This article was downloaded by: [70.83.24.169] On: 27 October 2014, At: 18:44
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA

Interfaces

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

# SPRINT: Optimization of Staff Management for Desk Customer Relations Services at Hera

Daniele Vigo, Claudio Caremi, Angelo Gordini, Sandro Bosso, Giuseppe D'Aleo, Beatrice Beleggia

To cite this article:
Daniele Vigo, Claudio Caremi, Angelo Gordini, Sandro Bosso, Giuseppe D'Aleo, Beatrice Beleggia (2014) SPRINT: Optimization of Staff Management for Desk Customer Relations Services at Hera. Interfaces 44(5):461-479. http://dx.doi.org/10.1287/inte.2014.0763

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 © 2014, INFORMS

Please scroll down for article—it is on subsequent pages

INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

Interfaces
Vol. 44, No. 5, September–October 2014, pp. 461–479
ISSN 0092-2102 (print) | ISSN 1526-551X (online)

http://dx.doi.org/10.1287/inte.2014.0763
© 2014 INFORMS

Downloaded from informs.org by [70.83.24.169] on 27 October 2014, at 18:44 . For personal use only, all rights reserved.

# SPRINT: Optimization of Staff Management for Desk Customer Relations Services at Hera

Daniele Vigo
University of Bologna, 40136 Bologna, Italy; and Optit srl, 40026 Imola, Italy, daniele.vigo@unibo.it

Claudio Caremi, Angelo Gordini
Optit srl, 40026 Imola, Italy {claudio.caremi@optit.net, angelo.gordini@optit.net}

Sandro Bosso, Giuseppe D'Aleo, Beatrice Beleggia
Hera Comm srl, Gruppo Hera, 40026 Imola, Italy
{sandro.bosso@gruppohera.it, giuseppe.daleo@gruppohera.it, beatrice.beleggia@gruppohera.it}

In this paper, we discuss a decision support system for optimizing staff management of desk customer relations services at Hera, a large Italian multiutility company. The system, SPRINT, which is based on state-of-the-art demand forecasting, implements a novel two-phase optimization procedure based on adaptive staffing. The processes developed proved to be superior to other state-of-the-art approaches. After using the system for two years, Hera has considerably improved its planning and management processes, achieved a significant level-of-service improvement of its desk customer services, and substantially increased staff productivity.

*Keywords*: customer relations management; staff management; forecasting; integer linear programming.

Hera Group, the second-largest Italian multiutility company, provides environmental services (collection, treatment, and disposal of urban and industrial wastes), water services (aqueducts, sewerage, and purification), and energy services (natural gas, electricity, and district heating). Hera, which operates mostly in the Emilia-Romagna region in northern Italy and serves more than 3.3 million citizens, had a 2012 turnover (i.e., sales) of almost 4.5 billion euros. Hera's two main business areas are environmental and water services, in which it ranks first and second, respectively, in the Italian market, and its energy sector is growing rapidly. Hera is currently the fourth-largest provider in the national gas market, and the seventh-largest provider in the electricity market (Gruppo Hera 2013).

As in other service organizations in the private sector, customer relations management (CRM) has become a major element of competitiveness in modern multiutility companies in the public or semipublic sector; these companies are increasingly viewing their users as customers. Reaching appropriate levels of CRM service is no longer a constraint imposed only by controlling authorities; it is now a fundamental tool to keep customer loyalty in a market that is proceeding toward full deregulation of services in Italy, as it is in many other countries.

The key elements of success in modern CRM systems for multiutility companies lie in effectively matching demand, consisting of the contacts the company has with existing and potential customers and supply, comprising front office CRM services, such as customer contact desks (CCDs), call centers, websites, and email, mail, and mobile phone services (see Figure 1). Components of CRM supply are integrated with the components of the large back office infrastructure that completes the service requests initiated through the various front office channels; for example, employees react to complaints that need further investigation, complete contracts when missing documents are available, or perform proactive sales activities.

Hera Comm is the Hera company-dedicated commercializing electrical energy and natural gas. Hera Comm also has CRM responsibility for the group that supports all Hera Group's main business lines (i.e., environment, water, and energy). It manages the front office infrastructure (CCDs, call centers, website, mail, and email services) and the corresponding back office. In 2012, Hera's CRM services handled almost 3.1 million user



RIGHTSLINK

Downloaded from informs.org by [70.83.24.169] on 27 October 2014, at 18:44 . For personal use only, all rights reserved.

