

Reading Report: Satyanarayanan17

Ricard Medina Amado

June 4, 2023

Upload your report in PDF format.

Use this LaTeX template to format the report, keeping the proposed headers.

The length of the report must not exceed **5 pages**.

1 Content

1.1 *Identify the genre¹ of the document, its purpose, and its target audience.*

Es tracta d'un paper dirigit a una audiència amb un coneixement mínim sobre aspectes de TI.

1.2 *Summarize the document, indicating the key concepts².*

2 Summary

- **Origin and background** Les arrels de la computació a l'extrem s'estenen fins a la dècada dels 90, quan Akamai va introduir les xarxes de lliurament de continguts (CDN) per accelerar el rendiment web. Un CDN utilitza nodes propers als usuaris per emmagatzemar contingut web i personalitzar-lo segons la ubicació. La computació a l'extrem amplia aquest concepte utilitzant "cloudlets" en l'extrem de la xarxa, que poden executar codi arbitrari. Això permet aprofitar la computació en el núvol a prop dels usuaris finals. Els investigadors han demostrat el valor de la computació a l'extrem per a la informàtica mòbil i han proposat el terme "cyber foraging" per a l'amplificació de les capacitats informàtiques dels dispositius mòbils. Tot i que la computació en el núvol és popular, la seva centralització pot implicar una separació significativa entre els dispositius mòbils i els centres de dades del núvol. Això pot generar retard i no és adequat per a aplicacions que requereixen un control precís de la latència.
- **Why proximity matters** A mesura que explorem noves aplicacions i casos d'ús per a la informàtica mòbil i l'IoT, les virtuts de la proximitat es fan cada vegada més evidents. En el món físic, la importància de la proximitat mai ha estat en dubte. L'axioma antic sobre els tres

¹Genres: book, article, essay, report, review, manual, white paper, data sheet, weblog, etc.

²The summary should help you to answer the questions about the reading in the exam.

principals determinants del valor immobiliari, "ubicació, ubicació i ubicació", captura aquesta observació de manera precisa. En el món ciber, la connectivitat perfecta que ofereix Internet ens ha adormit en una falsa indiferència per la proximitat física. Com que la proximitat lògica de la xarxa es caracteritza completament per una baixa latència, una baixa variabilitat i un ample de banda elevat, la pregunta "Quan de prop és prou físicament?" no es pot respondre de manera abstracta. Depèn de factors com les tecnologies de xarxa utilitzades, la contenció de la xarxa, les característiques de l'aplicació i la tolerància de l'usuari a una resposta interactiva deficient.

La proximitat física afecta la latència de punta a punta, l'ample de banda econòmicament viable, l'establiment de la confiança i la supervivència. Amb prou esforç i inversió de recursos, la falta de proximitat es pot ocultar parcialment. Per exemple, una connexió de fibra òptica directa pot aconseguir una baixa latència i un ample de banda elevat entre punts distants. No obstant això, hi ha límits a aquest enfocament. La velocitat de la llum és un límit físic evident de la latència. La necessitat d'utilitzar una estratègia de xarxa multi-salt per cobrir una àrea geogràfica gran amb molts punts d'accés imposa un límit econòmic tant a la latència com a l'ample de banda. Cada salt introdueix retard de cua i encaminament, així com un creixement dels búfers.

Serveis en el núvol altament reactius. La proximitat física d'un "cloudlet" a un dispositiu mòbil facilita aconseguir una baixa latència de punta a punta, un ample de banda elevat i una baixa variabilitat en els serveis ubicats al "cloudlet". Això és valuós per a aplicacions com l'AR (realitat augmentada) i la realitat virtual que desvien la computació al "cloudlet".

Escalabilitat mitjançant anàlisi a l'extrem. La demanda acumulada d'ample de banda d'entrada al núvol des d'una gran col·lecció de sensors IoT d'ample de banda elevat, com ara càmeres de vídeo, és considerablement menor si les dades brutes s'analitzen als "cloudlets". Només la informació i metadades extretes (molt més petites) s'han de transmetre al núvol.

Aplicació de les polítiques de privadesa.

Actuant com a primer punt de contacte en la infraestructura per a les dades dels sensors IoT, un "cloudlet" pot aplicar les polítiques de privadesa del seu propietari abans de l'alliberament de les dades al núvol.

