

Università degli Studi di Bologna

FACOLTÀ DI INGEGNERIA

Corso di Laurea specialistica in Ingegneria Informatica

RETI DI CALCOLATORI LS

**STRATEGIE DI ELABORAZIONE DI
TRACCIATI GPS PER RETI VEICOLARI**

Tesi di Laurea di:

ROBERTO CARISI

Relatore:

Chiar.mo Prof. **ANTONIO CORRADI**

Correlatori:

Dr. **GIOVANNI PAU**

Sessione Terza

Anno Accademico 2008/2009

PAROLE CHIAVE:

- **ELEMENTI TOPOLOGICI DELLE MAPPE**
- **TRACCIATI GPS**
- **RETI AD HOC**
- **RETI VEICOLARI**
- **ELABORAZIONE DINAMICA TRACCIATI GPS**

*Ad Andrea, Pina e Stefano,
a cui devo quello che sono.*

Indice

Ringraziamenti	vii
Introduzione	ix
1 Scenario Applicativo	1
2 Mobile Ad hoc NETWORKS	7
2.1 Evoluzione storica	7
2.2 Le reti 4G	9
2.3 Caratteristiche delle MANET	13
2.4 Scenari Applicativi	15
3 Ad-Hoc Networks	19
3.1 Tecnologie Attuali	19
3.1.1 Bluetooth	21
3.1.2 IEEE 802.11	23
3.2 Location Services	24
3.3 Routing e Forwarding in reti ad hoc	25
3.4 Location-Aware Routing	28
4 Global Positioning System, GPS	31
4.1 Evoluzione storica e funzionamento	31
4.2 Fonti di errore	38
4.3 Esperimenti in scenario urbano	46
4.3.1 Incidenza dell'ambiente urbano	46
4.3.2 Esperimenti in mobilità	47
4.3.3 Esperimenti con ricevitore stazionario	51
5 Crowdsourcing	55

6 Estrazione informazioni topologiche: Semafori e Stop	62
6.1 Pre-processing della mappa	63
6.2 Selezione delle tracce utili	65
6.3 Mappatura delle tracce con Reverse Geocoding intelligente	68
6.3.1 Reverse Geocoding	70
6.3.2 Reverse Geocoding Intelligente	74
6.3.3 Eliminazione di discontinuità temporali o di percorso dalle tracce	79
6.4 Analisi delle tracce ed estrazione informazioni caratteristiche	82
6.4.1 Svolte	83
6.4.2 Rallentamenti	84
6.4.3 Fermate	86
6.5 Identificazione di Stop, Semafori e temporizzazioni	88
6.5.1 Segnali di Stop	89
6.5.2 Semafori	91
6.5.3 Gestire la mancanza di informazioni	92
7 Valutazione del sistema	94
7.1 Caratteristiche dei Data Set	94
7.2 Risultati Sperimentali	97
Conclusioni	106
A Pre-processing delle mappe	108
A.1 Il Database TIGER Census	108
A.2 Il Database OpenStreetMap	111
B Standard dei Dati GPS	113
B.1 Ricevitore GPS Holux GR-213 SiRF Star III	113
B.2 NMEA	114
B.3 GPX	117
B.4 Sistemi di Coordinate	120
B.4.1 Universal Transverse Mercator System	120
B.4.2 East, North, Up	120
Bibliografia	122

Ringraziamenti

Bologna, 18 marzo 2010

Per una persona poco abituata a dire grazie, può essere proprio questa la parte più difficile da scrivere. È giusto però dedicarla a quelle poche persone a cui devo molto.

Prima di tutto, ad Andrea e Pina, grazie a cui ho imparato a crescere, e che coi loro sacrifici mi hanno permesso di raggiungere tanti traguardi, con la certezza di poter sempre contare su di loro, ma senza mai essere invadenti.

A Stefano, che anche se mi piace stuzzicarlo un po', è l'esempio perfetto di fratello minore che darei a chi mi chiedesse come lo vorrei.

A zio Mario, senza cui probabilmente non avrei mai deciso di allungare un po' il passo e fare le esperienze all'estero che invece mi porterò sempre nel cuore, e probabilmente mi hanno cambiato la vita.

A Giovanni, Rosina e Duilio, che mi han guidato per mano, riempiendo i miei primi 25 anni di ricordi che per tutta la vita mi faranno sorridere.

A Michela, che ancora mi tiene per mano in questo cammino, e con l'aiuto dei suoi pranzetti mi ha permesso di raggiungere quest'obiettivo in tempi insperati.

Ai miei cugini, cugine, zii e zie, che magari non vedo molto, ma ogni anni aspettiamo solo che sia la Vigilia o il giorno di Natale per trascorrerlo tutti insieme e ricordarci quanto è bello essere una famiglia così numerosa e felice.

Ai miei amici di sempre, gli “sbregaz della taverna” e i “regaz della discoteca”, con cui abbiamo condiviso tante pagine di vita e scritto tanti capitoli di storia, grazie a cui anche quando sono all’altro capo del mondo riesco a sentirmi come a casa.

Ad Elena, che riesce a strapparmi un sorriso anche a sei mila miglia di distanza, semplicemente guardandomi negli occhi, e mi sopporta nel mio essere burbero e scontroso quando sono sotto stress, facendo finta di niente.

Ai miei compagni di corso, quelli veri, quelli che quando pensi all’università per prima cosa ti ricordi le loro facce, e non i libri; ricordi i picnic improvvisati a vinaccia e affettati, gli aperitivi e i calcetti, non gli esami.

Ad Eugenio, Ryad, Fabio, Michele, Carlo, Piero, Pasquale, Raphael e Simone, per aver reso ogni giorno all’NRL un giorno speciale, da non dimenticare. Ai ragazzi italiani degli altri lab, con i quali ho lasciato a LA un pezzo di cuore.

A Giovanni, per gli insegnamenti, la stima e l’affetto, più che per la possibilità di ricerca datami all’UCLA.

Ai ragazzi della Ginzburg, che mi hanno ricordato il vero motivo per cui frequentare le biblioteche e mi han fatto sentire da subito uno di loro.

Ai compagni del liceo, che chi l’avrebbe detto, si cresce, si invecchia, si matura (questa la mettiamo ai voti eh...) e siamo sempre qui, tutti insieme come 10 anni fa.

A Stella e Rocky, perché spesso quando sei a casa non puoi avere al tuo fianco chi vorresti, ma loro ci sono sempre, e guardandoli capisci cosa vuol dire volersi bene.

Come promesso, un grazie va anche alla patata buona del Paleotti, che dava sempre i giusti stimoli per resistere fino alle 22 in biblioteca; all’ucraina che, con la pasta fatta a mano la mattina, mi permetteva di affrontare col giusto ritmo il pomeriggio di studio, supportata però dal Vez delle caldaroste, eroico con noi fino a metà febbraio.

Introduzione

Il proliferare di dispositivi di elaborazione e comunicazione mobili (telefoni cellulari, computer portatili, PDA, wearable devices, personal digital assistant) sta guidando un cambiamento rivoluzionario nella nostra società dell'informazione. Si sta migrando dall'era dei Personal Computer all'era dell'Ubiquitous Computing, in cui un utente utilizza, parallelamente, svariati dispositivi elettronici attraverso cui può accedere a tutte le informazioni, ovunque e quantunque queste gli si rivelino necessarie [32].

In questo scenario, anche le mappe digitali stanno diventando sempre più parte delle nostre attività quotidiane; esse trasmettono informazioni vitali per una pletora di applicazioni che acquistano maggior valore grazie alla localizzazione, come Yelp, Flickr, Facebook, Google Maps o semplicemente le ricerche web geo-localizzate. Gli utenti di PDA e Smartphone dipendono sempre più dai GPS e dai Location Based Services (LBS) per la navigazione, sia automobilistica che a piedi [36]. Gli stessi servizi di mappe stanno inoltre evolvendo la loro natura da uni-direzionale a bi-direzionale; la topologia stradale è arricchita da informazioni dinamiche, come traffico in tempo reale e contenuti creati dagli utenti.

Le mappe digitali aggiornabili dinamicamente sono sul punto di diventare un saldo trampolino di lancio per i sistemi mobili ad alta dinamicità ed interattività, che poggiando su poche informazioni fornite dagli utenti, porteranno una moltitudine di applicazioni innovative ad un'enorme base di consumatori [37]. I futuri sistemi di navigazione per esempio, potranno utilizzare informazioni estese su semafori, presenza di stop ed informazioni sul traffico per effettuare una ottimizzazione del percorso che valuti simultaneamente fattori come l'impronta al carbonio rilasciata, il tempo di viaggio effettivamente necessario e l'impatto della scelta sul traffico locale.

In questo progetto si mostra come i dati GPS raccolti da dispositivi fissi e mobili

possano essere usati per estendere le mappe digitali con la locazione dei segnali di stop, dei semafori e delle relative temporizzazioni. Queste informazioni sono infatti oggi rare e locali ad ogni singola municipalità, il che ne rende praticamente impossibile il pieno reperimento. Si presenta quindi un algoritmo che estrae utili informazioni topologiche da agglomerati di tracciati gps, mostrando inoltre che anche un esiguo numero di veicoli equipaggiati con la strumentazione necessaria sono sufficienti per abilitare l'estensione delle mappe digitali con nuovi attributi. Infine, si mostrerà come l'algoritmo sia in grado di lavorare anche con dati mancanti, ottenendo ottimi risultati e mostrandosi flessibile ed adatto all'integrazione in sistemi reali.

Nel primo capitolo di questo lavoro, verrà presentato più nel dettaglio lo scenario in cui si inserisce il sistema proposto, con una prima introduzione alle reti mobili ed al valore delle mappe digitali.

Nel secondo capitolo, dopo alcuni cenni storici sull'evoluzione delle reti mobili ad hoc (*MANET*), analizzeremo le principali caratteristiche delle reti di quarta generazione, in cui si inseriscono appunto le MANET, ed i loro possibili scenari applicativi.

Nel terzo, verranno mostrate le tecnologie attualmente presenti sul mercato riguardanti il mondo delle reti ad hoc e le possibili soluzioni di routing e forwarding, prestando particolare attenzione alla tipologia di routing location-aware.

Nel quarto capitolo invece verranno presentate brevemente le caratteristiche tecniche dei ricevitori GPS disponibili in commercio, la loro precisione e gli errori più comuni che si possono ritrovare nei dati ottenuti.

Nel quinto si introduce il concetto di crowdsourcing, in prima analisi come nuovo modello di business e successivamente come risorsa fondamentale per ottenere ed inferire nuova conoscenza dagli utenti stessi che diventano quindi pro-sumer.

Nei capitoli sei e sette si descrive invece il design e la fase di valutazione del sistema prodotto per trasformare semplici tracciati GPS registrati dagli utenti in utili informazioni topologiche.

Capitolo 1

Scenario Applicativo

La migrazione della società attuale dall'era dei Personal Computer a quella dell'Ubiquitous Computing è guidata nettamente dalla crescente disponibilità di dispositivi mobili a basso costo con sempre maggiori capacità computazionali. La natura onnipresente di questi dispositivi rende le reti wireless la soluzione più semplice per la loro interconnessione e, conseguentemente, anche il campo wireless ha assistito ad una crescita esponenziale nelle scorse due decadi.

Gli utenti mobili possono utilizzare il telefono cellulare per controllare la posta o fare ricerche in rete; chi viaggia col laptop può navigare in internet da aeroporti, stazioni, café o altri luoghi pubblici adeguatamente attrezzati; i turisti possono usare terminali Global Positioning System (GPS) per ottenere indicazioni stradali o individuare luoghi di interesse; i ricercatori possono scambiarsi file ed altre informazioni connettendosi a reti wireless mentre partecipano a videoconferenze multi utente, che siano essi in viaggio o in una camera d'albergo dall'altra parte del mondo; in casa infine, gli utenti possono facilmente sincronizzare dati fra dispositivi portatili e fissi. Un esempio della pletora di dispositivi che possono collegarsi alla rete in mobilità per soddisfare l'esigenza degli utenti è mostrato in figura 1.1.

Non solo i dispositivi mobili stanno diventando sempre più piccoli, economici, convenienti e dotati di potenza di calcolo, ma hanno anche un supporto a livello

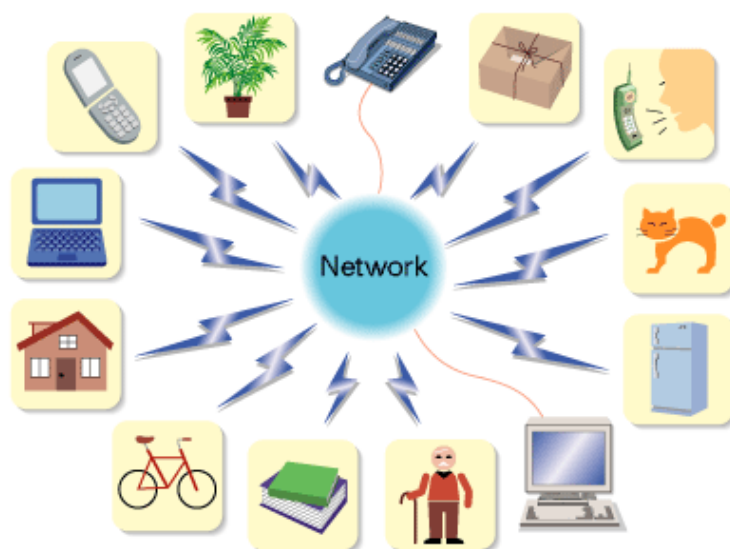


Figura 1.1: Esempio di dispositivi che concorrono a realizzare l'*ubiquitous computing*

applicativo e di servizi di rete in costante crescita, che alimenta incessantemente la loro diffusione esponenziale sul mercato [34]. Con questo ritmo di crescita, è lecito aspettarsi che il numero totale di utenti internet mobili surclassi presto quello degli utenti di linea cablata.

Fra tutte le applicazioni ed i servizi disponibili su dispositivi mobili, le connessioni di rete ed i servizi correlati sono senza dubbio i più richiesti dagli utenti. Secondo studi svolti dal Cahners In-Stat Group [58], il numero di iscritti a servizi di dati mobili, così come il numero di messaggi wireless scambiati nella scorsa decade è aumentato con ritmo più che esponenziale. Al momento, la maggioranza delle connessioni fra dispositivi mobili sono ottenute attraverso un service provider basata su un'infrastruttura fissa, o tramite reti private. Ad esempio, la connessione fra due telefoni cellulari avviene grazie a Base Station Controller (BSC) e Mobile Switching Center (MSC), mentre i computer portatili accedono alla rete internet passando tramite access points wireless.

Se da un canto le reti basate su infrastruttura forniscono un facile modo per i dispositivi mobili per accedere a servizi di rete, dall'altro richiedono tempo e potenzialmente alti costi per la creazione ed il mantenimento dell'infrastruttura stessa, che si ripercuotono poi sull'utente finale con alti costi di connessione. Ci sono, inoltre, situazioni in cui la connettività richiesta dall'utente non è disponibile

tramite un'infrastruttura in date aree geografiche, e fornire la necessaria connettività in queste situazioni diventa una vera sfida. Un esempio di tali situazioni sono i territori di guerra, o anche gli stessi ambiti urbani in seguito a una catastrofe o un attacco terroristico che fa venir meno l'infrastruttura esistente.

Proprio da queste osservazioni, recentemente hanno cominciato ad emergere metodi alternativi di fornitura di servizi di rete. Questi si basano sul fatto che i dispositivi mobili si connettano agli altri dispositivi nel loro raggio di trasmissione attraverso una configurazione automatica, costruendo una rete ad hoc mobile che è al tempo stesso flessibile e potente. In questo modo, non soltanto i nodi mobili possono comunicare vicendevolmente, ma possono anche ricevere servizi internet attraverso un nodo che abbia questa connettività e funga da gateway, estendendo effettivamente i servizi internet all'area priva di infrastruttura.

In una Mobile Ad hoc NETWORK (MANET) [1] di questo tipo, i dispositivi mobili degli utenti sono la rete stessa, e devono perciò fornire cooperativamente le funzionalità solitamente fornite dall'infrastruttura (es. routers, switches, servers). Possiamo considerare questi dispositivi come un'evoluzione degli odierni telefoni cellulari, come smartphones provvisti delle necessarie interfacce di rete wireless. L'unica risorsa esterna necessaria per la loro riuscita è la banda trasmissiva, ma molto spesso viene utilizzata la banda Industrial, Scientific and Medical (ISM) [79], che è senza vincoli di licenza. Terminali vicini fra loro possono comunicare direttamente sfruttando ad esempio tecnologie wireless LAN, mentre apparecchi che non sono in comunicazione diretta possono farlo inoltrando il traffico attraverso una sequenza di ricevitori intermedi.

Le Vehicular Ad hoc NETWORKS (VANET) rappresentano una particolare tipologia di rete mobile che consente la comunicazione tra veicoli in movimento e che ha tra i maggiori scopi quello di fornire sicurezza e comfort per i passeggeri durante la guida. Se i nodi mobili sono vicini, possono comunicare direttamente tra loro, altrimenti inoltrano il traffico ad una serie di dispositivi intermedi che inoltrano il traffico fino a destinazione. Inoltre tali nodi possono anche accedere ai servizi Internet, nel caso sia presente un dispositivo che funga da Internet gateway. Tutti questi vantaggi fanno delle reti mobili ad hoc il futuro delle reti wireless.

In campo aperto, senza tener conto delle problematiche introdotte da un ambiente cittadino, le tematiche di routing e forwarding dei pacchetti fra i nodi componenti una VANET sono relativamente semplici da gestire con tecniche ampiamente

descritte in letteratura e studiate da alcuni anni.

Lo scenario però si complica fortemente se vogliamo implementare le stesse tecniche in ambito cittadino, in quanto la topologia della mappa sottostante diventa fortemente vincolante. Non si può infatti più supporre di spedire liberamente pacchetti in qualsiasi direzione con le stesse proprietà di diffusione, ma bisogna tener conto della disposizione delle strade e dei palazzi che si trovano fra essi. Il routing diventa quindi strettamente dipendente da queste due componenti, e un maggior numero di informazioni a riguardo porta ad analisi e scelte più accurate e performanti. Infatti, i nodi che compongono la rete (i veicoli appunto) possono trovarsi soltanto sulle strade, ed i segnali radio scambiati non possono passare attraverso gli strati di cemento che formano i perimetri dei palazzi. Vedremo nel seguito come routing di tipo geografico necessitano di accurate informazioni per prendere le decisioni, e come l'aggiunta di informazioni di alto livello alle mappe disponibili al momento possa migliorarne le prestazioni sotto diversi aspetti.

Guardando oltre alle VANET, la tendenza attuale del mercato mostra un crescente bisogno di informazioni geo-referenziate aggiornate come componente chiave nel design di sistemi e servizi mobili, come veicoli interconnessi, percezione partecipativa, social networks mobili o personal fitness [43, 44].

Il sistema di navigazione introdotto da Google sul suo sistema operativo mobile Android [80] per esempio, basato su Google Maps, è uno dei primi servizi di navigazione per smartphones basato su prosumer. Utenti compartecipativi caricano dati ricavati dal GPS del proprio telefono sui server Google, in maniera trasparente. Questi contenuti forniti dagli utenti sono utilizzati nella stima del traffico attuale, congiuntamente a sorgenti di dati più tradizionali come sensori installati sulla superficie stradale, telecamere ecc [45, 46].

Il modello di business attuale, basato su una manciata di compagnie che pagano in anticipo gli alti costi necessari a costruire un database geografico globale, non è più in grado di sostenere la crescente domanda di nuove caratteristiche da associare alle mappe, come traffico in tempo reale, punti di interesse, recensioni degli utenti ed altro.

Navteq e Teleatlas, i due maggiori provider di mappe per la navigazione con strumentazione fissa o mobile, sono obbligati ad affidarsi gradualmente sempre più su dati forniti dagli utenti [47, 48]. Similmente, i provider dei servizi di navigazione veri e propri, come Tom Tom e Google, stanno utilizzando i dati forniti dagli

utenti per migliorare la precisione ed il dettaglio della propria offerta, tagliando al tempo stesso i costi di licenza per le mappe e riducendo la dipendenza dai fornitori cartografici [49]. Infine, i consumatori hanno guadagnato un ruolo essenziale nel processo di creazione di mappe digitali open source, come dimostrato dal progetto Open Street Map (OSM), un sito web progettato per costruire una mappa globale del pianeta, completamente editabile e gratuita, derivata da tracce GPS caricate da utenti di tutto il mondo [50].



Figura 1.2: Rappresentazione logica di un possibile sistema di nuova generazione che sfrutti le mappe e gli utenti in maniera bidirezionale

Gli apparecchi GPS introducono purtroppo per loro natura imprecisione e discontinuità nei dati registrati, rendendo l'identificazione delle caratteristiche cercate maggiormente difficoltosa. Per esempio, la campionatura del segnale può presentare discontinuità improvvise, cambi nella frequenza di campionamento, o ancora errori di localizzazione come la rilevazione di movimento anche quando un veicolo è in realtà fermo in coda a un semaforo. Si può pensare di superare tali difficoltà effettuando un'analisi correlata spazio/tempo sulla traccia GPS per identificare

pattern unici su ogni segmento stradale ed intersezione, migliorando la precisione dei dati registrati.

La figura 1.2 mostra l'architettura ad alto livello di un possibile sistema per una nuova generazione di servizi basati sulla geo-localizzazione, che sfruttino la natura dinamica delle applicazioni mobili.

Si è visto quindi quali possono essere gli scenari in cui va a collocarsi il sistema sviluppato, e quale sia l'importanza di avere informazioni accurate e stratificate sulle mappe digitali. Si procede quindi guardando ora più da vicino le reti mobili ad hoc, che si è visto essere alla base dello sviluppo dei nuovi scenari appena introdotti.

Capitolo 2

Mobile Ad hoc NETWORKS

In questo capitolo verranno introdotte le Mobile Ad hoc NETWORKS (MANET), partendo una panoramica storica che presenta i motivi principali che hanno portato al loro sviluppo e alla loro diffusione. Verranno poi mostrate le reti mobili di quarta generazione (4G), in cui le MANET assumono un ruolo da protagonista. Infine verrà fatta una descrizione delle caratteristiche delle reti mobili ad hoc e dei loro principali scenari applicativi.

2.1 Evoluzione storica

Le reti mobili ad hoc sono state utilizzate inizialmente in ambito militare per migliorare le comunicazioni in territori di guerra. La natura dinamica delle operazioni militari infatti spesso non consentiva la comunicazione tra i militari con mezzi tradizionali, poiché in molti territori di guerra mancava un'infrastruttura di comunicazione. Da qui, la nascita di reti wireless costituite da sensori senza un'infrastruttura centralizzata. I limiti della comunicazione wireless diretta tradizionale sono dovuti alle interferenze a cui sono soggetti i segnali radio e alla loro difficile propagazione oltre la line of sight (LOS) a frequenza maggiori dei 100 MHz [17]. Le reti

mobili ad hoc invece danno luogo ad una rete multi-hop, consentendo la connettività anche oltre la LOS.

Da un punto di vista storico, le reti ad hoc hanno mosso i primi passi nel progetto Packet Radio Network (PRNet) sviluppato dal Defense Advanced Research Projects Agency (DARPA) del 1972 [17]. Il progetto PRNet prevedeva un'architettura distribuita, con la presenza di diversi nodi nella rete ma con un minimo controllo centralizzato; per condividere il canale radio si fece ricorso ad una combinazione dei protocolli di accesso al mezzo Aloha [35] e CSMA [51]. Inoltre, usando le tecniche di routing multi-hop store-and-forward, si ebbe la possibilità di comunicare anche in aeree geografiche molto estese.

Nel 1983 DARPA sviluppò le Survivable Radio Networks (SURAN) per aggiungere alle caratteristiche di PRNet quelle relative alla scalabilità, alla sicurezza, alla capacità di processing e alla gestione dell'energia. Gli obiettivi erano lo sviluppo di algoritmi per consentire ad una rete di scalare in termini di decine di migliaia di nodi e al contempo di essere a basso costo, resistente agli attacchi, a basso consumo di energia e in grado di supportare protocolli Packet Radio complessi. Tutto ciò portò nel 1987 alla nascita della tecnologia Low-cost Packet Radio (LPR) [75]. Successivamente vennero sviluppati una serie di protocolli avanzati per il network management e venne introdotto il concetto di gerarchia di rete per migliorare la scalabilità.

Agli inizi degli anni 90 la crescita dell'infrastruttura Internet e lo sviluppo dei microcomputer portarono all'idea delle reti wireless. In particolare nel 1994 il Dipartimento della Difesa intraprese il programma DARPA Global Mobile (Glo-Mo) Information Systems, che mirava a fornire connettività multimediale attraverso dispositivi wireless in qualsiasi momento e in qualsiasi luogo [76].

Molti furono i progetti sviluppati successivamente, tra cui il progetto Wireless Internet Gateways (WINGs) dell'Università della California di Santa Cruz (UCSC), che prevedeva un'architettura di rete peer-to-peer, e il progetto Multimedia Mobile Wireless Network (MMWN) del GTE Internetworking, che si basava su un'architettura di rete gerarchica per migliorare la scalabilità. Nel 1997 l'US Army realizzò il progetto Tactical Internet (TI), che poteva essere visto come la più grande implementazione di una rete wireless ad hoc multi-hop [17]. Nel 1999, il Corpo della Marina sviluppò il progetto Extending the Littoral Battle-space Advanced Concept Technology Demonstration (ELB ACTD), una rete wireless che consentiva la co-

municazione wireless tra le navi in mare e la Marina a terra attraverso una rete di 20 nodi opportunamente configurati.

A metà degli anni '90, con la definizione di standard come l'IEEE 802.11 [57], il mondo del wireless entrò definitivamente nel mercato consumer e i ricercatori iniziarono a volgere la propria attenzione sui possibili vantaggi che si potevano ottenere con l'uso delle reti mobili wireless ad hoc anche fuori dall'ambito militare.

2.2 Le reti 4G

Uno dei principali obiettivi a sospingere l'evoluzione delle reti di quarta generazione (4G) è il fornire un ambiente computazionale pervasivo, che possa supportare in ogni luogo ed istante l'utente nel completare i suoi compiti, nell'accedere ad informazioni o nel comunicare con altri utenti [32]. In un simile ambiente, i tradizionali computer vengono spinti fortemente sullo sfondo; potenza di calcolo e connettività di rete sono integrati virtualmente in ogni dispositivo, per portare possibilità di computazione agli utenti, senza preoccuparsi di dove siano o sotto quali circostanze stiano lavorando. Queste nuove periferiche si personalizzano in nostra presenza per fornirci ciò di cui necessitiamo.

La nuova tendenza è aiutare gli utenti nei compiti di ogni giorno utilizzando tecnologie ed infrastrutture a lui trasparenti, senza richiedere alcun cambiamento rilevante nei loro comportamenti e nelle loro abitudini. Questa nuova filosofia è alla base dell'idea di Ambient Intelligence [2]. La parola chiave di questa intelligenza ambientale è l'integrazione: integrazione fra dispositivi digitali di vario genere nella vita di tutti i giorni, al fine di rendere accessibili, tramite interazioni semplici e naturali, una moltitudine di servizi ed applicazioni. L'utente è messo al centro della società dell'informazione, non ci interagisce più soltanto come un agente esterno.

Questa visione poggia pesantemente sulle reti 4G e sulle comunicazioni in mobilità. 4G è infatti incentrato totalmente su una rete globale integrata, basata su un approccio di sistema aperto. Integrare fluidamente differenti tipi di reti wireless con una backbone cablata, facendo convergere traffico voce, multimediale e di dati su una singola rete centrale basata su IP, è l'obiettivo principale. Con la disponibilità di una banda fino a 100Mbps, i servizi multimediali possono essere supportati efficientemente.

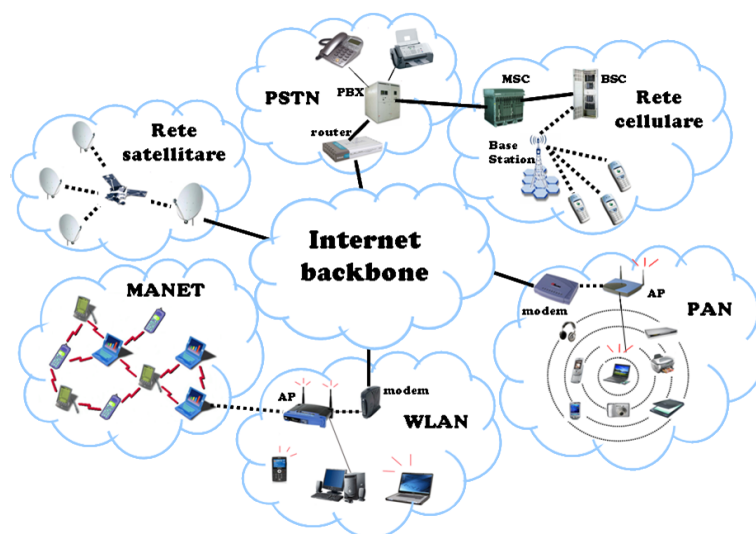


Figura 2.1: Schema di una rete 4G

La figura 2.1 mostra le reti ed i componenti all'interno di un'architettura di rete 4G, i cui punti di forza possono essere riassunti in:

- Integrazione di reti differenti.
- Condivisione del protocollo IP fra tutte le reti.
- Riduzione dei costi ed aumento dell'efficienza.
- Fornitura di connessione ad alta velocità e applicazioni multimediali.
- Sfruttamento intelligente della località.

Integrazione di reti differenti Le reti 4G sono famose come reti ibride a banda larga che integrano differenti topologie e piattaforme di rete, come mostrato in figura 2.1. Vi sono due livelli di integrazione. Il primo è l'integrazione di reti wireless eterogenee con varie caratteristiche trasmissive come Wide Area Networks (WAN), Wireless Local Area Networks (WLAN), Personal Area Networks (PAN) o Mobile Ad hoc NETWORKS (MANET). Al secondo livello invece troviamo l'integrazione di reti wireless con con l'infrastruttura di rete di backbone, Internet e PSTN. Molto lavoro rimane però da fare per garantire coesione fra esse, per esempio estendendo il protocollo IP a tutti i dispositivi mobili.

Condivisione del protocollo IP fra tutte le reti Il 4G parte con l'assunzione che le reti del futuro saranno interamente a commutazione di pacchetto, usando protocolli evoluti da quelli in uso nell'internet odierno [26]. Una rete interamente basata

su IP ha dei vantaggi intrinseci rispetto al passato. IP è al contempo compatibile ed indipendente con l'attuale tecnologia di accesso radio, il che significa che il nucleo della rete 4G può essere progettato ed evolversi in maniera indipendente dalle reti di accesso vere e proprie.

È possibile inoltre sfruttare direttamente il vasto insieme di protocolli e servizi già disponibili, ad esempio la convergenza di traffico voce e dati può essere gestita con un insieme di protocolli già disponibili per il VoIP, come MEGACOP, MGCP, SIP, H.323, SCTP, ecc [59]. Questa evoluzione verso una totale convergenza porterà a una notevole semplificazione della rete ed una riduzione dei costi dovuti al mantenimento di reti diverse per tipologie di traffico diverse.

Riduzione dei costi ed aumento dell'efficienza Le reti 4G basate su IP saranno più economiche ed efficienti delle odierne 3G. Prima di tutto, le attrezzature necessarie sono dalle 4 alle 10 volte meno costose delle controparti equivalenti per reti 2G o 3G. Un ambiente aperto pienamente convergente verso IP inoltre riduce costi di creazione e mantenimento della rete. Non sarà necessario neanche acquisire nuovo spettro frequenziale, in quanto può essere riutilizzato quello del 2G/3G, e la maggioranza delle frequenze necessarie a WLAN e WPAN è pubblica e non richiede licenza.

Fornitura di connessione ad alta velocità e applicazioni multimediali I sistemi 4G mirano a fornire velocità di trasmissione elevate, fino a 100Mbps in movimento e 1000Mbps da fermi, 50 volte maggiori di quelle 3G. Questo balzo nella banda fornita permetterà servizi mobili a banda larga, permettendo agli utenti di guardare la TV, ascoltare la musica, navigare in Internet, accedere a server e programmi aziendali, fare videoconferenze o qualsiasi altra applicazione multimediale, come se fossero comodamente seduti sul divano di casa o alla scrivania del proprio ufficio.

Sfruttamento intelligente della località Per supportare i requisiti dell'Ubiquitous Computing, i terminali 4G devono essere più intelligenti in termini di identificazione della posizione e dei bisogni dell'utente, non limitandosi al riconoscere che essi varino la propria posizione, ma adattandosi a questo cambiamento, offrendo sempre servizi location-based aggiornati [5]. Esempi di servizi location-based sono la ricerca del ristorante o del cinema più vicino, di offerte speciali nella propria

zona, di altri utenti; o ancora, la ricezione di informazioni su situazioni di traffico o meteorologiche pericolose nei paraggi o l'invio di pubblicità localizzate (figura 2.2).

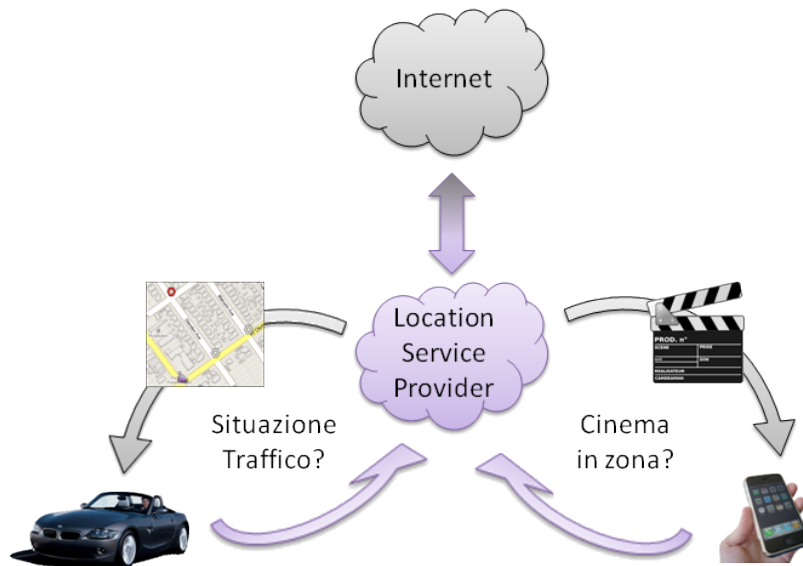


Figura 2.2: Esempi di servizi che sfruttano in maniera intelligente la località dell'utente

All'aperto, le applicazioni wireless possono utilizzare il GPS per ottenere le informazioni di posizione. Il GPS è un sistema satellitare che può fornire informazioni posizionali facili e discretamente precise quasi ovunque sulla terra. Diverse implementazioni GPS sono disponibili, inclusa la sua integrazione in cellulari e smartphones, spesso nella forma di A-GPS o E-OTD. Per applicazioni al chiuso invece, dato che il segnale GPS non può essere ricevuto attraverso gli edifici, sono studiate altre tecnologie come Infrarossi, Ultrasuoni o Onde Radio.

Le MANETs stanno prendendo piede in questo contesto 4G perchè concorrono a realizzare servizi di rete per utenti mobili in aree prive di infrastruttura precedente, o quando l'utilizzo di tale infrastruttura richiede un'estensione wireless [13, 18]. I nodi ad-hoc possono anche collegarsi ad una backbone attraverso un gateway dedicato, che abilita i servizi IP in aree in cui i servizi Internet non sono disponibili a causa di una mancanza dell'infrastruttura preinstallata. Tutti questi vantaggi rendono le reti ad hoc un'opzione molto invitante per le future reti wireless.

