutture complesse:** sicuramente il motivo più importante. Con csv non è possibile rappresentare strutture complesse, parliamo ad esempio di oggetti all'interno di altri oggetti. Sebbene per come sia ora strutturato il progetto una rappresentazione cvs è ancora possibile, ciò potrebbe non esserlo in futuro nell'ottica di espansione del progetto

Per esplicitare meglio il concetto di struttura complessa non fattibile è possibile focalizzarsi sul seguente esempio. Segue la rappresentazione di un pacchetto di rete così come viene utilizzato nel progetto, in questo caso una lista contenente un solo elemento:

```
[
  {
    "A": 1,
    "B": 2,
    "t": "2024-03-22 12:30:00",
    "d": 1518
  }
]
```

Codice 2.1. pacchetto di rete rappresentato in json

Dove A e B sono gli endpoints interessati, t è il timestamp di creazione pacchetto e d la dimensione del payload in bytes(). In csv questa struttura è rappresentabile come segue:

```
A,B,t,d
2,1,2024-03-22 12:30:00,1518
2,1,2024-03-22 12:30:00,1518
```

Codice 2.2. pacchetti di rete rappresentati in csv

Tuttavia se si dovesse rendere questa struttura più complessa, avremmo problemi a realizzarla in csv, basterebbe l'aggiunta di un campo che a sua volta necessita di informazione strutturata, come ad esempio inserire le informazione dell'header:

```
[
  {
    "A": 1,
    "B": 2,
    "t": "2024-03-22 12:30:00",
    "d": 4000,
    "ip_header": {
      "version": 4,
      "ihl": 5,
      "type_of_service": 0,
      "total_length": 300,
      "identification": 98765,
      "flags": {
        "reserved_bit": false,
        "dont_fragment": true,
        "more_fragments": false
      },
      "fragment_offset": 0,
      "time_to_live": 64,
      "protocol": 6,
      "header_checksum": "2B5A",
      "source_address": "192.168.1.1",
      "destination_address": "192.168.1.2"
    }
  }
]
```

Codice 2.3. pacchetto di rete maggiormente strutturate in json

In questa struttura abbiamo ben due oggetti nidificati, ip header e flags. Questo tipo di struttura è difficilmente replicabile in csv che si presta maggiormente per strutture piatte mentre json e yaml permettono molta più libertà di strutturazione. La scelta del formato json per rappresentare i dati del traffico di rete è quindi legata alla possibilità di espandere la rappresentazione con strutture più complesse in ottica di espansioni future del progetto, ma è indubbiamente legata alle prestazioni che questo formato riesce a dare. Json è un formato ampiamente supportato e la sua semplicità permette di ottenere risultati ottimi per le operazioni di parsing, lettura e scrittura per strutture aventi grandi quantità di dati. Sebbene l'applicazione supporti sia json che yaml per quanto riguarda il memorizzare i dati legati al traffico, la scelta preferenziale ricade su json. Questo perché, nonostante json produca files di dimensioni maggiori rispetto a yaml, la schiacciante velocità di elaborazione di json rende la creazione di un file di dimensioni maggiore ampiamente giustificabile, considerando soprattutto la natura del progetto in cui non vi è una criticità d'uso nella dimensione dei files. Come vedremo in seguito, i test eseguiti sui tempi di lettura e scrittura rispettivamente della stessa struttura ricreata sia in json che in yaml, sapranno ben dimostrare quanto appena affermato. Per impiegare questi formati sono state usate rispettivamente la libreria json [12], appartenente alla libreria standard di Python, e la libreria esterna PyYAML [2], un parser ed emitter di yaml per Python. Infine giungiamo ai motivi per cui si è scelto invece esclusivamente yaml per rappresentare i dati relativi alle informazioni che descrivono la rete. Le motivazioni consistono sempre nella volontà di garantire un formato che si presti a espansioni future e quindi che possano richiedere strutturazioni complesse, ma soprattutto, in questo caso specifico, nell'alta leggibilità che yaml offre. Generalmente i dati che servono a descrivere una rete non sono mai paragonabili a quelli necessari per registrare tutto il traffico, si voleva dare quindi uno strumento molto chiaro da leggere per fare in modo che la configurazione di rete fosse sempre molto chiara, intuitiva e di facile accesso, sia in scrittura che in lettura. Per dare una stima di grandezza, il file di configurazione network che produce il generatore di pacchetti, basato su 50 switch collegati come un grafo mesh pesa appena 9 KB. Seguirà la descrizione approfondita del progetto, principalmente suddiviso in tre parti; la prima parte riguarda la generazione di traffico di rete fittizio, la seconda è l'analizzatore dei dati di configurazione di rete e relativo traffico mentre la terza parte riguarda il visualizzatore grafico vero e proprio.

Capitolo 3

Generatore di traffico e configurazione di rete

Il progetto è composto da tre anime, una di esse è un generatore di traffico di rete e relativa configurazione. Lo scopo del generatore è da ricercare principalmente in motivazioni di testing, ma può essere usato anche in ambito didattico per scopi esemplificativi. Con esso è stato possibile testare l'analizzatore e conseguentemente il visualizzatore grafico, nonché si è creato uno strumento esemplificativo di determinati funzionamenti collegati al traffico di rete. La configurazione di rete consiste nella produzione di una struttura dati che descriva le caratteristiche della rete, la quale verrà memorizzata in un file dal nome "network" e avente estensione yaml, mentre per quanto riguarda il traffico, crea una struttura che consiste in una lista di pacchetti e memorizzata in un file denominato "packets".

3.1 Funzionalità e setup

Lo script permette di eseguire quanto descritto mettendo a disposizione due modalità, auto e user. Entrambe le modalità produrranno due files, network.yaml conterà le caratteristiche della rete e il file packets.json (o packets.yaml a seconda della scelta) conterrà il traffico vero e proprio di tutti i pacchetti generati dalla simulazione. La modalità auto chiede all'utente il numero di switch, la capacità dei link e la tipologia del grafo con il quale rappresentare la rete (completo, mesh, torus) e imposterà in modo del tutto automatico la disposizione degli switch in base alla scelta del grafo effettuata. Per poter operare, in entrambe le modalità, lo script ha bisogno di leggere il file sim_setup.yaml, il quale contiene le informazioni necessarie per la configurazione. Il file di setup è impostato come segue:

```
averageDelta: 1000
updateDelta: 100
creationDelta: 100
startSimTime: 2024-03-22 12:30:00
simTime: 5
packetSize: 4000
colorblind: "no"
dotsSize: "fixed"
trafficVariation: random
packetsFile: json
```

Codice 3.1. esempio di setup file

Andremo ora a spiegare il significato dietro ogni voce:

