UPS Optimizes Delivery Routes

Chuck Holland, Jack Levis, Ranganath Nuggehalli, Bob Santilli, Jeff Winters

To cite this article:

Full terms and conditions of use: http://pubsonline.informs.org/page/terms-and-conditions

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article’s accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2017, INFORMS

Please scroll down for article—it is on subsequent pages
UPS Optimizes Delivery Routes

Chuck Holland,a Jack Levis,b Ranganath Nuggehalli,b Bob Santilli,b Jeff Wintersa

a UPS, Atlanta, Georgia 30328; b UPS, Timonium, Maryland 21093

Contact: cholland@ups.com (CH); jlevis@ups.com (JL); rnuggehalli@ups.com (RN); bsantilli@ups.com (BS); kjwinters@ups.com (JW)

Abstract. UPS, the leading logistics provider in the world, and long known for its penchant for efficiency, embarked on a journey to streamline and modernize its pickup and delivery operations in 2003. This journey resulted in a suite of systems, including a metaheuristic optimization system, which it called “On Road Integrated Optimization and Navigation” (ORION). Every day, ORION provides an optimized route for each of UPS’ 55,000 U.S. drivers based on the packages to be picked up and delivered on that day. The system creates routes that maintain the desired level of consistency from day to day. To bring this transformational system from concept to reality, UPS instituted extensive change in management practices to ensure that both users and executives would accept the system. Costing more than $295 million to build and deploy, ORION is expected to save UPS $300–$400 million annually. ORION is also contributing to the sustainability efforts of UPS by reducing its CO2 emissions by 100,000 tons annually. By providing a foundation for a new generation of advanced planning systems, ORION is transforming the pickup and delivery operations at UPS.

Keywords: simulated annealing • OR practice • local search • vehicle routing and scheduling • consistent solutions

UPS has a rich history of operational innovations and is steeped in a culture of analytics. Anyone researching in its archives will quickly notice references to operations research (OR) dating back to the 1950s. Software, which UPS calls “On Road Integrated Optimization and Navigation” (ORION), now tells its 55,000 U.S. service providers (i.e., local-delivery van drivers) the sequence in which they should pick up and deliver packages for the customers that are assigned to them that day. UPS required nearly 10 years to develop and deploy this system.

In 1907, Jim Casey used a borrowed $100 to found the company as a messenger service company in Seattle, Washington; today UPS is a world leader in logistics. It offers a broad portfolio of services that range from manufacturing to warehousing to distribution to repair services. In 2015, its total revenue was $58 billion and its net profit was $4.8 billion. With 444,000 employees globally, it operates in more than 220 countries and territories. On a typical day, it delivers 18 million packages and documents to 10 million customers and collects packages from 1.8 million customers. Using a fleet of 237 aircraft it owns and another 413 it charters, UPS operates one of the world’s largest civilian airlines, using 728 airports throughout the world.

Its activities are wide-ranging and international in scope. A customer may never know that UPS produced his (her) last set of custom golf clubs, repaired a laptop, or fulfilled an order for Valentine Day roses. It is a full-service logistics provider, which enables global commerce. By an internal estimate, two percent of global GDP and six percent of U.S. GDP flow through the UPS network.

Both Jim Casey, who served as the first UPS CEO, and George D. Smith, its second CEO, played significant roles in shaping UPS. Their efforts resulted
in transforming the small messenger company into a leading transportation company, which focuses on innovation, service, and operational efficiency. By 1972, UPS had the largest industrial engineering (IE) department of any company in the world. IE is a corporate function and the planning arm of UPS. Process improvement through analysis and work measurement became an essential element of its operations.

**UPS Small-Package Operations**

The UPS U.S. small-package business is the company’s oldest and largest business segment. In 2015, it accounted for 63 percent of total UPS revenue and 62 percent of UPS operating profit. Two main groups comprise small-package operations. The first group, the transportation group, is responsible for moving packages from origin cities to destination cities. This group is comprised of UPS airline and UPS ground transportation. UPS airline is responsible for moving mostly premium-service time-sensitive packages, which must be delivered in one or two days, and for international transportation. Its main hub is in Louisville, Kentucky. Ground transportation is responsible for moving nonpremium packages between cities, either by truck or by train. UPS is one of the largest customers for many railroads, and has built an extensive ground network consisting of many consolidation hubs. At the principal ground hub in Chicago, employees can simultaneously load and unload more than 1,000 trucks.

The second group, pickup and delivery, is responsible for the local pickup and delivery of packages to customers. UPS operates about 1,400 package delivery centers (i.e., package centers) in the United States. The package centers are the gateway between customers and the UPS network. Early each morning, packages in these centers are loaded into delivery vans. A driver leaves a package center after all packages have been loaded into his (her) delivery van, makes deliveries for the major part of the day, and toward the end of the day, collects packages from customers for delivery to other customers. Some of these packages may be delivered in the same city the following day, or they may be transported to other cities for delivery. Depending on the origin-destination distance and the type of service (e.g., premium, saver, ground), a U.S. package may be in the UPS network for between one and five days. On a typical day, about 55,000 UPS drivers deliver more than 16 million packages in the United States. A typical driver serves about 140–160 customers (or stops in UPS terminology). The number of packages to be delivered and picked up can change significantly depending on the day of the week, the week of the month, and the month or season of the year. During the period between Thanksgiving and Christmas, the volume often reaches twice the normal volume. In 2015, UPS delivered 34 million packages on its peak day. Building and maintaining a physical network that can operate efficiently at both ends of the capacity spectrum represents a considerable challenge.

The U.S. network has evolved over time and is highly integrated. Because of the density of the UPS network, a premium-service package may travel entirely by ground transportation. At the destination, one driver delivers packages of all types of services in his (her) assigned service area, irrespective of how they were transported to the area. Its integrated network provides UPS with operational advantages, but also makes the network complex to manage.