2.3 Caratteristiche delle MANET

In generale, le reti mobili ad hoc sono formate dinamicamente da sistemi autonomi di nodi mobili, connessi da link radio, senza utilizzare un'infrastruttura di rete o un'amministrazione centralizzata. I nodi sono liberi di muoversi casualmente e organizzarsi fra loro arbitrariamente; per questo, la topologia della rete può cambiare rapidamente e in maniera imprevedibile. Una rete di questo tipo può operare in maniera indipendente, o può essere collegata ad Internet. Le MANET sono reti prive di infrastruttura poichè non richiedono alcuna infrastruttura fissa, come ad esempio una base station, per il loro funzionamento.

In generale, i percorsi dei pacchetti fra i nodi possono includere il passaggio per più di un nodo intermedio, per questo è appropriato chiamare queste reti anche Multi-hop Wireless Ad Hoc Networks. Ogni nodo può comunicare direttamente con gli altri nodi che siano nel suo raggio di trasmissione, mentre per comunicare con nodi più distanti deve sfruttare nodi intermedi su cui far rimbalzare i messaggi passo dopo passo.

Attraverso opportune configurazioni è possibile mettere in comunicazione reti mobili ad hoc differenti come mostrato in figura 2.3.

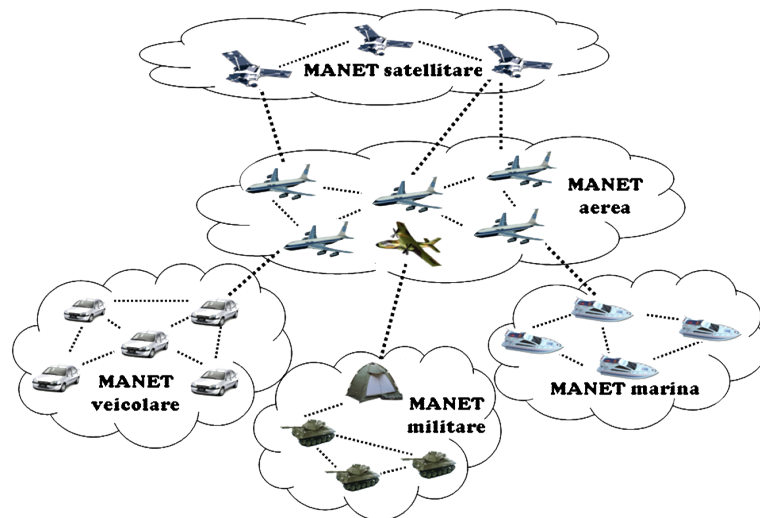


Figura 2.3: Comunicazione fra reti mobili ad hoc di natura differente

La comodità e flessibilità delle reti ad hoc non vengono però a costo zero. Esse infatti ereditano gli stessi problemi delle connessioni wireless in genere [22]:

- Il mezzo wireless non ha confini osservabili né tantomeno ben definiti. all'in-fuori dei quali si sa con certezza che i pacchetti non saranno ricevibili.
- Il canale non è protetto dai segnali esterni.
- Il canale wireless è significativamente meno affidabile che una connessione cablata.
- Il canale ha proprietà di propagazione dipendenti dal tempo ed asimmetriche.
- I fenomeni di interferenza noti come hidden-terminal ed exposed-terminal sono possibili.

A questi problemi, alla natura multi-hop e alla mancanza di un'infrastruttura fissa si aggiunge poi un numero di caratteristiche, gradi di complessità e limiti di progettazione specifici delle reti ad hoc [13, 14]:

- Autonomia e assenza di infrastruttura.
- Routing multi-hop.
- Cambiamenti dinamici della topologia di rete.
- Variazione della capacità dei link e dei nodi.
- Limitazione nell'uso dell'energia.
- Scalabilità.

Autonomia e assenza di infrastruttura Le MANET non dipendono su alcuna infrastruttura o amministrazione centralizzata. Ogni nodo opera in modalità peer-to-peer distribuita, agendo come un router e generando pacchetti indipendenti dagli altri. La gestione di rete deve essere distribuita sui diversi nodi, il che aggiunge complessità nel riconoscimento degli errori e nella gestione stessa.

Routing multi-hop Non c'è un router di default, ogni nodo agisce come router e inoltra ogni pacchetto che riceve per abilitare la condivisione di informazioni fra i diversi hosts della rete.

Cambiamenti dinamici della topologia di rete Poiché i nodi possono muoversi arbitrariamente, la topologia di rete, che è tipicamente multi-hop, può cambiare frequentemente ed in maniera imprevedibile, riflettendosi in cambi di rotta dei pacchetti, frequenti partizioni della rete e possibile perdita di pacchetti.

Variazione della capacità dei link e dei nodi Ogni nodo può essere equipaggiato con una o più interfacce radio che hanno capacità di ricezione e trasmissione variabili, ed operano su diverse frequenze di banda [11, 12]. Questa asimmetria nella capacità radio dei nodi può risultare in collegamenti asimmetrici fra i nodi. Inoltre, ogni nodo può avere una configurazione hardware / software specifica, che genera una capacità computazionale differente.

Progettare protocolli ed algoritmi di rete per una rete così eterogenea è complesso e richiede un'adattività dinamica alle condizioni del sistema (potenza e condizione del canale, variazioni del carico e della distribuzione del traffico, congestioni, ecc.).

Limitazione nell'uso dell'energia Poiché le batterie dei dispositivi mobili hanno un'energia finita, la potenza di calcolo è limitata a sua volta, il che limita servizi ed applicazioni che possono essere supportati da ogni nodo. Questo diventa un problema ancora più grande nelle reti mobili ad hoc perché, dato che ogni nodo si comporta contemporaneamente sia da dispositivo terminale che da router, è richiesta energia aggiuntiva per inoltrare i pacchetti provenienti da altri nodi.

Scalabilità Ad oggi, i comuni algoritmi di gestione delle reti sono disegnati per lavorare su reti fisse o wireless relativamente piccole. Molte applicazioni delle MANET invece coinvolgono vaste reti, con decine di migliaia di nodi, come ad esempio nelle reti di sensori o nelle reti tattiche militari [17]. La scalabilità è quindi un punto critico nel loro sviluppo.

Il passo verso una grande rete formata da nodi con risorse limitate non è semplice, ma presenta diverse sfide che devono ancora essere superate in aree come: indirizzamento, routing, gestione della località, gestione delle configurazioni, interoperabilità, sicurezza, sviluppo di tecnologie wireless ad alta capacità, ecc.

2.4 Scenari Applicativi

Le prime applicazioni e implementazioni di reti mobili ad hoc, come avviene per molte nuove tecnologie all'inizio complesse e molto costose, sono state in ambito militare con le Tactical Networks [72]. Solo l'introduzione sul mercato di massa

di nuove tecnologie a basso costo come il Bluetooth, l'IEEE 802.11 e le Hyperlan hanno facilitato lo sviluppo delle reti ad hoc fuori dal contesto militare e nuove applicazioni per reti ad hoc sono apparse nelle aeree riguardanti:

Reti Veicolari Reti che si vengono a costituire quando due o più veicoli, dotati della opportuna elettronica di bordo, si trovano l'uno in prossimità dell'altro. Le auto, stabilendo tra loro una connessione ad hoc, possono scambiarsi informazioni sul traffico, sulle condizioni della strada o dell'ambiente, su eventuali incidenti e pericoli, ecc.

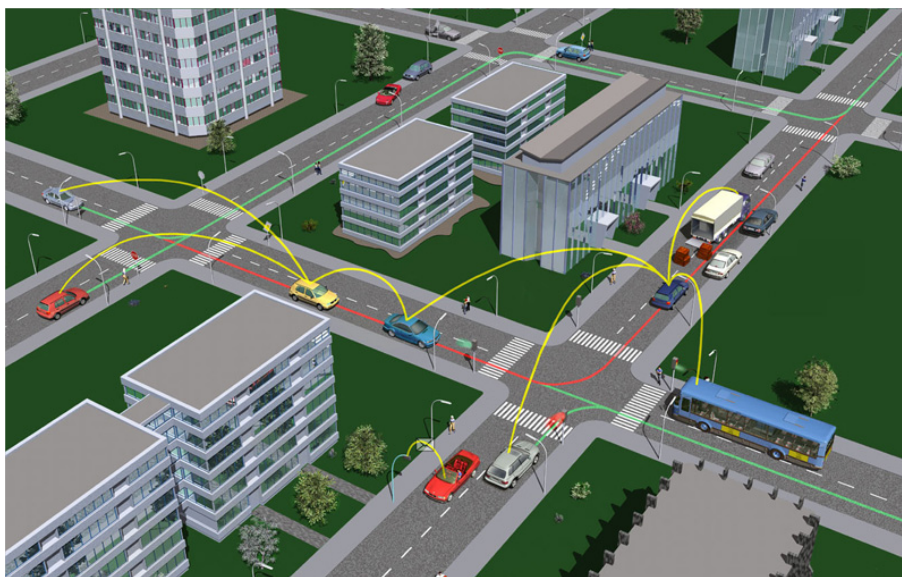


Figura 2.4: Esempio di rete veicolare utilizzata in caso di incidente: in giallo si vede come è propagata l'informazione, in rosso le traiettorie che i veicoli avrebbero seguito prima della notifica dell'incidente, ed in verde le nuove traiettorie calcolate per evitarlo

Un esempio è mostrato nella figura 2.4: grazie alla rete ad hoc tra i veicoli, i conducenti vengono a conoscenza dell'incidente appena avvenuto e scelgono percorsi differenti, evitando la strada in cui è avvenuto l'incidente e permettendo al tempo stesso alle forze dell'ordine e ai soccorsi di giungere più velocemente sul posto. Reti di questo tipo permettono anche nuove frontiere per il mondo dell'intrattenimento, con lo scambio di musica o filmati in base ai gusti degli utenti della rete magari, così come possibilità nuove per il campo videoludico. Si può infatti portare in auto l'esperienza di gioco on-line a cui siamo ormai sempre più abituati in casa.

La recente esperienza, due casi su tutti: iPhone e Nintendo Wii, ha insegnato come qualsiasi cosa mostri una possibile innovazione dal punto di vista dell'esperienza utente, abbia un enorme consenso da parte del mercato, che la porta ad avere una crescita esponenziale, anche a discapito di altre tecnologie o concorrenti tecnicamente più valide o economiche.

Reti di Sensori Reti di questo tipo trovano la loro applicazione sia in campo domestico, come la gestione di diversi dispositivi della casa in maniera locale o remota da parte degli utenti, sia ambientale, come ad esempio tracciare i movimenti di animali, individuare la presenza di sostanze chimiche in fiumi e laghi o rilevare specifiche condizioni atmosferiche e attività terrestri che possono essere preludio di terremoti o eruzioni vulcaniche come mostrato in figura 2.5.

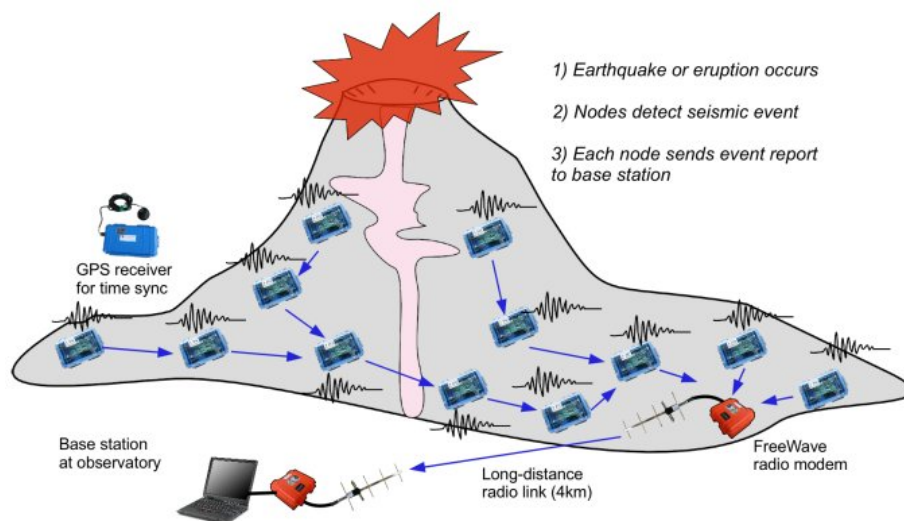


Figura 2.5: Esempio di reti di sensori utilizzata per acquisire dati prima e durante un'eruzione vulcanica

Servizi di emergenza Operazioni di ricerca e recupero di persone smarrite o ferite in situazioni di disastro o calamità possono trarre vantaggio da queste reti ad hoc, così come in caso l'infrastruttura esistente venga meno e non si possa più utilizzare per comunicare, le MANET potrebbero permettere di mantenere in piedi un sistema di comunicazione funzionante.

Servizi Location Aware Per ottenere informazioni dinamiche riguardo la mia posizione (figura 2.6), come ad esempio gli hotel o i ristoranti più vicini, o le stazioni di benzina con self service aperte. O ancora, ottenere la guida turistica dinamica della specifica zona che sto visitando, evitando di dover girare con moltitudini di guide cartacee.

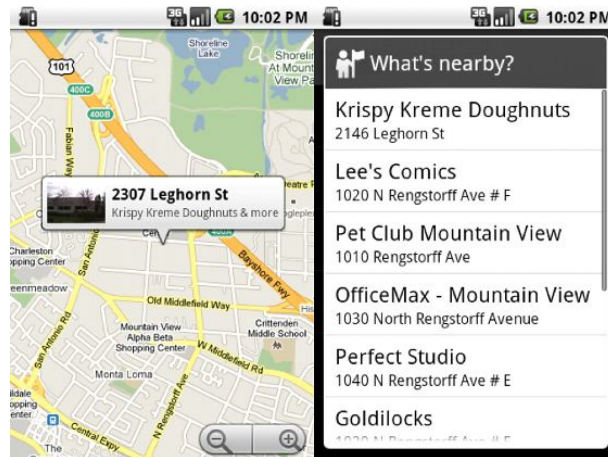


Figura 2.6: Esempio di servizio localizzato, che propone all'utente i punti di interesse vicini a lui

Intrattenimento Giochi multi utente in ogni circostanza, animali domestici robot che possono evolvere il loro comportamento in funzione delle informazioni di rete che ottengono, estensione della connettività a cui siamo abituati in casa anche all'esterno delle mura domestiche.

Capitolo 3

Ad-Hoc Networks

All'inizio di questo capitolo vengono presentate le tecnologie attualmente presenti sul mercato che riguardano le reti wireless, con particolare interesse per i due standard dominanti, Bluetooth e IEEE 802.11. Sono poi mostrate le possibili soluzioni di routing e forwarding note in letteratura per le reti ad hoc, con particolare attenzione al routing location-aware.

3.1 Tecnologie Attuali

Come mostrato in figura 3.1 possiamo classificare le reti ad hoc a seconda della loro area di copertura: Body Area Network (BAN), Personal Area Network (PAN), Wireless Local Area Network (WLAN), Metropolitan Area Network (MAN) e Wide Area Network (WAN). WAN e MAN sono reti mobili ad hoc multi-hop che presentano molte problematiche ancora da risolvere, come l'indirizzamento, il routing, la gestione di località e sicurezza, ecc.

Sul fronte opposto, ci si può attendere che reti mobili ad hoc con copertura inferiore siano presto una realtà. Infatti, reti single-hop BAN, PAN e WLAN sono già una realtà comune sul mercato [8] e costituiscono i mattoni con cui costruire

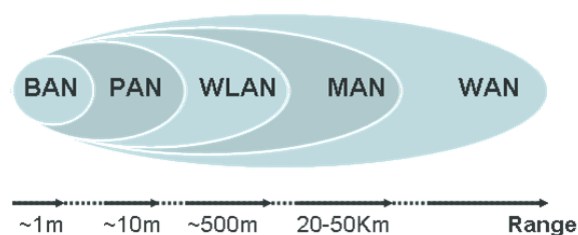


Figura 3.1: Classificazione delle attuali reti ad hoc in base al loro raggio di copertura

piccole reti multi-hop [13]. Una discussione dettagliata di questi tre tipi di reti può essere trovata in [8].

Una *Body Area Network* è fortemente correlata con i computer indossabili. Un computer indossabile distribuisce sul corpo i suoi componenti (es. display montato su occhiali, microfono, cuffie, ecc.), e la BAN fornisce connettività fra essi. Il suo raggio d'azione infatti corrisponde al corpo umano, non superando i 2 metri. Poiché cablare il corpo di una persona è solitamente scomodo ed ingombrante, le soluzioni wireless sono le migliori per interconnettere dispositivi indossabili.

Le *Personal Area Networks* connettono dispositivi trasportati dagli utenti con altri dispositivi mobili e stazioni base. Mentre una BAN è votata all'interconnessione di dispositivi di una stessa persona, una PAN è una rete che si estende nell'ambiente circostante l'individuo. Il suo raggio d'azione è intorno ai 10 metri, permettendo quindi la connessione fra BAN di persone vicine fra loro. L'utilizzo più comune di una PAN è la connessione di dispositivi come PDA, telefoni cellulari, computer portatili, stampanti o proiettori come mostrato in figura 3.2. La frequenza di banda solitamente utilizzata è quella ISM attorno ai 2.4 GHz.

Una *Wireless Local Area Network* ha un raggio comunicativo tipico di un singolo palazzo, o un piccolo blocco di palazzi, variando dai 100 ai 500 metri. Una WLAN dovrebbe soddisfare tutti i requisiti tipici di una LAN, compresa alta capacità, piena connettività fra i nodi collegati e capacità di broadcast. Per soddisfare questi obiettivi, le WLAN devono essere progettate per affrontare problematiche specifiche delle reti wireless, come la sicurezza del canale radio, il consumo di energia, la mobilità e la limitatezza di banda disponibile [30].

Il successo di una tecnologia è strettamente correlato allo sviluppo di dispositivi e prodotti ad un prezzo competitivo. Un fattore importante per ottenerli è la disponibilità di standard appropriati. Al momento, due standard principali dominano il



Figura 3.2: Esempio di *Personal Area Network*

settore: l'IEEE 802.11 per le WLAN [23] ed il Bluetooth [6] per le comunicazioni a breve raggio [3, 7, 27]. In aggiunta agli standard IEEE, l'*European Telecommunication Standard Institute* (ETSI) ha promosso una famiglia di standard per le WLAN chiamata HiperLAN [16] (*High Performance Radio LAN*). Fra questi, il più interessante è l'HiperLAN/2, che ha un'elevata velocità trasmissiva (dai 6 ai 54 Mbit/s) e supporta sia la modalità di connessione con infrastruttura sia quella ad hoc [15].

3.1.1 Bluetooth

La tecnologia Bluetooth si è affermata come uno standard de-facto per connessioni a breve raggio e basso costo fra computer portatili, telefoni cellulari ed altri dispositivi mobili [3, 27]. Le specifiche bluetooth sono rilasciate dal Bluetooth Special Interest Group (SIG), che riunisce lo sforzo congiunto di più di due mila ditte leader come 3Con, Ericsson, IBM, Intel, Microsoft, Lucent, Motorola, Nokia, Toshiba, ecc [6].

Un'unità bluetooth, integrata dentro un microchip, abilita connessioni wireless ad-hoc dati e voce fra dispositivi mobili o fissi [20]. Collegando ad esempio un microfono ed un auricolare bluetooth ad un telefono cellulare, è possibile per un guidatore rispondere ad una telefonata mantenendo entrambe le mani sul volante senza compromettere la sicurezza della guida, come mostrato in figura 3.3.

Grazie al suo costo ridotto, i chip bluetooth potrebbero essere a breve integrati in ogni dispositivo elettronico disponibile sul mercato. Come soluzione low cost a basso consumo energetico ma con esteso supporto industriale, la tecnologia blue-



Figura 3.3: Utilizzo di auricolare e microfono bluetooth per telefonare senza compromettere la sicurezza alla guida

tooth ha già iniziato a rivoluzionare il mercato della connettività, fornendo la libertà dai cavi che tutti gli utenti han sempre desiderato da quando miriadi di periferiche differenti han cominciato ad invadere il mercato, ognuna col suo tipico groviglio di cavi di collegamento.

I sistemi bluetooth possono gestire un piccolo numero di connessioni point-to-point e point to multi-point a basso costo su una distanza fino a 10 metri, con una potenza trasmittiva inferiore a 1mW. Essa opera globalmente nella frequenza ISM di 2.4 GHz, e applica una combinazione di tecnologie a commutazione di circuito e di frequenza per la trasmissione dei dati.

In una rete bluetooth, una base ha il ruolo di master, e tutti gli altri dispositivi sono invece slaves. Il master decide quale slave ha di volta in volta accesso al canale radio. Le unità che condividono lo stesso canale (sono cioè sincronizzate con lo stesso master) formano una *piconet*, che è il blocco fondamentale per costruire una rete bluetooth più ampia. Una piconet ha un bit rate di 1 Mbit/s, che rappresenta la capacità del canale ed include l'overhead introdotto dai protocolli e lo schema a polling utilizzato. Una piconet contiene un master e fino a 7 slaves *attivi* contemporaneamente (cioè che partecipano attivamente allo scambio di dati).

Dentro una piconet, le basi bluetooth possono stabilire fino a tre canali sincroni (voce) a 64 Kbit/s, o un canale dati asincrono che supporta un data rate di 723 Kbit/s se asimmetrico o 433 Kbit/s se simmetrico. Una presentazione maggiormente dettagliata a riguardo si può trovare in [3, 8, 27], mentre le performance complessive di una piconet sono investigate dettagliatamente in [4].

Una piconet costituisce una rete single hop. Interconnettendo più piconets si può invece ottenere una rete multi hop. Le specifiche bluetooth definiscono un metodo per la loro interconnessione: le *scatternet*. Nonostante le specifiche bluetooth definiscano la nozione di scatternet, non forniscono i meccanismi per costruirle. Una scatternet può essere dinamicamente costruita, in modalità ad hoc, quando alcuni nodi appartengono nello stesso momento a più di una piconet. Il traffico fra due piconet viene quindi recapitato attraverso i nodi comuni.

3.1.2 IEEE 802.11

Nel 1997 l'IEEE ha adottato il primo standard WLAN, chiamato 802.11, con data rates fino a 2 Mbps [21]. Da allora, diversi gruppi di specifiche (identificati dalle lettere a, b, c, ecc.) sono stati creati per estendere lo standard originale. Le estensioni 802.11b e 802.11a sono state molto rilevanti [23], e sono spesso identificate dal nome Wireless Fidelity (Wi-Fi).

L'802.11b estende lo standard portando il data rate fino a 11 Mbps su una frequenza di banda di 2.4 GHz, mantenendo però la retrocompatibilità. Questo standard, pubblicato nel 1999, è diventato un successo globale, con svariati prodotti disponibili sul mercato. L'802.11a invece opera sulla frequenza di 5 GHz e fornisce una velocità fino a 54 Mbps utilizzando la modulazione di tipo QAM-64; due anni dopo, l'802.11g prese il suo posto, offrendo le stesse prestazioni, ma operando alla frequenza standard di 2.4 GHz.

Lo standard 802.11, indipendentemente dalla sua specifica estensione, permette due modalità operative: infrastrutturata e ad hoc. Le interfacce di rete possono essere impostate per lavorare in entrambe le modalità, ma non contemporaneamente. La modalità infrastrutturata è simile a quella usata per le reti cellulari, ed è quella comunemente usata per costruire spot Wi-Fi con cui fornire accesso internet wireless (figura 3.4). Nella modalità ad hoc, ogni dispositivo che sia nel raggio d'azione di un altro, dopo una prima fase di sincronizzazione, può iniziare a comunicare con esso. Non è richiesto alcun access point, ma se uno dei due è connesso anche ad una rete cablata, tutti i dispositivi che sono parte della rete ad hoc guadagnano accesso wireless ad internet.

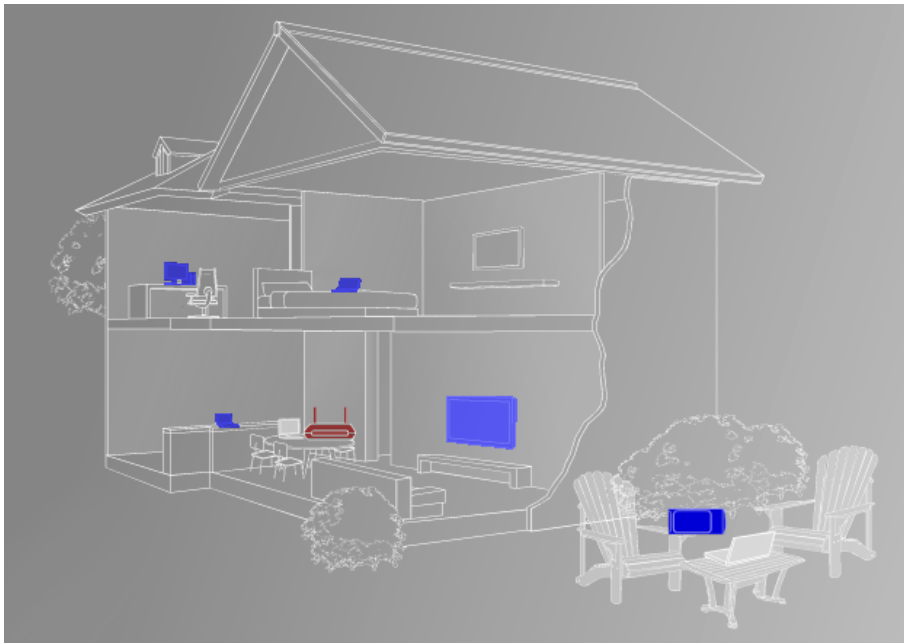


Figura 3.4: Esempio di spot Wi-Fi che fornisce accesso ad internet a diversi dispositivi di casa

Wi-Fi Abbreviazione di Wireless Fidelity, è un termine che indica dispositivi che possono collegarsi a WLAN basate sulle specifiche IEEE 802.11. Un dispositivo, anche se conforme a queste specifiche, non può utilizzare il logo ufficiale Wi-Fi se non ha superato le procedure di certificazione stabilite dal consorzio Wi-Fi Alliance (Wireless Ethernet Compatibility Alliance), che testa e certifica la compatibilità dei componenti wireless con gli standard della famiglia 802.11.

La presenza del marchio Wi-Fi su di un dispositivo garantisce quindi l'interoperabilità con gli altri dispositivi certificati, anche se prodotti da aziende differenti. Lo standard 802.11 prevede che la velocità trasmissiva venga adattata dinamicamente sulla base del rapporto segnale/rumore. Infine per evitare interferenze fra dispositivi radio che sfruttano la stessa banda, si ricorre a tecniche di spread spectrum, che consistono nel distribuire il segnale su una banda molto più larga del necessario, in modo che esso appaia come rumore ai dispositivi non interessati.

3.2 Location Services

Un Location Service risponde a query sulla posizione dei nodi, associando dinamicamente l'indirizzo logico di un dispositivo con la sua attuale posizione fisica. In

reti mobili legacy [25] (es. GSM), la presenza di un'infrastruttura fissa ha portato alla diffusione di schemi a due livelli per tener traccia della posizione dei nodi mobili. Esempi sono l'Home Location Register / Visitor Location Register delle reti GSM, o l'approccio ad Home Agent / Foreign Agent delle reti a Mobile IP. L'implementazione efficiente di questi approcci usa server centralizzati. Chiaramente, queste soluzioni si rivelano inutili in una rete mobile ad hoc, e nuovi approcci devono essere studiati [28].

Una semplice soluzione per la localizzazione dei nodi è basata sul *flooding* della query attraverso la rete. Chiaramente, il flooding non è un metodo scalabile, per cui questo approccio si rivela utilizzabile soltanto per reti di piccola dimensione, dove l'impatto sulle performance rimane limitato. Limitare l'area di flooding permette di raffinare un po' la tecnica, ma non basta. Il flooding è un approccio di tipo *reactive*, in cui non sono mantenute informazioni all'interno della rete fra una query e l'altra. Il costo di mantenimento del location service è trascurabile, e tutta la complessità è associata alle operazioni di query.

Opposto a questo, è l'approccio di tipo *proactive*, che suddivide la complessità nelle due fasi. Servizi proattivi costruiscono nella rete strutture dati che memorizzano le informazioni sulla posizione di ogni nodo, e sfruttando queste strutture dati le operazioni di query sono molto semplificate.

3.3 Routing e Forwarding in reti ad hoc

Le reti ad hoc hanno, come visto, due caratteristiche fondamentali:

- Autoconfigurazione: le reti ad hoc si costituiscono o sciolgono in modo automatico quando esiste o meno una determinata densità di nodi. Il movimento dei nodi è imprevedibile, e pertanto la rete deve essere in grado di reagire dinamicamente alla comparsa di nuovi elementi e alla scomparsa di altri.
- Capacità di instradamento dinamico e multi-hop: poiché a topologia fortemente variabile, la rete ad hoc deve essere dotata di soluzioni che permettano di calcolare il percorso migliore verso ciascuno dei suoi nodi nel numero ottimale di passi (hop), adattandosi rapidamente alle variazioni di configurazione.

Questi due aspetti aggiungono ulteriore complessità ai protocolli di routing, che dovrebbero avere anche una serie di funzionalità per ottimizzare le risorse limitate

e per gestire la scalabilità, la sicurezza, l'affidabilità, il supporto alla qualità del servizio (QoS), ecc.

Una prima classificazione, che non tiene ancora conto della mobilità, suddivide i protocolli di routing in Unicast, Geocast, Multicast, o Broadcast [29]. Il *Broadcast* è il metodo operativo base su un canale wireless, in cui ogni messaggio è ricevuto da tutti i vicini che distano un hop da chi l'ha trasmesso. Al contrario con *Unicast* si prevede una comunicazione one-to-one, verso quindi un solo destinatario.

Il *Multicast* entra in gioco quando più nodi vogliono ricevere lo stesso messaggio o lo stesso flusso di dati, entrando a far parte di un particolare gruppo multicast al quale la sorgente trasmette. I protocolli di routing multicast lavorano bene in condizioni di rete statica; sono in fase di studio numerose soluzioni per rendere possibile il forwarding multicast anche quando la topologia della rete è altamente dinamica come avviene nelle mobili reti ad hoc [10]. Il forwarding *Geocast*, infine, è utilizzato per trasmettere pacchetti a nodi che sono situati in una particolare area geografica.

Nelle reti mobili ad hoc i protocolli di routing sono in genere suddivisi in due grandi categorie: *proactive* e *reactive* (o *on-demand*). I protocolli di routing *proactive*, che derivano dai protocolli Distance-Vector e Link-State (figura 3.5), mantengono nella rete le informazioni di routing che vengono aggiornate propagando periodicamente degli update. Poiché di solito le informazioni sono mantenute all'interno di tabelle, spesso questi protocolli sono anche detti Table-Driven. La mobilità è un fattore molto critico per questo tipo di protocolli perché i cambiamenti della topologia della rete provocano una serie di update che, oltre ad aumentare l'overhead, possono dar vita a loop, impedendo il corretto forwarding dei pacchetti. Più i cambiamenti della topologia di rete sono frequenti, più i messaggi di update rischiano di diventare dominanti nel traffico di rete, finendo per utilizzare tutte le risorse che invece dovrebbero essere destinate al traffico dati effettivamente voluto.

Dall'altro lato, i protocolli *reactive on-demand* stabiliscono la route verso una destinazione solo quando è realmente necessario, ovvero c'è una richiesta effettiva per raggiungerla. Il nodo sorgente, attraverso un location service che fa partire il processo di route discovery, richiede il reperimento o il calcolo della route (figura 3.6). Una volta che viene stabilita, la route viene mantenuta fin quando la destinazione diventa irraggiungibile o la route stessa non viene usata per un certo intervallo di tempo (viene quindi dichiarata scaduta).

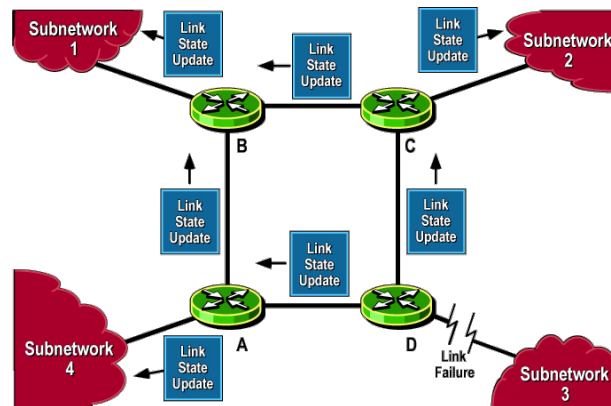


Figura 3.5: Esempio di routing basato su protocollo link-state, in cui le posizioni dei nodi vengono continuamente aggiornate nella rete

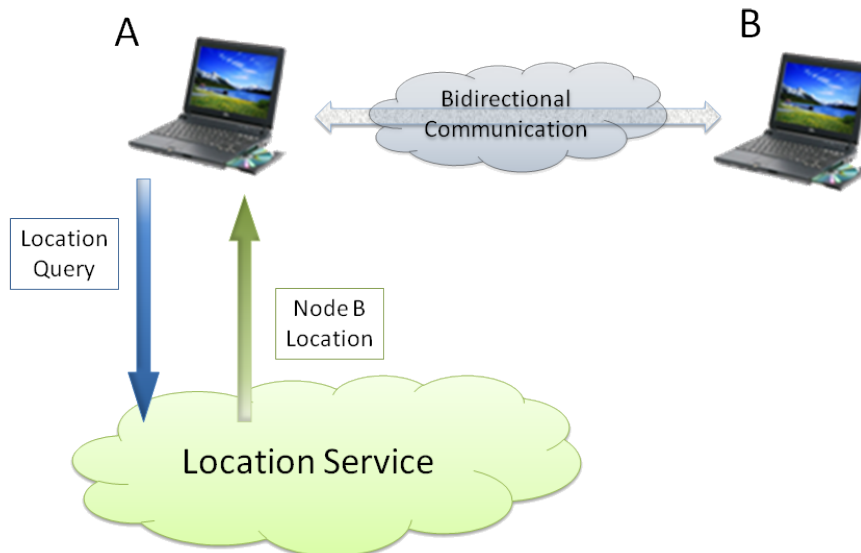


Figura 3.6: Esempio di routing basato su un protocollo reactive, in cui il nodo A interroga il *Location Service* per conoscere la posizione di B prima di poter iniziare la comunicazione con esso

I protocolli di routing geografici sono fondamentali nelle reti mobili ad hoc, per questo verranno descritti in modo più specifico nel prossimo paragrafo.

Esiste infine un'altra soluzione di routing, detta clustering (raggruppamento). I protocolli cluster-based assumono notevole importanza per la scalabilità di una rete di grandi dimensioni, in cui ci può essere una quantità elevata di informazioni di routing che può saturare la rete stessa. Inoltre, spesso i nodi sono molto diversi in termini di risorse disponibili e di prestazioni. Viene fatta quindi una gerarchia con-

siderando alcuni parametri come la capacità computazionale, la potenza trasmessa e l'energia disponibile, secondo cui alcuni nodi vengono dichiarati migliori rispetto ad altri per supportare le funzionalità delle reti ad hoc, come ad esempio il routing.

L'idea di base è quella di raggruppare i nodi della rete in un certo numero di cluster (o gruppi) sovrapposti e di eleggere per ogni gruppo un nodo di riferimento, chiamato clusterhead, che avrà conoscenza di tutti gli altri nodi del cluster. Ciò consente di aumentare la scalabilità della rete grazie ad un'aggregazione delle informazioni di routing, che contengono route solo tra i clusterhead e non tra tutti i nodi. Di conseguenza aumenta il tempo di vita delle route e si riduce l'overhead dovuto ai messaggi di controllo della topologia di rete.

3.4 Location-Aware Routing

I protocolli di routing con consapevolezza della posizione usano, durante le operazioni di forwarding, la posizione dei nodi fornita da GPS [24] o altri meccanismi [9, 31]. Precisamente, un nodo sceglie il prossimo hop per il forwarding usando la posizione fisica dei suoi vicini diretti e del nodo destinazione. I pacchetti sono quindi inoltrati ad un nodo in direzione del destinatario finale; per questa ragione questi protocolli di routing sono spesso chiamati anche approcci position-based o geografici. Solitamente, viene usato un location service per risolvere query riguardo la posizione dei nodi della rete.