- **averageDelta:** rappresenta l'intervallo temporale in millisecondi delle medie percentuali di traffico da calcolare nell'analizzatore di traffico.
- **updateDelta:** rappresenta ogni quanti millisecondi debbano essere aggiornate le medie averageDelta nell'analizzatore
- **creationDelta:** rappresenta ogni quanti millisecondi debbano essere creati i pacchetti nel generatore
- **startSimTime:** è il datetime dell'inizio della generazione nel formato YY:MM:DD HH:MM:SS, inoltre è possibile specificare anche un eventuale tempo avente millisecondi da specificare come ad esempio 12:30:00.500
- **simTime:** è la durata della simulazione in secondi
- **packetSize:** è la dimensione in bytes di un pacchetto, i pacchetti nella simulazione avranno questa dimensione
- **colorblind:** è una stringa "yes" o "no" che abilita se posta su "yes" una visualizzazione compatibile per persone daltoniche di cui parleremo successivamente quando discuteremo della parte relativa al visualizzatore grafico.
- **trafficVariation:** è la variazione di traffico per secondo che si desidera impostare; può essere il valore "random" oppure uno dei seguenti [5, 10, 20, 25, 50] e di conseguenza determinerà la variazione percentuale di traffico che avviene ogni secondo di simulazione. Il valore "random" sceglie casualmente una percentuale ogni secondo con un valore che va da 0 a 100
- **packetsFile:** specifica quale tecnologia si vuole utilizzare per il file packets, si può scegliere tra json e yaml

La generazione di pacchetti è calcolata sulla base della capacità dei link fornita e su un valore casuale di percentuale di traffico che varia ogni secondo, per esempio avendo 6 links su 3 secondi di simulazione e il parametro "trafficVariation" settato a random potremmo avere delle assegnazioni di percentuali di traffico come le seguenti:

```
endpoints: [1, 2], sim second: 0, trafficPerc: 65
endpoints: [1, 3], sim second: 0, trafficPerc: 9
endpoints: [2, 4], sim second: 0, trafficPerc: 23
endpoints: [2, 5], sim second: 0, trafficPerc: 67
endpoints: [3, 6], sim second: 0, trafficPerc: 7
endpoints: [3, 7], sim second: 0, trafficPerc: 87
endpoints: [1, 2], sim second: 1, trafficPerc: 86
endpoints: [1, 3], sim second: 1, trafficPerc: 26
endpoints: [2, 4], sim second: 1, trafficPerc: 13
endpoints: [2, 5], sim second: 1, trafficPerc: 13
endpoints: [3, 6], sim second: 1, trafficPerc: 39
endpoints: [3, 7], sim second: 1, trafficPerc: 65
endpoints: [1, 2], sim second: 2, trafficPerc: 31
endpoints: [1, 3], sim second: 2, trafficPerc: 54
endpoints: [2, 4], sim second: 2, trafficPerc: 11
endpoints: [2, 5], sim second: 2, trafficPerc: 20
endpoints: [3, 6], sim second: 2, trafficPerc: 17
endpoints: [3, 7], sim second: 2, trafficPerc: 46
```

Codice 3.2. esempio di variazione traffico casuale

Una volta lanciato il generatore avremo le seguenti possibilità di scelte:

- **auto:** la modalità automatica, provvederà a disporre i nodi (switch) di rete in modo del tutto automatico. Dopo aver scelto la modalità auto si dovranno inserire i seguenti parametri via prompt:
 - numero di switch
 - capacità dei link
 - tipologia della rappresentazione grafica della rete (grafo completo, mesh, torus)
- **user:** una modalità in cui l'utente può personalizzare la configurazione di rete

La modalità user necessita di un file "custom_graph.yaml" con i parametri necessari a descrivere la rete del quale si vuole analizzare il traffico. Con questa modalità l'utente ha completa libertà nel personalizzare la rete ed è tenuto quindi a descriverne ogni suo aspetto. Il custom_graph.yaml prevede la struttura di un dizionario in cui a ogni chiave corrisponde un valore, mostriamo un esempio con possibili varianti che elencheremo:

```
---
data:
  graphType: mesh
  coordinates:
    - [1, 0, 2, 0]
    - [3, 0, 4, 5]
    - [6, 7, 8, 9]
  switches:
    1:
      ip: "123.123.123.0"
      switchName: Anthem
    2:
      ip: "123.123.123.1"
      switchName: Beta
    3:
      ip: "123.123.123.2"
      switchName: Cyber
    4:
      ip: "123.123.123.3"
      switchName: Dafne
    5:
      ip: "123.123.123.4"
      switchName: Eclipse
    6:
      ip: "123.123.123.5"
      switchName: Fox
    7:
      ip: "123.123.123.6"
      switchName: Gea
    8:
      ip: "123.123.123.7"
      switchName: H20
    9:
      ip: "123.123.123.8"
      switchName: Italy
  links:
    - { linkCap: 10, endpoints: [1, 3] }
    - { linkCap: 100, endpoints: [1, 6] }
    - { linkCap: 10, endpoints: [3, 6] }
    - { linkCap: 10, endpoints: [4, 8] }
```

```

- { linkCap: 10, endpoints: [5, 9] }
- { linkCap: 100, endpoints: [3, 5] }
- { linkCap: 10, endpoints: [6, 9] }
- { linkCap: 100, endpoints: [4, 5] }
- { linkCap: 10, endpoints: [6, 7] }
- { linkCap: 10, endpoints: [7, 8] }
- { linkCap: 100, endpoints: [8, 9] }
- { linkCap: 10, endpoints: [2, 4] }
- { linkCap: 10, endpoints: [2, 8] }
phases:
  2024-01-01 00:00:01: "phase1"
  2024-01-01 00:00:02: "phase2"

```

Codice 3.3. esempio di custom file per la configurazione

Descriviamo i vari parametri:

- **graphType:** identifica la tipologia del grafo da rappresentare, sono disponibili tre opzioni:
 - **mesh:** l'algoritmo individua in modo automatico gli archi (i link) che collegano i nodi (gli switch) adiacenti tra loro presenti nella matrice coordinates
 - **torus:** esegue lo stessa procedura usate per mesh e in addizione collega tra loro i nodi che si trovano alle estremità della matrice
 - **graph:** è la modalità più libera, collega gli switch tramite i link forniti dall'utente, indipendentemente da dove vengono collocati
- **coordinates:** rappresenta le coordinate dei vari switch, i quali vanno rappresentati con un id numerico che va da 1 a 1000. C'è una precisa motivazione di design per questa scelta ed è legata a una rappresentazione grafica ottimale del visualizzatore grafico. Gli zeri invece rappresentano uno spazio vuoto in cui non è presente uno switch.
- **links:** rappresenta i link della rete i quali possono essere specificati con i campi:
 - **linkCap:** esprime la capacità del link in Mbps
 - **endpoints:** esprime gli endpoints collegati al link
- **phases:** rappresenta le fasi temporali che accompagnano la durata dell'attività di rete, sono identificate tramite timestamp che ha come valore la descrizione della fase che parte dal timestamp stesso.