**Planning Package and Delivery Operations in the 1970s**

UPS experienced significant growth in the 1960s and 1970s. To cope with its increased growth and maintain its operational efficiency, the company established a planning process, which it called controlled dispatch. Through this process, UPS provided specific guidelines for planning and executing a multitude of its package and delivery (P&D) activities using historical data. Controlled dispatch assisted in determining: (1) the packages-to-van assignment; (2) the shelf location in the van for each package; and (3) the sequence in which the packages on a van should be delivered.

Internet and mobile phones were not available, as they are today; therefore, most planning had to be done without knowledge of which packages had to be delivered on a given day. Controlled dispatch enabled UPS to plan its delivery operations using historical information.

The entire service area of a package center was divided into smaller areas called loops (Figure 1(A)). An industrial engineer would first arrange the street segments in a loop’s AM group and then the street segments in the PM group so that they formed one continuous path, which was called the base route. The base
route was constructed with several objectives; examples include minimizing travel time, serving business customers in the early part of the day, and avoiding traffic congestion. Each street segment was assigned a specific position in the base route. The IE group was responsible for creating the loops and the base routes. A loop and base route would remain active for two years or more. Whenever the delivery needs of a geographical territory significantly changed, an industrial engineer would recreate the loops and the base routes.

The local planner, the package dispatch supervisor (PDS), was responsible for day-to-day planning, including determining the number of vans needed and assigning geographical territories to each van. The PDS would combine the package-volume forecasts (i.e., number of packages expected to be delivered and picked up) from the corporate office and the historical demand patterns (i.e., probability that a street segment may have one or more deliveries) to simulate the pickup and delivery needs on a given day. The result would be a simulated list of customers aggregated by street segments. The simulated list would be used to estimate the delivery and pickup needs within a given loop. Using the simulated list and starting from the apex, the local planner would divide the base route into three to five separate subareas, with each subarea defining one or more contiguous geographical areas (Figure 1(B)). Each subarea would become a route, and would be assigned to a delivery van. By assigning segments from the base route, each van was assigned a geographical territory; the van’s driver had to deliver all packages destined for that territory. The person loading the van (i.e., preloader) merely had to know the territory that was assigned to each van. The preloader would look at the address on a package and load the package into the van whose territory included that package’s address.

The van-territory assignments were made using package volume estimates; therefore, the actual number of packages assigned to a van (after they had been loaded) usually differed from the planned number of packages. If the actual number was below or above the specified range, adjustments would be made during preload operations by adding packages to or removing packages from another van.

The base route also guided the driver in delivering the packages. A driver was expected to roughly follow the base route, because doing so provided a level of consistency for the drivers and the customers. Since the base route construction did not consider the...
time-committed deliveries, and not all street segments would need to be visited on a given day, strictly following the base route was not always optimal. Hence, the drivers were taught to use their judgment and make adjustments.

From experience, UPS had learned that placing packages in the vans in the order of their delivery would increase delivery efficiency. Drivers would not have to waste time searching for packages. Because the drivers would roughly follow the base route, the preloader would use the base route to determine where to place a package in a van. Typically, a preloader was responsible for loading up to three vans. Hence, to load the vans quickly, the preloader had to memorize one to three base routes.

A driver would not know which packages were in a van. Before leaving the package center to make deliveries, the driver would climb into the van, look at the first few packages, mentally determine a route, rearrange those packages if necessary, and deliver them in the sequence in which they were arranged (or rearranged). Once the first few packages were delivered, the process would be repeated.

Most of the planning was done manually on paper, and the plans were often communicated orally. Because of the growing volume and increased management complexity, UPS continually increased its reliance on following its established methods; however, establishing methods and strictly adhering to them had a downside. Making any changes to the system or adopting a new one to replace an outdated system was difficult.

Because of its methods, UPS grew and prospered, and its delivery people became among the highest paid in the transportation industry. This is still true today.

The Need for a New Planning Paradigm

UPS went public in 1999. Changing business conditions made the company reevaluate its planning process. The proliferation of Internet-enabled e-commerce resulted in increased volume, a need for more customized services, and more variability in the demand pattern. With the introduction of multiple premium services (e.g., delivery and pickup within a specified time window, customized instructions for delivery and (or) pickup, service on demand), planning P&D activities became more complex. UPS had to compete with other companies that did not have its high labor expenses and complex integrated networks. In addition to managing the physical flow of the packages, managing the information about the packages became important. Changing demographics also made attracting and retaining good preloaders difficult. In summary, the company’s logistics became more complex.

Computerizing the Process

In 2003, to maintain its operational efficiency as it tried to manage increasing complexity, UPS began deploying a suite of systems that collectively have come to be known as package-flow technologies (PFT). The initial goal was to computerize the UPS planning process and simplify the job of the preloader.

Using PFT enabled UPS to leverage its controlled dispatch methods and practices. The planners could plan, verify, and communicate electronically, and a planner could electronically make van and shelf-positioning assignments for packages. Hence, changing and fine-tuning the plans to meet the P&D needs for a specific day became easier.

The preload simplification was done using the preload assist label (PAL) (Figure 2). When a package was unloaded from a trailer, an employee scanned the printed address label. PFT electronically matched the address to the assigned delivery van, found its assigned shelf position in the delivery van, printed a PAL, and attached it to the package.

A preloader would look at the PAL (Figure 2), see the van number and the shelf position, and load the package at the specified shelf position in the assigned van.
Thus, a planner could easily change the van-assignment and package shelf-position plans as needed, and the preloader did not have to memorize the base route. PFT also provided the driver with a list of packages that were in the van; this list was displayed on a handheld computer, that is, a delivery information acquisition device (DIAD), in the order specified in the base route. The driver no longer had to look at the packages in the delivery van to determine the sequence in which to deliver them.

By 2011, UPS estimated that PFT had helped it to reduce annual travel by 85 million miles, save 8.5 million gallons of fuel, and reduce its CO$_2$ emissions by 85,000 tons annually.

From its beginning, the PFT suite was designed with the explicit objective of integrating OR tools and techniques into the UPS planning process. The objectives were to simplify the manual planning process, develop data models, and then supplement and (or) replace the planning tasks with optimization algorithms. PFT was a multiple-stage integration project. The ORION project was initiated to execute the first phase of algorithm development and integration.