Il routing geografico non richiede la creazione ed il mantenimento di percorsi, e nessuna informazione di routing è memorizzata. L'uso di informazioni geolocalizzate evita ricerche estensive nella rete, in quanto sia pacchetti dato che di controllo sono mandati verso le coordinate geografiche note della destinazione. Queste caratteristiche rendono i protocolli di routing location-aware velocemente adattativi a cambi di percorso, tipici delle MANET, e maggiormente scalabili di altri protocolli unicast come AODV o DSR [19].

Fra i protocolli di routing geografico possono essere identificate tre strategie principali [33]: greedy forwarding, directed flooding e hierarchical routing. Nel *Greedy Forwarding* un nodo cerca di inoltrare i pacchetti ad uno dei suoi vicini che sia più vicino di lui alla destinazione finale. Se esistono più nodi che soddisfano il requisito, il modo con cui è scelto uno di essi non è uno soltanto. Si può ad esempio scegliere il nodo più vicino alla destinazione, come in figura 3.7, preferendo

il miglioramento ottenuto con un singolo passo, piuttosto che quello più vicino al nodo trasmittente, salvaguardando maggiormente il consumo energetico.

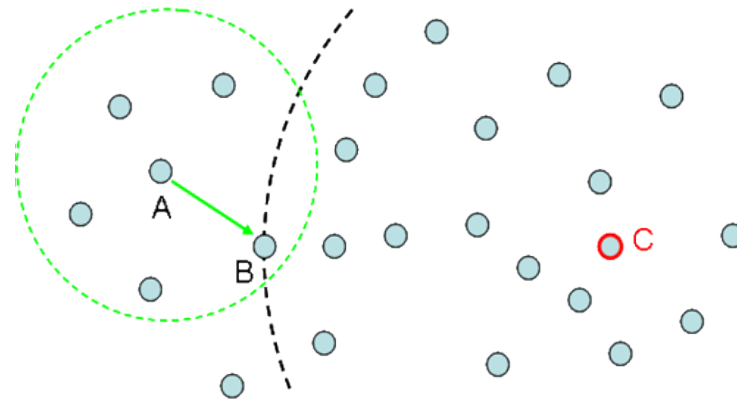


Figura 3.7: Esempio di *Greedy Forwarding* in cui il nodo A sceglie, fra i suoi vicini, quello che minimizza la distanza dalla destinazione C

Se, invece, nessun nodo soddisfa tale requisito, altre regole sono incluse nella strategia greedy per trovare un percorso alternativo. Le politiche Greedy entrano in questa situazione di deadlock quando i pacchetti raggiungono un ottimo locale, ovvero non esiste alcun vicino che sia più prossimo del nodo in questione all'obiettivo.

Nel caso di *Direct Flooding* i nodi inoltrano i pacchetti verso tutti i vicini che sono in direzione del nodo di destinazione, non scelgono quindi quello più promettente. Questo meccanismo può introdurre un grande overhead, perciò va usato in maniera ridotta nell'inoltro dei pacchetti, mentre può essere utilizzato in maniera leggermente più intensa per le route discovery, che coinvolgono messaggi di dimensioni inferiori e meno frequenti.

Il *Routing Gerarchico* infine, prevede un meccanismo di routing suddiviso in due fasi a seconda della distanza dal nodo destinatario: se il nodo è distante, viene usato il routing location-aware, come ad esempio l'algoritmo greedy, mentre quando il pacchetto arriva vicino alla destinazione viene adottato uno dei protocolli di routing di tipo proattivo basato sullo schema distance vector.

Si è visto quindi, in questi due capitoli, come le informazioni sulla posizione dei dispositivi siano sempre più fondamentali per l'evoluzione delle reti mobili ad hoc, indipendentemente dallo scenario in cui esse si collochino. Per questo motivo,

dopo aver compreso qual è il potenziale di questo tipo di reti, è utile procedere ad analizzare maggiormente i dispositivi di localizzazione a disposizione ai giorni nostri, ovvero i GPS.

Capitolo 4

Global Positioning System, GPS

In questo capitolo presenteremo in maniera dettagliata il sistema di posizionamento globale, diffuso ormai universalmente e conosciuto comunemente con il suo acronimo GPS. Per prima cosa sarà presentata la storia evolutiva di questo sistema e ne sarà spiegato in maniera generale, per addentrarsi poi maggiormente ad analizzare quali sono i suoi limiti e le principali fonti di errore. In ultima analisi, verrà analizzato il suo comportamento in un tipico scenario urbano, grazie allo studio effettuato congiuntamente ad alcuni colleghi dell'University of California Los Angeles.

4.1 Evoluzione storica e funzionamento

Il Global Positioning System (GPS) è un sistema di posizionamento su base satellitare, a copertura globale e continua, gestito dal dipartimento della difesa statunitense, che nasce in sostituzione del precedente sistema, il Transit [73]. Transit nasce nel 1960 dalla necessità della U.S. Navy per consentire a navi e sommergibili di determinare la propria posizione in mare in qualsiasi condizione meteorologica. Viene poi reso disponibile anche all'uso civile nel 1967, per poi cessare completamente di essere operativo nel 1996, completamente soppiantato dal GPS.

Il servizio GPS venne aperto al mondo nel 1991 dagli USA con il nome SPS (Standard Positioning System), con specifiche differenziate da quello militare denominato PPS (Precision Positioning System). La differenza sostanziale era l'introduzione della cosiddetta Selective Availability (SA), che introduceva errori intenzionali nei segnali satellitari allo scopo di ridurre l'accuratezza della rilevazione, consentendo precisioni solo nell'ordine di 100-150 m. Tale degradazione del segnale è stata disabilitata l'1 maggio 2000, grazie a un decreto dell'allora Presidente degli Stati Uniti Bill Clinton, mettendo così a disposizione degli usi civili la precisione attuale di circa 10-20 m. Un esempio di ricevitore GPS militare è mostrato in figura 4.1.



Figura 4.1: Ricevitore GPS per uso militare

Nei modelli per uso civile devono però tutt'ora essere presenti delle limitazioni: massimo 18 km per l'altitudine e 515 m/s per velocità, per impedirne il montaggio su missili. Questi limiti possono essere superati, ma non contemporaneamente. L'Unione Europea ha in progetto il completamento di una propria rete di satelliti, il Sistema di posizionamento Galileo, per scopi civili. Il sistema Galileo è un sistema duale cioè nato per compiti civili e militari.

Questo progetto ha un'evidente valenza strategica in quanto la rete americana è proprietà dei soli Stati Uniti d'America ed è gestita da autorità militari, che, in particolari condizioni, potrebbero decidere discrezionalmente e unilateralmente di ridurre la precisione o bloccare selettivamente l'accesso al sistema; la condivisione dell'investimento e della proprietà da parte degli stati utilizzatori garantisce continuità, accessibilità e interoperabilità del servizio.

Funzionamento del Sistema In figura 4.2 è mostrato lo schema di funzionamento base del sistema GPS, in cui il ricevitore calcola la sua posizione partendo dalle informazioni che riceve da almeno tre satelliti. Queste però non sono le uniche

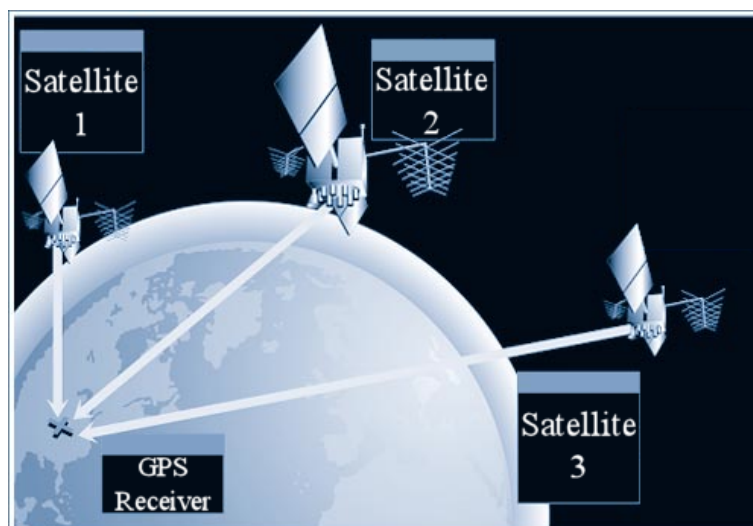


Figura 4.2: Schema di funzionamento base GPS che coinvolge un ricevitore ed almeno tre satelliti

componenti coinvolte, infatti il sistema di navigazione completo basato su GPS si articola delle seguenti componenti:

- Un complesso di minimo 24 satelliti (attualmente sono 31 quelli attivi), divisi in gruppi di almeno tre o quattro su ognuno dei sei piani orbitali.
- Una rete di stazioni di tracciamento (tracking station).
- Un centro di calcolo (computing station).
- Due stazioni di soccorrimento (injection stations).
- Un ricevitore GPS.

Satelliti Sono disposti su 6 piani orbitali inclinati di 55° rispetto al piano equatoriale (quindi non coprono le zone polari) a forma di ellissi a bassa eccentricità. Ogni piano orbitale ha 3 o 4 satelliti, e i piani sono disposti in modo tale che ogni utilizzatore sulla terra possa ricevere i segnali di almeno 5 satelliti. La loro quota è di 20 200 km e compiono due orbite complete in un giorno siderale.

Ciascun satellite emette sulle frequenze di 1,5 e 1,2 GHz derivate da un unico oscillatore ad alta stabilità di clock pari a 10,23 MHz che viene moltiplicato per 154 e 120 per ottenere la frequenza delle due portanti. Lo scopo della doppia frequenza è

quello di eliminare l'errore dovuto alla rifrazione atmosferica. Su queste frequenze portanti, modulate in fase, viene modulato il messaggio di navigazione che ha un bit-rate pari a 50 bps, contenente effemeride, almanacco e stato della costellazione, tempo GPS e parametri di correzione ionosferica.

Ogni satellite trasmette l'almanacco (parametri orbitali approssimati) dell'intera costellazione, ma esclusivamente le effemeridi relative a se stesso. La parte relativa alle effemeridi dura 18 secondi e viene ripetuta ogni 30 secondi, mentre per scaricare completamente l'almanacco dell'intera costellazione sono necessari invece 12,5 minuti.

Le effemeridi (dalla parola greca *ephemeris* = giornaliero) sono tabelle che contengono valori calcolati, nel corso di un particolare intervallo di tempo, di diverse grandezze astronomiche variabili quali: magnitudine, parametri orbitali, coordinate di pianeti, comete, asteroidi e satelliti artificiali.

Il ricevitore GPS, mentre effettua il conteggio doppler, riceve i parametri dell'orbita da cui deriva la posizione del satellite: viene così a disporre di tutti gli elementi necessari a definire nello spazio la superficie di posizione.

In orbita vi sono un minimo di 24 satelliti per la trasmissione di dati GPS, più 3 di scorta. Da questo si evince che da un punto del globo terrestre il ricevitore riesce a vedere solo la metà di essi, quindi 12; in realtà, non li vedrà mai tutti e 12 per via della loro inclinazione rispetto all'equatore. In più il ricevitore GPS stesso fa una discriminazione dei satelliti in base alla loro geometria e alla stima degli errori su ciascuno privilegiando quelli che forniscono maggior precisione.

Ogni satellite è dotato di 4 oscillatori ad altissima precisione (4 perché così si garantisce il funzionamento di almeno uno di essi), di cui 2 al cesio e 2 al rubidio; di reattori per effettuare le correzioni di orbita; di due pannelli solari di area pari a $7,25 m^2$ per la produzione di energia ed infine di batterie di emergenza per garantire l'apporto energetico nei periodi in cui il sole è eclissato. Pesa circa 845 kg ed ha una vita di progetto di 7,5 anni (figura 4.3).

Stazioni di tracciamento e centro di calcolo Il tracciamento dei satelliti comprende tutte quelle operazioni atte a determinare i parametri dell'orbita. A ciò provvedono 4 stazioni principali, site nei pressi dell'equatore, dette appunto di tracciamento (main tracking stations) e un centro di calcolo (computing center),



Figura 4.3: Satellite usato nel sistema GPS

tutti situati in territorio USA, ed in particolare a Wahiova (Hawaii), Point Mugu (California), Prospect Harbur (Maine) e Rosemount (Minnesota).

Point Mugu è anche sede del centro di calcolo. Ogni volta che ciascun satellite nel suo moto orbitale sorvola il territorio americano le stazioni di tracciamento ne registrano i dati doppler che vengono avviati al centro di calcolo e qui valorizzati per la determinazione dei parametri orbitali. Per risolvere questo problema è stato necessario venire in possesso di un fedele modello matematico del campo gravitazionale terrestre. La costruzione di questo modello è stato uno dei problemi di più ardua soluzione nello sviluppo del progetto Transit da cui è derivato l'attuale Navstar GPS (*NAVigation Satellite Time And Ranging Global Positioning System*), comunemente abbreviato semplicemente in *Global Positioning System* (GPS).

Stazioni di soccorrimento I parametri orbitali di ciascun satellite, appena determinati presso il centro di calcolo, sono riuniti in un messaggio che viene inoltrato al satellite interessato mediante una delle stazioni di soccorrimento. Il satellite registra i parametri ricevuti nella sua memoria e li utilizzerà poi per comporre i messaggi da inviare agli utenti.

Ricevitore GPS Il principio di funzionamento si basa su un metodo di posizionamento sferico, che consiste nel misurare il tempo impiegato da un segnale radio a percorrere la distanza satellite-ricevitore. Poiché il ricevitore non sa quando è partito il segnale dal satellite per il calcolo della differenza dei tempi il segna-

le inviato dal satellite è di tipo orario grazie all'orologio presente sul satellite e il ricevitore calcola l'esatta distanza di propagazione dal satellite a partire dalla differenza (dell'ordine dei microsecondi) dell'orario pervenuto con quello del proprio orologio perfettamente sincronizzato.

Conoscendo il tempo impiegato dal segnale per giungere al ricevitore e l'esatta posizione di almeno 3 satelliti per avere una posizione 2D (bidimensionale), e 4 per avere una posizione 3D (tridimensionale), è possibile determinare la posizione nello spazio del ricevitore stesso. Tale procedimento, chiamato trilaterazione, utilizza solo informazioni di distanza ed è simile alla triangolazione, dal quale tuttavia si differenzia per il fatto di fare a meno di informazioni riguardanti gli angoli.

La precisione può essere ulteriormente incrementata grazie all'uso di sistemi come il WAAS (statunitense) o l'EGNOS (europeo), perfettamente compatibili tra di loro. Consistono rispettivamente in due e tre satelliti geostazionari che inviano dei segnali di correzione.

La modalità invece Differential-GPS (DGPS) utilizza un collegamento radio per ricevere dati DGPS da una stazione di terra e ridurre l'errore sulla posizione ad un paio di metri. Questa stazione di terra infatti, conosce perfettamente la propria posizione, e confrontandola con quella che riceve dai satelliti, riesce a calcolare qual è l'errore introdotto da essi, e propaga questa informazione, cosicché i vari ricevitori possano elaborare la posizione tenendone conto.

La modalità DGPS-IP sfrutta, anziché onde radio, la rete Internet per l'invio delle informazioni di correzione.

Diffusione dei dispositivi GPS I moderni ricevitori GPS hanno raggiunto dei costi molto contenuti. Dopo il telefono cellulare stiamo assistendo alla diffusione di un nuovo cult: quello del navigatore satellitare personale. Il mercato offre ormai soluzioni a basso costo per tutti gli impieghi e per tutte le tasche che si rivelano efficaci non soltanto per la navigazione satellitare in sé e per sé, ma anche per usi civili, per il monitoraggio dei servizi mobili e per il controllo del territorio. Esistono varie soluzioni, mostrate anche in figura 4.4:

- **Integrate:** sono dispositivi portatili All-in-One che incorporano un ricevitore GPS, un display LCD, un altoparlante, il processore che esegue le istruzioni, date solitamente da un sistema operativo proprietario ed uno slot per schede di memoria ove memorizzare la cartografia.

- Ibride: sono dispositivi portatili (PC, Palmari, SmartPhone) che, nati per scopi diversi, sono resi adatti alla navigazione satellitare attraverso il collegamento di un ricevitore GPS esterno (Bluetooth o via cavo) e l'adozione di un software dedicato, in grado di gestire la cartografia. Addirittura, molti di questi dispositivi integrano direttamente al loro interno già il chip GPS.



(a) Navigatore All-in-one

(b) Smartphone con chip GPS

Figura 4.4: Due esempi di soluzione integrata (a) ed ibrida (b) per la navigazione GPS

Con la diffusione capillare dei sistemi GPS, e il conseguente abbattimento dei costi dei ricevitori, molti produttori di telefoni cellulari hanno cercato di inserire un modulo GPS all'interno dei loro prodotti, aprendosi quindi al nuovo mercato dei servizi LBS (*Location Based Service*, servizi basati sul posizionamento) [63, 64, 65]. Tali servizi vengono sempre più sfruttati per offrire anche sul web dei servizi molto utili.

Tuttavia, la relativa lentezza con cui un terminale GPS acquisisce la propria posizione al momento dell'accensione (in media, tra i 45 e i 90 secondi), dovuta alla necessità di cercare i satelliti in vista, ed il conseguente notevole impegno di risorse hardware ed energetiche, ha frenato in un primo momento questo tipo di abbinamento.

Negli ultimi anni, però, è stato introdotto in questo tipo di telefoni il sistema *Assisted GPS*, detto anche *A-GPS*, con cui è possibile ovviare a tali problemi: si fanno pervenire al terminale GPS, attraverso la rete di telefonia mobile, le informazioni

sui satelliti visibili dalla cella a cui l'utente è agganciato. In questo modo un telefono A-GPS può in pochi secondi ricavare la propria posizione iniziale, in quanto si assume che i satelliti in vista dalla cella siano gli stessi visibili dai terminali sotto la sua copertura radio. Tale sistema è molto utile anche come servizio d'emergenza, ad esempio per localizzare mezzi o persone ferite in seguito ad un incidente.

4.2 Fonti di errore

Le sorgenti di errore nella stima della posizione effettiva di un ricevitore GPS sono varie, e ognuna di esse ha un'incidenza differente sul calcolo finale. Gli errori introdotti dalle diverse sorgenti sono noti e ben documentati, e possono riassumersi in [60]:

- Selective Availability.
- Geometria Satellitare.
- Orbita Satellitare.
- Effetto di percorso multiplo.
- Effetti Atmosferici.
- Errori di sincronizzazione ed arrotondamento.
- Effetti relativistici.

Selective Availability Quello che fino al decennio scorso è stato il fattore più rilevante di errore nei sistemi GPS non è in realtà oggi più presente. Difatti, alla mezzanotte dell'1 maggio del 2000, la cosiddetta Selective Availability (SA) è stata disabilitata. SA era una falsificazione artificiale del tempo trasmesso dal satellite.

Nei ricevitori GPS civili questo portava ad una degradazione della precisione del segnale GPS variabile dai 50 ai 100 metri (figura 4.5). Inoltre, anche i dati all'interno dell'effemeride erano trasmessi con precisione inferiore, non rispecchiando esattamente la posizione reale del satellite al momento dell'invio. In questo modo un'imprecisione complessiva fra i 50 ed i 150 metri poteva essere ottenuta per svariate ore. Questo considerando il posizionamento orizzontale. Nel caso del posizionamento verticale invece, l'impatto era ancora maggiore, raggiungendo i 500 metri e rendendolo di fatto inutilizzabile.

La ragione alla base dell'impiego della Selective Availability erano preoccupazioni a livello di sicurezza. Ad esempio, organizzazioni terroristiche non avreb-

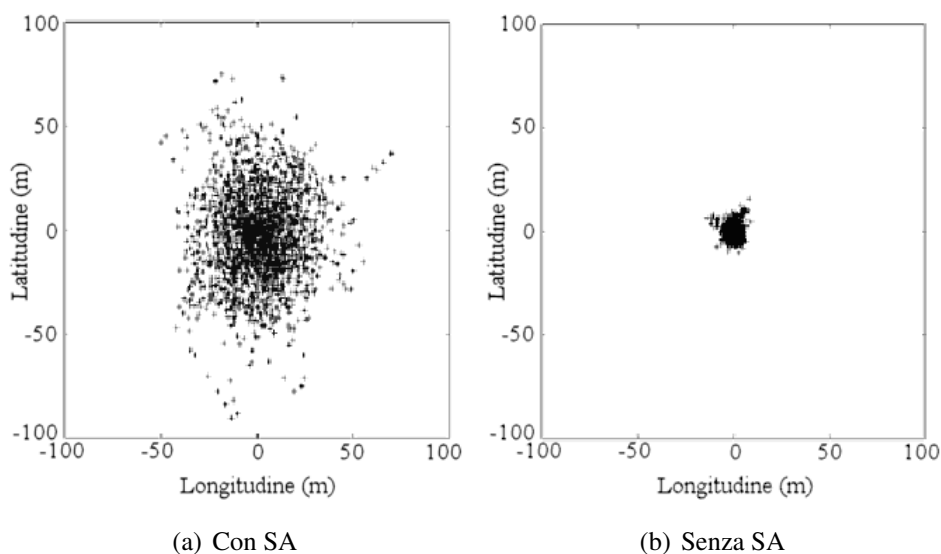


Figura 4.5: Grafico che mostra la posizione rilevata nell'arco di 24 ore, con (a) e senza (b) Selective Availability. Nel primo caso il 95% dei punti ha un errore inferiore a 45 metri, nel secondo la stessa percentuale di punti ha un errore inferiore ai 6.3 metri

bero dovuto essere fornite della possibilità di localizzare con precisione monumenti o palazzi importanti con armi a controllo remoto prodotte in proprio. Paradossalmente, durante la prima guerra del golfo nel 1990, SA dovette essere disattivato parzialmente (ed in una zona circoscritta), poiché non erano disponibili per le truppe americane abbastanza ricevitori militari, e si vedettero costrette ad utilizzare 10'000 ricevitori civili (forniti da Magellan e Trimble). Con la disattivazione di SA, anche questi dispositivi civili si rivelarono estremamente precisi nel deserto, senza alcun altro punto di riferimento.

La crescente diffusione di dispositivi D-GPS (presentati in precedenza) e la grande efficienza con cui essi riuscivano ad arginare gli errori introdotti dalla Selective Availability, furono uno dei principali motivi che spinsero ad eliminarla.

Geometria Satellitare Il secondo fattore ad influenzare la precisione con cui viene identificata la posizione è la *Geometria Satellitare*. In una descrizione semplificata, essa descrive la posizione relativa dei satelliti dal punto di vista dell'osservatore / ricevitore. Se un ricevitore vede 4 satelliti che sono tutti disposti nella stessa zona, ad esempio nord-ovest, questo porta ad una cattiva geometria.

Nel caso peggiore, non è possibile la determinazione di alcuna posizione, se tutte le determinanti della distanza puntano nella stessa direzione. Anche se la posizione fosse determinata, la sua imprecisione sarebbe nell'ordine di 100 - 150 metri. Se invece i 4 satelliti sono ben distribuiti sull'intero firmamento, la posizione calcolata sarà molto più precisa. Vediamo un esempio grafico del caso semplice bidimensionale in figura 4.6.

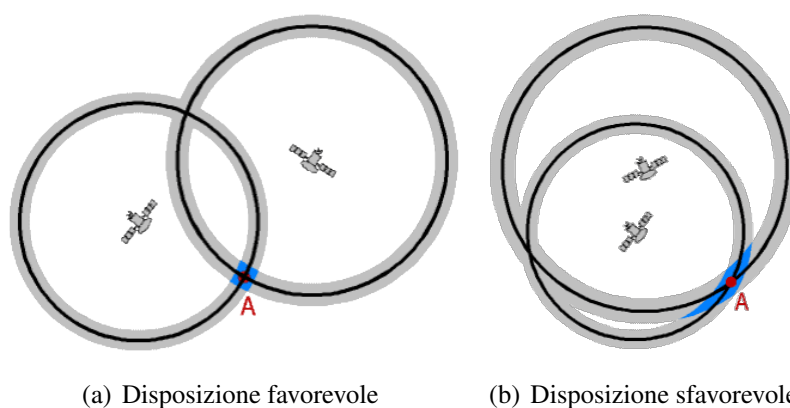


Figura 4.6: Esempio di come la disposizione dei satelliti visti dal ricevitore possa influenzare la precisione nel calcolo della posizione

Se i due satelliti si trovano in una posizione vantaggiosa dal punto di vista dell'osservatore, possono essere visti con un angolo approssimativamente di 90° l'uno dall'altro. Il tempo di propagazione del segnale non può essere stimato con precisione assoluta, come visto in precedenza, per cui la posizione possibile è evidenziata dal cerchio grigio, di spessore non trascurabile. Il punto di intersezione delle informazioni provenienti dai due satelliti, che determina la possibile posizione attuale, è evidenziato dal quadrato blu ed è relativamente piccolo 4.6(a).

Se i due satelliti sono invece posizionati grosso modo sulla stessa linea visiva dal punto di vista del ricevitore, il piano di intersezione delle possibili posizioni è considerevolmente maggiore, con conseguente aumento dell'imprecisione della determinazione della posizione 4.6(b).

La geometria satellitare è rilevante anche quando il ricevitore è utilizzato all'interno di veicoli o vicino a palazzi alti. Se alcuni dei segnali sono oscurati ad esempio, i satelliti che rimangono visibili determinano la qualità dell'identificazione. Più è ampia la porzione di cielo oscurata, più diventa difficile la determinazione

della posizione. Per indicare la qualità della geometria satellitare, sono comunemente usati gli indici DOP (Dilution of Precision). In base a quale fattore venga considerato per il calcolo del DOP, si parla di diversi DOP:

- GDOP (Geometric Dilution Of Precision): precisione complessiva, coordinate 3D e tempo.
- PDOP (Positional Dilution Of Precision): precisione posizionale, coordinate 3D.
- HDOP (Horizontal Dilution Of Precision): precisione orizzontale, coordinate 2D.
- VDOP (Vertical Dilution Of Precision): precisione verticale, altezza.
- TDOP (Time Dilution Of Precision): precisione temporale, tempo.

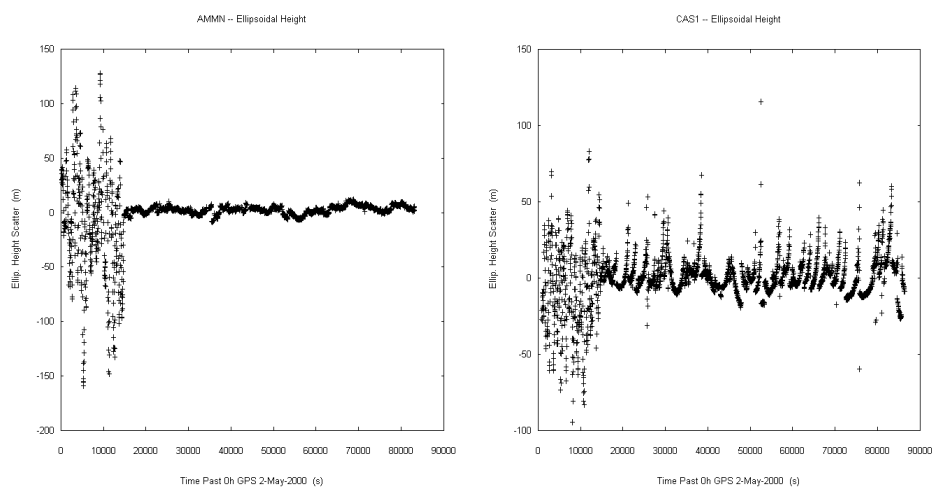
La geometria satellitare di per se non genera imprecisioni nel posizionamento che possano essere misurate direttamente in metri, bensì i valori di DOP influenzano le imprecisioni introdotte da altre fonti. Elevati valori DOP semplicemente amplificano maggiormente gli errori rispetto a valori DOP bassi.

La latitudine del ricevitore influenza anch'essa l'errore dovuto alla geometria satellitare, come si può notare bene nell'esempio in figura 4.7. Sono qui riportati i grafici dell'errore GPS verticale rilevato la notte dell'1 maggio 2000 da due differenti stazioni fisse di riferimento. La prima parte del grafo mostra l'errore dovuto alla Selective Availability, ma concentrandosi sulla seconda parte dello stesso si può notare come una volta eliminata quella, il comportamento sia sostanzialmente differente.

Ad Amman, in Giordania (figura 4.7(a)), ad una latitudine di 32.03° Nord, l'errore residuo è praticamente costante attorno a qualche metro. Qui infatti può in ogni momento essere trovata una costellazione di satelliti ottimale.

La stazione Casey invece, in Antartide (figura 4.7(b)), con una latitudine di 66.28° Sud, ha un comportamento molto più variabile e spesso l'errore raggiunge picchi di 50 metri. Questo a causa della copertura non sempre ottimale dei satelliti e dell'effetto atmosferico, che induce una falsificazione maggiore man mano che ci avviciniamo ai poli.

Orbita Satellitare Nonostante i satelliti siano posizionati con estrema cura su orbite estremamente precise, leggere deviazioni delle stesse sono possibili a causa



(a) Stazione di Amman, Giordania

(b) Stazione di Casey, Antartide

Figura 4.7: Errore di stima della posizione verticale a differenti latitudini

delle forze gravitazionali. I dati delle orbite sono controllati e corretti regolarmente, e sono mandati ai ricevitori nell'insieme dei dati contenuti nell'effemeride. Per questo l'incidenza finale sul calcolo della posizione è molto bassa, inferiore ai 2 metri.

Effetto di percorso multiplo L'effetto di percorso multiplo è dovuto alla riflessione dei segnali satellitari, che sono onde radio, sugli oggetti. È concettualmente lo stesso effetto che generava immagini fantasma nella televisione quando le antenne sui tetti erano più diffuse dei ricevitori digitali odierni e si affollavano le une vicino alle altre.

Nei segnali GPS, questo effetto si presenta maggiormente nelle vicinanze di grossi palazzi o altre strutture molto elevate (figura 4.8). Il segnale riflesso impiega più tempo di quello diretto per raggiungere il ricevitore, e per questo si genera errore nella stima della distanza del satellite. L'errore risultante è però anche qui contenuto entro il paio di metri.

Effetti Atmosferici Un'altra sorgente di imprecisione è la ridotta velocità di propagazione dei segnali radio nella troposfera e nella ionosfera, rispetto allo spazio vuoto in cui si propagano con la velocità della luce (figura 4.9). L'errore dovuto alla ionosfera è per la maggior parte corretto direttamente dai ricevitori, in quanto

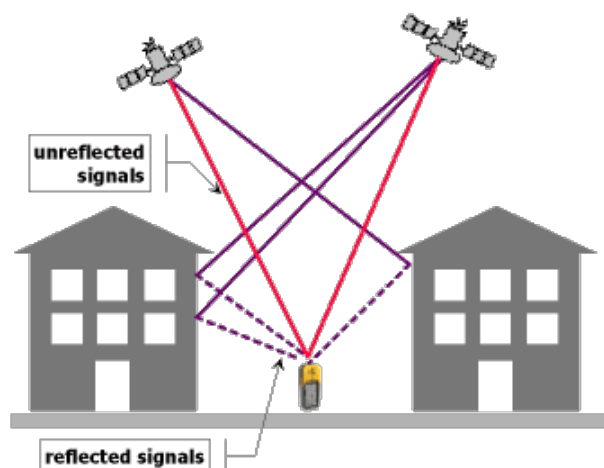


Figura 4.8: Effetto del percorso multiplo sull'errore GPS

la variazione di velocità tipica dell'attraversamento della ionosfera per onde ad alta e bassa frequenza è ben noto in condizioni standard.

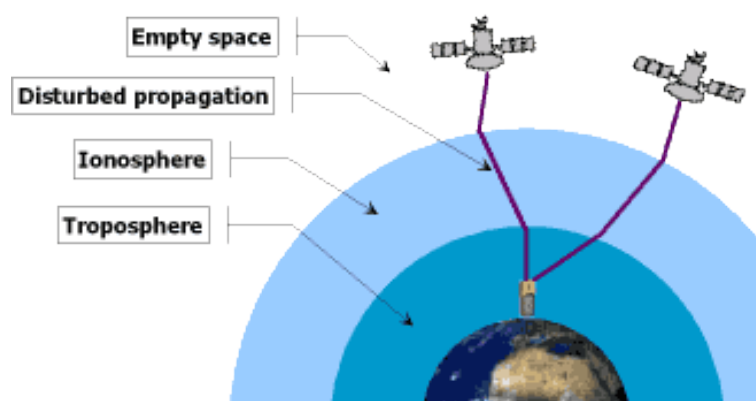


Figura 4.9: Differente propagazione dei segnali radio nella troposfera e nella ionosfera terrestre

Queste variazioni sono tenute da conto per ogni calcolo della posizione, anche se i ricevitori ad uso civile non sono in grado di correggere cambiamenti dinamici imprevisti, come ad esempio quelli generati da forte vento solare. I ricevitori militari invece, che lavorano su due frequenze distinte, possono colmare questa lacuna essendo in grado di calcolare l'effettivo tempo di attraversamento della ionosfera confrontando i dati delle due diverse onde.

L'effetto della troposfera concorre ad aumentare ulteriormente il tempo impiegato dalle onde elettromagnetiche per raggiungere la destinazione, per mezzo della rifrazione. Il motivo alla base della rifrazione è la differente concentrazione di va-

pori acquei nella troposfera, causata dalle diverse condizioni meteorologiche. L'errore generato in questo modo è inferiore a quello introdotto dalla ionosfera, ma purtroppo non può essere eliminato grazie a calcoli specifici. Può infatti essere soltanto approssimato tramite un modello di calcolo generale.

Con l'implementazione di WAAS (*Wide Area Augmentation System*) ed EGNOS (*European Geostationary Navigation Overlay Service*), già presentati in precedenza, è possibile definire mappe delle condizioni atmosferiche su regioni del globo differenti e mandarle ai ricevitori, migliorando così notevolmente la precisione ottenibile.

Errori di sincronizzazione ed arrotondamento Malgrado la sincronizzazione del clock del ricevitore con l'ora satellitare durante la determinazione della posizione, la rimanente imprecisione sui tempi porta ad un errore di circa 2 metri nella stessa. Errori di arrotondamento e di calcolo da parte del ricevitore poi aggiungono un altro metro circa.

Effetti relativistici Nella vita di tutti i giorni, siamo generalmente inclini a non renderci conto, o meglio, a non considerare l'onnipresenza della teoria della relatività. Nonostante questo, essa influenza moltissimi processi, fra cui anche il corretto funzionamento dei sistemi GPS. Come fin qui abbiamo visto, il tempo è un fattore molto rilevante nella navigazione GPS, e deve essere preciso nell'ordine di 20 - 30 nanosecondi per assicurare la necessaria precisione. Perciò, il veloce movimento dei satelliti (circa 12000 km/h) deve essere considerato.

Come insegnato dalla teoria della relatività, il tempo scorre più lento durante movimenti molto veloci. In un satellite che si muove ad una velocità di 3874 m/s, questo si traduce in una differenza temporale, se osservato dalla terra, di 7,2 microsecondi al giorno. La teoria della relatività insegna inoltre che il tempo scorre più lentamente col crescere del campo gravitazionale a cui si è sottoposti. Per il solito osservatore sulla terra quindi, l'orologio a bordo di un satellite scorrerà più velocemente (per la precisione, questo effetto è 6 volte più forte del precedente per un satellite posto a circa 20'000 km dalla terra).

