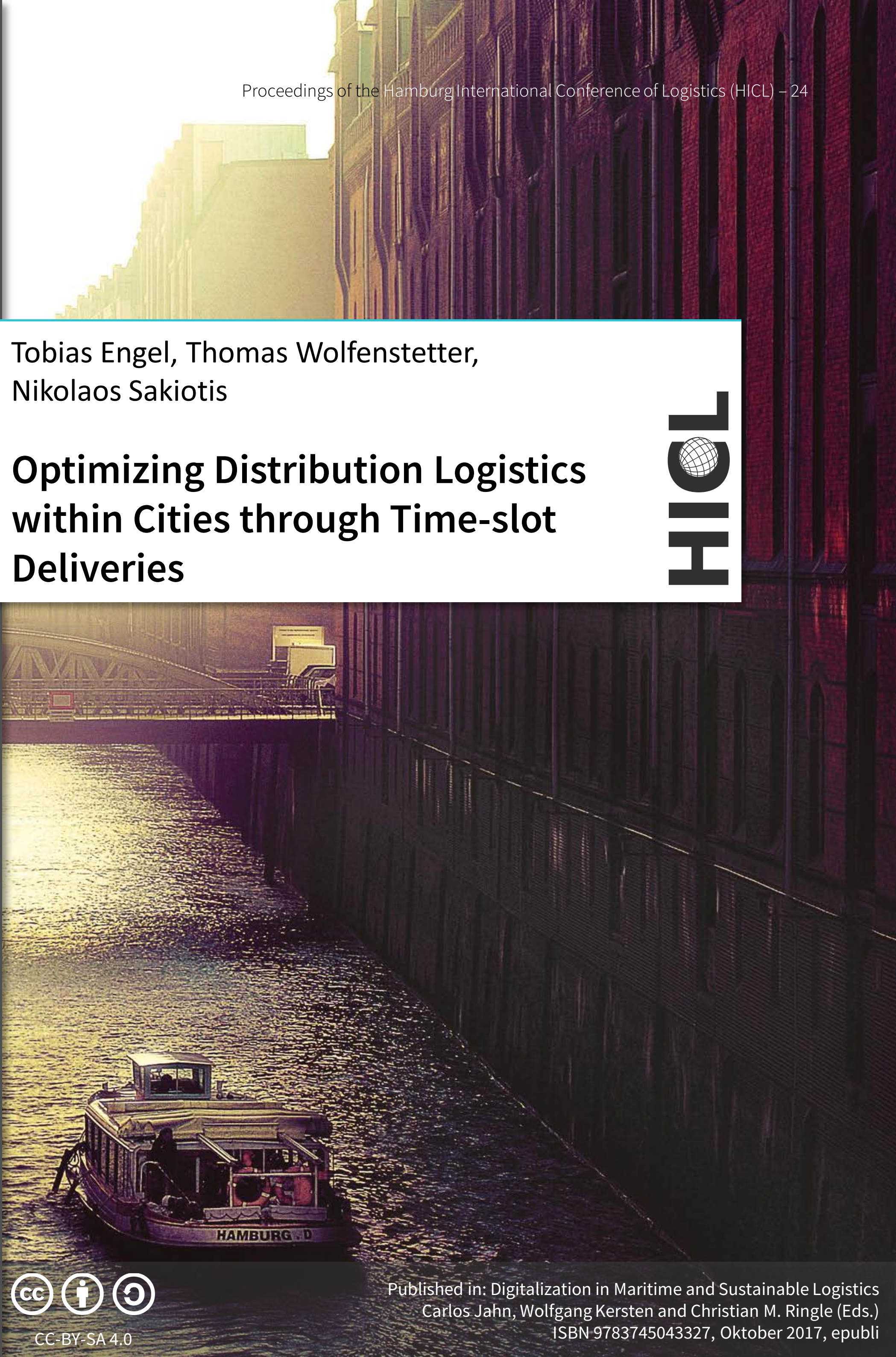
Proceedings of the Hamburg International Conference of Logistics (HICL) – 24

Tobias Engel, Thomas Wolfenstetter,
Nikolaos Sakiotis

Optimizing Distribution Logistics within Cities through Time-slot Deliveries



Published in: Digitalization in Maritime and Sustainable Logistics
Carlos Jahn, Wolfgang Kersten and Christian M. Ringle (Eds.)
ISBN 9783745043327, Oktober 2017, epubli

Optimizing Distribution Logistics within Cities through Time-slot Deliveries

Tobias Engel[1], Thomas Wolfenstetter[2], Nikolaos Sakiotis[3]

1 – University of Applied Science Neu-Ulm

2 – IUNO - Nat. Referenzprojekt - IT-Sicherheit in Industrie 4.0

3 – Capgemini Deutschland GmbH

Technological evolvement creates new business opportunities such as the optimization of distribution networks by using big data to calculate optimal delivery routes respecting individual time windows. The optimized network enhances firms to improve customer satisfaction, reduce delivery costs, and increase the utilization of the truck load by integrating reverse flows. Following the design science approach, we review and evaluate relevant data and scientific theories. Subsequently, requirements and system design are analyzed. On this basis we implement a prototypical information system to realize the optimized distribution network. The prototype consists of a mobile user-app and a back-office-system. Major functionalities of the prototype include the data analysis and the specification of settings for the desired delivery times and locations being actively influenced by the parcel recipients. The results of this research indicate that several tradeoffs have to be faced against the optimal use and analysis of available data, in order to provide a prototype that aims at improving customer satisfaction and reducing distribution costs at the same time. Hence, the focus has been laid on the analysis of dedicated, user-defined delivery time and location time windows by using a customized heuristic approach in order to decrease the complexity of the distribution process. We simulated the results, while a field study would enhance more insights. Practitioners can use the prototype to reduce their transportation costs to improve the utilization of the truck load or for new city services such as same day deliveries or as part of cyber-physical (distribution) systems (CPS).

