Karlsruher Institut für Technologie

Datenschutz und Privatheit in vernetzten Informationssystemen

Kapitel 8: Ubiquitous Computing – Lokationsbasierte Dienste

Erik Buchmann (buchmann@kit.edu)

IPD, Systeme der Informationsverwaltung, Nachwuchsgruppe „Privacy Awareness in Information Systems“

Inhalte und Lernziele dieses Kapitels

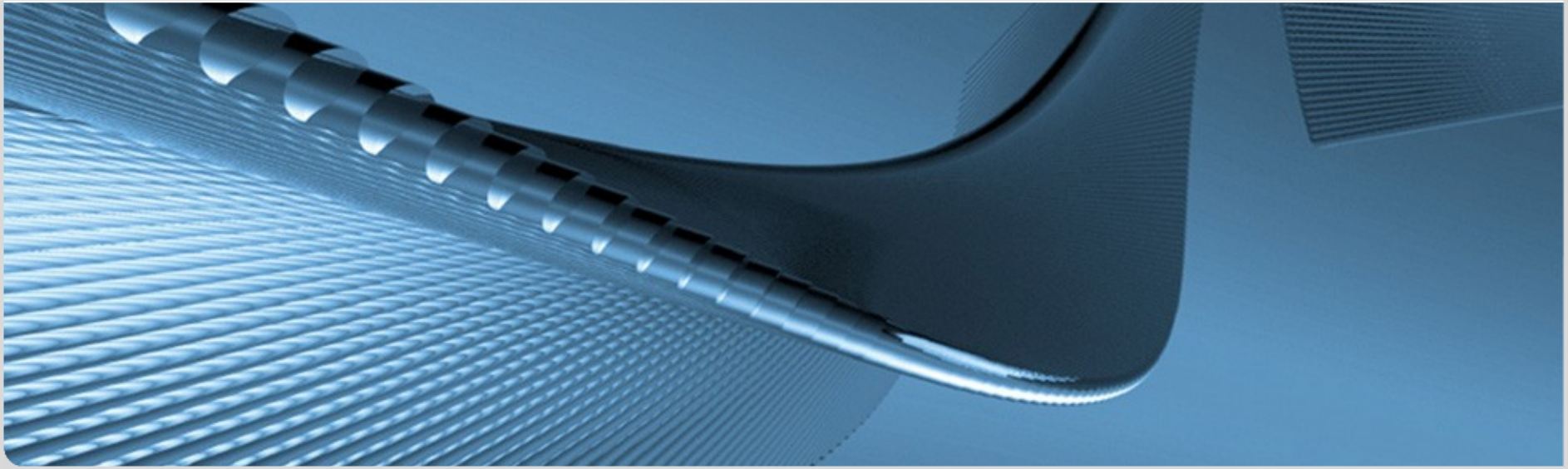
- Einführung: Lokationsbasierte Dienste
  - Methoden der Positionsbestimmung, LBS-Anwendungen
  - Datenschutzprobleme
- Verfahren zum Datenschutz
  - zentrale Verfahren
  - dezentrale Verfahren
  - Datenselbstschutz
- Abschluss

- Lernziele
  - Sie können die besonderen Herausforderungen erläutern, die Systeme mit Ortsbezug an den Datenschutz stellen.
  - Sie können je nach den Erfordernissen einer LBS-Anwendung geeignete Datenschutzmechanismen vorschlagen und diskutieren.

Karlsruher Institut für Technologie

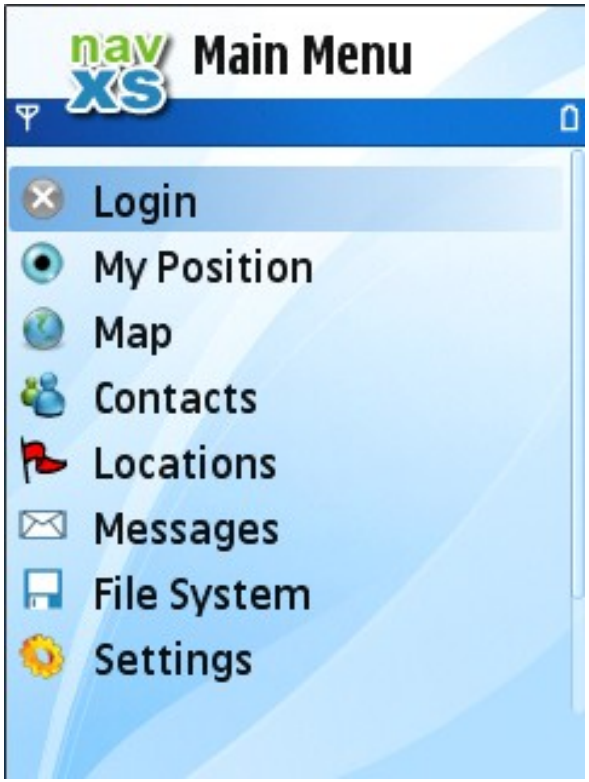
Lokationsbasierte Dienste

IPD, Systeme der Informationsverwaltung, Nachwuchsgruppe „Privacy Awareness in Information Systems“



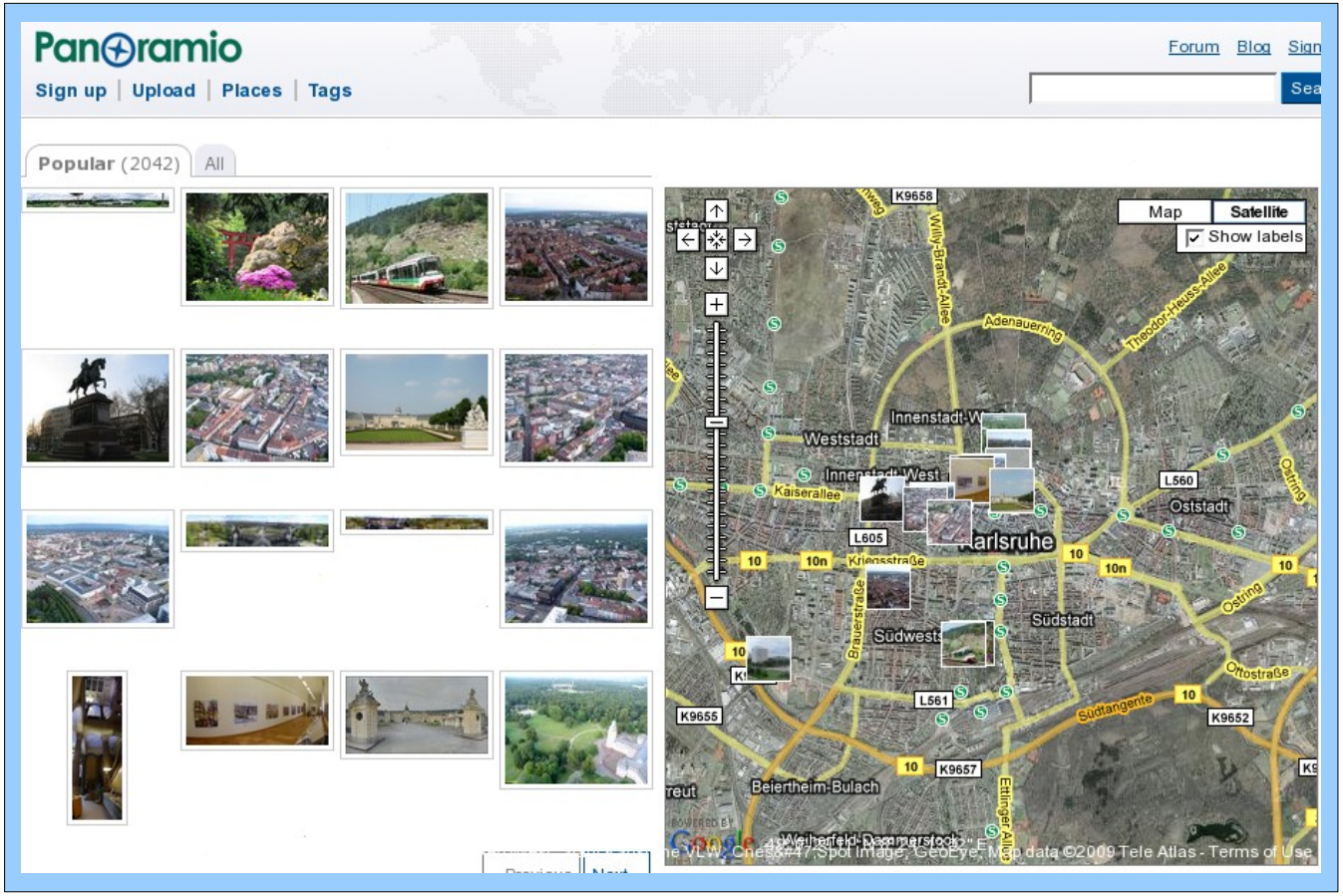
Beispiel: NavXS

- einfacher Austausch von Ortsinformationen
  - vergleichbar mit Instant Messaging, z.B. Skype, ICQ
- läuft auf Mobiltelefonen mit GPS



