Edge Computing tra innovazione e sicurezza

Author : Andrea Boggio

Date : 21 Ottobre 2019

> "*Tendiamo a sovrastimare gli effetti della tecnologia nel breve periodo e a sottostimarli sul lungo periodo*".
> Roy Charles Amara

La fiera del 5G

Nella realtà globale in cui viviamo, dominata dalle tecnologie dell'informazione e della comunicazione e caratterizzata da dinamiche inarrestabili di progressiva digitalizzazione e *divenire-software* del mondo[1], il consumo e la produzione di dati hanno raggiunto dimensioni ragguardevoli. Si tratta di un processo in forte crescita: alcune stime suggeriscono che entro il 2023 l'utente medio disporrà di una media di 6-8 dispositivi personali. Questi ultimi ammonteranno a oltre 30 miliardi e genereranno circa 100 *Exabyte* al mese di traffico (1 *Exabyte* equivale a 100,000,000,000,000 *Megabyte*)[2].

Come noto i dati sono il "*nuovo petrolio*" (secondo l'espressione attribuita nel 2006 al matematico e *data scientist* Clive Humby, noto anche per aver architettato la carta fedeltà della Tesco e le modalità di analisi dei dati generati da consumatori) e una delle sfide del prossimo futuro verterà su come estrarne effettivamente valore.

La frontiera dell'innovazione tecnologica rappresenta l'elemento fondamentale per abilitare il disegno di nuovi scenari sociali, culturali e di business. Nella grande narrazione collettiva degli ultimi anni il protagonista assoluto della scena – quella che Clayton M. Christensen chiamerebbe *disruptive technology***[3]** – è un'entità generale che per comodità e facilità di comunicazione viene divulgata a tutti i livelli con il nome **5G**, che a tutti gli effetti si configura come un'iperonimia che include un'intera famiglia di tecnologie eterogenee.

Dal punto di vista delle caratteristiche tecniche il 5G si fonda su alcuni elementi precisi:

- **velocità maggiori**, fino a 100 volte superiori rispetto alle attuali. La velocità possibile

sarà pari a 10 Gigabit per secondo e potremo scaricare dalla rete un film HD in meno di 10 secondi (oggi sono necessari 10 minuti);
- **latenze minori**: riscontreremo molto meno ritardo nelle comunicazioni e *lag* nell'utilizzo dei telefoni e dei dispositivi connessi. In una rete 4G la latenza si attesta tipicamente intorno a 40-50 millisecondi, ma nelle reti 5G sarà inferiore a 1 millisecondo, rendendola di fatto impercettibile per l'utente finale;
- **capacità maggiore**: le reti saranno in grado di supportare numerose applicazioni *high demanding* oggi non realizzabili: dalle automobili connesse ai dispositivi dell'*Internet of Things* (*IoT*), dalle esperienze di realtà virtuale allo *streaming* video HD;
- **affidabilità**: il 5G sarà "*ultra-reliable*", il che significa che potranno essere realizzate applicazioni per implementare casi d'uso critici (medicina digitale, controllo dei droni, veicoli a guida autonoma);
- **flessibilità**: il *Network Slicing* permette di dividere una rete fisica in reti virtuali multiple, in modo tale che l'operatore possa decidere quale fetta (*slice*) di rete utilizzare sulla base di requisiti eterogenei e dipendenti, ogni volta, dagli specifici casi di utilizzo.

Lo sviluppo del 5G promette possibilità entusiasmanti ma non ci troviamo dinnanzi a una trasformazione che avverrà dal giorno alla notte: si tratta di un *viaggio*. Così come il 4G è stato caratterizzato dall'utilizzo crescente degli *smartphone*, ci si aspetta che il 5G veda fiorire l'*Internet of Things*. Le caratteristiche di bassa latenza alta affidabilità delle operazioni tipiche del 5G aprono le porte a nuove applicazioni, quali:

