

# Diplomarbeitsthemen

## Simulation

Beim deutschen LKW-Mautsystem werden Positionen der LKWs per GPS-Technik ermittelt. Die sog. On-Board-Unit (OBU, oder Client) berechnet aus der aktuellen Position anhand von gespeicherten kartographischen Informationen gemäß einem komplexen Regelwerk die zu zahlende Maut. Die Summe wird per Mobilfunktechnik an ein Computerzentrum übermittelt. Zur korrekten Berechnung muss die Karteninformation regelmäßig aktualisiert werden, dies geschieht ebenfalls für die gesamte Flotte per Mobilfunktechnik. Außerdem besteht über dieselben Kanäle die Möglichkeit die Software der OBUs zu aktualisieren um Systempflege betreiben zu können. Zusätzliche Funktionen und Änderungen der Kommunikationsstruktur sind denkbar.

Zu den beschriebenen Aufgaben ergeben sich verschiedene Themen für Diplomarbeiten, die kombiniert oder gesplittet werden können, je nach Interesse und Fähigkeit des Kandidaten.

Die Simulationstechnik basiert auf Multi-agenten und Discrete-Event. Es soll in erster Linie das hausinterne Modell (ML Designer) erweitert werden. Gleichzeitig soll auch freie Software (Matsim) auf den betrieblichen Einsatz getestet werden.

Ein Themenblock befasst sich mit der Aktualisierung/dem Update-Prozess (Remote Device Management), ein weiterer mit Erhebungsverhalten (Einsammeln der Maut). Bisher ist der Input der event-getriebenen Simulation ein statistisches Modell der Kommunikationsstruktur, das aus der TC Datenbank erstellt wurde. Deterministische Modelle sind von hohem Interesse (basierend auf ökonomischen/logistischen Modellen) und sollten entwickelt werden. Weitere Themen, s.u.

### **1. Update: GUI**

Zum besseren betrieblichen Einsatz soll mit MLDesigner ein GUI mit den nötigen Analysefunktionen und einer verbesserten Ausgabe der Ergebnisse zum Import in die hausinterne Datenbank entwickelt werden. Eine Mitentwicklung der Anforderungen (UML) kann ein Bestandteil der Diplomarbeit sein.

### **2. Variation Update-Szenarien**

Das bei Toll Collect entwickelte Modell spiegelt den Ist-Zustand des Systems wider. Für zukünftige Szenarien und Systemverbesserungen sollen verschiedene Richtungen der Weiterentwicklung bzw. Änderung des Update-Prozesses entwickelt und untersucht werden. Beispielsweise kann an eine Änderung der Mobilfunk-Bandbreite (GSM/UMTS) oder eine genaue Untersuchung der verwendeten Protokolle gedacht werden. Weiterhin ist es von Interesse, die Auswirkungen von konzeptionellen Änderungen, wie serverseitiger Steuerung der Aktualisierung statt Client-seitiger Steuerung zu studieren.

### **3. Verkehrssimulation (Statistischer Input)**

Im Modell des Update-Prozesses ist keinerlei Information über die tatsächliche Fahrt einer OBU enthalten, sondern nur die jeweiligen An/Aus-Zeiten, da nur diese relevant sind.

Für eine allgemeine Simulation des Verkehrsaufkommens ist es aber wichtig, auch das räumliche Fahrprofil von LKWs zu modellieren.

Im ersten Schritt werden anhand der TC Datenbank typische Autobahn-Belastungen statistisch generiert und dann in einem Simulationsmodell umgesetzt. Dazu kann auf vorhandene Modelle zurückgegriffen werden (MATSIM, etc.)

Im zweiten Schritt wird das vorhandene Modul zur Aktualisierung eingebaut, um einen Konsistenztest durchführen zu können.

Im dritten Schritt wird das sog. Erhebungsverhalten modelliert. Dabei handelt es sich um die Simulation der Kommunikation Client-Server im Rahmen der Übermittlung der zu zahlenden Maut. Das angewandte Regelwerk ist recht komplex, so dass eine realitätsnahe Modellierung eine anspruchsvolle Aufgabe ist. Hier gibt es Kopplungen und Wechselwirkungen mit der Aktualisierung, die zu nichttrivialen Systemzuständen führen können.

In einem weiteren Schritt wird die Reaktion der OBUs auf Änderungen bzw. Störungen der GPS-Konstellation modelliert. Es kann hier auf das SatLab-tool von MLDesigner zurückgegriffen werden oder bei Bedarf eine eigene, spezialisierte Bibliothek entwickelt werden.

