Datenschutz und Ubiquitous Computing
Am Beispiel von Location Privacy

Thierry Bücheler

July 12, 2004

Abstract

Wie können wir in einer Zeit, in welcher ortsabhängige Applikationen beginnen, unsere Bewegungen im Namen der Bequemlichkeit zu verfolgen, unsere Privatsphäre schützen?

Heute können Positionstechnologien relativ einfach den Aufenthaltsort eines Subjekts bestimmen: Sensornetze und andere UbiComp-Technologien erweitern die Möglichkeiten, automatisch Daten zu sammeln. Spezialisierte Technologien zur Ortsbestimmung haben eine drastische Preisreduktion erfahren. Verbesserte Datenbank- und Speichersysteme erlauben permanentes Aufnehmen und eine weite Verbreitung dieser Daten. Moderne Data-Mining-Techniken ermöglichen das Auffinden von Bewegungsmustern.

Dennoch ist die Nutzung der anfallenden Daten kaum geregelt oder sie wird trotz Datenschutzgesetzen missachtet. Experten warnen davor, dass leicht ein Markt entstehen könnte, auf dem mit Ortsinformationen unbedachter Nutzer ohne deren Wissen rege gehandelt wird.

Ein Ziel der Forschung ist die Minimierung personenbezogener Informationen. In diesem Paper werden zwei Ansätze vorgestellt, welche es erlauben sollen, die Privatsphäre eines Individuums ausreichend zu schützen. Der erste Ansatz "verwischt" Daten in einem Sensornetz, bereits bevor die Daten an einen (potentiell bösen) Server geleitet werden. Der zweite Ansatz verwendet häufig ändernde Pseudonyme, um einen Nutzer zu anonymisieren, ohne die Verwendung personalisierter Dienste zu verhindern.

1 Einführung: Location Based Services

Location Based Services (LBS) sind Mehrwertdienste, die sich auf den aktuellen Standort des Users beziehen. Optimalerweise wird der Standort des Users dabei automatisch erkannt. Seine Positionsdaten in Verbindung mit dem personalisierten Benutzerprofil des Nutzers erlauben ein massgeschneidertes Angebot.

Die Benützung von LBS ist aber nicht immer freiwillig. Bei der Benutzung eines GPS-Geräts beispielsweise kann ein User trivialerweise auf die Herausgabe seiner Ortsinformation verzichten, da er das GPS-Gerät mit den darauf enthaltenen Informationen kontrolliert.

Dies ist aber nicht möglich bei grösseren Netzwerken wie z.B. GSM-Interfaces (Mobiltelefon-Tracking) oder gesichtserkennenden Kameraprojekten (z.B. an Flughäfen).

2 Privacy und Anonymity

Die *Universal Declaration of Human Rights* der UNO von 1948 deklariert, dass jedermann ein Recht auf Privatsphäre (Privacy) zu Hause, in der Familie und in seiner Korrespondenz hat.

Location Privacy ist eine spezielle Art von sogenannter Information Privacy [1], definiert als die *Fähigkeit, andere daran zu hindern, den gegenwärtigen oder vergangenen Aufenthaltsort einer Person zu erfahren* [1].

Der Schutz der Privatsphäre wird als genügend erachtet, wenn die Verfolgung einer Person (Ortstracking) einen vergleichbaren Aufwand benötigt wie traditionelle Überwachungsmassnahmen (Beschatten, Funksender an Auto haften etc.) [2].

3 Gegenwärtige Probleme

Daher sind einige Probleme gegeben, die ein User von LBS lösen will:

- Er will den Zugang zur Ortsinformation kontrollieren
- Gleichzeitig will er aber nicht den gesamten Zugang sperren

Da dies auf den ersten Blick (und auch auf den zweiten) gegensätzliche Ziele sind, muss ein trade-off zwischen Bekanntgabe des Ortes und Kontrolle dieser Information gefunden werden.

