

## MSA - 31.03.2021

Abbiamo presentato lo standard Mobile IP, lo abbiamo esaminato nella versione base e dopo nei suoi raffinamenti che facevano riferimento alla versione 4 del protocollo IP, e poi nella seconda parte della lezione di ieri abbiamo visto come il protocollo si è evoluto rispetto alla versione 4. Abbiamo tirato le somme su come si collocavano le soluzioni all'interno della griglia generale vista in precedenza.

E' ora di tirare le somme da un altro punto di vista, guardando con un occhio un po più critico a questa soluzione. Aspetti problematici di queste soluzioni sono quelli che vedete evidenziati. Una cosa che risulta evidente è che essendo una soluzione host based come discusso ieri, questa soluzione per essere implementata richiede una modifica dello stack protocollare di ogni singolo nodo mobile. Non soltanto può risultare problematico a fronte di un parco già esistente di nodi mobili che non sono stati rilasciati nativamente con questa soluzione già preinstallata, ma da un punto di vista operativo significa che ogni nodo mobile si deve fare parte di una parte del lavoro necessario ad attivare le funzionalità previste dal protocollo. Altro problema può essere la latenza per completare un'operazione di hand-off tra una sotto rete e l'altra. Come conseguenza del fatto che sono i nodi mobili ad essere coinvolti direttamente nella realizzazione delle funzioni del protocollo, può avere come effetto collaterale quello di aumentare il tempo di hand-off. Queste sono alcune delle ragioni per cui, se uno va a guardare nella realtà, quello che uno osserva è che il successo di questo protocollo non è stato quello che i suoi ideatori speravano. In parte queste problematiche, ma anche per altre ragioni, non tecniche ma economico-politiche in un certo senso, che hanno contribuito al mancato successo di questa soluzione. Il successo di una tecnologia non dipende necessariamente dai suoi meriti intrinseci, ma può dipendere anche da fattori esterni, esogeni, che contribuiscono al suo successo o insuccesso. Al di là di queste considerazioni esterne, limitandosi al fatto tecnico, questi problemi evidenziati sono reali. Nell'ambito dell'IETF stesso si è cercato di trovare una soluzione alternativa.

La soluzione alternativa è un altro standard proposto e prende il nome di Proxy Mobile IPv6. I suoi obiettivi sono quelli mirati a superare gli aspetti problematici del Mobile IP esaminato: consentire di essere installata senza richiedere modifiche a nodi mobili preesistenti, uso più efficiente delle risorse wireless e le operazioni di hand-off. I vantaggi, avendo questi obiettivi, non richiede modifiche dunque più facile il deployment, dal punto di vista delle prestazioni riduce la quantità di messaggi che circolano nella rete essendo una soluzione network based, la riduzione del traffico di controllo causato dalla messa in atto di questa soluzione esclude categoricamente la necessità di creare tunnel che si estendono anche sulla parte wireless. C'è anche un altro vantaggio che la soluzione offre e saremo in grado di apprezzarlo quando parleremo dopo dell'architettura della rete cellulari, dal punto di vista dei provider che gestiscono le reti. La soluzione è molto più vicina alla soluzione adottata nelle reti cellulari. Questa variante dei protocolli previsti all'interno dell'IETF è poi quella che ha trovato la sua applicazione nel mondo delle reti cellulari.

Breve carrellata sulla terminologia. Entità che entrano in gioco. LMA è il punto di aggancio dei nodi mobili rispetto al resto della rete internet. E' un router il cui prefisso di rete

identifica la posizione occupata dai nodi mobili all'interno della rete internet. Ogni nodo mobile nel momento in cui si aggancia ad un dominio di rete il cui ingresso è gestito dall'LMA, riceve un indirizzo IP che appartiene topologicamente a quel dominio. A questa entità singola fanno capo più MAG, che sono le entità a cui fisicamente si aggancia un nodo mobile, ognuna di queste avrà un indirizzo IP specifico, e l'indirizzo IP che identifica questi MAG è l'indirizzo usato da LMA per inoltrare all'interno del suo dominio i messaggi indirizzati verso il nodo mobile. Ad ogni nodo mobile è assegnato un indirizzo IP fisso fin quando il nodo mobile sta all'interno del dominio di rete, LMA. Anche se cambia MAG, il suo indirizzo identificativo rimarrà costante. Informazione aggiornata dal LMA nella sua tabella. Ogni MAG agisce da proxy per ogni nodo mobile dal punto di vista della raggiungibilità. I tunnel vanno da LMA al MAG. Se vogliamo fare un parallelo con i discorsi fatti in precedenza, LMA svolge il ruolo analogo a quello dell'HA, ed i MAG un ruolo analogo a quello dei FA.