Schließlich kann die Tiefe der Modellierung der zentralen Systeme mit den Prozessen, die an den Eingang von Nachrichten von den Clients angeschlossen sind, weiter ausgebaut werden. Hier handelt es sich um Server und Datenbanken, die die eingehenden Nachrichten weiterverarbeiten, wie zum Beispiel Rechnungserstellung und Analyse bzw. Systemüberwachung.

Ein sehr reizvolles Thema beschäftigt sich mit der vollen Modellierung des LKW Verkehrs anhand von Regeln zur Lieferung und Wechselwirkung mit anderen Verkehrsteilnehmern. Hier ist grundsätzliche Forschung nötig mit Sichtung aktueller Arbeiten, Erstellen entsprechender Regeln, und räumlicher Simulation auch auf den untergeordneten Straßen (Nicht-BAB).

## Data Mining

### 4. Verkehrsaufkommen: Einzelstrecken-Statistiken, Fahrtzeiten

Aus der TC Data Base können anhand von anonymisierten Fahrdaten statistische Auswertungen zum Verkehrsaufkommen bzw. dem Fahrverhalten der LKWs vorgenommen werden. Insbesondere ist es betrieblich wichtig eine automatisierte Pflege von Verteilungen und Kenngrößen zu betreiben. Dies ist ein Ziel der Arbeit.

Bisherige Auswertungen erfassen Cluster von Einzelfahrten. Eine Fahrt wird charakterisiert durch Anfang und Dauer. Zusätzlich kann die Dauer der Pause zwischen zwei Fahrten zur Charakterisierung hinzugezogen werden.

Der nächste wichtige Schritt besteht in der Bestimmung von bedingten Wahrscheinlichkeiten: Wie hoch ist die Wahrscheinlichkeit, dass eine OBU am Tag  $d$  zur Zeit  $t$  aktiv wird, *unter der Bedingung*, dass sie vorher am Tag  $d_0$  zur Zeit  $t_0$  aktiv wurde, und zu einer Zeit kleiner als  $d$  ausgeschaltet wurde.

## **5. Verkehrsstatistik: Quelle-Zielverteilung**

Aus den Leistungsdaten der TC Data Base, die auch zum Billing genutzt werden kann die Quelle-Zielverteilung des (12 t ) LKW-Verkehrs bestimmt werden. Details wie die entfernungsabhängige Zieldichte sollen ebenfalls bestimmt werden. Die statistisch saubere Aufarbeitung der Daten und ein Vergleich mit analytischen Modellen soll erreicht werden. Es ist wichtig, Verfahren zu entwickeln, die eine Korrektur der Daten im Bezug auf systematische Fehler beinhalten. Das beinhaltet u.U. das Einbeziehen von Mustererkennungsverfahren, wie sie auch schon bei TC im Einsatz sind.

## **6. Verkehrsstatistik: Kachelsystem**

Eine Möglichkeit das Datenaufkommen im Rahmen des Aktualisierungsprozesses gering zu halten besteht darin, den Fahrzeugen individuelle kartografische Informationen zur Verfügung zu stellen. Dabei kann das Maut-Betreibergebiet in sog. Kacheln zerlegt werden. Diese Kacheln sollten von den OBU's „nach Bedarf“ geladen werden. Anhand der Quelle-Zielverteilung und den Spezifikationen zu den Kacheln wird eine optimale Kachelgröße bestimmt. Das beinhaltet das Aufstellen einer Kostenfunktion, die von der Kachelgröße abhängt und das Suchen deren Minimums unter Variation der Kachelgröße.

## **7. Distanzmatrix**

Es wird eine Pflege der Fahrtzeiten und –geschwindigkeiten der einzelnen Abschnitte, die im TC System genutzt werden angestrebt. Hierzu kann auf verschiedene Datenquellen zurückgegriffen werden. Es muss also ein Verfahren entwickelt werden, das die Daten automatisch verarbeitet und einen –evtl. Tages- und tageszeitabhängige Geschwindigkeit auf dem Autobahnnetz zu bestimmen. Fehlerrechnung und statistische Auswertung von Verteilungen ist beinhaltet. Als Nebenprodukt könnten Rückschlüsse auf Staus auf der Autobahn gezogen werden.