[1] definiert in der Global Internet Liberty Campaign, www.gilc.org/privacy/survey

3.1 Attacken

Auf den User von LBS-Dienstleistungen können verschiedene Attacken geführt werden:

Locating/Lokalisierung Das Subjekt kann auf verschiedene Arten lokalisiert werden, z.B. Aufdecken von LBS-Informationen oder Wireless-Signale triangulieren.

Identification/Identifikation Der Real-World-Name oder z.B. die Adresse des Subjekts wird bekannt. Netzwerkadressen werden nicht als Identifikation des Subjekts bezeichnet, sondern als Pseudonyme, welche es ermöglichen können, ein Subjekt zu identifizieren.

Lokalisierung und Identifikation sind z.T. verknüpft: So kann ein Attackierender beispielsweise den Ort einer Person feststellen und mittels Geocode [2] die physische Adresse aufgrund der Koordinaten (und damit evtl. das Wohnhaus des Subjekts) bestimmen. Somit kann er vom Ort der Person auf deren Identitt schliessen.

3.2 Ziele

Es stehen also drei Ziele im Mittelpunkt:

- Ein User möchte von den Vorteilen des Ortstrackings (Bestimmung des Ortes und/oder Verfolgung gewisser Schritte) profitieren.
- Die Privatsphäre des Users soll dabei gewährleistet/geschützt bleiben.
- In vielen Fällen macht es Sinn, die wahre Identität des Users geheim zu halten.

3.3 Anonymität

Eine gegenwärtig schon eingesetzte Praxis ist es, dem User sogenannte *Privacy Policies* vorzulegen, welche den User über die Datenverwendung des Serviceproviders informieren und dem User als Entscheidungsbasis dienen, Informationen freizugeben. Privacy Policies können aber sehr umfangreich und kompliziert sein. Selbst wenn sie im Stil standardisiert würden, hätte der User wahrscheinlich keine Lust, diese ständig zu lesen, bevor er einen Service nutzt.

[2] http://www.geocode.com

Anonymität kann einen hohen Grad an Privacy ermöglichen, rettet Service-Nutzer davor, sich mit den Privacy Policies der Serviceprovider herumzuschlagen und reduziert die Anforderungen der Serviceprovider, private Information zu schützen [1, 2]. Daher wird Anonymität als geeignetes Mittel empfunden, die Privatsphäre eines Nutzers ausreichend zu schützen.

3.4 K-Anonymität

Bevor die zwei Lösungsvorschläge vorgestellt werden, soll hier noch das Prinzip der "K-Anonymität" vorgestellt werden, welches bei beiden Vorschlägen eine Rolle spielt. Der Begriff stammt von Samarati [6] und wird folgendermassen definiert:

Daten sind *k-anonym*, wenn jeder Ortsevent (erfasstes Subjekt) aus dem Netzwerk ununterscheidbar ist von mindestens k-1 anderen Ortsevents (Subjekten).

4 Lösungsvorschlag 1

Daten sind schwieriger zu schützen, wenn sie einmal auf einem System gespeichert sind. Der Vorschlag von Gruteser et al. [3] erhöht die Privacy durch verteilte Anonymitätsmechanismen direkt im (Sensor-)Netz. Im Paper wird ein Sensornetz als Beispieltechnologie verwendet. Das Prinzip kann aber auch auf andere Technologien angewendet werden.

Bei vielen vorgeschlagenen Privacy-Lösungen besteht das Problem, dass die Middleware von nicht vertrauenswürdigen Institutionen gesteuert werden kann, was die Anonymität brechen kann.

Die Mechanismen in diesem Paper anonymisieren die Daten direkt in einem Sensornetz. Somit muss sich der Serviceprovider nicht um die Sicherheit der Daten kümmern.

