

## MSA - 2.03.2021

Diamo qualche motivazione. Detto in pochissime parole, questa frase che vedete proiettata, l'ho estratta dall'editoriale del direttore di una rivista dedicata ai temi su cui questo corso si focalizza, uscito pochissimi mesi fa. In pochissime parole da un'idea del perché all'interno del corso di laurea in ingegneria informatica magistrale esiste un corso come questo. Un corso che si concentra sulle infrastrutture ed applicazioni che ruotano attorno all'utilizzazione tramite dispositivi mobili, questo dominio applicativo non sono semplicemente la versione in taglia minore di infrastrutture ed applicazioni che girano sul mondo fisso, meglio attrezzato, ma hanno delle caratteristiche particolari che meritano di essere approfondite. Le ragioni di fondo che hanno consentito la definizione, lo sviluppo, di questo tipo di dominio applicativo sono quei tre aspetti riassunti nel primo periodo: a partire da un certo punto nell'evoluzione delle infrastrutture informatiche, hanno trovato piena realizzazione la possibilità di effettuare comunicazioni utilizzando canali wireless, la disponibilità di nodi alimentati da batterie e gli avanzamenti tecnologici che hanno consentito di rendere portabili i dispositivi di calcolo di capacità ragionevoli. Detto in tre parole è ciò che sta alla base di un corso di questo tipo.

L'esigenza di interessarci di queste tematiche le possiamo datare ad una 30ina di anni fa, inizio anni '90. Da allora è passata abbastanza acqua sotto ai ponti, e quelle che erano problematiche che si ponevano allora hanno lentamente preso una forma diversa, si sono arricchite di nuove sfaccettature, che in parte poi toccheremo durante il corso. Questa è una citazione dal documento introduttivo di una conferenza dall'edizione di due anni fa. Quelli che vedete evidenziati sono in questo documento introduttivo le sfide e gli argomenti sui quali si è interessati a conoscere lo stato dell'arte della ricerca. Alcuni punti cruciali, descrivono abbastanza bene lo stato dell'arte della situazione attuale: i dispositivi mobili, portatili, stanno continuamente accrescendo le loro capacità computazionali; abbiamo un'infrastruttura di rete che permette ai dispositivi di interagire tra di loro, che sta transitando da un panorama con dispositivi dedicati ad un mondo dove tutto è software, dove l'hardware diventa di general purpose; infrastruttura cloud, ma a questo tipo di infrastruttura sta evolvendo in direzione di una maggiore distribuzione, diffusione nell'ambiente, come edge e fog computing. Cambio di scenario da architetture per dispositivi fissi a mobili.

Tra gli aspetti rilevanti che il cambio di scenario comporta è la gestione della connettività in movimento, come consentire ad entità, fisiche o virtuale, la cui collocazione fisica nello spazio non è fissa, come garantire la raggiungibilità dei nodi, avendo a che fare con un'infrastruttura di comunicazione che è quella internet, che invece è nata quando la mobilità non era concepibile. Altro aspetto architeturale importante che ha un impatto nel modo in cui le applicazioni vengono progettate è la disponibilità, più o meno diffusa, di risorse computazionali, che possono fornire l'occasione di demandare a dispositivi esterni rispetto a quello mobile una parte del carico computazionale o di memorizzazione, dove questi nodi esterni possono essere piattaforme cloud remote ma anche nodi più in prossimità. Tutto questo ha implicazioni su come le applicazioni vengono progettate. La tematica mobile computing è estremamente vasta, non basta un singolo corso per trattare

in maniera esaustiva ed in profondità tutte le tematiche connesse. Ne toccheremo alcune, e l'approccio generale che utilizzerò è un approccio fondazione, per quanto possibile cercheremo di andare a guardare sotto la superficie delle cose, al di là delle istanziazioni correnti a certi problemi, qual'è il nocciolo del problema che si vuole risolvere, e come magari nel corso del tempo possiamo leggere i tentativi di soluzione che sono stati dati in passato ed imparare da queste esperienze.