Beispiel: Panoramio

- Geotagging
  - Annotieren von Positionen mit Fotos und Texten



Weitere Beispiele

- Routenplanung
  - *Wie komme ich von hier nach Heidelberg?*
- Points-of-Interest
  - *Wo ist das nächste griechische Restaurant?*
- Track-your-Child
  - *Wo ist der Nachwuchs grade?*
- Flottenmanagement
  - *Wo befinden sich die Firmenwagen?*
- Abrechnung der Straßenmaut
  - *Fahrzeuge melden gefahrene Strecken an Zentrale.*

Möglichkeiten zur Positionsbestimmung

- Im Freien:
  - GPS
  - Triangulierung von Funkbaken
  - Datenbank von WLAN- oder Mobilfunkstationen (z.B. Google Latitude)
- Innerhalb von Gebäuden:
  - Infrarot-Baken (z.B. Active Badge)
  - Ultraschall (z.B. Active Bat)

GPS

- seit 1995 in Betrieb
- 31 (min.24) Satelliten 20km über dem Erdboden
- Ortung durch Empfang von mindestens 4 Satelliten
  - Genauigkeit: besser als 10m in 90% aller Messungen
- Verwendete Funkfrequenzen können Wasser oder Beton nicht durchdringen
  - kein Inneneinsatz
  - Probleme bei (starkem) Unwetter

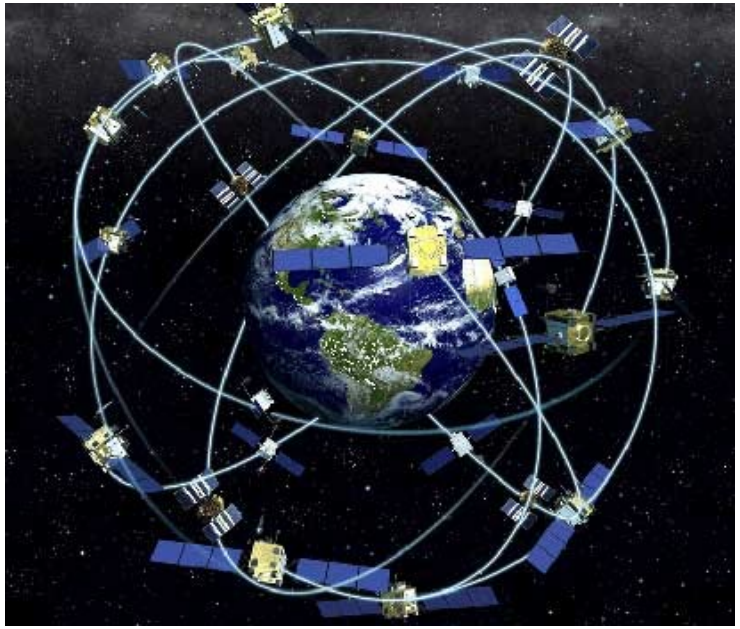
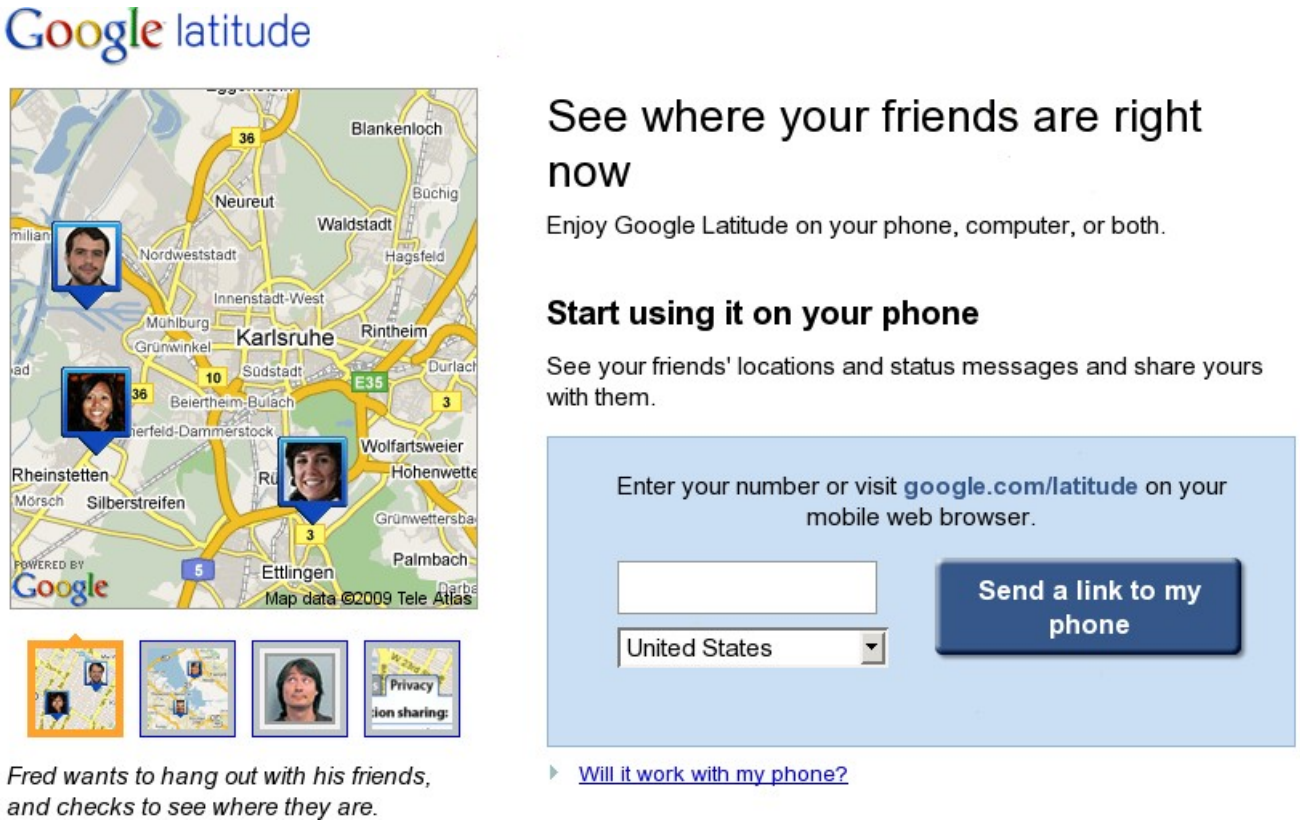


Bild: www.gpsmagazine.com

Google Latitude

- Ortung über Datenbank von Mobilfunkzellen im Handy
  - Genauigkeit bestenfalls 200m, in dünnbesiedelten Gegenden (wenig Mobilfunkstationen) schlechter



Fred wants to hang out with his friends, and checks to see where they are.

Active Bat

- Person trägt Ultraschallsender
- System von Ultraschall-Sensoren im Gebäude
  - Sensoren messen Entfernung zum Ultraschallsender
- 720 Sensoren decken 1000qm auf 3 Gebäudeetagen ab
- Auflösung 3cm, 75 Objekte pro Sekunde
- mit mehreren Sendern pro Person kann auch Ausrichtung im Raum (Blickrichtung) bestimmt werden

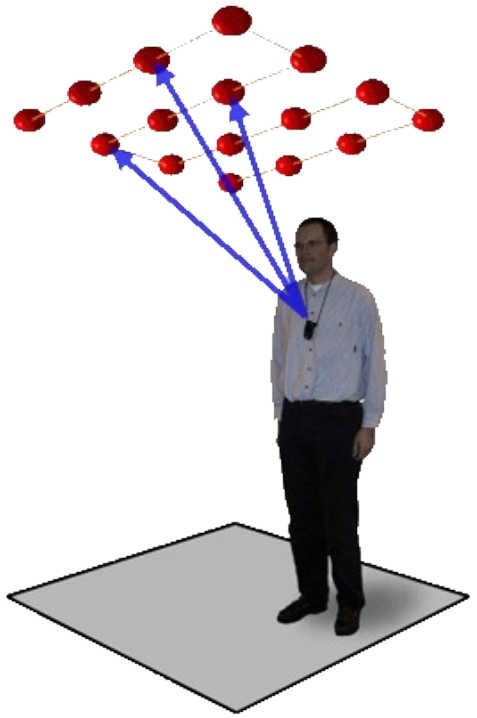
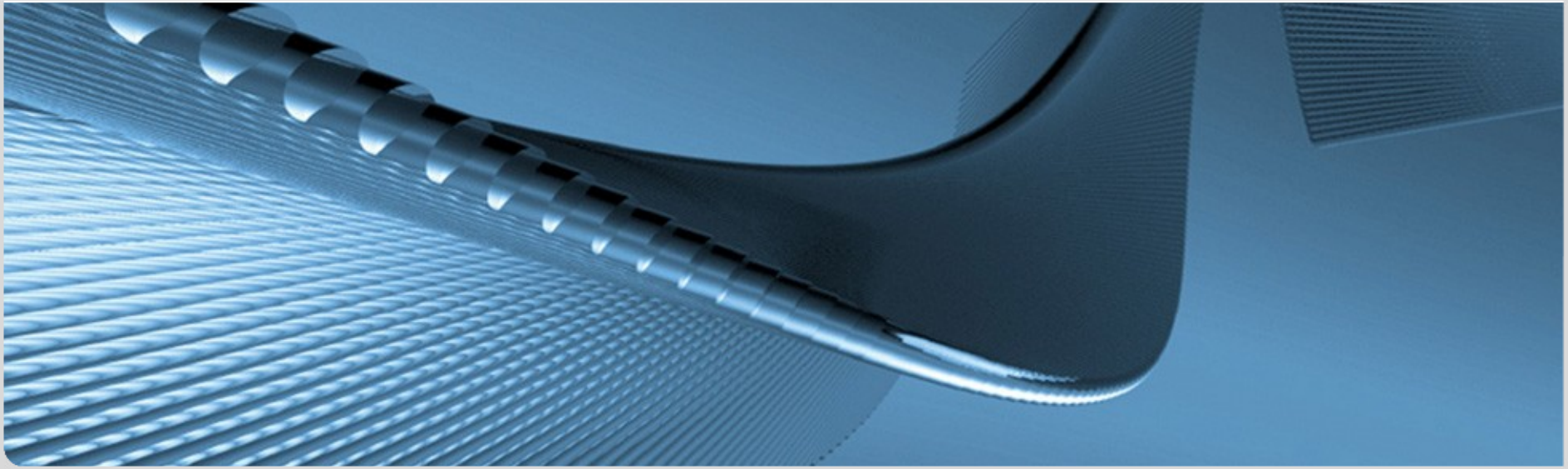