4.1 Welche Art von Applikationen können geschützt werden

Gewisse Applikationen benötigen aggregierte Statistiken über die Popularität gewisser Orte (z.B. im Supermarkt, um den besten Standort für die Lancierung eines neuen Produkts zu bestimmen), aber nicht unbedingt präzise Informationen über den genauen Standort einer

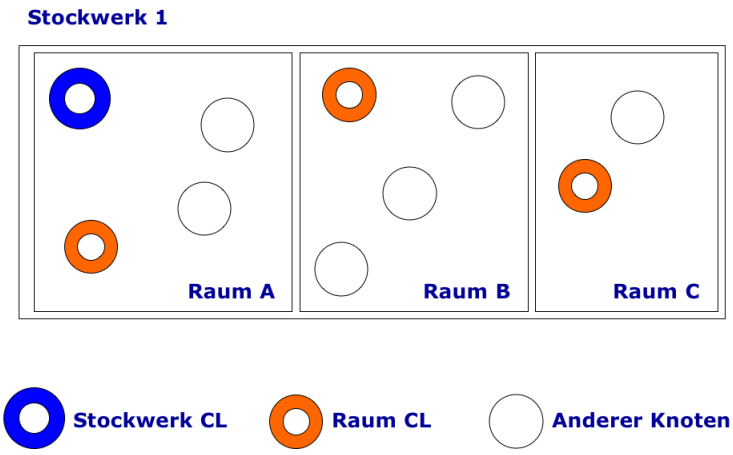

Figure 1: Die Hierarchische Gliederung

einzelnen Person zu einem bestimmten Zeitpunkt. Ich nenne sie hier '"Bulk-Applikationen". Genau diese Art von Applikationen können mit dem Vorschlag von Gruteser et al. geschützt werden.

Drei Voraussetzungen werden bei diesem System erfüllt:

- Daten können aggregiert werden
- Daten werden gecloaked (verwischt)
- Die Kommunikation zwischen Knoten des Netzes ist sicher

4.2 Hierarchische Aggregation

Um eine Datenaggregation zu ermöglichen, wird das Sensornetz hierarchisch gegliedert. Auf jedem Hierarchielevel wird mittels eines geeigneten Protokolls ein coordination leader (cl) gewählt (ein mögliches Protokoll wird im Paper beschrieben).

Die IDs (Namen) der Knoten werden mit einem "Präfixcode" bestückt, d.h. Knoten in einem Supermarkt haben beispielsweise dieselbe Stockwerk- und Raum-ID codiert. Die IDs an sich sind aber alle eindeutig (unique).

4.3 Data Cloaking

Der Hauptnutzen, den man aus der Hierarchie der Knoten ziehen kann, ist das verwischen der genauen Ortsinformation. Eine weitere

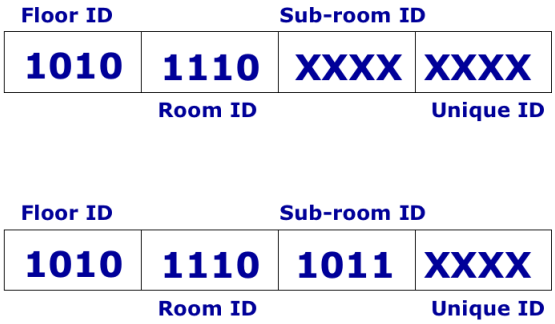

Figure 2: Der Ort wird "unscharf" durch Abschneidung von Bits

Möglichkeit, die aber nicht von der Hierarchie abhängt, ist die Verwischung der Information, die ein Knoten sendet. Es gibt also zwei Möglichkeiten, Daten ungenauer zu machen:

- Biete weniger örtliche Genauigkeit: Cloake die ID (genauer Ort des Sensors), liefere präzise Daten
- Verwische die Anzahl gezählter Subjekte in der Zone: Cloake die Daten, liefere dafür die genaue Knoten-ID

Zwischen diesen zwei Varianten wird eine ausgewählt, indem die Anzahl Subjekte mit einem festgesetzten Anonymitätslevel k verglichen wird. übersteigt die Anzahl Subjekte k, werden die Daten gecloaked und die ID wird präzis geliefert (und umgekehrt).