Puntualizzazioni di alcuni fatti da tenere presenti. Un primo fatto che è sotto gli occhi di tutti quanti è che attualmente i dispositivi mobili sono diventati per molti di noi, per tanti di noi, il principale punto di accesso a tutte le tecnologie informatiche a nostra disposizione. Da alcuni anni il numero di dispositivi mobili a disposizione degli utilizzatori ha superato il numero di dispositivi di calcolo fissi, detti desktop. E' anche abbastanza usuale che un singolo utente disponga di più dispositivi tramite i quali l'utente può usufruire ed interagire con applicazioni e sistemi informatici. Cerchiamo di non considerarci il centro del mondo. Ciò che noi personalmente sperimentiamo non è detto che corrisponda all'esperienza della maggior parte degli esseri umani. Aumenta anche la diffusione delle tecnologie indossabili, a partire dagli smartwatch fino ad altri dispositivi che si utilizzano in domini più limitati come medico e sportivo. Più in generale c'è l'idea, ormai si avvia a diventare realtà, dell'IoT, la capacità di gran parte degli oggetti che ci circondano di agire come produttori e consumatori attivi di informazione. Alla fine l'ecosistema a cui possiamo pensare di fare riferimento per le tematiche di cui ci occuperemo nell'ambito di questo corso può essere riassunto in questa figura. Il tutto si basa sul protocollo internet, e l'infrastruttura è lo strumento con cui entità di tanti tipi diversi sono in grado di interagire, e la necessità di interagire e favorire queste capacità interattive portano con se tutta una serie di problematiche, evidenziate dalle parole chiavi citate nella figura.

Questi erano alcuni fatti, positivi. Ma, dobbiamo anche avere una visione ad ampio raggio e tenere presente limiti, vincoli di cui tenere conto per non farci trainare dall'entusiasmo. Un primo fatto è che è vero che i dispositivi mobili, le piattaforme che abilitano l'utilizzo in mobilità di funzionalità informatiche, rispetto agli albori di questa era che datano ad una 30ina di anni fa, le capacità computazionali sono cresciute e continuano a crescere. Ma se confrontiamo, nello stesso istante di tempo, le capacità di un'infrastruttura mobile rispetto a quella della contemporanea infrastruttura fissa, abbiamo tutt'ora un gap. C'è il problema della disponibilità di una sorgente portatile di energia, una batteria, che in assenza di alimentazione è soggetta ad esaurimento. Tipicamente le interazioni tra dispositivi mobili ed il resto di un'infrastruttura tipicamente hanno una latenza superiore a quello che accade tra dispositivi fissi oltre alla possibilità di disservizi. Dobbiamo tenere conto che questi limiti verranno sempre più messi alla prova da un aumento quasi esponenziale nella richiesta di risorse che abbiamo a livello applicativo. Il mondo dell'IoT, miliardi di dispositivi in questo scenario sono o saranno connessi o desiderosi di connettersi ad un'infrastruttura generale, che provocherà l'introduzione nell'infrastruttura di quantità di dati inimmaginabili fino a poco tempo fa, andando a minare la possibilità di reggere il traffico. In fine, ultimo aspetto, in questa crescita continua di esigenze e di risorse che vengono messe in campo, dobbiamo cercare di non dimenticarci che il mondo in cui

viviamo è un mondo finito. Ci dobbiamo preoccupare di guardare all'aspetto della sostenibilità che pensiamo di progettare ed introdurre in questo mondo.

Tematiche vaste, sfidanti, non avremo tempo e modo di trattarle tutte. Nell'ambito di questo corso presenteremo alcune problematiche sottostanti la realizzazione di sistemi, infrastrutture ed applicazioni fruibili in mobilità, dove gran parte delle problematiche che esamineremo sono legate alla presenza di quei tre punti evidenziati all'inizio: comunicazioni wireless, dispositivi portabili e la presenza di tanti dispositivi che possono generare o consumare dati ed informazioni. Tutto ciò, se lo inquadrriamo che è il tipico modello concettuale di architettura a strati, tutto ciò ha un impatto su tutti i livelli del modello. Nell'ambito di questo corso esamineremo tematiche che si collocano ai vari livelli del modello. Trascureremo il modello più basso, però a partire dal livello logico, delle comunicazioni e fino al livello di rete e trasporto, per tutti questi livelli affronteremo alcune tematiche ed esploreremo alcune soluzioni date a problemi emersi in ognuno di questi livelli.