Come si può notare dalla figura 3.4, gli switch sono identificati tramite un valore numerico, questa è una precisa scelta di design e mira a mantenere il concetto di leggibilità sempre presente. In questo caso la scelta deriva dal semplificare e rendere immediatamente chiaro il campo coordinates (rappresentato da una matrice quadrata), e l'occhio identifica immediatamente la rappresentazione della disposizione. Questa leggibilità è ottenuta grazie al formato yaml, il quale minimizza i caratteri da scrivere. Così facendo si rende intuitivo il posizionamento degli switch nello spazio, inoltre è una caratteristica importante per quanto riguarda il visualizzatore grafico, poichè rappresentando i nodi con un valore numerico che va da 1 a 1000 (spiegheremo

il perchè di questa scelta a breve), è possibile rappresentare su schermo una grande quantità di elementi per far sì che possano essere visualizzati contemporaneamente e rimanere leggibili ma soprattutto, distinguibili. Uno dei problemi maggiormente riscontrati nelle applicazioni discusse nella sezione Stato dell'arte 1.3 è che, le varie soluzioni impiegate per la rappresentazione dei nodi, risultano essere confusionarie per via di una sovrapposizione massiva di link che collegano i nodi, mentre uno degli scopi del progetto è di rendere altamente leggibile l'interpretazione del traffico di rete ed è per questo che si è posta una particolare attenzione in termini di sovrapposizione degli elementi, spaziatura adeguata, scelta dei colori giusti e altri aspetti che approfondiremo nella sezione apposita riguardante la parte di visualizzazione grafica. Qualora la capacità dei link sia uguale per tutti è possibile inserire il campo `linkCap` per poter così evitare di avere lo stesso valore nei vari link rappresentati nel campo `links`, potendo rappresentare solo gli endpoints:

```
data:
  graphType: torus
  coordinates:
    - [1, 0, 2, 0]
    - [3, 0, 4, 5]
    - [6, 7, 8, 9]
  linkCap: 10
  switches:
    1:
      ip: "123.123.123.0"
      switchName: A
    2:
      ip: "123.123.123.1"
      switchName: B
    3:
      ip: "123.123.123.2"
      switchName: C
    4:
      ip: "123.123.123.3"
      switchName: D
    5:
      ip: "123.123.123.4"
      switchName: E
    6:
      ip: "123.123.123.5"
      switchName: F
    7:
      ip: "123.123.123.6"
      switchName: G
    8:
      ip: "123.123.123.7"
      switchName: H
    9:
      ip: "123.123.123.8"
      switchName: I
  links:
    - [1, 3]
    - [1, 6]
    - [3, 6]
    - [4, 8]
    - [5, 9]
    - [3, 5]
    - [6, 9]
    - [4, 5]
    - [6, 7]
    - [7, 8]
    - [8, 9]
    - [2, 4]
```

```
- [2, 8]
phases:
  2024-01-01 00:00:01: "phase1"
  2024-01-01 00:00:02: "phase2"
```

Codice 3.4. esempio di custom file con capacità uguali a tutti i link

Un'altra possibilità è quella di lasciare che il programma ricavi in automatico i link, in questo caso basterà specificare solo linkCap che sarà uguale per tutti i link:

```
---
data:
  graphType: mesh
  coordinates:
    - [1, 2, 3]
    - [4, 5, 6]
    - [7, 8, 9]
  linkCap: 10
  switches:
    1:
      ip: "123.123.123.0"
      switchName: Anthem
    2:
      ip: "123.123.123.1"
      switchName: Beta
    3:
      ip: "123.123.123.2"
      switchName: Cyber
    4:
      ip: "123.123.123.3"
      switchName: Dafne
    5:
      ip: "123.123.123.4"
      switchName: Eclipse
    6:
      ip: "123.123.123.5"
      switchName: Fox
    7:
      ip: "123.123.123.6"
      switchName: Gea
    8:
      ip: "123.123.123.7"
      switchName: H20
    9:
      ip: "123.123.123.8"
      switchName: Italy
  phases:
    2024-01-01 00:00:01: "phase1"
    2024-01-01 00:00:02: "phase2"
```

Codice 3.5. esempio di custom graph file con link di eguale capacità

Come si può notare, questo tipo di approccio concede una certa libertà di azione, dando la possibilità di prestarsi anche a soluzioni ibride di grafi canonici. Possiamo avere un esempio di quanto descritto nell'opzione torus. Il tipo di grafo torus generalmente è associato ad un grafo in cui ogni nodo ha quattro nodi adiacenti come mostrato in figura:

Tuttavia, le opzioni messe a disposizione dall'applicazione permettono una certa libertà di movimento, per esempio avendo una configurazione come quella riportata in figura 3.4, una volta analizzati i dati tramite l'analizzatore e forniti come input al visualizzatore, potremmo ottenere un risultato in cui i nodi nella matrice che abbiano un valore pari a zero, non vengano collegati:

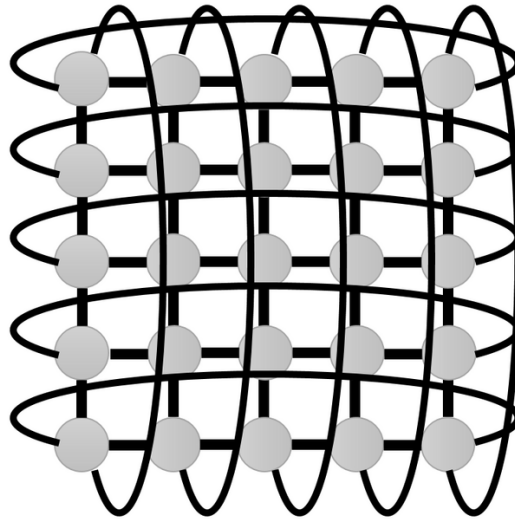


Figura 3.1. esempio di torus graph [18]

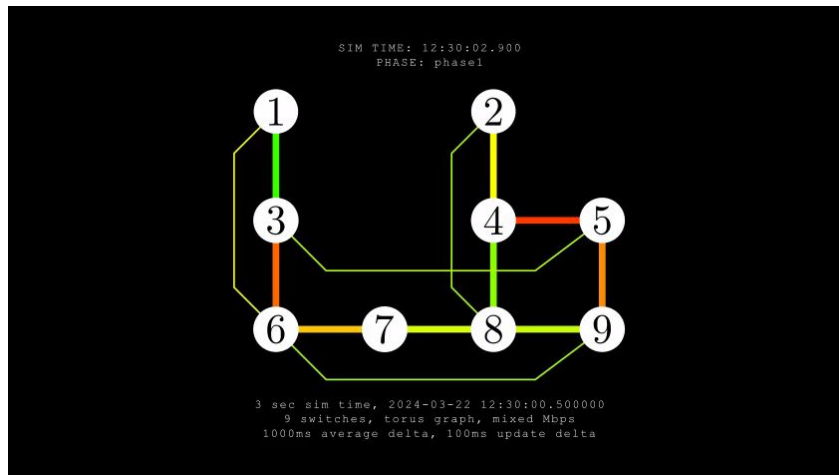


Figura 3.2. risultato grafico del visualizzatore per un custom graph file

L'ultima possibilità di personalizzazione consiste nel poter personalizzare la matrice coordinates come si preferisce, seguendo sempre la regola che laddove c'è un valore numerico, esso sarà un nodo, mentre laddove ci sarà uno zero verrà inteso come spazio vuoto. Quindi se si vuole rappresentare un grafico che non sia di tipo mesh o torus, bisogna specificare nel campo graphType il valore "graph", mostriamo un esempio:

```
---
data:
  graphType: graph
  coordinates:
    - [0, 0, 0, 0, 1, 0, 0, 0, 0]
    - [0, 0, 2, 0, 0, 0, 3, 0, 0]
    - [0, 4, 0, 5, 0, 6, 0, 7, 0]
  linkCap: 10
  switches:
    1:
```

```

    ip: "123.123.123.0"
    switchName: A
  2:
    ip: "123.123.123.1"
    switchName: B
  3:
    ip: "123.123.123.2"
    switchName: C
  4:
    ip: "123.123.123.3"
    switchName: D
  5:
    ip: "123.123.123.4"
    switchName: E
  6:
    ip: "123.123.123.5"
    switchName: F
  7:
    ip: "123.123.123.6"
    switchName: G
links:
  - [1, 2]
  - [1, 3]
  - [2, 4]
  - [2, 5]
  - [3, 6]
  - [3, 7]
phases:
  2024-01-01 00:00:01: "phase1"
  2024-01-01 00:00:02: "phase2"