Das Sensornetz verwischt die Daten, sodass das k-Anonymitäts-Kriterium erfüllt wird. Im Idealfall wird nur minimal verwischt, damit die Daten ihre Brauchbarkeit für möglichst viele Applikationen behalten.

4.4 Sichere Kommunikation

Damit keine altbekannten Attacken auf das Sensornetz geführt werden können, sollen die Knoten mit verschlüsselten, authentifizierten Datenpaketen kommunizieren. Um Traffic Analysis zu verhindern, muss jeder Knoten dabei mindestens ein Paket pro Datensammlungsintervall versenden, auch wenn kein Event eintritt.

4.5 Fazit Lösung 1

Lösungsvorschlag 1 ermöglicht v.a. eine *Ortsunschärfe* (die Verwischung der Daten agiert sozusagen als Zückerchen) und schützt vor *Lokalisierungsattacken*. Es ist aber schwierig, diese Technologie auf einzelne Individuen und deren Tracking anzuwenden.

5 Lösungsvorschlag 2

Der zweite Lösungsvorschlag bedient sich der Hilfe von Pseudonymen. Bezüglich Pseudonymen gibt es verschiedene Arten von Location Based Services: Solche, bei denen die genaue Identität des Users wichtig ist (z.B. "Wo im Gebäude ist XY?"), solche, bei denen die Identität unwichtig ist (z.B. "Nenne mir in einer Tankstellenzone den Benzinpreis"), und schlussendlich solche, bei denen zwar absolute Anonymität nicht möglich ist, ein internes Pseudonym aber genügt (z.B. "Zeige mir auf diesem Computer meinen Desktop").

5.1 Statische Pseudonyme

Bei statischen Pseudonymen ist es bei vielen Anwendungen relativ einfach, die Identität eines Users zu bestimmen. Beispielsweise kann das Pseudonym in einem Bürogebäude verfolgt werden und der User aufgrund seines häufigsten Aufenthaltsortes (wahrscheinlich sein Büro, ein sogenanntes *Home*) identifiziert werden.

Eine Lösung dieses Problems ist die Registrierung eines ganzen Sets von Pseudonymen. Diese Pseudonyme werden dann während des Trackings gewechselt. Falls die räumliche und zeitliche Auflösung des Lokalisationssystems aber gross genug ist, können die Homes ähnlich den statischen Pseudonymen bestimmt werden, indem die Pseudonyme bei genauer Beobachtung in den Homes "zusammengehängt" werden können. Alastair Beresford und Frank Stajano [1] haben mit den *Mix Zones* eine Lösung für dieses Problem gefunden.

Bevor diese aber vorgestellt wird, hie noch etwas zu LBS und Zonen:

5.2 Bindung an Zonen

Normalerweise will man nicht, dass LBS-Informationen für einen bestimmten Ort einer Applikation an einem anderen Ort bekannt wer-

den. So will ein User beispielsweise nicht, dass sei Chef am Arbeitsplatz weiss, dass der Mitarbeiter ber Mittag im Vergngungsviertel war. LBS-Informationen können also optimalerweise an Zonen gebunden werden (z.B. die Bro-Zone). Wichtig ist es also, den Austausch von persönlichen Informationen zwischen diesen Zonen zu verhindern.

Beresford und Stajano setzen, um das Zusammenhängen von Pseudonymen zu verhindern, sogenannte *Mix Zones* ein. Die Mix Zones gehen auf das Konzept der *Mix Networks* und *Mix Nodes* von David Chaum zurück.