Giusto per darvi un'idea sommaria del percorso che seguiremo, inizieremo a partire da oggi con una brevissima caratterizzazione di quello che è il livello 0, a partire dal quale ci si inizia a porre problematiche del mondo mobile. Successivamente faremo una rapida panoramica di alcuni aspetti fondamentali del mobile computing nelle reti wireless. Fatto questo inizieremo ad entrare nel merito delle problematiche. La direzione rispetto alla pila, al modello a strati, sarà dal basso verso l'alto in una certa misura. Inizieremo a guardare a livello delle comunicazioni, soprattutto alcune problematiche relative alla realizzazione di comunicazioni usando un supporto wireless. Tematica connessa è come gestire la raggiungibilità dei nodi che non hanno una posizione fissa, problema che si potrebbe porre in assenza di connessione wireless. Sono due tematiche che in modi diversi guardano all'aspetto della interazione tra dispositivi mobili.

Fatto questo passeremo al livello applicativo. Quello che guarderemo saranno aspetti che riguardano la fondazione della progettazione, quindi aspetti che riguardano la definizione di un'architettura pensata per un mondo mobile, ed esploreremo due strategie architetturali proposte come possibili approcci per risolvere una serie di problemi di questo livello. Come tematiche trasversali che toccheremo in maggiore o minore misura in tutti questi punti ci sono le problematiche dell'adattamento, il mondo mobile è un mondo soggetto a cambiamenti continui, soluzioni in grado di adattare il nostro oggetto di interesse ai continui cambiamenti. Un'altra tematica importante è quella del risparmio energetico, che ha comunque una valenza generale, a prescindere dalla specificità del mondo mobile, dove assume una valenza particolare in quanto tipicamente molti dei dispositivi con cui avremo a che fare saranno alimentati da batterie.

---

## Baseline

Giusto per avviare il discorso, questo primo set di discorsi che faremo servono un po' a tracciare quello che possiamo considerare il livello 0 rispetto al quale ci confronteremo.

Questa pietra di paragone può essere un sistema distribuito classico, in cui i nodi di elaborazione hanno tipicamente una posizione fissa nello spazio, e le infrastrutture di comunicazioni utilizzate per interagire è basata su cavi. Puntualizzeremo alcuni punti caratterizzanti di un sistema di questo tipo, rispetto al sistema che noi prenderemo in considerazione.

Questo tipo di sistema che stiamo considerando è un sistema i cui aspetti fondamentali sono stati definiti un mezzo secolo fa. L'evoluzione, la definizione abbastanza completa delle soluzioni architetturali, è un'evoluzione che ha avuto il suo sviluppo nelle due decadi dagli anni '70 fino ai primi anni '90. Le soluzioni proposte sono state dettate da quello che all'epoca era l'evoluzione della tecnologia, quindi in particolare ad inizio anni '70 si passava da un mondo basato su pochi grossi mainframe, si iniziava a transitare in modo più evidente ad un mondo in cui comparivano personal computer che si connettevano in rete tra di loro. Una caratteristica delle infrastrutture che emergevano è che queste infrastrutture erano omogenee, sia in termini di hardware, abbastanza usuale che i dispositivi connessi ad una singola rete fossero più o meno tutti dello stesso tipo; inoltre questi sistemi venivano progettati sotto la così detta assunzione di un mondo chiuso, le cui caratteristiche possono essere considerate note a priori e non soggette a cambiamenti, su una scala di tempi ragionevolmente lunga.

