Alma Mater Studiorum · Università di Bologna

SCUOLA DI SCIENZE
Corso di Laurea Magistrale in Informatica

Progetto e sviluppo di un'estensione del protocollo MQTT per la gestione di dati geospaziali con meccanismi di privacy

Relatore:
Chiar.mo Prof.
Marco Di Felice

Presentata da:
Marco Silvestri

Sessione II
Anno Accademico 2019/2020

A Caterina.

Parole chiave

MQTT

Privacy

Geofencing

Servizi di geolocalizzazione

Internet of Things

"We can only see a short distance ahead, but we can see plenty there that needs to be done."

Alan Mathison Turing

Scritta in LaTeX

Abstract

Molte aziende nel mondo tengono traccia dei movimenti motivando la loro attività affermando che gli utenti stessi acconsentono a essere tracciati, i dati geolocalizzati memorizzati sono anonimi e al sicuro. É doveroso ricordare infatti che i dati sulla posizione contengono miliardi di punti senza informazioni identificabili come nomi o indirizzi e-mail; il problema in materia sussiste in quanto è stato dimostrato essere fattibile il collegamento di nomi reali ai punti che appaiono sulle mappe nel mondo. Si possono quindi ricordare dozzine di aziende che traggono profitto da tali dati ogni giorno in tutto il mondo, raccogliendoli direttamente dagli smartphone. Per questo motivo in questo progetto di tesi è stato affrontato il tema della privacy nella geolocalizzazione, proponendo un'estensione di un protocollo molto utilizzato in vari ambiti dell'Internet of Things (IoT): MQTT. Il protocollo è stato scelto per le sue ottime prestazioni in sensori e dispositivi nell'ambito dell'IoT, come lo stato dell'arte riportato conferma. Il focus nel mondo dell'IoT è avvenuto in seguito alla presenza continua di questi dispositivi in automobili, ambienti di lavoro, domestici e molti altri, rendendo così un'esigenza fondamentale la regolazione del trattamento dei dati di geolocalizzazione all'interno di questi sensori. Si è realizzato un meccanismo di abbonamento ad un servizio fornito in base ad un geofence precedentemente progettato da determinati utenti. Si rilevano l'entrata, l'uscita e la permanenza nei geofence stessi al fine di avvertire gli utenti con messaggi da loro preimpostati e con notifiche relative al servizio a cui si sono volontariamente abbonati.

Indice

Elenco delle figure

Introduzione

"Every minute of every day, everywhere on the planet, dozens of companies — largely unregulated, little scrutinized — are logging the movements of tens of millions of people with mobile phones and storing the information in gigantic data files." É così che l'articolo pubblicato sul New York Times il 19 Dicembre 2019 da Stuart A. Thompson e Charlie Warzel motiva l'attenzione verso il tema della privacy, argomento oggigiorno di grande attualità e rilevanza. Gli autori affermano anche che "In the United States, as in most of the world, no federal law limits what has become a vast and lucrative trade in human tracking". Come riportato dall'articolo stesso, molte aziende tengono traccia dei movimenti degli utenti, giustificando tali attività sulla base di tre affermazioni: le persone hanno dato il loro consenso a essere tracciate, i dati sono anonimi e sono gestiti in maniera sicura. É corretto affermare che i dati sulla posizione contengono miliardi di punti senza informazioni identificabili come nomi o indirizzi e-mail; tuttavia, molti studi dimostrano la possibilità di risalire ad informazioni sensibili a partire da punti che appaiono sulle mappe. Paul Ohm, professore di diritto e ricercatore sulla privacy presso il Georgetown University Law Center ha affermato che "è assolutamente impossibile rendere anonime le informazioni di geolocalizzazione longitudinali veramente precise."
Sulla base di tali motivazioni, in questo progetto di tesi è stato affrontato il tema della privacy nella geolocalizzazione, proponendo un'estensione di un protocollo molto utiliz-

zato nell'Internet of Things (IoT): MQTT. Il protocollo è stato scelto per le sue ottime prestazioni in sensori e dispositivi in ambienti intelligenti, come confermato dallo stato dell'arte. Nello specifico, si è realizzato un meccanismo di subscribe ad un servizio in base a un geofence precedentemente definito. Il meccanismo di sottoscrizione geospaziale consente di rilevare eventi quali: l'entrata, l'uscita e la permanenza nei geofence, inviando notifiche preimpostate agli utenti. In aggiunta, sono state studiate soluzioni per massimizzare la privacy dell'utente che condivide la posizione con il broker MQTT. L'elaborato è strutturato in due parti: la prima passa in rassegna lo stato dell'arte in ambito IoT, MQTT e privacy, mentre la seconda parte espone il progetto realizzato.
Nel primo capitolo viene dettagliato il protocollo MQTT con le relative specifiche. Nel secondo capitolo viene analizzata brevemente la normativa sulla privacy nello stato Italiano. Si pone particolare attenzione al capitolo terzo, in cui viene affrontato il tema della gestione della privacy di un dato geospaziale, focus del progetto di tesi. Nel quarto capitolo vengono inoltre dettagliate le principali tecniche di privacy che possono essere applicate nel mondo della geolocalizzazione. Con l'inizio del capitolo numero cinque, si affronta la progettazione della parte progettuale: vengono esposti gli obiettivi e le esigenze imposte dall'implementazione del progetto. Inoltre, è presente in dettaglio una panoramica delle tecniche utilizzate e dei limiti delle stesse. Vengono esposti anche i principali strumenti ottenuti. Nel sesto capitolo viene affrontata l'implementazione, ponendo attenzione alle tecnologie utilizzate e ai servizi realizzati. Infine, il progetto è stato oggetto di una dettagliata analisi che ha posto la validazione dei risultati come obiettivo fondamentale nello studio dell'elaborato. Nell'ultima parte sono state realizzate considerazioni di quello che è stato il lavoro svolto e di eventuali possibili sviluppi futuri.

Parte I

Stato dell'arte

Capitolo 1

Protocollo MQTT

1.1 L'avvento dell'IoT

Un ambiente intelligente è un contesto in cui tutti i dispositivi sono connessi e lavorano in modo collaborativo per rendere confortevole la vita delle persone. Con "ambiente" si intende semplicemente tutto ciò che circonda l'utente. É corretto anche specificare cosa si intende con il termine "intelligente": la capacità di ottenere e applicare in modo autonomo la conoscenza. Un ambiente intelligente ha la capacità di acquisire conoscenza dal luogo in cui opera (es. attraverso sensori), ed è in grado di adattare le proprie caratteristiche alle esigenze dei suoi utilizzatori, affinché essi possano migliorare la loro esperienza all'interno dell'ambiente stesso. Un elemento abilitante del concetto di ambienti intelligenti è quello dell'Internet of Things (IoT). Con IoT si intende un'evoluzione dell'uso della rete internet: gli oggetti (le "cose") si rendono riconoscibili e acquisiscono intelligenza grazie al fatto di poter comunicare dati su se stessi e accedere ad informazioni aggregate da parte di altri [25]. Si cita un articolo, molto utile nel fornire una panoramica generale di come l'IoT sta cambiando molti settori [6].

1.2 Un protocollo operante nell' IoT: MQTT

Si introduce ora un protocollo molto utilizzato nell'Internet of Things, ed oggetto di interesse dell'elaborato di tesi: il protocollo MQTT (Message Queuing Telemetry Transport Protocol). MQTT è un protocollo di messaggistica leggero progettato per la telemetria M2M (machine to machine) [8] in ambienti con risorse limitate. É stato proposto inizialmente da Andy Stanford Clark (IBM) e Arlen Nipper nel 1999 per il collegamento di sistemi di telemetria per oleodotti via satellite. In seguito è stato rilasciato da Royalty free nel 2010 e come standard OASIS nel 2014. "MQTT è un protocollo di messaggistica standard OASIS per Internet of Things (IoT). É progettato come un trasporto di messaggistica di pubblicazione/sottoscrizione estremamente leggero, ideale per connettere dispositivi remoti con un ingombro di codice ridotto e una larghezza di banda di rete minima. MQTT oggi è utilizzato in un'ampia varietà di settori, come automobilistico, manifatturiero, delle telecomunicazioni, petrolio e gas, ecc." [26]. MQTT è basato sul protocollo TCP: mantiene aperta la connessione TCP tra un client e un broker il più a lungo possibile, tramite messaggi PINGREQ. É progettato per connessioni con postazioni remote dove è richiesta un basso impatto energetico e una larghezza di banda limitata [73]. La descrizione tecnica del protocollo può essere recuperata dai documenti di specifica forniti da OASIS [1] [2] [3]. Il paradigma su cui si basa è quindi il paradigma "Publish/Subscribe" [4].

1.2.1 Paradigma Publish/Subscribe

Come già anticipato in precedenza, il paradigma consente la comunicazione tra più nodi ed è molto utilizzato oggigiorno. Dal punto di vista architetturale, il protocollo coinvolge tre attori:

- Publisher: hanno il ruolo di produrre dati sotto forma di eventi e inviarli al broker;
- Subscriber: hanno il ruolo di dichiarare interesse sui dati pubblicati con iscrizioni (per eventi specifici) e ricevere notifiche quando è disponibile un nuovo messaggio per l'argomento;
- Broker: svolgono servizio di notifica degli eventi (ENS). Notifica a ciascun abbonato ogni evento pubblicato che corrisponde ad almeno uno dei suoi abbonamenti avvenuti tramite iscrizione. Questi quindi possono filtrare i dati in base all'argomento e distribuirli agli abbonati.

É importante ricordare che i ruoli di editori/abbonati sono puramente logici. Il paradigma è generale e può essere applicato su molti casi d'uso diversi di sistemi distribuiti di rete. Con "argomento" si definisce il contesto del messaggio.
Altro aspetto importante da ricordare è che non esiste comunicazione diretta tra i client (i messaggi di dati vengono sempre inoltrati tramite il broker).

1.2.2 Problemi risolti dal paradigma Publish/Subscribe

Rispetto al paradigma Client-Server, abbiamo numerosi vantaggi:

- **Modello di comunicazione molti a molti**: abbiamo delle interazioni in un ambiente in cui diversi produttori e consumatori possono comunicare tutti contem-

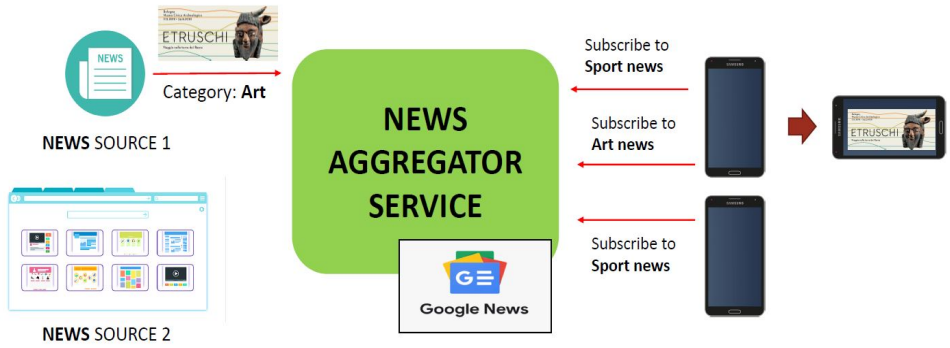


Figura 1.1: Esempio d'uso del paradigma Publish/Subscribe. Sorgente [66]

poraneamente. Ogni informazione può essere consegnata nello stesso momento a vari consumatori, che ricevono informazioni da vari produttori;

- **Disaccoppiamento spaziale**: le parti che interagiscono non hanno bisogno di conoscersi. L'indirizzamento dei messaggi si basa sul loro contenuto;

- **Disaccoppiamento temporale**: le parti che interagiscono non hanno necessità di partecipare attivamente all'interazione contemporaneamente. La consegna delle informazioni è mediata da una terza componente;

- **Disaccoppiamento della sincronizzazione**: il flusso di informazioni dai produttori ai consumatori è mediato. Non è necessaria quindi la sincronizzazione tra le parti che interagiscono: queste non devono conoscersi. L'indirizzamento dei messaggi si basa sul loro contenuto;

- **Interazioni push/pull**: sono consentiti entrambi i metodi (sia push sia pull).

Queste caratteristiche rendono il paradigma perfettamente adatto per applicazioni distribuite che si basano sulla comunicazione incentrata sui documenti.

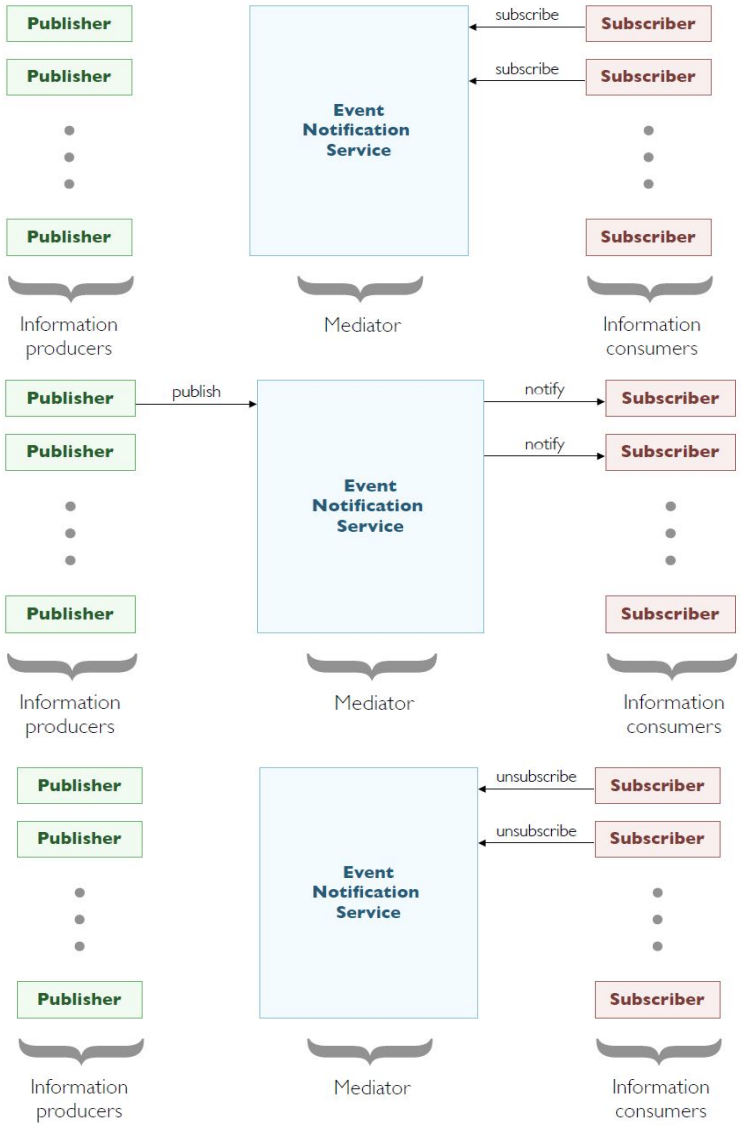


Figura 1.2: Immagine del funzionamento paradigma Publish/Subscribe. Sorgente [4]

1.3 Sicurezza del protocollo

In [43], viene progettato "il primo studio sistematico sulla protezione dei principali cloud IoT (ad esempio, AWS, Microsoft, IBM) messo in atto per il protocollo di messaggistica probabilmente più popolare MQTT." Nel documento hanno scoperto che "le aggiunte di sicurezza di queste piattaforme al protocollo sono tutte vulnerabili, consentendo all'avversario di ottenere il controllo del dispositivo, lanciare un attacco denial-of-service su larga scala, rubare i dati segreti della vittima e falsificare la vittima". Mentre in [48] il sistema proposto basato su MQTT prevede di lavorare anche tramite l'utilizzo di una rete non collegata ad internet. Di conseguenza, si può affermare che il sistema è protetto e non è possibile accedervi da software illegale o da altre reti. Infine, in [49], viene proposta una nuova tecnica di tutelare i dati utilizzando uno standard di interoperabilità aperto per l'IoT, che permette di mantenere il controllo sui dati personali (ovviamente abbiamo all'interno del sistema l'uso del protocollo MQTT).

1.4 Formato del pacchetto di controllo

Si analizza ora in dettaglio come è composto il pacchetto che il protocollo permette di scambiare. Questo pacchetto di controllo è formato principalmente da 3 sezioni:

1. **Fixed**: la prima parte è fissa e composta da solo 2 bytes: contiene il file tipo di pacchetto, la dimensione del carico utile e il livello della Qualità del Servizio (QoS);

2. **Variable**: la seconda parte è opzionale, ha una dimensione variabile;

3. **Payload**: la terza e ultima parte è a sua volta anche essa opzionale e ha una dimensione massima pari a 256 megabyte.

Le ultime due sezioni contengono i parametri aggiuntivi in base al tipo di comando. Nell'intestazione del messaggio publish/subscribe si trova il campo topic. Il topic è un campo stringa, senza un formato specifico, che adotta solo convenzioni di denominazione.

1.5 Livelli QoS del protocollo

MQTT utilizza diversi modelli per lo scambio di messaggi. Questi sono noti essere "Quality of Service" (QoS) [5]. Abbiamo vari profili QoS:

1. **QoS0**: il publisher invia i dati al broker e non attende il riconoscimento (ACK) dal broker. In questo profilo non esistono ritrasmissioni, e se i dati inviati non vengono ricevuti vengono persi;

2. **QoS1**: dopo aver pubblicato i dati, il publisher attende un ACK (chiamato PUBACK) dal broker. Passato un tempo predefinito, se non si riceve l'ACK, si ritrasmettono i dati. In confronto al profilo QoS0, si migliora l'affidabilità ma si aumenta l'overhead.

3. **QoS2**: il publisher invia un dato al broker e attende il messaggio PUBREC (PUBlish RECeived) da esso. Quando riceve il PUBREC, l'editore scarta il riferimento ai dati pubblicati per poi inviare un PUBREL (PUBLish RELeased) al broker. Il broker a sua volta compie la stessa procedura. Quando il flusso è completato, entrambe le parti possono essere sicure che il messaggio è stato consegnato.

1.6 Comparazione con altri protocolli

Come citato nell'articolo [8], nell'Internet Of Things non si può fare affidamento ad un unico protocollo per tutte le sue necessità delle applicazioni. Non a caso, sono disponibili numerosi protocolli di messaggistica per lo scambio dati in contesti IoT. Il criterio di scelta si può affermare essere un requisito richiesto dal sistema IoT.
Possiamo citare quindi quattro protocolli fondamentali ampiamente accettati ed emergenti per sistemi come questo: MQTT (già trattato), CoAP, AMQP e HTTP. Prima di procedere con l'analisi del confronto tra i protocolli, è doveroso evidenziare come questo confronto può variare in alcune circostanze in base alle componenti IoT utilizzate, e potrebbe, di conseguenza, riflettere risultati comparativi diversi. Questo studio riportato non considera le condizioni di rete dinamiche e gli overhead che si verificano nella ritrasmissione dei pacchetti, che possono anche modificare i risultati del confronto. Riportiamo nella tabella seguente la comparazione presa dall'articolo appena citato.

1.6.1 Analisi della dimensione del messaggio e del sovraccarico del messaggio

HTTP, come da figura, è il protocollo con dimensione del messaggio e overhead (sovraccarico) dello stesso più elevato. MQTT, AMQP e HTTP vengono eseguiti su TCP, quindi c'è un sovraccarico per la creazione e la chiusura della connessione. MQTT anche se ha un header piccolo, compensa grazie al requisito minimo di connessione TCP che consente di aumentare l'overhead complessiva e quindi l'intera dimensione del messaggio.

Criteria	MQTT	CoAP	AMQP	HTTP
1. Year	1999	2010	2003	1997
2. Architecture	Client/Broker	Client/Server or Client/Broker	Client/Broker or Client/Server	Client/Server
3. Abstraction	Publish/Subscribe	Request/Response or Publish/Subscribe	Publish/Subscribe or Request/Response	Request/Response
4. Header Size	2 Byte	4 Byte	8 Byte	Undefined
5. Message Size	Small and Undefined (up to 256 MB maximum size)	Small and Undefined (normally small to fit in single IP datagram)	Negotiable and Undefined	Large and Undefined (depends on the web server or the programming technology)
6. Semantics/ Methods	Connect, Disconnect, Publish, Subscribe, Unsubscribe, Close	Get, Post, Put, Delete	Consume, Deliver, Publish, Get, Select, Ack, Delete, Nack, Recover, Reject, Open, Close	Get, Post, Head, Put, Patch, Options, Connect, Delete
7. Cache and Proxy Support	Partial	Yes	Yes	Yes
8. Quality of Service (QoS)/ Reliability	QoS 0 - At most once (Fire-and-Forget), QoS 1 - At least once, QoS 2 - Exactly once	Confirmable Message (similar to At most once) or Non-confirmable Message (similar to At least once)	Settle Format (similar to At most once) or Unsettle Format (similar to At least once)	Limited (via Transport Protocol - TCP)
9. Standards	OASIS, Eclipse Foundations	IETF, Eclipse Foundation	OASIS, ISO/IEC	IETF and W3C
10. Transport Protocol	TCP (MQTT-SN can use UDP)	UDP, SCTP	TCP, SCTP	TCP
11. Security	TLS/SSL	DTLS, IPSec	TLS/SSL, IPSec, SASL	TLS/SSL
12. Default Port	1883/ 8883 (TLS/SSL)	5683 (UDP Port)/ 5684 (DLTS)	5671 (TLS/SSL), 5672	80/ 443 (TLS/SSL)
13. Encoding Format	Binary	Binary	Binary	Text
14. Licensing Model	Open Source	Open Source	Open Source	Free
15. Organisational Support	IBM, Facebook, Eurotech, Cisco, Red Hat, Software AG, Tibco, ITSO, M2Mi, Amazon Web Services (AWS), InduSoft, Fiorano	Large Web Community Support, Cisco, Contiki, Erika, IoTivity	Microsoft , JP Morgan, Bank of America, Barclays, Goldman Sachs, Credit Suisse	Global Web Protocol Standard

Figura 1.3: Tabella di comparazione dei protocolli operanti in IoT: MQTT, CoAP, AMQP e HTTP. Sorgente [8]

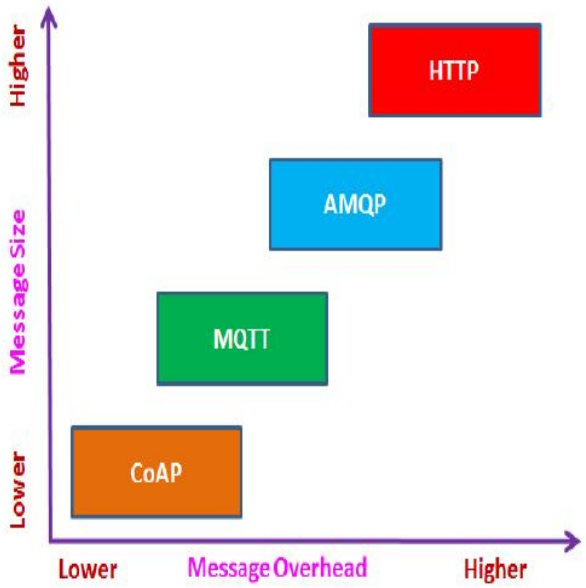


Figura 1.4: Analisi della dimensione del messaggio e del sovraccarico del messaggio. Sorgente [8]

1.6.2 Analisi del consumo di energia e del fabbisogno di risorse

HTTP richiede risorse più elevate rispetto agli altri protocolli. CoAP e MQTT invece sono progettati per dispositivi con larghezza di banda ridotta e risorse limitate, anche se si è dimostrato che CoAP ne consuma in quantità minore. AMQP, si differenza dagli ultimi due, perché richiede una quantità maggiore di energia e risorse a causa dell'esecuzione di altre operazioni necessarie per il provisioning e l'affidabilità.

1.6.3 Analisi della larghezza di banda e della latenza

HTTP introduce l'utilizzo di banda e la latenza maggiori rispetto a qualsiasi altro protocollo. Il protocollo TCP non aiuta a migliorare la latenza, motivo per cui MQTT, AMQP e HTTP non utilizzano completamente la larghezza di banda. Ciò è dovuto al fatto che si vuole evitare la congestione della rete. Si è dimostrato che MQTT richiede

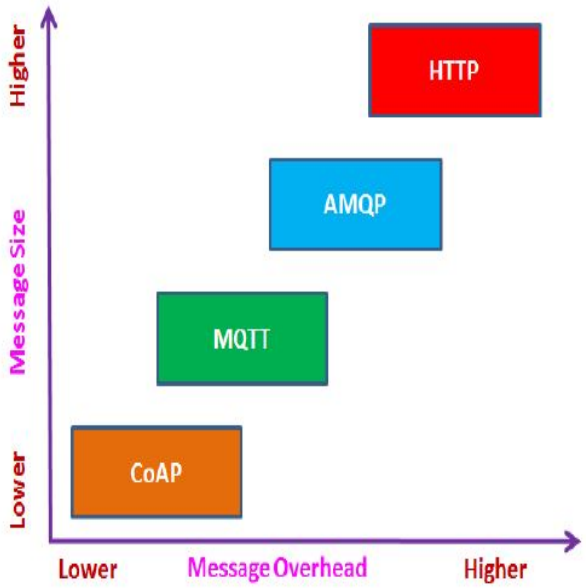


Figura 1.5: Analisi del consumo di energia e del fabbisogno di risorse. Sorgente [8]

approssimativamente quasi il doppio della larghezza di banda rispetto a CoAP. Si ricorda che anche i servizi extra di AMQP richiedono una larghezza di banda e una latenza maggiore.