Bild: http://www.cl.cam.ac.uk

Karlsruher Institut für Technologie

Datenschutzprobleme in Lokationsbasierten Diensten

IPD, Systeme der Informationsverwaltung, Nachwuchsgruppe „Privacy Awareness in Information Systems“



Gespeicherte Daten

- **Einzelkoordinaten**
  - einzelne Anfragen nach POI (z.B. nächste Tankstelle)
- **Tracks**
  - Aufzeichnung des GPS-Datenstroms
  - Sequenzen von aufeinanderfolgenden Einzelanfragen, z.B. nach aktuellen Staumeldungen
- **Dienstbezogene Inhalte**
  - Anfragen, Annotationen beim Geotagging, etc.
- **Metadaten**
  - Nutzer, Zeit der Aufzeichnung
  - Dienstbezogene Metadaten, z.B. EXIF-Informationen in Fotos

Einzelkoordinaten

- Erlauben Rückschlüsse auf
  - aktuellen bzw. vergangenen Aufenthaltsort,
    z.B. das eigene Haus, Treffpunkt mit Freunden, Urlaubsziele
  - persönliche Vorlieben
    z.B. Restaurant, Kino, Sehenswürdigkeiten
  - Lebensumstände
    z.B. Arbeitgeber, Medizinische Einrichtungen
  - soziale Kontakte
    z.B. Analyse, ob sich unterschiedliche Personen zur gleichen Zeit an gleichen Koordinaten aufhalten

Tracks

- Erlauben (zusätzlich zu den Einzelkoordinaten) Rückschlüsse auf
  - Reiseverlauf
  - Geschwindigkeit
    wichtig z.B. bei Verkehrsregeln
  - Reiseziel
    z.B. durch extrapolierten der Route
  - Gewohnheiten
    z.B. immer wiederkehrende Routen wie den Arbeitsweg

Dienstbezogene Inhalte

- Beispiel Panoramio:
  Wer und was wurde fotografiert?
  - Rückschlüsse auf
    - Lebensumstände
    - soziale Kontakte
    - Interessen, Hobbies
    - Fähigkeiten (gute vs. schlechte Bilder)
    - etc.

Metadaten

- generisch, z.B. Zeit der Anfrage, Nutzer
  - generieren von Bewegungsprofilen möglich
- anwendungsspezifisch
  - Beispiel: Fotos in Panoramio speichern EXIF-Daten
    - Aufnahmezeit (kann sich von der Zeit der Dienstnutzung unterscheiden)
    - Kamera-Modell (teure Kamera?)
    - Aufnahmeparameter (unfähiger Fotograf?)

- Metadaten können als Quasi-Identifier dienen, d.h. Fingerabdruck, der Anwenderverhalten pseudonym nachverfolgbar macht

Datenschutz in LBS

- Inhalte und Metadaten meist nicht zu vermeiden
- Tracks durch Anfragefolgen ebenfalls oft unvermeidlich

- Je weniger genau die Positionsangaben, desto schlechter die Dienstqualität
  - grobe Ortsposition
    - Dienste geben ungenaue Ergebnisse zurück
    - Privatheit des Anwenders sichergestellt
  - genaue Ortsposition
    - hohe Dienstgüte
    - Standort des Anwenders wird Dritten kenntlich gemacht

- → **Kompromiss erforderlich**

k-Anonymität auf Positionsdaten

- Wiederholung: k-Anonymität
  - *Eine Person ist k-Anonym, wenn sie von k-1 anderen nicht unterschieden werden kann.*
- Auf Koordinaten
  - ändere die Koordinaten von k Nutzern so, dass
    - sie ununterscheidbar werden
    - die Dienstqualität nur minimal sinkt
    - keine Zuordnung zu exakten Koordinaten möglich ist
- Methoden (Auswahl)
  - Koordinaten runden oder durch Region ersetzen
  - Zufallswerte addieren
  - Positionen durch Landmarken ersetzen

Angreifermodell

- Für uns relevant: Angreifer möchte an die Positionsdaten des Nutzers gelangen
  - fremden Dienst übernehmen oder eigenen Dienst anbieten, der Positionsdaten sammelt
  - unbemerktes Mithören der Kommunikation, z.B. als man-in-the-middle
  - Angriff auf eventuell vorhandenen Anonymisierungsdienst
    - selbst als Anonymisierungsdienst ausgeben, z.B. in Peer-basierten Verfahren
    - sich ins Anonymity Set einschleusen, z.B. als einer der k Mitglieder bei k-Anonymity

Kriterien für Datenschutzansätze

- technisch
  - Angriffsmöglichkeiten
  - Skalierbarkeit
  - Kommunikationsaufwand
  - Rechenaufwand
  - für Anonymisierung erforderliche Mindest-Nutzerdichte
  - Anwendbarkeit auf Anfragefolgen
- nicht-technisch
  - benötigtes Hintergrundwissen,Komplexität des Verfahrens
  - mentaler Aufwand, Aufmerksamkeit

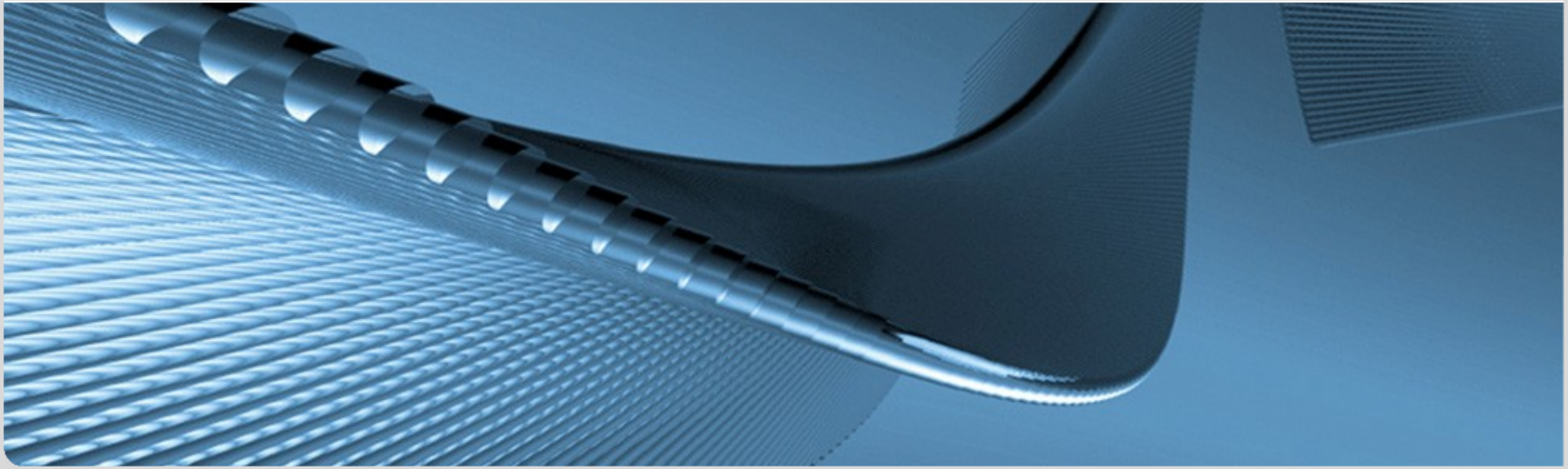
Im Folgenden

- Zentralisierte Verfahren
  - CliqueCloak
  - Mix-Zones
- Dezentrale Verfahren
  - Routing Through the Mist
  - Peer-to-Peer Anonymisierung
- Verfahren zur Selbstanonymisierung
  - Anonymisierung durch Dummies

Karlsruher Institut für Technologie

Zentralisierte Datenschutzansätze für LBS

IPD, Systeme der Informationsverwaltung, Nachwuchsgruppe „Privacy Awareness in Information Systems“