Quando un nodo mobile entra all'interno di un dominio di rete gestito dal protocollo ed intende agganciarsi dal punto di vista logico ed essere raggiungibile. Prima cosa il nodo mobile entra in contatto con il MAG che gestisce il pezzo di rete a cui si è connesso fisicamente. MN invia richiesta al MAG. Superata con successo questa fase, il nodo mobile chiede che gli venga associato un indirizzo IP che lo renda raggiungibile all'interno del dominio di rete. Il MAG lo inoltra all'LMA. A questo punto LMA assegna al nodo mobile un prefisso di rete che sia topologicamente corretto appartenente al dominio di rete dominato dall'entità. Aggiunge alla tabella l'informazione sul prefisso e l'informazione sul MAG a cui il nodo mobile è associato. L'informazione viene inoltrata al nodo mobile, che si auto configura, si dota di un suo indirizzo IP unico, e la cosa finisce qui, si completa l'aggancio del nodo mobile. A questo punto pacchetti che arriveranno ad LMA, verranno incapsulate usando le informazioni presenti nella tabella, quindi si costruirà un tunnel da LMA a MAG, il tunnel termina sul MAG, decapsula il pacchetto e lo invia al nodo mobile.

Se il nodo mobile cambia punto di aggancio, avviene un'operazione di distacco, il vecchio MAG rileva il fatto che il nodo mobile non è più agganciato a lui. Manda un messaggio di deregistrazione ad LMA che fa partire un timer, il tempo che viene concesso al nodo mobile per avviare una nuova procedura di aggancio ad un nuovo MAG. Scaduto il timer, LMA rimuove dalla sua tabella la riga relativa al nodo mobile in questione. Serve ad evitare informazioni di nodi mobili usciti completamente dal nodo di rete. Se il nodo mobile si sta agganciando ad un altro MAG, si ripete la procedura vista in precedenza.

Riassumendo, dal punto di vista del nodo mobile, tutto questo dominio di rete appare come una singola rete di livello 2, perché una volta che ha ricevuto un indirizzo IP che lo rende raggiungibile all'interno di questa rete, per quanto grandi gli spostamenti il suo indirizzo IP non cambia. Continua ad essere sempre connesso allo stesso dominio di rete. L'unica cosa che può cambiare è l'entità di livello 2 a cui si sta agganciando. Per un nodo corrispondente che vuole parlare con il nodo mobile, i pacchetti andranno all'LMA, da qui viaggeranno sul tunnel verso il MAG e poi, tolto l'header che realizza il tunnel, vengono indirizzati al nodo mobile. Il problema della triangolazione che nel caso di Mobile IP veniva risolto nel modo che abbiamo visto, consentendo al nodo corrispondente di indirizzare i messaggi direttamente verso il nodo mobile, in questo caso questa soluzione non è più possibile. La scelta di tenere fuori il nodo mobile dalla gestione della sua raggiungibilità ha come lato

negativo il fatto che non è più possibile applicare quella soluzione. L'instradamento segue comunque un percorso triangolare. L'unica ottimizzazione concessa è se due partner che interagiscono tra di loro fanno capo allo stesso MAG, non è necessario risalire fino allo stesso LMA.

Se vogliamo classificare questa soluzione, dal punto di vista del tracciamento, lungo questa dimensione, è una soluzione puramente pro attiva. Dal punto di vista del livello a cui viene applicato è una soluzione che si colloca a livello 3 dello stack mobile. Riguardo alla portata scendiamo di un livello. Mobile IP ha un'ambizione globale, ovunque si sposti il nodo all'interno dell'architettura complessiva di internet, quella soluzione consente di raggiungerlo. Qui invece questa soluzione riesce a gestire soltanto movimenti che si svolgono all'interno al dominio dominato da un LMA. Dal punto di vista delle entità coinvolte è una soluzione puramente network based.