All'interno di un sistema che si caratterizzava in questo modo sono emerse varie problematiche a cui sono state date risposte che ormai fanno parte del bagaglio culturale di ogni informatico, rientrano tutte sotto l'ombrello generale di calcolo distribuito. Se andiamo a guardare un po' più nel dettaglio come si caratterizzano le entità in gioco all'interno di questo sistema, partendo dal livello fisico, questi sistemi erano caratterizzati da nodi di elaborazione fissi, in una posizione che non cambiava nel tempo, con capacità computazionali notevoli, memoria centrale di dimensioni sempre più considerevoli; dal punto di vista delle modalità di interazione con utilizzatori umani, dispositivi sempre più sofisticati ed amichevoli. Dal punto di vista della disponibilità di energia, con tutte le dovute cautele, in prima battuta questi dispositivi assumono la disponibilità illimitata di energia. Non sono progettati pensando di dover avere il risparmio energetico come problema primario da tenere in considerazione.

Sempre rimanendo ancorati al livello fisico, dal punto di vista degli strumenti fisici di comunicazione, questi erano cavi di varia natura, con varie capacità che sono state via via evolvendo. Questi canali di comunicazione sono caratterizzati dal punto di vista delle prestazioni, con latenza molto bassa. Dal punto di vista dell'affidabilità, garantiscono tassi di errore estremamente bassi. Iniziando a salire nella pila protocollare, passando al livello data link, esistono degli standard ormai consolidati, con tutta la famiglia 802.x, una famiglia di protocolli di livello 2, alcuni dei quali ancora vivi e vegeti che hanno tutt'ora un ruolo primario nella gestione della comunicazione. Continuando a salire a livello network sappiamo che abbiamo il protocollo IP, con tutte le sue tematiche connesse di protocolli di instradamento. Continuando a salire di livello il protocollo IP è quello che realizza la comunicazione tra nodi di un sistema. A livello trasporto abbiamo la comunicazione tra processi che girano più o meno in parallelo su singoli nodi. A questo livello c'è tutta una famiglia di protocolli che si è consolidata nel tempo. Ognuno copre esigenze comunicative differenziate.

Poi a livello applicativo, sempre in quegli anni, si sono consolidati una serie di domini applicativi, ognuno mirato a specifiche esigenze caratterizzato dai suoi propri tipi di dati, cioè modi di rappresentare informazione, operazioni disponibili su queste rappresentazioni, dedicate a specifiche esigenze comunicative. Database, con tutti i loro meccanismi di algebra relazionale, file system più o meno distribuiti, applicazioni di tipo multimediale, sistemi iper testuali e così via. In questa pila protocollare nel corso del tempo si è aggiunto uno strato ulteriore che generalmente viene indicato con lo stato del middleware, tra applicazioni e trasporto, che permette di ridurre la distanza tra i due livelli. In questo livello si collocano tutta una serie di funzioni di livello più o meno alto. Questo è il nostro punto di riferimento, il livello 0.

---

## Focus

A partire a questo livello 0, iniziamo a vedere dove ci andremo a muovere, quale sarà il nostro oggetto di interesse.

Quello che dicevo nella slide introduttiva, quindi nel set precedente, le motivazioni per cui alla fine ci troviamo qui a parlare è che le tecnologie, quindi le infrastrutture informatiche si sono evolute nel tempo, e da una situazione con pochi grandi mainframe, si è passati alla situazione in cui possiamo assumere di avere un dispositivo di calcolo personale per ogni utente, alla situazione attuale in cui un singolo utente umano può avere a disposizione una varietà di possibilità di strumenti con cui interagire. E questo, sono grafici che si fermano ad un po di anni fa, ma diciamo che in ordinata ci sono le vendite per anno ed è possibile vedere come il mondo del mainframe ha avuto un picco negli anni '70.