CliqueCloak

- Zentralisiertes Verfahren, das zwischen Nutzer und Dienstanbieter geschaltet wird (Thrusted Third Party)
- Algorithmus zum effizienten anonymisieren
- Eingaben:
  - *Positionen* (Anfragen) vieler Teilnehmer
  - *Minimaler Grad an Anonymität*,
    d.h., mindest-k für k-Anonymity
  - *Minimale räumlich-zeitliche Auflösung*, die an einen Dienst weitergegeben wird
- Ausgabe:
  - Spatio-temporal Cloaking Box (SCB) mit k Nutzern

B. Gedik, L. Liu: *Location Privacy in Mobile Systems: A Personalized Anonymization Model.* ICSCS'05

Anforderungen an CliqueCloak

- Spatial Containment
  - SCB muss die Position des Nutzers beinhalten
- Spatio-temporal Resolution
  - SCB darf nicht größer sein als vom Nutzer vorgegeben → Dienstgüte
- Content Preservation
  - der Inhalt einer Anfrage oder Eingabe darf nicht verändert werden

Ansatz

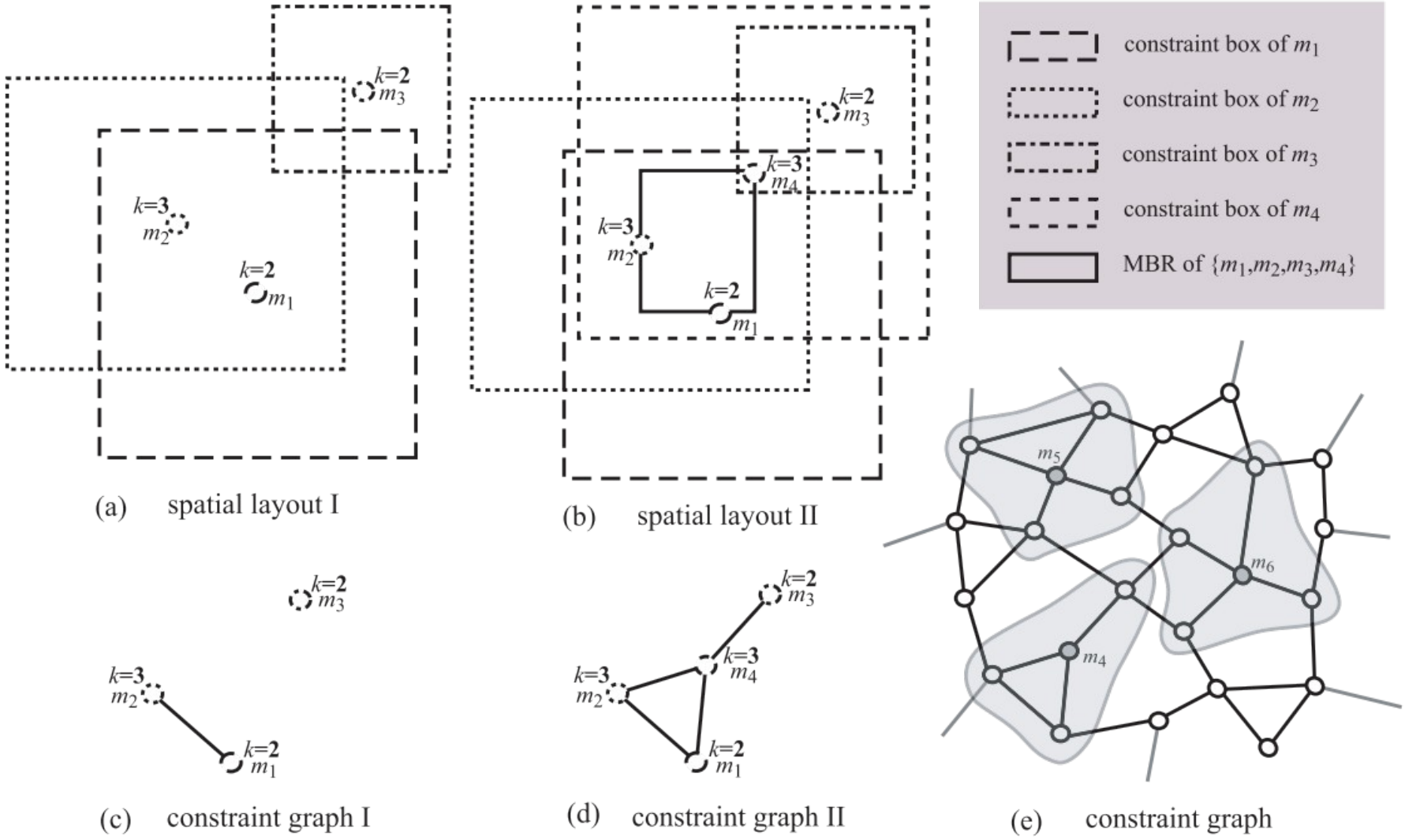
- Menge von Nachrichten m = <u_{id}, r_{no}, {t,x,y},k,{d_t,d_x,d_y},C>
  - C: Inhalt der Nachricht
  - u_{id}, r_{no}: Nutzer-ID, Nachrichten-ID
  - {t,x,y}: Zeit, Koordinaten
  - {d_t,d_x,d_y}: Toleranzen für die Constraint Box
    - Rückgabe soll nicht mehr als +/- d/2 vom tatsächlichen Wert abweichen
  - k: minimale Zahl von Personen in Constraint Box

- Ziel: *Finde in der Menge der Nachrichten die Minimale Constraint Box für jeden Nutzer*

Algorithmus

- füge jede Nachricht als Knoten in einen *Constraint-Graphen* ein
  - Kante von Knoten A zu Knoten B, wenn A innerhalb der Toleranzgrenzen $\{d_t, d_x, d_y\}$ von B liegt oder umgekehrt
- für jede Nachricht
  - suche im Constraint-Graphen eine Clique mit Kardinalität k
  - entferne die Clique aus dem Graphen
  - ersetze {t,x,y} durch das Minimum Bounding Rectangle der Clique
- wenn d_t abgelaufen bevor Clique gefunden, informiere Nutzer und lösche Knoten aus Constraint Graph

Beispiel


k=2
m3
k=3
m2
k=2
m1
(a) spatial layout I
k=2
m3
k=3
m4
k=3
m2
k=2
m1
(b) spatial layout II
constraint box of m1
constraint box of m2
constraint box of m3
constraint box of m4
MBR of {m1,m2,m3,m4}
k=2
m3
k=3
m2
k=2
m1
(c) constraint graph I
k=2
m3
k=3
m4
k=3
m2
k=2
m1
(d) constraint graph II
m5
m6
m4
(e) constraint graph

Kritik

- Feingranulare Einstellung der Privatheit

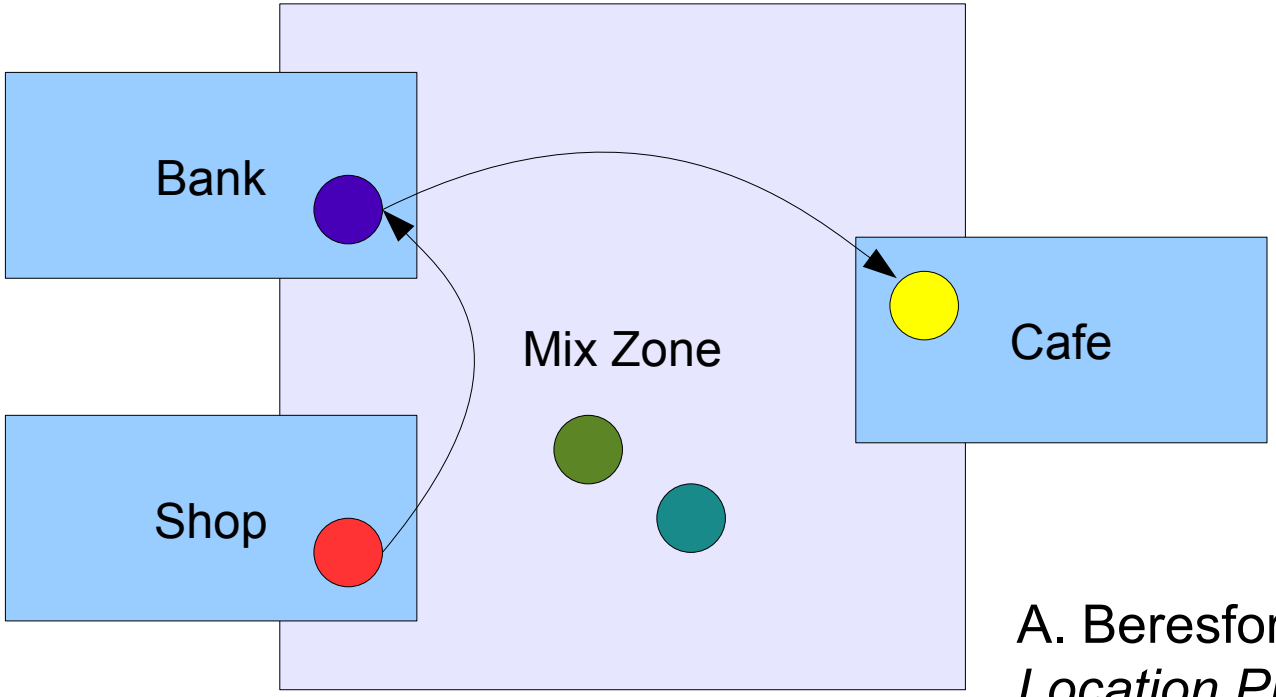
- Single Point of Failure
- Cliquen-Suche ist sehr teuer
  - Verfahren skaliert schlecht
- bekannte Schwächen der k-Anonymität
  - z.B. wenn alle k Nutzer an exakt der gleichen Position
- Parametrisierung erfordert genaue Kenntnis der eigenen Privatheitsbedürfnisse → schwierig
- bei geringer Nutzerdichte Kompromisse bei k oder $\{d_t, d_x, d_y\}$ erforderlich
- Funktioniert nicht bei Anfragesequenzen