Keywords: Distribution Logistics; Time-Slot Deliveries; City Logistics; Prototype

1 Introduction

Delivery of parcels has been part of everyday life of the population over the last decades. Especially due to the evolution of online retail activities over the last 15 year, millions of parcels are distributed worldwide (Fang and Zhang, 2005). However, it is commonly observed that there is a lack of coordination between the delivery schedules of the logistics companies and the schedules of the recipients (Engel, Sadovskyi et al., 2014). When a parcel is delivered, its recipient may sometimes not be present at the delivery location. This results from, either the recipient's unawareness about the delivery time in advance. However, even in a few cases, where logistics companies provide a 3-hour time range of estimated delivery in advance, the recipient may still not be able to be present at the delivery location and receive his parcel. In consequence, we have a need to optimize distribution networks (Shekhar, Gunturi et al., 2012, Valerio, D'Alconzo et al., 2009, Waller and Fawcett, 2013). This phenomenon is enhanced by the rapid increase of single-person households, as well as the amount of employed members in families (Nations, 2000).

This situation is primarily a problem for the logistics companies. Due to the heavy increase of parcels, ordered from online retailers, as well as the extra delivery attempts, the following problems have risen: Firstly, time and resources are wasted and distribution costs increase; trying to reach absent recipients and their neighbors subsequently. Additionally, the extra distribution routes lead to increased exhaust emissions, which are harmful for the environment. Furthermore, in order to decrease operational and distribution costs, the logistics companies provide their employees with insufficient working conditions and low salaries. Moreover, the most significant challenge for these companies, originating from the problematic situation described above, is, that their customer base lacks of customer satisfaction, as the current services that are offered do not meet their needs and requirements in the desired degree (Fang and Zhang, 2005, Weltevreden, 2008).

Despite all the problems, logistics companies have not made radical changes on their distribution processes. However, observing the evolution of online retail activities (Fang and Zhang, 2005) and the estimated increase of the internet users globally (Stats, 2013), these problems are expected to become more intense. Therefore, customer-oriented adjustments are needed. Customers need to have more delivery options considering their personal needs and schedule. They need high quality services, which let them choose whether their order is delivered at one of their addresses (work or home), at a parcel station, or even at a family member, a friend or a person of trust and moreover, the need to be able to combine those options for a single delivery, by specifying a time window for each one of them (Shekhar, Gunturi et al., 2012, Valerio, D'Alconzo et al., 2009, Waller and Fawcett, 2013). Such services would significantly contribute to the enhancement of customer satisfaction, which constitutes the major common goal of the supply chain partners, as well as to the optimization of distribution logistics processes.

Along with the evolvement of new technologies, such as real-time processing of huge data volumes new opportunities arise in the area of distribution logistics (Engel, Sadovskyi et al., 2014, Sadovskyi, Engel et al., 2014). The necessary data can be provided by the supply chain partners and originate from different sources (e.g. upstream and downstream inventory information, current and past geo-positions, traffic and weather information , historical data etc.). Moreover, focusing on the distribution logistics field, a customer-oriented approach is vital for any organization's success (Monczka, Handfield et al., 2008, Waller and Fawcett, 2013), considering the customer dynamism evolvement over the last years. Motivated by the steady increase of online retail activities (Fang and Zhang, 2005), we propose a novel approach as well as a prototypical solution to improve customer satisfaction avoiding unnecessary delivery attempts through real-time calculation of optimal delivery routes that take into account the current availability of the recipient.

The prototype allows firms to offer flexible, customer friendly, and effective services within the field of distribution logistics. Further, the optimization of distribution processes enhances firms to realize cost savings. Moreover, the prototype can be used as base for a distribution platform connecting various independent actors such as logistic distribution firms or taxis and smart things being equipped with sensors such as cyber-physical systems (CPS) creating a link to the Internet-of-Things. Using the prototype within a platform would enhance the society, e.g., to reduce gas emissions or improve the working conditions in the field of distribution logistics.

The paper is structured as follows: We briefly review related work as we base the prototype on an existing concept, followed by the conceptual approach – theoretical base for the prototype – and its development. In the fifth chapter we evaluate the prototype. Next we discuss the results, implications, and future research opportunities.

2 Related Work

A related concept has been described by Engel, Sadovskyi et al. (2014), proposing the development of a dedicated smartphone application, in order to use data from the customer's personal calendar and the GPS sensor, which operate on the smartphone, as source of his prospective location. Furthermore, the use of real-time information about weather, traffic and current location of the distribution vehicle is recommended (Waller and Fawcett, 2013). However, there is no optimization process until this last step of the supply chain, after the parcels have been loaded on the distribution vehicles. In order to provide a more effective solution for the problem described above, we have developed our prototype based on the work of Engel, Sadovskyi et al. (2014). The prototype aims at optimizing distribution logistics using big data, consisting of a dedicated smartphone application and a back-end system.