**ORION: Optimizing the Routes**

While PFT was in its initial stages of deployment, we started to develop the optimization algorithms, which had to solve two principal problems. The first was to assign packages to delivery vans (i.e., solve the van-assignment problem); the second was to determine the sequence in which each driver could service all his (her) assigned customers without violating the time-window commitments (i.e., solve the routing problem). These two problems are the components of the vehicle routing problem with time windows (VRPTW), a well-researched problem (El-Sherbeny 2010). Most VRPTW research treats these two problems as a single problem because these two decisions are interdependent; however, because of our experience in implementing advanced planning systems, we decided to treat them as two separate problems. We focused on optimizing the routing problem while retaining the existing computer-assisted van-assignment procedure. That is, we would monitor the packages that were loaded into a van, and generate an optimized delivery sequence after all packages had been loaded. We would then combine the delivery sequences with map data to provide a route (path) with specific directional guidance to the driver via a DIAD.

**First Attempt: Traveling Salesman Problem with Time Windows**

The objective of ORION is to create a low-cost route that enables a driver to deliver and pick up all packages while meeting the time-window requirements. By creating a route, we mean determining the sequence and (or) order in which all the assigned customers should be served (while recognizing that the driver’s path between stops will be determined later).

The UPS routing problem is similar to the traveling salesman problem with time windows (TSPTW), which is a variant of the well-known traveling salesman problem (TSP). The general case of TSPTW is known to be NP-complete. Although the TSPTW could be modeled as an integer programming problem (we provide details in the appendix), our tests indicated that this was not a viable option for us because the integer programming model would take too long to solve.

To determine a good route, we need complete information about all packages that must be delivered and picked up. The only time we know with 100 percent certainty all the addresses that a driver has to visit on a given day is after the van has been loaded completely. Immediately after a van has been loaded, its driver departs the package center. One minute of nonproductive time per day per driver across the entire U.S. network would cost UPS $15 million annually. Hence, the routing algorithm must be extremely fast. A review of research literature on TSPTW solution procedures, whether optimal or heuristic, revealed that no publicly available algorithm could meet our needs.

Hence, we developed a meta-heuristic consisting of local-search algorithms to solve our routing problem. Initial lab tests indicated that using the meta-heuristic would result in cost savings. We formed a team consisting of people from industrial engineering, operations research, and P&D operations to field test the algorithm. The OR group developed a tablet computer-based simulator for the field test. A test consisted of the following steps: (1) The simulator downloaded a set of stops from one of the actual routes and optimized it. (2) The team then followed the recommended route in a car to evaluate whether a UPS driver could
follow the prescribed route and meet all service constraints. (3) After simulating serving a customer, the simulator recalculated the arrival time at the remaining customers, and reoptimized the route if the current route became infeasible (i.e., missed time windows for one or more customers) because of reasons such as traffic delays. After doing extensive testing over several months and making a number of enhancements, the team concluded that the optimized routes were difficult to implement and would not be usable by our delivery drivers. Testing was halted, and a 12-month deadline was set to allow us to either prove the concept or to shut down the project. We categorize this first attempt (i.e., the attempt in which we failed) as Round 1 of the first phase.

**An Anatomy of Failure**

We analyzed the reasons for our failure to provide a usable solution. We grouped the problems into three categories: (1) software, (2) data, and (3) existing methods and metrics. Problems existed in each category. Controlled dispatch practices had trained a generation of UPS drivers to follow a regular pattern that would enable these drivers to efficiently serve their customers. For example, they could avoid being near a school when it releases its students. UPS has a long-standing practice of delivering to commercial customers early in the day and picking up packages from them at the end of the day; the driver then returns to the package center. Following a regular pattern facilitates loading packages close to the order of their delivery and increases the delivery efficiency. The routes produced by the algorithm did not follow a consistent pattern from day to day, and did not consider how the delivery van was loaded. Hence, the driver required more time to find the packages when the van arrived at the customer location. Contrary to customer expectations, the algorithm would schedule commercial deliveries near the end of the day when the drivers were supposed to visit them to pick up packages. Time windowing the commercial stops was difficult and created other problems. We also found that the map data did not have the necessary accuracy; for example, the customer location would differ from the location shown on the map, or the travel time and distances differed significantly from those indicated on the map. Additionally, we realized that some of our long-used methods and performance metrics were not pertinent, and were sometimes counter to the objectives of the ORION project. We had to accept the need for changing some business practices.

**Second Time Around: Adding Practical Constraints**

Based on the analysis we describe above, we proceeded to Round 2 of the first phase. We analyzed a route associated with our Lancaster, Pennsylvania package center, created some rudimentary tools to correct the map data, and then collected data for that route over multiple days. We developed a computer program to help us interactively build the routes. Although UPS had been delivering packages for about a century, few rules were in place to define what constituted a good route. Therefore, we had to discover the characteristics of good, implementable routes. We handcrafted multiple routes for each day and analyzed them. As a result of these analyses, we determined that incorporating the base route while sequencing the deliveries would eliminate some of the problems in our solutions.

We modified the TSPTW formulation to incorporate the consistency constraints (we provide details in the appendix) by adding upper-bound and lower-bound constraints on the customer positions on the route; that is, the order of delivery on the route (e.g., first delivery, second delivery).

We obtained the initial bounds by analyzing the delivery history. Since the number of packages and number of customers vary from day to day, the total time required to serve all the customers in given segment of the base route varies. Consequently, the position of a customer along the route also varies. For example, if we strictly follow the base route over a period of time, a specific customer may be the third customer served on the route on one day and the tenth customer served on that route on another day. Thus, by mapping the historical delivery data to the base route, we could determine the bounds on the position of a specific customer. Different customers may have different bounds. The planner can modify the initial bounds to meet the delivery needs on a given day. During the Christmas season when the van is heavily loaded, the route may need to closely follow the base route. We used these bounds to constrain the position of a customer on the route. These constraints are enforced as soft constraints, just as the time-window constraints are. Although modifying the
formulation was easy, developing a heuristic to incorporate the route-position bounds required considerable experimentation.