```

Codice 3.6. esempio di custom graph file, graphType posto al valore graph

Come possiamo intuire osservando i valori nel campo coordinates, l'esempio ricrea la struttura di un albero, i links sono appunto specificati nel campo link e una volta generato il traffico con il generatore, analizzato e dato come input al visualizzatore otteniamo un risultato come in figura 3.3.

Vien da sè che si ottiene così una certa libertà di espressione bidimensionale, con

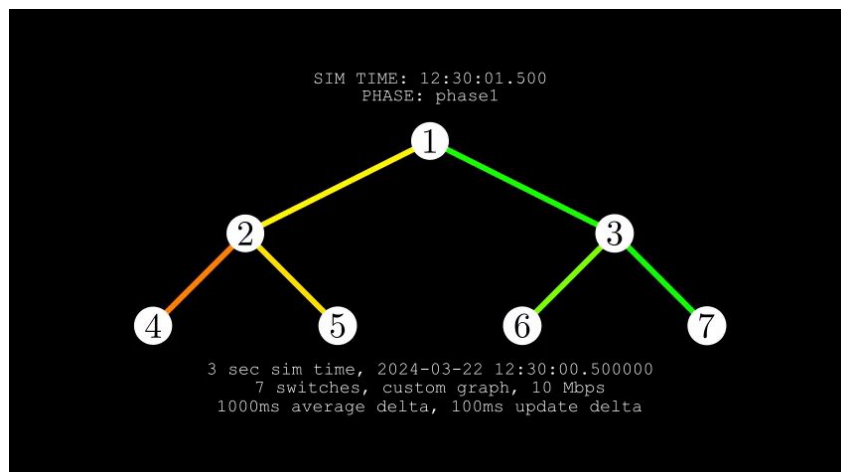


Figura 3.3. risultato grafico del visualizzatore per un custom graph file in modalità "graph"

la quale l'utente può scegliere la rappresentazione che più si addice alla rete che si vuole realizzare.

3.2 Esecuzione del generatore

Il generatore può essere pensato come suddiviso in sezioni logiche per quanto riguarda la stesura del codice nello script; la prima riguarda la fase di interazione con l'utente, il codice strutturato in una serie di richieste per l'utente e controlli sul corretto inserimento dei dati, successivamente vi è la parte di caricamento del file contenente i parametri di rete 3.1, segue la creazione dei links, la generazione del traffico creando tutti i pacchetti, si creano le informazioni inerenti agli switch e il tutto viene memorizzato sui files. Vedremo più in dettaglio questi aspetti, a livello di codice, nel capitolo dedicato allo studio della complessità computazionale e prestazioni, mentre in questo capitolo ci focalizzeremo sulla descrizione su come viene generato il tutto. Descriveremo ora il funzionamento dell'esecuzione vera e propria dello script relativo alla generazione del traffico e dei files che produrrà. All'avvio l'utente sarà chiamato a scegliere tra due modalità, come accennato precedentemente, *auto* e *user*; scegliendo *auto* verrà chiesto all'utente di scegliere il numero di switch che si vuole attribuire alla rete, da un minimo di due a un massimo di mille. Questo intervallo numerico nasce per poter rappresentare una rete in un generico datacenter, il quale generalmente ospita fino a un migliaio di switch. Ciò non toglie che, in un ottica futura di espansione del software, questo range non possa essere espanso con le dovute modifiche. Successivamente alla scelta nel numero di switch l'utente sarà chiamato a scegliere la tipologia di rete potendo scegliere tra i valori "c" (completo), "m" (mesh) e "t" (torus). Per grafo completo si intende quando un nodo è collegato a tutti i nodi rimanenti, è il grafo computazionalmente parlando più oneroso da ricreare per via della grande crescita del numero di archi da rappresentare al crescere dei nodi. Avendo infatti n nodi avremo un numero di link l pari a:

$$l = \frac{n(n-1)}{2} \quad (3.1)$$

Ogni nodo è collegato a tutti gli altri meno se stesso, $n(n-1)$, e si divide per due perchè gli archi non vanno contati due volte. Si parla quindi di una crescita esponenziale al variare del numero di nodi. Per grafo mesh si è inteso una rappresentazione in cui ogni nodo ha quattro nodi adiacenti, fatta eccezione di coloro che risiedono alle estremità, come raffigurato nell'esempio in figura creato dal visualizzatore 3.4. Nel caso della scelta di un grafico torus, il visualizzatore provvederà a disegnare la struttura come mostrata in figura 3.1. La scelta successiva in cui l'utente sarà invitato a fornire, riguarda la capacità che si vuole associare ai link appartenenti alla rete, potendo scegliere tra i valori 10, 100 e 1000 e vengono intese come grandezze espresse in megabytes (MB). Una volta inseriti questi parametri il generatore provvederà a creare la configurazione di rete e il rispettivo traffico, memorizzando i dati rispettivamente nei files `network.yaml` e `packets.json/yaml`. Scegliendo la modalità *user* non viene richiesto alcun dato via prompt, questo perchè, come descritto in precedenza, tutti i dati di cui ha bisogno il generatore devono essere presenti nel file `custom_graph.yaml` 3.4. Un diagramma che illustra le fasi e le scelte cui è possibile fare è rappresentato in figura 3.5.

Il generatore si serve di una serie di funzioni di utilità per creare le strutture in base alle scelte dell'utente. Tutte le tipologie di rappresentazioni avvengono seguendo una logica di distribuzione dei nodi in forma matriciale, fatta eccezione