3 Conceptual Approach

Considering the proposed distribution process by Engel, Sadovskyi et al. (2014), the following data sources have to be used for smarter routing of distribution vehicles: Customer geo-location (current and past), distribution vehicle geo-location, schedule of the customer (smartphone calendar events), past purchases and deliveries historical data from customers, order priority data, and real–time weather and traffic data. For the development of the prototype, it is necessary to evaluate the data sources based on criteria like information security, complexity, and usability.

3.1 Customer current geo-location data

Customer current geo-location can be tracked from the app using the GPS sensor of the smartphone in the following format: a pair of decimal values, representing the current location's latitude and longitude. Considering that a user potentially faces privacy violation and information security issues by providing personal information and data, user's current geo-location should be optional and not be used in key functionalities like the initial route calculation or the creation of user's time windows.

However, the customer's geo-location can be used for the detection of a delivery plan divergence. The app will detect right before the delivery time, if the user is located further than a specified distance away from the delivery address, in case it would not be possible to cover this distance at the planned delivery time. Then the system will recalculate vehicle routes, as it is described in the concept and requirements of the prototype. In addition, customer current geo-location data should only be analyzed in the app, and not be stored or transferred to the server. Further, there are no data size or complexity issues that should be taken into consideration.

3.2 Customer past geo-location data

Distribution logistics could be further optimized by the analysis of customer past geo-location data. For example, the system can predict customer's future location, recognize patterns, and habits that could be used for route calculation along with user-defined time windows etc. using big data analysis techniques. While this allows firms to increase the efficiency, the complexity increases. In addition, customer past geo-location analysis would require permanent storage of this data raising the same information security issues as described above. As this would be in discrepancy with the decisions taken about customer current geo-location data, we neglected to use customer past geo-locations data at all.

3.3 Vehicle geo-location data

Geo-location data can be used for optimization being provided by the navigation system of the distribution vehicle's, e.g. its current location. This data can be analyzed in real time, as well as stored and aggregated for later analysis being beneficial for the estimation of the delivery time or route recalculation processes.

However, navigation systems are not provided in the context of this prototype, so there is no opportunity to integrate vehicle geo-location data. While the possibility of generating geo-location test data has been considered in order to integrate this functionality, it would be very difficult to assure that this data is realistic enough to meet evaluation requirements. In addition, it would make the testing phase too complicated, particularly in case of a real time geo-location analysis. That is the reason for our decision to exclude the vehicle geo-location data from the prototype. Despite that, the usage of geo-location data is recommended for the final product.

3.4 Smartphone calendar events data

In order to be able to create time windows for the route, the user's preferred delivery addresses and calendar events are needed being accessible via the customer's smartphone. However, there are plenty of reasons why smartphone calendar events' data do not meet evaluation criteria such as missing start and end time or the event's location. Finally, users are probably not willing to share their personal calendar events, due to information security and privacy violation issues. Therefore, we exclude the direct access to customer's calendar data from our prototype. Instead, we use a platform with an interface allowing firms to harmonize time-window wishes from customers with the distribution route. This allows us to ensure that every time window has a start and an end time, a delivery location is chosen, and the existence of uniform data format.

3.5 Customer past purchases and deliveries historical data

The next data category to be evaluated is order-specific data. Originating from their Data Warehouse and Customer Relationship Management (CRM) systems, companies in the distribution logistics branch have data available suitable for optimizing distribution logistic networks. As our prototype is not developed in collaboration with one of those companies, it is very complex to generate and analyze realistic such data in order to evaluate the effectiveness of the prototype.

As long as delivery time windows and addresses are exclusively created by the customers based on their schedule, there is no need for historical data integration in this process. In case of a delivery plan divergence, when the time windows data are not sufficient for a successful delivery, historical data will be used, regarding successful and unsuccessful deliveries in specific time slots in the past.

3.6 Order priority data

Due to mathematical constraints, there is a need for sorting the orders at several steps of the distribution algorithm (see chapter 4). Furthermore, orders that match certain situation-based criteria shall be given priority. For example, (1) the customer has paid an extra fee for an early/next day delivery service, (2) the order contains food, beverages, or medicines, (3) the order has already been unsuccessfully delivered, or (4) the order contains fragile products or products of high value. This allows firms to differentiate their business model according to their needs.

The values of this criteria are either Boolean or scalar integers and the total priority is calculated as a decimal number with the weighted arithmetic mean method. For the context of this research, decimal values between 0 and 5 where generated, with 5 indicating the highest priority and 0 the lowest.

3.7 Real–time weather and traffic data

Another idea for optimizing distribution logistics is to include real-time weather and traffic data into the route calculation process. By that, streets with high current traffic, due to bad weather conditions, accidents, or roads under construction, may be avoided; or at least delivery delays can be detected, estimated and taken into consideration. However, there is always the challenge to balance the need for computation power and needed complexity when using such data.

