

UNIVERSITÀ DEGLI STUDI DI MILANO
FACOLTÀ DI SCIENZE E TECNOLOGIE

DIPARTIMENTO DI INFORMATICA
GIOVANNI DEGLI ANTONI



Corso di Laurea triennale in Informatica

PATTERN DI ASSEGNAIMENTO OTTIMIZZATI PER
L'EFFICIENZA ENERGETICA IN EDGE COMPUTING

Relatore: Prof. Alberto Ceselli
Correlatore: Prof. Marco Premoli

Tesi di Laurea di:
Manuel Parati
Matr. 958584

ANNO ACCADEMICO 2021-2022

Ringraziamenti

Questa sezione, contiene i ringraziamenti.

Indice

Ringraziamenti	i
Indice	ii
1 Introduzione	1
1.1 Obbiettivi	1
1.2 Struttura dell’elaborato	1
2 Modellazione del sistema	2
2.1 L’infrastruttura Mobile Edge Computing	2
2.2 Modello di orchestrazione proposto	3
Bibliografia	5

Capitolo 1

Introduzione

...

1.1 Obbiettivi

...

1.2 Struttura dell'elaborato

...

Capitolo 2

Modellazione del sistema

Questo capitolo fornisce una panoramica riguardo l'infrastruttura del Mobile Edge Computing ed introduce il modello di orchestrazione proposto.

2.1 L'infrastruttura Mobile Edge Computing

All'interno di un'infrastruttura Mobile Edge Computing (MEC) si trovano i cluster di virtualizzazione, spesso chiamati 'MEC facility' o più semplicemente 'facility', e gli access point (AP). Le facility sono il luogo in cui avviene la virtualizzazione, e sono composte dalle macchine virtuali (VM) su cui si eseguono le applicazioni degli utenti finali. Gli AP sono invece i dispositivi (per esempio le antenne wireless) a cui gli end point si collegano per poter ricevere il servizio. Ogni AP è associato ad una facility, a cui inoltra tutto il traffico che riceve: questo significa che ogni facility avrà in esecuzione nelle proprie VM tutte le applicazioni utilizzate dagli utenti connessi agli AP a lui associati.

Data la natura mobile degli end point, che sono per esempio smartphone o laptop, il traffico a cui sono sottoposti gli AP varia nel tempo e di conseguenza cambia la domanda rivolta alle facility. Per questo motivo l'assegnamento viene effettuato dinamicamente, e lo strumento incaricato di tale compito è l'orchestratore, che implementa la logica definita dal modello di orchestrazione. Le azioni che svolge vengono chiamate 'orchestrazioni' e consistono nell'assegnare un AP ad una facility diversa da quella attuale, provocando il ridimensionamento della potenza delle VM in termini del numero di processori e memoria utilizzata, e la migrazione del loro stato. Il ridimensionamento è dovuto alla variazione di domanda da soddisfare, mentre la migrazione dello stato è necessaria per avere in esecuzione in ogni facility le applicazioni utilizzate dagli utenti connessi agli AP che le sono assegnati e quindi ridurre la latenza. Il costo prodotto dalla migrazione prende il nome di 'costo di migrazione'.

Nel sottocapitolo successivo viene illustrato il modello di orchestrazione proposto in questo lavoro.

2.2 Modello di orchestrazione proposto

Il modello di orchestrazione proposto è l'equivalente a quello presentato negli articoli [1], [2] in cui nelle scelte di orchestrazione viene tenuta in considerazione anche la gestione dell'energia da parte delle facility.

Come illustrato in [2], all'interno del modello viene introdotta una discretizzazione temporale che permette di effettuare migrazioni solo in determinati istanti, per esempio ogni 15 minuti. Questa scelta è ragionevole, dato che in caso contrario si pagherebbe un costo troppo alto in termini di migrazione e spese generali dovute alla gestione del traffico. L'orizzonte temporale viene quindi rappresentato da un insieme di time slot della stessa durata.

L'obiettivo del modello è quello di effettuare assegnamenti che permettano di ottenere un buon compromesso tra: una buona qualità del servizio (cioè bassa latenza per l'utente finale), costi di migrazione contenuti e buona gestione dell'energia da parte delle facility. Nell'effettuare le scelte bisogna inoltre tenere in considerazione che la domanda degli AP varia nel corso del tempo e che ogni facility possiede un limite massimo di domanda che è in grado di soddisfare simultaneamente.

Dal punto di vista energetico, le facility utilizzano in ogni slot time una quantità di energia direttamente proporzionale alla domanda complessiva che gli viene rivolta. A tale scopo, ognuna di esse possiede un certo numero di pannelli fotovoltaici che le permettono di generare energia variabile nel tempo, che può essere direttamente utilizzata oppure immagazzinata all'interno delle proprie batterie per i time slot successivi. Le batterie hanno una capacità limitata e nel caso in cui l'energia prodotta non basti a soddisfare la domanda, è possibile acquistarla ad un prezzo che varia in base all'istante temporale.