1.6.4 Analisi dell'affidabilità/QoS e della interoperabilità

MQTT in questo caso offre il più alto livello di qualità dei servizi con la minima interoperabilità tra i quattro. HTTP, anche se fornisce la massima interoperabilità sul web, non include l'affidabilità come caratteristica di base. Si ricorda che i tre protocolli basati su TCP prevedono la garanzia di consegna di un pacchetto. A proposito di questo, MQTT definisce nei tre livelli QoS standard tre tipologie di garanzia differenti: in QoS0 prevede la sola garanzia TCP, in QoS1 prevede la garanzia MQTT con conferma, mentre in QoS2 è stata prevista la garanzia MQTT con handshake [51]. AMQP invece fornisce due livelli di QoS simili a quelli di MQTT (QoS0 e QoS1). CoAP compensa inaffidabilità

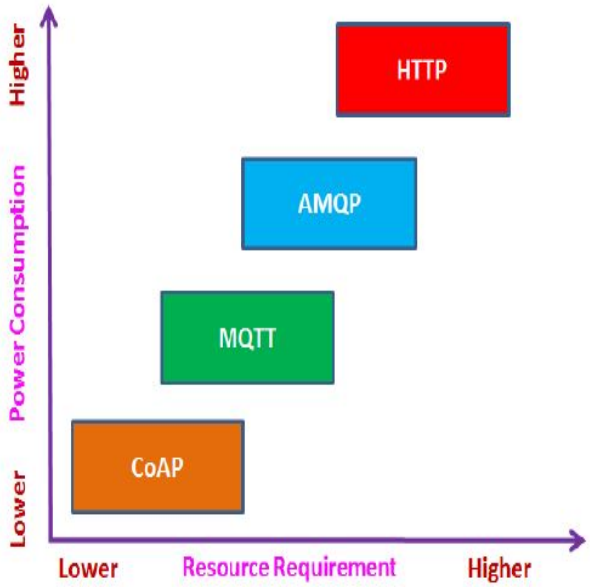


Figura 1.6: Analisi della larghezza di banda e della latenza. Sorgente [8]

del protocollo UDP definendo un meccanismo di ritrasmissione e di scoperta delle risorse che assomiglia al meccanismo previsto dal livello QoS0 e QoS1 di MQTT. Mentre HTTP si affida alla sola garanzia di base di TCP. Per quanto riguarda l'interoperabilità, si fa riferimento nuovamente al problema di base dell'IoT. Tutti e quattro i protocolli hanno problemi fondamentali in questo caso.

1.6.5 Analisi della sicurezza e del provisioning

Con questa analisi, il protocollo AMQP ha il più alto livello di supporto per la sicurezza e i servizi aggiuntivi. Si può considerare invece MQTT un protocollo di messaggistica base che supporta il livello più basso di sicurezza e servizi aggiuntivi. MQTT ha caratteristiche di autenticazione minime e si basa solo su nome utente e password. CoAP e HTTP prevedono due meccanismi di autenticazione, garantendo più sicurezza. AMQP fornisce massima sicurezza grazie alle sue tecniche. MQTT inoltre non offre servizi extra

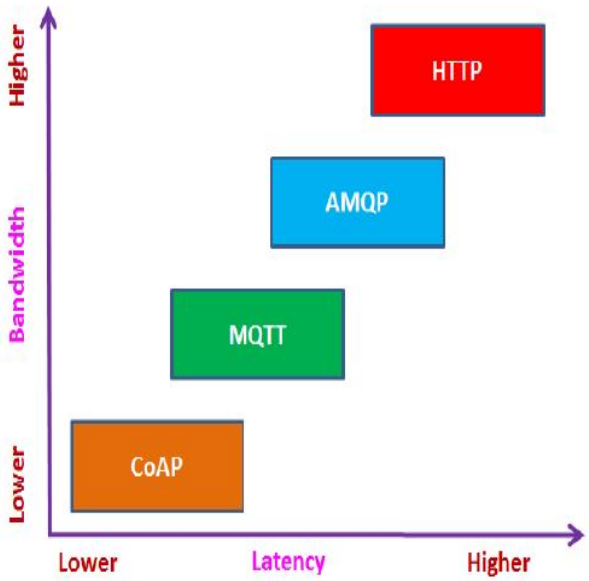


Figura 1.7: Analisi dell'affidabilità/QoS e della interoperabilità. Sorgente [8]

nemmeno per l'etichettatura dei messaggi: infatti questo necessita che i suoi client conoscano i formati dei messaggi in anticipo per consentire la comunicazione [52] (aspetto molto utilizzato nei capitoli seguenti nel progetto sviluppato). HTTP è uno standard web completo e offre diversi servizi. AMQP è quello maggiormente scelto dalle aziende per la sua sicurezza.

1.6.6 Analisi sull'utilizzo M2M nell'IoT e della standardizzazione

MQTT è stato utilizzato da un gran numero di organizzazioni, ma non è ancora uno standard globale. HTTP invece si può definire uno standard globale per il web, ma non nel mondo IoT. Si può affermare, infatti, che l'utilizzo di HTTP nell'IoT è limitato a causa delle sue complessità di gestione e della latenza.

MQTT è un protocollo M2M consolidato ed è stato utilizzato e supportato da un gran

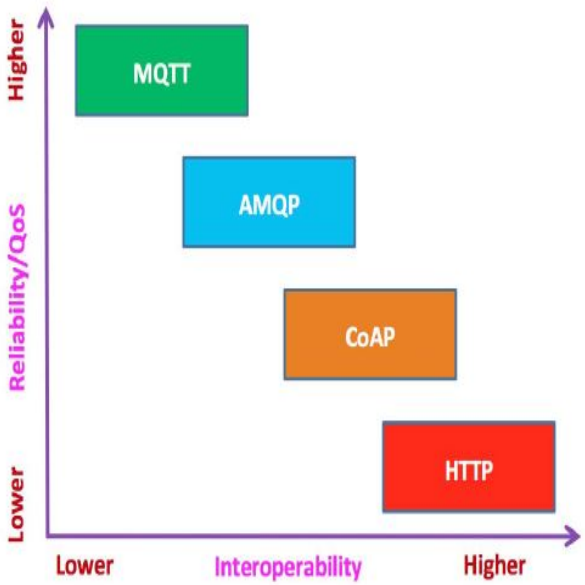


Figura 1.8: Analisi della sicurezza e del provisioning. Sorgente [8]

numero di organizzazioni come IBM, Facebook, Eurotech, Cisco, Red Hat, M2Mi, Amazon Web Services (AWS) e InduSoft. AMQP è il protocollo IoT di maggior successo impiegato nei più grandi progetti al mondo. Anche CoAP ha guadagnato rapidamente popolarità ed è supportato da molte grandi aziende.

MQTT si può definire infine come un protocollo emergente. É un protocollo de facto per l'IoT ed è supportato dagli standard aperti OASIS e dalla Eclipse Foundation [52] [53]. AMQP è sua volta uno standard internazionale adottato da OASIS ISO / IEC 19464: 2014 [7]. CoAP è uno standard IETF appositamente progettato per integrare IoT e Web e supportato da Eclipse Foundation [52]. Infine, HTTP è uno standard IETF e W3C, ed è riconosciuto essere uno standard globale per il Web [54].

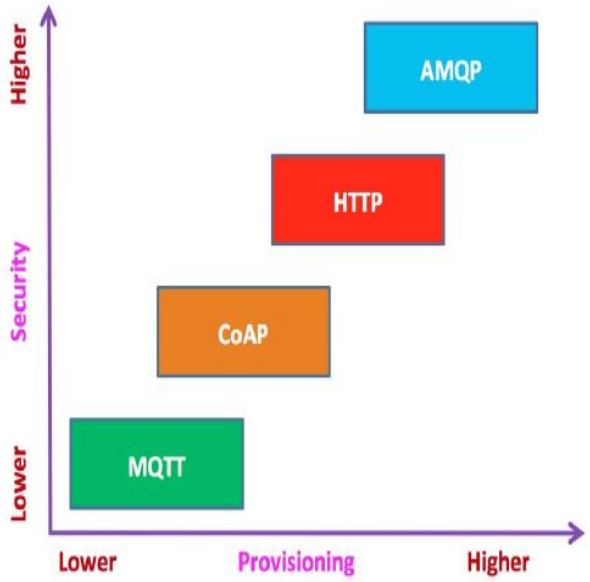


Figura 1.9: Analisi sull'utilizzo M2M nell'IoT e della standardizzazione. Sorgente [8]

1.7 La gestione dei dati geospaziali in MQTT

MQTT viene spesso utilizzato anche in ambito della geolocalizzazione. In [44] viene creata una rete di servizi cloud veicolari utilizzando il protocollo MQTT. Il sistema proposto cerca di evitare il più possibile problemi di traffico nelle aree di parcheggio molto trafficate come i centri commerciali e i parcheggi a lato della strada. Il termine coniato per l'aggiunta della geolocalizzazione al protocollo MQTT si chiama MQTT-G [45]. Nel documento [45], il protocollo proposto può essere utilizzato per offrire la geolocalizzazione come parte dell'infrastruttura di publish/subscribe, aiutando così le applicazioni in tempo reale per cui può essere utilizzato. Offrono versioni sia per l'ambiente C/C++ sia a client Android mobile. In [46] viene presentata la progettazione di un dispositivo incorporato che è destinato all'uso nelle carrozze dei treni per avvertire i pendolari quando arrivano a destinazione. La comunicazione è prevista attraverso protocolli come MQTT, e la tecnologia utilizzata è il GPS. Un altro importante contributo è stato quello

fornito in [47]. MQTT è ben noto per il suo basso consumo di energia e larghezza di banda che lo rende altamente adatto per le situazioni di messaggistica Green Internet of Things (IoT) in cui il consumo di energia è limitato, o in dispositivi mobili come telefoni, computer incorporati o microcontrollori. Anche in questo articolo, sempre [47], è stato presentato un nuovo protocollo chiamato MQTT-G, che include la geolocalizzazione nel suo funzionamento.

Capitolo 2

Privacy

2.1 Normativa sulla privacy

La privacy è un argomento di grande attualità. Si ricorda la prima definizione di A.F. Westin nel 1967: "La privacy è la pretesa di individui, gruppi o istituzioni di determinare da soli quando, come e in che misura le informazioni che li riguardano vengono comunicate ad altri...". Di seguito verranno proposte una serie di definizioni previste dalla legge che vanno a definire i dati oggetti della tutela della Privacy. Esistono molti interessanti temi affrontati in moltissimi articoli che trattano la Privacy nel mondo IoT. Alcuni analizzano la sicurezza del traffico di rete affermando che il traffico stesso viene utilizzato per identificare azioni e attività degli utenti con precisione elevata (anche con dati crittografati) [10]. In [11] vengono proposte quattro strategie ai produttori di sensori per proteggere la riservatezza degli utenti. Molti articoli propongono modelli e framework per tutelare la privacy valutando il livello di intrusione nella riservatezza degli utenti quando si utilizzano, ad esempio, cinque sensori IoT [9]. Questo aiuta a comprendere come in ogni tematica del mondo dell'IoT si pone attenzione alla letteratura legata al

tema della Privacy. É doveroso anche ricordare chi tutela la riservatezza degli italiani. Il Garante per la protezione dei dati personali è un'autorità amministrativa indipendente istituita dalla cosiddetta legge sulla privacy (legge 31 dicembre 1996, n. 675), poi disciplinata dal Codice in materia di protezione dei dati personali (d.lg. 30 giugno 2003 n. 196), come modificato dal Decreto legislativo 10 agosto 2018, n. 101. [15]. Si può quindi sottolineare l'importanza di questa autorità nell'elaborato. Il Garante non ha solo il compito di garantire la tutela dei diritti e delle libertà fondamentali, ma anche di tutelare il rispetto della dignità nel trattamento dei dati personali [16].
Si fornisce una panoramica sul diritto della privacy [13].

2.1.1 Dato personale

Il dato personale viene definito come "qualsiasi informazione riguardante una persona fisica identificata o identificabile (interessato); si considera identificabile la persona fisica che può essere identificata, direttamente o indirettamente, con particolare riferimento a un identificativo come il nome, un numero di identificazione, dati relativi all'ubicazione, un identificativo online o a uno o più elementi caratteristici della sua identità fisica, fisiologica, genetica, psichica, economica, culturale o sociale." come da Regolamento (UE) n. 2016/679 (noto come Regolamento generale per la protezione dei dati), Art. 4 comma 1.
Come dato personale si intende anche un'informazione costituita da suoni o immagini, ad esempio, un'intervista, un colloquio o qualsiasi manifestazione del pensiero proveniente dal soggetto interessato.
A seguito della creazione delle banche dati e dell'informatizzazione dell'informazione, è stato necessario sempre più normarne la circolazione e proteggere i dati [76]. Infatti,

"con l'evoluzione delle nuove tecnologie, altri dati personali hanno assunto un ruolo significativo, come quelli relativi alle comunicazioni elettroniche (via Internet o telefono) e quelli che consentono la geolocalizzazione, fornendo informazioni sui luoghi frequentati e sugli spostamenti." [14].

La riservatezza non è un elemento che caratterizza necessariamente il dato personale. I dati personali possono essere distinti a seconda che consentano di identificare un soggetto in maniera immediata o mediante il collegamento con un'informazione riferibile a questi o a terzi.

Il Codice in materia di protezione dei dati personali non include i dati anonimi, dati che non possono essere associati ad un interessato identificato o identificabile, anche privi di ogni riferimento e anche indiretti, al soggetto a cui sono collegati.

2.1.2 Dato sensibile

I dati sensibili costituiscono una sotto categoria dei dati personali che riguarda la personalità etico-sociale dell'individuo e le sue caratteristiche psico sanitarie. L'elencazione dei dati sensibili é "chiusa", cioè già definita, vista la rigorosa disciplina prevista per la tutela di questo sottogruppo. Inoltre, sui dati sensibili possono essere compiute operazioni di trattamento finalizzate al raggiungimento di singoli scopi, ma si deve verificare periodicamente se i dati sensibili utilizzati siano pertinenti e non in eccesso. Nel caso in cui i dati sensibili siano registrati in banche dati, elenchi o registri non cartacei, questo è permesso solo se necessario e deve avvenire mediante l'utilizzo di codici identificativi al fine di individuare gli interessati.

2.1.3 Trattamento dei dati personali

Il Codice in materia di protezione dei dati personali si applica al trattamento di dati sia con mezzi informatici sia con mezzi cartacei. Sono sottoposti a questa legge tutti i dati personali ad eccezione di quelli anonimi.

Comunicazione e diffusione

Tra le operazioni ricomprese nel concetto generale di trattamento, la comunicazione e la diffusione di dati personali hanno ricevuto una disciplina più rigida a causa dei maggiori rischi ad esse associate. Entrambe le operazioni si riferiscono alla trasmissione di dati personali, ma la comunicazione differisce dalla diffusione per la possibile determinazione del destinatario, che rimane al contrario indeterminato nel secondo caso.
Di seguito l'immagine contenente il ciclo del dato personale.

2.2 Privacy nella geolocalizzazione

"Un tipo speciale di privacy delle informazioni riguarda la pretesa degli individui di determinare da soli quando, come e in che misura le informazioni sulla loro ubicazione vengono comunicate ad altri". Questa è la definizione che M. Duckham ha fornito nel 2006 introducendo un nuovo tema: quello della privacy inerente alla geolocalizzazione.
Si ricordano innanzitutto alcuni fattori principali che sono necessari nella tutela della privacy. Ad esempio, in [34], viene ricordato come il sistema operativo e i software che sfruttano la geolocalizzazione dovrebbero essere aggiornati periodicamente, le patch dovrebbero essere implementate tempestivamente e i backup dei sistemi dovrebbero essere eseguiti regolarmente. Inoltre, dovrebbero essere implementati controlli di accesso fisico che limitino gli accessi non autorizzati. Tutto questo perché le tracce di mobilità umana

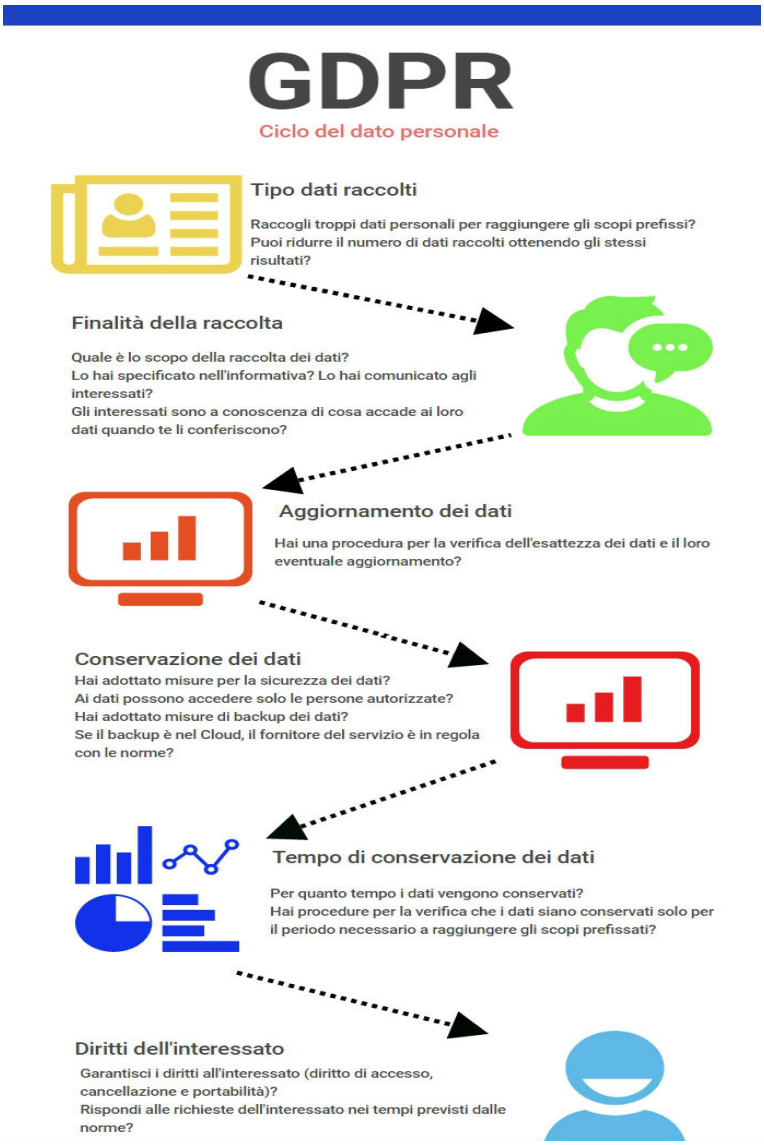


Figura 2.1: Ciclo del dato personale. Sorgente [18]

sono altamente uniche, cioè sono identificabili da una minima conoscenza preliminare. Nel mondo attuale si parla spesso della sicurezza nei dati geospaziali, quando pubblichiamo una foto con dato spaziale su un social o veniamo citati in qualche post contenente dati di questo tipo su internet. Infatti è importante sottolineare anche il ruolo dell'utente: dovrebbe comprendere le reali capacità di questa tecnologia. Gli utilizzatori dovrebbero tenersi aggiornati e aumentare la consapevolezza propria e dei propri cari relativamente alla gestione dei dati di geolocalizzazione: le azioni di familiari, amici e colleghi potrebbero rivelare informazioni basate sulla posizione che si vorrebbero voler mantenere private [20].

Possiamo anche affermare l'importanza di questo tema nelle aziende: molte imprese trattano dati di questo tipo. Il Garante per la protezione dei dati personali, nella newsletter n. 441 del 29 maggio 2018, ha evidenziato la possibilità del trattamento di dati personali mediante un sistema di localizzazione geografica dei dispositivi aziendali. L' articolo [12] tratta dell'importanza del Garante e del potere che detiene sulle aziende. Esiste molto materiale su provvedimenti presi dal Garante e discussi dai mass media nel campo della privacy dei dati di localizzazione, si veda [19]. Spesso il Garante fornisce anche dettagli precisi sulla gestione dei dati. In un provvedimento per un'azienda, il Garante ha stabilito come il servizio standard di geolocalizzazione fornito agli utenti doveva essere rimodulato, con particolare riguardo agli intervalli temporali di rilevazione della posizione geografica dei veicoli. Questi venivano fissati tra i 30 e i 120 secondi, mentre i tempi di conservazione dei dati venivano stabiliti in 365 giorni (come la memorizzazione e messa a disposizione delle mappe di tutti i percorsi effettuati). La società era tenuta anche ad informare i propri clienti della possibilità di adattare le caratteristiche del servizio alle concrete finalità perseguite. Inoltre, la funzione che permetteva la disattivazione del Gps doveva essere resa disponibile per tutti i tipi di abbonamento al servizio senza costi

aggiuntivi eccessivi [24].

Nella letteratura si parla molto di strategie ottimali per garantire ai nostri utenti una protezione maggiore dei loro dati geospaziali. L'obiettivo è quello di trovare un LP-PM (algoritmo di offuscamento della posizione che troviamo alla base di un efficace meccanismo di conservazione della privacy della posizione) ottimale che tenga conto di:

- requisiti di privacy degli utenti;
- le conoscenze e le capacità dell'avversario;
- il peggioramento della qualità del servizio offerto derivante dalla strategia utilizzata.

É fondamentale ricordare il trade-off che esiste tra l'esigenza della privacy della posizione e la correttezza della localizzazione. Ci sono vari studi che permettono di gestire quest'aspetto nel modo migliore: alcuni, per esempio, propongono strategie basate sulla teoria dei giochi [50].

Di seguito si definisce un "dato geospaziale" [17].

2.2.1 Dato di localizzazione

I dati di localizzazione, o dati sulla posizione (anche dati di mobilità) sono le informazioni trattate da una rete di comunicazione elettronica o da un servizio di comunicazione elettronica che indicano la posizione geografica dell'apparecchiatura terminale (es. smartphone o qualsiasi altro dispositivo in grado di gestire questa tipologia di dato) di un utente del servizio di comunicazione elettronica. In particolare sono tutti quei dati relativi alla:

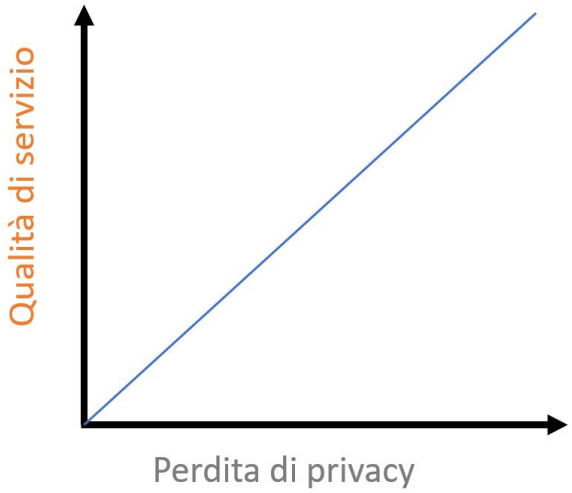


Figura 2.2: Compromesso tra privacy e corretta geolocalizzazione della posizione

- latitudine;
- longitudine;
- altitudine;
- direzione di marcia;
- ora di registrazione della posizione.

É fondamentale ricordare anche che i dati di localizzazione storici sulla posizione fanno ancora parte di questa categoria, e vengono tutelati in quanto permettono di dedurre comunque informazioni sull'utente, anche se rappresentano dati passati.

Possiamo suddividere i dati di localizzazione o geolocalizzazione in tre principali categorie: Geotagging, Geocoding e Georeferencing [20].

Il geotagging, o GeoTagging, è il processo di aggiunta di metadati di identificazione geografica a vari media come una fotografia o un video con geotag, siti web, messaggi

SMS, codici QR o feed RSS. Questa è una forma di metadati geospaziali. I geotagging sono dati molto spesso costituiti da coordinate di latitudine e longitudine. Alcune volte possono includere anche altitudine, rilevamento, distanza, dati di precisione, nomi di luoghi e qualche volta anche un timestamp [21].

La geocodifica è un processo che a partire da una descrizione testuale di una posizione (come un indirizzo o il nome di un luogo) restituisce le coordinate geografiche (tipo coppia latitudine/longitudine) per identificare una posizione sulla superficie terrestre. La geocodifica inversa, invece, converte le coordinate geografiche in una descrizione di una posizione, solitamente nel nome di un luogo o di una posizione indirizzabile [22].