- veicoli a guida autonoma, in grado di assolvere a tutte le funzioni critiche della guida e capaci di controllare in tempo reale le condizioni della strada;
- medicina digitale, con la possibilità di operare i pazienti da qualsiasi parte nel mondo tramite chirurgia robotica controllata remotamente, oppure di monitorare i pazienti per condizioni critiche quali epilessia, ictus e infarti;
- *smart city* e sicurezza, con la possibilità di potenziare la gestione del traffico tramite l'utilizzo di sensori di congestione combinati con impianti di illuminazione e segnali stradali intelligenti per semplificare i flussi. In scenari di emergenza – o per ragioni di ispezione e mappatura del territorio – i droni potranno raggiungere velocemente luoghi inaccessibili agli esseri umani e procedere alla cattura di informazioni preziose;
- fabbriche interconnesse, in un comparto industriale che sta già esplorando processi di manutenzione predittiva[4] e di controllo adattivo[5]. Il 5G ha la capacità di trasformare queste possibilità in realtà produttiva, superando alcune limitazioni tecniche, economiche o di compatibilità ambientale dei tradizionali meccanismi di connettività fissi e mobili. In una fabbrica, ad esempio, le caratteristiche proprie del 5G permetteranno a macchine, strumenti, parti e persone di essere perfettamente in sincronia innalzando la produttività e facilitando la personalizzazione di massa.

Ulteriori applicazioni appaiono imprevedibili e limitate unicamente dalla fantasia umana.

Nuove capacità della rete per nuove applicazioni

Le dimensioni tecnologiche indirizzate dal 5G sono essenzialmente tre:

1. la capacità, caratterizzata da prestazioni estremamente elevate in termini di velocità, trasferimento dati e densità;
2. il *Network Slicing*, caratterizzato dalla possibilità di “affettare” dinamicamente porzioni di rete per dedicarle a specifici servizi ritagliati sulle necessità reali degli utenti finali;
3. Il *Mobile Edge Computing (MEC),* caratterizzato da latenze minime, da aree avanzate di applicazione tecnologica (robotica, realtà virtuale, *smart city*) e dal progressivo spostamento delle applicazioni all’interno dell’infrastruttura di rete.

La dimensione di maggior interesse dal punto di vista della sicurezza è la terza: il MEC sposta il *data center* dal centro verso i bordi periferici (*edge*) della rete, garantendo la prossimità fisica delle risorse computazionali necessarie a ridurre i tempi di latenza della comunicazione.

Come descritto in precedenza, la velocità massima e l’affidabilità delle reti aumenteranno, i tempi di latenza diminuiranno e per un enorme numero di dispositivi sarà possibile connettersi alle celle della rete mobile. Le applicazioni e i casi d’uso del 5G si sposano perfettamente con alcune caratteristiche del *cloud computing* (supporto della connettività ubiqua, elasticità, scalabilità delle risorse e facilità di *deployment*) rendendo questi ambienti di calcolo estremamente interessanti. Il trend emergente dei *deployment* IoT sta però introducendo nuovi requisiti (distribuzione geografica, bassa latenza, conoscenza della posizione e supporto della mobilità) che gli attuali ambienti *cloud* non possono soddisfare adeguatamente.

Per indirizzare i nuovi e sfidanti requisiti introdotti dalle applicazioni e dai casi d’uso 5G la comunità di ricerca ha proposto tecnologie che possiamo concepire come un *cloud esteso*: l’obiettivo è consentire alle necessità di calcolo di essere soddisfatte il più vicino possibile alla sorgente dei dati. Questo scenario aumenterà la qualità dei servizi offerti poiché ridurrà drasticamente i tempi di scambio delle informazioni tra i nodi finali della rete e il *cloud*.

Il cosiddetto *edge computing* si riferisce proprio a questo: i dati sono processati sull’*edge* di una rete. Per esempio, i sensori delle applicazioni industriali IoT possono catturare flussi di dati per ottimizzare i processi di produzione connettendosi ai *Programmable Automation Controller* (*PAC*) che gestiscono l’elaborazione e la comunicazione. Nel caso dell’*edge computing*, le capacità di *cloud computing* sono fornite nell’ambito della *Radio Access Network* (*RAN*): in questo modo è possibile offrire applicazioni consapevoli del contesto.

Storicamente i principali *cloud service provider* hanno costruito un numero limitato di *data center* colossali dislocati in varie parti del pianeta dotati di risorse di calcolo tali da poter soddisfare un numero elevatissimo di utenti. Questa centralizzazione delle risorse implica un’importante separazione tra i dispositivi degli utenti e i loro *cloud*, cosa che determina, in media, un aumento dei tempi di latenza e di *jitter*[6]. A causa della distanza fisica i servizi *cloud* non sono in grado di accedere direttamente a informazioni locali di contesto quali, ad esempio, la posizione precisa dell’utente, le condizioni di rete locali o altre informazioni relative al comportamento in mobilità degli utenti.