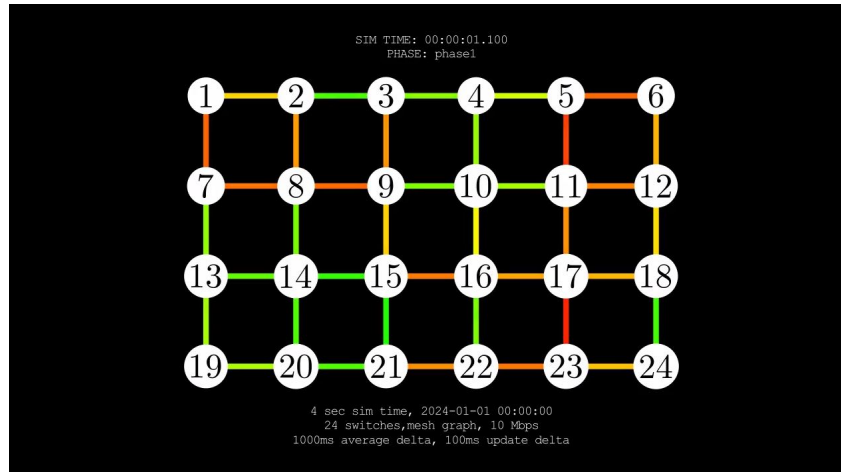


Figura 3.4. risultato grafico del visualizzatore per un auto mesh graph

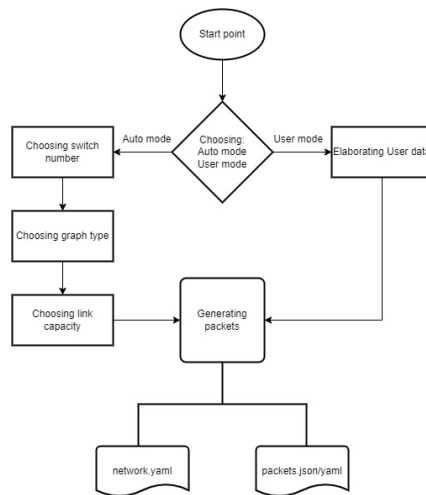


Figura 3.5. diagramma di flusso di esecuzione del generatore

della tipologia di grafo completo (spiegheremo nella parte dedicata al visualizzatore il perché), come descritto in dettaglio nel prossimo paragrafo.

3.3 Creazione dei links e nodi

Dopo aver effettuato le scelte via prompt e dopo che il programma ha caricato i parametri dal file `sim_setup.yaml` avviene la creazione dei link e dei nodi. Per i link viene creata una struttura che consiste in un dizionario in cui ogni chiave sarà associata ai parametri del link con un identificativo come chiave:

```

{
  "endpoints": sorted((switch_a, switch_b)),
  "capacity": link_cap,
  "trafficPerc": 0
}

```

Codice 3.7. esempio di rappresentazione di un link

Il campo `enpoints` rappresenta i nodi che il link collega, `capacity` è la capacità del link e `trafficPerc` è la percentuale di traffico del link. Le percentuali di traffico del link sono pensate per variare ogni secondo di generazione e servono solo ai fini della creazione della quantità di traffico, questo specifico campo non sarà riportato nel `network.yaml` poichè è un dato che non servirà più, di fatti la struttura che rappresenta i link verrà memorizzata come una lista di dizionari come segue:

```
- capacity: 10
  endpoints:
    - 1
    - 2
- capacity: 10
  endpoints:
    - 2
    - 3
- capacity: 10
  endpoints:
    - 1
    - 4
- capacity: 10
  endpoints:
    - 4
    - 5
```

Codice 3.8. esempio di struttura finale dei links

Il campo `capacity` identifica la capacità del link espressa in MB, mentre `endpoints` è una lista di due identificatori numerici compresi tra 2 e 1000 che rappresentano i nodi che li collega. Ci sono diverse configurazioni che lo script sarà tenuto a considerare in base alle scelte dell'utente, possiamo individuarle in:

- **Modalità auto** ci sono 3 configurazioni possibili, complete graph, mesh e torus.
 - **Complete graph:** i links vengono in modo tale che ogni nodo sia collegato a tutti i nodi rimanenti. Non c'è un riferimento esplicito al posizionamento poichè, come vedremo successivamente, sarà compito del visualizzatore grafico a disporre i nodi nello spazio.
 - **Mesh graph:** i links vengono creati in modo tale che ogni nodo abbia 4 nodi adiacenti rispetto a una distribuzione spaziale su matrice rettangolare, la cui verrà descritta in dettaglio nel prossimo paragrafo, fatta eccezione di quelli che risiedono all'estremità.
 - **Torus graph:** i links vengono creati in modo da replicare il più possibile la struttura di un torus graph completo, come mostrato in fig. 3.1 utilizzando lo stesso sistema di disposizione usato per il mesh graph basato su matrice rettangolare. Qualora il numero di nodi scelto non riesca a riempire la matrice, l'algoritmo non creerà i link di nodi che non risultino essere alle estremità.
- **Modalità user:** i links vengono semplicemente caricati dal file e copiati nella struttura, non vi è niente da generare.

In questa fase i nodi (ovvero gli endpoints) sono definiti tramite il campo `enpoints` dei vari links, solo successivamente verrà aggiunto l'indirizzo di ciascun endpoint, un

suo nome e l'ID. Nel caso di modalità auto vengono creati indirizzi automaticamente a partire da 10.0.0.1, l'identificativo è un valore numerico crescente a partire da 1 e il nome è semplicemente la composizione della parola *switch* seguita dall' ID, come ad esempio *switch1*. Nel caso di modalità user tutte le informazioni inerenti agli switch saranno copiate nella struttura così come sono memorizzate nel file `custom_graph.yaml`

3.4 Distribuzione dei nodi: aspect ratio

Come accennato precedentemente, la distribuzione dei nodi è pensata per essere disegnata in forma matriciale, questo segue una precisa scelta di design e riguarda l'aspect ratio adottato per disegnare al meglio gli elementi su schermo tramite la libreria Manim, la quale produce video in 16:9, quindi è sorta l'esigenza di sfruttare al massimo questo formato visivo per poter inserire nel modo più coerente e ottimizzato possibile i vari elementi della rete. La soluzione adottata è stata quella di pensare al numero di nodi come un'area di un quadrato (matrice quadrata) e successivamente trovare l'area equivalente di un rettangolo (matrice rettangolare) considerando un aspect ratio di 16:9. Consideriamo il numero di nodi A come l'area del quadrato e con n un suo lato $n = \sqrt{A}$. Denotiamo con l la lunghezza e con h l'altezza del rettangolo. La condizione di proporzione si può esprimere come

$$\frac{l}{h} = \frac{16}{9}.$$

Dato che l'area del rettangolo deve essere uguale a quella del quadrato, abbiamo che

$$l \cdot h = n^2.$$

Utilizzando la proporzione, possiamo esprimere l in termini di h come

$$l = \frac{16}{9}h.$$

Sostituendo questa espressione nell'equazione dell'area, otteniamo

$$\frac{16}{9}h \cdot h = n^2,$$

che si semplifica in

$$\frac{16}{9}h^2 = n^2.$$

Da qui, isoliamo h ottenendo

$$h^2 = \frac{9}{16}n^2 \implies h = n \cdot \frac{3}{4}.$$

Risostituendo il valore di h nell'espressione di l , abbiamo

$$l = \frac{16}{9} \cdot n \cdot \frac{3}{4} = n \cdot \frac{4}{3}.$$

Abbiamo ottenuto la base l , corrispondente al numero di colonne della matrice, e l'altezza h , corrispondente al numero di righe, del rettangolo in 16:9. Tuttavia c'è

l'esigenza di rappresentare numeri interi e la radice dell'area A può non essere un numero intero. Per ovviare al problema si è deciso di arrotondare il lato s all'intero superiore, così facendo, qualora si avesse il lato s avente parte decimale, sarà sempre verificato:

$$(l + 1)(h + 1) > lh \quad (3.2)$$

riuscendo a contenere l'area e rimanendo in proporzione.