Con georeferenziazione intendiamo il sistema di coordinate interne di una mappa o di un'immagine di una foto aerea che può essere correlato a un sistema di coordinate geografiche. Le trasformazioni delle coordinate rilevanti sono generalmente memorizzate nel file immagine [23].

2.2.2 Dispositivi di raccolta

Esistono vari dispositivi che permettono la raccolta di dati geospaziali. Nella maggioranza dei casi tali dati derivano dai dispositivi che un utente indossa, come smart band o fitness tracker, da dispositivi mobili come smartphone, tablet, navigatori satellitari o anche dalle apparecchiature wi-fi. Inoltre, è importante definire un LBS (Location Based System): acronimo con cui si intendono tutte quelle "applicazioni che forniscono agli utenti informazioni basate sulla loro posizione geografica, che potrebbero essere ottenute dal dispositivo mobile su cui stanno accedendo al servizio, o utilizzando una posizione definita manualmente" [55].

2.2.3 Tecnologie di localizzazione

I dati sulla posizione possono essere recuperati tramite le seguenti tecnologie:

- GPS (Global Positioning System), la tecnologia che comunemente viene citata come "rete satellitare". L'accuratezza varia a seconda della situazione e può arrivare fino a 5 metri. Questa precisione dipende da fattori come le condizioni meteorologiche o interferenze di vario genere. Si può abbinare anche ad altre tecnologie;

- le torri cellulari: utilizzate per la fornitura del servizio di comunicazione cellulare. I gestori di telefonia conoscono la posizione approssimativa di un dispositivo poiché questo dialoga continuamente con le torri. Dalla presenza in una "cella" e dalla forza del segnale può essere dedotta una probabile posizione del dispositivo. Di questo tracciamento l'operatore tiene un registro che può essere consultato solo dalle forze di polizia. Non si rilevano quindi problemi legati all'utilizzo di questi dati;

- le reti Wi-Fi; i dispositivi mobili possono ricavare la loro posizione eseguendo la scansione delle reti Wi-Fi (o dei punti di accesso) nelle vicinanze;

- i Beacon Bluetooth: sono piccoli trasmettitori radio che utilizzano segnali Bluetooth unidirezionali. Questi possono essere collegati a vari oggetti e installati in vari luoghi (es. musei). Se l'utente ne acconsente la connessione Bluetooth, possono trasmettere informazioni; Questo ci permette di dedurre la posizione del dispositivo;

- infine, può essere svolta una combinazione di segnali: i moderni smartphone combinano più segnali dalle fonti sopra indicate per calcolare la posizione in maniera più precisa, anche accorpando le informazioni fornite dagli innumerevoli sensori come altimetro e accelerometro.

I dati sulla posizione sono fondamentali per il funzionamento della connettività di base, cioè la capacità di inviare e ricevere contenuti sui dispositivi. Ricordiamo come i fornitori di servizi wireless, che utilizzano le torri cellulari della rete di telefonia mobile, sono sempre in grado di identificare un dispositivo tramite l'indirizzo IP, e quindi in grado di stabilire la sua posizione nello spazio.

2.2.4 Trattamento dei dati geospaziali

Come è già stato citato, tra i dati personali si annoverano i dati che trattano la posizione. Questi risultano essere soggetti ad una tutela rafforzata poiché espongono le persone ad un maggior rischio di diffusione di informazioni riguardanti la loro vita personale. Negli stati membri dell'Unione europea la direttiva "ePrivacy" è finalizzata a regolamentare l'accesso ai dati di localizzazione. Questi ultimi possono essere trattati, secondo la normativa, soltanto dai fornitori di servizi di comunicazione o fornitori di servizi a valore aggiunto. I dati possono essere trattati o in forma anonima oppure in base al consenso dell'interessato nel caso in cui l'elaborazione sia necessaria per la finalità relativa a servizi a valore aggiunto.

Questi servizi sono definiti tali poiché richiedono il trattamento di dati di traffico o di localizzazione, oltre a quello necessario per la trasmissione di una comunicazione. Appartiene a questa categoria di servizi il dato che utilizza l'informazione della posizione dell'utente chiamante al fine di fornirgli assistenza in caso di emergenze o problematiche del veicolo con cui si sposta.

Il consenso dell'utente deve essere specifico per il servizio e la relativa informazione

della localizzazione dell'utente deve essere corredata dall'indicazione riguardante i tipi di dati sulla posizione che verranno trattati: per quale scopo, per quanto tempo e se sono comunicati a terzi.

2.2.5 Dati di geolocalizzazione anonimi e aggregati

Se i dati di localizzazione sono anonimi, l'accesso ad essi è libero anche se dovrebbe essere riservato a casi essenziali. Infatti, rendere anonimi i dati di localizzazione, è estremamente difficile poiché utilizzandoli in sequenza si può facilmente ricostruire il movimento di una persona nel tempo (per esempio il tragitto casa-lavoro) e tale operazione nella maggior parte dei casi consente la de-anonimizzazione. Per di più, l'informazione relativa alla posizione dell'abitazione è facilmente individuabile poiché corrisponde al luogo dove un soggetto si trova generalmente di notte. Per tentare di risolvere il problema dell'anonimizzazione è stato proposto l'utilizzo di dati aggregati, risultato utile per migliaia di persone. Ma la classica "heat map" permette di ottenere informazioni private anche nel caso di dati aggregati.

2.2.6 Chi può trattare i dati sulla posizione

I dati sulla posizione sono trattati in genere da varie aziende tra le quali:

- operatori di telefonia mobile (carrier), che utilizzano la posizione dei dispositivi per poter permettere la comunicazione attraverso le torri cellulari;

- fornitori di sistemi operativi per dispositivi mobili (Android e iOS), che usufruiscono di queste informazioni a seguito dell'installazione di servizi o di specifiche funzionalità (in caso di richiesta dell'utente) ma anche per fornire all'utente pub-

blicità personalizzata a seconda del luogo in cui si trova, mettendo in evidenza i negozi più vicini;

- fornitori di App e partner, nel momento in cui l'utente installa e attiva una App che presenta della funzionalità basate sulla posizione (esempio avvisi meteo). Questi possono condividere le informazioni di posizione con i loro partner;
- fornitori di servizi di analisi della posizione che utilizzano queste informazioni al fine di ottenere un'analisi sul numero di soggetti che si trovano in un determinato luogo, come, ad esempio, un negozio o un aeroporto.

Possiamo ritenere esenti da tutte queste limitazioni chi opera nell'ambito dei servizi di emergenza (112) o nel campo degli allarmi di emergenza rivolti dalle autorità pubbliche competenti ai cittadini.

Inoltre la capacità di fornire dati di posizione precisi e tempestivi, di associare ad essi elementi di interesse (per esempio foto) e di utilizzare le coordinate di posizione come chiave di ricerca in un database costituiscono le proprietà caratteristiche di un fiorente mercato del software per applicazioni che funzionano su piattaforme mobili.
A livello sociale, la geolocalizzazione comporta, come qualsiasi tecnologia, numerosi vantaggi ma anche imponenti rischi. Infatti viene spesso in aiuto alle Forze dell'ordine, ma, d'altro canto viene sfruttata in azioni criminali o viene utilizzata impropriamente per il controllo ingiustificato di attività individuali o aziendali.

2.2.7 Interesse pubblico nella geolocalizzazione

Si possono riportare molti articoli che citano l'interesse del mondo nel tema affrontato, come in [36]. Come citato in [29], dai social network si riesce a dedurre le posizioni delle

case e quando gli abitanti trascorrono periodi di ferie fuori casa per poter rubare all'interno dell'abitazione. In [30] si evidenzia come un'applicazione, che permette di gestire le attività motorie dei suoi utenti, ha permesso di localizzare le aree militari in America. Premettendo la definizione di "trackmageddon" come le vulnerabilità che trattano informazioni come "coordinate GPS, numeri di telefono, modello di dispositivo e tipo, numeri IMEI, nomi assegnati personalizzato, foto e registrazioni audio caricati dagli dispositivi di tracciamento posizione" [35] si può citare l'articolo [31] che pubblica le principali violazioni. Tra queste si ricorda anche l'app "Tinder", applicazione di incontri sentimentali, che condivideva più dati sulla posizione degli utenti di quanto si rendessero conto gli utilizzatori stessi [32]. Si ricorda anche che Google ha previsto un meccanismo per salvaguardare la riservatezza della posizione degli utenti, utilizzando il dispositivo in modalità incognito [33].

2.2.8 Vantaggi commerciali della geolocalizzazione

Anche a livello commerciale, la geolocalizzazione offre numerosi vantaggi che possono essere sfruttati, da ogni tipo di impresa, in un'ampia varietà di settori, tra i quali la vendita al dettaglio, i servizi finanziari, le assicurazioni, i trasporti ed i servizi pubblici. I veicoli a guida autonoma e la consegna patrimoniale sono due esempi dell'uso aziendale dei dati di posizione, cosi come la pubblicità mirata all'utente e l'avvento della realtà aumentata. Inoltre, la combinazione della geolocalizzazione con le tecnologie mobili permette al cliente di avere un'esperienza migliore e consente alle aziende di sfruttare ulteriormente i dati di posizione integrandoli con le tecnologie dei social media.

2.2.9 Tipologia di violazione della privacy

Dedurre i punti di interesse degli utenti (POIs), la semantica dei POI e dei comportamenti di mobilità permette di capire cosa rappresentano i POI stessi e di conseguenza consente di trarre informazioni [39]. La violazione della privacy di dati geospaziali provoca anche una violazione nelle relazioni sociali, permettendo la deduzione di amicizie e la classificazione gli impegni di ogni individuo [40]. Altro aspetto da non sottovalutare è la previsione della mobilità futura: cioè prevedere i luoghi dove gli utenti possono andare, ad esempio sfruttando le informazioni sui check-in [41]. Ultima violazione della privacy su questi dati da ricordare è quella che permette la ri-identificazione degli utenti in base al percorso effettuato, anche per utenti ritenuti VIP [42].

Minacce alla geolocalizzazione

Di seguito [65] le possibili minacce alla privacy nella geolocalizzazione. Nel documento [38] sono argomentate in modo esauriente:

- **Minaccia di monitoraggio**: un avversario con aggiornamenti dell'utente a sua disposizione potrebbe essere in grado di identificare i modelli di mobilità dell'utente stesso (tragitti frequentemente percorsi) e prevedere la sua posizione presente e futura con elevata precisione;

- **Minaccia di identificazione**: un avversario può utilizzare le tracce dell'utente come identificatori per rivelare la sua identità in un set di dati anonimo. Questo accade anche se l'avversario accede alla posizione dell'utente solo sporadicamente, esponendo il pericolo di identificare l'utente, e riuscendo quindi, a dedurre anche informazioni come la posizione della casa e del lavoro;

- **Minaccia di profilazione**: l'avversario potrebbe profilare dati sensibili a società terze;

- **Identificazione dell'utente dalle sue tracce**: combinando comportamenti, preferenze dell'utente e posizione i dati sono utili per proporre determinati prodotti o servizi.

Capitolo 3

Tecniche di privacy applicabili nella geolocalizzazione

Esistono molti approcci che permettono di tutelare la privacy nella geolocalizzazione: alcuni sono progettati per proteggere l'identità dell'utente durante l'invio di query, altri si concentrano in modo molto attento sulla posizione dell'utente ed altri ancora offrono protocolli per offuscare anche le query [37]. Nell'articolo [56], viene nominato un framework che fornisce un modello della privacy delle query in modo specifico. Anche controllare chi accede è una modalità progettata per tutelare gli utenti da richieste indesiderate delle varie applicazioni in base ad eventi specifici, tuttavia non è sempre ammissibile rinunciare al servizio come descritto in [57].

Si introduce innanzitutto la suddivisione citata nell'articolo [27] che prevede l'analisi della privacy legata alla posizione in base al caso d'uso:

1. **Caso d'uso in tempo reale**, gli utenti interrogano un servizio basato sulla localizzazione e si aspettano una risposta immediata. L'interrogazione della posizione può essere:

   (a) privata su dati pubblici quando l'ubicazione dell'utente è privata e gli oggetti di interesse sono pubblici;

   (b) privata su dati privati quando sia l'ubicazione dell'interrogazione sia gli oggetti di interesse sono privati;

2. **Caso d'uso offline**, gli utenti inviano i loro dati ad un servizio basato sulla localizzazione. Questo creerà e pubblicherà un set di dati condivisi per vari motivi. I dati potranno essere utilizzati per fini commerciali e non;

3. **Caso d'uso batch**, gli utenti inviano i loro dati ad un servizio basato sulla localizzazione che pubblicherà un set di dati aggregati.

Una seconda suddivisione è quella prevista dall'articolo [28], in base alle esigenze dell'utente:

1. **Privacy della posizione dell'utente**, gli utenti vogliono tutelare la propria privacy nascondendo le informazioni sulla loro posizione e le informazioni sulle loro richieste;

2. **Richiesta di informazioni sulla privacy dell'utente**, gli utenti sono obbligati o sono d'accordo a rivelare la loro posizione, questo però comporta comunque il fatto che gli utenti vogliano nascondere le loro richieste;

3. **Privacy della traiettoria**: gli utenti non si preoccupano di rivelare poche località, tuttavia, vogliono evitare di collegare tra loro queste posizioni. Questo permetterebbe la ricostruzione di una traiettoria.

Si può definire ora un'altra distinzione tra tecniche di privacy basate:

- sulla posizione, tecnica denominata "Location privacy";

- sull'identità e la posizione dell'utente, tecnica denominata "Hybrid privacy";
- sull'identità dell'utente, tecnica denominata "Identity privacy".

Si mostra un'immagine per spiegare questo.

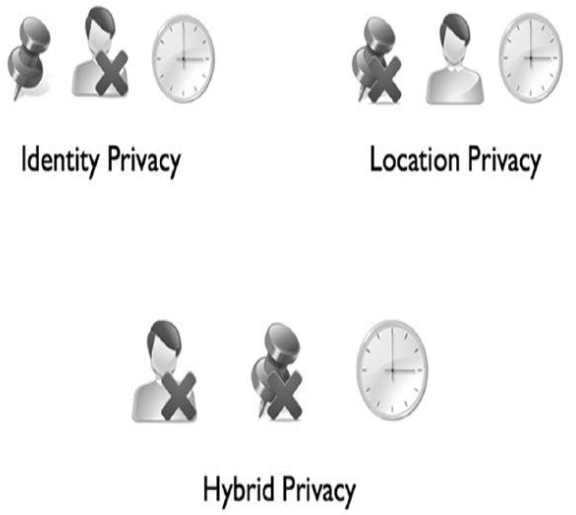


Figura 3.1: Prima distinzione tra tecniche di privacy basate sulla posizione, identità e orario dell'utente. Sorgente [37]

Inoltre, l'orario è fondamentale nel ricostruire i dati sensibili dell'utente, come possiamo vedere nel seguente esempio. Senza l'informazione sull'orario, si potrebbe recepire in modo errato lo schema della posizione dell'utente. L'immagine di esempio [37] permette di percepire l'importanza di ogni informazione legata ad ogni utente.

Ora viene proposta una distinzione fondamentale che consente di fare riferimento a due categorie principali di servizi di localizzazione differenti in base al loro comportamento. Si può denominare la prima come "Location Tracking" o "monitoraggio della posizione". Mentre la seconda si può denominare "POI Search" o "Ricerca POI", acronimo presentato in precedenza, composto da "Point Of Interest" o "punto di interesse".

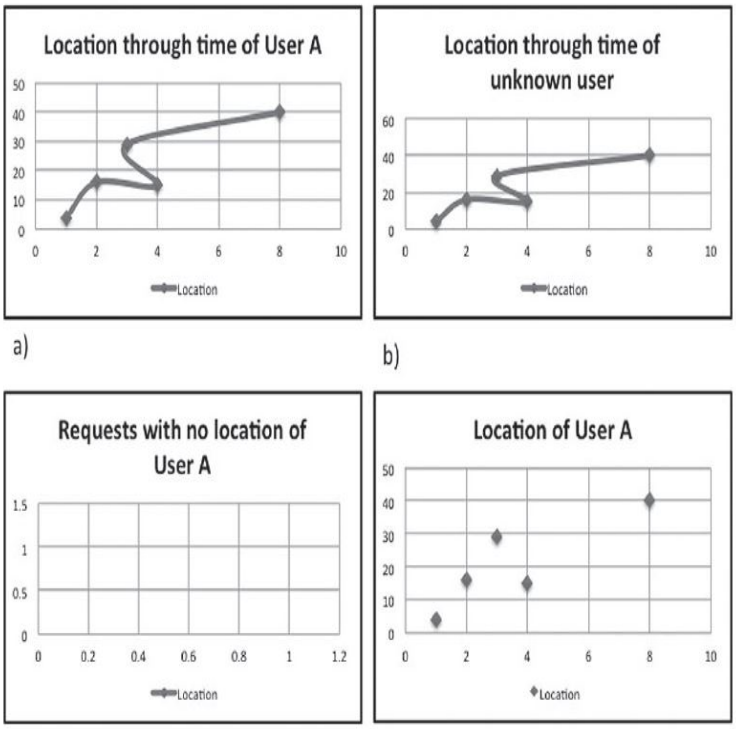


Figura 3.2: Casi d'esempio sull'utilità delle informazioni legate a posizione, orario e identità dell'utente. Sorgente [37]

Inoltre, si ricorda che un punto di interesse è un punto geolocalizzato a cui l'utente è interessato.

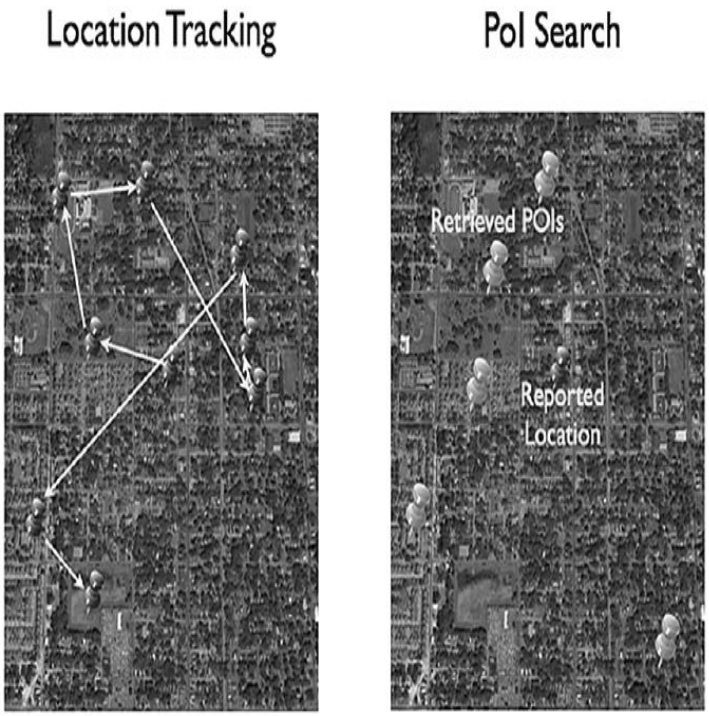


Figura 3.3: Due categorie principali di tipi di servizi di localizzazione in base al loro comportamento. Sorgente [37]

3.1 Location Tracking

I servizi che agiscono in background ascoltando ripetutamente la posizione degli utenti vengono inseriti in questa sezione. Questi consentono agli utenti di tracciare la posizione degli stessi. Identifichiamo i seguenti attori:

- **Monitoring User** (utente monitorato) è un utente monitorato (si tiene traccia delle posizioni);

- **Tracked device** (dispositivo tracciato) è il dispositivo che comunica costantemente la sua posizione al LBS per essere osservato da chi autorizzato;

- **Server** è la parte che mette a disposizione una piattaforma per la comunicazione tra utenti e dispositivi tracciati, memorizza posizioni e non solo;

Questi servizi possono essere, ad esempio, il monitoraggio delle merci, il monitoraggio del traffico, applicazioni per la ricerca di amici, navigazione, geomarketing, geofencing, ecc.
In questa sezione non interessa la posizione precisa dell'utente (ad esempio per chi è agli arresti domiciliari) interessa solo determinare quando il dispositivo tracciato supera un'area delimitata dall'utente di monitoraggio.

3.2 POI Search

I servizi di ricerca di un POI sono progettati per consentire agli utenti di interrogare un sistema che fornisce informazioni sui luoghi vicini in base agli interessi e alla posizione dell'utente. Questi servizi si concentrano sull'elaborazione delle richieste e funzionano in modo reattivo, non funzionano in background. Questo perché richiedono aggiornamenti costanti della posizione dell'utente. In questo servizio possiamo identificare i seguenti attori:

- **Requesting User** (Utente richiedente), è un utente che richiede il tracciamento della posizione e si iscrive per ricevere notifiche da luoghi vicini a lui di suo interesse;

- **Server**, è la parte del servizio che fornisce agli utenti iscritti informazioni rilevanti sulle loro richieste. Il servizio dovrebbe fornire agli utenti luoghi di loro interes-

se nelle vicinanze, informazioni su questi luoghi o anche offerte che potrebbero interessargli;

- **Places of Interest** (luoghi di interesse), rappresentano le informazioni sulla posizione che l'utente vorrebbe ricevere, non necessariamente è il tipo di luogo in cui si trova l'utente.

3.3 Architetture di LPPM per la privacy

Ci sono varie tipologie di LPPM che variano in base all'architettura progettata per la tutela della privacy del dato geospaziale:

- **Architettura locale**. Il dispositivo finale implementa completamente il LPPM prima di inviare i dati al Location Based Service (LBS);

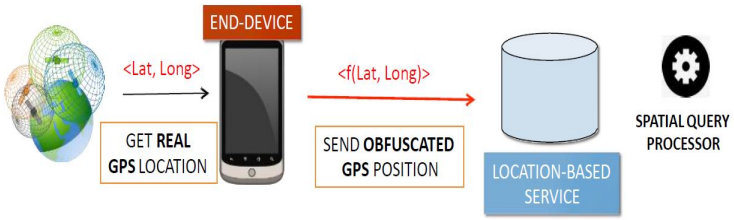


Figura 3.4: Architettura di un LPPM locale. Sorgente [67]

- **Architettura basata su proxy**. LPPM è implementato da un server, che possiamo denominare "proxy", che risulta essere di terze parti ma di fiducia. Il server trusted invia i dati al servizio di localizzazione;

- **Architettura Peer to Peer (P2P)**. Gli utenti sono impegnati in un protocollo di privacy collaborativo prima di inviare i dati stessi al Location Based Service (LBS).

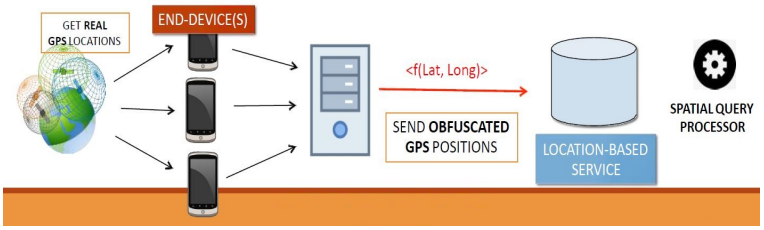


Figura 3.5: Architettura LPPM basata su Proxy. Sorgente [67]

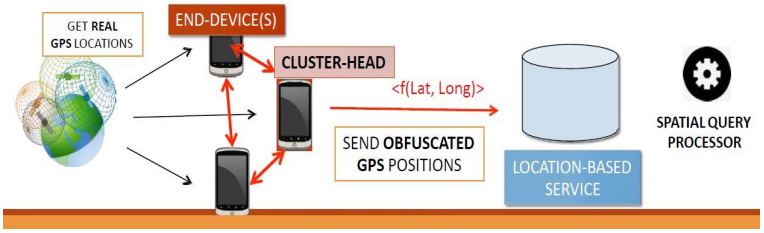


Figura 3.6: Architettura LPPM Peer to Peer. Sorgente [67]

3.4 Tipi di LPPM

Gli LPPM sono progettati per tutelare la privacy della posizione degli utenti utilizzatori del sistema. Il focus dell'analisi è quello di mantenere alto il livello di accuratezza del dispositivo assicurando allo stesso tempo la privacy dell'utente.