Per queste ragioni negli anni recenti sono emersi vari paradigmi, quali il *Fog Computing*, il *Mobile Edge computing* e il *Mobile cloud computing*. Il denominatore comune di questi paradigmi *edge* è il dispiegamento (*deployment*) delle capacità del *cloud computing* sul bordo

della rete. I *data center edge* implementano un'infrastruttura tecnologica virtualizzata *multi-tenant*[7]. Ogni Cliente –fornitori di servizi terzi, utenti finali, fornitori di infrastruttura stessi – può utilizzare i servizi di questi *data center*, che possono agire in autonomia o cooperare tra loro. È quindi possibile creare un'architettura gerarchica a più livelli interconnessa tramite infrastruttura di rete sottostante (*core network*) che fornisce vari meccanismi di supporto quali piattaforme di gestione e servizi di registrazione degli utenti. Un dominio di *trust* (ad esempio un'infrastruttura *edge* posseduta e gestita da un operatore) può cooperare con altri domini di *trust*, creando un ecosistema aperto che ha la capacità di servire moltitudini di utenti.

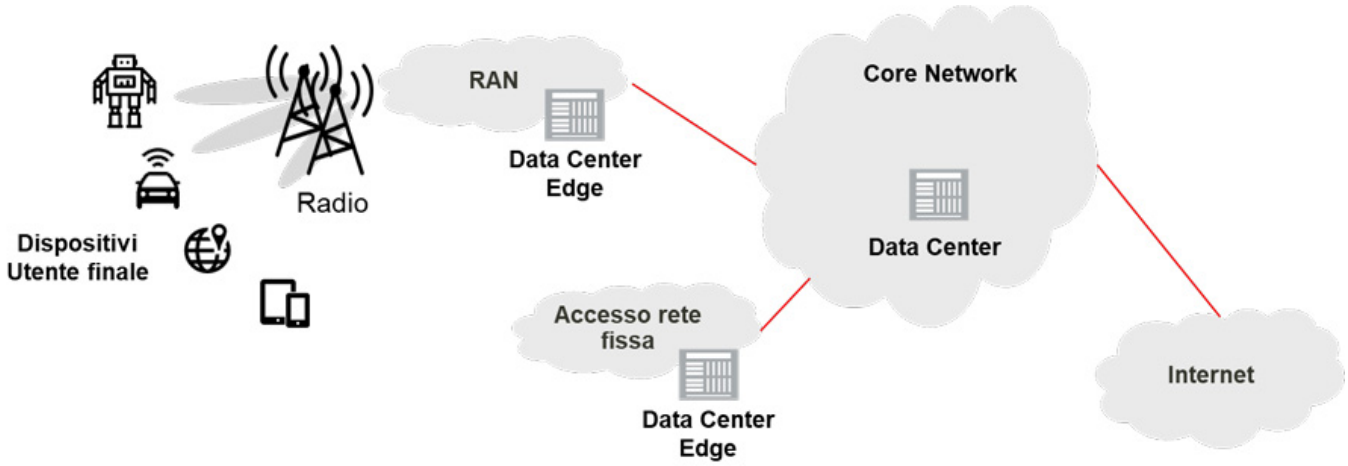

Figura 1 - Esempio di architettura Mobile Edge Computing

Paradigmi Edge

FOG Computing

Il concetto di *Fog Computing* è stato introdotto da Cisco nel 2012[8] e nella sua definizione iniziale era considerato una "estensione del paradigma del *cloud computing* (che) fornisce servizi di calcolo, rete e *storage* tra dispositivi finali e *server cloud* tradizionali". Il *Fog Computing* non cannibalizza il *cloud computing*, ma ne rappresenta il complemento: l'architettura *fog* facilita la creazione di infrastrutture gerarchiche laddove l'analisi dell'informazione locale è effettuata "sul campo" e il coordinamento e l'analisi globali sono effettuate nel *cloud*. Il *deployment* dei servizi *cloud* avviene sull'*edge* della rete, ma può avvenire anche altrove, ad esempio nel *backbone MPLS*[9]. L'infrastruttura di rete *fog* è eterogenea e in essa convivono collegamenti ad alta velocità e tecnologie *wireless*. Il modello di *Fog Computing* è stata concepito come estensione del *cloud* ed è caratterizzato dall'utilizzo di piattaforme altamente virtualizzate che possono soddisfare i requisiti dell'IoT. I dispositivi IoT consistono di sensori che inviano dati ai nodi *fog*, i quali possono ricevere *feed* in tempo reale, eseguire applicazioni per effettuare operazioni sui dati collezionati (*analytics*) e fornire conservazione di breve periodo degli stessi. Inoltre i nodi *fog* possono inviare informazioni aggregate al *cloud*, solitamente a intervalli predefiniti. Il *cloud* aggrega e colleziona i dati inviati dai nodi *fog* e li processa, per esempio

effettuando analisi a supporto dei processi decisionali. Il *cloud* può anche inviare direttamente nuove direttive ai nodi *fog* sulla base dei risultati dell'analisi dei dati per soddisfare specifici requisiti operativi.

MEC

Il *Mobile Edge Computing (MEC)* è stato inizialmente definito da IBM e Nokia Siemens[10] per descrivere le strutture di calcolo interne alla stazione mobile.