Dissimulació de fallades en el núvol. Si un servei en el núvol no està disponible a causa d'una fallada de xarxa, fallada del núvol o atac de denegació de servei, un servei alternatiu en un "cloudlet" proper pot dissimular temporalment la fallada.

- **Highly responsive cloud services** Les persones són extremadament sensibles als retards en el camí crític de la interacció, i el seu rendiment en tasques cognitives és sorprenentment ràpid i precís. Per exemple, en condicions d'il·luminació normals, el reconeixement facial triga de 370 a 620 ms, depenent de la familiaritat. El reconeixement de la parla triga de 300 a 450 ms per a frases curtes, i només requereix 4 ms per a reconèixer

que un so és una veu humana. Les aplicacions de realitat virtual que utilitzen sistemes de seguiment de cap requereixen latències inferiors a 16 ms per aconseguir estabilitat perceptual. Una latència de punta a punta de poques desenes de mil·lisegons és un objectiu segur però realitzable.

La Figura 1 il·lustra la importància dels "cloudlets" per als serveis de desplaçament de baixa latència. Els gràfics mostren la distribució acumulada dels temps de resposta mesurats per a una aplicació de realitat augmentada i un reconeixement facial en un dispositiu mòbil. Una imatge del dispositiu mòbil, que es troba a Pittsburgh, es transmet a través d'un primer salt Wi-Fi cap a un "cloudlet" o cap a un centre de dades d'Amazon Web Services (AWS). La imatge es processa al destí mitjançant codi de visió per ordinador que s'executa dins d'una màquina virtual. En el cas de la realitat augmentada, els edificis de la imatge són reconeguts i les etiquetes corresponents a les seves identitats es transmeten al dispositiu mòbil. En el cas del reconeixement facial, es retorna la identitat de la persona.

La corba ideal de la Figura 1 seria una funció de passos que salta a 1.0 a l'origen. Com mostra la figura, l'ideal s'aproxima millor a un "cloudlet". La latència de la xarxa de punta a punta entorpeix el rendiment, com indiquen les corbes de temps de resposta que empitjoren corresponents a ubicacions d'AWS més llunyanes. L'augment del temps de resposta també incrementa el consum d'energia per operació en el dispositiu mòbil. Aquest valor s'indica al costat de l'etiqueta corresponent a la llegenda de la figura. Per exemple, el dispositiu consumeix de mitjana 1.1 J per realitzar una operació de realitat augmentada al "cloudlet", però 3.1 J, 5.1 J, i així successivament quan es realitza a AWS-East, AWS-West, i així successivament. Es poden esperar resultats similars amb qualsevol servei de desplaçament que es centri en uns pocs grans centres de dades.

- **Scalability through edge analytics** Independentment de les consideracions de latència, els "cloudlets" també poden reduir el consum d'ample de banda cap a la núvol. Per exemple, considerem una aplicació en què molts usuaris col·locats estan transmetent de manera contínua vídeo des del seu telèfon intel·ligent a la núvol per a l'anàlisi de contingut. La taxa acumulada de dades fins i tot per a una petita fracció d'usuaris en una ciutat de mida modesta saturaria la seva xarxa metropolitana: 12.000 usuaris transmetent vídeo en resolució 1080p requeririen un enllaç de 100 gigabits per segon; un milió d'usuaris requeririen un enllaç de 8,5 terabits per segon.

La Figura 2 mostra com els "cloudlets" poden resoldre aquest problema. En el marc proposat de GigaSight, el vídeo d'un dispositiu mòbil només viatja fins a un "cloudlet" proper. El "cloudlet" executa anàlisis de visió per ordinador en temps quasi real i només envia els resultats (etiquetes de contingut, reconeixement de cares, etc.) juntament amb metadades (propietari, ubicació de captura, marca de temps, etc.) a la núvol. Això pot reduir el consum d'ample de banda cap a la núvol en tres a sis ordres de magnitud. GigaSight també mostra com les etiquetes i les metadades a la núvol poden guiar cerques més profundes i personalitzades del contingut d'un segment de vídeo durant el seu període de retenció en un "cloudlet".

Una càmera de vídeo és només un exemple d'un sensor d'alta taxa de dades a l'IoT. Un altre exemple és un avió modern, que pot generar gairebé mitja terabyte de dades de sensors durant un vol. L'anàlisi en temps real d'aquestes dades en un "cloudlet" a l'avió podria generar orientació oportuna per a manteniment preventiu, estalvi de combustible i altres avantatges.