Our tests also indicated the need to incorporate practical considerations. For example, making left turns and zigzagging (Figure 3) on a busy street are likely to increase the probability of accidents. Hence, to discourage left turns and zigzagging, we increased the objective function penalty for traveling between pairs of customer locations if such traveling would result in unsafe behavior.

In October 2007, after about 10 months of testing and refinement, we demonstrated the new algorithm to our industrial engineering vice president via a simulation ride. Midway through the demonstration ride, he stopped the ride and remarked, “Does it mean that all these years we have been telling our drivers to do the wrong thing and rewarding them for it?” His comment represented a turning point.

By considering the base route while optimizing, the ORION algorithm maintains a level of consistency from day to day in the delivery of UPS packages. This modification significantly changed the package delivery process, as many scholarly publications on the consistent vehicle routing problem illustrate. Examples include Campbell and Thomas (2008), Groër et al. (2009), and Sungur et al. (2010). UPS has been granted three patents (Zhong and Zaret 2008a, b; Zhong 2010) for the development of methods that build consistent delivery routes.

The ORION algorithm is a metaheuristic that contains variants of the Lin-Kernighan k-Opt procedure (Lin and Kernighan 1973), adoptive large-neighborhood search with variable-length neighborhood (Ropke and Pisinger 2006), Lagrangian relaxation (Lemaréchal 2001), and simulated annealing (Ingber 1993). Our routing problem had many side constraints; for example, some groups of stops must follow a strict predetermined delivery order. Therefore, we modified the above procedures to meet our specific needs. Although much of the ORION system is patented (Levis et al. 2009), UPS maintains the algorithm details as a trade secret.

**How ORION Reduces Cost**

In comparing the cost savings generated by a UPS driver-determined route with those of the algorithm, we can categorize the savings into two types of actions. The first one is reducing multiple visits to a customer and (or) neighborhood. The second is making small adjustments to the base route when those adjustments result in reducing the total cost. The base route is constructed assuming that each street segment contains a stop and that strictly following the base route every day is neither optimal nor necessary. Hence, a driver can make adjustments to the base route on any day that doing so serves the needs of that day’s customers. On average, ORION’s adjustments are better than those of the drivers.

In general, visiting a customer or a neighborhood more than once increases the cost; however, because of...
time windows, doing so is necessary sometimes. Figure 4 shows an example in which ORION outperformed a driver by reducing the number of visits to two neighborhoods and by more efficiently delivering to a third neighborhood. In cases in which ORION is unable to reduce the number of multiple visits, it attempts to lower the cost of doing so. For example, it may reduce multiple visits to distant neighborhoods, while increasing the number of multiple visits to neighborhoods that are close to its main service area.

**From Concept to Reality**

UPS is a company known for its measurements. In 2008, to determine the value of our optimization algorithm, we decided to use ORION to optimize all routes for a considerable length of time for a particular package center. Engineers in the UPS OR group and members of the information systems (IS) department collaborated to develop a prototype optimization system. At UPS, the IS department is responsible for maintaining all mission-critical systems, including PFT (and the prototype system had to interact with PFT). This was the first time at UPS that an application developed outside of the IS department was allowed to interact with a mission-critical system. We selected a small-package center with about 20 drivers in York, Pennsylvania. Optimizing and evaluating all of this center’s routes avoided the problem whereby some routes improve at the expense of other routes; however, the overall result did not improve initially. We added new team members to help us in this test, and tasked them with correcting the UPS internal data, validating and enhancing the geographical information (map data), simulating ORION routes, making necessary corrections, and training the drivers and package center management. Our evaluation of the results following these improvements showed that ORION provided significant benefits.

During the following year (2009), we extended our testing to two additional package centers with similar characteristics to verify that we could repeat the earlier success. UPS senior management started to prepare for the development and deployment of a fully integrated optimization system. In the first two years, all field tests were conducted by a corporate expert team. To deploy the system nationwide, we needed to add people who were not planning experts to the deployment team. Some project stakeholders (e.g., senior UPS managers) questioned whether nonplanners (e.g., operations personnel) could be trained successfully to deploy ORION. To learn more about the deployment process and to prove the ORION deployment was feasible, we organized a bigger team consisting of UPS employees with diverse operational backgrounds and located in various parts of the country. Using this team,
we then extended the testing to eight additional package centers. The additional testing helped us to fine-tune the deployment process and proved the viability of ORION beyond any doubt.

The field tests validated our belief that accurate map data were critical to ensuring the project’s success. After extensive investigation, we concluded that the commercially available map data lacked the precision we needed. Hence, we decided to buy a commercially available map database and edit it to meet our needs. Therefore, we built an infrastructure to enable us to edit and maintain the map data. Starting in 2005, we began to install GPS devices in both the DIADs and delivery vans. Each time a package is delivered, the DIAD automatically records the latitude and the longitude information. Similarly, the GPS device installed on the delivery van collects the travel path’s GPS tracking information and the van park positions. By developing a suite of data-mining and map-editing tools that leveraged the GPS data, we were able to increase the map-data precision to the required level. Today, our map database can locate our customers with very high accuracy, and is vital to the next phase (van assignment) of ORION.

In planning for the development of the fully integrated system, we realized that we would need four to five years of development time to integrate ORION with other UPS mission-critical systems. The prototype was not suitable for continued use in operations because it was not fully integrated with these systems; thus, the planner had to use different consoles to access the other systems. With this lack of integration, additional data preparation was required. Hence, we needed more staff. UPS senior management decided to develop a semi-integrated system by enhancing the prototype, and we deployed it as a stopgap measure until we could develop a fully integrated system. This represented a significant investment to UPS. The semi-integrated prototype deployment continued to provide significant cost reductions, thus keeping UPS management focused on the benefits of operations research. Because of subsequent enhancements to the algorithm and the deployment process, gains exceeded the initial estimates.