L'European Telecommunications Standards Institute (ETSI) ha lanciato un *Industry Specification Group (ISG)* alla fine del 2014 nello sforzo di giungere alla standardizzazione e sono stati prodotti documenti che forniscono una serie di specifiche sia per il *framework* sia per le architetture MEC. ETSI ha successivamente modificato l'acronimo MEC in *Multi-access Edge Computing* definendolo[11] come "capace di fornire […] un ambiente di servizio IT e capacità di *cloud computing* sull'*edge* della rete mobile per ridurre i tempi di latenza, assicurare operazioni di rete e consegna del servizio altamente efficienti offrendo una *User Experience* migliore". Il MEC offre sia un nuovo ecosistema sia una nuova catena del valore: gli operatori possono aprire l'*edge* delle proprie RAN all'utilizzo da parte di terze parti autorizzate, consentendo loro lo sviluppo rapido e flessibile di applicazioni innovative e servizi diretti a cittadini, imprese e segmenti verticali dell'industria. Tra i casi d'uso più importanti ETSI ricorda i video *analytics*, l'IoT, la realtà aumentata, il *caching* dei dati, i servizi di posizione e la distribuzione ottimizzata dei contenuti.

MCC

Il *Mobile Cloud Computing (MCC)* si focalizza principalmente sulla nozione di delega mobile: a causa delle risorse limitate dei dispositivi mobili, questi ultimi dovrebbero delegare lo *storage* di grandi volumi di dati e l'esecuzione di *task* di calcolo particolarmente onerosi a entità remote. Nel concetto originale del 2009[12] solamente le piattaforme *cloud* centralizzate erano considerate la soluzione sostenibile per implementare l'esecuzione remota di *task*. Successivamente, altri ricercatori hanno espanso l'ambito del MCC e in questa nuova visione i *task* possono essere delegati ai dispositivi localizzati sull'*edge* della rete. Al momento coesistono entrambe le visioni.

Caratteristica	MEC	Fog Computing	MCC	Cloud
Proprietà	Compagnie telefoniche	Entità private, individui		Entità private
Deployment	Edge	Vicino all'edge, edge	Edge, dispositivi	Core network
Hardware	Server eterogenei	Server, dispositivi utente		Server
Servizi	Virtualizzazione		Virtualizzazione, altri	Virtualizzazione
Architettura di rete	N-tier[13], decentralizzata, distribuita			Centralizzata
Mobilità	Sì			Non disponibile
Latenza e Jitter	Bassa			Nella media

Consapevolezza del contesto	Sì	Non disponibile
Disponibilità	Alta	
Scalabilità	Alta	Nella media

Tabella 1 – Confronto delle caratteristiche dei paradigmi edge[14]

I diversi paradigmi *edge* hanno tutti lo stesso obiettivo di base: portare capacità di calcolo *cloud* sull'*edge* della rete. Tutti forniscono supporto a qualche tipo di infrastruttura di virtualizzazione *multi-tenant* (nodo *fog*, server MEC, *cloudlet*) facilmente accessibile attraverso varie reti a banda larga (fibra ottica, comunicazioni *wireless*, reti mobile ad alta velocità). Queste infrastrutture possono aggiustare il rilascio delle capacità in base al luogo e alle necessità degli utenti e accedere, se necessario, a risorse computazionali.