Nel complesso quindi, l'orologio di un satellite sembra scorrere leggermente più veloce, all'incirca 38 microsecondi al giorno, che si rispecchierebbe in un errore cumulativo di circa 10 km al giorno. Per far sì che questi errori, stabili, non debbano

essere corretti costantemente, il clock dei satelliti è stato fissato a 10.229999995453 Mhz invece di 10.23 Mhz, ma essi sono trattati come se fosse 10.23 MHz. Con questo stratagemma l'effetto relativistico è compensato una volta per tutte.

Vi è anche un altro effetto relativistico, che non è però solitamente considerato nell'uso quotidiano del GPS. Questo effetto è chiamato effetto Sagnac ed è causato dal movimento dell'osservatore sulla superficie terrestre, che si muove anch'essa con una velocità fino a 500 m/s (sull'equatore) dovuta alla rotazione del globo. L'influenza di questo effetto, oltre ad essere molto ridotta, è molto complessa da calcolare in quanto dipende dalla direzione del moto, per questo è considerata in soltanto in casi particolari. Gli errori dei sistemi GPS sono schematizzati in tabella

Sorgente di Errore	Errore
Effetto della Ionosfera	± 5 metri
Scostamento delle orbite satellitari	± 2.5 metri
Errore di sincronia clock	± 2 metri
Effetto di percorso multiplo	± 1 metro
Effetto della Troposfera	± 0.5 metri
Errori di calcolo ed arrotondamento	± 1 metro

Tabella 4.1: Incidenza delle diverse cause d'errore sulla precisione della rilevazione GPS.

4.1. Tutti i valori non sono costanti, ma sono soggetti a varianza. Tutti insieme oggi generano un errore di circa ± 15 metri, mentre quando SA era ancora attiva, l'errore era nell'ordine di ± 100 metri.

Le correzioni introdotte da sistemi come WAAS ed EGNOS, che per lo più riducono l'effetto della ionosfera, oltre a migliorare gli errori di orbita e di clock, portano il range di errore a circa $\pm 3 - 5$ metri.

Sono inoltre disponibili sul mercato anche dispositivi che sono in grado di lavorare in multifrequenza, ricevendo cioè contemporaneamente più segnali inviati a frequenza diversa dallo stesso satellite, che tramite calcoli opportuni sono in grado di ridurre l'errore complessivo a meno di un metro, ma questo non è il caso dei dispositivi disponibili sul mercato di massa, in quanto i loro prezzi possono arrivare attorno ai 10'000 \$ per singolo ricevitore.

Tutti questi dati, come vedremo nel prossimo paragrafo, si intendono sempre in spazio aperto e con buona visibilità della porzione di cielo soprastante.

4.3 Esperimenti in scenario urbano

Vediamo ora come l'ambiente in cui viene utilizzato il GPS possa influenzare attivamente la sua precisione, e conseguentemente l'affidabilità delle informazioni ottenute.

Per prima cosa, viene presentato brevemente il modo in cui l'ambiente urbano influisce sulla precisione dei ricevitori GPS e sulla loro affidabilità, in quanto è questo lo scenario concreto in cui si andrà a collocare il sistema sviluppato.

Successivamente, verranno presentate due tipologie di esperimenti che sono stati svolti per quantificare l'incidenza di questo tipo di ambiente. La prima tipologia di esperimenti si focalizza sul caso di un ricevitore in movimento all'interno di scenari tipici urbani; il secondo invece si concentra sull'analisi del comportamento di un ricevitore fermo, monitorato continuativamente per diverse ore.

4.3.1 Incidenza dell'ambiente urbano

Come già anticipato, l'utilizzo del sistema GPS in un normale ambiente urbano può portare a un degrado delle prestazioni rispetto al caso in cui venga utilizzato in un'area aperta. Questo dipende in parte dall'errore dovuto al percorso multiplo effettuato dalle onde elettromagnetiche, che rimbalzano contro palazzi o strutture elevate presenti nelle vicinanze, ed aumenta con l'aumentare della dimensione di queste superfici d'ostacolo. In aggiunta a ciò, incide in maniera elevata anche la presenza dei cosiddetti *Canyon Urbani*, che possono introdurre errori anche di alcune centinaia di metri.

Un canyon urbano è una situazione in cui la presenza di palazzi alti ai bordi delle strade ricrea in un qualche modo la situazione che in natura si presenta nei canyon, appunto, con una valle profonda e stretta, dal cui interno si riesce a vedere liberamente soltanto una porzione ridotta di cielo (figura 4.10). Questo purtroppo incide fortemente sul funzionamento dei ricevitori GPS, perché una così stretta visione del cielo corrisponde a una copertura molto ridotta da parte dei satelliti, che si

trovano molto spesso ad essere offuscati e comparire e scomparire improvvisamente dal campo visivo.

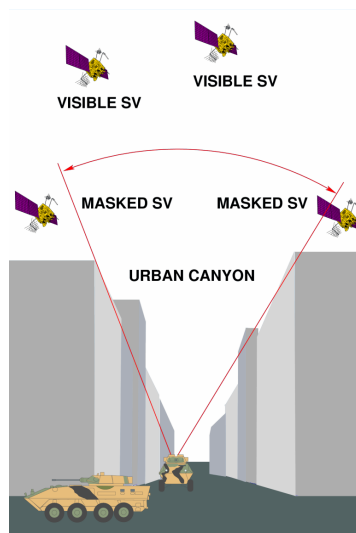


Figura 4.10: Effetto dei *Canyon Urbani* sulla ricezione GPS

Dato che il sistema sviluppato e presentato in questa tesi si basa sull'analisi di tracce registrate da ricevitori GPS, sono stati effettuati alcuni esperimenti reali per valutare l'incidenza di questi problemi, per avere una chiara idea della loro tipologia, e sviluppare un'adeguata metodologia per correggerli o quantomeno affrontarli nella maniera migliore.

4.3.2 Esperimenti in mobilità

Per prima cosa, sono stati svolti alcuni esperimenti nell'area di Los Angeles, California, rispettivamente lungo la freeway 10, che per conformazione si estende lungo una zona di cielo aperto, in cui la copertura satellitare è ottimale, e nella zona di Downtown Los Angeles, che invece coi suoi fitti grattacieli costituisce un ottimo esempio di canyon urbano [67]. Questi esperimenti sono stati effettuati con un ricevitore GPS stand-alone Hamlet HBTGPS20 a basso costo come quello in figura 4.11, collegato ad un computer portatile [56]. Questo tipo di ricevitore si basa su architettura SiRF Star III, come il ricevitore descritto nell'appendice B.1.

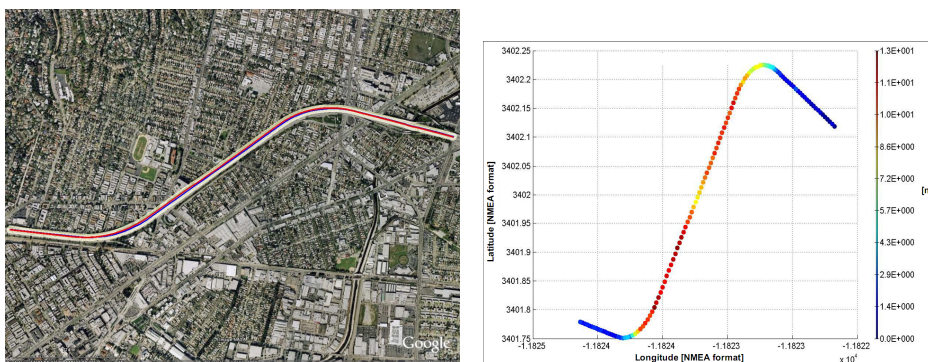
In figura 4.12(a) si mostra il percorso sulla freeway di Los Angeles. La linea rossa rappresenta il percorso effettivamente registrato dal ricevitore GPS, mentre la blu corrisponde a quello realmente effettuato dalla vettura. In questo scenario,



Figura 4.11: Ricevitore GPS stand-alone Hamlet HBTGPS20

soltanto lievissime differenze fra i due percorsi sono visibili. Queste rientrano nella tolleranza dello strumento dichiarata dal costruttore, di circa 10 metri in situazioni di cielo aperto.

In figura 4.12(b) invece è graficato l'errore del GPS misurato in ogni punto, da cui si può vedere che esso si mantiene sempre sotto i 20 metri. Inoltre la campionatura del GPS è costante, senza evidenti discontinuità nell'acquisizione della posizione, che avviene per lo più ogni secondo, che è la frequenza massima possibile per il ricevitore.



(a) Percorso effettuato

(b) Errore del ricevitore

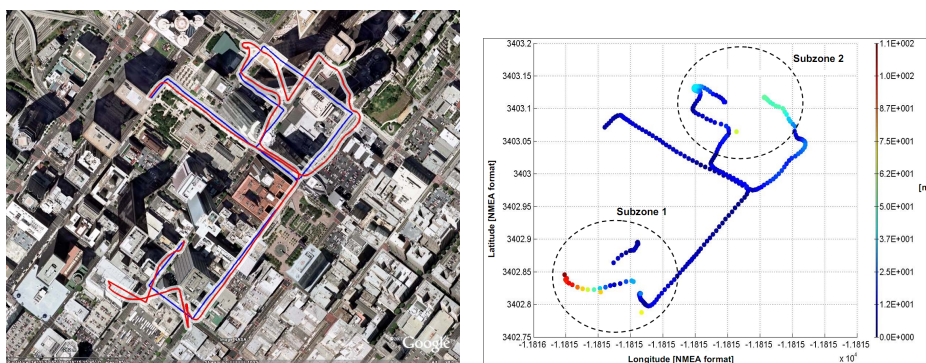
Figura 4.12: Percorso effettuato nei test sulla freeway 10 di Los Angeles (a) e relativo errore del ricevitore nella stima della posizione (b)

Il secondo scenario di test è, come detto, la zona di Downtown LA, in cui il percorso seguito (figura 4.13(a)) si snoda attraverso alcune strade contornate da alti grattacieli e scarsa visibilità della porzione di cielo soprastante, in particolare nelle due aree evidenziate in figura 4.13(b) come *Subzone 1* e *Subzone 2*.

Si può notare chiaramente in figura 4.13(a) come in questo scenario il percorso registrato dal ricevitore GPS non sia sempre vicino alla traiettoria realmente percorsa, com'era invece nel caso dello scenario di freeway.

Inoltre, attraversando le zone di canyon urbani, anche la frequenza di campionamento viene a decrescere, con un crescente numero di istanti in cui il ricevitore non è in grado di calcolare la posizione attuale, e quindi non campiona. Abbiamo quindi il duplice risultato di avere informazioni più imprecise dal punto di vista della posizione, e che non garantiscono continuità dal punto di vista temporale.

In figura 4.13(b) si può notare come l'errore raggiunga i 100 metri nella zona 1, caratterizzata da strade strette e palazzi di media altezza, e i 50 metri nella zona 2, che ha strade più larghe e grattacieli più elevati.



(a) Percorso effettuato

(b) Errore del ricevitore

Figura 4.13: Percorso effettuato nei test a downtown di Los Angeles (a) e relativo errore del ricevitore nella stima della posizione (b)

Osservando un'ultima volta il percorso registrato e quello effettivamente effettuato, si nota come l'errore introdotto dal ricevitore diventi molto difficile da mitigare in questo scenario, anche usando tecniche di map-matching, che basandosi sulla presenza di una mappa alla base del movimento registrato dal ricevitore GPS, cercano di correggere la posizione registrata dal ricevitore, cercando di mantenerla coerente con la topologia della mappa sottostante ed aderente alla strada che si sta percorrendo.

Se andiamo a vedere più da vicino la zona 2 ad esempio (figura 4.14) il ricevitore calcola una posizione che è addirittura sulla strada perpendicolare a quella

effettivamente percorsa. Qui, a causa dell'ambiente circostante, il ricevitore non è in grado di fornire una stima realistica della posizione dell'utente.

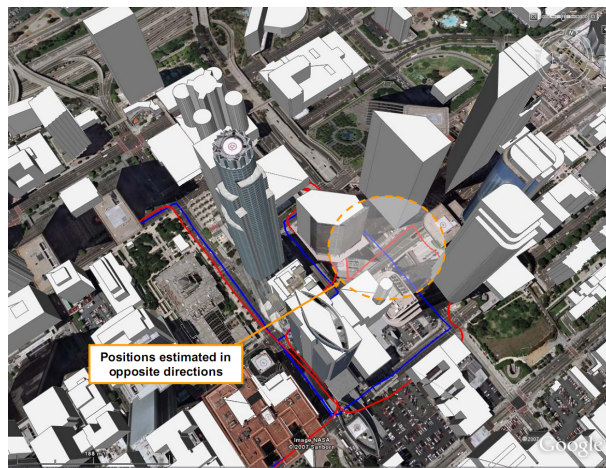


Figura 4.14: Zoom della zona con il canyon urbano più evidente. La linea rossa rappresenta il percorso registrato, la blu invece quello realmente percorso

I dati ottenuti sul campo vengono riportati in maniera compatta in tabella 4.2, in cui si evidenzia come in campo aperto non siano presenti errori sopra i 20 metri, e quasi il 65% degli errori sia localizzato in meno di 10 metri, mentre in condizioni di scarsa visibilità dei satelliti una frazione considerevole dei campionamenti abbia un errore superiore ai 20 metri.

Con un simile errore purtroppo i dati diventano fortemente inaffidabili e andrebbero scartati, però purtroppo se ci viene data soltanto la traccia GPS registrata dal ricevitore, senza il percorso che è stato effettivamente effettuato, non è così immediato discriminare un punto buono da uno da non considerare.

Range di Errore [m]	% in scenario Freeway	% in scenario Downtown
0 - 10	63.636%	50.673%
10 - 20	36.364%	22.642%
> 20	0%	26.685%

Tabella 4.2: Percentuale di errori di stima della posizione nei due scenari considerati.

4.3.3 Esperimenti con ricevitore stazionario

Oltre a quello in movimento appena presentato, è stato svolto un altro test di tipo intensivo sulla precisione delle informazioni registrate dal sistema GPS in due diverse situazioni di stallo. Per questi esperimenti è stato utilizzato un ricevitore HOLUX GR-213 dotato di interfaccia USB per il collegamento a notebook. Questo ricevitore è descritto più dettagliatamente nell'appendice B.1.

Per il primo esperimento, è stato posizionato il ricevitore GPS collegato ad un computer portatile per un'intera notte sul tetto del palazzo Boelter Hall dell'UCLA. Qui, è stato registrata continuamente la posizione acquisita dal ricevitore, in una situazione di cielo completamente visibile, in quanto non vi sono nei paraggi altri palazzi più alti di quello utilizzato che possano in qualche modo interferire con i segnali GPS inviati dai satelliti, o con la visione degli stessi da parte del ricevitore.

L'esperimento è durato 8 ore, 20 minuti e 14 secondi, e la posizione è stata campionata 29'468 volte in maniera distinta, indicando che soltanto 546 campionamenti non sono stati possibili, ovvero meno del 2%. Inoltre, come si può vedere dalla figura 4.15, la quasi totalità delle posizioni campionate ricade in meno di 10 metri di distanza dalla posizione effettiva del ricevitore, evidenziata in colore giallo.

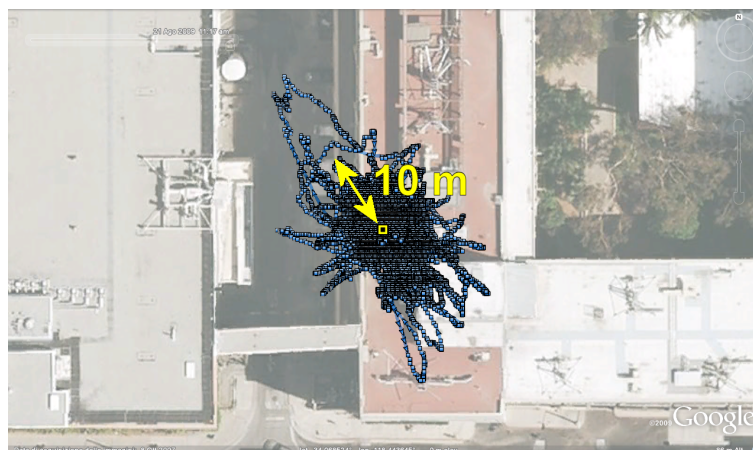


Figura 4.15: Posizione rilevata dal ricevitore GPS sul tetto dell'edificio Boelter Hall in 29'468 campionamenti

È utile in questa fase notare come il comportamento in caso di errore nella stima del GPS sia di tipo derivativo, ovvero il movimento campionato rassomiglia a quello di un corpo lasciato in mare che va alla deriva allontanandosi sempre più dalla

posizione originale. Questo finché, ad un certo punto, ricomincia a convergere verso risultati corretti.

Non vi è mai un outlier isolato, che si distacchi radicalmente dalle posizioni campionate prima e da quelle campionate dopo. Questo è importante perché dovremo tenerne conto quando cercheremo di correggere gli errori di posizionamento.

Il secondo esperimento, analogo, è stato effettuato posizionando la stessa strumentazione, invece che sul tetto, sul balcone del terzo piano affacciato sul cortile interno al palazzo. Qui, la situazione ricreata, è tale per cui la porzione di cielo visibile è molto ridotta.

L'esperimento è stato effettuato per 2 ore, 30 minuti e 27 secondi, e in tale periodo sono stati campionati 5376 punti distinti. Non è stato quindi possibile calcolare la posizione per ben 3651 volte, ovvero nel 40% dei casi circa.

Osservando la figura 4.16 si può notare come l'errore di stima sia molto superiore al caso precedente, variando comunemente in valori entro i 30 metri, ma arrivando facilmente anche a 50 o 60 metri. Inoltre, si può notare come l'errore sembra portare la stima del ricevitore lungo un asse ben preciso.

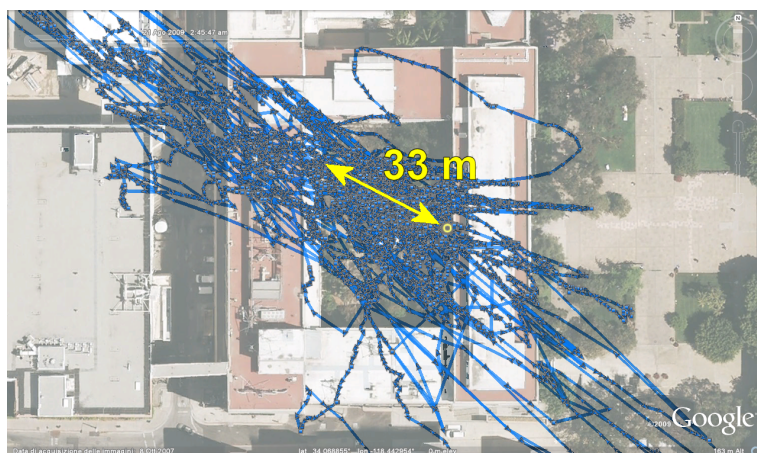


Figura 4.16: Visione locale delle posizioni rilevate dal ricevitore GPS sul balcone al terzo piano dell'edificio Boelter Hall in 2 ore di campionamenti

Analizzando i dati, si è visto che queste due punte sono confinate in due precise fasce orarie di circa un quarto d'ora l'una, in cui il GPS sembra dare risultati completamente privi di senso, che arrivano addirittura a comunicare una posizione che dista 2.8 km da quella attuale (figura 4.17). Questo problema è imputabile al fatto che essendo il GPS fisso, in alcune fasce orarie si trova ad avere una copertura

satellitare che non gli permette di calcolare la posizione in maniera corretta, non ricevendo sufficienti segnali dai satelliti. Dopo questi brevi lassi di tempo infatti, ricomincia a lavorare regolarmente con errori entro il centinaio di metri, come previsto.

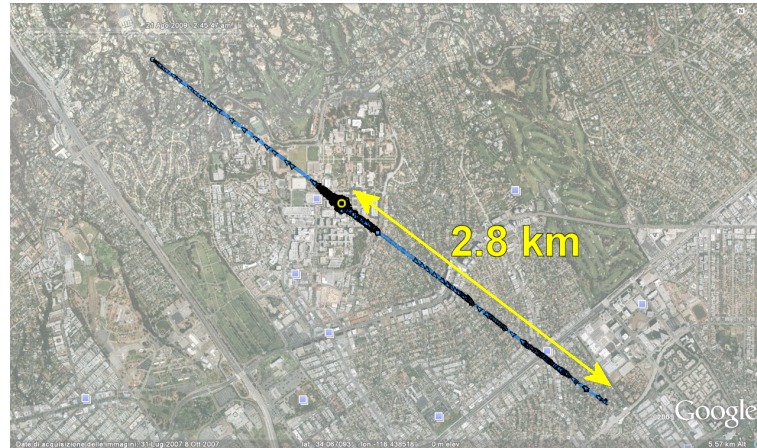


Figura 4.17: Visione completa delle posizioni rilevate dal ricevitore GPS sul balcone al terzo piano dell'edificio Boelter Hall in 2 ore di campionamenti

Questo esperimento ci permette di evidenziare come il comportamento del GPS, quando non ha una completa visibilità del cielo, sia alquanto variabile. È possibile che riesca a calcolare correttamente la posizione, come avviene per una frazione degli istanti di campionamento, come però è possibile che sia affetto da un errore, che può essere tollerabile o invece andare ad invalidare completamente il dato, in maniera però purtroppo del tutto imprevedibile da parte dell'utente.

L'unico sentore d'allarme in questo caso è una più bassa frequenza di campionamento, che indica che il ricevitore non è in grado di calcolare la posizione. Dai dati si nota infatti una correlazione fra il livello di imprecisione medio ed il numero di istanti in cui la posizione non è campionata. Del resto, è anche logico supporre che se in un istante il GPS non è addirittura in grado di calcolare l'informazione sulla posizione, non possa nell'istante successivo calcolarla in maniera precisa, con un errore contenuto entro la decina di metri.

Abbiamo visto quindi come il sistema di posizionamento globale *GPS* non possa essere considerato esatto nella totalità dei casi. I fattori che possono influenzarne la precisione sono molteplici, e trovano la loro giustificazione principalmente nella realtà fisica che circonda i ricevitori.

Proprio per questo motivo, l'utilizzo nello scenario urbano enfatizza i problemi, a causa dell'entropia introdotta dalla struttura complessa dello scenario stesso. Sono stati però compresi, grazie a diversi esperimenti pratici, i comportamenti errati più comuni in questo scenario; proprio da questa comprensione si partirà quindi per creare un sistema che sia tollerante ad essi, e possa perciò inserirsi nel campo dell'uso quotidiano.

Capitolo 5

Crowdsourcing

In questo capitolo si presenterà il significato della parola crowdsourcing, cercando di capire come mai questo termine stia diventando sempre più di moda, anno dopo anno. Inoltre, verranno analizzati i punti forti e le debolezze di questo metodo risolutivo, sfatando alcuni miti che girano attorno alla sua fama, ed affermando invece quali sono i punti cruciali che lo rendono, sotto certi aspetti, una forte innovazione.

Il termine crowdsourcing (da crowd = gente comune, folla; e outsourcing = esternalizzare una parte delle proprie attività) è un neologismo che definisce un modello di business nel quale un'azienda o un'istituzione richiede lo sviluppo di un progetto, di un servizio o di un prodotto ad un insieme distribuito di persone non già organizzate in un team [71].

Le reti sociali e l'evoluzione tecnologica stanno cambiando non solo il modo con cui le persone comunicano o quello con cui le persone lavorano, ma anche il modo con cui le imprese possono ridisegnare il perimetro delle loro attività di lavoro. Se ben sfruttato, il crowdsourcing velocizza i tempi di realizzazione del lavoro, diminuisce i costi e, soprattutto, crea valore aggiunto grazie alla struttura collaborativa della rete. Uno degli esempi più chiari, anche se non il più significativo, di questa tendenza è il Mechanical Turk di Amazon. Il sito consente a chiunque di pubblicare dei compiti da eseguire, fissare prezzo e tempo di esecuzione e sotto-

porlo a tutti gli iscritti al sito che, se completeranno il lavoro, verranno pagati per ciò che hanno fatto. Le proposte di lavoro pubblicate, chiamate Human Intelligence Tasks o Hit, sono svariate: sbobinatura di interviste in Mp3, valutazione dei risultati di un motore di ricerca o dell'usabilità di un sito, scrittura di brevi articoli e così via (figura 5.1).



Figura 5.1: Visione del crowdsourcing come la ricerca della soluzione ad un problema grazie al lavoro distribuito di un insieme di persone non organizzate fra loro

La “folla” può però essere usata anche per migliorare i propri prodotti. Nel febbraio 2007 il produttore di computer Dell ha pubblicato un sito chiamato IdeaStorm, con l'obiettivo di ricevere suggerimenti da parte degli utenti su come migliorare la propria linea di computer. Il sito ha avuto un buon successo, migliaia di navigatori hanno suggerito miglioramenti e caratteristiche tanto che la nuova linea Studio del produttore americano deriva direttamente dai suggerimenti ricevuti dagli utenti.

Ancora un esempio: Map Share è un sistema integrato nei navigatori Tom Tom che permette di modificare le mappe non aggiornate del proprio dispositivo e condividerle online con tutti gli utenti registrati al servizio. Secondo Tom Tom grazie a questo servizio sono state condivise 5 milioni di correzioni, migliorando la qualità del prodotto e l'esperienza degli utenti. Ma crowdsourcing non è solo un modo per migliorare il business di un'azienda ma può essere definito così qualsiasi cosa faccia leva sulla rete per suddividere in piccoli compiti grandi lavori e creare valore.

Il Crowdsourcing non è però da vedere soltanto come un modello di business, in cui un committente sceglie di affidarsi a una moltitudine di utenti, delle più svari-

ate caratteristiche e competenze, per ottenere la soluzione a un problema che fino a qualche anno fa avrebbe affidato a un dipendente specifico, o ad un team ben delineato. La definizione in termini economici della parola stessa gli dà un'accezione che è in qualche modo limitativa di quelle che sono le sue reali potenzialità. Un ruolo fondamentale del crowdsourcing infatti è la creazione di valore per l'utente finale, che diventa al tempo stesso fruitore e produttore di contenuti.

A settembre del 2007, l'imprenditore Steve Fosset si è perso con il suo aereo sul deserto del Nevada. Qualche giorno dopo DigitalGlobe, che possiede i satelliti che forniscono le foto della terra a Google Earth, ha fornito immagini aggiornate della zona. In 3 giorni, grazie al Web, 50 mila persone hanno esaminato oltre 300 mila immagini dell'area alla ricerca del noto miliardario. Steve Fosset è stato poi trovato solo ad inizio ottobre del 2008 ma la storia della ricerca collaborativa ha messo sulla bocca di molti una nuova parola: crowdsourcing, appunto. Non si può certo dire che questo sia stato strettamente un nuovo modello di business, ma piuttosto un primo passo nel portare nelle mani degli utenti la possibilità di risolvere partecipativamente problemi più grandi di quanto un uomo solo potrebbe risolvere, per lo meno in tempi utili.

Le parole di Dan Woods sono importanti in questa ottica per capire quali sono realmente i punti di forza della suddivisione del lavoro su una folla, e quali invece sono gli errori più comuni che si è portati a fare pensando al crowdsourcing, in modo tale da potersi concentrare nelle aree che dal crowdsourcing traggono reale beneficio, addirittura al punto da non poter essere sviluppate in maniera economicamente affrontabile senza ricorrere al crowdsourcing [78].

Nell'opinione popolare, e nella mente di milioni di persone, la parola crowdsourcing ha creato l'illusione che vi sia una folla che risolve i problemi meglio che singoli individui. La nozione di folla creatrice di soluzioni trova appiglio nel nostro desiderio di credere che lavorando insieme possiamo fare qualsiasi cosa, ma in termini di innovazione questo non è assolutamente vero. Non c'è reale folla nel crowdsourcing. Ci sono soltanto persone virtuose, solitamente dotate di talento unico ed altamente preparate che hanno lavorato per decenni in uno specifico campo. Frequentemente questi innovatori sono stati finanziati di continuo, fallimento dopo fallimento, in quanto è dalle loro fervide menti che scaturiscono nuove idee. La folla di per se non risolve nulla, non inventa nulla.

Ciò che realmente accade nel crowdsourcing così come è praticato in una varietà

di contesti, da Wikipedia all'Open Source alla ricerca scientifica, è che un problema è divulgato ad un ampio numero di persone con esperienza diversa. A questo punto, individui motivati dall'ossessione, dalla competizione, da motivi economici o anche da tutti e tre insieme, applicano il loro talento individuale per creare una soluzione.

Basta guardare ad alcuni casi di successo del crowdsourcing per vedere come la folla sia soltanto un'illusione.

Wikipedia per esempio sembra un buon esempio di un agglomerato di persone che hanno creato una grande risorsa. A una conferenza nel 2008 però, Dan Woods chiese al fondatore di Wikipedia, Jimmy Wales, informazioni riguardo come venissero realmente creati gli articoli presenti sull'enciclopedia online più famosa del mondo. Jimmy rispose che la grande maggioranza di essi sono il prodotto di un singolo individuo adeguatamente motivato. Una volta che l'articolo è stato creato, viene curato (corretto, migliorato ed eventualmente esteso) da varie persone differenti.

Alcuni articoli sono invece creazioni di gruppo che si sono evolute da una o due frasi iniziali. Ma se togliamo da Wikipedia tutti gli articoli che sono creazioni individuali, i rimanenti saranno davvero pochi rispetto alla completezza di informazioni a cui siamo abituati quando entriamo nell'enciclopedia.

Anche il caso dell'open source, se analizzato nel dettaglio, mostra un comportamento simile. Ad esempio, il suo esempio più eclatante, Linux, parte proprio dall'idea e dalla dedizione del geniale Linus Torvalds, che tutt'ora coordina il gruppo di programmatori che sviluppa il kernel Linux per conto dell'OSDL (Open Source Development Lab), un progetto di ricerca finanziato da industrie che hanno fondato parte del proprio business su GNU/Linux, come Computer Associates, Fujitsu, Hitachi, Hewlett-Packard, IBM, Intel, NEC [66].

Ciò che infastidisce così tanto Woods riguardo l'errata interpretazione che si sta diffondendo riguardo il crowdsourcing è proprio questa fede nella forza della folla, che finisce inevitabilmente per intaccare l'immagine dell'inventore, eroico per certi versi (figura 5.2). È necessario coltivare e finanziare gli inventori invece, dando loro il tempo di esplorare, provare e fallire. Una falsa idea del potere della folla invece riduce la motivazione verso questo tipo di investimento, con la supposizione che le ditte possano sfruttare la mente degli inventori a buon mercato.

Karim R. Lakhani, un professore associato alla Harvard Business School, chiama ciò che molti definiscono crowdsourcing semplicemente ricerca in broadcast.

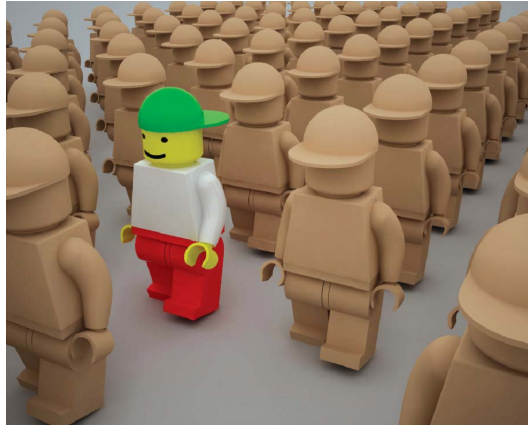


Figura 5.2: Visione del crowdsourcing come un ambiente in cui è il talento individuale di un singolo, che si differenzia dalla massa per capacità o motivazioni, a portare la folla alla soluzione

Dan Woods preferisce invece il termine ricerca virtuosa ad esempio. In fondo non conta il nome che si sceglie di dargli, ma è importante non cadere nell'errata supposizione che 10'000 Joes con medie capacità possano inventare prodotti migliori di Steve Jobs.

Ma allora, cos'è che realmente fa la forza del crowdsourcing, e cosa intendiamo noi esattamente con questo termine nel nostro lavoro? Il punto forte è la capacità di poter ottenere una moltitudine di dati, provenienti da utenti diversi. Questi dati rispecchiano i comportamenti, le abitudini, le situazioni in cui vengono a trovarsi nell'arco della loro giornata, o addirittura della loro vita, persone con caratteristiche che possono essere dalle più simili alle più diversificate fra loro. I dati raccolti possono essere del tipo più disparato, permettendo con elaborazioni successive di ricavare ciò che più interessa da essi. Si possono ad esempio fare analisi di mercato, analisi comportamentali, studi di mobilità. Tutto, insomma, può essere ricavato da una corretta raccolta di dati ed una ingegnosa analisi focalizzata. Come aggregare i dati e quali informazioni estrarre dalla loro analisi infatti è tutt'altro che immediato, e ciò che fa la differenza fra uno studio innovativo ed uno banale è proprio questo.

L'idea di utilizzare il crowdsourcing di tracce GPS per estrarre informazioni di alto livello è stata presentata recentemente in alcuni elaborati. Yu Zheng ad esempio propone di utilizzare tali dati per stimare il mezzo di trasporto utilizzato dall'utente mentre si sta spostando [81]. In questo modo, analizzando in tempo reale la traccia registrata dal suo GPS, si può dedurre se esso stia camminando, sia a bordo di

un'automobile o di un altro mezzo di locomozione con caratteristiche diverse.

Queste informazioni vengono poi usate per integrarsi con altri servizi a disposizione dell'utente, ad esempio la ricerca sul web, in modo da proporre i risultati più utili in relazione del mezzo di trasporto che si sta usando. Nel caso ad esempio che si cerchi una pizzeria in zona, verranno proposti prima i risultati a distanza di camminamento nel caso si stia camminando, mentre nel caso si sia a bordo di un'automobile, verranno presentati prima i risultati più comodi da raggiungere con quel mezzo, evitando ad esempio quelli in zone pedonali. In questo sistema in realtà il crowdsourcing non è presente nella fase descritta, ma bensì nella fase iniziale necessaria a modellare i comportamenti caratteristici della traccia GPS a seconda che l'utente sia a piedi, in bicicletta, in macchina o in autobus. Infatti, analizzando una moltitudine di tracce registrate da utenti diversi in varie situazioni, si costruisce il modello che verrà poi usato per identificare il mezzo di trasporto usato dal singolo utente in funzione della singola traccia GPS che esso sta registrando.

Un altro esempio di utilizzo del crowdsourcing di tracce GPS è invece quello proposto da Seth Rogers [69]. Qui l'idea di crowdsourcing è utilizzata direttamente in tutta la fase realizzativa, non solo per creare un modello iniziale. L'obiettivo di Rogers è quello di ampliare il modello stradale disponibile con informazioni aggiuntive riguardanti la dimensione e la disposizione delle differenti corsie di ogni strada. Analizzando infatti svariate tracce GPS passanti nelle diverse corsie, arriva ad inferire la topologia delle stesse.

Queste informazioni aggiuntive sul modello stradale permettono l'evoluzione di servizi sensibili alla specifica corsia di percorrenza. Alcuni esempi sono dispositivi di sicurezza passiva o attiva in grado di segnalare o intervenire quando il conducente sta attraversando erroneamente la linea di mezzzeria, o ancora sistemi di navigazione che permettono di guidare con adeguato anticipo il conducente nei cambi di corsia necessari per effettuare una successiva svolta, evitando così cambi di corsia improvvisi e pericolosi.

Un'ulteriore esempio di crowdsourcing applicato a tracce GPS è il database OpenStreetMap [50], che verrà descritto più nel dettaglio nei prossimi capitoli. OpenStreetMap è una mappa globale del pianeta creata tramite il crowdsourcing. Una moltitudine di utenti infatti partecipa attivamente al progetto, in una maniera molto simile a quanto succede per l'enciclopedia Wikipedia.