Package centers using ORION continued to show gains in the second year of use and beyond. In theory, once UPS has successfully deployed ORION, it should not see any additional gains when compared with the new baseline. In reality, because of ORION, the local planners can more accurately estimate the time that drivers need to serve their assigned customers. This knowledge has helped them to make better van assignments, and resulted in additional gains. They were also able to understand how the base route affects the ORION route. The planners started using the feedback they received from the ORION routes to modify the base route, which led to additional gains.

We started deploying the fully integrated system in 2014. Initially, our plan was to complete deployment throughout the United States by 2020. As a result of the benefits that ORION provided, we accelerated the deployment plan. As of this writing, UPS expects full U.S. deployment by the close of 2016. Table 1 gives the system usage since its inception.

As of this writing, 700 people are working full time in deploying ORION with support from a central support staff of 100. On average, a team member spends six days preparing (i.e., correcting, validating, and enhancing the UPS data, maps, and training) a single route to make it ready for ORION. Tom Davenport (Davenport 2013) considers ORION to be “arguably the world’s largest operations research project.”

To integrate ORION with additional legacy operational systems, we developed a hybrid infrastructure that is both local and cloud-based. The architecture combines data from multiple sources, both public and proprietary, to provide the necessary data for the ORION algorithm. The infrastructure supports 30,000 route optimizations per minute. To support this mission-critical system, UPS operates two mirrored data centers, each with 300 servers. If equipment in one of these data centers fails, processing immediately switches to the backup data center. To ensure high availability, 20 SQL databases are active across three clusters. Optimizations are handled by a farm of 168 blade servers with 16 cores each. The time and distance matrices are calculated by a farm of 63 blade servers with 16 cores each.

The infrastructure eases data maintenance and provides flexible processing capacity. When a correction to a map is made, the corrected data are available nationwide within 15 seconds of making that correction. In addition, we instituted processes that continue to improve the quality of the map data. This computer
Table 1. This table shows the number of drivers who used ORION each year between 2008 and 2016.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Total no. of drivers</td>
<td>21</td>
<td>68</td>
<td>268</td>
<td>268</td>
<td>1,697</td>
<td>7,150</td>
<td>20,094</td>
<td>38,456</td>
<td>55,000</td>
</tr>
</tbody>
</table>

infrastructure has also resulted in a patent (Hurley et al. 2014).

Although we need to optimize the route only once after all packages have been loaded into a van and before its driver leaves the package center, we do additional optimizations. The loading operation begins early each morning and spans about four hours. Once every minute, ORION gathers information about the packages that are loaded in each van and develops an optimized route from scratch (i.e., it ignores the previously optimized route) for each van. Thus, the routes are always optimized and available at a moment’s notice to the local planners and drivers. The planners continuously monitor the amount of work (i.e., the time they anticipate the drivers will need to deliver all packages in a van) for each van and adjust the assigned work if necessary. The drivers can use kiosks within the package centers to view their routes on a map before they leave the center.

Because of the multiple time zones in the United States, a processing capacity of 30,000 routes per minute is sufficient to meet the needs of the 55,000 drivers. By the time the drivers from the West Coast package centers begin loading their vans, the drivers on the East Coast are already making deliveries. During the peak Christmas season, we temporarily add additional processing capacity to manage the increased number of drivers.

**Teaching an Elephant to Dance**

ORION has dramatically changed the pickup and delivery operations of UPS. Change management was crucial to make this transformation happen. We used the strategies in the following list to manage this change.

(1) Separate research from development: One of the first tasks we undertook was to make the crucial distinction between research and development. Initially, we focused on proving the concept that algorithms can build implementable routes that cost less. Once we proved the concept, we started the full-scale development and deployment of ORION. Projects that have a significant research component frequently fail because they are planned and executed as though they are normal information technology projects whose elements are typically known with a high degree of certainty. When problems arise in research-based projects, and significant cost overruns occur, there can be pressure to take shortcuts, which often lead to failure. Once we proved the significant benefits that ORION could provide, obtaining funding for its full-scale development and deployment became easy. Extensive prototyping helped us to better estimate its costs and benefits. For us to have obtained approval for a project that would take 10 years and cost $295 million dollars to develop in a single step seemed inconceivable.

(2) Stage the implementations: We decided to focus on optimizing routes, while retaining the existing dispatching practice of assigning packages to vans. We felt that simultaneously changing both routing and dispatching would cause too much disruption in operations, and would make it difficult to get user acceptance. Now that we have a mechanism to create better routes, we are working on optimizing the van-assignment problem.

(3) Do not churn more than necessary: By incorporating the base route and consistency constraints into our model, the results were closer to the experiences of our drivers, thus easing their adjustment to the new system.

(4) Be creative: Few people within UPS believed that the P&D operations could be improved, let alone that a system such as ORION was needed. To prove that O.R. could produce better routes and significant cost savings, we developed a simulation game, which is similar to a video game and which we called the routing game. The user (i.e., player) is given a set of customers and a set of time windows. The task is to find the best route. The player can create a route by manually sequencing the customers (on the computer). After each selection (i.e., the customer to serve next), the computer optimizes delivery to the remaining customers and reports
the total route cost. A player who thinks that he (she) can produce a better result can undo the selection and select a different stop. The game allows multiple people to compete against each other and compete with the system. We demonstrated this game to UPS managers at all corporate levels in meetings we held in multiple UPS locations. This effort helped managers to realize the possibility of finding better ways to route the delivery vans.

(5) Simulate: To gain the support from executives and senior managers, and to demonstrate the effectiveness of the algorithm, we developed a simulation program, the ORION ride. This program is similar to the navigation systems that are commonly used in automobiles today, with one key difference. The ORION simulator could serve multiple customers and also incorporated many UPS business and operations rules. By optimizing a real route and following that route in a physical van, we were able to simulate ORION’s operations in real conditions, and demonstrate its effectiveness. The deployment team uses this simulator to demonstrate ORION’s effectiveness to drivers before they are required to follow the optimized routes. All drivers must take two simulated ORION rides in their vans. Our current CEO, David Abney, the president of U.S. operations, Myron Gray, and a few other senior vice presidents have been on the ORION ride. This approach helped in getting acceptance and backing from UPS management and staff at all levels of the organization.