Les avantatges de latència i ample de banda dels "cloudlets" són especialment rellevants en el context dels automòbils, per complementar els enfocaments de vehicle a vehicle (V2V) que s'estan explorant per al control en temps real i l'evitació d'accidents. En un futur previsible, la connectivitat a la núvol des d'un automòbil en moviment serà com a màxim 3G o 4G/LTE. Una pregunta important és si els "cloudlets" haurien d'estar en els automòbils o formar part de la infraestructura de telecomunicacions (potser un "cloudlet" connectat a través de fibra òptica a diverses torres de telefonia en una àrea). Totes dues alternatives tenen valor.

- **Masking cloud outages** A mesura que augmenta la nostra dependència de la núvol, també augmenta la vulnerabilitat a les fallades de la mateixa. En aquest sentit, s'identifiquen diversos contextos d'ús, com ara teatres d'operacions militars, regions en fase de recuperació després de desastres o països amb infraestructures de xarxa febles, on l'accés a la núvol pot ser considerat un luxe ocasional en lloc d'una necessitat bàsica.

Els "cloudlets" es presenten com una solució per a aquestes situacions. Són dispositius de computació de proximitat que tenen la capacitat de proporcionar serveis crítics durant les interrupcions de la núvol. A causa de la seva proximitat física amb els dispositius mòbils associats, tenen característiques de supervivència més properes als usuaris que la núvol llunyana. Això permet que un "cloudlet" actuï com a intermediari i realitzi serveis crítics durant les fallades de la núvol. Quan es restableix la connexió amb la núvol, les accions realitzades al "cloudlet" es poden propagar a la núvol per a la conciliació.

Es fa referència a l'exemple del Sistema de Fitxers Coda, que va proporcionar un accés de lectura i escriptura desconnectable a les dades compartides fa més de dues dècades. Aquest sistema utilitzava l'emmagatzematge en memòria cau, l'emulació i la reintegració com a passos essencials per a l'ús de les dades de la núvol en absència de connexió.

- **The road ahead** L'informàtica perifèrica presenta avantatges i desafiaments tècnics i no tècnics importants. Des del punt de vista tècnic, cal abordar qüestions sobre els mecanismes i algoritmes de programari per al control col·lectiu dels "cloudlets" i la gestió de la infraestructura dispersa. També és necessari abordar la seguretat dels "cloudlets" i desenvolupar solucions per mitigar les seves limitacions en comparació amb els centres de dades en núvol.

Des del punt de vista no tècnic, el repte més important és establir models de negoci viables per a la implementació dels "cloudlets". És necessària la col·laboració entre indústries, comunitats i organitzacions de normalització

per aconseguir això. Sense aplicacions úniques que aprofitin la informàtica perifèrica, no hi ha incentius per a la implementació dels "cloudlets". Però sense un desplegament suficient dels "cloudlets", no hi ha motivació perquè els desenvolupadors creïn noves aplicacions i serveis.

Per superar aquesta situació, és necessari fomentar la creació d'un ecosistema obert per a la informàtica perifèrica. Això es pot aconseguir a través de plataformes com OpenStack++, que permetin la innovació en maquinari, programari i serveis, i que serveixin com a catalitzador per a la implementació de "cloudlets".

L'informàtica perifèrica coincideix amb altres tendències importants en el camp de la informàtica i les comunicacions, com ara la xarxa definida per programari (SDN), les xarxes sense fils d'ultra-baixa latència i les capacitats informàtiques dels dispositius mòbils. L'informàtica perifèrica és una aliada natural de les xarxes 5G, ja que assegura una latència molt baixa en la connexió inicial. A més, l'informàtica perifèrica pot amplificar les capacitats dels dispositius mòbils i els sensors a través de la infraestructura.

En resum, l'informàtica perifèrica representa una nova fase en l'evolució de la informàtica, amb avantatges i desafiaments importants. Per superar els obstacles, és necessari desenvolupar solucions tècniques innovadores, establir models de negoci viables i fomentar la col·laboració entre diferents actors de la indústria i les comunitats.

3 Assessment

3.1 *Rate the readability of the document: easy, readable, difficult, unreadable.*

Ha estat un article fàcil de llegir.

3.2 *Give your opinion of the reading assignment, indicating whether it should be included in next year's course or not.*

La temàtica està bé, encara que ja sabia alguns conceptes degut a altres assignatures.