Anche se tutti i paradigmi hanno gli stessi obiettivi esistono differenze relative al modo in cui questi vogliono essere raggiunti. Per esempio, MEC limita il *deployment* delle piattaforme di *edge computing* alle infrastrutture di rete mobile, quali quelle del 5G. I nodi *fog* possono dispiegati anche in altri contesti, quali quelli dei *server* gestiti dagli utenti, gli *access point*, i *router*, i *gateway*. Questa differenza di *deployment* e gestione dei *data center edge* influenza la scelta di chi può diventare fornitore di servizio. Per esempio, nel paradigma MEC solamente gli operatori di telecomunicazione possono diventare fornitori perchè detengono l'infrastruttura di rete mobile che ospita i *data center edge*. Di contro, ogni utente può fare il *deployment* dei propri nodi *fog* e MCC diventando effettivamente parte dell'ecosistema di fornitura del servizio. Un'altra differenza collegata al punto precedente è il *deployment* di applicazioni *legacy*. Dato che i server MEC sono controllati dagli operatori di telecomunicazione e ospitati nella loro infrastruttura, è possibile per fornitori di servizi terzi lavorare in stretto contatto con gli operatori e sviluppare specifici servizi che possono essere testati ampiamente e possibilmente integrati in modi personalizzati.

Panorama delle minacce

La maggior parte dei paradigmi *edge* include tecnologie abilitanti quali reti *wireless*, sistemi distribuiti, *peer-to-peer* e piattaforme di virtualizzazione. È quindi necessario non solo proteggere tutti questi blocchi ma anche orchestrarli: garantire la sicurezza di tutte le parti non garantisce la sicurezza dell'intero sistema. Una volta che le capacità del *cloud computing* sono portate sull'*edge* della rete emergono situazioni la cui sicurezza non è stata ancora ben studiata: a causa delle specifiche caratteristiche l'intero sistema eredita *anche* le minacce di sicurezza negli scenari applicativi. L'impatto che un attacco andato a buon fine potrebbe causare all'intera società è considerevole proprio perché il numero di scenari applicativi dei paradigmi *edge* è enorme.

Praticamente ogni aspetto della nostra vita quotidiana può essere influenzato dalle applicazioni residenti su queste infrastrutture: le nostre informazioni private (foto, referti medici), le nostre abitudini quotidiane (trasporto, shopping), i nostri ecosistemi di impresa (industrie, *supply chain*), le nostre infrastrutture critiche (energia, sistemi di emergenza). L'estrema convergenza tra tecnologie digitali, mondo fisico e forti interazioni sociali sposta la tradizionale sicurezza

informatica nel campo della *cyber security*, accentuandone il carattere pervasivo.

Uno dei principali problemi è rappresentato dalla **mancanza di un perimetro globale di riferimento**. I *data center edge* sono in grado di fornire servizi senza dipendere continuamente da un'infrastruttura centrale, quindi tutti gli *asset* rilevanti - incluse l'infrastruttura di rete, di servizio (*data center edge, core network*), di virtualizzazione e i dispositivi degli utenti - sono controllati da attori diversi che devono cooperare tra loro. La conseguenza immediata è che ogni elemento dell'infrastruttura può essere oggetto di attacco in ogni momento: il principio dell'"*anything, anytime*" è ereditato dai blocchi elementari dell'architettura e dagli scenari applicativi.

Un *data center edge* (nodo *fog*, server MEC) fornirà servizi verso le entità geograficamente vicine: se questo significa limitare la superficie d'attacco ad un ambiente locale, d'altro canto un attaccante che prenda il controllo di un *data center edge* potrebbe governare tutti i servizi che sono erogati in quella specifica regione geografica.