(6) Carefully select team members: We paid careful attention to forming the initial team, which was responsible for development and field testing. The team members had to be sufficiently open minded to try new ideas; at the same time, they had to know which of the existing practices should be retained. Working through disagreements was necessary. During the initial stages of developing the algorithm, the team members from the OR group approached the problem as a TSPTW problem. The members from the IE group evaluated it based on current UPS methods. Although everyone (from team members up to senior management) recognized the need for some level of consistency, defining consistency in an unambiguous way was difficult. There were disagreements and heated discussions among the team members, and progress came to a standstill. But, they continued to work together. They decided to handcraft routes using the ORION routing game to understand the intricacies of delivery routes. At first, they spent months working on a single route with multiple days of data, and later added a few more routes. This work led them to come up with a key idea—incorporating the base route concept from its current practice.

(7) If the data are not available, create the information you need: The inaccuracy of the map data was a major problem in implementing ORION, and contributed significantly to our initial failure. A faction of UPS management felt strongly that acquiring the precision map data that we needed was impossible. Because we initially estimated that 17 person days would be required to prepare a route for ORION, we had difficulty in cost justifying the project. A significant part of the 17 days involved map corrections. The project was so important that we decided not to give up. Through diligent efforts, we developed methods and systems that reduced the preparation time to six days. This reduction in the preparation time made the project appealing to management. Today our map database is a prized asset.

(8) Deploy early: The decision to not wait until we had developed a fully integrated system was also a key part of our change management. We felt that keeping the organization focused on ORION was necessary. If we had waited five years before we started the full-scale deployment, we might have lost the support of the organization. Continued deployment helped the team gain experience and fine-tune the deployment process.

(9) Monitor deployment: Based on a previous experience in which a project failed because of a less-than-rigorous deployment process, during ORION’s deployment, we closely monitored the operational performance of the entire package center during all stages of deployment; that is, before, during, and after deployment. Each package center had to meet strict entrance criteria prior to the deployment. For example, at UPS, work measurements (e.g., motion and time studies) are considered crucial for accurate planning. However, some package centers had not maintained their data, and we decided to not begin deployment at these package centers until their work measurement data were updated. We also required the package centers to validate their loop structure and the base routes before
the deployment. Similarly, each package center had to meet exit criteria and show a defined level of improved performance for a minimum of two weeks before the deployment team could move on to the next center. Thus, the deployment schedule was driven by results instead of time.

(10) Develop a training program: Training and field support were planned carefully. We developed a certification program; to become a member of the deployment team, an applicant had to demonstrate a specified level of proficiency. To provide correct and consistent training, we developed a library of nearly 90 short video clips; each was limited to a maximum duration of four minutes. A team member who needed help and (or) clarification on a topic related to ORION could access a specific video and learn from it.

(11) Hide complexity from the user and allow the user to customize: One algorithm had to meet the needs of 55,000 drivers, each of whom make about 140 to 160 stops daily for 250 days each year. To manage the diverse practical considerations, ORION provides the planners with customizable options through simple graphical interfaces. For example, due to traffic congestion, some routes must be planned conservatively so that the committed packages are delivered well before their latest committed time; other routes could push the planned delivery time to be exactly at the end of its committed window. The planner can achieve the desired level of safety by selecting a time buffer. The buffer is used to modify the time windows sent to the algorithm, and the algorithm builds the route that maintains the desired level of safety buffer. In addition, few of the users are trained in operations research; therefore, after extensive testing and tuning, we programmed the algorithm to perform some initial analysis and automatically select some critical parameters. ORION hides the algorithm complexity from the planners by providing simple graphical interfaces in a language that is familiar to them. Hence, the users (i.e., planners and those who execute the plans) do not feel intimidated by the system.

(12) Provide field support: We developed a simple mechanism that enabled the deployment team and field users to get expert help from the OR group. For example, a user might want to understand the reasons underlying a particular occurrence or get help in creating a route to meet certain local conditions. By pressing a key, that user could send all relevant data to the OR group, including questions and (or) concerns. The OR group would use specially built programs to analyze the solution and would give feedback to the user. The same mechanism also helped the OR group to enhance the algorithm and find data-related issues. We made some of these tools available to the users in the field so that they could proactively correct the data.

Impact Value and Significance

The impact of ORION may be broadly categorized as follows: (1) financial impact to UPS, (2) nonfinancial and indirect impact to UPS, and (3) impact on the OR profession. Next, we describe each category.

(1) Direct financial impact to UPS: UPS estimates the total cost of ORION development and deployment to be $295 million. As of December 2015, the system had produced cumulative savings of $320 million. Based on the results seen during deployment, ORION’s savings projections have been revised upward twice. The current estimate (at full deployment) is an annual reduction of 100 million miles driven with driver-cost avoidance and fuel savings of between $300 million to $400 million annually.

These savings and benefits are in addition to the gains made earlier by the deployment of PFT. UPS has documented these savings in its reports to the financial agencies. The UPS Business Information Analysis (BIA) group has estimated and verified these benefits. BIA is an independent group that reports to the UPS chief financial officer and is not a stakeholder in the ORION project.

(2) Nonfinancial and indirect benefits to UPS: ORION provides the following five indirect benefits, as we discuss in (a)–(e) here.

(a) Some drivers feel that ORION has made their jobs safer. Because ORION provides an optimized delivery sequence that meets multiple operational constraints, the drivers are relieved of the complexity of determining how to make their deliveries; therefore, they can concentrate on driving safely.

(b) CO₂ emissions will be reduced by 100,000 metric tons annually, and the yearly fuel consumption will decrease by 10 million gallons.

(c) ORION helps to enable UPS to maintain its delivery workforce as one of the highest paid in the industry.
(d) ORION is enabling UPS to offer additional services. For a small fee, UPS customers can request a delivery or pickup within a specific time window.

(e) ORION is used to price new services, thus giving UPS a better mechanism to determine prices.