Mix Zones

- Szenario: Anfragesequenzen, pseudonym nutzbare LBS
  - Beobachtung: Anonymisierung nützt nichts, wenn der Nutzer z.B. bis zu seinem Haus oder seiner Arbeitsstelle nachverfolgt werden kann
- Idee:
  - Positionsdaten werden nicht kontinuierlich, sondern mit zeitlichem Abstand (Update Period)
  - zwischen den Updates wechseln Nutzer ihr Pseudonym in sogenannten *Mix Zones*
    - viele pseudonyme Personen gehen hinein, andere pseudonyme Personen gehen heraus
  - Nutzer interagieren mit Diensten in *Application Zones*

Beispiel

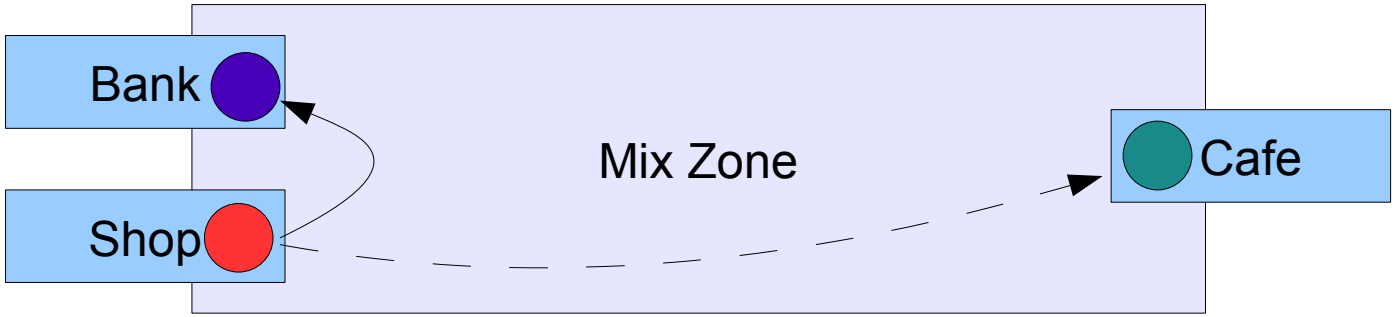
- Application Zones: Shop, Bank, Cafe



A. Beresford, F. Stajano
Location Privacy in Pervasive Computing,
Pervasive 2003

Effektivität des Konzepts

- Abhängig von der Update Period
  - wenn Updates von Positionen in zu rascher Folge, können Nutzer Mix Zone nicht durchqueren, bleiben trotz anderem Pseudonym verfolgbar
- Abhängig von der Raumaufteilung
  - bei ungünstiger Ausgestaltung Wechsel der Application Area zwischen Updates unmöglich



Kritik

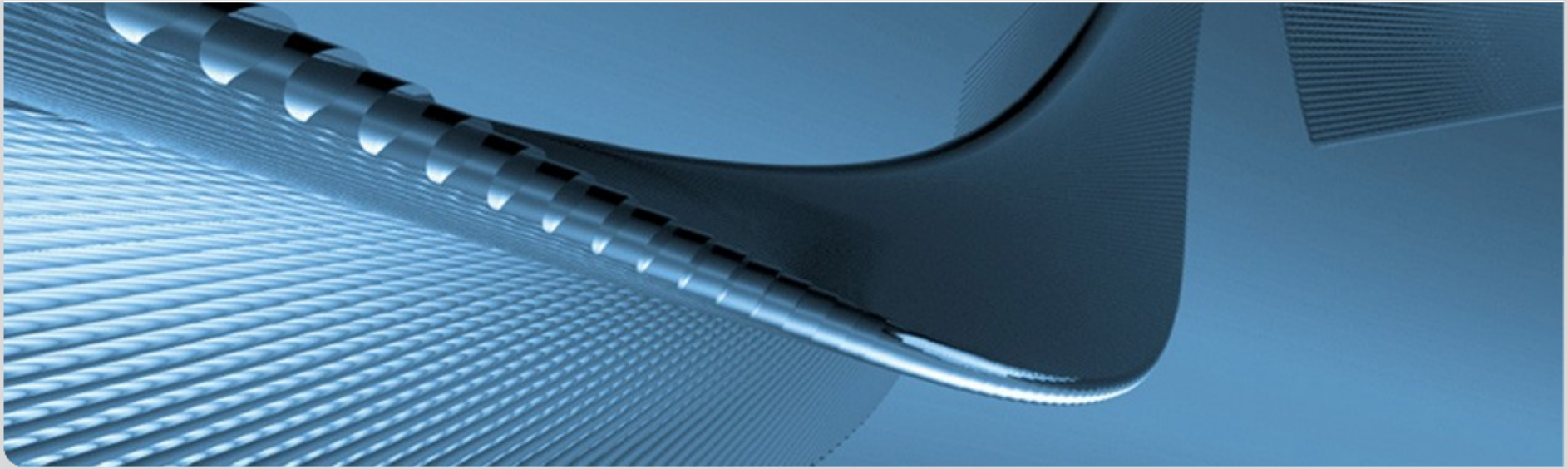
- funktioniert auch mit Anfragefolgen
- Skalierbar, wenig Rechenleistung und Kommunikation
- einfaches, verständliches Verfahren

- Single Point of Failure
- Starke Abhängigkeit von externen Faktoren
  - Raumaufteilung
  - Anzahl und Bewegung der Nutzer
  - Update Period
- statistische Angriffe möglich
  - wenn es unwahrscheinlich ist dass jemand bestimmte Orte nacheinander aufsucht
  - wenn Gehgeschwindigkeit der Personen bekannt

Karlsruher Institut für Technologie

Dezentrale Datenschutzansätze für LBS

IPD, Systeme der Informationsverwaltung, Nachwuchsgruppe „Privacy Awareness in Information Systems“

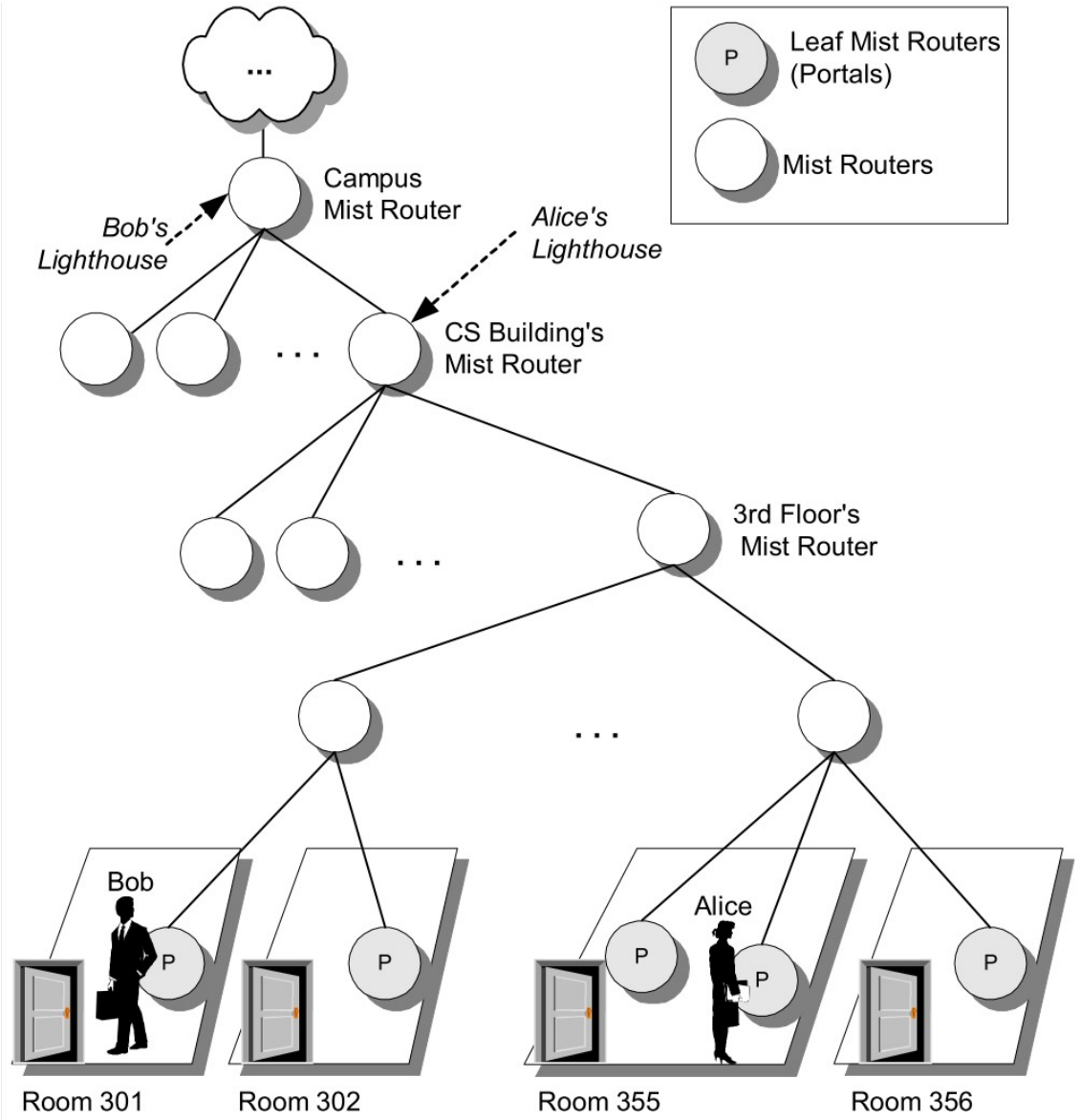


Routing Through the Mist

- “Durch den Nebel routen”: Ansatz zielt darauf ab, dass z.B. Mobilfunkstationen nicht wissen sollen, mit welchem Nutzer sie kommunizieren, aber wissen müssen wo er ist