Un'altra conseguenza della mancanza di perimetro globale è la natura dei diversi profili di attacco che prenderanno di mira i paradigmi *edge*. Anche se esisteranno sempre i tradizionali attaccanti esterni (che non controllano elementi dell'infrastruttura), esisteranno anche molti avversari che controlleranno uno o più elementi dell'infrastruttura: dispositivi utente, macchine virtuali, *server*, porzioni di rete, anche tutto il *data center edge*. Questa tipologia di avversario è sia interno sia esterno, perché controlla alcune parti dell'infrastruttura e non altre.

Asset	**Minacce**
Infrastruttura di rete	Denial of Service, Man-In-The-Middle, rogue gateway
Data Center Edge	Danno fisico, perdita della privacy, privilege escalation, manipolazione del servizio, rogue data center
Infrastrutture core	Perdita della privacy, manipolazione del servizio, infrastruttura rogue
Infrastrutture di virtualizzazione	Denial of Service, abuso delle risorse, perdita della privacy, privilege escalation, manipolazione delle Macchine Virtuali
Dispositivi utente	Injection di informazioni, manipolazione del servizio

Tabella 2 - Categorizzazione delle minacce nei paradigmi di edge computing[15]

Conclusioni

L'*hype* e l'eccitazione associate al 5G riflettono sia i benefici introdotti dalle nuove tecnologie sia i volumi di business attesi. Il *battage* pubblicitario è perfettamente giustificato da numeri che promettono di creare ricchezza e prosperità per tanti comparti industriali, senza dimenticare che sembriamo finalmente in procinto di concretizzare visioni ipertecnologiche degne della migliore letteratura *cyberpunk*. Tutto questo **sa di magia**.

Dal punto di vista della *cyber security,* la superficie complessiva d'attacco tende a espandersi in domini ben precisi: gestione delle identità digitali, dei processi di autenticazione e dei sistemi di controllo degli accessi, sicurezza delle infrastrutture di rete e dei protocolli utilizzati, gestione dei sistemi di virtualizzazione, del *trust* e della privacy. Il tutto associato alla storica e generalizzata difficoltà nel garantire un adeguato livello di sicurezza del software.

L'intersezione tra modelli di calcolo e attori diversi – necessaria per la realizzazione delle promesse innovative del 5G – è tipica dei paradigmi dell'*edge computing* e il risultato più evidente è l'aumento esponenziale dei punti di contatto tra entità diverse. Tra i vari paradigmi analizzati il MEC sembra essere quello più chiuso e, teoricamente, il meno insicuro (tutti gli elementi che concorrono all'erogazione del servizio sono gestiti e di proprietà di un'unica entità: l'operatore di telecomunicazione). Molti oggetti che faranno parte dell'ecosistema *edge*, inoltre, non hanno neanche meccanismi di protezione nativi.

Il tema merita la massima attenzione perché, se è vero che nel 2018 Gartner[16] scommetteva sul trend "*Cloud to the edge*" (entro il 2022 la metà delle grandi organizzazioni userà la tecnologia *edge*), nei fatti la sicurezza dell'*edge computing* vive ancora la propria infanzia.

Note

[1] https://www.afcearoma.it/images/icagenda/files/Presentazioni_Eventi/Anno2019/19_giugno_2019/Convegno_AFCEA_Andrea_Boggio.pdf

[2] https://www.ericsson.com/en/press-releases/2017/11/ericsson-predicts-1-billion-5g-subscriptions-in-2023

[3] https://hbr.org/2015/12/what-is-disruptive-innovation

[4] https://it.wikipedia.org/wiki/Manutenzione_predittiva

[5] https://it.wikipedia.org/wiki/Controllo_adattativo

[6] https://it.wikipedia.org/wiki/Jitter

[7] https://it.wikipedia.org/wiki/Multi-tenant

[8] https://www.cisco.com/c/en/us/solutions/enterprise-networks/edge-computing.html

[9] https://it.wikipedia.org/wiki/Multiprotocol_Label_Switching

[10] https://www-03.ibm.com/press/us/en/pressrelease/40490.wss

[11] https://www.etsi.org/technologies/multi-access-edge-computing

[12] Cloud Computing, Vol.5931 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2009

[13] https://it.wikipedia.org/wiki/Architettura_multi-tier

[14] https://www.sciencedirect.com/science/article/pii/S0167739X16305635

[15] *Ibid.*

[16] https://www.gartner.com/smarterwithgartner/gartner-top-10-strategic-technology-trends-for-2018/

Articolo a cura di **Andrea Boggio**