Ogni utente ha la possibilità di caricare i tracciati GPS che lui stesso ha re-

gistrato, piuttosto che procedere alla creazione, correzione o ampliamento di una parte della mappa attuale del pianeta. Chiaramente, questa modifica alla mappa può essere fatta in qualsiasi zona l'utente si ritenga competente, anche in micro aree poco esplorate, di cui magari lui è l'unico ad avere piena conoscenza. In questo modo, si può ottenere una mappatura dettagliata del pianeta, in cui l'interesse del singolo utente per una specifica zona va a vantaggio di tutta la comunità, cosa che solitamente non accade nei sistemi di mappe tradizionali in cui zone di poco interesse o poco abitate vengono solitamente trascurate.

Il lavoro presentato nel prossimo capitolo ha un denominatore comune con i tre appena presentati. Sfrutta infatti dati aggregati riguardo tracce GPS raw per ottenere informazioni di alto livello sulla topologia della mappa su cui tali tracce si collocano. La raccolta di dati distribuiti, propria del crowdsourcing, viene elaborata con la finalità di aumentare il potere informativo di una struttura consolidata, ovvero la mappa delle strade della città. Le informazioni ricavate riguardano i regolatori del traffico in prossimità degli incroci, arrivando ad inferire per ogni incrocio se esso è regolato da un semaforo o da uno stop, e ricavando nel primo caso le temporizzazioni caratteristiche dello stesso.

Capitolo 6

Estrazione informazioni topologiche: Semafori e Stop

In questo capitolo presenteremo nel dettaglio il lavoro svolto ed il sistema sviluppato. Il capitolo è articolato in diverse sezioni, corrispondenti rispettivamente ai vari step in cui si suddivide l'elaborazione, ed una sezione iniziale in cui si dà una visione generale di come è stato affrontato il problema e come si presenta sommariamente il procedimento risolutivo.

Il sistema sviluppato permette, nel suo insieme, di estendere le comuni mappe digitali oggi disponibili con informazioni topologiche sui regolatori di traffico presenti ad ogni incrocio urbano. Si è quindi in grado, con opportune elaborazioni, di inferire se un incrocio è regolato tramite segnali di Stop oppure tramite Semafori. Inoltre, in questo secondo caso, si è in grado di ricavare la temporizzazione dello stesso, per ogni singolo senso di marcia.

Per ottenere queste informazioni, si parte da semplici tracce GPS registrate da una moltitudine di utenti, con i comuni ricevitori disponibili per il mercato di massa. Queste tracce vengono analizzate al fine di capire quali sono stati i comportamenti degli utenti: dove hanno rallentato, dove si sono fermati, quali svolte hanno effettuato. Analizzando le tracce di più utenti in prossimità dello stesso incrocio, si può

fare un'analisi aggregata dei loro comportamenti, e da questa tramite opportune euristiche statistiche trovare le informazioni volute sulla mappa.

Per poter fare questo tipo di analisi comportamentale in relazione alla mappa sottostante, è necessario riuscire a legare in qualche modo le tracce alla struttura stradale sottostante. Per questo motivo, ogni traccia è elaborata prima con un algoritmo di *reverse geocoding*, che ne mappa ogni punto su una strada. Per sopperire agli errori di localizzazione di cui molti punti delle tracce soffrono, questo algoritmo di reverse geocoding viene poi evoluto in una versione più *intelligente*, che tiene conto del movimento del veicolo per localizzarlo, mettendo in relazione ogni punto con quelli che gli stanno attorno invece di considerarlo singolarmente per la sua localizzazione.

Le tracce utilizzate sono prese dal database di OpenStreetMap [50], per cui alcune non soddisfano i requisiti necessari al sistema. Per questo motivo, prima della loro analisi viene fatta una selezione preliminare.

Come visto, nel sistema si procede *mappando* le tracce sulle strade che percorrono. Proprio per questo motivo, è fondamentale disporre di una mappa su cui fare questa mappatura. La creazione di questa mappa è il primo passo del sistema, che partendo da una struttura fisica specifica del database di mappe utilizzato (TIGER [38] o OpenStreetMap) ne crea una rappresentazione logica strutturata semplicemente in strade ed incroci.

La figura 6.1 schematizza i passaggi logici appena descritti, così come si susseguono nel sistema; i prossimi paragrafi affronteranno nel dettaglio gli stessi uno ad uno.

6.1 Pre-processing della mappa

La miglior rappresentazione per una mappa digitale è un grafo orientato, con un insieme di nodi, rappresentanti gli incroci, collegati tramite archi orientati che rappresentano le strade ed il loro senso di marcia. In realtà, questa è proprio la rappresentazione utilizzata nella maggior parte dei database di mappe creati per rappresentare la struttura fisica delle strade, con finalità di navigazione o visualizzazione.

Mentre questa rappresentazione va molto bene nei precedenti due casi, non è però ottimale per i nostri scopi. Per l'estrazione di informazioni dalle tracce, serve infatti una rappresentazione della mappa che sia più vicina alla rappresentazione

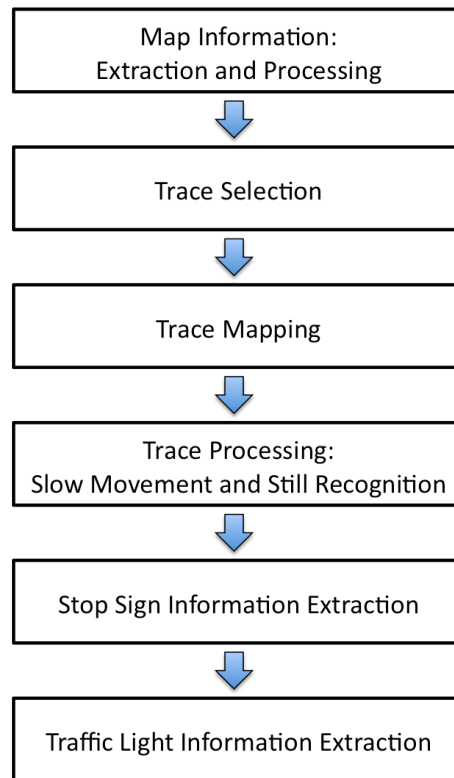
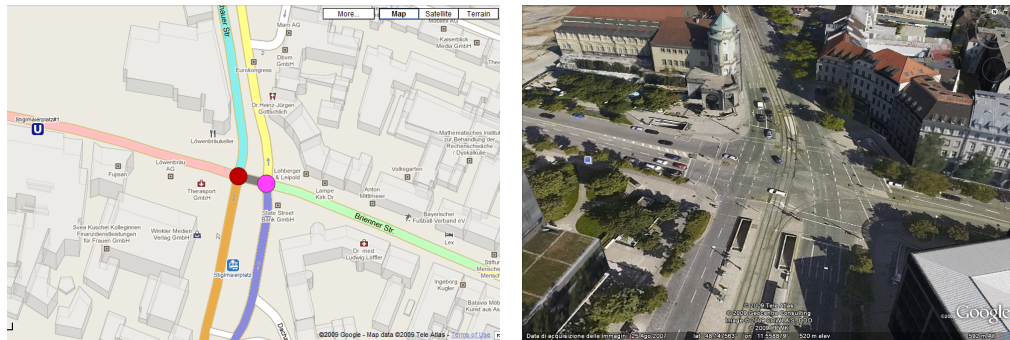


Figura 6.1: Schematizzazione del flusso di operazioni svolte nel sistema proposto

logica di strade ed intersezioni che c'è nella nostra mente. Il concetto di *Incrocio e Strada* deve infatti essere qualcosa di atomico, che possa essere direttamente usato senza ambiguità nel sistema.

L'incrocio mostrato in figura 6.2(a) sarebbe infatti salvato in un database comune, come ad esempio il database Open Street Map, come due differenti nodi e ben 7 segmenti rappresentanti altrettante sezioni di strada. Invece, logicamente, non è altro che un incrocio in cui convergono 4 strade, di cui due di esse con le carreggiate divise da uno spartitraffico solido (figura 6.2(b)).

Anche dal punto di vista urbano e della regolazione del traffico quello precedente è considerato come un singolo incrocio, ed è infatti regolato da un singolo semaforo. Si deve quindi fare in modo di creare una rappresentazione della mappa in memoria che rispecchi il punto di vista della regolazione del traffico, e quindi il punto di vista comune di un utente esterno. Si crea perciò una rappresentazione astratta formata da *Crossroads* e *Ways*, che possono comprendere, rispettivamente, diversi nodi e diversi archi, che verrà utilizzata nell'analisi del moto delle tracce



(a) Mappa digitale

(b) Vista satellitare

Figura 6.2: Incrocio fra Dachauer Str. e Brienner Str. a Monaco di Baviera, così come appare in una mappa digitale ed in una vista reale

GPS.

Questa rappresentazione può essere ottenuta partendo da qualsiasi mappa digitale, anche se il procedimento chiaramente è diverso da caso a caso. Per i dettagli implementativi della creazione di questo database si rimanda all'appendice A. Nel seguito di questo elaborato si considereranno due database di mappe universalmente noti e diffusi, quello chiamato TIGER/Line Shapefiles, estratto dal Census Bureau's MAF/TIGER database e reso pubblico gratuitamente, e OpenStreetMap, anch'esso pubblico e gratuito, ma ottenuto grazie alla collaborazione di migliaia di utenti in un progetto di tipo Open Source.

6.2 Selezione delle tracce utili

Le tracce analizzate nel sistema sviluppato, sono in maggior parte state prese dal database pubblico di OpenStreetMap [50]. Questo database raccoglie più di 300'000 tracce diverse, caricate da utenti di tutto il mondo. La finalità con cui nasce questo progetto Open Source vanno in realtà oltre questa raccolta di dati: l'obiettivo è ottenere una mappa globale di tutto il pianeta, con l'aiuto degli utenti. Le tracce caricate infatti vengono utilizzate come punto di partenza per disegnare la struttura delle strade su cui esse sono state registrate, o in alcuni casi, come per gli Stati Uniti d'America, per correggere e migliorare database esistenti.

Le tracce caricate sul sito dagli utenti, non hanno fundamentalmente alcun tipo

di vincolo, se non il formato che deve essere aderente allo standard GPX [53] (documentato in appendice B.3). Per questo, possiamo trovare tracce registrate a piedi, in macchina, in moto, fermi, ecc. che non sono tutte utili per gli scopi di questo lavoro. Inoltre, vi possono essere tracce con tempi di campionamento molto variabili, o con solo punti di interesse salvati in maniera scollegata l'uno dall'altro, o ancora, file che contengono diverse tracce agglomerate in una unica, nonostante siano state registrate a chilometri di distanza.

Per tutti questi motivi, è necessario svolgere un processo selettivo iniziale che, partendo dall'insieme di tutte le tracce disponibili, crei un sottoinsieme di tutte quelle effettivamente utili ai nostri scopi, che saranno poi le uniche considerate nel seguito del sistema. La caratteristica più cruciale da cui iniziare la selezione è sicuramente la frequenza di campionamento della traccia, in seguito denotata anche come *Sampling Rate*.

Come visto nel capitolo 4.3, una mancanza di campionamento è strettamente correlata con la mancanza di precisione nella localizzazione. Non tutti i dispositivi GPS però hanno una frequenza di campionamento di esattamente un secondo, come quello preso in esame nei nostri test. Alcuni dispositivi di navigazione per il mercato di massa hanno frequenza di campionamento, in caso si registri la traccia del proprio percorso, variabili, in un range che può estendersi anche ad una decina di secondi. Qualunque sia la frequenza di campionamento propria del dispositivo però, essa è generalmente costante se esso è in grado di localizzarsi correttamente.

Potremo quindi avere dispositivi che campionano ogni secondo, piuttosto che altri che campionano ogni 4 secondi, ma la cosa certa è che nel caso ideale di segnale GPS sempre ottimo, questa frequenza di campionamento deve essere costante per tutta la traccia.

Per prima cosa quindi, per ogni traccia andiamo a calcolare la frequenza di campionamento più comune, facendo un semplice clustering delle occorrenze dei campionamenti a distanza di un secondo. Avremo quindi il numero di campionamenti a 1 secondo di distanza l'uno dall'altro, a 2 secondi, a 3 e così via. Così facendo abbiamo quella che si suppone essere la frequenza propria del dispositivo (\overline{FC}), e possiamo andare a discriminare quali tracce sono abbastanza costanti attorno alla loro frequenza propria di campionamento, e quali invece no.

Per fare questo si è scelto di considerare buone le tracce che hanno almeno l'80% dei campionamenti che ricadono in una fascia di valori compresi tra $\frac{1}{2}\overline{FC}$

e $\frac{3}{2}\overline{FC}$, indicando quindi un campionamento abbastanza costante, sintomo di condizioni di registrazione in cui la precisione di localizzazione si può considerare accettabile. È da notare come questa sia soltanto una prima scrematura, in cui si scartano le tracce visibilmente troppo imprecise, che porterebbero più errori ed imprecisioni che né informazioni utili.

Si è consapevoli che l'utilizzo di questa flessibilità nella definizione di costanza della frequenza di campionamento lasci ancora vive situazioni di imprecisione nel campionamento, ma esse verranno corrette in seguito. Si deve ora cercare di avere un adeguato tradeoff, in cui si mantengono da una parte soltanto le tracce che con ragionevole probabilità saranno utilizzabili in seguito, e dall'altra si cerca di sfruttare il più alto numero possibile delle tracce disponibili sul database.

Oltre a questo, in funzione dell'interesse del sistema per comportamenti dell'utente quali rallentamenti e fermate, è necessario avere tracce con un sampling rate elevato, in quanto ad esempio se abbiamo informazioni soltanto ogni 10 secondi, non possiamo dire niente su cosa è successo fra due campionamenti successivi, poiché solitamente rallentamenti e fermate avvengono in finestre temporali di alcuni secondi. Per questo motivo, non vengono considerate tutte le tracce che hanno una frequenza di campionamento media maggiore di 4 secondi, che è verificato essere un valore limite per le informazioni di nostro interesse.

Al momento della lettura della traccia dal file, vengono calcolate per ogni suo punto alcune informazioni relativamente al punto successivo, e precisamente:

- La distanza dal punto successivo;
- Il tempo che intercorre col campionamento successivo;
- La velocità in direzione del punto successivo;
- L'accelerazione in direzione del punto successivo;
- L'azimut del movimento verso il punto successivo;

Queste informazioni vengono usate qui per discriminare ulteriormente le tracce che non sono registrate da veicoli in movimento nel loro ambiente abituale, come per esempio quelle registrate da pedoni, che verranno scartate, basandosi sulla velocità massima calcolata sull'intera traccia. Per maggiori dettagli tecnici riguardo alcune accortezze ulteriori necessarie nel corretto utilizzo di queste tracce, si rimanda all'appendice apposita B.3.

L'azimut (dall'arabo as-sûmut, le direzioni), citato poco fa, è un angolo che identifica esattamente la direzione del segmento congiungente due punti, direzionato dall'osservatore verso il punto di interesse. Nella convenzione, un segmento orientato verso il polo nord ha angolo 0° , e gli angoli crescono in senso orario.

6.3 Mappatura delle tracce con Reverse Geocoding intelligente

Una volta ottenuto un set di tracce significative ed adeguatamente affidabili, come descritto nel paragrafo precedente, è possibile andare a mapparle sul sistema stradale sottostante. Questo approccio si scompone in tre fasi, descritte accuratamente nel loro procedimento nei prossimi tre paragrafi.

Per prima cosa ci si basa su un efficiente algoritmo di *Reverse Geocoding* da noi implementato per mappare ogni singolo punto della traccia sulla mappa, poi si vanno a correggere gli errori di mappatura facendo un attento uso della correlazione spazio - temporale dei punti della traccia consecutivi fra loro, ed infine si correggono le discontinuità macroscopiche della traccia stessa.

Il Reverse Geocoding è il procedimento per cui si mappa un singolo punto, partendo dalle sue coordinate lat - lon, sulla strada su cui esso risiede nella topologia stradale. Questo procedimento in linea di principio richiede una scansione dell'intero database delle strade alla ricerca di quella più vicina al punto analizzato, ma la cosa è chiaramente molto inefficiente. Per questo è stata implementata un'indicizzazione che permetta di effettuare questa ricerca intensiva soltanto su una frazione di strade molto piccola rispetto al totale.

La mappatura di ogni singolo punto indipendentemente dal suo contesto però ha alcuni svantaggi, due su tutti:

1. Le tracce GPS sono affette da errori di campionamento improvvisi, che possono risultare in un moto apparente di un veicolo che salta da una strada all'altra, in maniera del tutto irrealistica (figura 6.3). Infatti, poiché una macchina è in movimento, l'ambiente circostante può improvvisamente oscurare alcuni dei satelliti, causando un conseguente calcolo della posizione errato (paragrafo 4.2).



Figura 6.3: Esempio di ricevitore GPS che per alcuni secondi da dei campionamenti errati, evidenziati in rosso, fuori dal percorso evidenziato invece dai campionamenti corretti, in verde

2. Una strada nelle mappe digitali è rappresentata da un segmento su un piano, senza alcuna informazione sullo spessore della strada. Per questo, in prossimità di un'intersezione, il punto di una traccia in moto lungo una strada dritta sarà facilmente mappato sulla strada incrociante [39, 40]. Difatti, è facile che la macchina si muova leggermente distante dal segmento rappresentante la strada (e questo è tanto più vero quanto più la carreggiata è larga ed ha un numero elevato di corsie), per cui raggiungendo l'intersezione il segmento rappresentante la strada incrociante può essere geometricamente più vicino, come mostrato in figura 6.4

Per queste ragioni è necessario introdurre un algoritmo di reverse geocoding che tenga conto della posizione spazio - tempo in cui si trova il punto da localizzare. Quello qui realizzato fa una doppia scansione della traccia, prima in avanti e poi all'indietro. In entrambi i casi considero, oltre al punto attuale, quello precedente. In questo modo al primo passaggio propago in avanti le informazioni sulla traiettoria passata, e al secondo propago all'indietro informazioni sul movimento futuro.

L'algoritmo tratta in maniera diversa gli istanti in cui il veicolo è fermo rispetto a quando esso è in moto (nel paragrafo 6.4.3 è spiegato come viene identificato il caso di un veicolo fermo). Nel primo caso vengono privilegiate soluzioni in cui la mappatura coincida con quella del punto precedente, mentre nel secondo si terrà conto

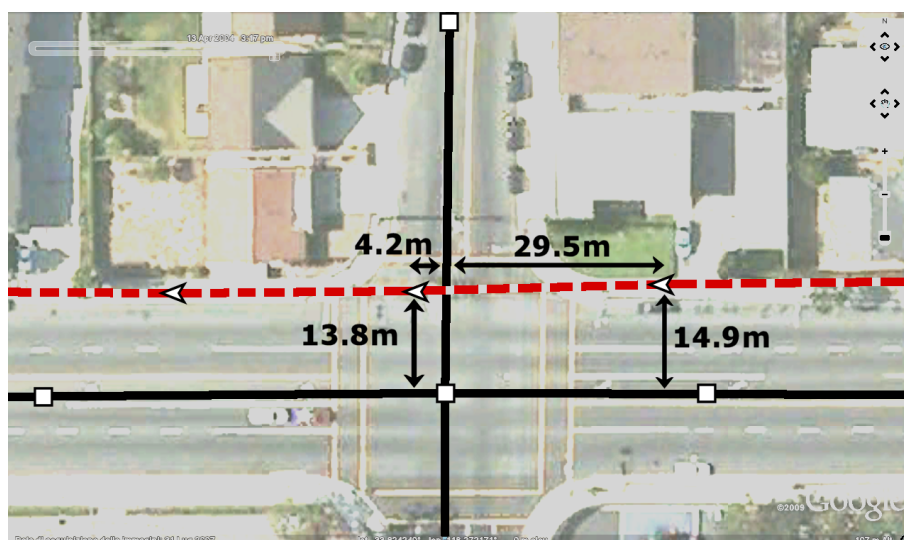


Figura 6.4: Errore di un comune procedimento di reverse geocoding in prossimità di un incrocio. Gli edges sono evidenziati in nero e la traccia percorsa in rosso

anche dell'orientamento delle strade, privilegiando poi quelle che sono collegate alla precedente. Il procedimento nel dettaglio è spiegato nel paragrafo 6.3.2.

Al termine di questa fase di reverse geocoding intelligente, è comunque possibile che alcuni punti non siano stati mappati su nessuna strada. Inoltre è possibile che ci siano marcati gap temporali fra due punti consecutivi, dovuti alla tolleranza sulla frequenza di campionamento usata nella precedente selezione delle tracce.

Per questo è necessario applicare un passaggio volto all'eliminazione delle discontinuità temporali, nonché dei punti privi di una corretta mappatura o mappati su un'autostrada, seguiti dall'eliminazione delle discontinuità spaziali. L'eliminazione di questi tipi di discontinuità è necessario perché nella successiva analisi non vogliamo tracce che balzino da una strada a un'altra, ma bensì vogliamo tracce che seguano percorsi attinenti alla topologia stradale, per cui si va a spezzare la traccia in due distinte ogni volta che ci si trova in presenza di una discontinuità.

6.3.1 Reverse Geocoding

Vediamo ora più nel dettaglio come è strutturato l'algoritmo di Reverse Geocoding puntuale implementato. Una prima parte dell'algoritmo consiste nel memorizzare la mappa di tutte le strade in una struttura dati che permetta un'indicizzazione effi-

ciente, in modo tale da non dover scorrere la lista di tutte le strade in memoria ogni volta che si deve posizionare un punto.

L'efficienza è fondamentale, perché questa operazione deve essere svolta in media migliaia di volte per ogni singola traccia (essendo questo il numero medio di punti da cui è formata) e dalla sua efficienza dipende quella generale del sistema.

La seconda parte dell'algoritmo riguarda invece il reperimento del risultato, ovvero come navigare questa struttura dati alla ricerca delle informazioni necessarie.

Creazione delle strutture dati Per prima cosa, viene letta la mappa della zona di interesse, e vengono salvate in due dizionari distinti gli *Edge* ed i *Nodi*, con tutte le informazioni ad essi associate (nome della strada, tipo di strada, range di civici, ecc.). Ogni edge rappresenta un pezzo di strada, e corrisponde fisicamente ad una linea spezzata, identificata da tutti i vertici interni ed i due vertici estremi. Ogni nodo invece rappresenta un vertice estremo di un edge, ed è solitamente condiviso da più edges, che si incrociano quindi su esso (si rimanda per un maggior approfondimento all'appendice A specifica della struttura delle mappe digitali). L'importante in questa fase è sapere che ogni edge ha l'informazione dei suoi vertici intermedi, ed ogni nodo ha l'informazione di quali edges lo hanno come vertice.

Sostanzialmente, cercare la strada più vicina ad un punto dato, corrisponde a trovare l'edge che fra i suoi segmenti costituenti ne ha uno che è il più vicino in assoluto al punto, rispetto a tutti i segmenti costituenti tutti gli edges del database.

Per questo, creiamo ora una lista di nodi che oltre a contenere tutti i nodi del precedente dizionario, contiene anche le informazioni riguardanti tutti i vertici intermedi; allo stesso modo, creiamo anche un nuovo dizionario che invece che contenere tutti gli edge, contiene adesso tutti i segmenti parziali che li compongono, che verranno chiamati *partial edges*, o più semplicemente *edge parts*. Questi edge parts contengono l'informazione sull'edge che contribuiscono a creare, sui propri due vertici, sulla propria lunghezza e sul proprio azimut. Nel caso che un edge part sia troppo lungo, verranno inseriti poi al suo interno tanti *fake nodes* quanti ne servono per far sì che venga spezzato in sotto segmenti che non superino la lunghezza desiderata (che si vedrà in seguito essere il doppio della precisione voluta nella localizzazione), e tali fake nodes verranno anch'essi aggiunti alla lista dei vertici intermedi precedente.

Ogni nodo, al momento della sua creazione, viene memorizzato con le sue coordinate in gradi di latitudine e longitudine. Oltre a queste però, vengono calcolate e salvate le sue coordinate anche in un altro sistema di coordinate, quello UTM (*Universal Transverse Mercator System*) che permette, entro zone confinate, di utilizzare la geometria euclidea per calcolare la distanza fra un punto ed un segmento, invece che la più complessa geometria geodetica terrestre. Nel caso tre punti si trovino malauguratamente in zone UTM differenti, si ricorrerà al calcolo delle coordinate ENU (*East, North, Up*) per quegli specifici punti quando dovrò calcolare la distanza fra uno ed il segmento delineato dagli altri due, utilizzando come origine e punto di tangenza del piano il punto interessato. Per maggiori dettagli su questi due sistemi di coordinate si rimanda all'apposita appendice B.4.

La lista di nodi così creata viene ordinata in maniera assoluta secondo la latitudine, ed in caso questa sia uguale fra due nodi, successivamente in base alla longitudine. Viene utilizzata la latitudine come primo criterio di ordinamento perché la relazione che lega una distanza in metri con una differenza di latitudine in gradi si mantiene grosso modo costante su tutto il globo, e questa caratteristica sarà utile in seguito per l'ottimizzazione.

La terra infatti non è una sfera perfetta, bensì è leggermente schiacciata ai poli. Questo fa sì che 1° di latitudine sia leggermente più grande, se trasformato in metri della superficie terrestre, ai poli piuttosto che né all'equatore. Si passa infatti da una lunghezza di 110.5673 km all'equatore, a 111.6993 km ai poli. Se si applica un ragionamento analogo alla longitudine invece, si osserva come i relativi meridiani convergano sui poli. Conseguentemente, 1° di longitudine avrà un valore molto diverso in base al parallelo su cui è calcolato, andando da una lunghezza di 111.321 km se calcolato sull'equatore, a 0 su uno dei due poli [61, 62].

Ricerca del risultato di Reverse Geocoding Ora che abbiamo creato le strutture dati ordinate che contengono tutte le informazioni su edge e nodi della mappa, vediamo come viene effettuata l'operazione di localizzazione di un punto, nel seguito definita più semplicemente *lookUp*.

Per prima cosa dobbiamo introdurre qui un parametro che va a incidere su questa fase, ovvero la *precisione* voluta nel reverse geocoding. Per precisione qui intendiamo la tolleranza che utilizziamo per dire che un punto sia localizzato o meno su una strada. Abbiamo visto come gli errori del GPS e l'assenza di informazioni

sulla larghezza di una strada, possano portare un punto ad esser campionato a vari metri dall'edge che rappresenta la strada. Non è però sensato pensare di associare ad una strada un punto che sia a centinaia di metri da essa, ad esempio al centro di un campo, soltanto perché il relativo edge è il più vicino al punto.

Si sceglie quindi un valore di tolleranza che rappresenta la distanza massima che un punto può avere da un edge per essere considerato situato su quella strada. Come valore di tolleranza sono stati scelti 25 metri, lasciando comunque sempre aperta la possibilità di cambiarlo ad ogni esecuzione dell'applicazione; tale valore può ad esempio ridotto a 20 o 15 metri nel caso la mappa presenti strade molto strette o vicine, in quanto comunque un errore maggiore non sarebbe tollerabile perché posizionerebbe i punti su strade vicine a quella reale.

Andiamo ora, tramite *binary search*, a trovare la posizione in cui si andrebbe ad inserire il punto analizzato all'interno della lista ordinata di nodi creata in precedenza, sulla base della sua latitudine. Da questa posizione, selezioniamo tutti i nodi della lista che sono compresi in una fascia orizzontale fra 25 metri nord e 25 metri sud del punto (perché 25 metri è la precisione che abbiamo deciso prima, ma può essere qualsiasi valore di precisione). Per fare questo, si sfrutta il fatto di aver precedentemente ordinato i nodi in base alla latitudine, in quanto si può facilmente trasformare la precisione/tolleranza data in metri nel suo corrispondente valore in gradi di latitudine, in maniera diretta per tutti i punti del globo.

Essendo la fascia di dimensione molto ridotta, i punti al suo interno saranno relativamente pochi, ed essendo la lista ordinata è facile recuperare i punti che ci interessano. A questo punto, sapendo che se un partial edge si trova a distanza minore di 25 metri dal punto considerato, almeno un suo vertice od un fake node si troverà nell'insieme di nodi appena estratto, possiamo procedere nel seguente modo.

Per ogni nodo dell'insieme, si considerano tutti i partial edge ad esso correlati (per cui cioè esso è un vertice o un fake node), e si calcola la distanza fra il punto cercato ed il partial edge. Una volta eseguita l'operazione, si considera il punto situato sul partial edge più vicino, e conseguentemente sull'edge ad esso associato. Se però la distanza è superiore alla precisione accettata (25 metri in questo caso), il punto si dichiara non situato su alcuna strada e il reverse geocoding non ritorna alcun risultato.

La precedente assunzione sui nodi presenti nella fascia analizzata, fa capire il

perché prima si fossero inseriti dei fake nodes laddove un partial edge superasse la lunghezza doppia della precisione voluta. Infatti, in questo modo, siamo sicuri che anche nel caso peggiore (segmento orientato nord-sud), se il segmento è a meno di *precisione* metri dal punto, ci sarà sempre un punto associato al segmento che ricade nella fascia di interesse, e quel segmento verrà considerato nel successivo confronto di distanza.

Oltre alla funzione di *lookUp* appena descritta, è stata definita anche una sua variante, denominata *lookUpAll*, che restituisce l'insieme di tutti gli edge parts a meno di 25 metri dal punto considerato (se ve ne sono), ordinati dal più vicino al più lontano. Questo metodo sarà utilizzato in seguito.

6.3.2 Reverse Geocoding Intelligente

Come visto in precedenza, un semplice algoritmo di Reverse Geocoding che localizzi i punti in maniera indipendente l'uno dall'altro non è abbastanza nell'analisi di tracce GPS, in quanto non tiene conto di un'informazione fondamentale: ogni punto dipende in maniera diretta dal punto precedente, o dall'insieme dei punti precedenti. Per questo, per ottenere un risultato più preciso, si deve tenere conto di questa relazione. È in quest'ottica che è stato progettata questa evoluzione del precedente algoritmo, che verrà identificata col nome di *lookUp-InTrace*.

Dal nome si può già intuire come non venga toccata la prima parte dell'algoritmo, in cui si creano le strutture dati, ma si lavori soltanto sulla seconda parte, di reperimento del risultato.

Per localizzare ogni punto della traccia sulla strada corretta, l'algoritmo applica due scansioni della traccia distinte, una in avanti ed una all'indietro, in cui vengono propagate rispettivamente informazioni sul passato e sul futuro movimento. Le due scansioni svolgono le stesse operazioni, per cui si descriverà ora nel dettaglio soltanto la scansione in avanti (*forward*).

Definiamo per semplicità come *pres* il punto presente che stiamo analizzando, e come *past* quello precedente, che supponiamo già analizzato e con una propria localizzazione.

Per prima cosa, ricaviamo la lista di tutti gli edge parts candidati alla localizzazione di *pres* con il metodo *lookUpAll* descritto in precedenza. Se tale lista è vuota, non possiamo localizzare questo punto e passiamo ad analizzare il successi-

vo. Supponendo che invece non sia così, procediamo distinguendo due casi: il punto si trova in una situazione di movimento; oppure il punto si trova in una situazione di assenza di movimento, in cui ad esempio il veicolo è fermo (sarà descritto nel prossimo paragrafo come si riesce a discriminare fra una situazione di movimento ed una di stallo).

Poiché l'informazione sul movimento coinvolge due punti, avremo in realtà alcune sfaccettature in più della situazione: possiamo infatti avere, oltre a un punto che si trovi in mezzo a una serie di punti fermi o a una di punti in movimento, anche un punto che sia a cavallo fra una serie di punti in movimento ed una di punti fermi, essendo così dotato di una direzione di arrivo ma non di una di uscita dalla posizione attuale; o viceversa, essendo a cavallo fra una serie di punti fermi ed una di punti in movimento, avendo una direzione di moto futuro ma non una di arrivo.

Per comodità, definiamo il primo caso quello di un punto che *si fermerà*, ed il secondo quello di un punto che *si muoverà*. Ricordiamo a questo punto che durante la scansione, si ha conoscenza soltanto del punto precedente, e che ogni punto contiene al suo interno la direzione del segmento che lo congiunge con il punto successivo nella traccia (azimut); in questo modo, per ogni punto *pres* conosciamo l'azimut del passo fatto per arrivare lì, che è l'azimut del punto *past*, e del passo che si farà per spostarsi da lì, che è l'azimut del punto *pres* stesso.

Nel caso il punto sia in una situazione di *movimento completo*, ovvero sia un punto in movimento situato in una serie di punti tutti in movimento, tengo tutti gli edge parts candidati che hanno azimut simile a quello di *pres* o di *past*, ordinandoli in base alla differenza di azimut dal movimento più simile fra i due. Nel dire che due azimut sono simili, qui si intende che la differenza fra i due valori è meno di 30° (in modulo 180 in quanto non interessa il vero ma solo la direzione della strada).

Nel caso il punto sia in una situazione di *stallo completo*, ovvero, il punto è fermo e si trova in una serie di punti definiti come fermi, il punto deve trovarsi sulla stessa strada su cui si trovava quello precedente. Se fra gli edge parts candidati quindi, ve ne è uno facente parte dello stesso edge su cui si trova il punto *past*, verrà scelto quello e la localizzazione è terminata, altrimenti non si può localizzare il punto e si passa a quello successivo. Anche nel caso il punto precedente non avesse localizzazione, non si può trovare un risultato e si passa al punto successivo.

Nei casi intermedi spiegati prima, se abbiamo un punto che *si fermerà*, lo consideriamo come un punto in movimento, utilizzando però solo l'informazione sulla

sua direzione di arrivo. Se invece il punto *si muoverà*, lo consideriamo dapprima come un punto in movimento utilizzando solo la direzione di moto futuro, e nel caso che questa non dia alcun risultato utile, lo consideriamo come un punto fermo e cerchiamo un match sulla stessa strada del punto precedente. Il comportamento apparentemente diverso nei due casi, trova una sua giustificazione nella doppia scansione della traccia, che al secondo passaggio tratterà i due casi in maniera complementare. La ragione del privilegiare la visione di questi punti come in movimento al primo tentativo, risiede nel cercare sempre di ottenere un risultato che sia al contempo corretto ed utile per l'analisi del punto successivo, identificando il prima possibile un cambiamento di strada.

A questo punto, come visto, nel caso che il punto fosse fermo abbiamo completato la sua analisi, sia che si sia stati in grado di localizzarlo, sia che non si sia stati in grado. Nel caso invece fosse in movimento, si procede ulteriormente. Se fra gli edge parts soltanto uno ha superato la prova del confronto degli azimut, viene scelto quello.

Se nessuno l'ha superata invece, per cui non vi è alcun edge vicino che sia all'incirca orientato come il moto entrante o uscente dal punto, è possibile che il punto si trovi nel mezzo di una svolta, in quanto è possibile che svoltando la direzione del moto in alcuni punti non sia simile a quella delle strade coinvolte (figura 6.5). In questo caso, il punto dovrebbe trovarsi o sulla strada in cui era prima di iniziare la svolta, o su quella che sarà una volta terminata la svolta. Per questo, ricordando il discorso delle due scansioni e della conoscenza solo passata, se l'edge part più vicino fra quelli possibili appartiene alla strada in cui faceva match il punto *past*, questo sarà il risultato della localizzazione del punto *pres*, altrimenti la localizzazione sarà lasciata nulla e si passa ad analizzare il punto successivo.

Nel caso più generale in cui più di un edge part sia abbastanza vicino al punto ed abbia un orientamento abbastanza simile al suo moto, devo scegliere il match migliore fra quelli possibili. Qui per prima cosa, partendo dall'edge part più vicino al punto, cerco una soluzione che si trovi sulla stessa strada del punto precedente, o comunque sul suo proseguimento. Questo perché, comunque, cerco di evitare di saltare da una strada ad una sua parallela ad esempio, privilegiando una continuità di moto, abbastanza comune nell'ambiente urbano (figura 6.6).