3.5 Generazione del traffico

Successivamente alla creazione dei link e dei nodi vi è la generazione del traffico, ovvero la creazione di tutti i pacchetti che vengono creati. Descriveremo ora il criterio di creazione del traffico; l'idea è quella di iterare su frazioni di tempo cui l'unità temporale è definita dal campo `creationDelta` (tutti i campi che citeremo provengono dal file `setup sim_setup.yaml`). L'algoritmo cicla per ogni frazione di tempo per la durata della generazione, definita nel campo `simTime` e per ciascuna calcola un quantitativo di pacchetti, proporzionale alla percentuale di traffico, per ogni link presente nella rete. Ogni secondo trascorso viene cambiata la percentuale di traffico di ogni link secondo la scelta fatta nel campo `trafficVariation`. Il tempo, data e ora, che definisce quando è cominciata la trasmissione lo ritroviamo nel timestamp `startSimTime`. Poichè per ogni frazione il calcolo dei pacchetti può non essere preciso, ovvero può capitare che ci sia un numero con una parte decimale, l'algoritmo provvede a recuperare questa rimanenza, tramutarla in pacchetto e aggiungerla. Questo fa sì di avere un'elevata approssimazione, tendente al 100% per quanto riguarda l'essere più fedeli possibili alle percentuali di traffico da generare.

3.6 Files prodotti

Una volta creati links e nodi, generato il traffico, costruito le strutture del traffico e della rete, lo script provvederà a salvare il risultato nei files descritti precedentemente. Un esempio di file `network` lo ritroviamo in figura 3.9.

```
- - capacity: 10
  endpoints:
    - 1
    - 2
- - capacity: 10
  endpoints:
    - 1
    - 3
- - capacity: 10
  endpoints:
    - 3
    - 4
- - capacity: 10
  endpoints:
    - 2
    - 4
- - address: 10.0.0.1
  switchID: 1
  switchName: switch1
- - address: 10.0.0.2
  switchID: 2
  switchName: switch2
```

```

- address: 10.0.0.3
  switchID: 3
  switchName: switch3
- address: 10.0.0.4
  switchID: 4
  switchName: switch4
- averageDelta: 1000
  colorblind: 'no'
  dotsSize: fixed
  graphType: mesh
  linkCap: 10
  packetsFile: json
  simTime: 5
  startSimTime: 2024-03-22 12:30:00
  updateDelta: 100
- coordinates:
  - - 1
    - 2
  - - 3
    - 4
- 2024-03-22 12:30:01: phase1

```

Codice 3.9. esempio di network file

Il file è una lista contenente tutti i valori relativi alla configurazione network, avendo i seguenti campi nell'ordine:

- Una lista di links ognuno dei quali descrive la capacità e gli endpoints al quale è collegato
- Una lista di switch ognuno dei quali descrive l'indirizzo, il proprio identificativo e un eventuale nome associato
- Un oggetto rappresentante un dizionario con i parametri di rete:
 - averageDelta: l'intervallo di tempo delle medie percentuali di traffico da calcolare con l'analizzatore espressi in millisecondi
 - colorblind: yes/no, abilita nel caso di scelta "yes" una palette cromatica nel visualizzatore grafico adatta a persone affette da daltonismo
 - dotSize: fixed/adaptive, un'opzione che rende uguale la dimensione a tutti gli elementi raffiguranti gli switch nel visualizzatore grafico (fixed), oppure che adatta la dimensione alla lunghezza dell'identificativo degli switch (adaptive)
 - graphType: la tipologia del grafo da visualizzare (complete, mesh, torus)
 - linkCap: valore compreso tra [10, 100, 1000]/mixed, un valore che identifica la capacità dei links qualora fossero tutti della stessa grandezza, mixed altrimenti
 - packetsFile: json/yaml, la scelta del formato con il quale si vuole memorizzare il traffico di rete
 - simTime: la durata complessiva delle operazioni di rete
 - startSimTime: timestamp, data e ora nel formato YY:MM:DD HH:MM:SS dell'inizio della trasmissione del traffico

- updateDelta: un intervallo temporale che identifica ogni quanto la media averageDelta debba essere aggiornata
- Una lista di liste che descrive la matrice del posizionamento degli switch. Nel caso di scelta di grafo completo questo campo è una lista vuota ([])
- Una lista di fasi temporali che descrivono specifici intervalli temporali nel formato "YY:MM:DD HH:MM:SS = descrizione fase"

Per il file packets abbiamo invece due possibili risultati in base alla scelta fatta nel file di setup 3.1, ovvero json e yaml. Il generatore crea i pacchetti secondo l'assunzione in cui essi vengano registrati in ordine di invio, avremo quindi una struttura corrispondente a una lista di pacchetti ordinati per timestamp di invio. Le strutture si presentano come in figura 3.10 per il formato json e in figura 3.11 per il formato yaml. Come vedremo nel prossimo capitolo, questi due files rappresentano l'esatto formato che riesce a leggere l'analizzatore, quindi un utente che voglia usare l'analizzatore e il visualizzatore grafico dovrà formattare accuratamente i suoi dati come descritto.

```
[
  {
    "A": 1,
    "B": 2,
    "t": "2024-03-22 12:30:00",
    "d": 4000
  },
  {
    "A": 2,
    "B": 1,
    "t": "2024-03-22 12:30:00",
    "d": 4000
  },
  {
    "A": 2,
    "B": 1,
    "t": "2024-03-22 12:30:00",
    "d": 4000
  }
]
```

Codice 3.10. esempio di packets file in formato json

```
- A: 1
  B: 2
  d: 4000
  t: &id001 2024-03-22 12:30:00
- A: 1
  B: 2
  d: 4000
  t: *id001
- A: 1
  B: 2
  d: 4000
  t: *id001
```

Codice 3.11. esempio di packets file in formato yaml

In cui in entrambi i casi abbiamo i seguenti campi:

- A: l'endpoint di partenza del pacchetto

- B: l'endpoint di arrivo del pacchetto
- d: la dimensione del payload del pacchetto espresso in Bytes
- t: il timestamp rappresentante la data e l'ora di invio del pacchetto

La scelta di usare un solo carattere per descrivere i vari campi può sembrare in contrasto con la volontà di rendere leggibile il tutto, ma c'è una valida motivazione dietro. Il risparmio di caratteri, in files che sono potenzialmente molto grandi, è un aspetto cruciale sia per quanto riguarda l'occupazione di memoria fisica, sia per motivi di efficienza. Consideriamo una rappresentazione di un pacchetto più descrittivo come in figura 3.13:

```
{
  "endpointA": 1,
  "endpointB": 2,
  "timestamp": "2024-03-22 12:30:00",
  "dimension": 1518
}
```