Qui per finire, le slide successive riproducono un semplice modello analitico che confronta varie soluzioni. Non ve lo illustro nel dettaglio, lo riporto soltanto per ribadire quello detto già altre volte. L'importanza di fare valutazioni quantitative quando abbiamo soluzioni alternative per lo stesso problema. A volte non serve costruirsi strumenti e tecniche di valutazione complicate per ottenere risposte significative ai dubbi e le domande che una persona si può porre. Modello analitico semplice. Descrive i messaggi scambiati ed il tempo necessario per il loro inoltro. Se guardiamo i risultati dei vari esperimenti analitici, abbiamo l'impatto sul tempo di hand-off a fronte di vari valori del ritardo causato dal canale wireless che connette il nodo mobile al resto del mondo. La linea corrispondente alla latenza dell'hand-off nel caso di Proxy mobile IP è nettamente inferiore rispetto agli altri protocolli. Questo vale anche il ritardo complessivo nelle comunicazioni tra il nodo mobile ed il suo partner. Anche l'impatto del ritardo con cui viene rilevato un cambio di posizione, proxy Mobile IP risulta vincente rispetto agli altri protocolli.

---

## Mobility management in wireless WAN

Vogliamo esplorare come il problema della mobilità viene affrontato e risolto in un'architettura diversa, quella delle reti cellulari. Rimaniamo nell'ambito della mobilità, come si può raggiungere un nodo la cui posizione non è fissa all'interno dell'infrastruttura. Parleremo di standard in uso, e questa volta il mondo in cui queste soluzioni sono nate ma è il mondo dei grandi provider che hanno messo in piedi le infrastrutture note come reti cellulari.

Abbiamo già fatto conoscenza con questo tipo di infrastruttura. Sappiamo che a partire dalla generazione 2 fino alla 3, questa è l'architettura gerarchica di riferimento. Alla radice ci sono le MSC, a cui fanno capo entità di livello più basso BSC, a cui fanno capo entità di livello ancora più basso i BTS. Spostarsi, muoversi, all'interno di una rete di un'infrastruttura realizzata in questo modo significa passare da una cella ad una regione adiacente coperta da un'antenna diversa. Come abbiamo visto, queste transizioni a seconda di come si collocano queste celle dal punto di vista dell'infrastruttura gerarchica che sta dietro, gli hand-off possono essere di vari tipi. Di livello 1 all'interno dello stesso BTS. Di

livello 2 all'interno dello stesso BSC. Di tipo 3 all'interno dello stesso MSC. Di livello 4 all'interno di MSC diversi. Di hand-off di tipo 3 c'è un'ulteriore articolazione che fa sì che ne possiamo distinguere due tipi diversi. Passando alla quarta generazione l'architettura cambia in quanto si appiattisce rimuovendo un livello, quello formato dalle antenne che realizzano il canale wireless e poi una struttura piatta retrostante che connette le varie antenne tra loro e rende possibile ad un'entità esterna la possibilità di parlare con un nodo mobile che si trova connesso ad un'antenna.

Le entità che entrano in gioco sono 2. MME ed SGW, che da un punto di vista di strutturazione gerarchica stanno allo stesso livello, presenti nella parte core della rete. Sono entità indipendenti tra di loro, soprattutto rispetto al rapporto che hanno con le antenne sottostanti. Come la figura mostra, due singole antenne posso fare capo allo stesso MME e/o stesso SGW. Prese due antenne e due celle adiacenti ci sono quattro configurazioni possibili a seconda se fanno capo allo stesso MME ed allo stesso SGW, con quattro hand-off possibili. MME si preoccupa di mantenere aggiornate le informazioni sulla posizione quando il nodo è in stato idle, SGW entra in gioco primariamente quando il nodo mobile cambia posizione ed è impegnato in un'interazione con il suo partner.