Google Maps API supports real-time weather and traffic data, but not in the free edition used for developing this prototype (Google, 2014). This fact prohibited the use of such data. However, if the prototype is further developed in the future in order to be launched in the market, there will be the need to use the commercial version of Google Maps API, where real–time weather and traffic data is fully integrated.

Table 1 1 provides and overview of used data and the source of the data proposing its relevance for the prototype.

3.8 Travelling Salesman Problem

For our distribution problem, we use the Travelling Salesman Problem (TSP) allowing us to optimize the route (Domschke and Scholl, 2010).

In our case, nodes represent the delivery address of the customer, while c_{ij} represent the travelling time between two addresses, as the overall goal of the

Table 1: Overview of type and source of data used for delivery optimization

Data	Source
Customer current geo-location data	Smartphone
Customer past geo-location data	Smartphone
Vehicle geo-location data	Navigation system
Smartphone calendar events data	Smartphone
Customer past purchases and deliveries historical data	CRM system
Order priority data	ERP system
Real–time weather and traffic data	External platform provider

prototype is maximization of customer satisfaction (if cost reduction through route optimization was the main goal of the prototype, then c_{ij} would represent the travelling distance between two addresses).

A more formal mathematical description of TSP is presented below:

Decision variables:

$$x_{ij}$$

$$(x_{ij} = 1 \quad \text{– if driving from i to j, 0 otherwise})$$

auxiliary variable:

$$z_i$$

Objective function:

$$\text{min}\, c = \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}$$

Constraints:

$$\text{drive to every location: } \sum_{i=1}^{n} x_{ij} = 1 \qquad j = 1, 2, ..., n$$

$$\text{leave from every location: } \sum_{j=1}^{n} x_{ij} = 1 \qquad i = 1, 2, ..., n$$

$$\text{avoid short cycles:} z_i - z_j + n * x_{ij} \leq n - 1 \qquad i, j = 2..., n; i \neq j$$

However, the generalized TSP algorithm would violate the constraint of specific time windows for each delivery. Therefore, an extension of TSP algorithm is demanded, known as Traveling Salesman Problem with Time Windows (TSPTW). Additionally, a further extension is needed, as every customer has declared a set of delivery addresses, from which only one should be chosen. This extension is known in combinatorial optimization as set TSP, group TSP, One-of-a-Set TSP, Multiple Choice TSP or Covering Salesman Problem.

There has been an effort to combine algorithms proposed by Solomon (1987)), Dumas, Desrosiers et al. (1995)), Dumitrescu and Mitchell (2001)) and Focacci, Lodi et al. (2002)). However, every approach would result to at least non-deterministic, polynomial-time hard (NP-Hard) complexity class, which is unacceptable for this prototype. Furthermore, as the prototype's approach involves real-time responses to delivery schedule changes, it is very difficult for the exact methods, which were reviewed, to be integrated into this approach.

Therefore, it was decided to simplify the routing process and face the tradeoff between complexity and optimization by developing a custom heuristic approach: Every parcel is initially delivered to the depot, where the order address belongs. Then, parcels are assigned to vehicle routes, by focusing on the declared time windows, as well as the priority of the parcel. Finally, the optimal routes are computed using an external service, which, on a later iteration, was decided to be Google Maps API, that uses an implementation of a TSP algorithm (Google, 2014).

3.9 Clustering Algorithm

While developing the distribution-routing algorithm, there was the need to apply a clustering method. Given were a predefined set of distribution K vehicles and a set of n parcels, which have to be assigned to the vehicles in such way that, the vehicles cover the minimal distance to deliver them. Moreover, the capacity of the vehicles I, as well as the delivery address of each parcel is also known. After reviewing relative data mining literature, it was decided to proceed with k-means clustering algorithm (Hartigan and Wong, 1979). However, this algorithm does not compute clusters of predefined capacity, therefore it was slightly extended based on Lloyd's algorithm (Lloyd, 1982). The concept of the algorithm is explained below:

As the address of each order is represented by a pair of decimal values, corresponding to the address' latitude and longitude, they can be projected as nodes on a 2-dimensional coordinate system.

1. K nodes are selected randomly as each cluster's centroids.
2. Each one of the n nodes (parcels' address) is assigned to the nearest centroid, by calculating the distance between them as a Euclidean distance. If the centroid is full (due to capacity), the next nearest centroid is chosen.
3. For each cluster that has been formed, a new centroid is computed, by finding the node with the minimal average squared distance to the other nodes of its cluster.
4. Steps 3 and 4 are repeated, until there are no changes regarding the centroids, or after a predefined time of iterations.

4 Prototype

The purpose of the project is to develop a prototype that optimizes distribution logistics process using big data. The prototype will be used by the customers and employees of distribution logistics companies. The prototype consists of a smartphone application, allowing the users to get informed about, as well as to participate in the delivery procedure of their orders, by defining their desired delivery time windows and locations and managing their favorite parcel stations and trusted persons. Furthermore, a back-end system should be implemented, operating as a server and being responsible for routing calculations, and data storage. The objective of the prototype is to maximize customer satisfaction, by providing a flexible parcel distribution service, which is managed via a user-friendly smartphone application. Moreover, it aims to an efficient route computation process, also considering real-time changes in the distribution schedule, in order to minimize transportation costs. Beside customer satisfaction and user experience, the major quantitative success criterion is a high percentage of orders delivered within the user-defined time windows, at the desired location. The deployment model is visualized in figure 1

4.1 Actions if recipient is not present

If the recipient is not present at the delivery location, a popup message is displayed on his smartphone app. Then he can choose between the following four options: 1) his order will be instead delivered to another of his registered addresses, which has to be located in the service area of the delivery vehicle, 2) his order will be instead delivered to a selected trusted person, that has to be located in the service area of the delivery vehicle, 3) his order will be instead delivered to a favorite parcel station, that has to be located in the service area of the delivery vehicle, 4) his order will be returned to the depot. If the user does not respond to the new delivery location proposal, it is checked if there is a favorite parcel station of him, located in the delivery vehicles service area. If a suitable parcel station is found, the order is delivered to this parcel station, otherwise it is returned back to the depot.