Se questo non è possibile, non essendoci alcun match con la strada del punto precedente, procedo cercando un match su una strada collegata alla precedente,



Figura 6.5: Esempio di traccia durante una svolta. In rosso è evidenziato il punto in cui il moto non è orientato come nessuno dei due edges su cui si muove il veicolo



Figura 6.6: I punti in rosso si trovano geograficamente più vicini alla strada parallela alla freeway, ed anche il loro moto è più simile all'orientamento di essa che né a quello della freeway. Privilegiando però un moto *connesso* topologicamente, essi verranno localizzati dal *reverse geocoding intelligente* sulla freeway, come i precedenti

quindi raggiungibile senza balzi, privilegiando però ora quella con orientamento

più simile al moto. Se trovo un risultato, lo salvo e passo al punto successivo.

Come visto, non è però scontato che si riesca a trovare un matching valido per un punto; è facile notare come sia praticamente impossibile trovare una soluzione nel caso il punto precedente non abbia fatto matching, in quanto quasi tutti i criteri di matching coinvolgono un confronto con il valore di matching del punto precedente. Per questo motivo, è stato introdotto un metodo con cui si può inizializzare l'algoritmo, che entra in funzione all'inizio della traccia ed ogni qualvolta si perde il fixing per uno o più punti.

Questa inizializzazione ha dei requisiti di precisione abbastanza stringenti riguardo il calcolo della posizione da parte del ricevitore, perché un errore in questa fase si protrarrebbe in avanti su tutte le decisioni che prendiamo anche nei punti successivi, ed è quindi meglio rimandare un po' la decisione di come localizzare i primi punti piuttosto che prendere una decisione errata, che rischia di andare a danneggiare tutti i calcoli successivi.

L'inizializzazione si svolge in questo modo: se abbiamo il punto in movimento, consideriamo l'edge part che gli è più vicino, e vediamo se il moto per arrivare in questo punto è simile all'orientamento di tale segmento. Qui, come detto, il vincolo è più stringente, per cui si considerano simili i due azimut se hanno una distanza angolare minore di 5° . In caso affermativo salviamo l'informazione di matching come provvisoria, e passiamo al punto successivo. Se questa condizione si verifica per tre volte consecutive, e l'edge più vicino è sempre lo stesso, significa che il movimento è strettamente posizionato su di lui, ed il posizionamento diventa stabile, e non più provvisorio.

I punti che non hanno trovato una loro corretta localizzazione in questa prima scansione, verranno localizzati dalla seconda scansione, che propaga all'indietro le informazioni sul moto futuro. Questa seconda scansione (*backward*) è identica alla precedente, ad esclusione del fatto che necessita anche del punto successivo *next* a quello analizzato (che è in realtà il precedente nella traccia così come è salvata in memoria), perché le informazioni di distanza ed azimut in avanti sono memorizzate lì, e non avrebbe senso ricalcolarsele.

In figura 6.7 è mostrato come l'algoritmo corregge la traccia presentata all'inizio del paragrafo, localizzando correttamente i punti in azzurro sulla strada percorsa, ed associando invece una localizzazione nulla ai punti fuori percorso, in rosso.

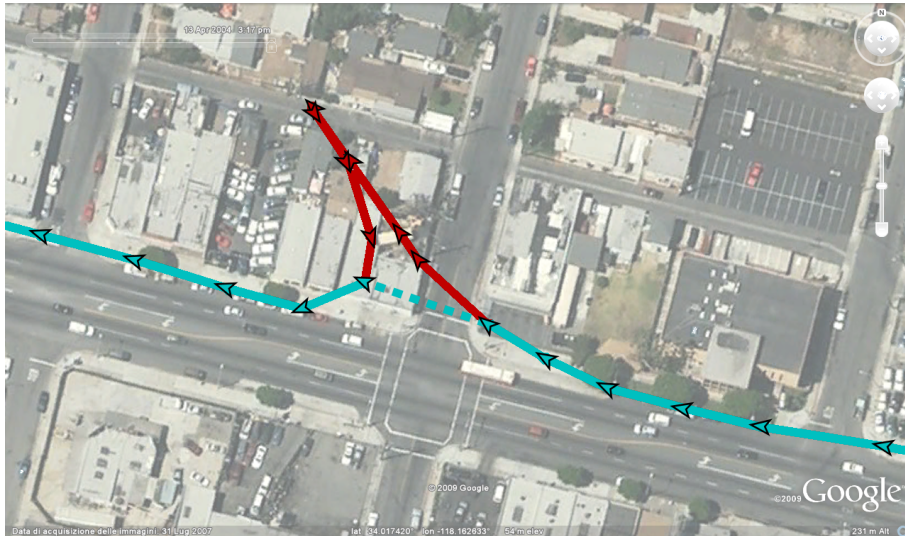


Figura 6.7: Esempio di come l’algoritmo per il reverse geocoding intelligente riesca a correggere una traccia con fixing del ricevitore errati. I punti errati e la traiettoria che generano, in rosso, sono eliminati, mentre quelli in azzurro vengono mappati correttamente sulle due strade in cui avviene lo spostamento

6.3.3 Eliminazione di discontinuità temporali o di percorso dalle tracce

A questo punto, tutti i punti possibili sono stati identificati sulla relativa strada della mappa. È però ancora possibile che ci siano:

1. Discontinuità temporali fra due punti successivi.
2. Punti che non fanno match su nessuna strada della mappa.
3. Discontinuità di percorso fra due istanti di campionamento contigui.

La correzione di queste tre situazioni è importante in quanto semplifica la successiva analisi, senza però eliminare alcuna informazione, infatti in presenza di una discontinuità o di un punto che non trova localizzazione su una strada, non potremmo in ogni caso avere alcuna informazione utile riguardo il moto del veicolo.

Correzione discontinuità temporali Per prima cosa, si procede spezzando la traccia ogni volta che è presente un tempo di campionamento superiore alla soglia massima tollerata, che era stata già definita in precedenza come $i \frac{3}{2}$ della frequenza di campionamento propria della traccia. Questo perché, in tale situazione, ci troviamo a non poter dire niente su cosa è accaduto in quel lasso di tempo, ed esso è

troppo grande per poter trascurare il movimento che può essere avvenuto; inoltre, è probabile che un tempo di campionamento lungo sia dovuto a problemi nel calcolo della posizione da parte del ricevitore.

Questa separazione della traccia però, non può essere fatta in maniera indiscriminata, in quanto è stato scoperto, analizzando diverse tracce, che l'impossibilità del ricevitore di calcolare la posizione attuale non è l'unico motivo alla base di una mancanza di registrazione della posizione. Alcune volte infatti, anche in presenza di ricezione buona, si assiste ad un abbassamento della frequenza di campionamento del GPS quando il veicolo è fermo.

Nei casi più comuni, si passa da un campionamento ogni secondo ad un campionamento ogni due o tre secondi ad esempio, ma in casi più estremi è stato verificato anche che il ricevitore non campiona per 20 o 25 secondi, per poi ricominciare a campionare normalmente non appena il veicolo riprende il suo moto. Questo comportamento non è documentato dai costruttori di ricevitori GPS, ed è in effetti abbastanza raro, in quanto la maggior parte delle volte il GPS campiona normalmente anche in caso sia fermo, però non può essere trascurato. Non considerandolo infatti, si andrebbero a perdere una considerevole percentuale di informazioni su tempi e posizioni di fermata dei veicoli.

Per questo motivo, non si andrà a separare la traccia nel caso che i due punti separati da un tempo di campionamento alto siano:

- Sullo stesso edge a meno di 50 metri di distanza.
- Su due edge collegati, ad una distanza minore della dimensione dell'incrocio.

Questi due casi infatti rispecchiano rispettivamente il caso in cui, dopo che il GPS non ha campionato per un po, riprenda rispettivamente a campionare sulla stessa strada o su una ad essa collegata, ad una distanza tale da non generare una discontinuità spaziale.

Questo processo di separazione andrà a creare un insieme di nuove tracce, prive di discontinuità temporali, partendo da quella originale. Fra queste nuove tracce, quelle troppo corte per essere realmente considerate tracce, come meglio descritto nell'appendice B.3, verranno scartate.

Eliminazione punti non situati su una strada A questo punto, andremo a considerare singolarmente ognuna delle nuove tracce create. In esse, andremo ad elim-

inare tutti i punti che sono privi del match su una strada, che fanno match ma il loro posizionamento non è coerente con il movimento registrato dal veicolo o che sono situati su una freeway.

I punti su una freeway non sono di nostro interesse in quanto questo tipo di strada non presenta regolatori del traffico come semafori o stop, ed inoltre questo tipo di strade presenta spesso ponti, passaggi sopraelevati o sotto passi, che vengono erroneamente identificati come incroci nelle mappe a nostra disposizione, che non hanno l'informazione sull'elevazione delle strade.

Nello specifico, si elimineranno quindi tutti i punti che:

- Non sono posizionati su una strada.
- Sono situati su una freeway.
- Risultano posizionati su una strada, ma il punto successivo e precedente invece no.
- Risultano posizionati su una strada che non è collegata né a quella del punto precedente, né a quella del punto successivo, mentre queste due sono fra loro collegate (una strada è per definizione sempre collegata anche a se stessa).
- Sono su una strada senza nome, ed il punto precedente è stato eliminato.

Per i punti rimasti nella traccia, una volta scandita nella sua interezza, verranno ricalcolate le informazioni che erano state calcolate al momento della lettura iniziale della traccia, ovvero distanza dal punto successivo, azimuth, velocità, ecc. Anche qui, come nel passaggio precedente, se la traccia rimanente è troppo corta, viene scartata.

Correzione discontinuità spaziali L'ultima fase della preparazione della traccia per le analisi successive, è la sua separazione in corrispondenza di discontinuità nel percorso tracciato. Questo tipo di discontinuità, che potrebbe già essere presente all'interno della traccia stessa a causa di errori di ricezione del GPS, è inoltre introdotto con buona probabilità ogni qualvolta di eliminano dalla traccia dei punti.

In linea di principio, separiamo la traccia in due distinte ogni volta che il posizionamento salta da una strada ad un'altra, senza seguire strettamente la topologia delle strade (figura 6.8). Nel dettaglio, si considera una traccia continua fra due punti quando:

- Si passa da una strada ad una ad essa collegata direttamente.

- Si passa da una strada ad una ad essa collegata per mezzo di una strada piccola, di dimensione minore di quella media di un incrocio.



Figura 6.8: Esempio di traccia che presenta una discontinuità spaziale. Tale traccia verrà spezzata in due dall'algoritmo in concomitanza delle due croci gialle

Se la velocità necessaria per effettuare questo spostamento da un punto all'altro è superiore a 40 m/s, o lo spostamento è superiore a 100 metri, la traccia viene comunque spezzata in questo punto, in maniera precauzionale, perché nell'ambiente urbano da noi analizzato valori di questo tipo sono certamente indotti da una cattiva ricezione del segnale satellitare, che generano a volte valori imprevedibili e totalmente fuori dalle unità di grandezza proprie dell'ambiente.

6.4 Analisi delle tracce ed estrazione informazioni caratteristiche

Il fine ultimo di questo lavoro è estrarre la posizione e la temporizzazione di regolatori di traffico come semafori o segnali di stop all'interno di uno scenario urbano.

Per ottenere questo risultato, si devono per prima cosa classificare i comportamenti dei conducenti, in modo da costruire un insieme di astrazioni che possano

essere utilizzate per trovare i regolatori. In particolare, ci si concentra inizialmente su tre comportamenti:

- Svolte effettuate ad ogni incrocio.
- Rallentamenti lungo il percorso.
- Fermate.

6.4.1 Svolte

Abbiamo visto nel paragrafo 6.1 sulla lettura della mappa, come creiamo una rappresentazione della stessa in memoria che permetta di evidenziare bene gli incroci e le strade che li mettono in collegamento. Abbiamo inoltre appena descritto come le tracce vengono ripulite per essere mappate correttamente sulle strade. Possiamo quindi ora agilmente analizzarle e vedere il loro comportamento in prossimità degli incroci, vedendo che svolta effettuano in base alla strada da cui sopraggiungono.

Bisogna qui precisare che per svolta si intende genericamente l'azione di entrare in un incrocio da una strada ed uscirne da un'altra, non per forza l'azione di curvare. Pertanto, una macchina che attraversa un incrocio andando dritto, ha comunque eseguito una svolta, da una strada ad un'altra.

Per analizzare le svolte, vi è in realtà ancora un problema, di cui dobbiamo tenere conto. Infatti è probabile che durante una svolta, la traccia venga a trovarsi non soltanto sulle due strade interessate, ovvero quella di ingresso e quella di uscita dall'incrocio, ma anche su qualche altra strada connessa all'incrocio attraversato. Per capirlo meglio, basta guardare l'esempio di figura 6.9.

Per risolvere questo tipo di problema, basta posticipare il salvataggio dell'informazione di svolta di un incrocio. Ovvero, invece di salvare una svolta quando realmente avviene, la salviamo quando avviene la svolta successiva.

- In ogni istante, teniamo traccia della strada su cui siamo (*Strada A*).
- Quando la strada cambia (*Strada B*), controlliamo se le due strade sono collegate da un incrocio.
- In caso affermativo, salviamo la svolta come provvisoria (*da A verso B*), relativamente a quell'incrocio.
- Al successivo cambio di strada (*Strada C*), controlliamo l'incrocio coinvolto nel cambio da B a C.

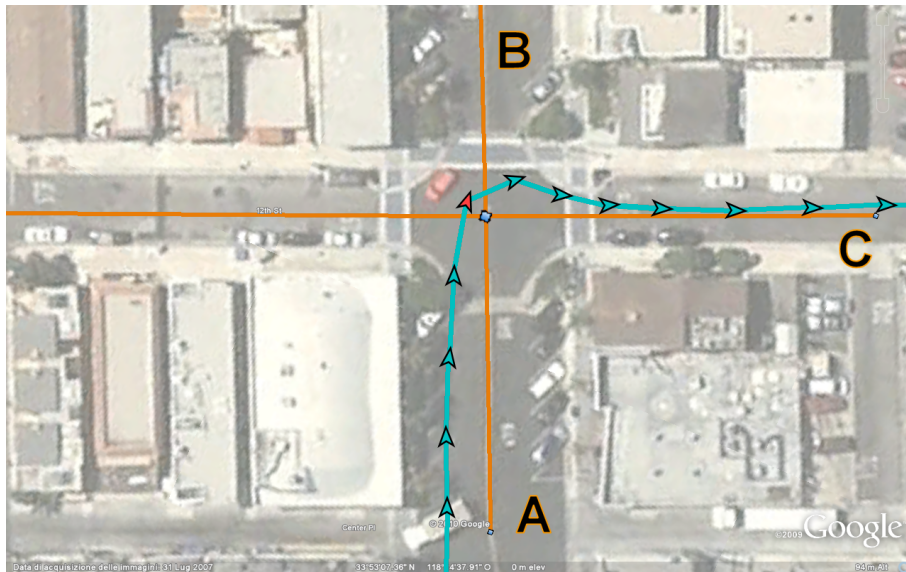


Figura 6.9: Traccia che nell'attraversare un incrocio fa match, col punto in rosso, anche sull'edge di una strada che in realtà non attraversa, quello Nord nell'incrocio

- Se l'incrocio coinvolto è lo stesso di prima, aggiorniamo l'informazione provvisoria associata, che diventa ora (*da A verso C*).
- Se l'incrocio coinvolto invece è differente, l'informazione sul primo incrocio (*da A verso B*) diventa definitiva, ed al nuovo incrocio si associa l'informazione provvisoria (*da B verso C*).

In questo modo, finché il veicolo imbocca strade che escono dallo stesso incrocio, si aggiorna soltanto l'informazione su quella svolta, evitando di salvare informazioni multiple ed errate su quell'incrocio (figura 6.9).

6.4.2 Rallentamenti

Per lo scopo di questo studio, gli unici rallentamenti di nostro interesse sono quelli che avvengono in prossimità di un incrocio. Infatti, nell'ambiente urbano, un veicolo può essere indotto a rallentare da una moltitudine di fattori, ma è soltanto agli incroci che tale comportamento è strettamente dipendente dalla topologia della mappa e dal tipo di regolatore di traffico presente.

Si tratta quindi ora di definire una distanza massima dall'incrocio entro cui il rallentamento deve avvenire, ed una velocità limite sotto la quale si può considerare un veicolo rallentato adeguatamente. Difatti, essendo la velocità di un veicolo

calcolata come la derivata del movimento da un punto di campionamento a quello successivo, essa non è praticamente mai costante, e analizzando le sue fluttuazioni si andrebbero a trovare rallentamenti ed accelerazioni continue.

Per scegliere questi due valori si sono analizzate centinaia di tracce in prossimità di incroci regolati da stop, e si sono svolti anche alcuni esperimenti sul campo per verificare i dati trovati dall'analisi teorica. Si è visto così che la totalità delle macchine che rallentano sopraggiungendo a uno stop lo fa entro 50 metri dallo stesso, e più del 95% di esse rallenta ad una velocità che è inferiore ai 5 m/s. Per questo motivo, sono stati scelti questi due valori. È da notare come questo, come tutti gli studi su un campione, non ci dia la certezza assoluta dei valori scelti, per cui si dovrà sempre considerare una certa tolleranza nel loro uso.

A questo punto, si procede analizzando la traccia nuovamente e considerando ogni punto in cui il veicolo rallenta sotto i 5 m/s, sopraggiungendo ad un incrocio a meno di 50 metri da esso. Oltre all'informazione di svolta quindi, verrà associata all'incrocio l'informazione sul fatto che il veicolo per compierla abbia rallentato o meno (figura 6.10).

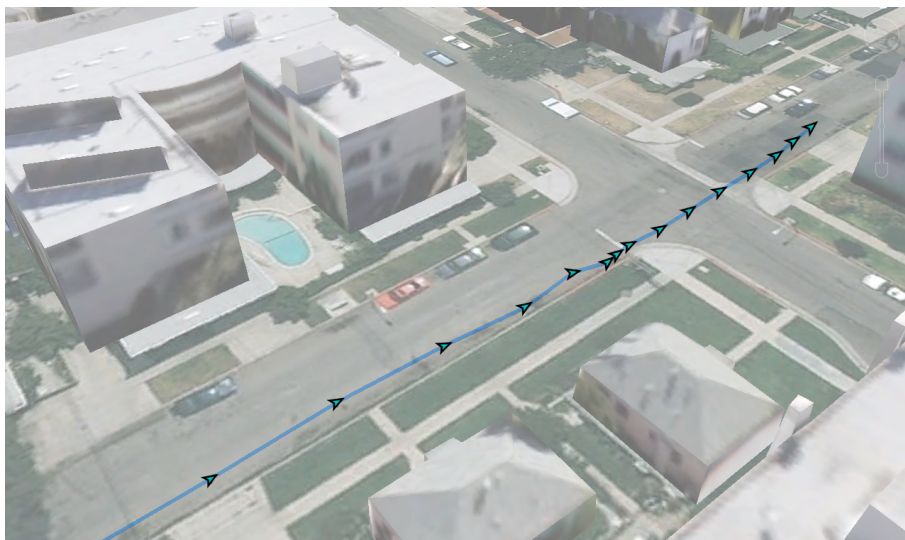


Figura 6.10: Traccia di un veicolo che, sopraggiungendo ad un incrocio regolato da stop, rallenta prima di proseguire dritto

Come si può facilmente immaginare, rallentamenti differenti possono avvenire a seconda che uno debba svoltare a destra, a sinistra o andare dritto, e che debba farlo dando o meno precedenza. Inoltre, nel caso un conducente rallenti più volte sopraggiungendo ad un incrocio, perché ad esempio fermo in coda, e quindi sot-

toposto a continue ripartenze, l'informazione verrà comunque salvata in maniera unitaria, perché quello che interessa è il numero di veicoli che hanno rallentato effettuando una specifica svolta, e non il numero di rallentamenti a cui uno è stato sottoposto, che potrebbe essere invece utile per fare un'analisi sul traffico.

6.4.3 Fermate

Il riconoscimento di situazioni in cui un ricevitore GPS è fermo sembra una delle operazioni più semplici. Nell'immaginario più immediato, ogni qual volta viene registrata una posizione uguale alla precedente, il ricevitore è fermo. Purtroppo invece, gli errori a cui è sottoposto il sistema GPS descritti in precedenza, ed il differente comportamento delle singole strumentazioni in assenza di moto, lo trasformano in un compito molto complesso, come mostrato in figura 6.11.



Figura 6.11: Esempio di spostamenti erroneamente registrati da un ricevitore GPS fermo

Per poter riconoscere quando un ricevitore è fermo, è stata perciò progettata un'euristica, che cerca di superare questi problemi. Questa si scompone in due fasi:

- Identificazione di punti candidati ad essere in una situazione di fermata;
- Analisi accurata di tali punti con una adeguata finestra temporale.

Identificazione Candidati Un punto viene identificato come possibile candidato per essere considerato fermo quando la sua velocità di spostamento verso il punto

successivo, o precedente, è minore di 4 m/s. Questo valore è stato scelto tenendo conto che vogliamo trovare un ricevitore GPS che sta fermo, per cui per fermarsi, deve prima rallentare, portando la sua velocità idealmente a zero.

Si definisce così una sequenza di punti candidati ad essere fermi, che parte da questo punto e contiene tutti i successivi, finché non si ha un punto che ha entrambe le velocità superiori a 4 m/s, che sarà il primo escluso dalla sequenza. Chiamiamo tale sequenza *SEK* per poterla identificare in seguito più facilmente, e procediamo all'analisi di tutti i punti che ne fanno parte.

Analisi con finestra temporale Per decidere in ogni punto se tale punto è fermo o no, non posso basarmi soltanto sulla sua posizione, o sul suo scostamento dal punto a lui subito precedente o successivo. Si è deciso perciò di utilizzare una finestra temporale entro cui valutare lo spostamento del ricevitore.

Preso un generico punto *P* all'interno di *SEK*, si centra su esso una finestra temporale di 10 secondi, e si va a verificare se la distanza fra i due punti estremi di tale finestra è inferiore a 20 metri. In caso affermativo, il punto si considera fermo, in quanto nel suo intorno il ricevitore ha mantenuto una velocità media di non più di 2 m/s per 10 secondi.

Si suppone, per scegliere questi parametri, che il ricevitore abbia una buona ricezione del segnale satellitare, e che l'errore di stima di posizione sia quindi contenuto entro una decina di metri dalla posizione reale. Presi due punti qualsiasi quindi, la loro distanza non è mai superiore a 20 metri per definizione. Non appena un veicolo comincia a muoversi nuovamente, questo viene rilevato dal metodo che quindi lo identifica come in movimento.

Chiaramente, nei punti estremi di *SEK* non è possibile centrare la finestra, che verrà quindi estesa in avanti per i primi punti ed indietro per gli ultimi, invece che essere centrata perfettamente. Inoltre, è possibile che la finestra non possa essere ampia esattamente 10 secondi, perché ad esempio l'intervallo complessivo *SEK* è meno esteso, o la frequenza di campionamento variabile fa sì che non esistano due campioni a 10 secondi di distanza; in questo caso, la distanza tollerata viene adattata sull'estensione temporale considerata, in modo tale che la velocità media rimanga sempre 2 m/s.

L'utilizzo di questa finestra temporale, introduce per costruzione qualche secondo di imprecisione nell'identificazione di una sosta. Infatti, i due o tre secondi

precedenti alla sosta vera e propria, in cui il veicolo sta ancora rallentando, possono venir già identificati come istanti di sosta se il movimento è molto lento. Lo stesso succede nella fase di ripartenza, anche se in maniera meno influente perché solitamente l'accelerazione è più netta. Questa imprecisione concettuale non crea la realtà grossi problemi nella pratica, in quanto la soglie di 4 e 2 m/s rispettivamente fanno sì che l'errore si mantenga nell'ordine di pochi secondi.

Una volta definiti i punti in cui il veicolo è fermo, si può anche calcolare la durata della sosta, semplicemente confrontando il tempo che intercorre fra quando l'utente si è fermato e quando è ripartito.

Tenendo sempre a mente l'obiettivo finale, ovvero individuare le temporizzazioni dei semafori, sono state scartate le soste con una durata superiore ai 300 secondi, in quanto più facilmente riconducibili a soste dipendenti dalle specifiche necessità del conducente che né al regolatore di traffico presente.

Questa informazione sulla fermata e sulla durata della stessa, verrà associata poi alla specifica svolta sullo specifico incrocio, come per il caso del rallentamento descritto nel paragrafo precedente. Nel caso uno stesso utente stia fermo due volte prima di effettuare una svolta, le durate di entrambe le svolte saranno salvate, ma si conterà ugualmente come un solo utente che si è fermato.

Nel considerare le fermate, non si introduce un vincolo di distanza dall'incrocio, come invece si era fatto per il calcolo dei rallentamenti, in quanto in caso di traffico medio / alto la coda ad un semaforo può estendersi facilmente anche a centinaia di metri dall'incrocio interessato. Se fra la posizione di fermata e l'incrocio che ne è responsabile vi è qualche altro incrocio, la fermata verrà erroneamente attribuita a quest'ultimo, e non a quello corretto. Si può considerare però questo un caso isolato, che viene scartato dalle euristiche definite in seguito per l'identificazione dei regolatori di traffico.

6.5 Identificazione di Stop, Semafori e temporizzazioni

Rallentamenti e fermate, così come descritti nel paragrafo 6.4, preparano la strada per un successivo livello di astrazione; essi infatti sono le fondamenta su cui si può

basare l'algoritmo per l'identificazione dei regolatori del traffico presenti in ogni intersezione.

Le due tipologie di regolatori su cui ci concentriamo sono i *semafori* ed i segnali di *stop*. A tale scopo, sono state progettate un insieme di euristiche, derivate direttamente da settimane di osservazioni concrete della dinamica del traffico stradale. Queste euristiche tengono conto degli artefatti che possono affliggere le tracce GPS, e che le stesse sono soggette a imprecisione e tempi di campionamento finiti.

Le due procedure per estrapolare gli stop ed i semafori sono state separate; in particolare, per ogni incrocio prima si controlla se è regolato da stop, ed in caso negativo si verifica se è regolato da semaforo. A causa di dati mancanti o incompleti, è possibile che entrambe le euristiche non restituiscano un risultato chiaro; in tal caso, quella fornita è una *previsione ottima* del regolatore presente, effettuata con i dati a disposizione.

6.5.1 Segnali di Stop

Osservazioni sul campo, così come le leggi sul traffico, suggerirebbero che ogni vettura raggiungendo un incrocio con un segnale di stop, rallenti, si fermi per alcuni secondi e quindi riparta. Questo, indipendentemente dalla condizione del traffico, che al massimo può riflettersi in una sosta più lunga.

Questa assunzione purtroppo, non trova riscontro nelle tracce GPS analizzate. Una prima spiegazione può essere il comportamento reale dei conducenti: molti non si fermano completamente, bensì rallentano solamente in prossimità dell'incrocio. Un secondo motivo, come discusso in precedenza, è che anche quando il veicolo è fermo, l'output del ricevitore GPS può essere fuorviante. Infine poi, la frequenza di campionamento della traccia introduce un limite inferiore alla lunghezza del periodo in cui il veicolo deve rimanere fermo affinché noi lo possiamo rilevare. Ad esempio, un sampling rate di 1 secondo, che è il meglio ottenibile coi dispositivi consumer, permette di individuare tempi di stop non inferiori a 2 secondi.

Per questo, si è deciso di utilizzare le informazioni sui rallentamenti, così come calcolate nel paragrafo 6.4.2, come indicazione della presenza di uno stop. Per ogni incrocio, si analizza singolarmente ogni strada in esso confluyente. Ognuna di esse si marca come *potenzialmente* regolata da stop se l'80% delle tracce entranti

nell'incrocio per quella strada rallentano. L'informazione ottenuta per ogni strada è quindi aggregata per includere tutte le strade dell'incrocio.

In un incrocio regolato da stop, al massimo due strade potranno avere la precedenza, mentre tutte le altre avranno l'obbligo di fermata, indipendentemente dal numero complessivo di strade confluenti. Per questo motivo, se *tutte*, o *tutte tranne una* strada appartenenti ad un incrocio sono marcate come *potenzialmente* regolate da stop, l'incrocio è identificato come *regolato da stop*, e tutte le strade marcate come stop potenziali, diventano ora regolate a tutti gli effetti da un segnale di stop.

Nel caso invece le strade marcate come stop potenziali fossero *tutte tranne due*, l'algoritmo non è in grado di dare una risposta immediata e rimanda la risposta alla verifica della presenza di un semaforo. Questo perché, osservazioni sul campo, hanno mostrato come all'intersezione fra una strada primaria ed una secondaria, spesso il compito del semaforo sia supportato ed influenzato con spire induttive, che fanno scattare il verde per la strada secondaria soltanto in presenza di traffico entrante. Questo, può venir facilmente scambiato per uno stop, infatti il traffico entrante dalle strade secondarie troverà sempre rosso al momento del suo arrivo, se non vi è davanti un veicolo che è già stato rilevato dalle spire, dovendo quindi rallentare (figura 6.12). Per questo motivo, la decisione in questo caso è rimandata.



(a) Stop a 2 vie



(b) Semaforo

Figura 6.12: Esempio di due incroci successivi su Westwood Blvd, il primo (a) con Mississippi Ave regolato da *Stop* sulla strada secondaria, ed il secondo (b) con La Grange Ave invece, regolato da *semaforo*

Nel caso più di due strade non siano marcate come potenziali stop, l'incrocio non è identificato come regolato da segnali di stop e si procede a verificare se sia

presente un semaforo.

Il valore usato per discriminare una strada dotata di stop da una che non lo è, cioè che l'80% delle vetture rallentino, è stato trovato tenendo conto che i parametri di velocità e distanza dall'incrocio introdotti per definire i rallentamenti sono stati scelti grazie ad una stima sul campo, per cui non si può supporre che abbiano una precisione del 100%. È ragionevole inoltre supporre che qualche automobilista non si attenga alla segnaletica stradale.

Test effettuati sulle tracce a disposizione hanno inoltre confermato essere questo il valore che meglio approssima il comportamento degli automobilisti in prossimità di uno stop.

6.5.2 Semafori

Il comportamento di un veicolo nel superare un semaforo dipende, ovviamente, dal fatto che esso sia verde o rosso. Analizzando le tracce in prossimità di un semaforo, come prevedibile, si può osservare che soltanto una parte di esse si ferma. Questa frazione, supponendo di avere una serie infinita di campioni, dipenderà dal rapporto tra il tempo di verde e quello di rosso; informazione quasi impossibile da avere per ogni semaforo. Per questo si decide di usare lo stesso approccio conservativo di prima: si marca una strada come *potenzialmente* regolata da semaforo se almeno il 15% dei veicoli che attraversano l'incrocio arrivando da essa si fermano.

In linea di principio, se un incrocio è regolato da un semaforo, tutte le strade facentene parte dovrebbero soddisfare questo criterio. Ciò nonostante, a causa del limitato numero di tracce solitamente a disposizione per ogni incrocio, della propensione ad avere preferibilmente situazioni in cui il traffico è regolato da qualcosa piuttosto che essere lasciato al suo moto anarchico e della casualità dei comportamenti dei conducenti, si definirà un incrocio come *regolato da semaforo* se la metà + 1 delle sue strade sono *potenzialmente* regolate da semaforo.

Nel caso si fosse arrivati a ricercare un semaforo perché in dubbio con uno stop a due vie (in cui cioè due ways hanno la precedenza e due hanno lo stop) fra strada secondaria e primaria, il caso verrà trattato in maniera particolare. Qui infatti, l'incrocio verrà definito come *regolato da semaforo* soltanto se entrambe le strade non identificate come potenziali stop sono *potenzialmente* regolate da semaforo. In qualsiasi altro caso, viene identificato come uno stop a due vie.

Una volta che l'incrocio è identificato come semaforico, si procede a stimare la durata del rosso su ogni strada entrante. A questo scopo, è utile tenere in mente alcune cose:

- Le vetture possono stare ferme al semaforo un tempo variabile da 0 secondi a tutta la durata del rosso.
- In rari casi, possono stare ferme per periodo più lunghi.
- Nella maggior parte dei casi, in situazioni di traffico regolare, il tempo di fermata sarà inferiore all'effettiva durata del rosso.

Per identificare la durata del rosso si è scelto quindi di prendere il 95° percentile di tutti i tempi di fermata registrati per quella strada. Così facendo, si trascurano i tempi brevi, ma si riescono a scartare anche eventuali valori outsider [41], che vadano oltre la reale durata del rosso (figura 6.13).

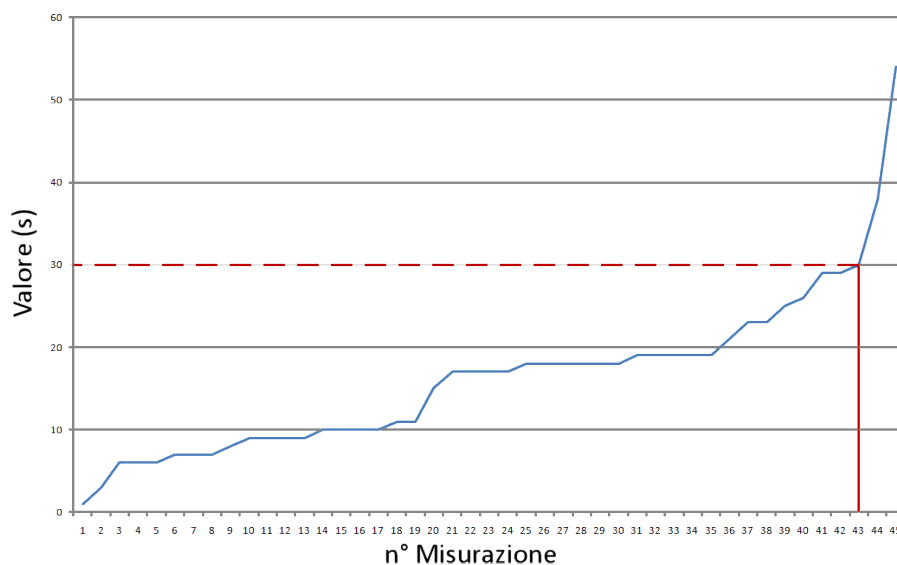


Figura 6.13: Esempio di estrazione del valore corrispondente al 95° percentile in una serie di 45 misurazioni

6.5.3 Gestire la mancanza di informazioni

Se un incrocio non può essere identificato né come regolato da stop, né da semaforo, può essere perché mancano informazioni su qualcuna delle strade che lo costituiscono. Nonostante l'elevato numero di tracce disponibili su database come Open-

StreetMap infatti, nella realtà è abbastanza comune che un incrocio abbia alcune strade prive di tracce. Per questo, si è cercato di trattare anche questi casi al meglio.

Nel caso che una o due strade non abbiano informazioni, si riprocesa l'incrocio alla ricerca del suo regolatore di traffico, questa volta senza considerare le strade senza informazioni, proprio come se non facessero parte dell'incrocio. I risultati di questa seconda elaborazione sono sicuramente un po' forzati, in quanto si devono fare delle supposizioni su dati non disponibili; per questo motivo vengono evidenziati in maniera diversa dai risultati certi, e definiti come *stime* al meglio delle nostre conoscenze. In ogni caso, come verrà mostrato nel capitolo 7 sulla valutazione sperimentale, il risultato è più che soddisfacente anche in questo caso.

La stima in presenza di dati mancanti viene utilizzata anche in un secondo scenario. Come visto, uno stop può venir identificato anche in assenza di informazioni su alcune strade. Potremmo quindi avere un incrocio che *sicuramente* è regolato da stop, di cui alcune strade sono marcate come *sicuramente* dotate di stop, alcune come *sicuramente* prive di stop, ed altre invece prive di informazione. Qui, si applica il criterio di *stima* per decidere se queste strade hanno o meno lo stop, in modo da poter dare un'informazione il più completa e dettagliata possibile.

Studiando la topologia degli incroci tipici delle strade americane, si vede come la maggior parte di essi coinvolga 2 strade che si incrociano a formare, appunto, una croce (+). In questo caso poi, la disposizione degli stop è simmetrica. Ovvero, se ad esempio solo la strada secondaria ha lo stop, esso sarà presente su entrambi i suoi lati, così come se un lato della principale ha la precedenza, lo stesso varrà per l'altro.