Ritorniamo indietro. Questa è una figura che riproduce l'architettura di una rete di seconda generazione, primariamente pensate per supportare traffico di tipo vocale. L'infrastruttura per la gestione della mobilità è formata da MSC, collegati da una rete su cui circolano messaggi di controllo, entità attivamente coinvolte. Di queste entità ne abbiamo una che svolge il ruolo privilegiato che è il punto di contatto dell'architettura con il mondo esterno, in questo caso è principalmente una rete telefonica pubblica. Oltre queste entità attive l'infrastruttura utilizza una struttura di memorizzazione per mantenere l'informazione sulla posizione dei nodi che è un database a due livelli. La radice della gerarchia è il database HLR, a cui fanno capo i VLR. Ogni VLR è associato in rapporto 1-1 con un singolo MSC. Il modo in cui questo database a due livelli viene utilizzato lo vedremo adesso.

Passando all'architettura dell'evoluzione della generazione successiva, la configurazione generale non cambia, abbiamo ancora i BTS e i BSC, che fanno capo al livello più alto della gerarchia. Le entità che entra no in gioco cambiano nome, MSC diventa SGSN. Quello che era l'MSC primario diventa GGSN connesso direttamente alla rete internet. Di SGSN ce ne saranno tanti in generale, e di GGSN uno solo invece, singolo punto di accesso verso la rete internet globale. Il colloquio tra SGSN e GGSN avviene tramite un backbone che utilizza il protocollo IP. Se ci pensiamo un attimo, questa organizzazione se andiamo a discorsi appena conclusi, comincia ad essere molto sovrapponibile con l'architettura del Proxy Mobile IP che abbiamo descritto. GGSN sovrapponibile ad LMA e SGSN sovrapponibile a MAG.

Nelle reti di quarta generazione l'architettura cambia, per cui le singole antenne si connettono direttamente al core. Ogni antenna si connette ad un MME e ad un SGW. Di SGW ce ne possono essere tanti. C'è una singola entità, il PGW che svolge un ruolo analogo a quello di GGSN. L'SGW svolge un ruolo analogo agli SGN.

Vediamo come vengono svolte le funzioni di mobilità management. I requisiti sono abbastanza ovvi. Come si realizza il problema del reperire un nodo mobile all'interno dell'architettura. Viene risolto il problema tramite due funzioni: quella di tracciamento per

tenere aggiornate le informazioni sulla posizione con un livello di precisione più o meno spinto, e l'altra è quella di localizzazione che viene attivata se l'informazione fornita dal tracciamento non è precisa, localizzare con massima precisione il nodo mobile.

Partiamo dalla generazione 2. A questa organizzazione fisica gerarchica dobbiamo sovrapporre un'ulteriore organizzazione logica, quella della ripartizione della totalità delle celle che si estendono sul territorio coperto dall'infrastruttura, le celle vengono ripartite tra regioni formate da celle contigue tra di loro, Location Area. Il partizionamento avviene da un sottoinsieme delle celle che sono sotto il controllo di un singolo MSC. Per effetto di questo partizionamento delle celle, abbiamo nella figura tre location area, due che fanno capo ad uno stesso MSC, una che fa capo ad un MSC diverso. Se vediamo i tipi di hand-off di tipo 2, corrispondono a passaggi da una cella ad una cella adiacente governata da un'antenna diversa entrambe che fanno capo allo stesso MSC. La presenza di partizione della Location Area crea i due sottocasi distinti negli hand-off di tipo 3, spostamenti tra celle adiacenti che fanno capo a due BSC diversi ma stesso MSC. Hand-off di tipo 3.1 è se due celle fanno capo a BSC diversi ma stessa LA. Hand-off di tipo 3.2 se le LA sono differenti tra due celle adiacenti. Qual'è l'informazione mantenuta nel database? Nel VLR, associato ad ogni MSC, viene mantenuta l'informazione per ogni nodo mobile un record che mantiene l'identità del nodo mobile e qual'è la LA in cui si trova attualmente il nodo mobile. L'informazione che viene mantenuta è un'informazione sull'informazione approssimativa di un nodo mobile. Se LA è formata da più celle adiacenti, sapere la LA in cui si trova il nodo mobile mi dà un'informazione approssimata sul nodo mobile. Nel livello superiore della gerarchia si mantiene un alto record che ha come chiave di accesso l'identificativo del nodo mobile, ma come informazione associata ha quale VLR in cui è attualmente presente questo record.