Codice 3.12. rappresentazione di un pacchetto con descrizione completa

Avremmo così 37 bytes necessari per i nomi contro i 4 per la rappresentazione di un pacchetto descritta precedentemente, avendo uno scarto di 33 bytes per pacchetto. Questa quantità, moltiplicata la grande quantità di pacchetti da rappresentare, può pesare non poco. Indipendentemente dal numero di bytes necessari a rappresentare un pacchetto in forma minimale, possiamo fare una stima di quanto sarebbe più grande un file packets avendo descrizioni di nomi più lunghi; supponiamo di avere un grafo completo costituito da 5 switch collegati tra loro con switch aventi capienza 1 Gbps (gigabit), il numero dei link, per l'equazione in fig. 3.1, risulta essere pari a 10. Supponiamo che tutta la rete stia trasmettendo al 50% della sua capacità trasmissiva per tutta la durata dell'attività di rete. Considerando una dimensione d di payload pari a 1518 Bytes uguale per tutti i pacchetti trasmessi, definiamo come pps i pacchetti al secondo, da cui, per un link in un dato secondo avremo:

$$\begin{aligned}
 1 \text{ Gbps} &= 10^9 \text{ bit} \\
 1 \text{ GBps} &= \frac{10^9}{8} = 125,000,000 \text{ bytes} & (1 \text{ byte} = 8 \text{ bit}) \\
 50\% \text{ capacity} &= \frac{125,000,000}{2} = 62,500,000 \text{ bytes} \\
 d &= 1518 \text{ bytes} \\
 pps &= \frac{62,500,000 \text{ bytes}}{d} \approx 41172.59
 \end{aligned}$$

$$\text{bytes di differenza} = 41172.59 \times 33 \approx 1.35 \text{ MBytes}$$

Ciò significa che per ogni secondo di traffico avremmo $1.35 \text{ MB} \times 10 \text{ links} = 13.5 \text{ MB}$ in più, in 100 secondi di traffico avremmo 1350 Mbytes, cioè 1.35 GBytes in più per una rete avente soltanto 10 links, quindi con la scelta di rendere mono carattere le descrizioni si ha un notevole risparmio in termini di data storage. Chiaramente in ottica di espansione futura del progetto, avendo strutture più complesse per i pacchetti, potrebbe essere necessario avere elementi maggiormente descrittivi, tuttavia ciò non toglie che una certa cura nella scelta di una descrizione minimale non possa essere comunque adottata in via preferenziale.

3.7 Aspetti realizzativi scartati: yaml flow e block style

Un approccio inizialmente realizzato e poi scartato, è stato quello di poter dare la possibilità di scegliere la modalità di salvataggio del file `packets.yaml`. La libreria `PyYAML` permette di salvare un file `yaml` secondo due modalità; la prima è definita `block style` e l'altra `flow style`. Di default, `PyYAML` sceglie lo stile di una `collection` a seconda che abbia a sua volta `collection` nidificate. Qualora ci fossero `collection` nidificate verrà assegnato lo stile `block`, il quale è il formato come lo abbiamo già visto in figura 3.11, altrimenti verrà assegnato lo stile `flow` che corrisponde al codice in figura.

```
[{A: 2, B: 1, d: 4000, t: &id001 !!timestamp '2024-03-22 12:30:00'}, {A: 1, B: 2, d: 4000, t: *id001}, {A: 1, B: 2, d: 4000, t: *id001}, {A: 2, B: 1, d: 4000, t: *id001}, {A: 2, B: 1, d: 3752, t: *id001}, {A: 1, B: 2, d: 4000, t: &id002 !!timestamp '2024-03-22 12:30:00.100000'}, {A: 2, B: 1, d: 4000, t: *id002}, {A: 1, B: 2, d: 4000, t: *id002}, {A: 1, B: 2, d: 4000, t: *id002}, {A: 2, B: 1, d: 3752, t: *id002}, {A: 1, B: 2, d: 4000, t: &id003 !!timestamp '2024-03-22 12:30:00.200000'}, {A: 2, B: 1, d: 4000, t: *id003}, {A: 2, B: 1, d: 4000, t: *id003}, {A: 1, B: 2, d: 4000, t: *id003},
```

Codice 3.13. rappresentazione di un pacchetto con descrizione completa

Inizialmente si era provato a dare la possibilità di scegliere, tramite file `sim_setup`, tra il formato `block` e `flow` poichè in fase di test si è visto che l'opzione `flow` riduce ulteriormente la dimensione del file `packets.yaml`. Tuttavia i test hanno dimostrato che le differenze delle grandezze dei files e i tempi necessari a realizzarli non erano così diversi tra loro, quindi si è deciso di scartare del tutto l'opzione. Possiamo vedere il risultato del test sulle dimensioni dei files in figura 3.6 e quello relativo ai tempi di scrittura/lettura in figura 3.7. Possiamo notare dai grafici che il risparmio di bytes è di circa 4 MB per il test effettuato su 10 switch, mentre i tempi di esecuzione, sia di scrittura che di lettura, sono praticamente sovrapponibili, infatti l'operazione di scrittura per `block` e `flow` style è identico, mentre lievemente più performante è il `block` style in lettura.