3.4.1 Cryptography-Based Mechanisms

La prima tecnica analizzata è quella che utilizzano gli LPPM per il tracciamento basato sulla crittografia. Questi metodi offrono comunicazioni sicure e preservano l'accuratezza delle informazioni sulla posizione; Utilizzano la crittografia simmetrica, in cui ogni utente pone una chiave univoca che viene condivisa con i propri amici e viceversa. Si ottiene così una tecnica sicura che può essere implementata in vari casi d'uso. Lo scambio delle chiavi viene eseguito tramite una comunicazione sicura prima di eseguire i protocolli. Per ogni aggiornamento di un utente viene utilizzata una chiave diversa. Tutto questo è possibile

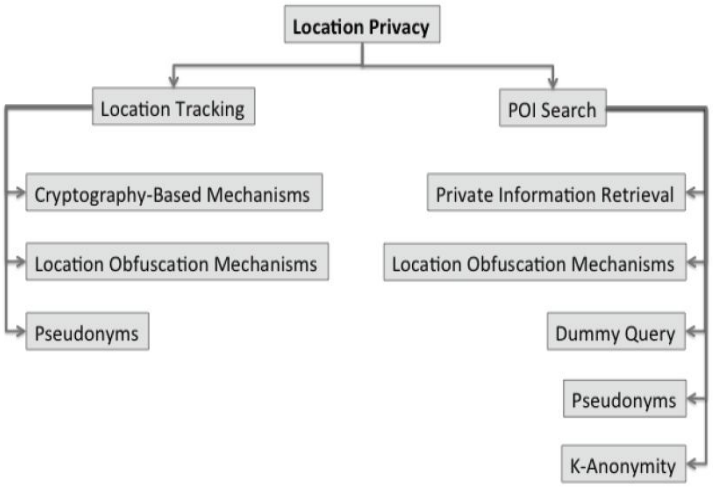


Figura 3.7: Varie tecniche di privacy previste. Sorgente [37]

grazie alla generazione di un key stream basato sulla chiave scambiata inizialmente dagli utilizzatori. Esistono due protocolli in materia: C-Hide e Seek, C-Hide e Hash.

3.4.2 Private Information Retrieval

Il recupero delle informazioni private è una tecnica utilizzata per tutelare la privacy della posizione nelle ricerche dei vicini più vicini. Formalmente, è stato definito per la prima volta nell'articolo [58] come "schemi che consentono a un utente di accedere a k copie replicate di un database (k maggiore o uguale di 2) e recuperare privatamente le informazioni memorizzate nel database. Ciò significa che ogni singolo database non ottiene informazioni sull'identità dell'elemento recuperato dall'utente ". Inizialmente si utilizzavano due database con limitazioni di comunicazione tra di loro e non uno solo. Questo perché si proponeva di fornire informazioni sulla privacy teorica: ciò richiedeva un avversario senza la conoscenza delle informazioni richieste e con risorse informatiche illimitate per gli attacchi. Esistono varie tecniche che implementano la PIR. Questa

tecnica è anche denominata "Differential privacy".
L'obiettivo dunque è quello di proteggere i dati di un individuo durante la pubblicazione di informazioni aggregate sul database. É permesso fare questo aggiungendo rumore controllato all'esito della query, in modo tale che la modifica dei dati di un singolo utente avrà un effetto trascurabile sulla risposta.

3.4.3 Noise-based or Location Obfuscation Mechanisms

L'LPPM basato sul rumore manipola la posizione dell'utente in modo tale che la posizione originale venga cambiata definitivamente. L'elemento fondamentale è comunque quello di fornire prestazioni accettabili, questo grazie al mantenimento della posizione risultante abbastanza vicina da quella reale. In [59] sono stati eseguiti test per dedurre gli indirizzi di casa dai log di tracciamento della posizione degli utenti. I risultati dello studio hanno rilevato che con una semplice tecnica di rumore gaussiano è necessaria una deviazione standard di due chilometri di rumore aggiuntivo al fine di ridurre quasi a zero la quantità di inferenze corrette negli attacchi. Questa tipologia di tecnica può essere implementata sul dispositivo senza interferire con LBS. Sono stati identificati tre termini per l'imperfezione delle informazioni spaziali: "La non accuratezza, l'imprecisione e la vaghezza". La non accuratezza riguarda la mancanza di corrispondenza tra informazione e realtà; l'imprecisione riguarda una mancanza di specificità nell'informazione; la vaghezza riguarda l'esistenza di casi limite nelle informazioni (testo tradotto da [60]). Ogni regione rappresenta una località. In più, una serie di relazioni rappresentano quanto è vicino il soggetto a ciascuna regione. Il processo di offuscamento consiste nel cambiare la relazione in una relazione più vaga, ovvero l'utente specifica di essere vicino alla posizione x.

3.4.4 Dummy Query

La Dummy Query è una tecnica utilizzata nel calcolo del POI che consiste nell'invio di n richieste false unite alla richiesta reale. Questo viene fatto per mascherare la vera posizione dell'utente. La tecnica ha lo svantaggio che prevede un'elaborazione del server nel calcolare n-query aggiuntive causando così costi di comunicazione e overhead dei server. Tuttavia ci sono alcune tecniche sviluppate sulla base di query non fittizie che riescono a diminuire tali costi. Le posizioni fittizie vengono generate in modo da formare tracce ammissibili di un utente normale. Nel primo turno di generazione delle false posizioni si generano casuali, per poi, nelle richieste successive, basarsi su quelle precedenti al fine di costruire n possibili tracce per quel determinato utente.

3.4.5 Pseudonyms

Gli pseudonimi sono un'alternativa che da sola non è sufficiente a fornire la privacy della posizione in un LPPM, questo poiché uno pseudonimo che rimane lo stesso nel tempo porterà alla fine all'identificazione di un utente.
Per questa tecnica, nell'articolo [61], è stata introdotta la denominazione di "zona mista". Il meccanismo previsto richiede un middleware di terze parti di fiducia che fornisce agli utenti alcuni pseudonimi affinché sia garantito che la vera identità degli utenti non venga rivelata al LBS. L'anonimizzazione dell'identità viene fornita cambiando gli pseudonimi nel tempo in zone miste designate.
Definiamo ora cosa si intende con "zone miste": sono zone in cui gli utenti non hanno alcuna applicazione sottoscritta e quindi possono cambiare lo pseudonimo. Gli svantaggi includono il caso in cui nessun utente è disponibile in una zona mista o questa potrebbe essere troppo grande in modo che il LBS possa identificare gli utenti al proprio interno.

3.4.6 K-anonymity and Spatial Cloaking

Questa tecnica rende un utente indistinguibile tra vari utilizzatori. In alcune implementazioni, l'utente è in grado di specificare quanti utenti operano nel sistema. Mentre in molte implementazioni di queste tecniche vengono utilizzate regioni ammantate che forniscono una tutela della privacy garantendo l'anonimato. All'interno di ogni area troviamo k utenti abbastanza simili. Questo non consente agli aggressori di identificare il vero emittente di una richiesta LBS, analizzando le regioni con una maggiore densità di utenti che risultano in aree occultate più piccole. Dal punto di vista degli LBS, questa tecnica può incorrere in un sovraccarico di calcolo, poiché i costi di elaborazione legati all'acquisizione di una regione anziché di un punto sono differenti. Il cloaking introdotto in [62] consiste in un'architettura centralizzata che offusca informazioni anonime sulla posizione. L'algoritmo citato nell'articolo fornisce un K-anonimato attraverso il cloaking spaziale. Il server centralizzato, chiamato anche server trusted, è una terza parte del sistema che riteniamo essere fidata. Questa conosce le posizioni degli utenti LBS in ogni momento e le utilizza per garantire che almeno K utenti siano contenuti nell'area segnalata. Nella fonte numero [64], viene citato il termine "fuga di informazioni". Questo è stato utilizzato per riferirsi alla quantità di informazioni sulla posizione rivelate nel cloaking spaziale per fornire prestazioni migliori.

Un altro metodo di cloaking utilizzato è il cloaking temporale che ritarda la richiesta fino a quando almeno K utenti non entrano nella stessa regione designata. In questo metodo non si utilizza la regione ammantata. Gli autori in [63] presentano un meccanismo sempre di questa categoria che utilizza però la privacy del cloaking nelle reti peer-to-peer. I membri che usano questa tecnica si scambiano informazioni sulla posizione tra di loro al fine di calcolare una regione occultata per essere segnalati al LBS. In questo modo si

riesce ad ottenere la stessa tecnica citata in precedenza, eliminando però la necessità di una terza parte fidata. Gli svantaggi della tecnica sono i costi di comunicazione e l'eventuale assenza di utenti utili per formare una regione occultata. Per di più, importante è ricordare anche il problema della condivisione delle posizioni degli utenti: chiunque all'interno del sistema conosce la posizione degli altri.

3.4.7 Progressive Retrieval

In questa tecnica si eseguono molte richieste per una singola interazione utente. Questo approccio mira a rivelare il minor numero possibile di informazioni sulla posizione per ottenere una maggiore tutela della privacy. Viene definito uno spazio di domanda che indica quanto dovrebbe essere vicino un PoI dalla posizione originale per essere accettato come rilevante dall'utente. In seguito, gli algoritmi utilizzano uno spazio di offerta. In primo luogo questo è solo l'ancora, in secondo luogo avverrà un aumento del raggio della regione rispetto all'ultimo PoI più vicino recuperato a ciascuna iterazione. Tutti gli algoritmi terminano quando lo spazio della domanda è completamente contenuto nello spazio dell'offerta, garantendo che il POI interessi veramente il client, senza rilasciare la posizione reale al server.

3.4.8 Confronto tra le tecniche

Un'implementazione ideale di un meccanismo di tutela della privacy dovrebbe essere sufficientemente generale da essere utilizzato sia per la ricerca sia per il monitoraggio dei punti di interesse. Purtroppo però la specializzazione delle tecniche non rende tutto sempre così semplice. Ricordiamo anche che l'hardware ha un ruolo fondamentale nella scelta di questa tecnica. Un fattore basilare e importante da tenere in considerazione è il

LPPM	Type of Technique	Allows Pol-Search	Allows Tracking	Requires Third Party/ Hardware	Reports Location Info to LBIS	MUR	Special Implementation in the LBIS
[39]	Cryptography	N	Y	N	N	ESD	Y
[36]		N	Y	N	N	ESD	Y
[29]	PIR	Y	N	SC	N	ESD	Y
[31]		Y	N	N	Region	CI/ CO	Y
[47]		Y	N	N	Region	CI	Y
[21]		Y	N	N	Region	CI	Y
[2]	Noise-Based	Y	LA	N	Region	NG	Y
[18]		Y	LA	N	Y	NG	Y
[33]		Y	LA	N	Y	NG	N
[24]		Y	LA	N	Y	NG	N
[48], [49]		Y	LA	N	Y	NG	N
[23]	Dummy Queries	Y	N	N	Y	CI	N
[43]		N	Anonymous	N	Y	CI	Y
[38]		Y	N	N	Y	CI	Y
[4]	Pseudonym	Y	Anonymous	Y	Y	N	N
[15]		Y	Anonymous	N	Y	N	N
[19]	K-Anonymity	Y	N	Y	Region	NG	Y
[7]		Y	N	N	Region	NG	Y
[50]	PR	Y	N	N	Y	NG	N
[32],[35]		Y	N	Density Map	Y	NG	N

Figura 3.8: Comparazione delle tecniche di privacy. Sorgente [37]

seguente: se l'applicazione deve riportare le informazioni sulla posizione, l'ideale sarebbe che la posizione esatta non fosse sempre condivisa ma solo quando necessario. Per le applicazioni che necessitano di una condivisione continua della posizione, le tecniche crittografiche e le tecniche che segnalano regioni sono consigliate, anche se si ottiene una posizione leggermente alterata.

Parte II

Parte progettuale

Capitolo 4

Progettazione

4.1 Obiettivi progettuali

L'obiettivo di questo progetto è quello di fornire una variante di MQTT in grado di gestire l'invio di dati geospaziali. Il sistema prevede l'adozione di uno specifico formato del Payload che gestisce il dato contenente informazioni di tipo geospaziale. In questo modo non è necessaria una ristrutturazione dell'header del pacchetto inviato da MQTT, agevolandone l'implementazione e l'utilizzo. Su questo sistema è stata gestita la riservatezza di un dato geospaziale all'interno del protocollo MQTT, obiettivo ottenuto implementando alcune tecniche citate nello stato dell'arte. Dopo la creazione di luoghi di interesse all'interno della città di Bologna utilizzando la tecnica dei geofence, si rende possibile un meccanismo di sottoscrizione a tali aree, garantendo la possibilità di essere avvertiti una volta entrati in essi. É previsto un meccanismo di entrata, uscita e permanenza per ogni geofence realizzato.

4.2 Recensire città per densità

Nelle tecniche esposte, viene proposta una gestione della privacy correlata ad un'analisi obbligatoria del caso di studio. Il sistema proposto richiede un'analisi della densità degli scenari. Questo permette una gestione ottimizzata delle posizioni correlate ai luoghi di analisi. Prendendo come esempio la città di Bologna, si deve suddividere in vari "settori" in base alla densità di ogni area cittadina. Queste informazioni possono essere raccolte su internet [78] ma non solo. É quindi richiesta un'analisi preliminare di ogni città in cui si intende offrire il servizio. Questo non vale solo per l'Italia, ma per qualsiasi area di analisi nel mondo in cui si vuole proporre il progetto. Ricordiamo l'operato di Google nel fornire il servizio di "Google Maps Street View" [79], che dimostra come sia possibile realizzare un'analisi dettagliata delle aree geografiche utilizzando supporti tecnologici.

4.3 Gestione dei geofence

"Il geofencing è un servizio basato sulla posizione in cui un'app o un altro software utilizza dati GPS, RFID, Wi-Fi o cellulari per attivare un'azione pre-programmata quando un dispositivo mobile o un tag RFID entra o esce da un confine virtuale impostato attorno a un'area geografica, nota come geofence." [99]. Questa tecnica è stata utilizzata per la gestione dei punti di interesse nel progetto. Un geofence può anche essere configurato dagli utenti finali, come in [98] viene citato, e su questa affermazione è stato previsto nel progetto un meccanismo di personalizzazione. L'utente creatore di quest'area di interesse può realizzarla con la forma e con la dimensione che preferisce (un qualsiasi poligono). Inoltre, l'utente personalizza l'accesso all'interno di un geofence di suo interesse andando a specificare se vuole essere notificato (o solo abbonato) e dettagliando anche il messaggio

che vuole ricevere una volta entrato. Nel progetto viene gestita anche la permanenza di un utente all'interno del geofence, fornendo messaggi personalizzati. Inoltre, una volta usciti, l'utente riceve una notifica affinché sia correttamente informato. Importante è ricordare come tutto questo dipende sempre dal tempo di aggiornamento della posizione: se questo avviene sporadicamente, il progetto perde della sua speciale utilità.

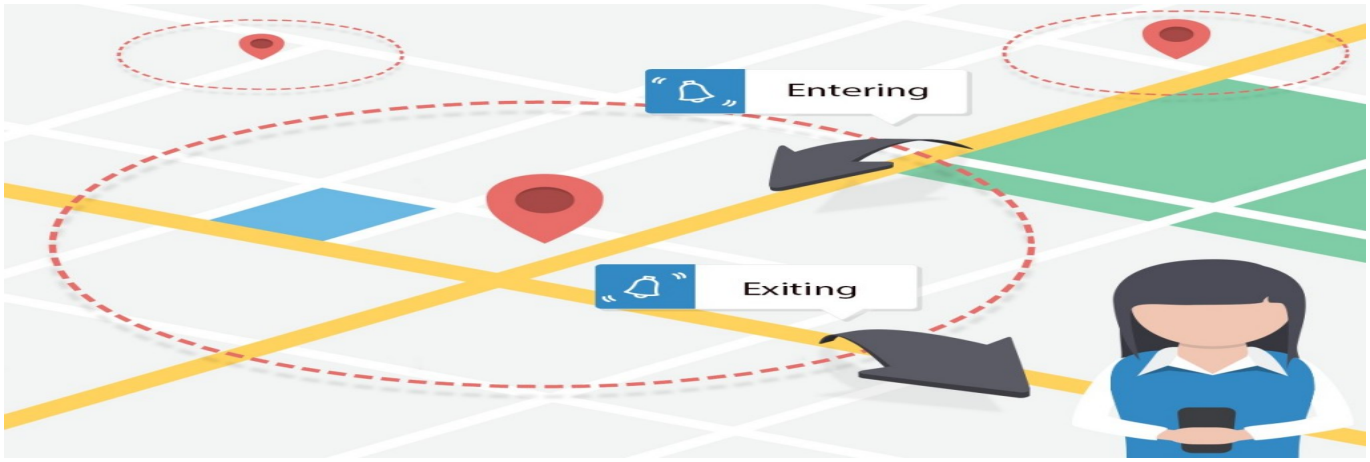


Figura 4.1: Immagine d'esempio della spiegazione di un geofence. Sorgente[99]

4.4 Pacchetto di MQTT

Il pacchetto inviato da MQTT include la geolocalizzazione del payload. Quindi l'unico requisito per utilizzare questo sistema, è quello di conoscere in dettaglio la composizione del pacchetto analizzato. Il formato utilizzato è il GeoJSON. Nel capitolo seguente verrà mostrato in dettaglio come questo pacchetto deve essere implementato.

4.4.1 Formati GeoJSON

GeoJSON è un formato utilizzato per la codifica di una varietà di strutture di dati geografici [93]. Si basa sul formato JSON [94]: JavaScript Object Notation. Questo è

uno standard aperto che fornisce un formato dei dati di interscambio per trasmettere oggetti di dati costituiti da coppie attributo-valore e tipi di dati di matrice (o qualsiasi altro valore serializzabile) [95]. GeoJSON supporta i seguenti tipi di geometria: Punto, LineString, Polygon, MultiPoint, MultiLineString, MultiPolygon e GeometryCollection [96]. Nel progetto è stato utilizzato il formato "poligono", che permette la creazione di geofence con confini particolari e personalizzati.

4.5 Tecniche di Privacy utilizzate

Il sistema realizzato si basa su alcune principali tecniche di privacy:

1. Perturbazione dopo troncamento;
2. Dummy updates con percolazione;
3. Assegnazione di pseudonimi random ad ogni coppia di coordinate (de-anonimizzare la richiesta);
4. Cloaking spaziale unendo richieste di vari utenti;

Le tecniche adottate sono state scelte in base al guadagno che potevano portare nell'implementazione del progetto. Le prime tre sono state implementate all'interno del client, mentre l'ultima è stata implementata all'interno del server.

4.5.1 Perturbazione dopo troncamento

In questa tecnica, vengono realizzate quattro perturbazioni applicabili su cinque settori. La suddivisione in cinque settori è stata realizzata grazie ad un'analisi preliminare della densità e delle aree di Bologna. Dopo l'analisi della città, si è deciso di adottare le

seguenti funzioni nel calcolo che hanno permesso la suddivisione in settori. Per quanto riguarda le coordinate che variano sui meridiani si è utilizzata la seguente funzione: $6x^x$. Mentre per le coordinate che seguono i paralleli si è utilizzata la funzione $2x^x$.

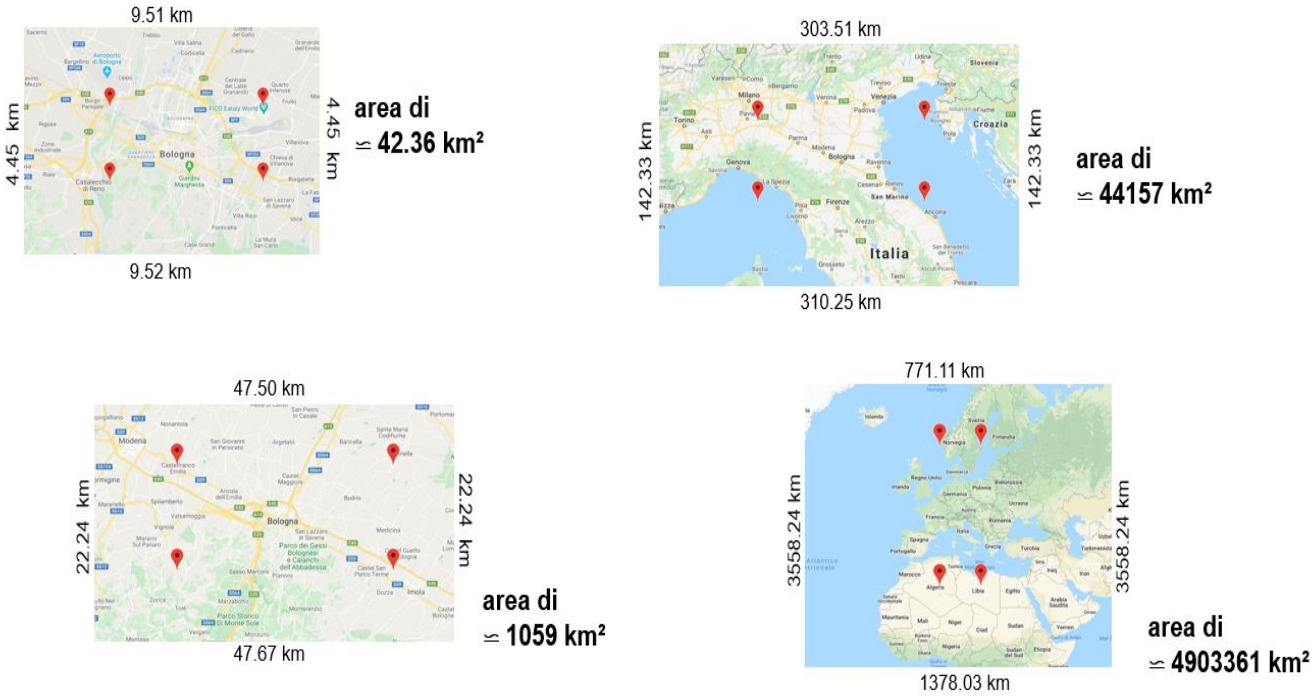


Figura 4.2: Suddivisione dell'area oggetto di interesse in settori.

Nell'immagine riportata viene mostrata la suddivisione della città di interesse nei settori presi in considerazione. Vengono quindi creati cinque settori:

1. il primo ha un'area di 42.36km^2 circa, e comprende il centro di Bologna. I geofence vengono creati all'interno di questa area secondo l'obiettivo perseguito;

2. il secondo ha un'area di 1059km^2 circa, e comprende il comprensorio della città;

3. il terzo ha un'area all'incirca di 44157km^2 e comprende le richieste che provengono dal nord Italia;

4. il quarto ha un'area di $4903361km^2$ e suddivide le richieste che provengono da un intorno dell'Italia e quelle provenienti dall'esterno;

5. il quinto viene spesso anche chiamato "settore zero" e comprende tutto il resto del mondo al di fuori del quarto settore.

Questa suddivisione permette di realizzare delle perturbazioni delle coordinate in base al settore in cui l'utente si geolocalizza. Questo deriva dalla densità dei luoghi: quando, per esempio, un luogo ha una densità elevata, geolocalizzarlo all'interno di essa significa avere una precisione elevata, che si traduce mediamente in pochi metri. D'altra parte, analizzare una posizione proveniente da un luogo con poca densità di popolazione, consente di geolocalizzarlo sicuramente all'interno dell'area se la precisione è elevata. É stato previsto, quindi, un meccanismo di troncamento delle coordinate, come viene mostrato nell'immagine seguente.

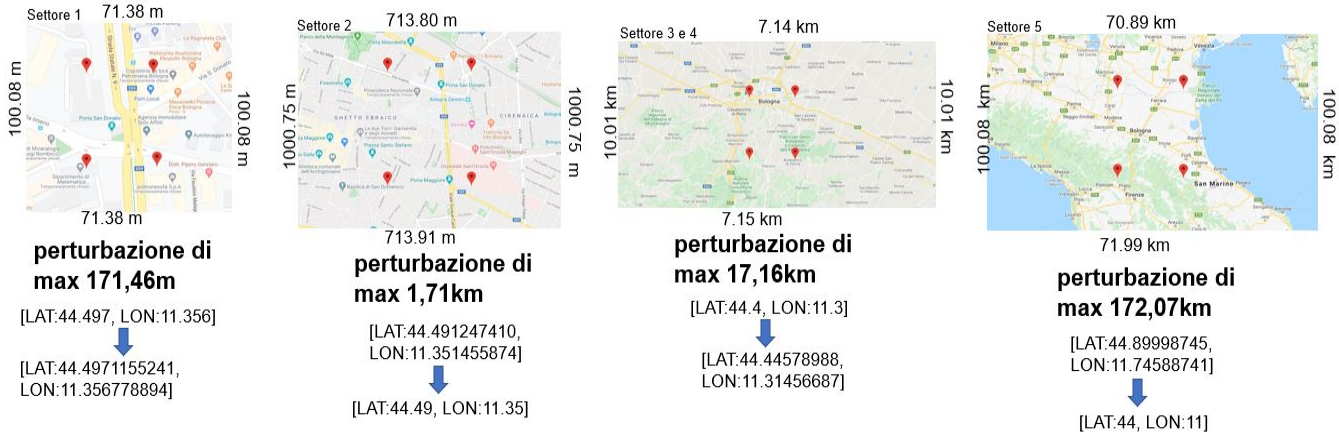


Figura 4.3: Comparazione delle aree perturbate per ogni settore.

Dopo aver troncato le coordinate, per rendere la posizione uguale alle altre, si concatenano la latitudine e la longitudine con una stringa di numeri casuali che la rendono lunga come prima, consentendo comunque la geolocalizzazione di una posizione precisa,

ma non di quella reale. Questo però non influisce sui risultati ottenuti, in quanto le aree di interesse dei geofence analizzati si trovano nel primo settore, quindi nell'area più interna.