Algoritmo di tracciamento. Viene attivato nel momento in cui un nodo mobile si accende e si connette ad una certa rete cellulare. Come prima cosa connettendosi alla rete, il nodo mobile si conatterà ad un certo access point, e connettendosi ad una certa antenna di conseguenza si determinerà all'interno del sistema la LA in cui il nodo mobile si trova attualmente, grazie ad un segnale di beacon emesso periodicamente dalle antenne. Quest'informazione a questo punto viene mandata al VLR di riferimento. Il VLR che è associato all'MSC a cui fa capo la stazione a cui il nodo mobile si è connesso, in questo VLR viene registrata l'identità del nodo mobile e la LA in cui si trova attualmente. Questa informazione il VLR la inoltra all'HLR, al livello superiore. Finché il nodo mobile rimane attivo, se il nodo mobile cambia LA, quello che succede dipende dal tipo di hand-off. Se è di tipo 1, di tipo 2, o di tipo 3.1, quindi non cambia LA non succede nulla. Se invece il cambio di cella ha causato anche un cambio di LA, hand-off di tipo 3.2 o di tipo 4, allora sicuramente va aggiornata l'informazione che è presente nel VLR, quindi la LA del nodo mobile. Il punto è quale VLR. Potrebbe essere lo stesso in cui era già registrata l'informazione precedente, nel caso di hand-off 3.2, allora questa è l'unica cosa da fare. L'aggiornamento del record posseduto da un VLR non coinvolge il livello superiore della gerarchia. Se però quest'informazione non deve essere più registrata nel vecchio VLR, ma da un nuovo VLR perché ho cambiato LA e MSC, hand-off di tipo 4, allora l'informazione inviata al nuovo VLR deve essere propagata al livello superiore della gerarchia all'HLR, e la rimozione dal vecchio HLR del record relativo al nodo mobile. Tutto questo finché il nodo

mobile è acceso. Nel momento in cui il nodo mobile è spento, nell'ultimo VLR viene registrato il fatto che il nodo mobile non è raggiungibile. Il vantaggio della riduzione del traffico viene pagato dalla mancanza di informazioni sulla precisione sulla posizione del nodo mobile.

Nel momento in cui qualcuno vuole contattare il nodo mobile, è necessario un algoritmo di reperimento del nodo mobile stesso. La richiesta si origina da un'entità esterna. Questa richiesta arriverà ad un certo MSC che sarà il punto di ingresso della rete. Questo MSC manda una query al livello superiore HLR per sapere qual'è il VLR che possiede le informazioni sulla posizione occupata dal nodo mobile, usando come chiave di accesso l'identità del nodo mobile. HLR riceve la richiesta ed inoltra la richiesta di localizzazione del nodo mobile al VLR di riferimento ed a questo punto il VLR di riferimento trova al suo interno l'informazione sulla LA, l'insieme di celle, in cui il nodo mobile si deve trovare. A questo punto l'MSC associato al VLR invia in parallelo a tutte le celle che fanno parte di questa LA un segnale che chiede al nodo mobile di farsi vivo. In tutte le celle-1 questi messaggi andranno a vuoto. Nella cella in cui si trova effettivamente la ricerca avrà successo ed a questo punto si può stabilire la connessione. Ovviamente questo potrebbe anche fallire questa ricerca, perché il nodo mobile volontariamente o involontariamente potrebbe risultare non raggiungibile. Se ci pensiamo un attimo, se combiniamo questa procedura con quella precedente, questo ci dà anche un'idea del fatto che il ritardo quando proviamo a contattare un oggetto connesso alla rete cellulare, il ritardo che con cui veniamo informati della non raggiungibilità può cambiare. Se il nodo è spento, allora l'informazione viene registrata nell'ultimo VLR a cui era associato quel nodo mobile, e quindi quando si avvia la procedura di localizzazione già a questo punto si è in grado di rispondere che il nodo mobile non è raggiungibile. Se il nodo mobile non è raggiungibile perché è entrato in una zona d'ombra, la procedura deve essere eseguita tutta quanta e c'è un timeout prima di indicare il terminale come non raggiungibile. Il tempo di attesa che sperimentiamo in questo caso è sensibilmente percepibile.