## **8. Update und Quelle-Zielverteilung: Zeiten aus Limit- und Statusanfragen, Korrelation mit Summendatentöpfen**

Um die Fahrprofile, die aus den Leistungsdaten bestimmt werden, zu schärfen und zu erweitern können die nicht ortsbezogenen Limit- und Statusanfragen zu Hilfe gezogen werden. Das Bestimmen der Zeiten fern der Autobahn und die eindeutige Zuordnung der Kommunikation ist nicht trivial, aus den entsprechenden Daten sollen Statistiken generiert werden, die zur Optimierung des Update-Prozesses benutzt werden. Aus den Zeiten kann mit Hilfe von Geschwindigkeiten eine maximale Entfernung von der Auffahrt auf die BAB bestimmt werden. Daraus wiederum kann die Quelle Ziel Verteilung verbessert werden für Orte außerhalb der BAB.

## **9. Fahrspuren: Map matching und Fahrspur-Ueberlapp**

Für die kartografische Modellierung der vermuteten Straßen ist es wichtig zeitlich hochaktuelle Daten zu erheben. Dazu können die entsprechenden Strecken abgefahren werden.

Befahrungen mit GPS-Positionsbestimmung unterliegen naturgemäß Schwankungen in der Positionsgenauigkeit.

Im ersten Schritt wird die Basis der existierenden Methoden recherchiert und dargestellt.

Im zweiten Schritt soll ein Verfahren entwickelt werden, das ein ortsaufgelöstes Maß und ein globales Maß für die Übereinstimmung von einer gemessenen Fahrspur und einem vorgegebenen Verlauf liefert. Dazu muss aus den Punkten der Fahrspur-Vermessung ein Graph (mit Fehlerschlauch) erzeugt werden, der mit dem vorgegebenen Verlauf verglichen wird. Verschiedene statistische Methoden dazu sollten geprüft werden.

Im dritten Schritt wird darauf aufbauend ein Verfahren entwickelt, das 2 Fahrspuren vergleicht. Die erhöhte Schwierigkeit liegt nun im Bestimmen des Überlapps zweier Graphen mit „Fehlerschlauch“.

Auf den vorhergehenden Schritten aufbauend soll studiert werden wie aus einem Bündel von Fahrtverläufen ein mittlerer Straßenverlauf bestimmt wird. Die Genauigkeitsabschätzung ist dabei sehr wichtig.

Schließlich soll die letzte Studie darin münden, die Größe der Stichprobe zu bestimmen, die nötig ist um mit vorgegebener Genauigkeit einen Straßenverlauf bestimmen zu können.

## **Datenbanken (reine Informatik)**

Im System der Toll Collect gibt es verschiedene Datenquellen, die zur Analyse des Systemzustands, bzw. der Fahrzeugflotte dienen. Zum einen sind dies Quellen des Produktivsystems:

Kontrollbrücken, Summendatentöpfe und manuelle Kontrolldaten; zum anderen senden Fahrzeuge des sog. Fahrbetriebs (das sind Fahrzeuge, die zur kartografischen Modellierung die Autobahnen abfahren) Streckendaten. Weiterhin senden Fahrzeuge, die mit modernen GPS-Empfängern ausgestattet sind Positions- und GPS-relevante Informationen. Schließlich existiert eine Flotte von Testfahrzeugen, deren OBUs zu Testzwecken weiterführende Informationen übertragen.

Aus jeder Datenquelle werden verschiedene Analysen abgeleitet. Aufgabe der Arbeit ist es eine Datenhaltung im Rahmen der TC Datenbank-Umgebung zu konzipieren, so dass Korrelationen der verschiedenen Quellen und konsistente Abfragen einfach möglich sind.

Weiterführend werden Mustererkennungsverfahren, die auf den verschiedenen Datenquellen arbeiten verfeinert und zusammengeführt. Typische Werkzeuge sind Regressions- und Korrelationsanalysen.

## **10. Analyse Datenbank und Design auf TC Zwecke**

Die TC-Datenbank besitzt eine spezifische Struktur, die auf die zeitnahe Verarbeitung von Massendaten mit anschließender Analyse abgestimmt ist. Zur Zeit ist eine Oracle Datenbank in Benutzung. Es ist von prinzipiellem Interesse zu untersuchen, ob dies das optimale design für den Nutzen von TC ist. Die Arbeit soll sich also damit befassen Vorschläge zur Optimierung des DB designs oder der Datenstruktur zu erarbeiten und ein Konzept zur Umsetzung vorzuschlagen.