Mix Networks sind Store-and-Forward-Netzwerke, die aus normalen Knoten und Mix Nodes bestehen. Die Mix Nodes nehmen n gleichlange verschlüsselte Pakete als Input und ändern deren Reihenfolge nach einer bestimmten Metrik (z.B. lexikografisch oder zufällig), bevor sie sie wieder verschlüsseln und weitersenden. Somit kann ein Beobachter keine Verbindung zwischen eingehenden und ausgehenden Paketen in den Mix Nodes herstellen.

5.3 Mix Zones

Eine Mix Zone ist eine örtliche Region maximaler Grösse, in welcher kein User einen LBS abonniert hat. Im Gegensatz dazu sind *Application Zones* Orte, an denen mindestens ein User einen LBS-Dienst registriert hat. Mix Zones können entweder a priori von einer Middleware definiert oder laufend berechnet werden.

Weil Applikationen keinerlei Ortsinformationen erhalten, während User in einer Mix Zone sind, werden Identitäten gemixt. Ein User ändert sein Pseudonym in einer Mix Zone und ein Beobachter kann keine Verbindung zwischen eingehenden und ausgehenden Usern herstellen. Die k-Anonymität wird erreicht, sobald k Subjekte zur gleichen Zeit in einer Mix Zone sind (oder die Mix Zone dementsprechend vergrössert wird). Ein User kann dann entscheiden, ob der Anonymitätslevel k ihm genügend Privatsphäre bietet.

Ein gewisses Problem besteht aber bei der Verwendung der k-Anonymität als Sicherheitsmesslatte. Verfügt nämlich ein Beobachter über historische Daten, kann er mittels statistischen Attacken die Sicherheit der k-Anonymität um ein Vielfaches unterbieten. So besteht beispielsweise eine starke Korrelation zwischen Eintritts- und Austrittspunkt in einer Mix Zone. Die k-Anonymität in einer Mix Zone ist nur ein gutes Mass, wenn alle Subjekte für einen bösen Beobachter gleich interessant sind und zum Beispiel die Ein- und Austrittspunkte

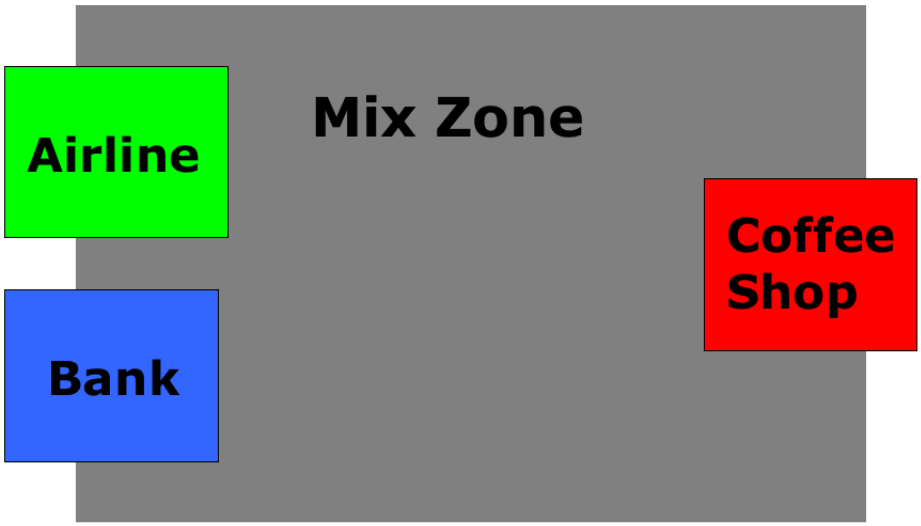

Figure 3: Ein einfaches Mix-Zone-Beispiel

der Subjekte gleich wahrscheinlich sind. Eine solche Attacke wird in [1] beschrieben.

5.4 Fazit Lösung 2

Lösungsvorschlag 2 ermöglicht v.a. eine *ID-Unschärfe* (mit Hilfe der Pseudonymisierung) und schützt v.a. vor *Identifikationsattacken*. Der Ansatz von Beresford und Stajano hat gezeigt, dass Pseudonymisieren nicht so einfach ist, wie es scheint und darum einige Anforderungen an die Sicherheit gestellt.