A plan divergence is automatically recognized from the smartphone app. By checking the user's current or last known location, using the GPS sensor of the smartphone or his mobile network, the app should recognize whether he is located further than 1 km away of the delivery address, 30 minutes before the estimated delivery time. If a plan divergence is recognized, as described in the requirement "Plan divergence recognition", the smartphone app should display a popup message, asking the user, if the divergence automatically recognized is about to take place or not. If the user confirms the divergence, actions described in the requirement "Actions if recipient is not present" will be executed.

4.2 Distribution-routing algorithm

To face the complexity challenge of Travelling Salesman family of problems, there was the need to meet a trade-off between optimization and complexity. The core of the distribution-routing algorithm lays therefore on the user-defined time win dows, on order's priority, as well as on historical data from past purchases and

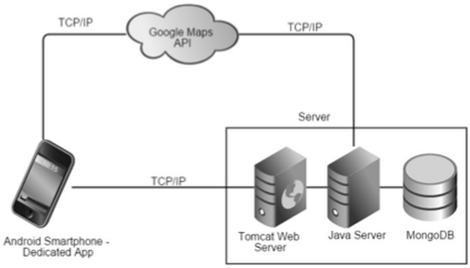


Figure 1: Deployment model

a clustering technique. The problem will be faced as a TSP, only at its last step, in order to calculate the optimal route for one particular vehicle. The algorithm is described in detail below, using natural language instead of code, in favor of comprehensibility. It is supposed to run the evening before the orders are delivered.

```
Preconditions
      Variable number of depots; Variable number of vehicles per
           depot
      Service time: 09:00 -19:00
      4 routes/time slots per vehicle per day: 09:00 -11:00,
           11:30 -13:30, 14:00 -16:00, 16:30 -19:00
      8 orders per route per vehicle (Limitation of Google Maps
           API free version)
      Algorithm runs for each depot in parallel
Step 1
      Separate orders without declared time windows
Step 2
      Sort orders by descending priority
      For each time slot:
            For each order:
                  For each order time window whose address be-
                       longs to the current 'depots service area
                        calculate time window midpoint: (end
                             time + start time)/2
```

```
                          If midpoint is in this time slot,
                               assign order to this timeslot
                          Sort assigned orders by descending
                               priority
                          Keep only as much as can be loaded in
                               the existing vehicles according to
                                its capacity
Step 3
      Sort unassigned orders by descending priority
      For each unassigned order
            For each order time window
                  Find the depot, in whose service area the time
                        window address belongs
                  Calculate time window midpoint
                  Choose the respective time slot for the
                       calculated midpoint
                  If time slot is not full, assign order to that
                        timeslot
Step 4
      Sort orders by descending priority
      For each unassigned order (about 5-10 %)
            According to past purchases, find the best timeslot
                 to deliver to order address
            If time slot is full, repeat step a and find the
                 next better timeslot
Step 5
      Sort orders without declared time windows by descending
           priority
      For each order According to past purchases, find the best
           timeslot to deliver to order address. If time slot is
           full, repeat step a and find the next better timeslot
Step 6
      For each time slot: Assuming that the depot has K vehicles,
      use the extension of the K-means clustering algorithm,
           described in Chapter 2, in order to distribute the
           ordersassigned in this time slot, into K clusters of
           maximal size of 8 nodes
Step 7
      For each time slot: For each route/cluster:
      Optimize route with Google Maps API: Depot address is set
           as start point, as well as destination of the route
           and the maximal 8 orders are set as waypoints. Google
```

```
Maps API uses an implementation of TSP, described in
Chapter 2, in order to calculate the optimal route
```

4.3 Real-time re-computation

As described in the concept, routes should partially adapt real-time to changes in the distribution plan. However, considering customer satisfaction, as well as the nature and context of the service provided, there is no possibility to re-compute all the routes and constantly notify the customers about an updated delivery address and estimated delivery time. Therefore, it was decided to perform a real-time re-computation in the following 2 cases:

(1) Case recipient is not present, when an order is delivered.

If there is a trusted person with a shared address located in the service area of the depot, deliver the order to that person. If the trusted person is also not present, return order to the depot

Else if there is a customer's favorite parcel station located in the service area of the depot, deliver the order to that parcel station

Else, return order to the depot

(2) Case of detected plan divergence

In case it is detected, using the smartphone's GPS sensor, that the recipient is located further than 1km from the delivery address, 30 minutes before the estimated delivery time, display a popup message. This should ask him to verify, if there is a plan divergence or not.