Nel particolare caso di un incrocio di questo tipo quindi, si procede propagando l'informazione di stop o precedenza sulle strade che non hanno informazioni, partendo dal loro proseguimento nel caso che esso abbia l'informazione. Per identificare una strada come il proseguo di un'altra, si confrontano i due orientamenti, e se la loro distanza angolare è inferiore a 10° si considerano le strade in tale relazione. Questa propagazione di informazione si rivela corretta in più del 90% dei casi studiati.

È stato fin qui descritto il sistema sviluppato, sia dal punto di vista dei passaggi logici seguiti nella sua realizzazione, sia nella spiegazione dettagliata del funzionamento di ogni sua singola parte. Si procederà quindi ora ad un'analisi del suo funzionamento, valutandone le prestazioni e l'accuratezza complessiva.

Capitolo 7

Valutazione del sistema

In quest'ultimo capitolo, viene presentata una valutazione del sistema sviluppato. Per questa valutazione, vengono utilizzati due differenti insiemi di dati; una prima sezione perciò presenterà le caratteristiche di questi data set, mentre successivamente verranno presentati ed analizzati i risultati sperimentali ottenuti.

7.1 Caratteristiche dei Data Set

Per verificare le performance del sistema sviluppato, lo si è testato su due insiemi di tracce GPS: tracce estratte da OpenStreetMap, ed un insieme di tracce che noi stessi abbiamo raccolto guidando veicoli adeguatamente equipaggiati per diversi giorni nell'area di Los Angeles.

OpenStreetMap OpenStreetMap è un database gratuito di mappe editabili dell'intero pianeta [50]. È costruito con il contributo di utenti da ogni parte del mondo. Concettualmente, gli utenti caricano le tracce GPS da loro registrate e, se vogliono, possono editare le strade che hanno percorso. In seguito, gli amministratori del sistema controllano i risultati dell'editing ed eventualmente validano le modifiche.

Questo approccio garantisce una mappa sempre aggiornata ad un costo virtualmente nullo. Inoltre, tutte le tracce GPS che vengono caricate, sono disponibili alla comunità di utenti per il download, in modo tale che un utente possa contribuire anche editando strade sulla base di percorsi registrati da qualcun altro. Al momento in cui sono stati svolti i test qui presentati sul database di OpenStreetMap erano disponibili 307'000 tracce, ma questo numero è in continua crescita, man mano che nuovi utenti si uniscono alla comunità.

In tabella 7.1 sono riportati i dati sul numero di tracce presenti in alcune regioni in cui la comunità è più attiva, in Germania e Stati Uniti. La Germania è, fra l'altro, lo stato in cui la comunità di utenti OpenStreetMap è più attiva in assoluto, e la sua mappa digitale è quasi interamente creata dagli utenti, nonché molto dettagliata e ricca di informazioni aggiuntive, come punti di interesse e descrizioni di alcune caratteristiche delle strade ad esempio.

Regione	Numero di Tracce
Baviera, Germania	37795
Los Angeles County, California	2584
Marion County, Indiana	1133
Orange County, California	876
Santa Clara County, California	848
Travis County, Texas	845

Tabella 7.1: Distribuzione delle tracce di OpenStreetMap nelle regioni più attive di Germania e USA.

Negli Stati Uniti invece la mappa è derivata dal database TIGER [38], con modifiche e correzioni introdotte dagli utenti. Nonostante il numero di tracce sia alquanto alto, la loro copertura geografica è molto lontana dall'essere completa. Basta infatti mettere mentalmente a confronto il numero di incroci stradali che possono esistere nelle aree considerate, col numero di tracce presenti, per capire come molti di essi non vengano in alcun modo coperti.

La figura 7.1 mostra la distribuzione cumulativa degli incroci in funzione del numero di tracce che li attraversano. Si può osservare come più del 95% degli incroci sia attraversato da meno di 10 tracce, anche nel caso più favorevole (la regione della Baviera). Questi numeri poi peggiorano ulteriormente se si considera come le

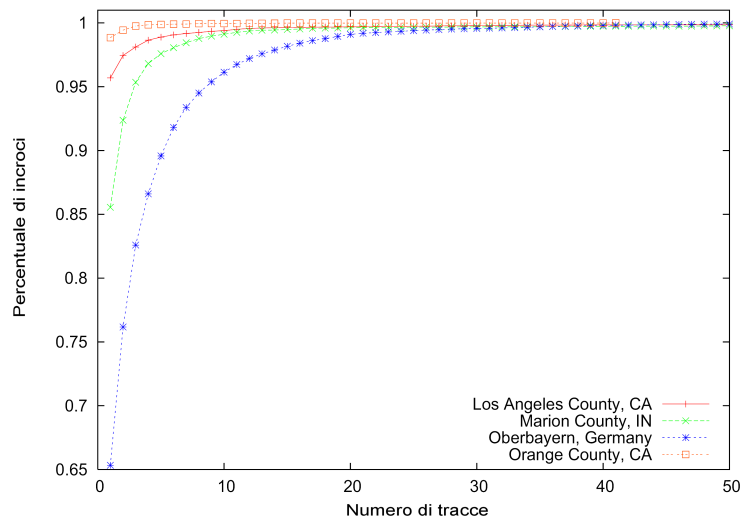
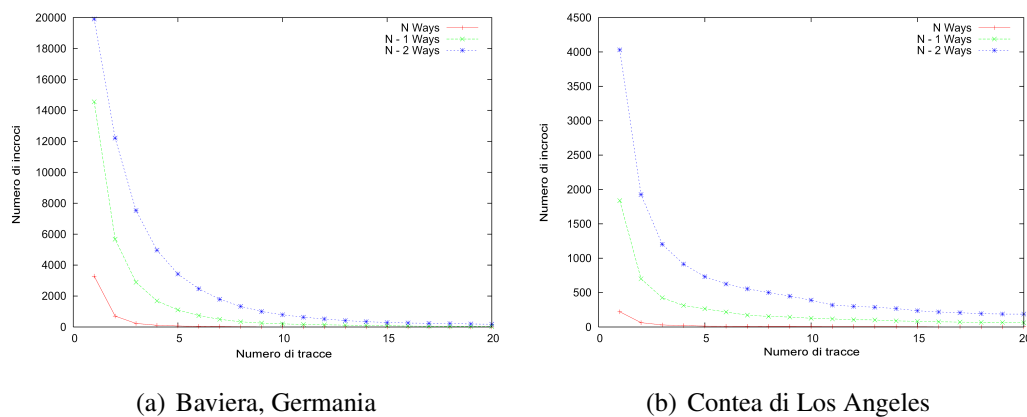


Figura 7.1: Distribuzione cumulativa degli incroci in funzione del numero di tracce che li attraversano

tracce attraversano gli incroci. In figura 7.2 si mostra il numero di incroci che sono attraversati da un numero minimo di tracce su tutte (N) le strade entranti, su tutte tranne una ($N-1$) e tutte tranne due ($N-2$) delle strade entranti per la regione della Baviera, Germania e per la contea di Los Angeles.



(a) Baviera, Germania

(b) Contea di Los Angeles

Figura 7.2: Distribuzione degli incroci in funzione del numero di tracce che li attraversano sulle diverse strade che vi confluiscono

Si può osservare come il numero di incroci che abbiano almeno 5 tracce su ogni strada entrante sia prossima a 0. Inoltre, i dati mostrano che la maggior parte degli incroci siano attraversati nella stessa direzione, che solitamente coinvolge la strada principale. È infatti abbastanza comune avere decine di tracce su un incrocio, senza

però avere alcuna informazione sulle strade secondarie che ne fanno parte. Per questo motivo nel nostro sistema è stata prevista una parte di *stima* del risultato più probabile, per poter affrontare questo problema.

Dataset CALI Come discusso nel paragrafo precedente, usando soltanto le tracce di OpenStreetMap, il numero di incroci attraversati da un adeguato numero di tracce, e che quindi possano essere considerati, è davvero esiguo. Inoltre, su questi pochi incroci, le tracce sono distribuite uniformemente fra strade primarie e secondarie. Per valutare il sistema invece, è necessario avere tracce sicure registrate in prossimità di regolatori di traffico. Per questo, si è deciso di registrare un nuovo insieme di tracce GPS, più completo, su cui valutare qual è l'insieme di informazioni minimo di cui il sistema necessita per fornire risultati affidabili.

Sono state registrate tracce GPS in due differenti aree della California. La prima area consiste di un quadrato di 800x800 metri quadrati, centrato nel punto di coordinate $34^{\circ} 02' 48.15$ N, $118^{\circ} 26' 2.93$ W, ed include 28 incroci, di cui 3 regolati da semaforo e 25 da segnali di stop; questo dataset verrà chiamato *CALI-I*. La seconda area, centrata nel punto di coordinate $34^{\circ} 01' 4.65$ N, $118^{\circ} 29' 35.48$ W, consiste solamente di 4 intersezioni consecutive, tutte regolate da semaforo; questo dataset verrà chiamato *CALI-II*. I percorsi in queste aree sono stati studiati in modo tale che le tracce fossero distribuite uniformemente in tutti gli incroci delle due aree.

Al fine di minimizzare possibili interferenze nell'acquisizione dei dati, è stato chiesto ad alcuni studenti del Network Research Lab, che ringraziamo, di guidare i veicoli equipaggiati con i necessari strumenti, in tre diversi giorni lavorativi dal 26 al 29 ottobre 2009 e nel giorno festivo del 26 novembre 2009. Inoltre, per coprire una finestra temporale adeguata, i veicoli sono stati guidati dalle 7 alle 12.30 e dalle 14 alle 18 in tutti i giorni degli esperimenti.

Il ricevitore GPS utilizzato per acquisire le tracce è l'Holux GR-213, già usato nei precedenti test, e presentato nel dettaglio nell'appendice B.1.

7.2 Risultati Sperimentali

In questa sezione si presentano i risultati della valutazione sperimentale del sistema. Per prima cosa, si valuta la robustezza del sistema alla mancanza di dati; dopo di che, si verificano le prestazioni complessive sul database di OpenStreetMap sia per

il riconoscimento dei regolatori del traffico, sia per la stima delle temporizzazioni dei semafori.

Robustezza alla mancanza di dati Il test sulla robustezza nei confronti di dati mancanti è stato svolto sia sugli incroci del dataset CALI, sia su quelli con un sufficiente numero di tracce del database OpenStreetMap. In entrambi i casi, consideriamo gli incroci che sono attraversati da almeno 10 tracce su ogni strada entrante. Tutti i 32 incroci del dataset CALI soddisfano questo requisito, mentre per la contea di Los Angeles nel suo insieme soltanto 8 incroci lo fanno.

I test sono stati svolti utilizzando informazioni su tutte, tutte tranne una e tutte tranne due delle strade convergenti nell'incrocio. In sostanza, per testare l'incrocio simulando di avere informazioni soltanto su $N-1$ delle strade, viene scelta a caso una strada per cui non si considera alcuna informazione, e fra tutte le tracce disponibili per ognuna delle altre strade, ne vengono selezionate casualmente X per ogni strada. Così si simula il caso specifico di avere X tracce per ogni strada dell'incrocio, ad eccezione di una per cui non si hanno informazioni.

Scelto uno specifico valore di X , e definito il numero di strade da considerare (N , $N-1$ o $N-2$), lo specifico test è stato ripetuto 100'000 volte, ogni volta selezionando a caso le strade da considerare e le X tracce per ognuna di esse.

La figura 7.3 mostra i risultati ottenuti sul dataset CALI. In particolare la figura 7.3(a) mostra la percentuale di risultati corretti analizzando soltanto gli incroci regolati da stop. Il sistema riconosce correttamente il regolatore di traffico in più del 90% dei casi se ha a disposizione informazioni su tutte o tutte tranne una strade dell'incrocio. Se non vengono usate informazioni su due delle strade invece, le performance calano, ma si mantengono ancora sopra l'80% di successo, che è un risultato molto buono considerando che sono utilizzate soltanto metà delle informazioni necessarie (tutti gli incroci considerati infatti coinvolgono 4 strade).

È da notare come le performance siano migliori nel caso di $N-1$ strade considerate, piuttosto che nel caso in cui vengano considerate tutte le strade. Questo è facilmente spiegabile se si pensa che è possibile che in una strada alcune tracce non rallentino adeguatamente in presenza di uno stop. Questo potrebbe portare a considerare, erroneamente, quella strada come una strada con precedenza. Il fatto di non considerarla invece (perché selezionata casualmente come la strada da ignorare), rimuove la fonte d'errore, e contemporaneamente abbassa i requisiti per

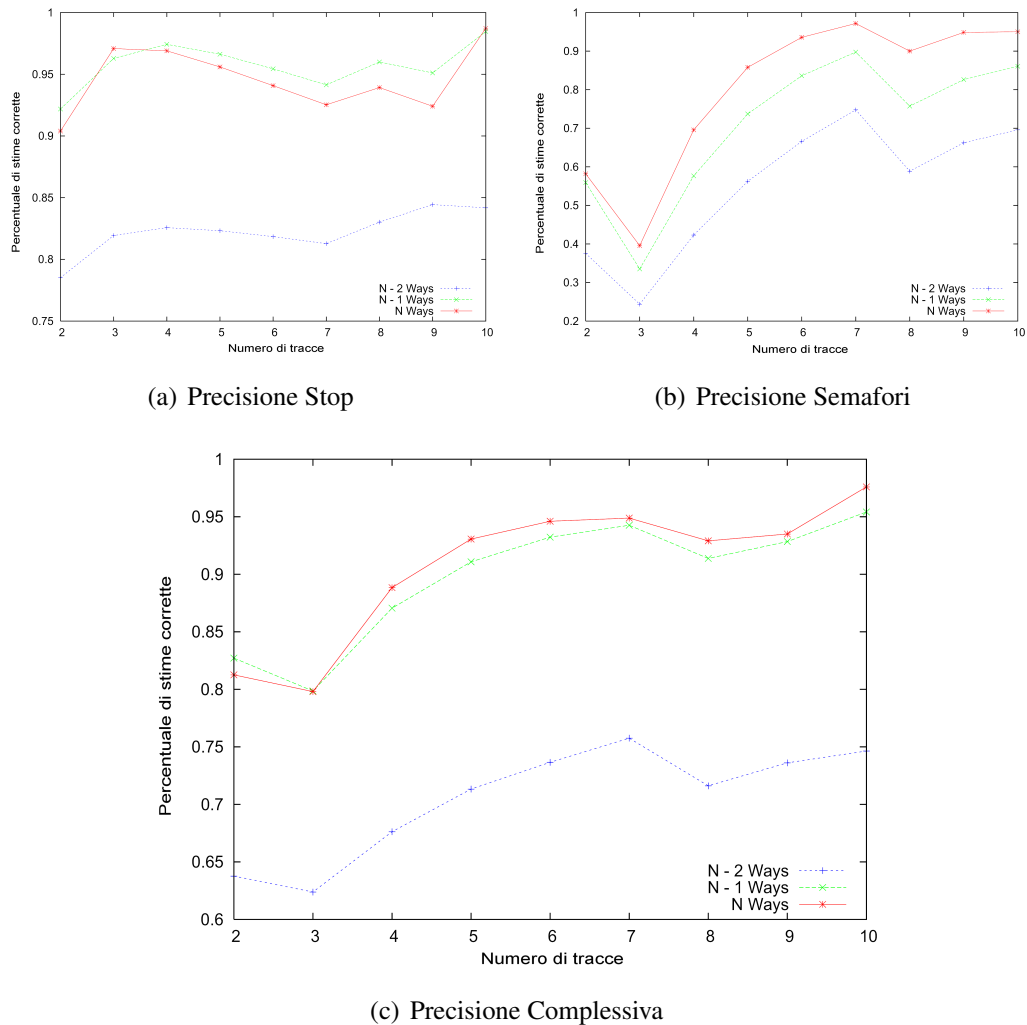


Figura 7.3: Risultati della robustezza del sistema alla mancanza di dati utilizzando il dataset CALI: numero medio di risultati corretti in funzione del numero di tracce utilizzate su N, N-1 ed N-2 strade convergenti nell'incrocio

identificare l'incrocio come regolato da stop (basta una strada con stop sulle tre totali considerate).

La figura 7.3(b) mostra invece la percentuale di successo nel riconoscimento di incroci regolati da semafori. Qui si può notare un calo delle prestazioni in concomitanza di 3 ed 8 tracce disponibili. Va ricordato che il sistema non verifica la presenza di un semaforo se l'incrocio è già stato identificato come regolato da stop, per cui se sovrastima la presenza di stop, sottostimerà quella di semafori. In effetti, sia per 3 che per 8 tracce, si nota un aumento del numero di stop riconosciuti.

La ragione di questo comportamento è comprensibile considerando che la soglia percentuale utilizzata nel valutare se una strada può presentare uno stop o meno deve essere discretizzata nella sua applicazione pratica, diventando quindi leggermente più permissiva nell'individuare gli stop con 3 sole tracce, o 8. La figura 7.4 mostra quali siano i valori realmente usati per le euristiche a causa della discretizzazione, al variare del numero di tracce.

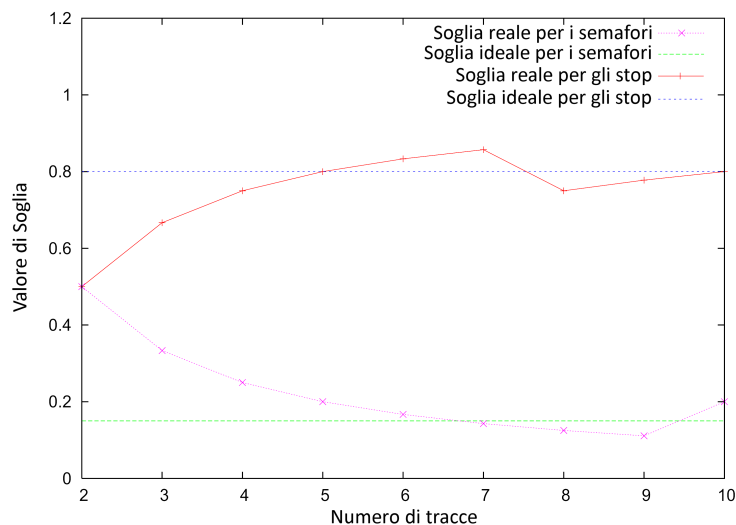


Figura 7.4: Valori realmente usati, a causa della discretizzazione, nelle euristiche per identificare stop e semafori, calcolati al variare del numero di tracce disponibili

Il caso con solo due tracce considerate è riportato sui grafici per completezza, ma non è di per se molto significativo, perché le euristiche si riducono ad un 100% di richieste di rallentamenti per presumere uno stop, o ad un 50% di fermate per inferire un semaforo, e la singola traccia può sbilanciare la scelta in una direzione piuttosto che un'altra.

La figura 7.3(c) mostra il risultato aggregato considerando tutti gli incroci, sia regolati da stop che da semaforo. Il sistema può correttamente valutare il regolatore presente in più del 90% dei casi con soltanto 5 tracce per singola strada.

La figura 7.5 mostra i risultati ottenuti sul dataset di OpenStreetMap con lo stesso insieme di esperimenti. Si può osservare come la tendenza delle prestazioni sia simile. In valore assoluto, esse sono in questo secondo test leggermente inferiori, e la ragione è imputabile ad una qualità delle tracce leggermente inferiore. In

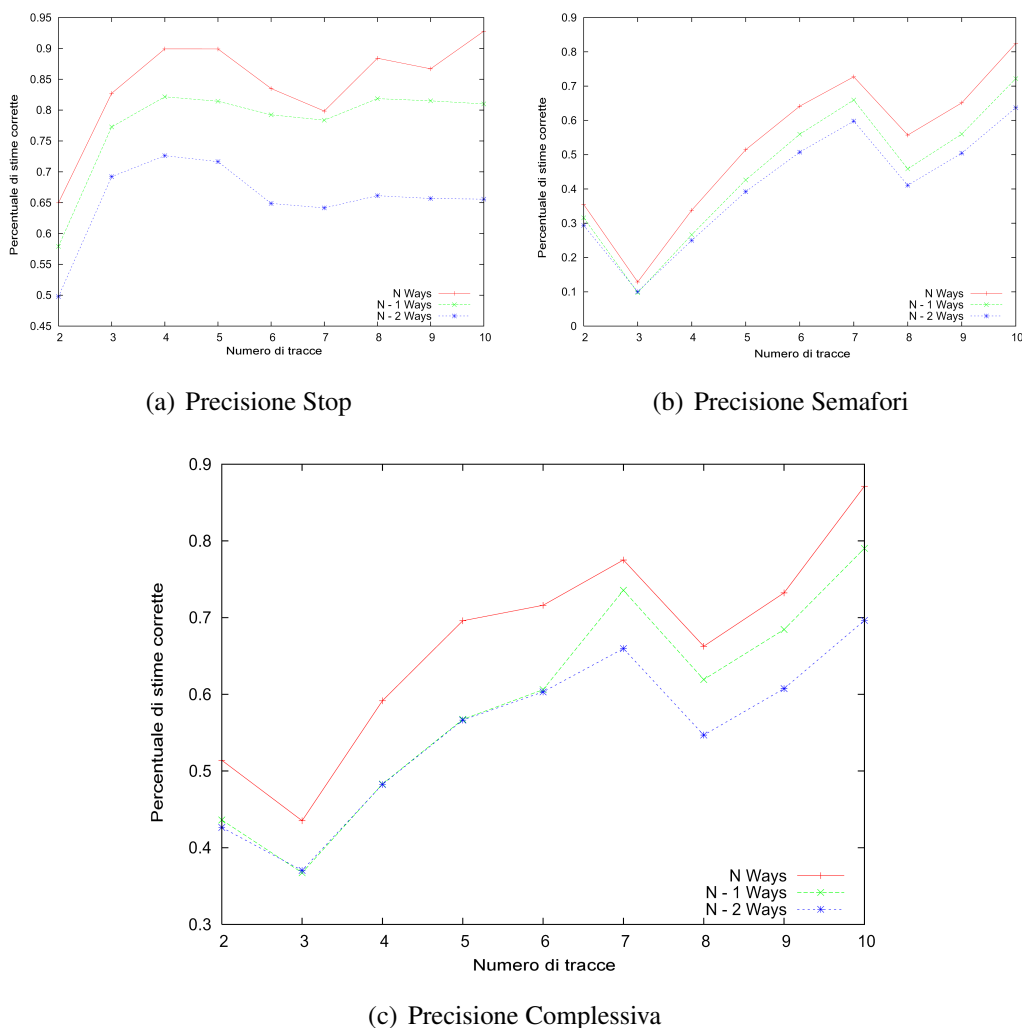


Figura 7.5: Risultati della robustezza del sistema alla mancanza di dati utilizzando il dataset OpenStreetMap: numero medio di risultati corretti in funzione del numero di tracce utilizzate su N, N-1 ed N-2 strade convergenti nell'incrocio

ogni caso, il sistema raggiunge una precisione complessiva del 90% quando sono disponibili almeno 10 tracce per ogni strada entrante nell'incrocio.

È da notare come la qualità delle elaborazioni, che qui è evidenziata dipendere dal numero di tracce disponibili per ogni strada dell'incrocio, dipenda anche dalla qualità delle stesse. Infatti, un piccolo numero di tracce che rispecchino fedelmente l'andamento del traffico su quella strada, possono dare informazioni più precise ed utili che né un alto numero di tracce, che però non rispettano, ad esempio, la segnaletica stradale in maniera rigorosa.

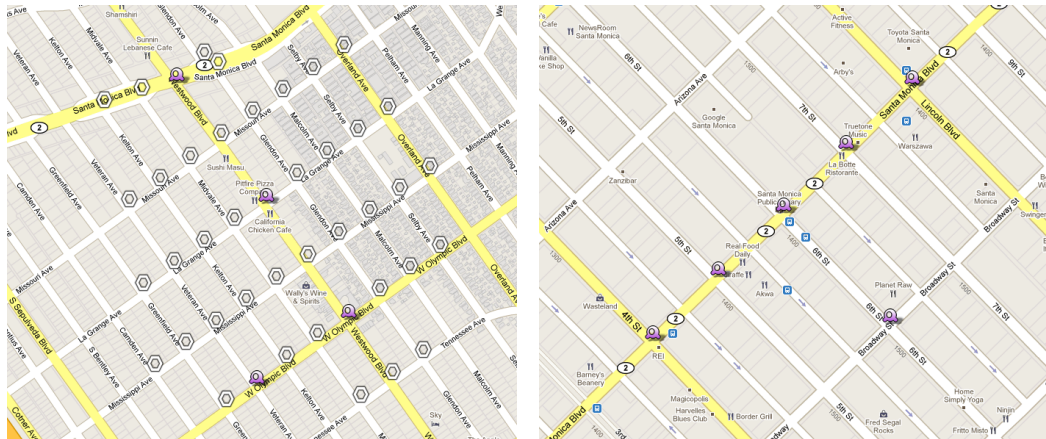
I test effettuati selezionando in maniera casuale le tracce per ogni strada considerata, cercano proprio di ripercorrere questo ragionamento, mostrando come, analizzando lo stesso incrocio, considerando le stesse strade e lo stesso numero di tracce, alcune volte si riesca a identificare correttamente il regolatore di traffico, ed altre no. Supponendo che le tracce che non rispecchino la reale segnaletica stradale, o il comportamento della comunità dei conducenti, siano casi parzialmente isolati, si può ipotizzare che al crescere del numero dei tracce a disposizione, la loro *representazione della realtà* diventi sempre maggiore, confinando il numero di quelle inaffidabili a una frazione sempre più piccola del totale. In quest'ottica, con l'aumentare del numero di tracce disponibili su ogni strada, aumenta la precisione nel calcolo del regolatore di traffico dell'incrocio coinvolto.

Un altro fattore da tenere in considerazione, e che viene attenuato con l'aumentare del numero di tracce, è che alcuni comportamenti degli autisti, pur rispettando la segnaletica stradale, possono essere parzialmente indipendenti da essa: per esempio, una vettura che deve svoltare a sinistra, attraversando in questo modo la carreggiata di marcia opposta alla propria, sarà solitamente obbligata a rallentare o fermarsi nel caso vi sia traffico su questa, anche se la strada percorsa ha precedenza su tutte le altre.

Tutti i fattori qui descritti sono, in uno studio approfondito della dinamica del traffico, marginali e completamente coperti dalle soglie percentuali introdotte con le euristiche sviluppate. Perché questo sia vero però, le tracce devono essere abbastanza per cui un caso isolato non ricopra una percentuale troppo alta dei campioni, come visto.

Valutazione generale delle performance Avendo dimostrato come il sistema sviluppato riesca a tollerare la mancanza di dati, si procede a valutare la sua bontà sulla totalità del database a nostra disposizione. Nella zona relativa al dataset CALI si ottiene un 100% di correttezza nella valutazione; infatti, il sistema riesce a riconoscere correttamente le 25 intersezioni regolate da stop e le 7 regolate da semaforo. Inoltre, il sistema fornisce una *stima* su 17 incroci circostanti l'area che si rivela essere corretta.

La figura 7.6 mostra un'immagine delle due zone considerate, con evidenziati su un livello aggiuntivo i regolatori trovati in ogni incrocio: gli esagoni bianchi rappresentano i segnali di stop, mentre i cerchi viola i semafori.



(a) Westwood Area

(b) Santa Monica Area

Figura 7.6: Regolatori del traffico identificati partendo dal dataset CALI

Come discusso nel paragrafo precedente, nel dataset di OpenStreetMap vi sono davvero pochi incroci che abbiano un adeguato set di tracce abbinato. Per questa ragione, si è stati costretti a scegliere casualmente 65 intersezioni che avessero almeno 5 tracce su almeno tutte le strade entranti tranne una. Si è imposto il limite di almeno 5 tracce e non meno, perchè dalla precedente analisi si è visto come un numero di tracce inferiori non dia risultati abbastanza affidabili. Per questo motivo inoltre, si è scelto di configurare il sistema in maniera tale che quando una strada ha meno di 5 tracce, esse vengano ignorate, considerando la strada come priva di informazioni. Infatti, è meglio ignorare tali tracce piuttosto che considerarle e ritrovarsi con valori che non sono realmente significativi ed affidabili.

Sui 65 incroci testati, il sistema riconosce correttamente il regolatore di traffico presente in 51 casi, con una percentuale di successo del 78%. Questo valore è perfettamente in linea con i dati ottenuti nei test precedenti.

Stima della durata del rosso semaforico Come ultima analisi, vengono presentati i risultati sulla stima dei tempi dei semafori. Nella tabella 7.2 si riporta la durata del rosso semaforico misurata, congiuntamente con la sua deviazione standard e la stima fornita dal sistema proposto.

Il sistema stima la durata del rosso del semaforo su una specifica strada come il 95° percentile dei tempi di fermata delle tracce entranti nell'incrocio per quella

Incrocio (Direzione)	Durata media (σ)	Stima
A (EO / OE)	17.75 (5.2)	24
A (NS / SN)	42.48 (5.98)	51
B (EO / OE)	24.36 (0.44)	29
B (NS / SN)	33.52 (0.29)	36
C (EO / OE)	22.49 (5.26)	31
C (NS / SN)	22.52 (7.48)	27
D (EO / OE)	22.35 (3.7)	29
D (NS / SN)	23.56 (1.64)	28

Tabella 7.2: Durata media del rosso semaforico su 10 misure, confrontata con la stima del sistema. σ è la deviazione standard.

strada [41]. Per costruzione quindi, il valore stimato è un limite superiore al tempo di rosso effettivo. Inoltre, come spiegato nel paragrafo 6.4.3, il metodo utilizzato per riconoscere le situazioni in cui un veicolo è fermo introduce alcuni secondi ulteriori alla misura. Si può notare nella tabella come i tempi siano sempre sovrastimati di circa 5 secondi, in accordo con quanto ci si attende. Questo in realtà non è un problema, in quanto conoscendo il fattore di sovrastima, è molto facile correggerlo introducendo una costante additiva opposta nella fase di calcolo dei tempi di rosso.

Si è visto quindi come il sistema sviluppato riesca ad ottenere ottimi livelli di precisione nella stima dei regolatori di traffico, necessitando per l'elaborazione di un numero di dati ristretto. Infatti, con solo 5 tracce per ogni strada confluyente nell'incrocio si riesce ad ottenere un risultato corretto in più del 95% dei casi. Inoltre, anche nel caso non tutte le strade siano coperte da un numero sufficiente di tracce, la percentuale di successo si mantiene superiore al 90%.

Questi valori sono però strettamente dipendenti dalle euristiche utilizzate per analizzare i comportamenti degli automobilistici, e dalla qualità delle tracce a disposizione. Per questo, si intuisce come una crescente disponibilità di tracce possa sicuramente giovare al sistema. Allo stesso modo, una taratura più aggressiva delle euristiche, magari plasmandole per adattarsi ad uno specifico dataset di tracce, potrebbe portare a percentuali di successo maggiori, a fronte però di un maggior numero di incroci per cui non si può trovare una soluzione.

I risultati sono nel complesso molto promettenti ed incoraggianti, e mostrano come quella dell'analisi aggregata di tracce GPS possa essere una strada valida da seguire per arricchire le mappe odierne con informazioni di alto livello, calcolate in maniera statica (come ad esempio regolatori di traffico presenti agli incroci) o dinamica (situazione del traffico in tempo reale).

Conclusioni

In un mondo come il nostro, in continua evoluzione ed in cui gli utenti sono sempre più in movimento, i dispositivi mobili stanno conquistando sempre più il mercato. Si assiste ad un passaggio dal tradizionale personal computer, ad un più vasto concetto di *ubiquitous computing*, in cui i servizi e la libertà di accesso alla rete si spostano con l'utente.

In questo nuovo scenario, proprio la posizione e la mobilità dell'utente diventano fondamentali per poter fornire servizi su misura che sfruttino queste due informazioni in maniera trasparente, portando valore aggiunto per l'utente. In questo lavoro di tesi si è presentato un semplice ma efficiente metodo per analizzare il movimento degli utenti, registrato tramite ricevitori GPS, e sfruttare le informazioni estratte per restituire loro delle informazioni di alto livello sull'ambiente che li circonda.

Nello specifico, analizzando tracce GPS, si arriva a riconoscere il tipo di regolatore di traffico presente in un incrocio, che esso sia un semaforo piuttosto che un segnale di stop. Nel caso di un semaforo, l'informazione viene poi arricchita con la temporizzazione dello stesso. Il risultato è ottenuto avendo a disposizione soltanto 5 o 7 tracce per ogni strada confluyente nell'incrocio, a seconda che si debba identificare rispettivamente uno stop od un semaforo; i test effettuati dimostrano inoltre che il sistema risponde correttamente anche in presenza di dati mancanti, con una precisione di oltre l'85% nel caso che una delle strade sia priva di informazione.

Si è visto come la qualità delle tracce GPS disponibili sia la chiave per poter estrarre informazioni dalla loro analisi. In particolare, un'alta frequenza di campionamento ed un numero di errori di campionamento casuali ridotti al minimo sono cruciali per poter estrarre informazioni utili. Per questo motivo, una prima parte del lavoro si è concentrata sulla selezione e correzione delle tracce da analizzare, mentre una seconda si è dedicata alla loro analisi vera e propria.

Il lavoro svolto, per quanto si focalizzi sull'estrazione di informazioni sui regolatori di traffico stradali, vuol essere un esempio delle possibilità che si celano dietro l'analisi di tracce GPS. Quello considerato infatti, è soltanto un esempio di come si possono usare le tracce registrate dagli utenti nei loro spostamenti per ottenere livelli di informazione aggiuntivi. Si può infatti pensare di applicare lo stesso procedimento per calcolare probabilità di svolta agli incroci, andamento del traffico in diverse ore del giorno, dislocazione di fermate degli autobus o attraversamenti pedonali, ecc., arricchendo così le mappe a disposizione degli utenti e la loro esperienza nel loro utilizzo.

Il contributo dell'utente è quindi fondamentale, la forza del sistema dipende dalla collaborazione di tutti. Avendo una comunità di utenti attiva, che mette costantemente a disposizione nuovi dati, le possibilità diventano infinite, a tutto vantaggio della qualità e della varietà dei servizi di cui gli utenti stessi potranno fruire.

Appendice A

Pre-processing delle mappe

In questa sezione presenteremo in maniera dettagliata la fase di pre-processing a cui sono sottoposte le mappe utilizzate nel nostro sistema per passare dalla rappresentazione dei due database utilizzati, TIGER e OpenStreetMap, a quella astratta di strade ed incroci usata dal sistema.

A.1 Il Database TIGER Census

Il TIGER Census Data Base è una raccolta di tutte le caratteristiche geografiche e demografiche degli Stati Uniti d'America, disponibile pubblicamente. L'ultima versione disponibile, aggiornata al 2009, fornisce i dati sotto forma di *Shape files* [42], organizzati gerarchicamente per stato e contea. Le informazioni contenute sono di vario tipo e molto estese, ma fra esse sono presenti anche le informazioni sulla topologia stradale, rappresentate da un insieme di nodi interconnessi da linee spezzate (*poly lines*). Come già spiegato, questa rappresentazione deve essere elaborata prima di essere utilizzata nel sistema, definendo strade (*ways*) ed *incroci* invece di segmenti e nodi.

Per prima cosa, dallo shapefile della contea si estraggono tutte le *poly lines* corrispondenti alle strade, scartando quindi quelle relative a fiumi, acquedotti o simile,

che nel database vengono salvate insieme. Per far questo si utilizza semplicemente il flag che associa alla poly line la sua tipologia. A questo punto si crea in memoria la rappresentazione di edge e node descritta nel paragrafo 6.3.1, che permette di effettuare tutte le operazioni di reverse geocoding. Da questa si procede per creare la mappa degli incroci e delle strade che vi confluiscono, considerando però soltanto la lista di nodi originali e di edge, non i nodi intermedi o gli edge parziali. Questo perché nel database TIGER, gli edge possono incontrarsi soltanto nei loro nodi terminali, e mai in un nodo intermedio.