(3) Impact on the OR profession and other organizations: ORION has brought significant recognition to the OR profession by attracting widespread media attention. Several prominent television programs, newspapers, business journals, and trade magazines have featured it. The media has used ORION extensively as an example of how OR can contribute to an organization’s profitability. During the past few years, UPS has hosted several groups from companies, large and small, that were interested in learning about how UPS is leveraging OR and (or) analytics. The U.S. Census Bureau is investigating the possibility of using OR and the ORION infrastructure to reduce the cost of the 2020 U.S. Census. The innovative feature of ORION, balancing between optimality and consistency, has sparked research interest in academic researchers. In a recent report by the Office of Net Assessment of the Department of Defense, UPS was featured as an example of how OR and system integration can alleviate resource constraints in the future (Ausubel et al. 2015). ORION was also featured in a TED Talk (Levis 2016).

Finally, ORION was responsible for the UPS sponsorship of the UPS George D. Smith Prize to encourage the administrators of academic programs to be effective and innovative in training future OR practitioners.

Why Did It Take So Long?

We frequently hear this question when we present ORION to outside groups. No doubt it took a long time, and some of it was necessary.

ORION taught us about the ignorance of our ignorance. When we started, we thought we knew everything there is to know about routing, only to realize that we did not. Few written rules—but many guidelines—were available, and the implementation of these guidelines was open to interpretation. We had to shed some of our preconceived notions and practices, and we made our share of mistakes; however, we tried not to repeat them. Sometimes, we had to act instinctively and improvise. Sometimes, we got lucky; other times, we paid the penalty.

Each time we overcame a hurdle, another would pop up. Few people at UPS believed that we could create maps to support ORION. It is ironic that UPS was a pioneer in digital map creation in the early 1990s, dissolved that part of its business in 1995 to concentrate on its core business, and then came back to it 15 years later. Time and again, we had to remind ourselves that if it was easy, someone else would have already done it. Once we realized the significance of ORION, we made a decision to ensure the deployment was right, and not to take any shortcuts in developing and deploying the system; consequently, we required more time than many would have wished.

People from many different parts of UPS, with different backgrounds and opinions, worked together, sometimes in spite of their strong disbeliefs and reservations, to make ORION happen. Once convinced off its value, they embraced it. Some deployment teams made their own ORION t-shirts, often sporting some comic variations of the ORION constellation.

How good is ORION? The bad news is that it is not as good as it can be. Being a heuristic, albeit a fast one, it misses many elements that a human would easily notice. We still have not completely succeeded in making one algorithm fully meet the diverse needs of 55,000 drivers. Some drivers can do better than ORION. In some cases, it is due to the shortcomings of our data models. In other cases, a human can act subjectively and a computer cannot. Not everyone has embraced ORION. A small percentage of drivers and managers still do not like it.

The good news is that ORION can only get better. In spite of its shortcomings, it has delivered substantial savings to UPS. With the increased deployment, ORION is gaining acceptance with less and less resistance. Just as it has detractors, it also has support from drivers who have embraced it because they view ORION as easing the complexity of their jobs. As we were completing our writing of this paper, we received the following message from a member of one of the deployment teams; none of us (this paper’s authors) knows the writer.

I enjoy all of your videos on ORION. I have been a part of the team for over 3 years now. I know all about looking like a fool to the operations managers and on road supervisors. I was a driver for 12 years before going into operations, so I get a fair amount of buyin [sic] from
the drivers in the centers I implement. Watching the lightbulb go off is a beautiful thing. When I’m showing 15 percent stop per mile gains, they can’t argue with the results. Numbers don’t lie. Thanks again for leading this project. Chaos has a way of working out when you bring forward knowledge, determination, and collaboration. I’m almost sorry to see this project come to an end this year.

Acknowledgments
The authors acknowledge people from the following UPS organizations who made ORION possible and successful. They saw the possibilities and took the risks.

- Operations research
- Industrial engineering
- Information technology
- 700 deployment team members
- U.S. operations management
- 55,000 drivers across the United States
- Management committee
The authors thank their coaches, Randy Robinson, and Ananth Iyer, for their enthusiastic support and invaluable guidance in preparing for the Edelman competition, and Hongsheng Zhong for his help in preparing this manuscript.

Appendix. ORION Routing Problem Formulation
We formulate the ORION routing problem using the TSPTW formulation proposed by Dash (Dash et al. 2012) and augment it with consistency constraints. In the vehicle routing literature, consistency is generally defined from one of three perspectives (Kovacs et al. 2014): (1) arrival-time consistency, (2) person-oriented consistency (i.e., customer served by the same driver or the same small set of drivers), and (3) delivery quantity and (or) frequency consistency. The problem we define in this paper addresses arrival-time consistency and differs from Kovacs et al. (2014) or any other problem described in the literature. Person-oriented consistency is addressed through the complementary controlled dispatch process for making van assignments. We formulated the problem to determine the sequence of customer stops for a single driver.

Let
V = Set of service locations (nodes), including customers, start locations, and end locations;
A = Set of all permissible sequencing arcs between nodes;
p = Special node added to represent the start location of the driver;
q = Special node added to represent the end location of the driver; start locations and end locations are usually the same;
n = Total number of locations that need to be visited, including the start and end locations.