If he confirms it, let him select an alternative delivery location, choosing between: i) an address of him, ii) a favorite parcel station iii) a trusted person's shared address iv) the depot. All of these option have to be filtered, by displaying only those, which are located within the depots service area

If he denies it, perform no action

If he does not respond, follow the steps of "case recipient is not present"

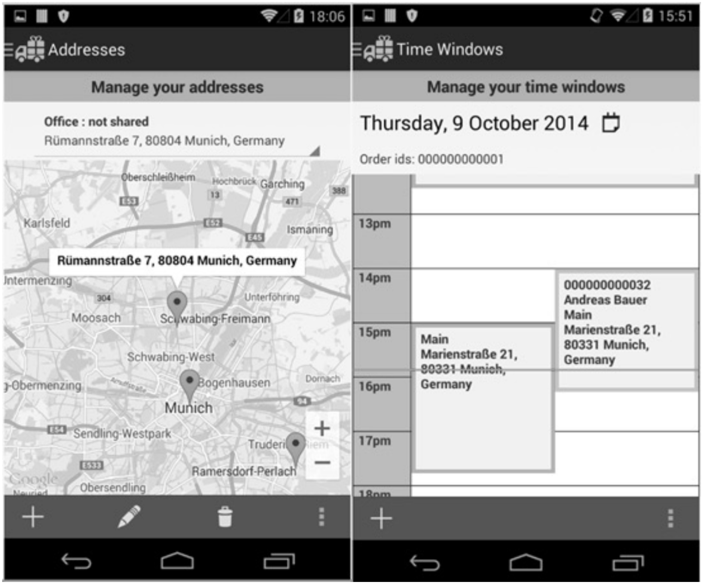


Figure 2: Screenshot of the Software Prototype

5 Evaluation

It was decided to assess customer satisfaction by calculating the following effectiveness metric: for each test run, the percentage of orders estimated to be delivered within a recipient-defined time window in comparison to the total number of order.

Test data

The evaluation of the prototype would be more accurate, if there was a big distribution logistics company available as business partner for this research. As long as no such partner is available, the test data had to be generated in case no suitable real data was available. However, the generated test data was as realistic as possible.

Service area

As service area, we chose the city of Munich, with a city population of 1.407.836 and an area of 310.43 km2 being a good example for medium-sized cities in the world.

Depots

In the service area, 74 different postal codes exist. Those postal codes where grouped by their first four digits, resulting in 19 groups. Some groups were merged with others to avoid groups covering very small area, which led to 17 groups/districts. The estimated population of the districts varies between 50.000 and 135.000 people. At the geographical center of each district a fictional depot was created, as well as the fictional customers. The number of customers of each district was calculated according to their population, also considering that at least 2 vehicles operate for each depot.

Customers

For each one of the 3000 customers, the following attributes were generated: A real main address (main addresses are distributed, corresponding to the demand for test customers of its depot)

- 0 to 5 further real addresses, randomly located within the whole service area
- 0 to 6 random favorite parcel stations
- 0 to 5 random trusted persons (randomly chosen between the rest customers)
- 0 to 1 pending order

Orders

For each one of the 150 orders (all estimated to be delivered at the same date) a random priority decimal value between 1 and 5 was generated, with intermediary steps of 0.1.

Parcel stations

47 parcel stations were created. Instead of randomly generating their locations, locations of actual parcel stations of DHL in Munich were used (DHL, 2014).

Purchase history

For each customer, a purchase history record was generated. It consists of a couple of integers for each one of the 4 delivery time slots, representing the sum of un-/successful delivery attempts in the past. The value of these integers varies between 0 and 10.

Time windows

Time windows were the most complex objects to generate. We used the following assumption: For each customer expecting an order at the specific day:

With probability of 10%, create no time windows at all

Starting with 9am, with probability of 15%, create a time window

Starting with minimum 2 hours duration, with probability of 80%, keep increasing its duration by 15 minutes.

With probability of 25%, choose randomly between the favorite parcel stations of the customer and set its address as the desired delivery address for that time window

With probability of 25%, choose randomly between the shared addresses of a randomly chosen trusted person of the customer and set this address as the desired delivery address for that time window, after checking that there is no conflict with the schedule of the trusted person.

With probability of 50%, choose randomly between the addresses of the customer and set it as the desired delivery address for that time window. A greater weight is given to the main address of the customer.

With probability of 85%, increase the time by 15 minutes and start over.

After several test runs, the metrics for assessing customer satisfaction corresponds to the average of those percentages, as well as average deviation between the defined time window and the estimated delivery time or time range. This is the factor that affects customer satisfaction the most, as the customers would expect of a service, supposed to optimize distribution logistics process, to deliver their orders at a time and location that fits to their schedule. However, when using that test scenario, it is not possible to integrate real time changes and route re-computations, which would probably slightly decrease the overall effectiveness metric, but at the same time providing a significant added satisfaction to the individual customers, whose schedule was changed.