Dal punto di vista delle infrastrutture di comunicazione, questa diapositiva è una specie di albero genealogico alla rovescia, in quanto parte da tanti rami diversi e poi nel corso del tempo queste diverse tecnologie sono confluiti verso di più un'infrastruttura che offre una visione unica delle modalità di interazione tra dispositivi. Dal punto di vista hardware si è passati da soluzioni fisicamente basate su tecnologie analogiche a tecnologie digitali. Di tutto questo emerge un sistema in cui quelli che propriamente intendiamo come dispositivi informatici, tablet, smartphone e poco più, sono i punti di accesso e di controllo di una varietà molto più ampia di dispositivi che sempre più sono dotati della capacità di interagire tra di loro, mediati da questi dispositivi centrali, e con esseri umani che usano questi dispositivi di accesso e di controllo.

Le problematiche poste dalla progettazione, realizzazione di questo tipo di infrastruttura e gli obiettivi che ci possiamo porre, li possiamo nominare a vari livelli di ambizione. Il termine, le parole chiave che si utilizzano per indicare le sfide che abbiamo davanti, l'obiettivo a cui tendono, a volte sono diversi. Intanto, piccola precisazione terminologica, è il termine di **pervasive computing**. L'idea dietro a questo termine è la sfida ottimamente descritta dalle parole di un ricercatore molto brillante. Queste parole citate sono estratte da un suo articolo. Qui presenta quella che è la sua visione di ciò che rende una particolare tecnologia una tecnologia di successo. In che modo noi possiamo misurare il grado di successo di una specifica tecnologia. La sua visione è che il suo livello di successo si misura

rispetto al suo livello di beneficio che riesce a portare a noi umani. Astraendo sda specifici domini applicativi, se vogliamo trovare un minimo fattore comune per trovare tecnologie diverse potrebbe essere quella dell'invisibilità. Più di successo, quanto meno noi umani per usufruire delle sue funzioni non ci rendiamo conto della presenza di quella tecnologia. Quanto meno una tecnologia è invasiva, tanto più è di successo. Esempio è la produzione di energia elettrica. Il più che ci viene chiesto è di infilare una spina in una presa, l'unico livello di coinvolgimento attivo richiesto. Siamo tutti in grado di intuire il livello di complessità che c'è dietro al semplice gesto di infilare una spina in una presa e godere della presenza di questa sorgente di energia. Se l'informatica mobile vuole aspirare allo stesso livello di successo, deve aspirare allo stesso livello di indistinguibilità. Le funzioni offerte da questo tipo di applicazioni si devono intrecciare nel tessuto della nostra vita umana e sparire, farci godere dei loro vantaggi senza richiedere un impegno fisico o mentale tangibile. Ancora non ci siamo in buona misura, ma lo possiamo porre come asticella.

Un altro termine potrebbe essere quello di ubiquitous computing, che mette in basso l'asticella. Sparisce la focalizzazione l'invisibilità e semplicemente, sotto questo ombrello generale, entrano in gioco le problematiche con il rendere disponibile ovunque, in ogni luogo e tempo possibilmente, la possibilità di usufruire di funzionalità informatiche. Altro dominio è quello del wearable computing, lo sviluppo di applicazioni che intendono supportare, aumentare o controllare funzioni umani grazie a dispositivi che sono più o meno incorporati nel nostro corpo fisico, sovrapposto anche alla superficie esterna, ma anche all'interno del nostro corpo come i pacemaker che ormai sono integrati nella nostra vita.

C'è una tematica trasversale che abbraccia tutti questi ambiti applicativi, che è quella della così detta context awareness, la consapevolezza del contesto. Se voglio realizzare applicazioni pervasive, ubiquitarie o a diretto contatto con il corpo umano, queste applicazioni devono essere dotate della capacità di distinguere tra contesti diversi. Per contesto si intende tutto quell'insieme di informazioni, parametri, eventi ecc che ci consentono di caratterizzare lo stato e la situazione di una certa entità. A seconda dello stato o della situazione, può essere più o meno opportuno consentire di fare delle cose piuttosto che altre.

Fatta questa introduzione proviamo a rifare il percorso fatto per il sistema fisso, di cui abbiamo parlato poco fa, proviamo a rifarlo in termini rispetto allo scenario mobile di cui ci vogliamo occupare.