- Hierarchie (Baum) von Routern
  - Blätter: einzelne konkrete Orte
  - Knoten: fassen Orte zusammen, z.B. Stadtteile, Städte, Länder
- Eingabe: Wahl eines Routers als “Leuchtturm”, der mit den eigentlichen Diensten kommuniziert
  - Leuchtturm tief im Baum: genauere Position
  - Leuchtturm weiter oben: gröbere Position
- Trennung von Position und Identität

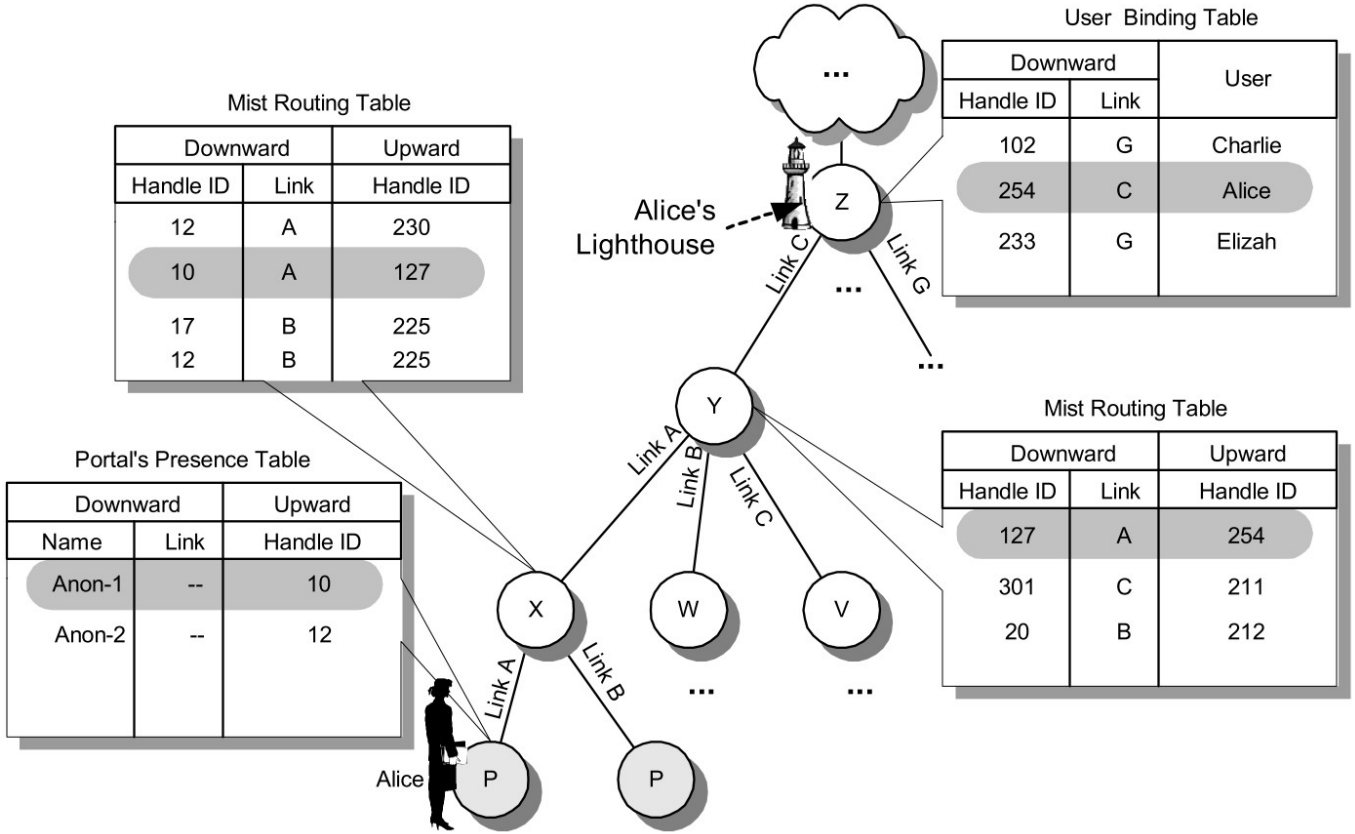
Beispiel



Quelle: Routing Through The Mist Al-Muhtad et al.

Trennung zwischen Identität und Position

- anonyme Kommunikation, Handle-IDs und Link-IDs
  - Portale kennen Nutzerposition, aber nicht Nutzer
  - Leuchttürme kennen Nutzer, aber nur eigene Position



Quelle: Routing Through The Mist Al-Muhtad et al.

Mobile Nutzer

- Portale
  - können leicht automatisch gewechselt werden
    - z.B. wenn Nutzer Raum verlässt, für den der Router zuständig war
- Leuchttürme
  - müssen manuell gewählt werden, wenn Nutzer Zuständigkeitsbereich des Leuchtturms verlässt
    - Wahl des Leuchtturms entscheidet über Anonymisierungsgrad

Kritik

- gut skalierbar, kaum zusätzlicher Rechenaufwand
- einfaches, verständliches Verfahren

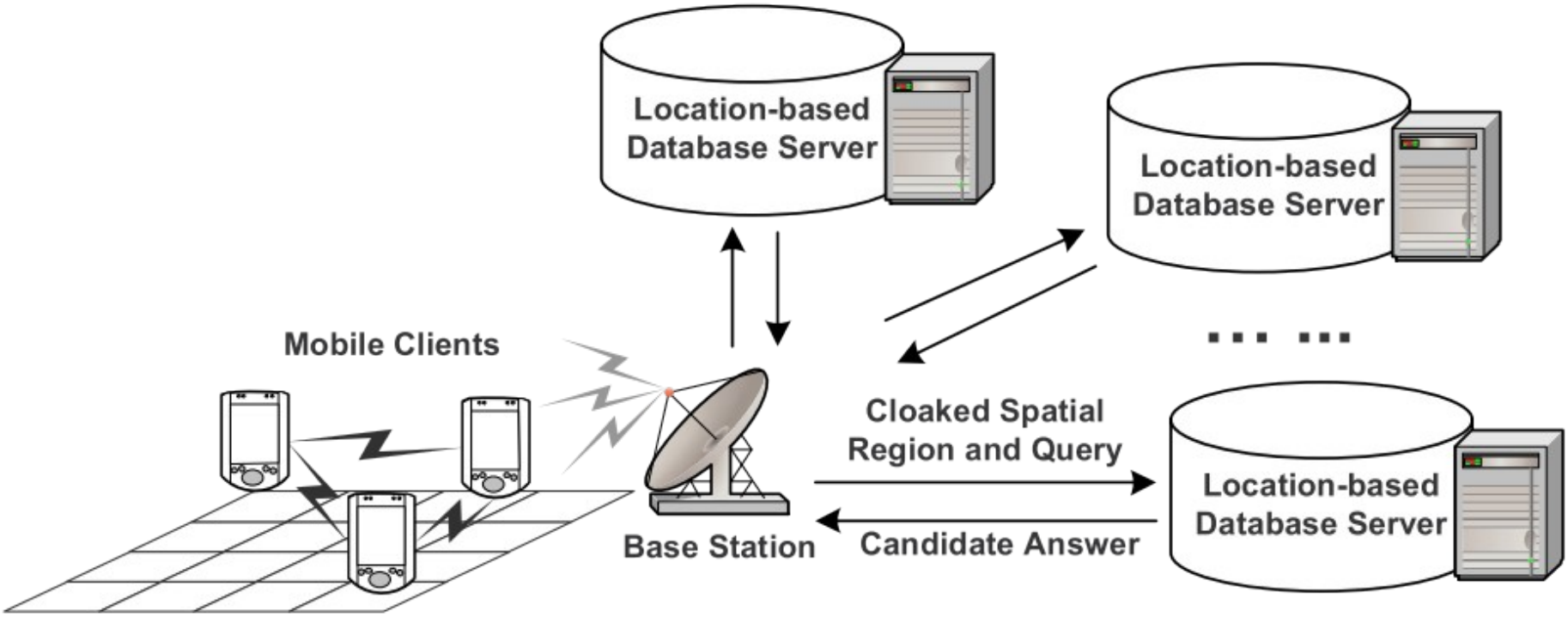
- Portale und Leuchttürme müssen vertrauenswürdig sein
  - dürfen Wissen nicht zusammenführen
  - Router egal: verschlüsselte Kommunikation
- Leuchtturmdichte ist entscheidend, Nutzerdichte ist egal
  - zuwenig Leuchttürme: unnötig grobe Positionsmeldungen
- erhöhter Kommunikationsoverhead
- Anfragefolgen generieren aussagekräftige Tracks, wenn Leuchttürme öfter gewechselt werden müssen