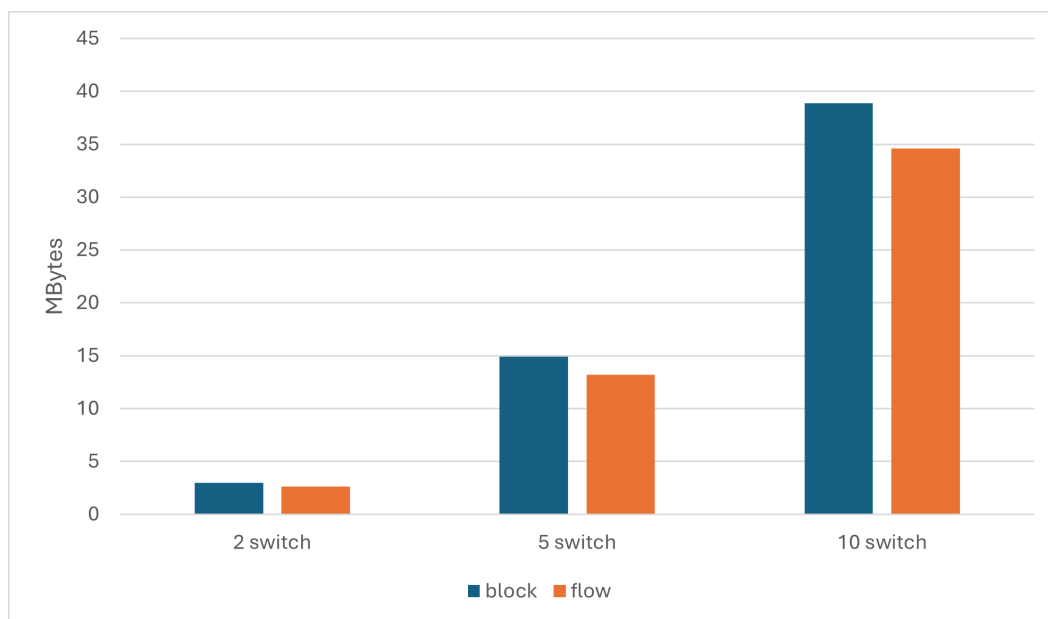


Figura 3.6. test yaml block/flow style, dimensione files

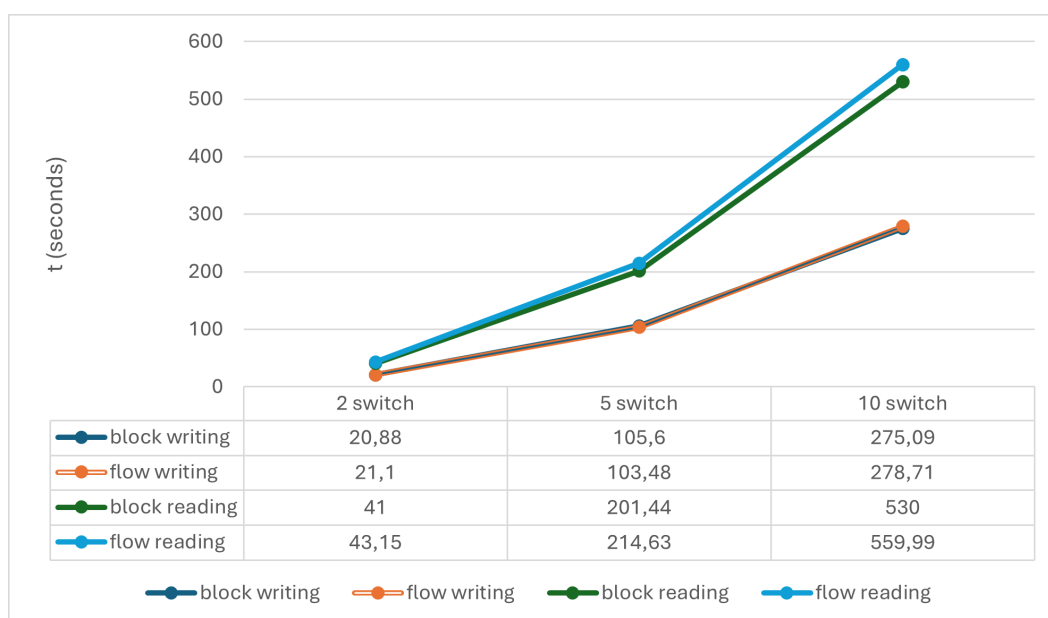


Figura 3.7. test yaml block/flow style, tempi di scrittura e lettura

Capitolo 4

Analizzatore di traffico

L'analizzatore di traffico è la seconda anima del progetto, è la parte che analizza il traffico calcolando le medie percentuali e producendo un file, `analyzed_data.yaml`, che conterrà sia le informazioni di rete come links, switches e parametri di rete, sia le registrazioni del traffico percentuale di ogni link. Per poter operare, l'analizzatore ha bisogno dei due files, `network.yaml` e `packets.yaml`, creati dal generatore o che siano creati e formattati dall'utente. Abbiamo descritto in dettaglio la composizione dei files nella sezione 3.6. L'esecuzione dello script parte in primo luogo caricando i files richiesti e creando le strutture

Capitolo 5

Visualizzatore grafico

Capitolo 6

Prestazioni e complessità computazionale

Per quanto riguarda la creazione automatica dei links e dei nodi per la tipologia di grafo completo, la funzione dedicata al compito è riportata in figura 6.1

```
def create_auto_complete_links(link_cap, switch_number):
    """ Create links for a complete graph

    Key arguments:
    link_cap: int -- link capacity
    switch_number: int -- switch number

    Returns
    dict -- links representation
    """
    # The arcs representing the links connecting the switches (nodes)
    links = {}
    # The link ID counter
    link_id = 1

    for i in range(1, switch_number + 1):
        for p in range(i, switch_number + 1):
            if i != p:
                if link_id not in links:
                    links[link_id] = link_format(i, p, link_cap)
                    link_id += 1

    return links
```

Codice 6.1. funzione per la creazione automatica di un grafo di rete completo

Capitolo 7

Conclusioni

Capitolo 8

Test

La complessità temporale dell'algoritmo è $O(m + n)$.

$$O(m + n) \tag{8.1}$$

.. .. .

Bibliografia

- [1] Stephen Cass Author. *The Top Programming Languages 2023*. <https://spectrum.ieee.org/the-top-programming-languages-2023>. Accessed: 2024-04-12. 2023.
- [2] Canonical. *PyYAML Documentation*. Accessed: 2024-04-12. 2024. URL: <https://pyyaml.org/wiki/PyYAMLDocumentation>.
- [3] Cytoscape Consortium. *Cytoscape: An Open Source Platform for Complex Network Analysis and Visualization*. <https://cytoscape.org>. Accessed: 2024-04-12. 2024.
- [4] Gephi. *Gephi: The Open Graph Viz Platform*. <https://gephi.org>. Accessed: 2024-04-12. 2024.
- [5] Graphviz. *Graphviz - Graph Visualization Software*. <https://graphviz.org>. Accessed: 2024-04-12. 2024.
- [6] igraph. *igraph - The Network Analysis Package*. <https://igraph.org>. Accessed: 2024-04-12. 2024.
- [7] Institute of Electrical and Electronics Engineers. *IEEE - Advancing Technology for Humanity*. <https://www.ieee.org>. Accessed: 2024-04-12. 2024.
- [8] Johns Hopkins University. *Homepage of Johns Hopkins University*. <https://www.jhu.edu/>. Accessed: 2024-04-12. 2024.
- [9] Johns Hopkins University Libraries. *Network Data Visualization Guide*. <https://guides.library.jhu.edu/datavisualization/network>. Accessed: 2024-04-12. 2024.
- [10] Pajek. *Pajek - Program for Large Network Analysis*. <http://mrvar.fdv.uni-lj.si/pajek/>. Accessed: 2024-04-12. 2024.
- [11] Jon Postel. *Internet Protocol*. RFC 791. Accessed: 2024-04-12. RFC Editor, 1981. URL: <https://www.rfc-editor.org/rfc/rfc791#page-11>.
- [12] Python Software Foundation. *JSON encoder and decoder – Python 3.10.8 documentation*. Accessed: 2024-04-12. 2024. URL: <https://docs.python.org/3/library/json.html>.
- [13] Python Software Foundation. *Python Programming Language – Official Website*. <https://www.python.org>. Accessed: 2024-04-12. 2024.
- [14] Y. Shafranovich. *Common Format and MIME Type for Comma-Separated Values (CSV) Files*. RFC 4180. Accessed: 2024-04-12. RFC Editor, 2005. URL: <https://www.rfc-editor.org/rfc/rfc4180.html#page-2>.

- [15] SocNetV. *Social Network Visualizer (SocNetV)*. <https://socnetv.org>. Accessed: 2024-04-12. 2024.
- [16] The Manim Community Developers. *Manim Documentation: Quickstart Guide*. Accessed: 2024-03-25. 2024. URL: <https://docs.manim.community/en/stable/index.html>.
- [17] UCINET. *UCINET Software*. <https://sites.google.com/site/ucinetsoftware/>. Accessed: 2024-04-12. 2024.
- [18] Wikipedia contributors. *File:Torus graph.png*. https://en.m.wikipedia.org/wiki/File:Torus_graph.png. Accessed: 2024-04-12. 2024.
- [19] YAML Core Team. *YAML Ain't Markup Language (YAML™) Version 1.2*. <https://www.rfc-editor.org/info/rfc9512>. Accessed: 2024-04-12. 2009.