I passaggi logici seguiti per arrivare ad una mappa di incroci e strade sono:

- Si trovano tutti i nodi che sono parte genericamente di un incrocio;
- Si trovano tutti i nodi che sono parte dello stesso incrocio e lo si definisce;
- Si trovano tutti gli edge che fanno parte dell'incrocio appena definito;
- Si trovano tutte le strade differenti identificate da questi edge, e le si definiscono agglomerando in esse gli edge che le formano.

Vediamo ora nel dettaglio come queste operazioni sono svolte. Per trovare l'insieme dei nodi che possono essere parte di un incrocio, si scartano tutti i nodi che non sono condivisi da tre o più edge differenti. Infatti, si considera che un incrocio debba mettere in comunicazione almeno tre strade per poter essere considerato tale, e conseguentemente anche un nodo che ne fa parte deve fare lo stesso. A questo punto, si parte considerando il primo nodo del maxi insieme appena creato. Si associa ad esso un incrocio, di cui lui è il primo nodo. A questo punto, inizia una fase ricorsiva in cui si vanno ad aggiungere via via all'incrocio gli altri nodi che ne fanno parte, nonché gli edges che compongono le strade che vi confluiscono.

Questa fase ricorsiva è così strutturata: vi è un insieme di nodi che devono essere controllati, che inizialmente contiene solo il primo nodo che ha iniziato la procedura di creazione dell'incrocio. Per ogni nodo dell'insieme, si analizzano tutti gli edges che risultano a lui connessi. Se l'edge è più lungo di d , lo si aggiunge alla lista di edge che compongono le strade confluenti nell'incrocio. In caso contrario, si aggiunge il nodo all'altro capo dell'edge all'insieme di quelli che devono essere controllati, e a quelli facenti parte dell'incrocio. La dimensione d deve essere scelta considerando lo scenario rappresentato dalla mappa, infatti riflette la dimensione media di un incrocio. Nello scenario urbano di Los Angeles ad esempio, è scelto

un valore di 25 metri, mentre a Monaco di Baviera il valore più adatto è inferiore, attorno ai 15 metri.

In questo modo si annettono all'incrocio tutti i nodi abbastanza vicini fra loro da essere considerati parte dello stesso incrocio, e tutti gli edge abbastanza lunghi da poter essere considerati una strada uscente, e non soltanto un segmento di congiunzione fra due nodi dell'incrocio. Tenendo traccia degli edge e dei nodi già considerati, si evita di considerare lo stesso elemento più volte, rischiando di generare loop infiniti.

A questo punto si calcola la posizione in cui localizzare l'incrocio, come il punto medio delle coordinate dei nodi che ne fanno parte, e si procede con la generazione delle strade che lo compongono partendo dagli edges a lui associati in precedenza.

Per fare questo si calcola l'azimuth del primo *edge part* connesso all'incrocio per ognuno degli edge, e successivamente si considerano come componenti la stessa strada tutti gli edges per cui l'azimuth così calcolato è uguale, entro 5° di tolleranza. In questo caso, l'azimuth è calcolato modulo 360 in quanto il verso della strada è importante, perché ci dice da che parte dell'incrocio si trova. Dopo di che, l'azimuth della strada stessa è calcolato come l'azimuth medio fra quelli appena trovati.

Come nome della strada invece si usa il nome associato al primo suo edge che abbia nome non nullo, e si aggiunge l'informazione sull'orientamento (E, O, S o N) ad indicare sommariamente dove si trova la strada rispetto all'incrocio considerato; questo è fatto in modo tale da poter ad esempio distinguere chiaramente il caso di una vettura che attraversa l'incrocio in direzione E - O sulla strada A che continua da un lato all'altro, dal caso in cui lo faccia con direzione O - E sulla stessa strada.

Per come è fatto il database TIGER utilizzato infatti, è possibile che due edge paralleli che rappresentano le due carreggiate della stessa strada siano identificati l'uno da un nome di strada e l'altro dall'assenza del nome della strada. È inoltre possibile che entrambi non abbiano nome, nel qual caso non si considera quella strada in quanto sarà una strada privata o di passaggio pedonale solitamente.

Abbiamo a questo punto identificato un incrocio e le strade che vi confluiscono. È ancora possibile a questo punto che l'incrocio non metta in relazione tre o più strade distinte anche se prima i suoi nodi collegavano tre differenti edges, perciò l'ultima fase consiste nel tenere soltanto gli incroci corretti, per i quali si tengono in memoria apposite strutture dati che mettono in relazione bidirezionale incroci con nodi e strade con edges.

La figura A.1 mostra in maniera schematica il risultato dell’elaborazione della mappa letta da tiger.



Figura A.1: Esempio di come è rappresentato in TIGER un incrocio come quello fra Slauson Ave e Sepulveda Blvd, a Los Angeles (a), con 5 nodi (evidenziati dalle spunte blu), 5 edges di collegamento fra essi e ben 7 edges uscenti, rappresentanti le carreggiate delle strade. In (b) è invece mostrato come lo stesso incrocio è salvato in memoria, in modo da rispecchiare la sua struttura logica, in cui vi sono solo 4 strade confluenti ed un unico incrocio

A.2 Il Database OpenStreetMap

OpenStreetMap, prima di essere un database di tracce GPS registrate da una moltitudine di utenti, è un database contenente una mappa gratuita ed editabile dell’intero pianeta. Essa è costruita con il contributo degli utenti di tutto il mondo. Anche questa mappa è disponibile tramite degli *shape files* [42] con formato standard.

Questi shape files però contengono informazioni caratteristiche diverse da quelle contenute nel database TIGER, quali ad esempio limiti di velocità e classificazioni delle strade, che vanno adeguatamente gestite. Inoltre, le poly lines che identificano le strade in OpenStreetMap sono definite per singola strada, e non da incrocio ad incrocio. Conseguentemente, si ha che gli incroci non sono situati soltanto sui nodi vertici degli edges, ma possono trovarsi anche in qualsiasi suo nodo intermedio.

Questo complica la creazione dell’astrazione della mappa utilizzata, perché bisogna controllare se è presente un incrocio anche fra tutti i punti intermedi degli edges. Inoltre, non si potrà più considerare il numero di edge che condividono un nodo, ma bisognerà considerare gli edge parts che vi incontrano, in quanto diven-

tano questi ora le unità che generano un incrocio. Allo stesso modo, per considerare le unità da aggregare a formare una singola strada, non ci si potrà concentrare su interi edges, ma si dovranno considerare soltanto loro sezioni, formate dall'insieme di edge parts che uniscono due incroci. Anche la dimensione media di incrocio utilizzata nei calcoli dovrà essere adattata, e nello specifico si è visto come il valore più corretto sia fra i 12 ed i 15 metri.

Le accortezze da usare per poter adattare il sistema alle mappe ottenute da OpenStreetMap non si esauriscono qui, in quanto anche alcune altre fasi come ad esempio il *reverse geocoding* devono essere adattate alla struttura diversa della mappa, ma comunque tutte le operazioni svolte dal sistema rimangono uguali dal punto di vista logico.

Appendice B

Standard dei Dati GPS

In questa sezione presenteremo per prima cosa il ricevitore GPS utilizzato per tutti i test da noi effettuati, descritti nel paragrafo 4.3, e per registrare tutte le tracce dei dataset CALI-I e CALI-II presentati nel paragrafo 7.1. Dopo di che, verrà presentato lo standard NMEA, utilizzato nello scambio di messaggi fra il ricevitore GPS ed il computer.

A seguire si presenta invece lo standard GPX, che è un formato XML appositamente creato per lo scambio di dati GPS tra applicazioni e servizi web differenti. Tale formato è utilizzato fra le altre cose per salvare le tracce GPS analizzate dal sistema proposto.

L'ultima parte invece approfondisce brevemente i due sistemi di coordinate più diffusi oltre a quello di coordinate geografiche lat - lon, che vengono usati per la maggior parte dei calcoli nel sistema sviluppato.

B.1 Ricevitore GPS Holux GR-213 SiRF Star III

Il ricevitore GPS Holux GR-213 è basato sul chipset SiRF Star III [70], che è appositamente studiato per l'utilizzo in ambiente urbano o boschivo fitto, grazie alla sua rapidità nell'acquisire la posizione e nel mantenerla anche in presenza di osta-

coli. Tale chipset è praticamente lo standard di mercato indiscusso, soltanto recentemente alcuni produttori di sistemi embedded hanno cominciato ad integrare chipset differenti.

Il ricevitore, mostrato in figura B.1, si connette al notebook tramite interfaccia USB, su cui trasmette dati in formato NMEA v2.2. Altre caratteristiche fondamentali di questo dispositivo sono la possibilità di tracciare fino a 20 satelliti, di avere una frequenza di campionamento massima di 1 Hz e di essere mono frequenza, ricevendo il segnale L1 in C/A code.



Figura B.1: Ricevitore GPS Holux GR-213, usato nella maggior parte dei test svolti

Per quanto riguarda invece le caratteristiche prestazionali fornite dal costruttore, è dichiarata una precisione nella localizzazione fra i 5 ed i 25 metri in movimento, ed una capacità di rilevare velocità di spostamento da 0.1 a 515 m/s [55].

B.2 NMEA

NMEA è uno standard di interfacciamento tra apparecchiature digitali, il cui nome deriva dallente che lo gestisce, la *National Marine Electronics Association* [77], ed ha origini e finalità prevalentemente nautiche. Nell'ambito del GPS il sistema NMEA viene impiegato per trasmettere dati da un ricevitore GPS verso un computer, in quanto il formato è efficiente e facilmente decodificabile dal software di un elaboratore.

Lo standard contempla i segnali elettrici, il protocollo e la temporizzazione della trasmissione e il formato dei dati, detti *sentences* ovvero *frasi*, operanti su un bus

dati seriale (tipicamente RS232) con baud rate pari a 4800 bps. Ogni bus dati può avere un unico trasmittente (*talker*) e molti riceventi (*listeners*).

Struttura delle frasi NMEA Tutte le frasi NMEA sono costituite da una sequenza di caratteri ASCII e strutturate in un prefisso, una serie di campi ed un checksum finale. L'inizio della frase è delimitato dal carattere \$, i singoli campi sono separati da virgole ed il checksum finale è direttamente successivo all'ultimo campo, senza virgola a delimitarlo. Una frase NMEA può contenere fino a 82 caratteri compreso il carattere \$. Se i dati per un campo non sono disponibili, il campo viene omissso, ma le virgole che lo delimiterebbero sono comunque presenti, senza spazi tra di loro [68].

Il prefisso Dopo il carattere \$ seguono due caratteri che identificano l'entità che genera i dati, ovvero il *talker*. Nell'ambito del GPS quindi, tutte le frasi iniziano con \$GP, dove GP denota che la frase è generata da un dispositivo GPS. I successivi tre caratteri del prefisso della frase indicano il tipo di frase; per esempio una frase che inizia con il prefisso \$GPGGA è inviata da un dispositivo di tipo GPS ed è del tipo *Global Positioning System Fix Data* (GGA).

Il checksum Per poter individuare eventuali errori sulla linea di trasmissione tra talker e listener, al termine di ciascuna frase NMEA viene posto un *checksum* (dopo un asterisco), che è lo XOR a 8-bit di tutti i caratteri della frase NMEA, comprese le virgole ed esclusi i delimitatori \$ e *. Il valore esadecimale dei 4 bit più alti e più bassi del risultato vengono convertiti in due caratteri ASCII, di cui viene trasmesso per primo il più significativo.

Le frasi utilizzate Il gruppo GPS è costituito da una sessantina di frasi, di cui le più usate in campo GPS sono:

- \$GPRMC - Recommended Minimum specific GPS/TRANSIT data
- \$GPRMB - Recommended Minimum navigation info
- \$GPGGA - Global Positioning System fix data
- \$GPGLL - Geographic Position - Latitude/Longitude
- \$GPGSV - Satellites in View

Nel nostro caso, le due frasi utilizzate per conoscere la posizione attuale dei nodi mobili sono quelle con prefisso \$GPGGA o \$GPRMC. Le informazioni contenute nelle due frasi sono simili, e precisamente entrambe contengono la posizione e l'orario a cui essa è stata registrata. Soltanto la seconda però contiene anche l'informazione del giorno in cui il campionamento è stato attuato, per questo motivo dove disponibile si preferisce usare la seconda frase piuttosto che la prima. Vediamo ora più nel dettaglio come sono strutturate le due frasi.

Una frase \$GPGGA di esempio è:

*\$GPGGA,173519.94,4857.038,N,01131.169,E,1,08,0.9,545.4,M,46.9,M,,*47*

ed in tabella B.1 è spiegato il significato di ogni suo campo.

Valore	Significato
GGA	Global Positioning System Fix Data
123519	Posizionamento ottenuto alle 12:35:19 UTC
4807.038,N	Latitudine 48 gradi 07.038' N
01131.000,E	Longitudine 11 gradi 31.000' E
1	Qualità rilevamento (0 - invalido, 1 - GPS, 2 - DGPS, 3 - PPS)
08	Numero di satelliti visti
0.9	HDOP (Horizontal Dilution Of Position)
545.4,M	Altitudine del ricevitore in metri
46.9,M	Altezza del geoide rispetto all'ellissoide WGS84
(campo vuoto)	Tempo in secondi dall'ultimo aggiornamento DGPS
(campo vuoto)	Identificatore della stazione DGPS (0000 - 1023)
*47	Checksum

Tabella B.1: Significato dei parametri in una frase GGA dello standard NMEA

Una esempio di frase \$GPRMC invece è:

*\$GPRMC,123519,A,4807.038,N,01131.000,E,022.4,084.4,230394,003.1,W*6A*

ed in tabella B.2 è spiegato il significato di ogni suo campo.

Valore	Significato
RMC	Recommended Minimum sentence C
123519	Posizionamento ottenuto alle 12:35:19 UTC
A	Stato A = Attivo o V = Nullo
4807.038,N	Latitudine 48 gradi 07.038' N
01131.000,E	Longitudine 11 gradi 31.000' E
022.4	Velocità rispetto a terra misurata in Knots
084.4	Direzione angolare in gradi
230394	Data - 23 Marzo 1994
003.1,W	Declinazione Magnetica
*47	Checksum

Tabella B.2: Significato dei parametri in una frase RMC dello standard NMEA

B.3 GPX

GPX o *GPS eXchange Format* è uno schema XML progettato per il trasferimento di dati GPS tra applicazioni software [53]. Può essere usato per descrivere punti, tracciati e strade. Il formato è aperto e può essere usato senza il bisogno di pagare per la licenza. I campi al suo interno memorizzano posizione, elevazione e tempo, in aggiunta ad un insieme di altre informazioni accessorie.

Essendo uno schema XML, vi è una struttura con tag annidati in maniera gerarchica, in cui il padre di tutti è il tag `<gpx>` appunto [54]. Esso può avere alcuni parametri, come la versione e l'autore, un tag *metadata* con informazioni aggiuntive, e poi uno o più tag che introducono il tipo di dati presente: *waypoint*, *route* o *track*, come si può vedere in questo schema:

```
<gpx
  version=1.1 [1] ?
  creator=xsd:string [1] ?>
  <metadata> metadataType </metadata> [0..1] ?
  <wpt> wptType </wpt> [0..*] ?
  <rte> rteType </rte> [0..*] ?
  <trk> trkType </trk> [0..*] ?
```

```
<extensions> extensionsType </extensions> [0..1] ?
</gpx>
```

Uno stesso file può contenere tutte e tre le tipologie di dati, anche più di uno per tipologia. È quindi possibile più tracce nello stesso file, precedute magari da una lista di punti di interesse. Per questo, nel leggere il file .gpx, si dovrà andare ad individuare quali sono le informazioni d'interesse, scartando le altre. Si considerano nello specifico soltanto i tag *<trk>*, che indicano una traccia. Questi hanno diversi tag al loro interno, che identificano una moltitudine di possibili caratteristiche ed informazioni aggiuntive, ma solitamente i più comuni sono *<name>*, che identifica il nome dato alla traccia, e *<number>* che identifica il numero della stessa.

Dopo di questi parametri, compare il tag più importante, *<trkseg>*, che introduce il raccoglimento di tutti i punti che formano la traccia. È possibile che vi siano più tag *<trkseg>* all'interno della stessa traccia, nel caso ad esempio che si sia interrotta la registrazione ad un certo punto per poi riprenderla dopo.

Il tag *<trkseg>* contiene quindi una moltitudine di tag di tipo *<trkpt>*, uno per ogni punto della traccia. Questi punti possono avere al loro interno più di 15 attributi differenti, che però non sono quasi mai presenti nella realtà e non vengono perciò considerati dal nostro sistema. Gli unici che devono per forza essere presenti, pena l'inutilità del punto ai nostri scopi, sono le coordinate di latitudine e longitudine, nonché l'istante di campionamento.

Un esempio di come si presenta un blocco traccia all'interno del tag padre *gpx* è il seguente:

```
<trk>
  <name>ACTIVE LOG 2</name>
  <number>1</number>
  <trkseg>
    <trkpt lat=40.423297221 lon=-79.919558549>
      <ele>337.560059</ele>
      <time>2007-09-05T11:57:07Z</time>
    </trkpt>
    <trkpt lat=40.423261095 lon=-79.919599704>
      <ele>334.195435</ele>
      <time>2007-09-05T11:57:30Z</time>
    </trkpt>
```



```
<trkpt lat=40.423260089 lon=-79.919600124>  
  <ele>334.195435</ele>  
  <time>2007-09-05T11:57:31Z</time>  
</trkpt>  
</trkseg>  
</trk>
```

Non vi è però purtroppo alcuna garanzia sulla qualità delle tracce memorizzate in un file GPX. Difatti, è possibile trovare punti con coordinate $lat = 0.0$ e $lon = 0.0$, sinonimo che si è cominciato a generare il file prima che il ricevitore GPS avesse identificato la posizione. Oppure, è possibile trovare punti doppi, o ancora punti con lo stesso istante di campionamento ma diversa posizione.

Anche l'informazione stessa sulla data poi può essere presenta in diversi formati, presentando ad esempio un valore che si ferma alla sensibilità di un secondo, piuttosto che spingersi fino al millisecondo.

La traccia stessa poi, nel suo complesso, può essere formata da una manciata di punti o da migliaia. Capita facilmente che lo stesso file contenga un paio di tracce iniziali vuote, o con un paio di punti, e poi una lunga realmente utile. Chiaramente, tracce di pochi punti devono essere scartate, in quanto inutili per rappresentare un movimento.

Per capirlo, basta pensare ad un esempio semplice: con un ricevitore che campiona ogni secondo, e si muove ad una velocità di 50 km/h, una traccia di 30 punti corrisponde ad uno spostamento di appena 400 metri. Riducendo la velocità a 35 km/h, ragionevoli in ambiente urbano se si incontra almeno una svolta, lo spostamento corrispondente non raggiunge i 300 metri.

Non va dimenticato poi il tempo medio necessario ad un ricevitore GPS per acquisire la posizione dopo esser stato spento per un po': dai 30 ai 45 secondi. Per questo motivo, tracce di lunghezza inferiore ad almeno 30 punti non vengono considerate, in quanto la probabilità di errore è superiore a quella di trovare informazioni utili, in quanto in un tragitto così breve sarebbe pressoché impossibile individuare comportamenti caratteristici del conducente da cui calcolare caratteristiche della topologia stradale.

B.4 Sistemi di Coordinate

Vediamo brevemente cosa sono i due sistemi di coordinate più diffusi per identificare i punti della superficie terrestre, oltre a quello comune di *coordinate geografiche* (latitudine e longitudine).

B.4.1 Universal Transverse Mercator System

È una proiezione, derivata dalla proiezione di Mercatore, della superficie terrestre su un piano, una delle soluzioni meglio riuscite al problema di rappresentare la superficie terrestre a due raggi di curvatura. La proiezione *Universal Transverse Mercator System* (UTM) si utilizza dal parallelo di 80° sud a quello di 80° nord, mentre per i poli viene utilizzata la proiezione UPS (*Universal Polar Stereographic*). Perché la proiezione di mantenga precisa agli estremi così come nel punto di tangenza fra il piano e la superficie terrestre, è necessario suddividerla in sezioni di dimensione contenuta, per questo la Terra viene divisa in 60 fusi di 6° di longitudine ciascuno, a partire dall'antimeridiano di Greenwich in direzione Est, ed in fasce di ampiezza di 8° di latitudine (figura B.2) [74].

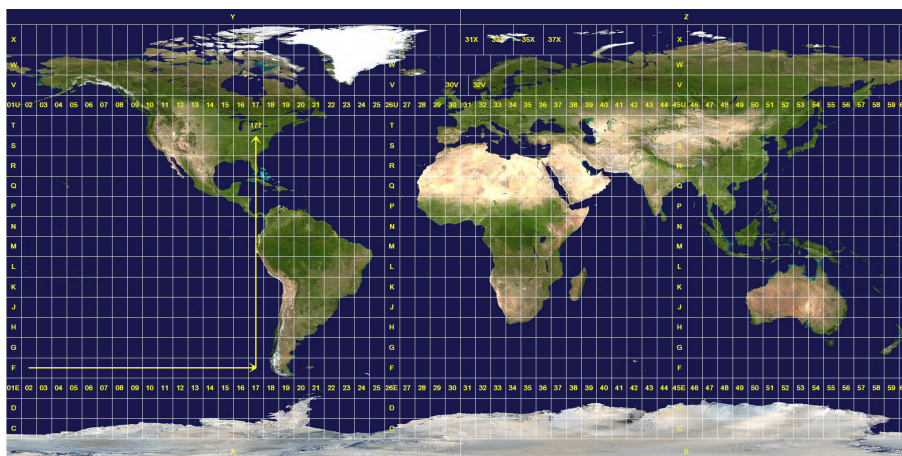


Figura B.2: Zone in cui la superficie terrestre è divisa dal sistema di coordinate UTM

B.4.2 East, North, Up

Sono simili alle coordinate UTM, soltanto la relazione con le coordinate geodetiche lat-lon non è fisso, in quanto il piano di tangenza va definito volta per volta, e

sulla base di dove lo si posiziona cambia il valore delle coordinate *East*, *North*, *Up* (ENU) che si ottengono. La proiezione di due punti sullo stesso piano di tangenza, mi permette di lavorare su quel piano con le regole della geometria euclidea. Va sempre tenuto a mente però, che a causa della curvatura della superficie terrestre, più il punto che vado a proiettare sul piano è lontano dal punto di tangenza, più la distanza fra i due calcolata sul piano sarà imprecisa rispetto a quella reale sulla superficie terrestre (figura B.3) [52].

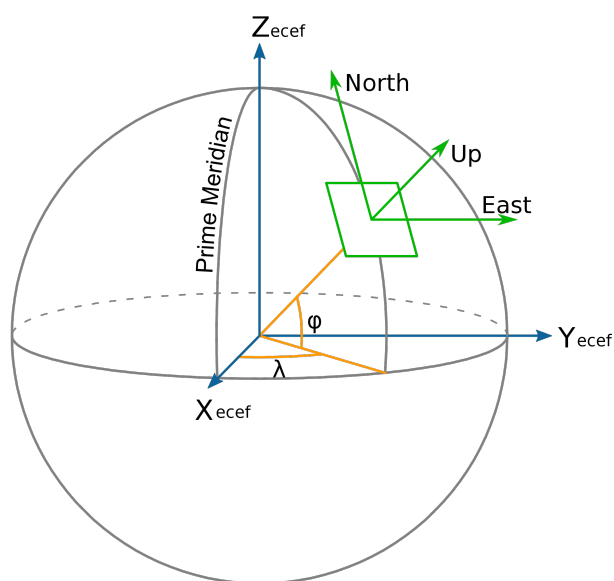


Figura B.3: Schema utilizzato per il calcolo delle coordinate ENU

Bibliografia

- [1] I. Chlamtac, M. Conti, J. Liu, *Mobile ad hoc networking: imperatives and challenges*, 2003.
- [2] J. Ahola, *Ambient Intelligence*, ERCIM (European Re-search Consortium for Information and Mathematics) NEWS, N. 47, October 2001.
- [3] C. Bisdikian, *An overview of the Bluetooth wireless technology*, IEEE Communication Magazine, December 2001.
- [4] R. Bruno, M. Conti, E. Gregori, *Bluetooth: architecture, protocols and scheduling algorithms*, Cluster Computing Journal 5, 2002.
- [5] S. Basagni, I. Chlamtac, A.V.R. Syrotiuk, *Location aware one-to-many communication in mobile multi-hop wireless networks*. In Proceedings of the IEEE Vehicular Tech-nology (VTC), Tokyo, Japan, May 2000.
- [6] Web site of the Bluetooth Special Interest Group: <http://www.bluetooth.com/> .
- [7] Specication of the Bluetooth System, Version 1.1, February 2001.
- [8] M. Conti, *Body, personal, and local wireless ad hoc networks*, in: M. Ilyas (Ed.), Handbook of Ad Hoc Networks, CRC Press, New York, 2003.
- [9] *Special issue on mobile ad hoc networking*, Cluster Computing Journal 5, April 2002.
- [10] I. Chlamtac, S. Kutten, *On broadcasting in radio networksproblem analysis and protocol design*, IEEE Transactions on Communications COM-33, 1985.

- [11] I. Chlamtac, A. Lerner, *Link allocation in mobile radio networks with noisy channel*, in: IEEE INFOCOM, Bar Harbour, FL, April 1986.
- [12] I. Chlamtac, A. Lerner, *Fair algorithms for maximal link activation in multi-hop radio networks*, IEEE Transactions on Communications COM-35, 1987.
- [13] M.S. Corson, J.P. Maker, J.H. Cernicione, *Internet-based mobile ad hoc networking*, IEEE Internet Computing 3, 1999.
- [14] C.-F. Chiasserini, R.R. Rao, *Pulsed battery discharge in communication devices*, in: Proceedings of The Fifth Annual ACM/IEEE International Conference on Mobile Computing and Networking (MOBICOM '99), Seattle, WA, August 1999.
- [15] E. Mingozzi, *QoS support by the HiperLAN/2 MAC protocol: a performance evaluation*, Cluster Computing Journal 5, 2002.
- [16] ETSI Technical Report 101 683, V1.1.1, *Broadband Radio Access Networks (BRAN): High Performance Local Area Network (HiperLAN) Type 2; System Overview*.
- [17] J. A. Freebersyser, B. Leiner, *A DoD perspective on mobile ad hoc networks*, in: Charles E. Perkins (Ed.), Ad Hoc Networking, Addison Wesley, Reading, MA, 2001.
- [18] S. Giordano, *Mobile ad-hoc networks*, in: I. Stojmenovic (Ed.), Handbook of Wireless Networks and Mobile Computing, Wiley, New York, 2002.
- [19] S. Giordano, I. Stojmenovic, *Position based ad hoc routes in ad hoc networks*, in: M. Ilyas (Ed.), The Handbook of Ad Hoc Wireless Networks, CRC Press, New York, 2003.
- [20] IBM Zurich Research Laboratory web site: <http://www.zurich.ibm.com/cs/wireless/bluetooth.html> .
- [21] IEEE standard for Wireless LAN-Medium Access Control and Physical Layer Specification, P802.11, November 1997.
- [22] IEEE P802.11/D10, January 14, 1999.

- [23] Web site of the IEEE 802.11 WLAN: <http://group-er.ieee.org/grups/802/11/main.html> .
- [24] E.D. Kaplan (Ed.), *Understanding GPS: Principles and Applications*, Artech House, Boston, MA, 1996.
- [25] Y. B. Lin, I. Chlamtac, *Wireless and Mobile Network Architectures*, Wiley, New York, 2000.
- [26] Y. B. Lin, Y.R. Huang, A. Pang, I. Chlamtac, *All-IP approach for third generation mobile networks*, IEEE Network Magazine, in press.
- [27] B.A. Miller, C. Bisdikian, *Bluetooth Revealed*, Prentice Hall, Englewood Clis, NJ, 2000.
- [28] M. Mauve, J. Widmer, H. Hartenstein, *A survey on position-based routing in mobile ad-hoc networks*, IEEE Network 15, 2001.
- [29] P. Kuosmanen, *Classification of ad hoc routing protocols*, Finnish Defence Forces, Naval Academy, Finland.
- [30] W. Stallings, *Local & Metropolitan Area Networks*, Prentice Hall, Englewood Clis, NJ, 1996.
- [31] A. Savvides, M. Srivastava, *Location discovery*, in: S. Basagni, M. Conti, S. Giordano, I. Stojmenovic (Eds.), *Ad Hoc Networking*, IEEE Press Wiley, New York, 2003.
- [32] M. Weiser, *The Computer for the Twenty-First Century*, Scientific American, 1991.
- [33] J. Widmer, M. Mauve, H. Hartenstein, H. Füßler, *Position-based routing in ad-hoc wireless networks*, in: Mohammad Ilyas (Ed.), *The Handbook of Ad Hoc Wireless Networks*, CRC Press, Boca Raton, FL, USA, 2003.
- [34] Wireless World Research Forum (WWRF): <http://www.ist-wsi.org> .
- [35] N. Abramson, *THE ALOHA SYSTEM: another alternative for computer communications*. In Proceedings of the November 17-19, 1970, Fall Joint Computer Conference. Houston, Texas, November 17 - 19, 1970.

- [36] ABI-Research, *Handset-based navigation for pedestrians; trends, market players, and forecasts*. Technical report, ABI Research Technology and Market Intelligence Company, September 2009.
- [37] ABI-Research, *Traffic information for navigation systems; services and infrastructure for real-time, historical, and predictive traffic data*. Technical report, ABI Research Technology and Market Intelligence Company, September 2007.
- [38] Census, *Topologically Integrated Geographic Encoding and Referencing system*. <http://www.census.gov/geo/www/tiger/> .
- [39] D.L. Zimmerman, X. Fang, S. Mazumdar, G. Rushton, *Modeling the probability distribution of positional errors incurred by residential address geocoding*. International Journal of Health Geographics, 2007.
- [40] N. Bicocchi, G. Castelli, M. Mamei, A. Rosi, F. Zambonelli, *Supporting location-aware services for mobile users with the whereabouts diary*. In Proceedings of the 1st international conference on MOBILE Wireless MiddleWARE, Operating Systems, and Applications. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering) ICST, Brussels, Belgium, Belgium, 2008.
- [41] J. Yoon, B. Noble, M. Liu, *Surface street traffic estimation*. In Proceedings of the 5th international conference on Mobile systems, applications and services, 2007.
- [42] E. Inc. ESRI Shapefile Technical Description. URL: <http://www.esri.com> , 1998.
- [43] ABI-Research, *Gps-enabled handsets; market, technology and business factors driving the growth of gps, a-gps and hybrid positioning for mobile handset location*. Technical report, ABI Research Technology and Market Intelligence Company, March 2009.
- [44] ABI-Research, *Sport and fitness gps devices and applications; gps watches, cycling computers, golf rangefinders, and smartphone applications*. Technical re-

port, ABI Research Technology and Market Intelligence Company, September 2009.

- [45] Google Inc. Google maps navigation (beta), <http://www.google.com/mobile/navigation/> , February 2010.
- [46] Dash Express Inc.; dash driver network -TruTrafic. <http://www.dash.net/product/> .
- [47] NAVTEQ Inc.; company site. <http://www.navteq.com> , February 2010.
- [48] TELEATLAS Inc.; whitepapers: Hd trafic -report n. ta-ct031959. <http://www.teleatlas.com/> , February 2010.
- [49] TomTom Inc.; tomtom plus services. <http://www.tomtom.com/plus/> , February 2010.
- [50] M. Haklay, P. Weber, *OpenStreetMap: user-generated street maps*. IEEE Pervasive Computing, 2008.
- [51] A. S. Tanenbaum, *Computer Networks*. Prentice Hall, Upper Saddle River, NJ (2003).
- [52] Geodetic System, Wikipedia, http://en.wikipedia.org/wiki/Geodetic_system .
- [53] GPX Explained, http://www.topografix.com/gpx_for_users.asp .
- [54] GPX 1.1 Schema Documentation, <http://www.topografix.com/GPX/1/1/> .
- [55] Holux GR-213 GPS Riceiver specification, http://www.holux.com/JCore/en/products/products_spec.jsp?pno=127 .
- [56] Hamlet HBTGPS20 GPS recevier, <http://www.hamletcom.com/ProductDetails.aspx?sid=aac8d3f2d8254146b1cec7fef7aa58ae&ProductId=3589> .
- [57] IEEE 802.11, The Working Group Setting the Standards for Wireless LANs, <http://www.ieee802.org/11/> .

- [58] Cahners In-Stat Group, <http://www.instat.com/> .
- [59] P. Inchingolo, *Presentation on Information Society Technologies and Methodologies in Logistics*, Trieste, January 2007.
- [60] GPS explained: Error sources, Kowoma, <http://www.kowoma.de/en/gps/errors.htm> .
- [61] Texas Commission on Environmental Quality, *Horizontal Accuracy Standards for TCEQ GIS Positional Data*.
- [62] A. H. Robinson, *Elements of Cartography*, 5th ed. New York: John Wiley & Sons, 1984.
- [63] S. Steiniger, M. Neun, A. Edwardes, *Foundations of Location Based Services*, University of Zurich, 2006.
- [64] GSM Association, *Permanent Reference Document SE.23: Location Based Services*, 2003.
- [65] S. Wang, J. Min, B. K. Yi, *Location Based Services for Mobiles: Technologies and Standards*, IEEE International Conference on Communication (ICC) 2008, Beijing, China.
- [66] Linus Torvalds, Wikipedia, http://it.wikipedia.org/wiki/Linus_Torvalds .
- [67] S. Savasta, M. Pini, G. Mara, *Performance Assessment of a Commercial GPS Receiver for Networking Applications*. In Proceedings IEEE International Consumer Communications and Networking Conference, Las Vegas, NV, USA, 2008.
- [68] Dale DePriest, *NMEA data*, <http://www.gpsinformation.org/dale/nmea.htm> .
- [69] S. Rogers, P. Langley, C. Wilson, *Mining GPS data to augment road models*. In Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM New York, NY, USA, 1999.
- [70] SiRF Technology, http://www.sirf.com/products/gps_chip.html .

- [71] Il Sole 24 ORE, *Crowdsourcing: quando la rete ...trova la soluzione*, <http://www.ilsole24ore.com/art/SoleOnLine4/Economia%20e%20Lavoro/2009/02/crowdsourcing-rete-soluzione.shtml>
- [72] E. Edlund, *Tactical ad-hoc networks, Model suggestions and simulations*, Swedish Defence Materiel Administration, Royal Institute of Technology, May 2003.
- [73] Navy Navigation Satellite System, <http://sd-www.jhuapl.edu/Transit/>
- [74] S. Dutch, *The Universal Transverse Mercator System*, Natural and Applied Sciences, University of Wisconsin, 2000.
- [75] W. Fifer, F. Bruno, *The low-cost packet radio*, Proceedings of the IEEE, 1987.
- [76] B. Leiner, R. Ruth, A.R. Sastry, *Goals and challenges of the DARPA GloMo program*, IEEE Personal Communications, 1996.
- [77] R. Stewart, *Note sullo standard NMEA 0183 in relazione al GPS*, 2003.
- [78] D. Woods, *The Myth of Crowdsourcing*, September 2009.
- [79] International Telecommunication Union, *Explanation of industrial, scientific and medical (ISM) applications*, <http://www.itu.int/ITU-R/terrestrial/faq/index.html>
- [80] Android Operating System Website, <http://www.android.com/>
- [81] Y. Zheng, L. Liu, L. Wang, X. Xie, *Learning Transportation Mode from Raw GPS Data for Geographic Applications on the Web*, Microsoft Research Asia, Beijing, China, 2008.