Peer-to-Peer Anonymisierung

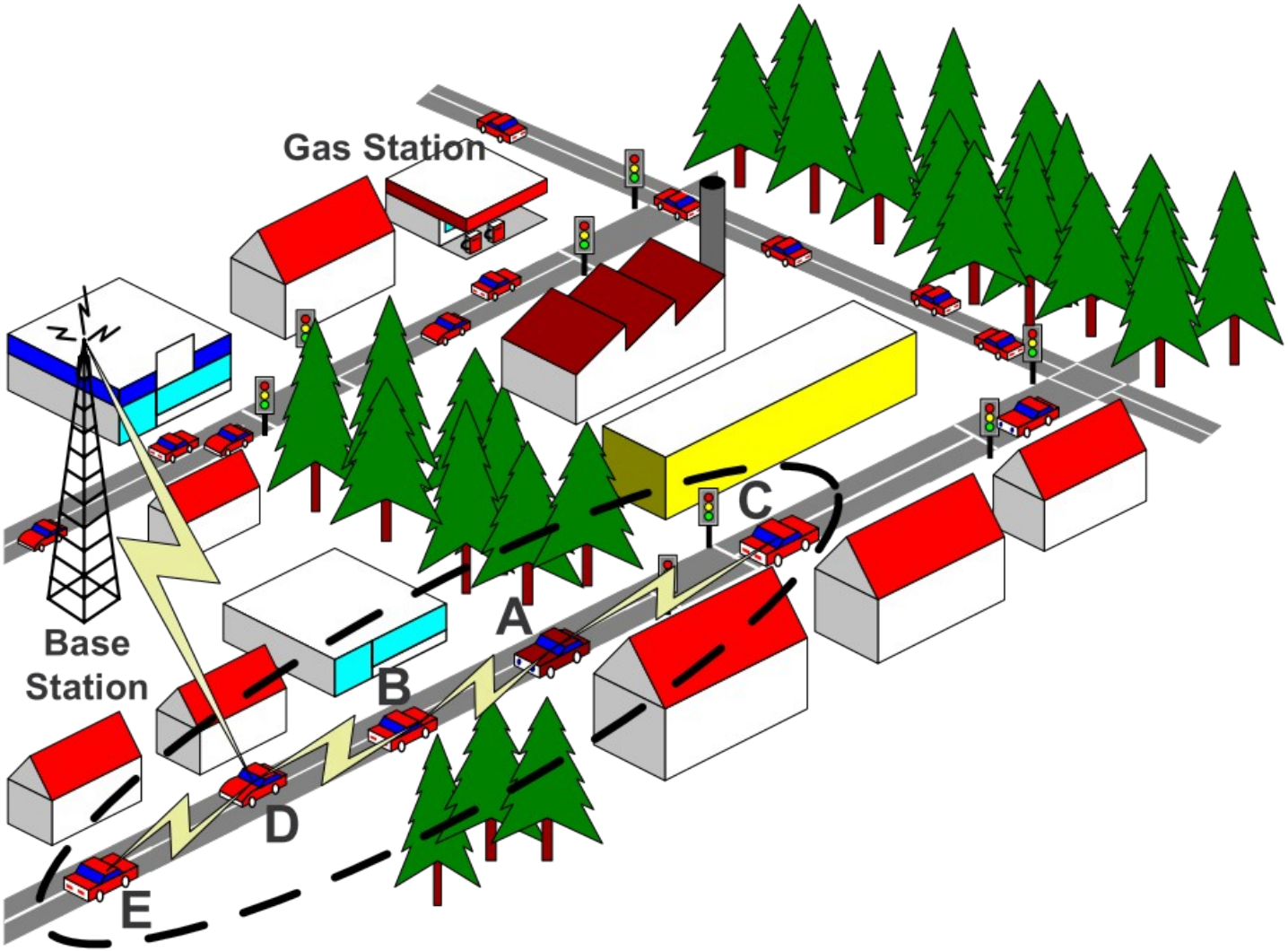
- Szenario: mobile Anwender, die über Ad-Hoc-Netze kommunizieren und Geodaten an Dienste schicken
- Gruppen von Nutzern annonymisieren ihre Positionsdaten gegenseitig
  - Anonymisierungsmenge besteht aus den Nutzern, die grade in Funkreichweite sind
  - Dienst erhält bereits anonymisierte Daten, keine vertrauenswürdige 3. Partei erforderlich
- Eingabe:
  - k für Größe der Anonymisierungsmenge
  - A_{min} als kleinstmögliche Anonymisierungsregion

C. Chow, M. Mokbel, X. Liu: *A peer-to-peer spatial cloaking algorithm for anonymous location-based service*. GIS 2006

Systemarchitektur

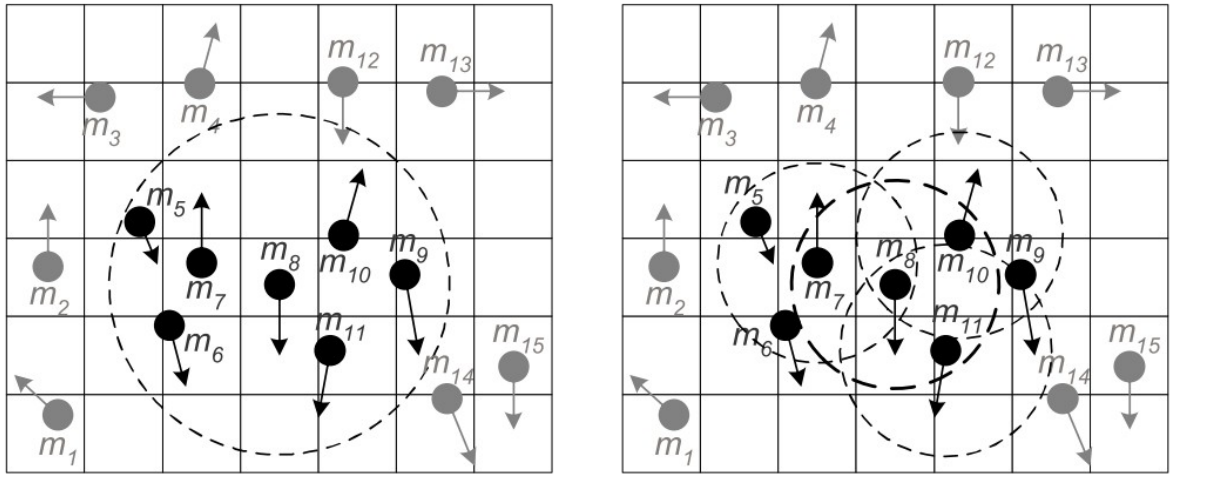


Beispiel



Verfahren (1/2)

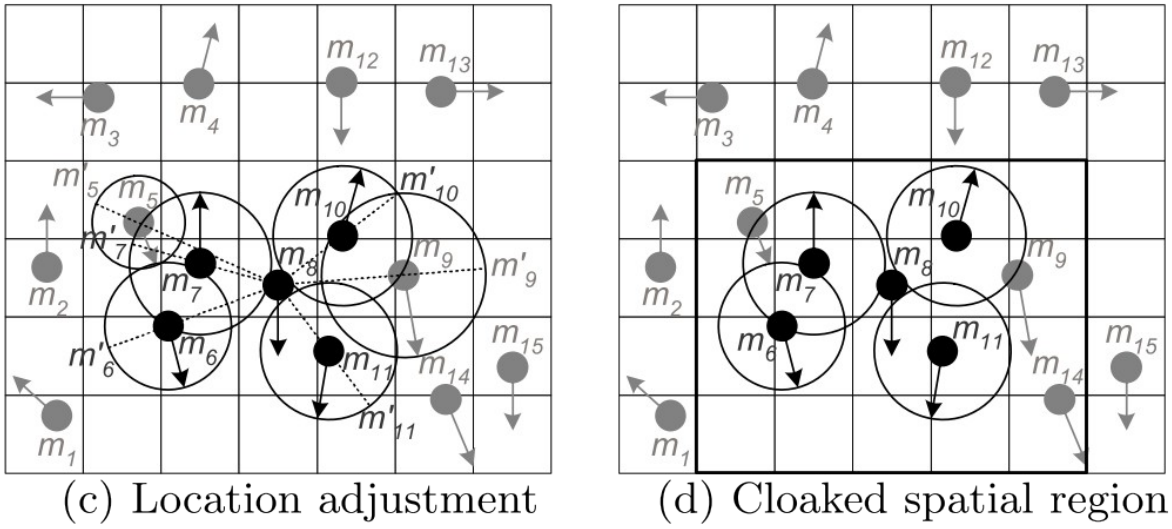
- Suche Teilnehmer in direkter Funkreichweite (Abb. a)
- Berechne Bewegungsvektoren, um Teilnehmer zu finden die in die gleiche Richtung unterwegs sind
  - Overhead durch Neuberechnung sparen
- Wenn weniger als k-1 Teilnehmer gefunden werden, suche multi-hop (Abb. b)