6 Was wird eigentlich gelöst

Die Lösung von Gruteser und Grunwald stärkt den Privacy-Schutz im Vergleich zu Lösungen auf dem Datenbank-Level, weil er das Sammeln von Privacy-sensitiven Daten verhindert. Bei Bulk-Applikationen (und allgemeiner Applikationen, die eine Ortsunschärfe erlauben) kann der Ort eines Einzelevents (-Subjekts) verschleiert werden. Man kann eine gewisse (unscharfe) Awareness schaffen, die bei Bedarf (z.B. Emer-

gency, Ausbrechen eines Brandes in einem Gebäude) geschärft werden kann (die Sensoren sind ja präzis verteilt und identifizierbar).

Der Ansatz von Beresford und Stajano ermöglicht es einem Individuum, das Tracking zu verhindern. Es wird ein Mass für Pseudonym-Switching definiert und gezeigt, dass die k-Anonymität einer Mix-Zone nicht als Qualitätsmass genügt. Es ist möglich, pseudonymisierte Services zu nutzen, solange man ab und zu mixt.

Applikationen der Art "Wo ist X?" oder "Ich möchte ein Taxi an meine Adresse" funktionieren mit beiden Ansätzen nicht. Wahrscheinlich können diese Applikationen auf dem technischen Level nicht geschützt werden, da Identität und Ort gleichzeitig bekannt sein müssen, um den gewünschten Service zu erbringen.

Für viele LBS ergibt die Verwischung von Ortsinformationen wenig Sinn. In diesem Fall ist es angebracht, an der Verwischung der Identität zu forschen, um den User anonym zu halten. Für verschiedene LBS-Arten müssen verschiedene Privacy-Modelle verwendet werden. Ein sinnvolles Ziel ist die Datensparsamkeit (data minimization). Man will nur so viele Daten verbreiten wie wirklich benötigt werden.

7 Weitere Forschung

Was kann getan werden, um die Forschung am Datenschutz zu vereinfachen?

Hier einige Erkenntnisse zusammengefasst, die aus den Papers im Literaturverzeichnis gewonnen wurden:

Benötigt wird ein formales, wenn möglich probabilistisches Modell für Lokationsanonymität, welches einen kontinuierlichen Datenstrom liefert. Dies würde eine bessere Evaluation ermöglichen. Ein besseres Verständnis über die Anforderungen an die Genauigkeit von Lokationsdaten für verschiedene Klassen von Applikationen würde eine Analyse des Anonymitätslevels für LBS- und Sensornetz-Applikationen erleichtern.

Evtl. kann eine Standardisierung und Vereinfachung der Privacy Policies gefunden werden, damit auch LBS, die eine genaue Identifikation und Lokalisierung bentigen, einfacher geschtzt werden knnen.

References

[1] Alastair R. Beresford, Frank Stajano: *Location Privacy in Pervasive Computing*, Security and Privacy Magazine, 2003

[2] Marco Gruteser and Dirk Grunwald: *Anonymous Usage of LBS through Spatial and Temporal cloaking*, 2003

[3] Gruteser, Schelle, Jain, Han, Grunwald: *Privacy-Aware Locatio Sensor Networks*, Proceedings of HotOS IX, 2003

[4] Marco Gruteser and Dirk Grunwald: *A Methodological Assessment of Location Privacy Risks in Wireless Hotspot Networks*, 2003

[5] Pfitzmann and Köhntopp: *Anonymity, Unobservability, and Pseudonymity - A Proposal for Terminology*, Draft v0.14, 2003

[6] Perangela Samarati: *Protecting Respondents' Identities in Microdata Release*, IEEE Transactionns on Knowledge and Date Engineering, 13(6), 2001