4.5.2 Dummy updates con percolazione

In questa tecnica viene prevista la creazione di altre quattro posizioni fittizie da inviare insieme a quella reale. É giusto ricordare che nel sistema realizzato, arrivata a questo step quella reale ha già subito il meccanismo di perturbazione, rendendo di fatto ancora più complicata una diretta correlazione dell'utente alla posizione geospaziale. Le altre quattro posizioni vengono generate all'interno dello stesso settore in cui l'utente si trova. É importante ora ricordare la generazione delle quattro posizioni fittizie nel tempo. Quando l'utente aggiorna la posizione allo step $n + 1$, la differenza della latitudine e della longitudine viene applicata a tutte le altre posizioni fittizie. In questo modo viene gestita una percolazione a macchia d'olio dell'utente. In base all'utilizzo della tecnica di percolazione appena esposta, si può prevedere il troncamento di questo aggiornamento delle posizioni in base all'obiettivo della creazione dei geofence. Dipende dall'utilizzo reale dell'app, che nel progetto è stata mantenuta per cinque step ($n = 5$), perché si è gestito un caso esempio generale.

4.5.3 Assegnazione di pseudonimi random ad ogni coppia di coordinate

Ad ogni richiesta viene associato uno pseudonimo random: questo permette la de-anonimizzazione della richiesta, perdendo qualsiasi riferimento diretto all'utente. "Lo pseudonimo, anche noto come nome in codice, è un'informazione associata all'identifica-

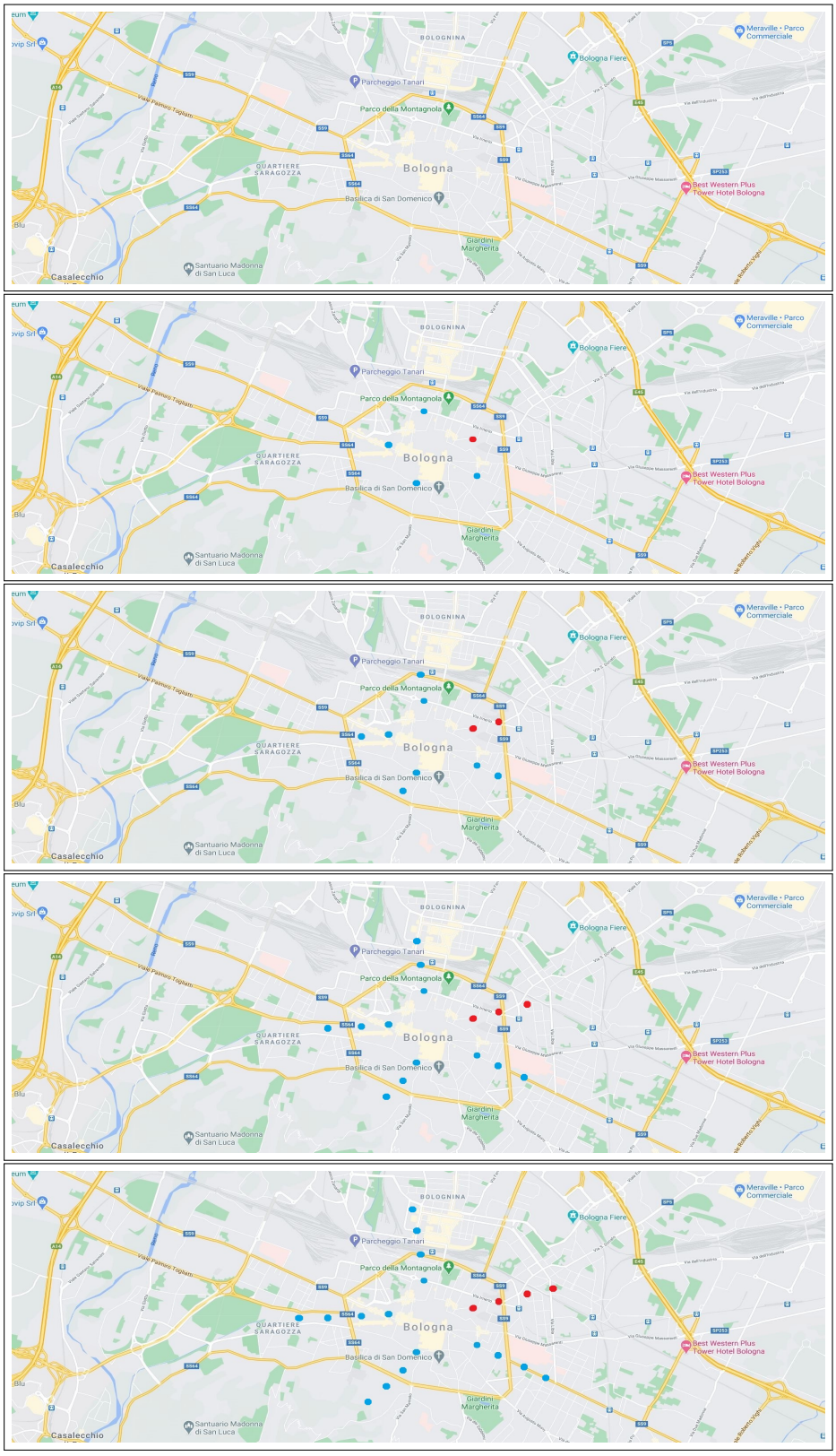


Figura 4.4: Creazione delle posizioni "Dummy" e gestione di esse a "macchia d'olio".

tivo di un individuo o ad altri tipi di dati personali (come i dati relativi all'ubicazione)" [80]. Nel progetto è prevista l'associazione del dato ad uno pseudonimo casuale che non rende possibile rintracciare un individuo da esso.

4.5.4 Cloaking spaziale unendo richieste di vari utenti

L'ultima tecnica considerata si basa sul cloaking spaziale ed è riferita all'unione di varie geolocalizzazioni di vari utenti. In base al settore preso in analisi, si uniscono più posizioni di utenti, permettendo così l'invio solo di alcune di esse. Per tutelare la precisione del sistema, questo viene fatto nel settore 3, 4 e 0. Si attende la ricezione di posizioni da parte di tre utenti (quindi 15 posizioni), per poi estrarre in maniera random cinque tra queste. Queste cinque estratte vengono poi inviate contenendo tutti gli pseudonimi assegnati, così da non aver "smarrito" nessuna richiesta da parte degli utenti. Se trascorrono due minuti senza ricevere tre richieste, queste vengono inviate senza questo tipo di tecnica, per non rallentare troppo il sistema realizzato. Questa tecnica è stata adottata dopo la realizzazione di test appositi che hanno evidenziato queste esigenze nel sistema (si veda capitolo sui test).

4.6 Framework

Sono stati utilizzati vari framework e siti internet per la progettazione dell'elaborato. In [81] vengono esposte tutte le funzionalità di sistema di Google Maps, sistema che ha permesso di effettuare tutti i test e la creazione dei settori, aree, e tecniche appena esposte. In [82], [83] e [84] è possibile semplificare il passaggio da una coordinata GPS al corrispettivo nome del luogo preso in considerazione, utile per la realizzazione di geofence e localizzazione di possibili posizioni dell'utente.

4.7 Architettura

Il progetto realizzato è composto da varie componenti che permettono di realizzare servizi che rispondono agli obiettivi prefissati nel progetto: nel capitolo seguente verrà introdotta l'implementazione di ogni singola parte di questo. I servizi progettati in blu permettono

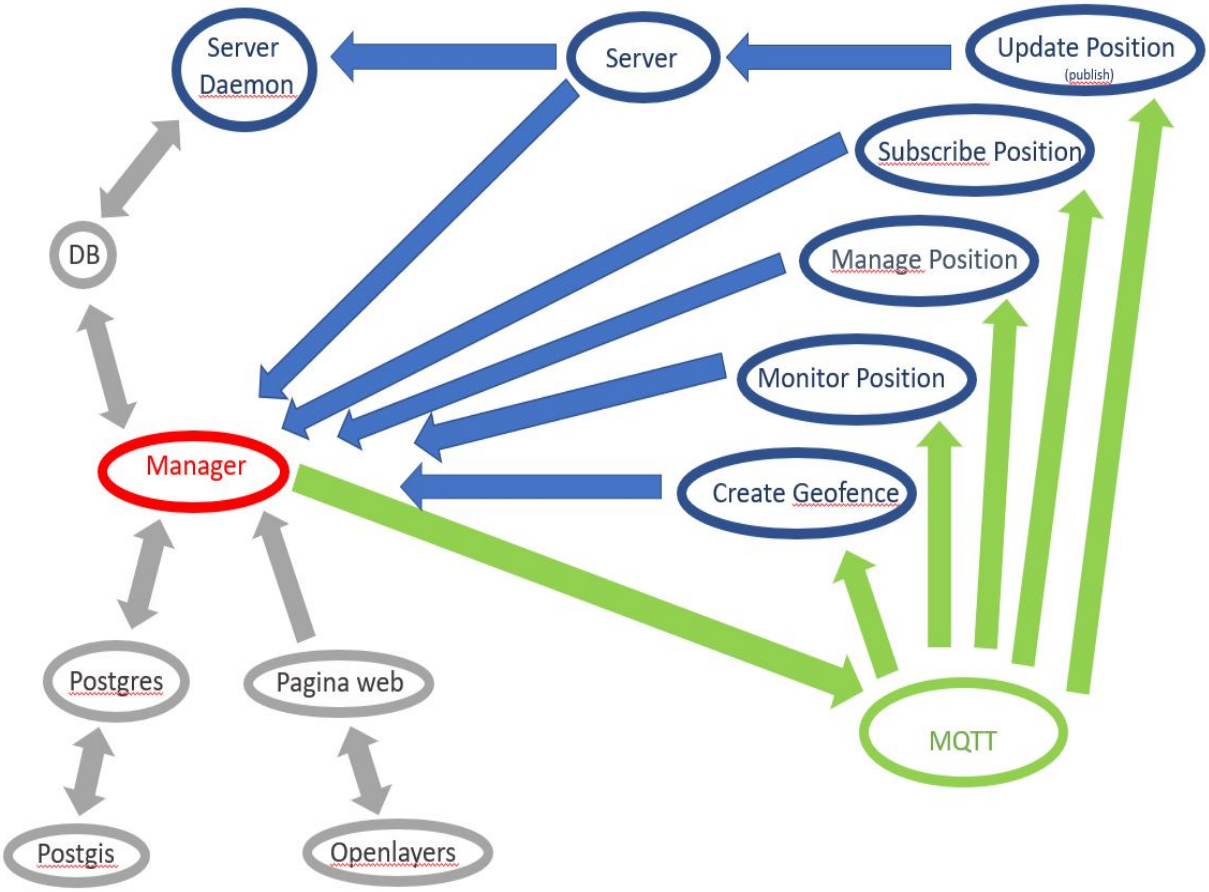


Figura 4.5: Architettura del progetto

di inviare messaggi al Manager. Nel caso del "Manage Position", questo permette di ricevere la lista contenenti tutti i geofence disponibili per un utente. "Subscribe Position" comunica al Manager l'intenzione di abbonarsi ad un servizio, mentre "Update Position" consente di aggiornare la propria posizione. "Monitor Position" consente di restare in

ascolto del traffico del canale, e, ovviamente, "Create Geofence" permette la creazione di un geofence. Il Manager gestisce la comunicazione tra queste componenti e il broker MQTT. Di fatto, anche il manager assume il ruolo di client nei confronti del broker, inoltra i messaggi da inviare per tutti i client. Per ogni utente dedichiamo un solo canale MQTT, da cui il nostro utente riceverà poi comunicazione diretta dal broker. Quindi, le frecce blu regolano una comunicazione che avviene in socket TCP, mentre quelle in verde regolano una comunicazione che avviene nel protocollo MQTT. Infine, in grigio, descriviamo tutti i componenti con cui il manager si interfaccia: database e pagine web. Per ognuno di questi servizi ne verrà descritta l'implementazione in dettaglio nel capitolo seguente.

4.8 Dashboard Web

La Dashboard web progettata permette di gestire i dati ricevuti dal manager. Consente di interpretare le coordinate GPS visualizzandole direttamente su una pagina web. La dashboard ha una home e una pagina denominata "CoordinateGPS". All'interno di CoordinateGPS si può visualizzare la cronologia delle coordinate ricevute e, soprattutto, le ultime ricevute dal sistema. Infine, si può scegliere di visualizzare la mappa in due modalità grafiche differenti.

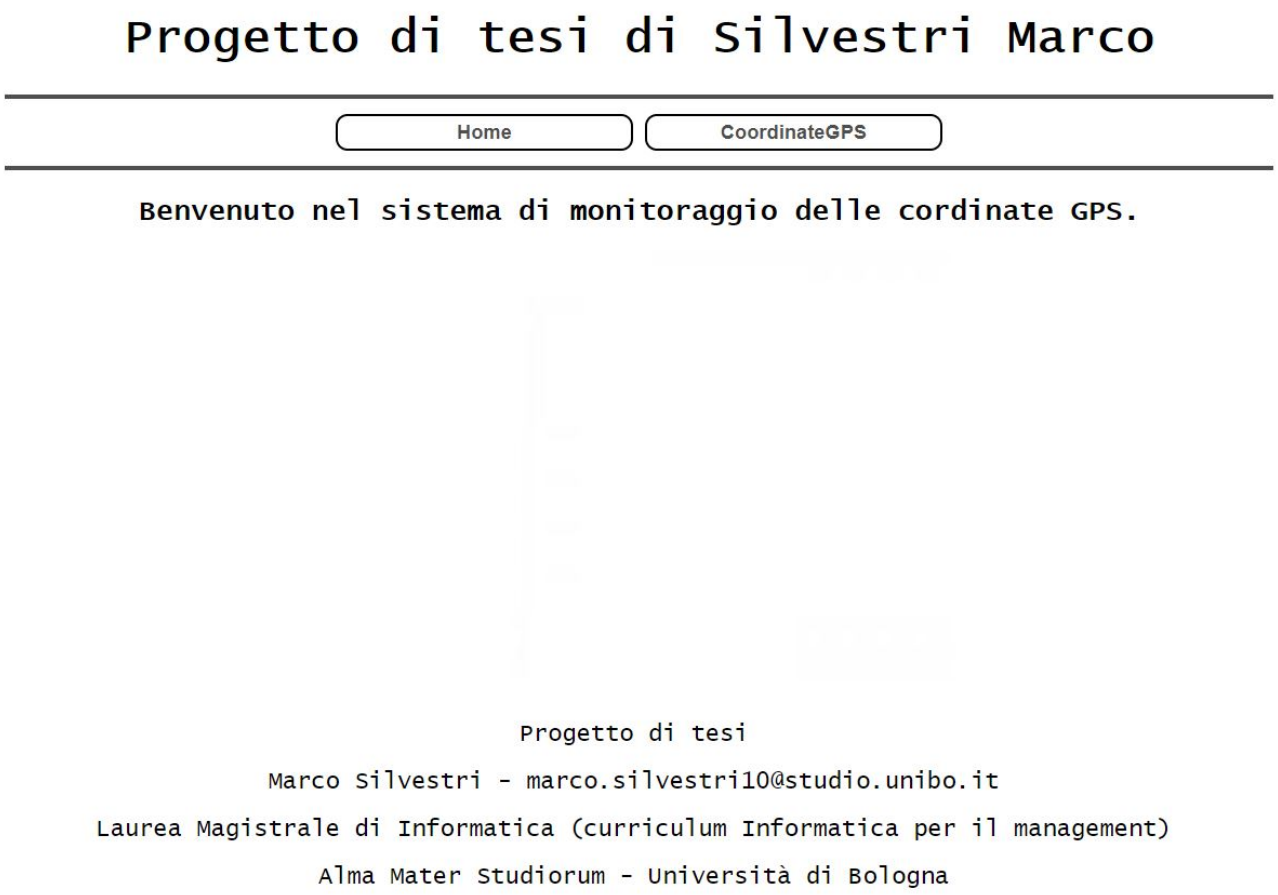


Figura 4.6: Dashboard Web realizzata, pagina Home.

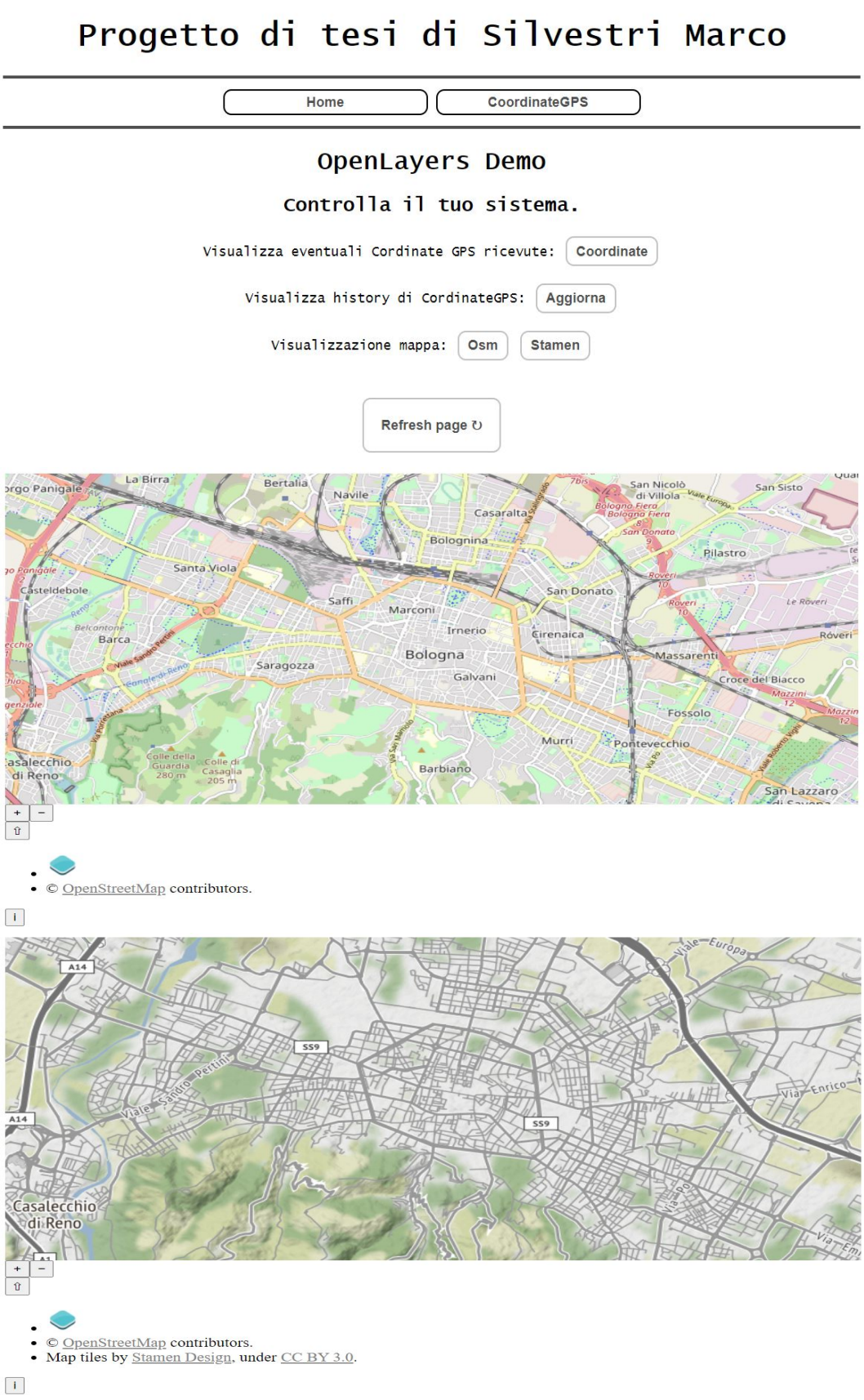


Figura 4.7: Dashboard Web realizzata, pagina Coordinate GPS.

Capitolo 5

Implementazione

5.1 Tecnologie utilizzate

Per la realizzazione di questo progetto di tesi sono state utilizzate diverse tipologie di tecnologie.

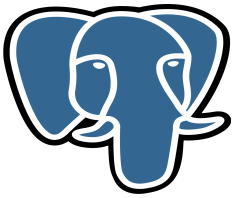
5.1.1 Node e Javascript

Per la realizzazione del server LBS è stato utilizzato **Node.js** e **Javascript** poichè i numerosi moduli di Node permettono un'integrazione rapida ed agile tra tecnologie diverse. Javascript è stato utilizzato anche per la creazione e "customizzazione" della dashboard web del LBS in cui vengono forniti in output le posizioni ricevute dall'utente [68]. Importante è ricordare che "JavaScript è un linguaggio di programmazione orientato agli oggetti e agli eventi, comunemente utilizzato nella programmazione Web lato client (esteso poi anche al lato server) per la creazione, in siti web e applicazioni web, di effetti dinamici interattivi tramite funzioni di script invocate da eventi innescati a loro volta in vari modi dall'utente sulla pagina web in uso." [75]. In [74] si può trovare anche

un test editor online.

5.1.2 PostgreSQL e Postgis

É stata gestita la persistenza dei dati tramite l'utilizzo del database relazionale **PostgreSQL** e la sua estensione spaziale **Postgis** [70]. Questa estensione introduce un tipo di dato spaziale non nativo di Postgres con cui si è riuscito a creare, interrogare e manipolare il database utilizzato. L'accesso a PostgreSQL è stato fatto mediante il client database grafico **pgAdmin III**: questo è il client ufficiale per PostgreSQL, integra il linguaggio SQL per la manipolazione dei dati [69].



5.1.3 OpenLayers e CSS

Per la realizzazione della dashboard web del LBS è stata utilizzata la libreria **OpenLayers** e il linguaggio **CSS**. OpenLayers è una libreria di JavaScript che permette la visualizzazione di mappe interattive nei browser [71].

5.1.4 Handlebars e HTML

Handlebars è un linguaggio caratterizzato da template che genera formati html e permette di mantenere vista e codice separati. Nel progetto è stato utilizzato per gestire la parte grafica della dashboard web [72].

5.1.5 Node-Postgres

Node-Postgres è una collezione di moduli di Node.js che permette di interagire con i database PostgreSQL. Tramite questo tool è stato possibile interfacciare il gestore con i database e calcolare i "Points of Interest" tramite query spaziale.

5.1.6 Google Maps

L'applicazione **GoogleMaps** è stata utilizzata come strumento ausiliario allo sviluppo del nostro progetto. In particolare, GoogleMaps è stato utilizzato nella fase di valutazione (si veda l'ultimo capitolo).

5.2 Framework utilizzati nell'implementazione

Sono state visionate molte fonti per l'implementazione del sistema. Oltre ai siti appena visionati, ricordiamo alcuni di questi che hanno permesso di risolvere punti cruciali del sistema. Molto utili sono stati i siti [85], [87] e [88], che permettono di capire e interpretare il funzionamento del protocollo MQTT. In [86] viene esposto molto dettagliatamente come si realizza la comunicazione a livello TCP in Node. Fondamentale è stato il sito [89] che ha consentito una gestione ottimale dei geofence realizzati: viene esposto soprattutto anche il formato utilizzato, il "geojson". Questo viene fatto in maniera molto più dettagliata in [90]. Inoltre per la gestione dei database presi in oggetto sono state molto utili le seguenti fonti: [91] e [92].

5.3 Servizi realizzati

Nell'implementazione del progetto sono stati realizzati vari servizi che elenchiamo di seguito.

5.3.1 Servizio Aggiorna Posizione

É stato implementato un servizio denominato "Aggiorna Posizione" che permette l'aggiornamento della posizione di un utente utilizzatore del sistema. Questo servizio consente di inviare la posizione dell'utente per poter aggiornare il sistema progettato. Permette anche al sistema di gestire l'ingresso, la permanenza e l'uscita da un geofence. L'utente stesso decide se essere notificato all'ingresso di un geofence, e definisce il messaggio da ricevere al momento dell'ingresso nello stesso.
All'interno di questo servizio avviene la prima gestione della riservatezza del dato perso-

nale. Il dato viene contestualizzato all'interno di un settore dell'area analizzata. Nell'esempio preso in considerazione, si analizza l'area di Bologna, per gestire la creazione dei geofence all'interno dell'area della stessa. Il servizio tronca le coordinate GPS in base al settore, andando a realizzare un'area di "sicurezza" intorno alla posizione dell'utente. Viene applicata una funzione random che permette la generazione di un punto qualsiasi all'interno di questa area presa in considerazione. In seguito, vengono create altre quattro posizioni "fittizie" all'interno del settore in cui la posizione originale si trova. Quindi, se la posizione GPS è all'interno del settore, per esempio il primo, dopo aver perturbato e offuscato la posizione originaria, avviene la creazione di altre quattro posizioni all'interno del primo settore. Queste vengono memorizzate in un database, consentendo così la realizzazione della tecnica di percolazione: allo step $n + 1$, le altre quattro verranno generate nuovamente in base alla differenza di quella reale dallo step n e quella allo step $n + 1$. Vengono inoltre inviati cinque username fittizi per ogni posizione, l'utente mantiene memorizzato quale corrisponde alla sua posizione. In questo modo, quando si ricevono aggiornamenti, si sa interpretare la risposta di suo interesse.

