

Raccolta e modellazione di dati reali di rete per scenari Cloud-Edge

Proposta

Le prestazioni dei sistemi distribuiti cloud-edge dipendono in modo determinante dalle caratteristiche della rete che collega nodi, siti e regioni: latenze tra punti geografici diversi, banda disponibile, variabilità temporale dei ritardi, congestione, perdita di pacchetti e struttura topologica della rete. Questi fattori influenzano direttamente decisioni operative come il posizionamento dei servizi, la replica su più siti o la migrazione di componenti tra cloud ed edge.

Nello stato dell'arte esistono dataset e misurazioni reali relativi a reti WAN, infrastrutture edge/MEC e topologie di rete, ma tali dati sono spesso separati per contesto applicativo, raccolti con metodologie diverse e descritti tramite formati e metriche non uniformi. Di conseguenza, è difficile utilizzarli in modo integrato per costruire scenari di rete realistici utili allo studio di sistemi distribuiti e strategie di gestione delle risorse.

L'obiettivo della tesi è raccogliere, analizzare e integrare dati reali di rete costruendo un dataset coerente, pulito e strutturato che descriva in modo omogeneo topologie, nodi, collegamenti e metriche di rete in scenari cloud-edge. Il dataset dovrà permettere di rappresentare, ad esempio, relazioni tra siti, ritardi tra nodi, capacità dei link e variabilità delle prestazioni nel tempo, così da poter essere usato come base per simulazioni e studi su orchestrazione e sistemi distribuiti.

Il lavoro prevede l'analisi di dataset pubblici relativi a topologie e metriche di rete, la definizione di un modello dati unificato per descrivere nodi, link e proprietà di rete, la pulizia dei dati (normalizzazione delle unità di misura, gestione di valori mancanti, conversione dei formati) e, dove possibile, l'allineamento temporale di misurazioni provenienti da fonti diverse. Una parte significativa della tesi sarà dedicata alla caratterizzazione statistica delle reti, ad esempio studiando la distribuzione delle latenze, le differenze tra collegamenti edge-cloud e intra-cloud, e la variabilità delle prestazioni nel tempo. Il risultato sarà un dataset documentato e riutilizzabile, accompagnato da un'analisi che ne evidenzia le principali proprietà e limiti.

Strumenti e tecnologie

- Python
- Pandas / NumPy
- Dataset pubblici di rete e topologie
- Matplotlib / Seaborn
- JSON / CSV / Parquet

Obiettivi

1. Analizzare e selezionare fonti dati rilevanti su topologie e metriche di rete.
2. Progettare uno schema dati unificato.
3. Implementare pipeline di integrazione e pulizia dei dati.
4. Effettuare analisi statistiche su latenze e bande.
5. Produrre un dataset documentato e riutilizzabile.

Prerequisiti

Buona conoscenza di Python.

Nozioni di base di reti di calcolatori.

Interesse per sistemi distribuiti e analisi dati.

È inoltre richiesto di mostrare un caso d'uso concreto che utilizzi il dataset prodotto.