Our evaluation results indicate that for 63.8% of the orders, the estimated delivery time concurs with a customer-defined time window. Regarding the rest orders, the average deviation between the estimated delivery time and a time window is 101.64 minutes. This value may seem quite high for the customer-oriented service, our prototypes aims to provide. However, in most cases this deviation is

Table 2: Results of the Evaluation

Metric	Value
orders, for which time windows are defined	86.2%
orders, for which time windows are not defined	13.7%
orders, for which the estimated delivery time concurs with a defined time window	63.8%
Average deviation between the estimated delivery time and the limits of a defined time window	101.64 minutes
orders, for which the estimated delivery time range concurs with a defined time window	48.4%
orders, for which the estimated delivery time range overlaps a defined time window	31.4%
orders, for which the estimated delivery range does not concur with a defined time window 20.2%	
Average deviation between estimated delivery time range and time window	33.67 minutes

no longer than 20 minutes, but the value is highly affected by extreme cases. On those cases, the total time of defined time windows for a particular order do not extend the minimum of two hours, which usually results in significantly higher deviation values.

This can be derived from the following set of metrics, regarding not the exact estimated delivery time, but the estimated delivery time range, with duration of 60 minutes. The estimated delivery time range of 48.4% of the orders concurs with a customer-defined time window, while another 31.4% of the delivery time ranges overlaps with a time window. These sums up to 79.8% of all orders, while only by 20.2% of the orders the delivery deed not meet a defined time window at all.

6 Discussion

The results indicate the availability of big data originating from different supply chain partners being relevant to improve distribution networks, e.g. processes or services. However, challenges such as information security, accuracy, complexity, optimization, or the user experience have to be regarded and tradeoffs have to be balanced between overall goals and the purpose of each service or product.

In the case of this prototype a lot of opportunities in the area of big data were applied to develop a prototype for optimizing distribution networks. The major component of the distribution logistics optimization process is the analysis of user-defined time windows, providing information about the desired delivery time and location of the customers. The prototype fulfills the majority of its initially defined functional and non-functional requirements providing a flexible, user friendly, and effective service. The prototype represents an effective solution considering customer needs, improving customer satisfaction, and providing firms with the opportunity to realized cost savings.

While our prototype does not use big data analysis techniques, the prototype has laid the technical foundations and relations between data for future distribution solution. Therefore, our prototype allows firms to offer customer-oriented delivery solutions plus being cost-effective and -efficient. Based on the evaluation results, the prototype could be used as an extension for existing distribution services or products. In addition, the prototype is being setup to integrate further data from other sources such as cyber-physical-systems or autonomous cars.

Using the prototype requires firms to adjust the prototype and adapt their supply chain strategy, especially from a pricing perspective. Firms need to define their planned margin, costs, and revenues. Firstly, in case of a new entry in the market, a niche player or a logistics company with considerable market share, which aims to extend its customer base, the offer of the service for free is recommended (as standard delivery option). Secondly, if the logistics company is not prepared for the full integration of this service as standard delivery, but simultaneously wants to face the margin pressure and not left behind by its competitors, a yearly or monthly subscription pricing strategy or a pay-per-use pricing strategy is recommended. Finally, a combination of the two previous pricing strategies is proposed, following a freemium model, by offering the service for free for some times monthly, but if the customer extends that limit, a pay-per-use model should be then used.

6.1 Implications

Regarding the implications of this research, the prototype is a practical application of a recent theoretical concept (Engel, Sadovskyi et al., 2014) and identified needs within the field of logistics (Waller and Fawcett, 2013), based on the latest technology evolvement representing a contribution in the field of information systems. Moreover, the heuristic distribution algorithm, which was developed to approach the set Traveling Salesman Problem with Time Windows and is explained in Chapter 2, constitutes a scientific contribution to the field of supply chain management. Rather than extending existing algorithms, in order to slightly decrease the complexity of this (at least) NP-Hard family or problems, the approach of this research focused on facing the tradeoff between optimization and complexity. The problem has been by notably simplified, using available data, and approach it as a TSP only, at one of its latest stages.

Furthermore, this prototype could serve as a model basis and be integrated in the delivery services currently offered by the logistics companies, which mostly do not take into consideration the schedule and needs of the recipient (Lienbacher, Waldegg-Lindl et al., 2013). This allows firms to increase service effectiveness by optimization of the routes, distribution costs will be reduced, and customer satisfaction will be on the foreground, constituting the major goal of distribution logistics companies. Such a customer-oriented approach of high quality is vital for any organization's success (Monczka, Handfield et al., 2008), considering the evolvement of customer dynamism over the last years. Finally, if this prototype is widely adapted in the future, it may also indirectly contribute to society, since it will affect gas emissions and working conditions in the field of distribution logistics.

6.2 Limitations

Considering the limitations of this research, all the analysis and development phases have been highly affected in several ways by the prototype's context, its planned duration and the available resources. Having set customer satisfaction, user experience and compliance with requirements of customers as major goals of the prototype, as well as the lack of a logistics company as business partner, prevented the analysis and use of data, which could be highly beneficial for the optimization of distribution logistics. There has also not been an opportunity to integrate actual distribution vehicles into the prototype, which could provide the optimization process with a considerable amount of big data, originating from their navigation systems or NFC sensors. Furthermore, the use of the free version of Google Maps API, prohibited the integration of real time weather and traffic information, the calculation of routes between more than 10 nodes, as well as the use of distribution vehicles of varying capacity. Regarding the smartphone application, the focus has been laid on implementing functionalities for managing a user's orders, time windows, addresses, parcel stations and trusted persons, as well as on creating a simple and friendly user interface. There is plenty of room for improvements, including, for example, a push notifications system and a more efficient data persistence management, before a future market launch. Finally, the most significant limitations of this research are observed at the testing phase. The real-time nature of the service provided by the prototype also demands a real-time and realistic testing environment, to be fully and properly tested. Considering the context and resources of this prototype, such a testing phase could not be realized, and therefore a different testing scenario was chosen. Using this scenario, customer satisfaction, which is the major goal of the prototype, is being assessed by calculating the percentage of orders being delivered with a customer-defined time window. However, as customer satisfaction is a qualitative metric, further evaluation actions are recommended.