Parametri che influenzano le prestazioni di questi algoritmi. Dipendono dalla dimensione della LA. Al minimo può essere formata da una singola cella, nulla esclude che possa coincidere con una singola cella. All'estremo opposto può essere formato dalla totalità dei nodi che fanno capo all'MSC. Altro parametro che può influire è il tipo di LA, se definita in modo statico indipendentemente dalle azioni del nodo mobile, o se definita in modo dinamico in relazione alle attività del nodo mobile. Questi due parametri, dimensione e tipologia, hanno un impatto sicuramente sulle prestazioni degli algoritmi che abbiamo descritto. Dove l'impatto può essere misurato in termini di traffico di controllo generato, di ritardo che si spende nel localizzare con precisione la posizione del nodo mobile ed anche energia che deve essere consumata dal nodo mobile a causa del suo coinvolgimento in queste operazioni.

Riguardo l'impatto causato dalla dimensione della LA, riguardo al traffico generato per fare tracciamento quando il nodo si sposta, chiaramente quanto più piccole sono le LA, maggiore è il traffico. Dal punto di vista del costo del tracciamento sembrerebbe convenire scegliere LA quanto più grandi possibili. Il problema è che dal traffico di localizzazione, che corrisponde ad inondare tutte le celle che fanno parte di una singola LA di messaggi di ricerca del nodo, quanto più grande è la LA quante più celle riceveranno il messaggio di

localizzazione. Dal punto di vista del consumo energetico va in parallelo con il tracciamento, quindi LA più grandi richiedono meno consumo energetico a causa del coinvolgimento del nodo mobile per l'attivazione dell'algoritmo di tracciamento. Dove collocarsi? Dipende dalla tipologia di traffico e di mobilità dei nodi che devono essere supportati da una certa rete. Bisogna analizzare il livello di mobilità dei nodi, se si muovo spesso o con un ampio raggio, se si muovono poco, se si indirizzano molte o poche richieste di connessione. A seconda dei profili di uso e di mobilità, può convenire dare peso ad un fattore o all'altro.

Dal punto di vista del tipo della modalità di definire la LA, la modalità più semplice è di definirla in modalità statica. Come ripartire questa architettura fisica in aree logiche adiacenti invece è una cosa che può essere decisa, magari una volta per tutte in modo statico, che è la cosa più semplice da fare. L'inconveniente che si può creare è che si possono creare situazioni anomale, nel caso in cui magari per evoluzioni non previste nei profili di uso e di mobilità degli utenti che utilizzano questa infrastruttura, si crea un eccessivo traffico lungo la linea di confine tra due LA. Si può creare il così detto effetto ping-pong, dunque se tanti utenti si spostano frequentemente da una posizione a cavallo della linea di confine, questo può generare traffico di controllo e di tracciamento notevole, il che non è un bene dal punto di vista della tenuta del sistema. Come si può ovviare a questo? Un modo è quello di definire le LA non in modo statico ed indipendente, ma in maniera dinamica. A questo punto la LA diventa un concetto legato al nodo mobile, ed ognuno ha la sua LA. Ad ogni nodo mobile viene associato un insieme di celle all'interno delle quali il nodo si trova. Nel momento in cui il nodo si accede, viene associata una LA formata dalla cella in cui si trova e dalle celle circostanti. Fin quando il nodo mobile si trova all'interno di questo insieme non succede nulla. Quando esce dall'insieme, l'informazione deve essere aggiornata e ri-centrata la LA prendendo come centro la nuova posizione occupata e definendo un nuovo intorno. L'aggiornamento può essere fatto in modo reattivo, quando esco dalla LA, ma anche pro attivamente prevenendo questo transito tenendo conto degli spostamenti del nodo mobile, ricentrando la LA intorno alla posizione attuale. L'idea p che ho una posizione attuale ed una LA definita da un intorno di celle circostanti. In questa idea di non aspettare che il nodo mobile esca dai confini e cercare di seguire i suoi spostamenti, se riesco a seguirlo riduco la necessità di aggiornare la posizione del nodo mobile. L'idea sul quando fare il riaggiustamento, sono state proposte varie soluzioni. Questa soluzione è stata inizialmente esplorata in maniera teorica, ma a partire dalla quarta generazione questo concetto è stato riesumato.