(a) Single-hop peer searching (b) Multi-hop peer searching

Verfahren (2/2)

- Wenn k-1 Teilnehmer gefunden, bilde Anonymisierungsregion (Rechteck in Abb. d)
  - wähle Region so, dass kein Teilnehmer den Bereich während der Anfrage verlässt, selbst wenn er sich mit Maximalgeschwindigkeit (Kreise in Abb. c) direkt vom Anfrager entfernt (gestrichelte Linien in Abb. c)



(c) Location adjustment (d) Cloaked spatial region

Kritik

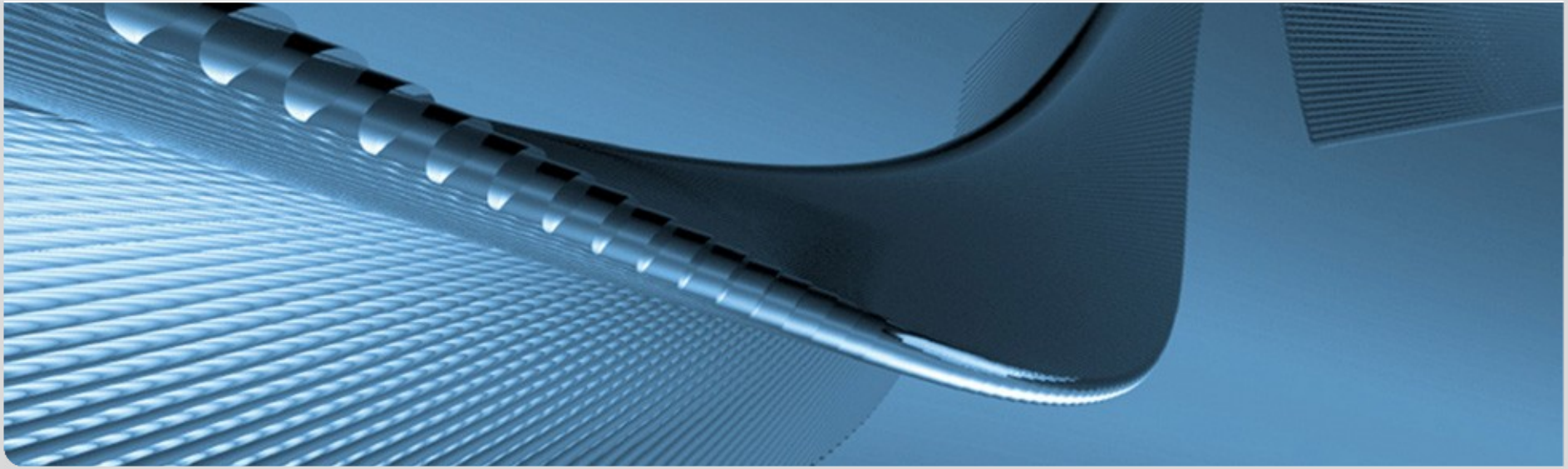
- dezentrales Verfahren, kein Single Point of Failure
- skaliert gut, da durch jeden Peer neue Ressourcen hinzukommen
- intuitiver Ansatz

- hoher Kommunikationsaufwand
- Peer Group muss vertrauenswürdig sein
  (andererseits muss sich Angreifer beim Opfer befinden, kennt dessen Position also sowieso)
- Nutzerdichte ist problematisch, k Teilnehmer in direkter Funkreichweite erforderlich
- Wenn sich die Nutzer oft auseinanderbewegen, können einzelne Teilnehmer durch Anfragesequenzen identifiziert werden

Karlsruher Institut für Technologie

Selbstanonymisierungsverfahren für LBS

IPD, Systeme der Informationsverwaltung, Nachwuchsgruppe „Privacy Awareness in Information Systems“



Anonymisierung durch Dummies

- Szenario: Sequenzen von Geo-Anfragen an einen Dienst
- Idee: Selbstanonymisierung durch das Generieren von k-1 'gefälschten' Nutzern
  - Dienst erhält k unterschiedliche Koordinaten
    - eine echt, k-1 künstlich generiert
  - Dienst antwortet mit k Ergebnissen
    - nur Anwender weiß, welche Koordinate 'echt' ist, und kann das korrekte Anfrageergebnis
- Problem: realistische Dummies

H. Kido et al., *An Anonymous Communication Technique using Dummies for Location-based Services*, ICPS'05

Realistische Dummies

- Dummies nicht von echten Daten unterscheidbar
  - Autos fahren nicht, wo keine Straßen sind
  - realistische Bewegungsprofile

- *Moving in a Neighborhood*
  - nächste Position eines Dummies in Nachbarschaft zuvoriger Position
- *Moving in a Limited Neighborhood*
  - nächste Position eines Dummies in Nachbarschaft zuvoriger Position
  - zusätzlich: Dummies, die in einem zu dicht besetzten oder ungeeigneten Gebiet generiert werden, werden verworfen und durch neue ersetzt

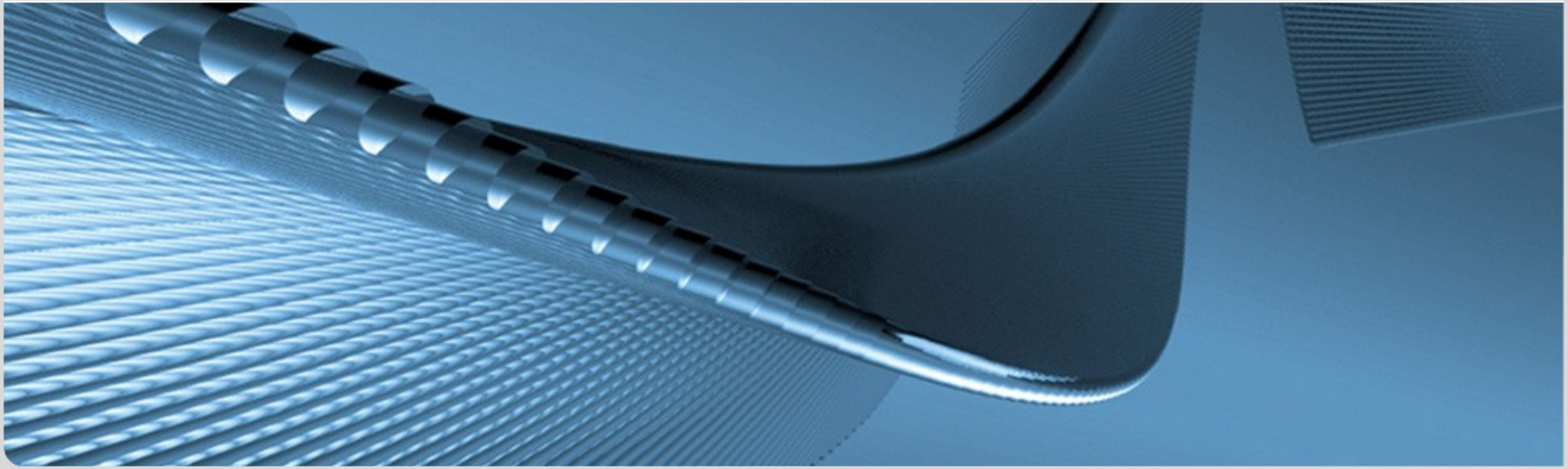
Kritik

- verteiltes Verfahren, da von jedem Nutzer selbst durchzuführen
- funktioniert auch mit Anfragefolgen
- skaliert linear, da nur k-1 zusätzliche Anfragen je Nutzer gestellt werden

- hoher Rechen- und Kommunikationsoverhead auf dem Mobilgerät
- sehr schwierig, realistische Dummies zu erzeugen die sich nicht statistisch herausfiltern lassen
  → nicht intuitiv *sicher* anwendbar

Karlsruher Institut für Technologie

Abschluss



Zusammenfassung

- Lokationsbasierte Dienste sind eine der attraktivsten Ubiquitous-Computing-Anwendungen
  - Aber: Datenschutzprobleme unvermeidlich
    → Kompromiss zwischen Privatheit und Dienstgüte
- Besonders problematisch: Tracks und Anfragefolgen, die das Erstellen von Tracks erlauben
  - Nutzer lassen sich 'herausmitteln', wenn jede Anfrage ein anderes Anonymity Set generiert
- Vorgestellte Verfahren
  - zentrale Ansätze
  - dezentrale Ansätze
  - Methode zum Datenselbstschutz

Literatur

[1] Jürgen Czerny, *Datenschutz in lokationsbasierten Diensten,* Seminararbeit am IPD, 2007
http://dbis.ipd.uni-karlsruhe.de/img/content/SS07Czerny_DatenschutzLBS.pdf