7 Future Research

Besides facing the time-based and resource-based limitations, the focus should be firstly laid on further optimization of the distribution and routing algorithm, in terms of efficiency and accuracy. Furthermore, a partnership with a big distribution logistics company is highly recommended in order to access a large amount of big data. This allows researchers to make use of analytics for optimizing the distribution network. Moreover, the business partner could provide the appropriate testing environment and conditions, in order to face the major limitation of the current research. Finally, being in an era of rapid and continuous technological evolvement, new domain/related opportunities, as well as new customer requirements will arise. Their iterative review will be critical for future research in the field of distribution logistics optimization.

References

DHL (2014). *DHL Standorte*.

Domschke, W. and A. Scholl (2010). *Logistik: Rundreisen und Touren*. Oldenbourg Verlag.

Dumas, Y., J. Desrosiers, E. Gelinas, and M. M. Solomon (1995). "An optimal algorithm for the traveling salesman problem with time windows". In: *Operations research* 43.2, pp. 367–371.

Dumitrescu, A. and Joseph S. B. Mitchell (2001). “Approximation algorithms for TSP with neighborhoods in the plane”. In: 365417: Society for Industrial and Applied Mathematics, pp. 38–46.

Engel, T., O. Sadovskyi, R. Heininger, M. Böhm, and H. Krcmar (2014). “A Conceptual Approach for optimizing Distribution Logistics using Big Data”.

Fang, L. and C.-q. Zhang (2005). “The e-Logistics Framework in e-Commerce”. In: *Proceedings of the 7th International Conference on Electronic Commerce*. ICEC ’05. Xi’an, China: ACM, pp. 408–412.

Focacci, F., A. Lodi, and M. Milano (2002). “A Hybrid Exact Algorithm for the TSPTW”. In: *INFORMS Journal on Computing* 14.4, pp. 403–417.

Google (2014). *Google Directions API*.

Hartigan, J. A. and M. A. Wong (1979). “Algorithm AS 136: A k-means clustering algorithm”. In: *Applied statistics* 28.1, pp. 100–108.

Internet World Stats (2013). *Internet Usage Statistics - The Internet Big Picture*.

Lienbacher, E., M. B. Waldegg-Lindl, and M. M. Waldegg-Lindl, eds. (2013). *Last Mile: Die letzten–und teuersten–Meter zum Kunden im B2C ECommerce*. Handel in Theorie und Praxis. Berlin, Wiesbaden, and Heidelberg: Springer.

Lloyd, S. (1982). “Least squares quantization in PCM”. In: *Information Theory, IEEE Transactions on* 28.2, pp. 129–137.

Monczka, R., R. Handfield, L. Giunipero, and J. Patterson, eds. (2008). *Purchasing and supply chain management*. Boston, MA and USA: Cengage Learning.

Sadovskyi, O., T. Engel, R. Heininger, M. Böhm, and H. Krcmar (2014). “Analysis of Big Data enabled Business Models using a Value Chain Perspective”. In: *MKWI 2014 - Multikonferenz Wirtschaftsinformatik. Relativistic groups and analyticity*. Proceedings of the Eighth Nobel Symposium. (Feb. 26–28, 2014). Ed. by D. Kundisch, L. Suhl, and L. Beckmann. Paderborn, pp. 367–377.

Shekhar, S., V. Gunturi, M. R. Evans, and K. Yang (2012). “Spatial Big-data Challenges Intersecting Mobility and Cloud Computing”. In: *Proceedings of the Eleventh ACM International Workshop on Data Engineering for Wireless and Mobile Access*. MobiDE ’12. Scottsdale, Arizona: ACM, pp. 1–6.

SODA ’01: Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms (2001). Washington, D.C., USA: Society for Industrial and Applied Mathematics.

Solomon, M. M. (1987). “Algorithms for the vehicle routing and scheduling problems with time window constraints”. In: *Operations research* 35.2, pp. 254–265.

United Nations (2000). *The World’s Women: 2000: Trends and Statistics*. United Nations Publications.

Valerio, D., A. D’Alconzo, F. Ricciato, and W. Wiedermann (2009). “Exploiting Cellular Networks for Road Traffic Estimation: A Survey and a Research Roadmap”. In: *VTC Spring 2009 - IEEE 69th Vehicular Technology Conference*, pp. 1–5.

Waller, M. A. and S. E. Fawcett (2013). “Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management”. In: *Journal of Business Logistics* 34.2, pp. 77–84.

Weltevreden, J. W. J. (2008). “B2c e-commerce logistics: the rise of collection-and-delivery points in The Netherlands”. In: *International Journal of Retail & Distribution Management* 36.8, pp. 638–660.