5.3.2 Servizio Creazione Geofence

Il servizio di creazione dei geofence consente la progettazione e la realizzazione di geofence attraverso richiesta esplicita al manager. La richiesta avviene grazie all'invio di una socket TCP, formattando correttamente il pacchetto di gestione di MQTT. Il geofence può essere di varie dimensioni, per questo il sistema accetta poligoni come input.

5.3.3 Servizio Gestire Posizione

Il servizio di gestione della posizione permette di ricevere la lista contenenti tutti i geofence disponibili per un utente. Questo consente a qualsiasi utente di visualizzare tutti i geofence che potrebbero interessargli, affinché gli sia consentito in qualsiasi momento di abbonarsi ad un servizio per migliorare la percezione con il sistema.

5.3.4 Servizio Monitorare Posizione

Per monitorare la posizione è stato realizzato questo servizio, che permette agli utenti di ascoltare il traffico del canale. É corretto quindi affermare l'importanza cruciale di questo servizio: permette di ascoltare il traffico del canale, soddisfacendo il requisito base del protocollo MQTT, ovvero l'ascolto di eventuali pacchetti da parte del broker.

5.3.5 Server Privacy

Il server per la Privacy ha un compito solo: gestisce il cloaking spaziale all'interno dell'algoritmo. Quando le cinque posizioni arrivano al server in questione, la maggior parte delle tecniche privacy pensate sono già state applicate. Rimane solo la necessità di unire più posizioni tra loro per creare un ulteriore filtro privacy nel progetto. Questo servizio permette di salvare le coordinate ricevute in un database che serviranno ad un altro servizio che lavora in background denominato "Server Privacy Daemon". Questo inserisce nel database l'orario di arrivo delle posizioni e il settore in cui queste si geolocalizzano. Si analizza in dettaglio di seguito questo servizio.

Server Privacy Daemon

Per realizzare questa tecnica, è stato implementato nel server un demone, cioè "un programma eseguito in background" [77]. All'interno del demone è stato gestito il tempo: ogni due minuti tutte le posizioni in attesa vengono inviate. Se durante questi due minuti di attesa vengono salvate nel database altre dieci posizioni presenti nello stesso settore, queste vengono inviate "unendo" gli interessi. Il servizio quindi invia casualmente alcune delle posizioni ricevute assegnando però tutti gli username in ingresso. Permettendo così un totale offuscamento della posizione, una gestione ottimizzata degli username stessi, offrendo comunque un metodo di offuscamento che può essere utilizzato solo per le posizioni ricevute dal settore 3, 4 e 0 (cioè i settori più esterni dell'area di interesse). Non è stato implementato nel settore 1 e 2 per la seguente semplice motivazione: questo tipo di gestione effettuata in questi settori avrebbe portato una totale confusione del dato, che è già perturbato in un'area di dimensioni notevoli.
In dettaglio, il metodo prevede di prendere, in maniera random, posizioni nei settori inviandone cinque al broker MQTT con tutti gli username salvati.

5.3.6 Servizio sottoscrizione posizione

Il servizio "subscribe" della posizione fornisce la possibilità agli utenti di sottoscrivere un servizio, rendendosi utenti "abbonati" dello stesso. Questo avviene tramite la comunicazione al manager della volontà di abbonarsi. Qualsiasi utente può interessarsi alla lista completa dei geofence disponibili utilizzando il servizio prima descritto. Come già citato in precedenza, nel momento in cui l'utilizzatore si abbona ad un geofence, inserisce anche il testo che vuole visualizzare nel momento in cui entrerà nello stesso.

5.3.7 Manager

L'elemento cruciale del progetto qua esposto è il manager. Questa componente si occupa della gestione di tutto il sistema: comunica con il broker MQTT e con tutti i servizi descritti precedentemente. Si ricorda innanzitutto un compito molto interessante di questo servizio: la gestione della console web che permette la visualizzazione dei dati operanti nel sistema. Come elemento fondamentale da affrontare nell'analisi dettagliata di questo servizio, si trova la gestione del database Postgres di memorizzazione dei geofence. Questo viene messo a disposizione per la soddisfazione di alcune richieste effettuate da parte dei vari servizi prima citati. Un altro database gestito dal manager è quello che prevede il salvataggio delle coordinate GPS per coordinare l'ingresso, la permanenza e l'uscita di un utente da un geofence, e questo è realizzato tramite MySQL in locale. Un altro database realizzato sempre con MySQL è quello utilizzato per la memorizzazione dei vari geofence per cui i vari utenti si abbonano. Infine l'ultimo database utilizzato è quello già citato nella sezione del demone sulla gestione della privacy: il server salva i dati che riceve sul database stesso così da consentire al servizio in background di esaminare la permanenza della richiesta nello stesso e l'eventuale invio al broker MQTT.
Infine, si ricordano le varie tipologie di messaggi che il manager riceve, si riportano quindi le strutture dei payload.

Payload

Si riporta il formato dei payload più interessanti, in quanto, per utilizzare il sistema, si deve uniformare e adattare tutto a questo formato. In tutto il codice inserito nella parte seguente, è stato utilizzato il carattere "X" al posto delle coordinate da inviare, ad esempio "44.497145, 11.356137". Mentre il carattere "Y" è utilizzato per sostituire gli

username. Per ora, la velocità non è stata utilizzata, in futuro si potrebbe però utilizzare per molti fini utili alla gestione del progetto. In "type" si indica il tipo di servizio implementato.

Il primo formato riportato è quello inviato dal servizio che aggiorna la posizione:

```
'{"type": "update",
  "features": [ {
        "type": "Feature", "subscribe" : null,
        "properties": {
               "Username": "Y1, Y2, Y3, Y4, Y5",
               "Speed": "0"
        },
        "geometry": {
               "type": "Polygon",
               "coordinates": "[ [ '+ X +' ] ]"
        }
  } ]
}'
```

Un altro formato è quello che gestisce il demone della Privacy per l'invio delle 5 coordinate GPS (le nominiamo X1, X2, X3, X4 e X5):

```
'{"type": "update",
  "features": [ {
        "type": "Feature", "subscribe" : null,
        "properties": {
```

```
                "Username": "'+ Y[n] + '",
                "Speed": "0"
        },
        "geometry": {
                "type": "Polygon",
                "coordinates": "[ [ ['+ X1 +'],
                        ['+ X2 +'], ['+ X3 +'],
                        ['+ X4 +'], ['+ X5 +'] ] ]"
        }
  } ]
}'
```

Il terzo formato da riportare è quello del servizio utilizzato per sottoscriversi e abbonarsi. L'utente inserisce se vuole essere notificato una volta entrato, e specifica anche il messaggio che vuole visualizzare all'ingresso del geofence di interesse. In subscribe si inserisce l'id del geofence:

```
clientServerTCP.write(
'{"type": "subscribe",
  "features": [ {
        "type": "Feature",
        "subscribe" : "400",
        "properties": {
                "Username": "100",
                "Speed": "0",
                "Avverti" : "1",
```

```
10                "Messaggio" : "Sei entrato nel Geofence!"
11        },
12        "geometry": {
13                "type": "null",
14                "coordinates": "null"
15        }
16   } ]
17 }'
```

Un quarto formato è quello che gestisce la posizione:

```
1  clientServerTCP.write(
2  '{"type": "managePosition",
3    "features": [ {
4         "type": "Feature",
5         "managePosition" : "si",
6         "properties": {
7                 "Username": "100",
8                 "Speed": "100"
9         },
10        "geometry": {
11                "type": "null",
12                "coordinates": "null"
13        }
14   } ]
15 }'
```

E infine quello della creazione del geofence:

```
'{"type": "creoGeofence",
  "features": [ {
        "type": "Feature", "subscribe" : null,
        "properties": {
               "Username": "400",
               "Speed": "100"
        },
        "geometry": {
               "type": "Polygon",
               "coordinates": " X1, X2, ... , Xn"
        }
  } ]
}'
```

Inoltre, si riporta di seguito un esempio del codice implementato.

5.4 Il broker MQTT

Il broker MQTT è Mosquitto [97]. Il messaggio inviato dal broker MQTT contiene nel payload il messaggio elaborato dal manager.

```
clientServerTCP.write(
    '{"type": "update",
     "features": [ {
           "type": "Feature", "subscribe" : null,
           "properties": { "Username": "100, 101, 102, 103, 104", "Speed": "0"},
           "geometry": { "type": "Polygon", "coordinates": "[ [ '+ CoordinateDaInviare +' ] ]" }
     } ]
    }');
```

```
clientServerTCP.write(
    '{"type": "update",
      "features": [ {
            "type": "Feature", "subscribe" : null,
            "properties": { "Username": "'+ usernArray[1] + '", "Speed": "0"},
            "geometry": { "type": "Polygon",
                          "coordinates": "[ [ ['+ dataDaInviareCompleta[0] +'],
                                              ['+ dataDaInviareCompleta[1] +'],
                                              ['+ dataDaInviareCompleta[2] +'],
                                              ['+ dataDaInviareCompleta[3] +'],
                                              ['+ dataDaInviareCompleta[4] +'] ] ]"
                          }
                    } ]
     }');
```

```
clientServerTCP.write(
    '{"type": "subscribe",
     "features": [ {
          "type": "Feature", "subscribe" : "400",
          "properties": { "Username": "100", "Speed": "0", "Avverti" : "1", "Messaggio" : "Sei entrato nel Geofence!"},
          "geometry": { "type": "null", "coordinates": "null" }
     } ]
    }');
```

```
clientServerTCP.write(
    '{"type": "managePosition",
      "features": [ {
            "type": "Feature", "managePosition" : "si",
            "properties": { "Username": "100", "Speed": "100"},
            "geometry": { "type": "null", "coordinates": "null" }
      } ]
    }');
```

```
clientServerTCP.write(
    '{"type": "creoGeofence",
     "features": [ {
           "type": "Feature", "subscribe" : null,
           "properties": { "Username": "400", "Speed": "100"},
           "geometry": { "type": "Polygon", "coordinates": " 11.332998275756836 44.49411760680279,
                                                             11.324329376220703 44.49779098829304,
                                                             11.310596466064453 44.49405638181746,
                                                             11.336603164672852 44.47911556412261,
                                                             11.349048614501951 44.49405638181746,
                                                             11.340208053588865 44.498158313714924,
                                                             11.332998275756836 44.49411760680279 "
                       }
                  } ]
    }');
```

Figura 5.1: Esempio di implementazione dei payload inviati al manager

Figura 5.2: Esempio di messaggio inviato dal broker MQTT

Capitolo 6

Validazione dei risultati

6.1 Metriche di valutazione

La fase di valutazione è stata eseguita dopo aver implementato le tecniche di privacy nel progetto. Prima di introdurre le varie valutazioni, è doveroso elencare le metriche utilizzate:

1. Accuratezza dei vari geofence, cioè in quanti casi il sistema rileva correttamente l'ingresso di un utente nel geofence. Questo è stato analizzato sia dal punto di vista dei geofence sia dal punto di vista dei settori;

2. Errore legato alla distanza dal geofence in base alla dimensione dello stesso. Cioè, quanto è lontano il geofence suggerito dalla posizione ottenuta dopo l'applicazione delle tecniche di privacy;

3. Precisione, con cui si intende il valore della distanza tra la posizione reale e quella simulata.

6.2 Geofence

I geofence presi in considerazione hanno una dimensione di 500 metri, 1 kilometro e 2 kilometri. Vista la dimensione dell'area metropolitana di Bologna, oggetto di test nel progetto, queste tre sembravano dimensioni ragionevoli per possibili casi di utilizzo in futuro del progetto.

6.3 Tecniche di privacy valutate

É corretto inizialmente analizzare le tecniche utilizzate per capire come sono stati eseguiti i test. Si ricordano ora brevemente che le quattro principali tecniche di privacy implementate:

1. Perturbazione dopo il troncamento;
2. Dummy updates con percolazione;
3. Assegnazione di pseudonimi random ad ogni coppia di coordinate;
4. Cloaking spaziale unendo richieste di vari utenti;

Innanzitutto si può affermare che il tema di studio sono i geofence creati nel settore 1 e 2, quindi nei settori più esterni non interessa la creazione di geofence.
Inizialmente sono stati condotti i test per ogni tecnica implementata singolarmente. Poi sono state unite più tecniche per analizzare come queste peggioravano o miglioravano le valutazioni ottenute. Le valutazioni più importanti sono state riportate, non tutte si sono rivelate essere influenti nel progetto. Analizzando ogni tecnica possiamo riportare la seguente spiegazione dettagliata:

- valutando singolarmente la tecnica numero 1 si ottengono risultati uguali all'unione della tecnica 1 e 4. Questo perché il cloaking spaziale ha effetto nei settori 3, 4 e 0, i settori in cui il geofence non è di nostro interesse. Nell'accuratezza si è tenuto comunque conto della comparazione del settore 3, in quanto circa nel 10% delle richieste la posizione si inserisce nei geofence, ma questo solo quando non si ha la perturbazione. Quindi l'unica differenza tra la tecnica 1 e l'unione delle tecniche 1 e 4 , è il settore 3. Non riportiamo la differenza delle valutazioni in quanto si valutano i geofence inseriti nei primi due settori.

- valutando singolarmente la tecnica numero 2, è stata analizzata la media delle 5 posizioni ottenute dalla Dummy Updates. In questo modo si è tenuto conto della media aritmetica delle cinque latitudini e longitudini ottenendo così una posizione probabile dell'utente. Ovviamente non è stato possibile analizzare l'accuratezza, in quanto in tutti i test la posizione media risultante si posizionava all'esterno dei geofence, ma in questo caso è tollerato visto che si vuole rendere la posizione visibile solo per il diretto interessato. É stata inoltre testata anche l'unione di questa tecnica con la numero 1: tenendo conto della media delle 5 posizioni ma considerando quella reale perturbata;

- non è stato possibile valutare la tecnica numero 3, in quanto l'assegnazione di pseudonimi random ad ogni coppia di coordinate non varia le coordinate GPS inviate al sistema. Questo perché si "de-anonimizza" la richiesta, senza alterare effettivamente le coordinate ricevute;

- per la tecnica numero 4, come già anticipato, si sono condotte alcune analisi. Queste hanno riportato che si può ritenere sicura la geolocalizzazione dei settori esterni in quanto avviene una tecnica di randomizzazione della posizione nei settori più

esterni. Questo rende impossibile una valutazione completa mantenendo l'obiettivo del progetto: la tecnica influisce nella gestione della privacy senza però intaccarne i due settori obiettivo di analisi.

Tutte le altre valutazioni non sono state riportate in quanto si sono rivelate essere inutili al fine della valutazione dell'obiettivo del progetto.

6.3.1 Analisi della perturbazione

In questa sezione è stata analizzata inizialmente l'accuratezza. Si riportano le figure rappresentanti i test: ovviamente maggiore è la dimensione del geofence in analisi e maggiore è l'accuratezza del sistema. La perturbazione a seguito del troncamento delle coordinate varia in base al settore di analisi. Con l'aumentare del settore, aumenta l'area di perturbazione delle coordinate, di conseguenza se si aumenta anche la dimensione dei geofence, si hanno più probabilità che la posizione si geolocalizzi correttamente. Questo è il motivo per cui con l'aumentare della dimensione del geofence si aumenta anche la percentuale di correttezza sull'asse delle ordinate. Si riporta anche l'analisi del settore 3, in quanto questa è l'unica metrica in cui ha senso riportare i risultati, in quanto i risultati analizzati sono solo positivi o negativi (nelle altre metriche i numeri sono troppo grandi per essere presi in considerazione).

I risultati ottenuti sono di seguito riportati. Quando il geofence è di 500 metri, la percentuale rappresentante la possibilità per l'utente nel settore 1 di geolocalizzarsi nel geofence sale a 80%. In media, contando il 90% come livello di confidenza, questa percentuale di casi in cui il geofence rileva l'entrata di un utente è compreso fra il 60% e il 100% circa. Mentre se l'utente si trova nel settore 2, la probabilità dell'utente si abbassa al 30%, con un intervallo di confidenza del più o meno 24%. Questo sempre

mantenendo il livello di confidenza a 90. Infine, alle stesse condizioni, si può affermare che, per il settore 3, la probabilità scende al 10%, con un intervallo più o meno del 16%. Aumentando la dimensione del geofence, ma mantenendo sempre gli stessi parametri di confidenza, si ottiene che nel settore 1 la percentuale sale al 90% con un intervallo di più o meno 15%. Nel settore 2 aumenta notevolmente, arrivando al 70% con un intervallo del 24%. Mentre il terzo settore rimane invariato da quanto ottenuto nel caso di analisi del geofence da 500 metri.
Infine, se il geofence preso in considerazione è di dimensione pari a 2 kilometri, allora si può affermare che il 100% delle posizioni nel settore 1 viene correttamente rilevato, mentre nel settore 2 il 90% con un intervallo di più o meno 16%. Mentre la probabilità nel settore 3 sale al 20% delle geolocalizzazioni. L'intervallo di confidenza è del 21%. Questo perché l'area di perturbazione delle coordinate nel settore 3 è più grande del geofence stesso, quindi non si può avere precisione in questo settore.

Un'ulteriore valutazione è stata fatta analizzando la percentuale introdotta precedentemente confrontando però l'aumentare della dimensione dei geofence in ogni settore di analisi. Questo permette di porre attenzione non più sui risultati al variare dei settori presi in analisi, ma all'aumentare della dimensione dei geofence stessi.

Inoltre, si possono mettere a confronto tutti i risultati appena ottenuti, prima confrontando l'accuratezza all'aumentare dei geofence, poi confrontandola nei vari settori di analisi. Si ricorda che con "Geo" si intende il geofence preso in analisi, e con "S1" il settore 1, con "S2" il settore 2 e infine con "S3" il settore 3.
Si ricorda nuovamente che, se si considera la tecnica numero 4, il cloaking spaziale applicato al settore 3 azzererebbe tutti i risultati rendendo l'analisi non oggetto di interesse nello studio. Per questa ragione non viene riportato alcun test a riguardo.

La seconda metrica analizzata è l'errore legato alla distanza del geofence. Questa

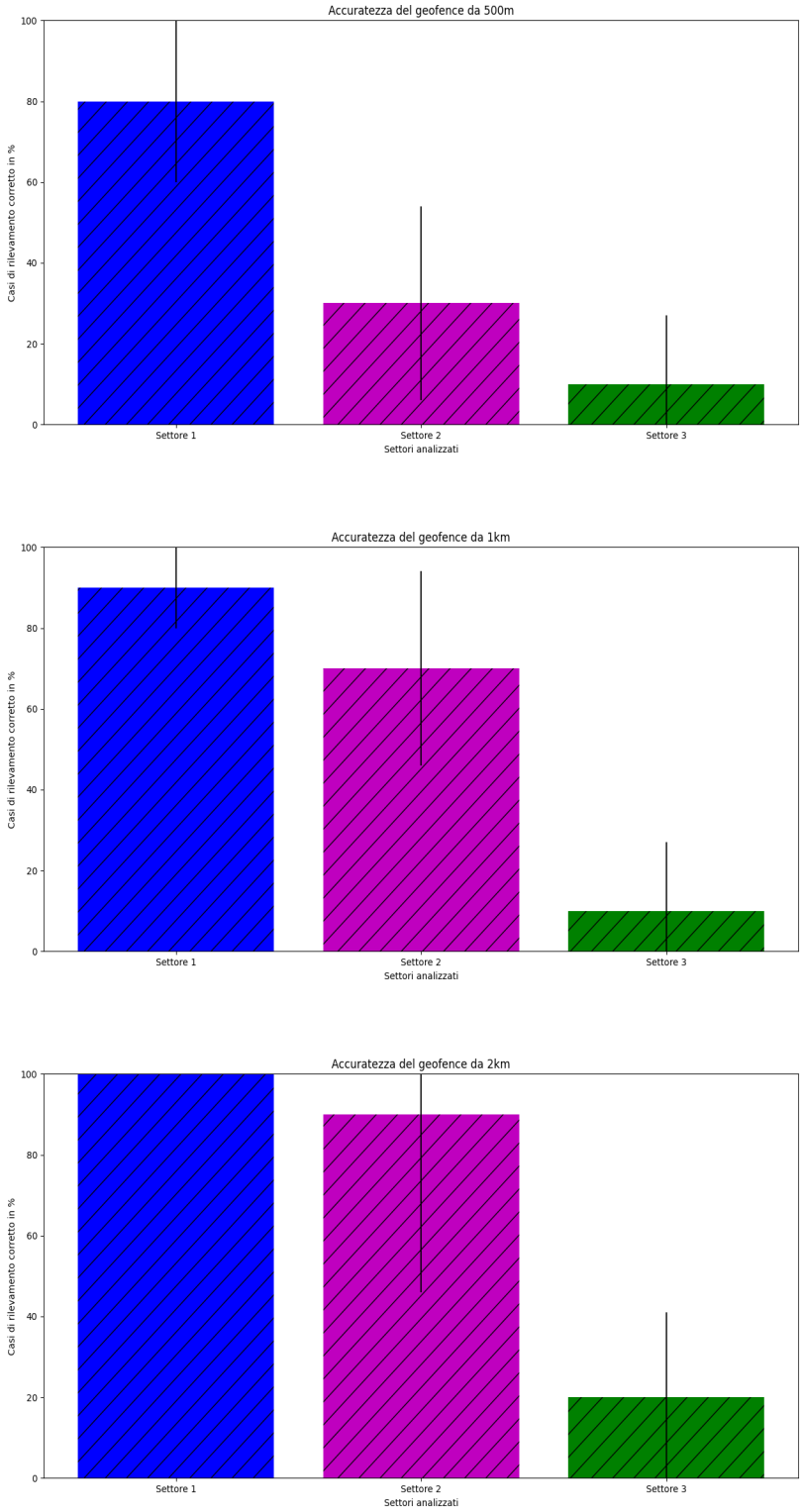


Figura 6.1: Accuratezza della posizione al variare dei settori nell'analisi della perturbazione

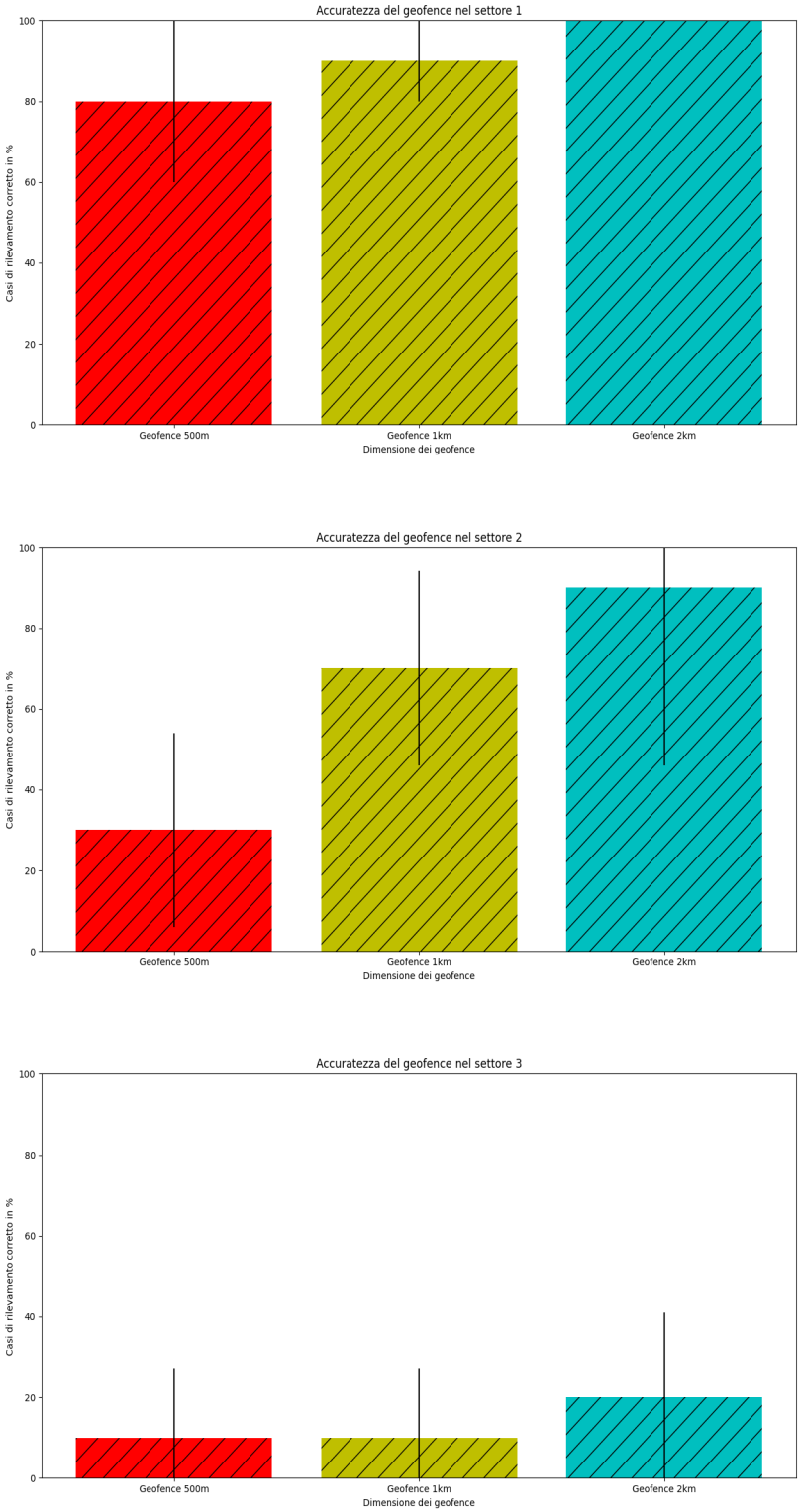


Figura 6.2: Accuratezza della posizione al variare dei geofence nell'analisi della perturbazione