Data

\[ C_{ij} = \text{Cost of serving customer } i \text{ and traveling from } i \text{ to customer } j; \]
\[ C^w = \text{Unit cost of driver wait time}; \]
\[ C^{LT}_i = \text{Unit penalty cost of lateness to customer } i; \]
\[ C^{LB}_i = \text{Unit penalty cost for delivering earlier than the earliest preferred route position for customer } i; \]
\[ C^{PD}_i = \text{Unit penalty cost for delivering later than the latest preferred route position for customer } i; \]
\[ t_{ij} = \text{Time required to serve customer } i \text{ and travel from } i \text{ to customer } j; \]
\[ ET_i = \text{Earliest time at which the customer } i \text{ may be served}; \]
\[ LT_i = \text{Latest time at which the customer } i \text{ may be served}; \]
\[ UDO_i = \text{Upper bound on the route position for customer } i; \]
\[ LDO_i = \text{Lower bound on the route position for customer } i; \]
\[ M = \text{A very large constant used to enforce and (or) relax a constraint.} \]

Decision Variables

\( x_{ij} = 1, \) if customer \( j \) is served immediately after serving customer \( i, = 0, \) otherwise;
\( S_i = \text{Service start time at customer } i; \)
\( D_i = \text{Position of customer } i \text{ in the route}; \)
\( P = \text{Position of customer } i \text{ in the route}; \)
\( W_i = \text{Wait time before serving customer } i; \) because there is a lower bound on the service start time, a driver may have to wait sometimes before starting the service;
\( OT_i = \text{Lateness in serving customer } i; \)
\( PD_i = \text{Deviation from the lower bound on the route position for customer } i; \)
\( PD^*_i = \text{Deviation from the upper bound on the route position for customer } i. \)

Minimize
\[
\sum_{(i,j) \in A} C_{ij}x_{ij} + \sum_{k \in V} C^W_k + \sum_{k \in V} C^{LT}_k OT_k
+ \sum_{k \in V} C^{PD}_k PD^*_k + \sum_{k \in V} C^{PD}_k PD^*_k
\tag{1}
\]
subject to
\[
\sum_{j \in \{V \backslash p\}} C_{ij}x_{ij} = 1, \quad \forall \; i \in \{V \backslash q\}; \tag{2}
\]
\[
\sum_{j \in \{V \backslash q\}} C_{ij}x_{ij} = 1, \quad \forall \; j \in \{V \backslash p\}; \tag{3}
\]
\[
S_i + t_{ij} - (1 - x_{ij})M + W_j \leq S_j, \quad \forall (i, j) \in A; \tag{4}
\]
\[
S_i \geq ET_j, \quad \forall \; j \in V; \tag{5}
\]
\[
S_i - OT_j \leq LT_j, \quad \forall \; i \in V; \tag{6}
\]
\[
D_i + 1 - (1 - x_{ij})M \leq D_j, \quad \forall (i, j) \in A; \tag{7}
\]
\[
D_p = 1, \tag{8}
\]
\[
D_q = n, \tag{9}
\]
\[
D_i + PD^*_i \geq LDO_i, \quad \forall \; i \in \{V \backslash (p, q)\}; \tag{10}
\]
\[
D_i - PD^*_i \leq UDO_i, \quad \forall \; i \in \{V \backslash (p, q)\}; \tag{11}
\]
\[
x_{ij} \in \{0, 1\}, \quad \forall \; (i, j) \in A; \]
\[
D_i, PD^*_i, PD^*_i \geq 0, \quad \text{and integer } \forall \; i \in V; \]
\[
S_j, W_i, OT_j, M \geq 0, \quad \forall \; i \in V. \]
The objective function minimizes the sum of the travel cost, wait-time cost, penalty cost associated with lateness in delivering a package, and penalty for violating the bounds on the service position of the customer. The travel cost consists of the driver time and mileage costs. At each customer location, time is needed to provide service. For simplification, the service time is included in the travel time from customer \( i \) to customer \( j \). Hence, the travel cost also includes the service-time cost. Constraint (2) states that some customer \( j \) has to be served immediately after a customer \( i \) is serviced, except after the last customer, which is the end depot \( q \). Similarly, constraint (3) requires that each customer \( j \) has to be served immediately after some customer \( i \), except for the starting depot \( p \). Constraint (4) states that when customer \( j \) is served immediately after serving customer \( i \), the service start time at customer \( j \) should not be less than the arrival time at customer \( i \) and the travel time from \( i \) to \( j \) (which also includes the service time at \( i \)), and any wait time at customer \( j \). Constraints (5) and (6) enforce the time-window constraint. Constraints (7)–(11) together enforce the bounds on the delivery position of customer \( i \) in the route. Constraints (1)–(6) represent the TSPTW problem, and constraints (7)–(11) represent the consistency constraints.

References


Chuck Holland is the vice president of industrial engineering at UPS. Chuck is responsible for defining and executing a vision for the optimization of UPS operations. His portfolio is responsible for process re-engineering, design and development of operations planning and control systems, and for providing operations research support. He has held multiple operations, engineering, and technology assignments (including vice president of engineering within UPS’ Supply Chain Solutions subsidiary), and his varied experience over his 39-year career perfectly positions him to engineer the next generation of UPS operations planning, execution, and control systems.

Jack Levis is the senior director of the UPS Package Process Management. A 40-year UPS veteran, Jack drives the development of operational technology solutions that make extensive use of advanced analytics to streamline and improve UPS package operations. During his distinguished career, Jack has overseen the integration of multiple planning and control systems, resulting in expansive process re-engineering. Jack has been very active in promoting advanced analytics, and was the INFORMS vice president for practice from 2009 to 2012.

Ranganath Nuggehalli is a principal scientist in the Operations Research group of UPS, which is responsible for developing advanced planning systems for use in UPS operations. He joined UPS in 1990 after completing a PhD in operations management from Purdue University. His research interests are routing and scheduling, network modeling, location analysis, system integration, and computer-human interfaces. He is an active member of INFORMS.

Bob Santilli, Director of Process Management, is responsible for providing operational technology solutions. The projects that Bob’s group manages have reengineered current systems in effort to streamline processes and maximize productivity. Under Bob’s guidance, UPS has completed
integration of multiple operations systems, requiring extensive system engineering and usability provisions. These systems ultimately synchronize the flow of data throughout UPS, allowing the seamless movement of goods, funds and information.

Jeff Winters is the director of the Small Package Operations Research Division at UPS. With 31 years of experience at UPS, he leads a team responsible for providing routing solutions for UPS’ Brown vehicle fleet. Within INFORMS, Jeff is the Roundtable 2016 VP Meetings and has served two previous terms as Roundtable President. Jeff received a BSBA degree in finance from the University of Denver and a master’s in project management from George Washington University.