Una semplice applicazione di esempio è presentata nella Figura 1. Come si può vedere, è presente una rete MEC con due facility (K1 e K2) e quattro AP (A, B, C, D) considerata in due time slot consecutivi ($t=0, 1$). Gli end point connessi agli AP sono rappresentati con dei piccoli cerchi. Si supponga che per gestire la domanda di ogni utente sia necessario utilizzare una singola unità di energia e che inizialmente le batterie siano vuote in entrambe le facility. Nel primo time slot ($t=1$) gli AP A e B sono associati alla facility K1 mentre C e D a K2, quindi tutti gli utenti che sono connessi ad A e B avranno le proprie applicazioni in esecuzione su una VM presente in K1, mentre quelli connessi a C e D le avranno in K2. Questi assegnamenti permettono di avere una buona latenza, dato che ogni AP è connesso alla facility più vicina, ma anche una buona gestione energetica,

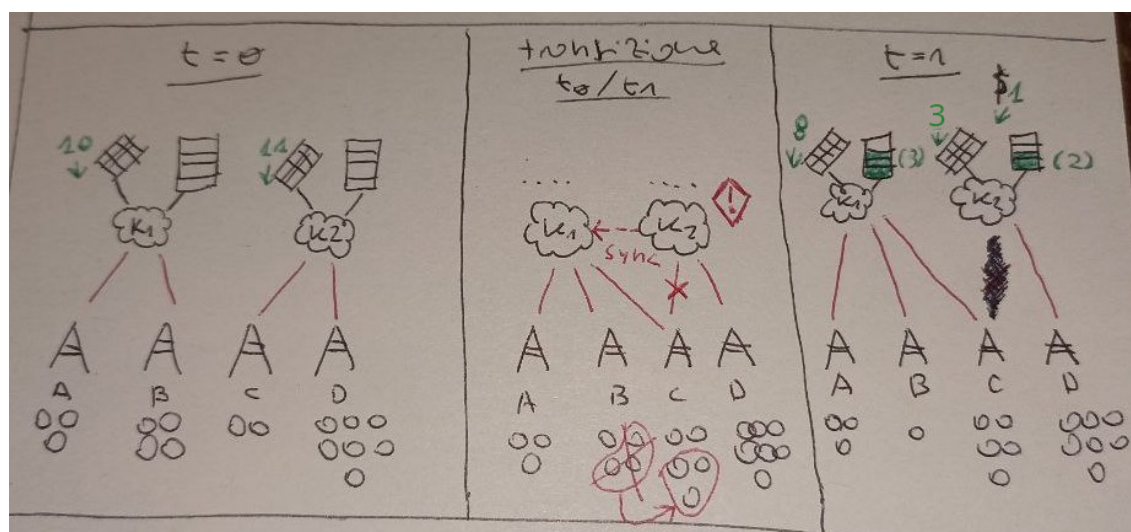


Figura 1: Esempio funzionamento del modello di orchestrazione.

in quanto nella facility K1 vengono prodotte 10 unità di energia ed utilizzate solo 7 mentre in K2 se ne producono 11 e utilizzano 9. In entrambi i casi avanzano delle unità (3 in K1 e 2 in K2) che verranno immagazzinate nelle batterie per poter essere spese nei prossimi time slot. Successivamente tre utenti connessi all'AP B si spostano e vengono agganciati da C: a questo punto K2 riceve una richiesta di domanda eccessiva che non riesce a gestire e di conseguenza è necessario effettuare un'azione di orchestrazione per ribilanciare il traffico. Questo comporta assegnare C a K1, ridimensionare le VM (aumentarne la potenza in K1 ed abbassarla in K2) e sincronizzare le facility trasferendo lo stato delle VM riguardanti tutti gli utenti connessi a C da K2 a K1. In un contesto reale, il modello deve prevenire una situazione di questo tipo effettuando orchestrazioni che tengano in considerazione la futura domanda dei vari AP, dato che questa circostanza va ad inficiare la qualità del servizio offerto. Nel secondo time slot ($t=1$) si avranno quindi gli AP A, B e C assegnati alla facility K1, mentre D a K2. Dal punto di vista energetico, la facility K1 riesce a soddisfare la domanda di traffico (9 unità) utilizzando l'energia presente nella batteria e quella prodotta in questo time slot (8 + 3 unità complessive). Per quanto riguarda invece K2, l'energia disponibile (3 + 2 unità) non è sufficiente per colmare la domanda, e quindi è necessario acquistare una ulteriore unità. Da notare come l'energia potrebbe avere costi diversi nel corso dei time slot, e di conseguenza per il modello potrebbe risultare conveniente acquistarla quando il costo è basso per immagazzinarla nelle batterie ed utilizzarla successivamente.

Bibliografia

- [1] Alberto Ceselli, Marco Fiore, Marco Premoli, and Stefano Secci. Optimized assignment patterns in mobile edge cloud networks. *Computers & Operations Research*, 106:246–259, 2019.
- [2] Alberto Ceselli, Marco Fiore, Angelo Furno, Marco Premoli, Stefano Secci, and Razvan Stanica. Prescriptive analytics for mec orchestration. In *2018 IFIP Networking Conference (IFIP Networking) and Workshops*, pages 1–9, 2018.