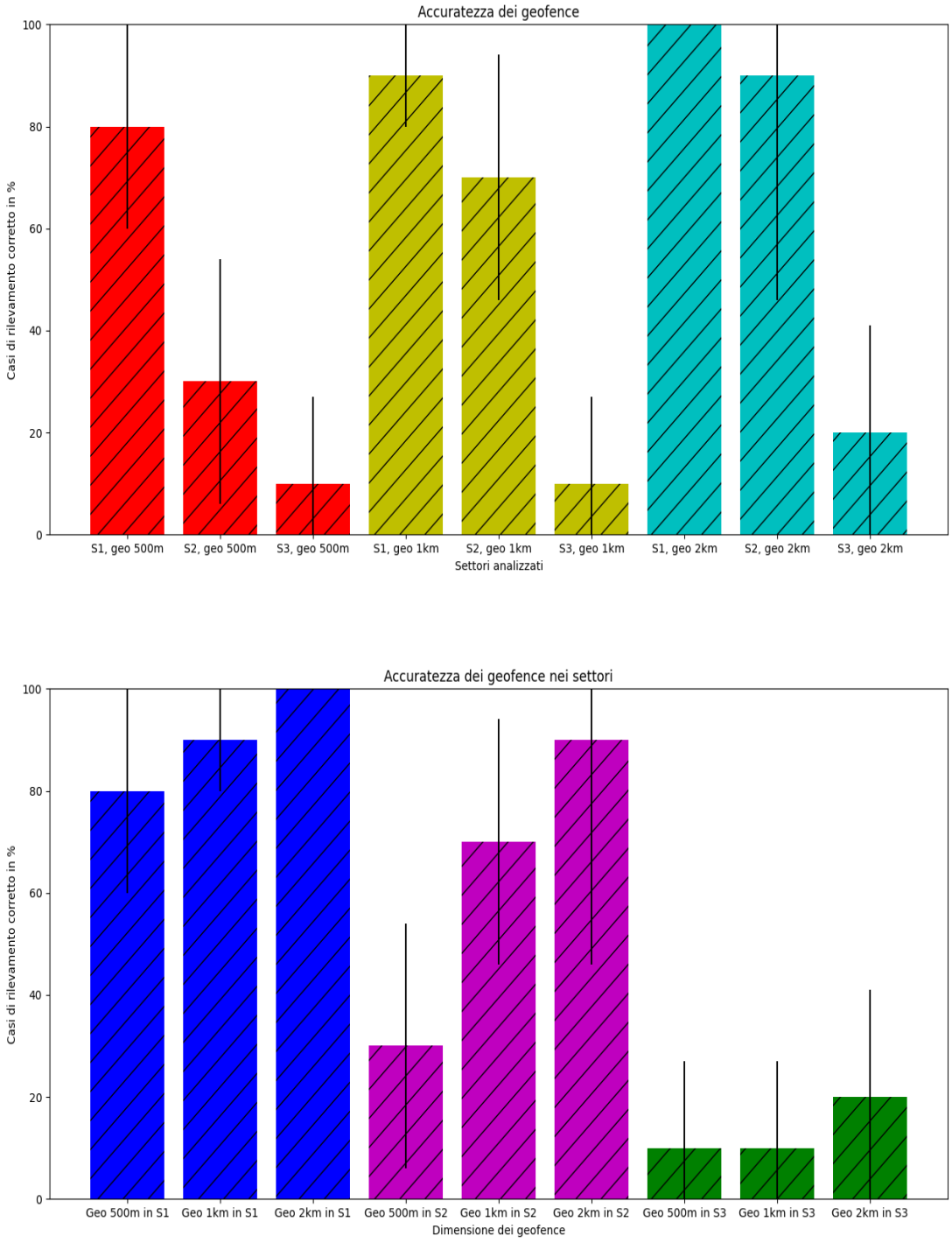


Figura 6.3: Comparazione dell'accuratezza raggruppandola in base alla dimensione dei geofence e ai settori presi in analisi

analisi permette di confrontare i metri di lontananza di un geofence suggerito dalla posizione dell'utente. Si ottengono buoni risultati, soprattutto per il settore 1, il centro dello studio.

I risultati permettono inoltre di evidenziare come all'aumentare della dimensione dei geofence, la probabilità che l'utente si trovi realmente nel geofence suggerito aumenta. Infatti, nel caso di un geofence di 2km, è molto probabile che l'utente si geolocalizzi correttamente al proprio interno. Questo perché l'area di perturbazione delle coordinate nel settore 1 è minore della dimensione del geofence, cioè 2km. I risultati ottenuti si possono riassumere nel seguente elenco. Si ricorda che l'intervallo di confidenza utilizzato è del 90%:

- se il geofence analizzato è di 500 metri, la media è di 85 metri con un intervallo di più o meno 39 metri per il settore 1. Mentre per il settore 2 la media è pari a 462 metri con un intervallo di 191 metri;

- se il geofence analizzato è di 1 kilometro, la media aumenta a 100 metri con un intervallo di più o meno 49 metri per il settore 1. Mentre per il settore 2 la media migliora, diminuendo a 350 metri con un intervallo di 152 metri;

- se il geofence analizzato è di 2 kilometri, il settore 1 ha la percentuale di precisione del 100%, risultato ottimo per l'analisi presa in considerazione. Mentre per il settore 2, la media è di 200 metri, con un'intervallo molto grande, di 98 metri. Questo perché la varianza è molto alta nei casi presi in considerazione.

L'ultima metrica valutata è la precisione: cioè la distanza tra la posizione reale e quella simulata. Ovviamente più si aumenta la perturbazione delle coordinate e più si aumenta l'area di interesse. Infatti nel grafico si riporta questa situazione, ottenendo una

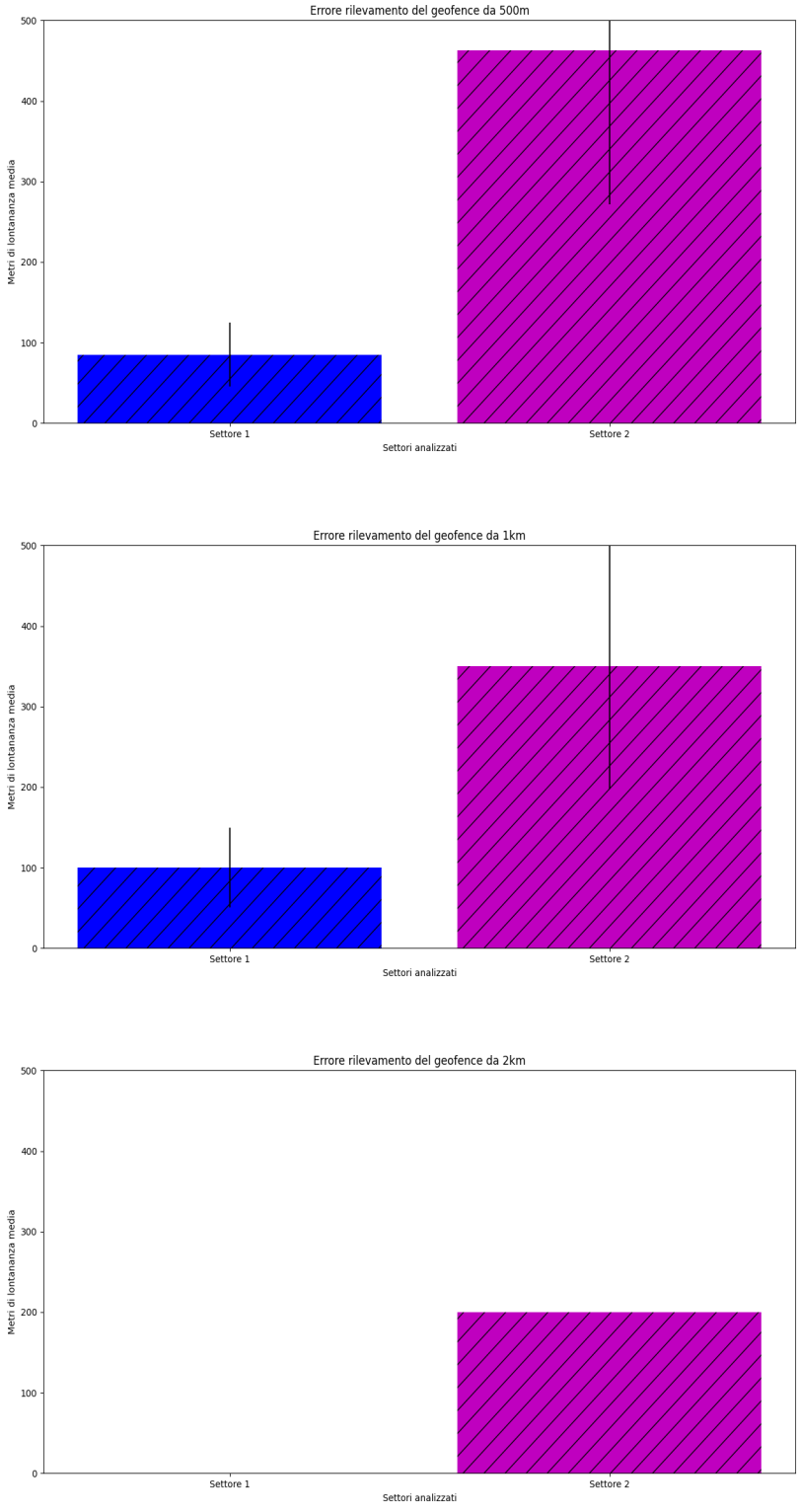


Figura 6.4: Errore legato alla distanza del geofence in base alla dimensione nell'analisi della perturbazione

media di 104 metri per il primo settore e 840 metri per il secondo. Anche gli intervalli di confidenza sono buoni: il primo è di più o meno 22 metri, mentre il secondo è di 254 metri.

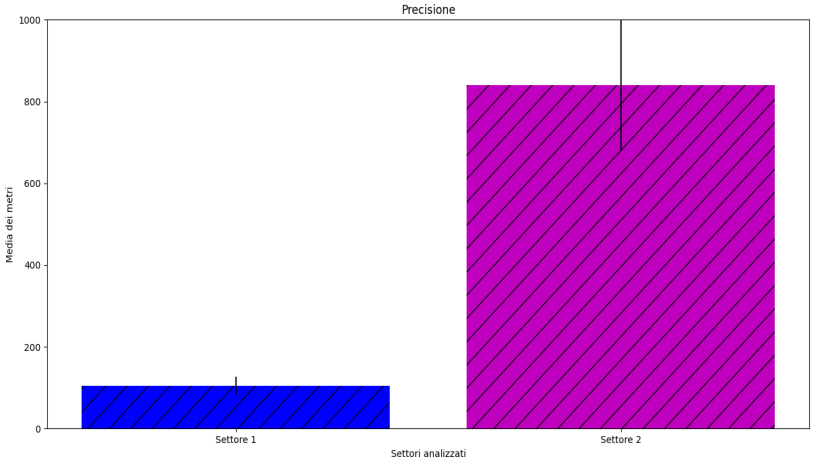


Figura 6.5: Precisione nell'analisi della perturbazione

6.3.2 Analisi dei Dummy Updates

La seconda analisi è stata effettuata implementando singolarmente il meccanismo dei Dummy Updates. Per valutare questa tecnica è stata utilizzata la media delle 5 posizioni che il sistema riceve. Facendo questo, diventa inutile calcolare l'accuratezza, in quanto ovviamente nessuna posizione media calcolata viene geolocalizzata nel geofence (affermazione a seguito di test effettuati che sostengono la toeria). Anche in questa metrica l'intervallo di confidenza considerato è pari a 90%. La prima valutazione effettuata è stato il calcolo dell'errore rappresentato dalla distanza dal geofence alla posizione rilevata dal sistema. Considerando il geofence di 500 metri, la media nel settore 1 è di 1745 metri, con un intervallo di più o meno 414.13 metri. Mentre nel settore 2 la media è

ovviamente più alta, di 1745 metri, con un intervallo che varia più o meno di 2597 metri. Per quanto riguarda il caso di analisi con un geofence di 1 kilometro, nel settore 1 la media è di 1236 metri, con un intervallo di confidenza di più o meno 495 metri. Nel settore 2 invece si è calcolata una media di 7925 metri, con un intervallo che varia più o meno di 2568 metri.
Se si considera invece un geofence di 2 kilometri, la media calcolata diminuisce, avendo un'area che ha la maggior probabilità di comprendere l'utente. La media è pari a 734 metri nel primo settore, con un intervallo di confidenza di più o meno 419 metri. Nel settore 2 la media è pari a 7159 metri, con un intervallo che varia da più o meno 2529 metri.

La seconda e ultima metrica considerata è la precisione della posizione ricevuta nel sistema. Questa metrica è la più interessante per la tecnica di privacy che si sta analizzando: la differenza tra la posizione reale e la media tra le cinque permette di percepire l'offuscamento della tecnica con dei dati reali. Nel settore 1 la media è molto alta, e questo è un fattore positivo, in quanto è una tecnica che permette un buon offuscamento. La media è di 1963 metri, con un intervallo di più o meno 449 metri. Mentre nel secondo settore la media è pari a 8591 metri, con un intervallo più o meno di 2632 metri.

6.3.3 Analisi dei Dummy updates con perturbazione della posizione reale

L'ultima valutazione è stata effettuata sul progetto finale: di fatto, le tecniche influenti che possono variare i test presi in considerazione sono 2. La prima è l'analisi dei Dummy Updates, e la seconda è la perturbazione della posizione reale. Ora si considera quindi l'unione delle due tecniche citate precedentemente: le 4 posizioni fittizie vengono gene-

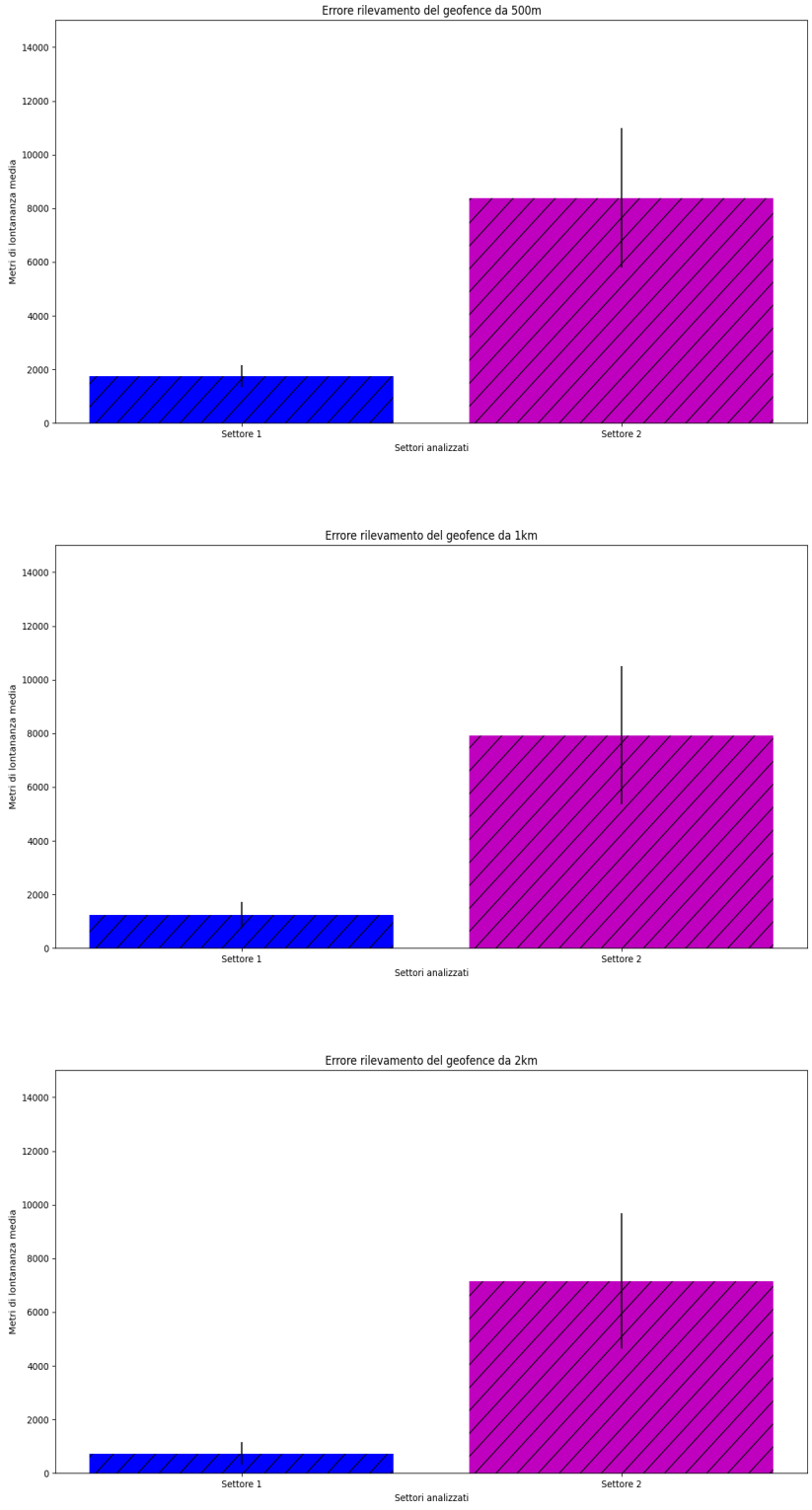


Figura 6.6: Errore rappresentato dalla distanza dal geofence nei Dummy Updates

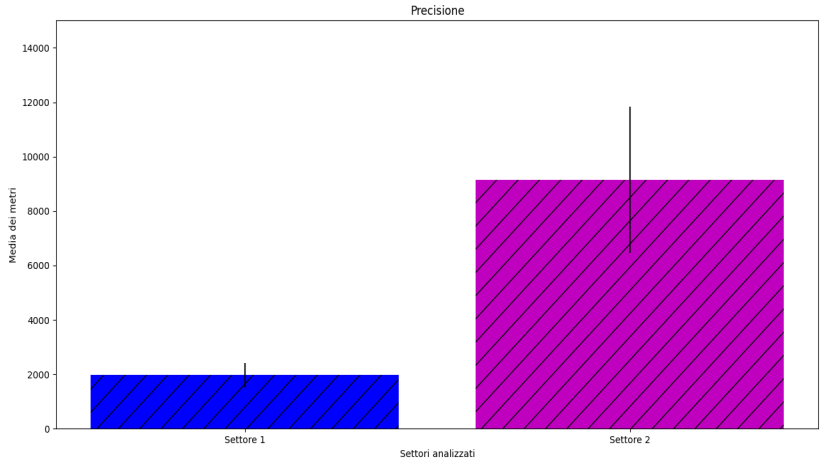


Figura 6.7: Precisione dei Dummy Updates

rate a seguito della perturbazione della posizione reale. Questo si concretizza nei test in qualche variazione di quello precedente. I dati aumentano, ma non notevolmente. Anche in questa metrica l'intervallo di confidenza considerato è pari a 90%. La prima valutazione presa in considerazione è stato l'errore rappresentato dalla distanza dal geofence. Si inizia considerando il geofence di dimensione pari a 500 metri. La media calcolata nel primo settore assomiglia molto a quella precedente, pari a 1709 metri. Anche l'intervallo assomiglia a quello precedente: più o meno di 473 metri. Mentre nel secondo settore la media è di 8857, con un intervallo più o meno di 2674 metri. Anche qua è aumentato di poco rispetto a prima. Nell'analisi dei risultati nei test che riguardano i geofence di 1 kilometro, la media risultante nel primo settore diminuisce rispetto a prima, raggiungendo la cifra di 1275 metri, con un intervallo di 454 metri. Anche nel secondo settore la media diminuisce ed è pari a 8455 metri, con un intervallo di 2620 metri. La motivazione è molto semplice: come già detto in precedenza, aumentando la grandezza del geofence, diminuiscono le distanze da esso. Infine, stesso ragionamento segue i risultati dei test

effettuati con geofence di dimensione pari a 2 kilometri. La media del settore 1 è di 7689 metri, e nel settore 2 è di 7689 metri. La prima ha un intervallo di confidenza di più o meno 427 metri, mentre la seconda di più o meno 99 metri, risultato ottimo per la metrica considerata.

Infine l'ultima valutazione del sistema consente di rapportare la posizione reale e quella simulata del sistema, andando a valutare la precisione del sistema finale di gestione della privacy. Nel settore 1 la media è molto alta, e questo è un fattore positivo, in quanto è una tecnica che permette un buon offuscamento. La media è di 1969 metri, con un intervallo di più o meno 448 metri. Mentre nel secondo settore la media è pari a 9151 metri, con un intervallo più o meno di 2691 metri.

Si può concludere che sono state migliorate notevolmente le prestazioni del sistema senza peggiorare di molto i test ottenuti.

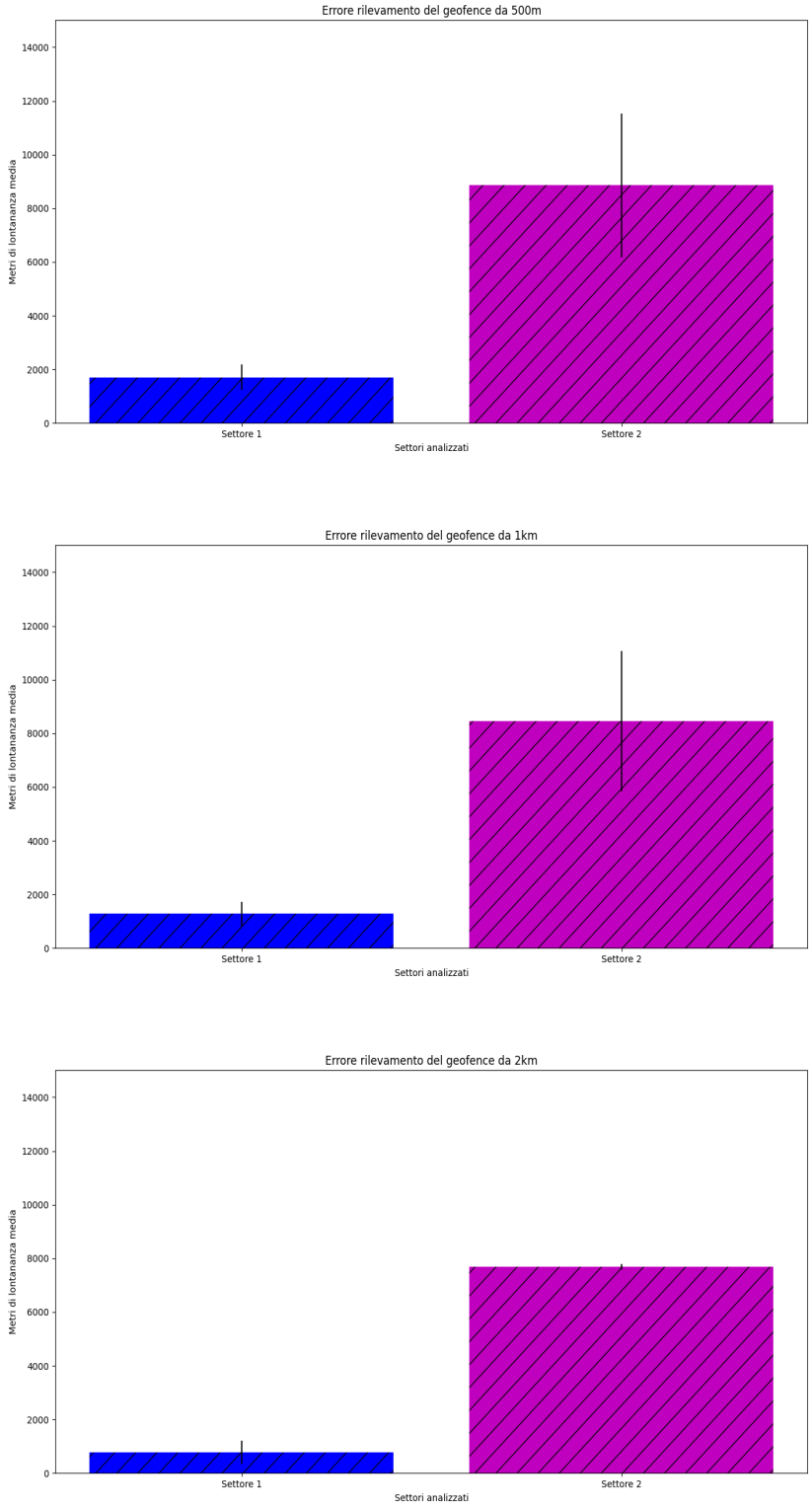


Figura 6.8: Errore rappresentato dalla distanza dal geofence nella tecnica di Privacy finale adottata

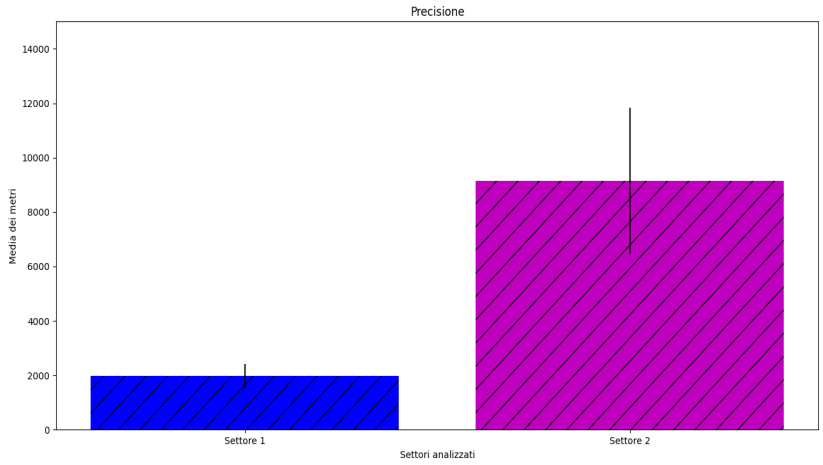


Figura 6.9: Precisione nella tecnica di Privacy finale adottata

Conclusioni e Sviluppi futuri

Un articolo del 10 Dicembre 2018 del New York Times di Jennifer Valentino-Devries, Natasha Singer, Michael H. Keller e Aaron Krolikafferma, afferma "Your Apps Know Where You Were Last Night, and They're Not Keeping It Secret". Questo ricorda come il tema della gestione della privacy del dato geolocalizzato sia un argomento molto discusso in tutto il mondo e a prescindere dal momento attuale. Nell'articolo del 23 Aprile 2020 dell'Economist, scritto da Charlemagne, si tratta il tema della privacy anche in momenti storici delicati come la pandemia da Coronavirus del 2020. L'articolo stesso afferma "If the EU had an official religion, it would be privacy", frase che sottolinea l'importanza e il ruolo della Privacy nel mondo Europeo. In questo progetto di tesi è stata proposta una nuova tecnica di gestione della privacy di un dato geolocalizzato. Prima di tutto è stata fornita una tecnica per la gestione del dato geospaziale nel protocollo MQTT: è stato proposto un formato a cui il payload di un messaggio MQTT deve uniformarsi per consentire l'invio di questa tipologia di dati. Inoltre, è stato proposto un meccanismo di gestione dei geofence che consente ad un utente di abbonarsi a determinati servizi e ricevere aggiornamenti a lui interessanti. Infine, è stata realizzata una tecnica di gestione del dato geospaziale, che regola la riservatezza del dato stesso in un contesto aziendale e non. Come i test hanno dimostrato, questa tecnica di gestione della privacy ha riportato buoni risultati nel caso d'uso d'esempio.

Si possono riportare vari sviluppi futuri possibili nel progetto. Quelli riportati interessano gli elementi principali del progetto: il primo tratta un'eventuale modifica nel formato proposto di un pacchetto nel protocollo MQTT, mentre il secondo riguarda le tecniche di privacy adottate nel sistema. Per quanto riguarda MQTT, si lascia aperto il sistema a ulteriori aggiunte nel formato del pacchetto: già il parametro della "Speed" introdotto, potrebbe permettere un'analisi avanzata nella gestione del geofence, andando a prevedere e regolare eventuali ingressi e uscite dal geofence stesso. Questo permetterebbe di fare deduzioni sul movimento dell'utente, fornendo anche aggiornamenti più specifici. In generale, l'inserimento di altri parametri che potrebbero gestire vari aspetti nel sistema può essere ritenuto un buon sviluppo futuro. Si ricorda il numero elevato di sensori che operano nel mondo dell'IoT: un'eventuale inclusione di altri dati provenienti da questi sensori fornirebbe un maggior dettaglio su cui effettuare analisi di vario genere.

Un'altra proposta nella tutela della privacy consiste nell'estendere il progetto a molte città, andando a correlare i dati sulla densità territoriale con i settori implementati nel sistema. Questo permetterebbe di ampliare il progetto realizzato in vari contesti italiani, gestendo abbonati da tutto il territorio nazionale. In questa fase, uno studio sulla popolazione è necessario: la densità delle aree geografiche svolge un ruolo fondamentale in questo elaborato.

Il progetto è estensibile a vari ambiti. Molti tipi di servizi possono essere implementati e gestiti con questo sistema. L'articolo del 5 Ottobre 2020 della BBC Future, scritto da David Hambling, "What would the world do without GPS?", ricorda come le coordinate GPS hanno un ruolo fondamentale oggigiorno, essendo parte integrante del futuro tecnologico. Per questo motivo, fornendo un'estensione del protocollo MQTT e ponendo particolare attenzione al tema della privacy, è stato fornito un sistema in grado di gestire in modo efficace i dati di geolocalizzazione.

Bibliografia e Sitografia

1. MQTT For Sensor Networks (MQTT-SN) Protocol Specification - Version 1.2 Andy Stanford-Clark and Hong Linh Truong - November 14, 2013

2. MQTT Version 3.1.1 OASIS Standard 2 - 9 October 2014

3. MQTT Version 5.0 OASIS Standard 07 March 2019

4. Università di Roma "La Sapienza" - Dipartimento di Informatica e Sistemistica - Middleware Laboratory - Publish/Subscribe Systems - Leonardo Querzoni

5. 2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a better tomorrow (RTSI) - Comparing Application Layer Protocols for the Internet of Things via Experimentation - Stefan Mijovic, Erion Shehu, Chiara Buratti - DEI, University of Bologna

6. IEEE Wireless Communications October 2016 - Enabling Wireless Communications and Networking Technologies for the Internet of Things - Smart Environments: State of the Art, Taxonomy, and Open Research Challenges - Ejaz Ahmed, Ibrar Yaqoob, Abdullah Gani, Muhammad Imran, and Mohsen Guizani

7. Internet Engineering Task Force (IETF) Z. Shelby - Request for Comments: 7252 ARM - Category: Standards Track K. Hartke - ISSN: 2070-1721 C. Bormann -

Universitaet Bremen TZI - June 2014 - The Constrained Application Protocol (CoAP)

8. Choice of Effective Messaging Protocols for IoT - Systems: MQTT, CoAP, AMQP and HTTP - Nitin Naik - Defence School of Communications and Information Systems - Ministry of Defence, United Kingdom

9. Privacy Invasion through Smarthome IoT Sensing - Ravishankar Chamarajnagar* and Ashwin Ashoky - *VMware Inc., Department of Computer Science, Georgia State University - Department of Computer Science, Georgia State University 2019 IEEE - SECON 2019 workshop on Security Trust and Privacy in Emerging Cyber-Physical Systems

10. Tigist Abera Abbas Acar, Hossein Fereidooni and Amit Sikder . Peek a boo : I see your smart home activities, even encrypted ! arXiv:1808.02741v1, 2018.

11. Dillon Reisman Noah Apthorpe and Nick Feamster . Closing the blinds four strategies for protecting smart home privacy from network observers. arXiv:1705.06809, 2017.

12. Garante privacy: trattamento di dati personali mediante geo-localizzazione dei dispositivi aziendali - `http://www.dottrinalavoro.it/notizie-c/garante-privacy-trattamento-di-dati-personali-mediante-geo-localizzazione-dei-dispositivi-aziendali`

13. Giusella Finocchiaro - Diritto di Internet - Zanichelli Bologna

14. Cosa intendiamo per dati personali? - `https://www.garanteprivacy.it/home/diritti/cosa-intendiamo-per-dati-personali`

15. L'autorità - https://www.garanteprivacy.it/home/autorita

16. Garante per la protezione dei dati personali - https://it.wikipedia.org/wiki/Garante_per_la_protezione_dei_dati_personali

17. GPS, dati di localizzazione e privacy - https://protezionedatipersonali.it/gps-dati-localizzazione-privacy

18. Dato personale - https://protezionedatipersonali.it/dato-personale

19. Il titolare del trattamento dei dati è il soggetto le cui finalità vengono perseguite attraverso il trattamento medesimo - https://www.diritto.it/titolare-del-trattamento-dei-dati-soggetto-le-cui-finalita-vengono-perseguite-trattamento-medesimo/

20. Geolocalizzazione, tutto ciò che devi sapere: pericoli e sfide - https://www.cybersecurityup.it/blog/2-news-cyber-security/132-geolocalizzazione,-tutto-ci-che-c--da-sapere-pericoli-e-sfide

21. Geotagging - https://en.wikipedia.org/wiki/Geotagging

22. Geocoding - https://en.wikipedia.org/wiki/Geocoding

23. Georeferencing - https://en.wikipedia.org/wiki/Georeferencing

24. Il gps deve rispettare la privacy degli utenti: decisione storica del Garante - https://www.corrierecomunicazioni.it/privacy/il-gps-deve-rispettare-la-privacy-degli-utenti-decisione-storica-del-garante/

25. Internet of Things - IoT - Internet delle cose - https://it.wikipedia.org/wiki/Internet_delle_cose

26. Il protocollo MQTT - https://mqtt.org/

27. V. Primault, A. Boutet, S. B, Mokhtar and L. Brinoe, The long road to computational location privacy: a survey, IEEE Communications Survey and Tutorials, 2019

28. Mohamed F. Mokbel, Privacy Preserving Location Services, Proc. of IEEE MDM 2008 - https://wwwusers.cs.umn.edu/mokbel/tutorials/icdm08

29. Please Rob Me: Site Tells The World When You're Not Home - https://www.huffpost.com/entry/please-rob-me-site-tells_n_465966

30. Strava heat map exposes secret military locations, sparks security fears - https://geoawesomeness.com/strava-heat-map-exposes-secret-military-locations-sparks-security-fears/

31. Privacy of location tracking device owners threatened by 'Trackmageddon' flaws, Bradley Barth, Rene Millman - https://www.scmagazine.com/home/security-news/vulnerabilities/privacy-of-location-tracking-device-owners-threatened-by-trackmageddon-flaws/

32. Tinder dating app was sharing more of users' location data than they realised - https://www.theguardian.com/technology/2014/feb/20/tinder-app-dating-data-location-sharing

33. Use Google Maps in Incognito mode - https://support.google.com/maps/answer/9430563?hl=en&ref_topic=6384263

34. Yi Song, Daniel Dahlmeier, and Stéphane Bressan. Not so unique in the crowd: a simple and effective algorithm for anonymizing location data. In Luo Si and Hui

Yang, editors, Proceeding of the 1st International Work shop on Privacy Preserving IR: When Information Retrieval Meets Pri vacy and Security, volume 1225 of CEUR Workshop Proceedings, pages 19-24

35. Trackmageddon: Servizio di localizzazione GPS di monitoraggio - https://sensorstechforum.com/it/trackmageddon-gps-hackable/

36. Many Popular Android Apps Leak Sensitive Data, Leaving Millions Of Consumers At Risk, AJ Dellinger - https://www.forbes.com/sites/ajdellinger/2019/06/07/many-popular-android-apps-leak-sensitive-data-leaving-millions-of-consumers-at-risk/

37. A Survey on Privacy in Location-Based Services - Revisión en privacidad en servicios basados en localización - Mayra Zurbarán, Liliana González, Pedro Wightman Rojas, M. Labrador

38. Konstantinos Chatzikokolakis, Ehab Elsalamouny, Catuscia Palamidessi, Anna Pazii. Methods for Location Privacy: A comparative overview. Foundations and Trends in Privacy and Security , Now publishers inc, 2017, 1 (4), pp.199-257. ff10.1561/3300000017ff. ffhal-01421457v2f

39. S. Gambs et al., "Show Me How You Move and I Will Tell You Who You Are", Transactions on Data Privacy 4 2011) 103 126 , http://www.tdp.cat/issues11/tdp.a078a11.pdf

40. Igor Bilogrevic et al.. Inferring Social Ties in Academic Networks Using Short Range Wireless Communications. 12th ACM Workshop on Privacy in the Electronic Society (WPES), Nov 2013, Berlin, Germany. https://hal.archivesouvertes.fr/hal00853975/document

41. Anastasios Noulas, Salvatore Scellato, Neal Lathia, Cecilia Mascolo, Mining User Mobility Features for Next Place Prediction in Location based Services, 12 th International Conference on Data Mining, 2012

42. Riding with the Stars: Passenger Privacy in the NYC Taxicab Dataset - `http://content.research.neustar.biz/blog/differential-privacy/QueriesWidget.html`

43. 2020 IEEE Symposium on Security and Privacy -Burglars' IoT Paradise: Understanding and Mitigating Security Risks of General Messaging Protocols on IoT Clouds - Yan Jia;Luyi Xing;Yuhang Mao;Dongfang Zhao;XiaoFeng Wang;Shangru Zhao;Yuqing Zhang

44. Intelligent parking Cloud services based on IoT using MQTT protocol - Prarna Dhar, Poonam Gupta - 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT)

45. MQTT-G: A Publish/Subscribe Protocol with Geolocation - Robert Bryce;Thomas Shaw;Gautam Srivastava - 2018 41st International Conference on Telecommunications and Signal Processing (TSP)

46. Design of auto wakeup alarming system for commuters in railway sleeper coaches - T. M. Hayath;Dadapeer;S. G. Tejashwini;N. M. Indravasan - 2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon)

47. Green Communication Protocol with Geolocation - Gautam Srivastava;Andrew Fisher;Robert Bryce;Jorge Crichigno - 2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring)

48. A Cyber-Secured MQTT based Offline Automation System - Nahian Ibn Hasan;Md. Tasnimul Hasan;Nazmul Haque Turja;Rishad Raiyan;Shuvagata Saha;Md. Farhad Hossain - 2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)

49. Work-in-Progress: Towards an International Data Spaces Connector for the Internet of Things - Michael Nast;Benjamin Rother;Frank Golatowski;Dirk Timmermann;Jens Leveling;Christian Olms;Christian Nissen - 2020 16th IEEE International Conference on Factory Communication Systems (WFCS)

50. Protecting Location Privacy: Optimal Strategy against Localization Attacks - Reza Shokri, George Theodorakopoulos, Carmela Troncoso, Jean-Pierre Hubaux, and Jean-Yves Le Boudec - https://www.researchgate.net/publication/236670805_Protecting_location_privacy_Optimal_strategy_against_localization_attacks https://users.cs.cf.ac.uk/G.Theodorakopoulos/papers/protecting-ccs12.pdf

51. S. Bandyopadhyay and A. Bhattacharyya, "Lightweight internet protocols for web enablement of sensors using constrained gateway devices," in Computing, Networking and Communications (ICNC), 2013 International Conference on. IEEE, 2013, pp. 334–340.

52. T.Jaffey.(2014,February)MQTTandCoAP,IoTprotocols.[Online].Available: https://eclipse.org/community/eclipsenewsletter/2014/february/article2.php

53. OASIS.org.(2015,December10)MQTT3.1.1.editedbyAndrewBanksandRahulGupta.29october2014.OASISStandard.[Online].http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/mqtt-v3.1.1.html

54. I.Grigorik,œMakingthewebfasterwithHTTP2.0,CommunicationsoftheACM,vol.56,no.12,pp.4249,2013.

55. P. Kalnis, G. Ghinita, K. Mouratidis e D. Papadias, "Preventing LocationBased Identity Inference in Anonymous Spatial Queries", Transazioni IEEE sulla conoscenza e ingegneria dei dati, vol. 19, n ° 12, pp. 1719-1733, dicembre 2007.

56. D. Micciancio, "A First Glimpse of Cryptography's Holy Grail", Commun. ACM, vol. 53, n.3, pp. 96-96, marzo. 2010.

57. K. W. Tan, Y. Lin, and K. Mouratidis, "Spatial cloaking revisited: Distinguishing information leakage from anonymity", in Advances in Spatial and Temporal Databases. Springer, 2009, pp. 117-134.

58. B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan, "Private information retrieval", Journal of the ACM, vol. 45, n.6, pp. 965-981, Nov. 1998.

59. C. S. Jensen, H. Lu, and M. L. Yiu, "Location privacy techniques in clientserver architectures", in Privacy in location-based applications. Springer, 2009, pp. 31-58.

60. M. Duckham and L. Kulik, "A formal model of obfuscation and negotiation for location privacy," in Pervasive Computing. Springer, 2005, pp. 152-170.

61. A. R. Beresford and F. Stajano, "Mix zones: User privacy in location-awar services", in Pervasive Computing and Communications Workshops, 2004. Proceedings of the Second IEEE Annual Conference on, 2004, pp. 127-131."

62. M. Gruteser and D. Grunwald, “Anonymous usage of location-based services through spatial and temporal cloaking”, in Proceedings of the 1st international conference on Mobile systems, applications and services, 2003, pp. 31-42.

63. Y. Che, Q. Yang, and X. Hong, “A dual-active spatial cloaking algorithm for location privacy preserving in mobile peer-to-peer networks”, in Wireless Communications and Networking Conference (WCNC), 2012 IEEE, 2012, pp.209-2102.

64. S. Mascetti, C. Bettini, D. Freni, and X. S. Wang, “Spatial generalisation algorithms for LBS privacy preservation”, Journal of Location Based Services, vol. 1, n. 3, pp. 179-207, 2007.

65. Kassem Fawaz and Kang G. Shin. Location privacy protection for smartphone users. In Proceedings of the 21st ACM Conference on Computer and Communications Security (CCS 2014), pages 239–250. ACM Press, 2014.

66. The Internet of Things: Messaging Protocols - L. Bononi, M. Di Felice, Department Of Computer Science And Engineering, University Of Bologna, Italy - `http://site.unibo.it/iot`

67. Privacy in Location Context Aware Systems - Prof. Marco Di Felice - Department of Computer Science and Engineering - University of Bologna - `http://www.cs.unibo.it/difelice/cas/`

68. Node - `https://nodejs.org/it/`

69. Postgresql - `https://www.postgresql.org/`

70. Postgis - `https://postgis.net/docs/using_postgis_dbmanagement.html`

71. Openlayers - https://openlayers.org/

72. Handlebars Javascript - https://handlebarsjs.com/

73. Il protocollo MQTT - https://en.wikipedia.org/wiki/MQTT

74. Javascript, fonte ufficiale - https://www.javascript.com/

75. Javascript - https://it.wikipedia.org/wiki/JavaScript

76. Dati personali - https://it.wikipedia.org/wiki/Dati_personali

77. Demone informatica - https://it.wikipedia.org/wiki/Demone_(informatica)

78. Fonte ISTAT - https://www.istat.it/it/files//2018/12/C01.pdf

79. Google Street View - https://it.wikipedia.org/wiki/Google_Street_View

80. Tecniche di pseudonimizzazione e migliori pratiche - Novembre 2019 - ENISA Europa - enisa.europa.eu

81. Piattaforma di Google Maps Documentazione - https://developers.google.com/maps/documentation?hl=it

82. Trovare Google Maps coordinate rapidamente e facilmente - https://www.mapcoordinates.net/it

83. Coordinate GPS - Coordinate Geografiche - https://www.coordinate-gps.it/

84. Strumenti per GPS e Mappe - https://www.faureragani.it/mygps/mygps.aspx

85. Libreria MQTT per Javascript - https://www.npmjs.com/package/mqtt

86. Socket TCP in Node.js - https://hackerstribe.com/2016/socket-tcp-in-node-js/

87. Libreria MQTT - The definitive source of the best JavaScript libraries, frameworks, and plugins. - https://www.javascripting.com/view/mqtt-js

88. Eseguire MQTT per Javascript - https://docs.cloudplugs.com/kb/Developer-Guides/MQTT-API/Javascript-Examples

89. Creazione di dati in formato geojson - https://geojson.io/

90. Using GeoJSON with Leaflet - https://leafletjs.com/examples/geojson/

91. Node.js e Database: come interagire con MySQL e SQL Server - https://www.mrwebmaster.it/javascript/node-js-database-come-interagire-mysql-sql-server_13040.html

92. Node-db, gestire database MySQL con Node.js - https://www.html.it/pag/33419/node-db-gestire-database-mysql-con-nodejs/

93. Formato GeoJSON - https://geojson.org/

94. Specifiche GeoJSON - https://en.wikipedia.org/wiki/GeoJSON

95. Definizione di JSON - https://en.wikipedia.org/wiki/JSON

96. Specifiche sul formato GeoJSON - https://tools.ietf.org/html/rfc7946

97. Corso Iot, Università di Bologna, dipartimento DISI - https://site.unibo.it/iot/en/teaching-1/the-iot-course

98. What is geofencing? Putting location to work, Sarah K. White - https://www.cio.com/article/2383123/geofencing-explained.html

99. Geofencing: definizione, funzionamento, applicazioni e utilizzo. Franscesco Destri - https://www.cwi.it/mobile-wireless/app-mobile/geofencing-definizione-e-utilizzo-109720

Ringraziamenti

Inizio ringraziando il professore Di Felice che mi ha aiutato nel mio progetto di tesi. Ritengo sia stato veramente fondamentale e incisivo il suo contributo nel mio percorso di studi. Lo ringrazio molto per la collaborazione avuta in tutti questi mesi.
Vorrei inoltre ringraziare anche tutte le persone che hanno contribuito in questo mio percorso accademico. Lo farò citandone solo alcune, ma non in ordine di importanza. Vorrei innanzitutto ringraziare i miei genitori. Mio padre, roccia della nostra famiglia e uomo esemplare, per la sua presenza e fermezza. La sua sobrietà e tranquillità mi hanno insegnato tanto. La sua intelligenza e cultura sono sempre state fonte di ispirazione nella mia vita. Mia madre, senza cui non sarei al mondo e sarei nulla. A colei che ha sempre visto il meglio in me e mi ha sempre accolto a braccia aperte amorevolmente. Alla donna che ha sempre fatto tante rinunce per i suoi figli, donando anima e impegno. Ha sempre creduto nelle mie potenzialità, sempre più di chiunque altro. Grazie, davvero. Tutto questo è anche opera loro. Vorrei ringraziare Greta, la mia compagna, perché, come le ho sempre detto, ogni mio traguardo è un po' anche suo. Ci auguro mille altri di questi giorni. Questo è solo l'inizio.
Un grande ringraziamento va a Daniele, fratello paziente che mi ha sempre sorretto e supportato. A Simone, Denise, Luca e Eleonora: un esempio modello di famiglia. La vostra solidità e bellezza vi rendono esempio di vita. Un enorme ringraziamento va a mia

cugina Laura, a cui ho sempre pensato come una sorella. Un punto fisso, una certezza. Anche ad Elena, piccola ma grande forte donna, spero di essere un buon esempio per te. A mia zia Eleonora, che mi ha sempre aiutato con tutte le sue forze. Vorrei poter ricambiare tutti i viaggi per prendere il treno e tutte le fatiche pre esame e post esame che le ho fatto fare. A tutti i miei cugini: in particolare a Maicol e il suo ingegno. A tutti i miei zii e parenti. A Gabriella, che con il suo cibo e la sua attenzione mi ha nutrito e viziato in questi anni. A Clelia, che mi ha sempre ospitato ed elogiato. A Claudio, Giorgia e Nello, sempre disponibili in tutto. A Lorenzo, persona straordinaria che, nel tempo, si è rivelato essere un vero amico e sempre presente. A Debora, alla sua risata e al suo essere. A Martina, Lorenzo, Valentina, Chiara, Michele, Elisa e Gianluca, che mi hanno sopportato giorno dopo giorno. Ai miei compagni di università, sia triennale sia magistrale: Elena, Stefano, Miriana, Giada, Francesca, Salvatore, Luca, Alessandro, Andrea, Irene e tutti gli altri che mi hanno aiutato in tutto il mio percorso. A Riccardo, appuntamento settimanale che ha creato un legame unico e forte; diventato un amico fondamentale e uno studente modello. Al gruppo ragazzi e a tutti i restanti miei amici. Dulcis in fundo, ai miei compagni di viaggio. Ai miei coetanei che mi hanno sempre sostenuto in tutto il mio percorso. Ricordo chi non ho già nominato, Francesco, Greta, Alessandro, Alessandro, Simone, Carlo e Martina. Il gruppo creato mi ha aiutato ogni settimana a continuare i miei studi dando sempre il massimo.

Un ultimo grande grazie va ai miei nonni, in particolare a Luigi, sperando di renderlo orgoglioso del mio percorso accademico. Ma un ringraziamento speciale va a Caterina, a cui è stato permesso di supportarmi nel mio percorso. Non scorderò mai tutto quello che ha sempre fatto per me. Alla donna che si è sempre dimostrata sangue del mio sangue. Esempio di vita e di ispirazione. Colei che porterò sempre nel mio cuore e nei miei pensieri per tutta la mia vita.