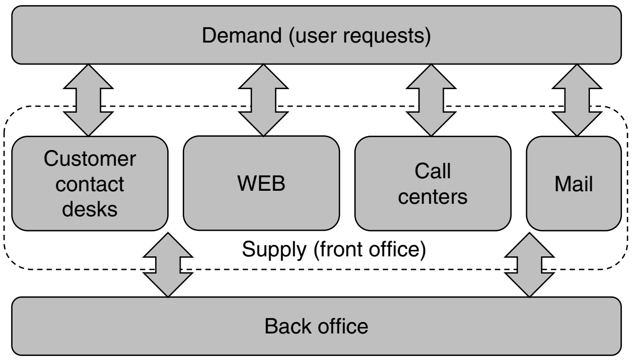


**Figure 1: In Hera's CRM system, the demand comprises user requests that are processed through the front office channels and form the supply. The back office infrastructure completes the services initiated by the front office.**

contacts, corresponding to four million service requests (one user can contact Hera for CRM-related services for multiple reasons simultaneously). Figure 2 shows the distribution of user contacts among the various front office services.

In this paper, we focus on desk CRM services, which we define as services in which customers come into a CCD and have face-to-face contact with representatives. These services are provided by Hera Comm through a network of about 80 CCDs that employ almost 200 people, serve 650,000 customers per year, and annually handle more than 1.1 million service requests. Depending on the size of the associated service territory, CCDs are classified into three categories. Eight large CCDs are in major towns, 20 medium-size CCDs are in smaller towns, and the remaining CCDs are small and located in rural areas or other special sites; the map in Figure 3 shows the location of the CCDs that Hera Comm manages. Large and medium CCDs include several (i.e., up to 20) counters at which dedicated employees service user requests. These CCDs are supervised by desk managers who plan and direct the work of the employees, including day-by-day management of the CCDs (e.g., they determine the opening and closing of specific counters in the office within a given day).

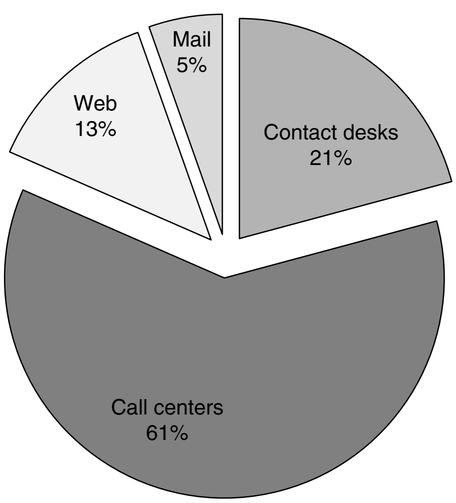


**Figure 2: The breakdown of the user contacts for Hera's CRM services during 2012 shows that contact desks serve more than 650,000 users per year (i.e., 21 percent of three million users).**

The CCDs serve both residential and business users and support customers in the resolution of a wide variety of requests, such as opening and closing accounts, checking and correcting billing issues, providing information about services, and resolving customer complaints. When not occupied with front office activities, CCD employees may handle back office tasks, such as the completion of unfinished front office requests and proactive sales activities, thus reducing the workload of the back office employees.

Given the size and the geographical distribution of its CCD network, Hera Comm maintains a central planning unit, comprised of a group of central planners who supervise and coordinate the CCD activities (see Figure 4). This unit is responsible for the system's long-term planning, including the following activities:

- Monitoring the CCD network's geographic and demand coverage.
- Determining the location and size of CCDs.
- Defining annual performance objectives or key performance indicators (KPIs); some KPIs relate to elements of service level agreements (SLAs), such as the mean waiting time of users.

In addition, the central planning unit supports desk managers in medium- and short-term planning of CCDs; examples of these activities include the following:

- Providing medium-term forecasts of the expected service demand.
- Defining staff assignments to CCDs, including scheduling holidays and vacations.
- Monitoring the performance of the system, including objectives achieved and number of employee hours,

RIGHTSLINK

Downloaded from informs.org by [70.83.24.169] on 27 October 2014, at 18:44 . For personal use only, all rights reserved.

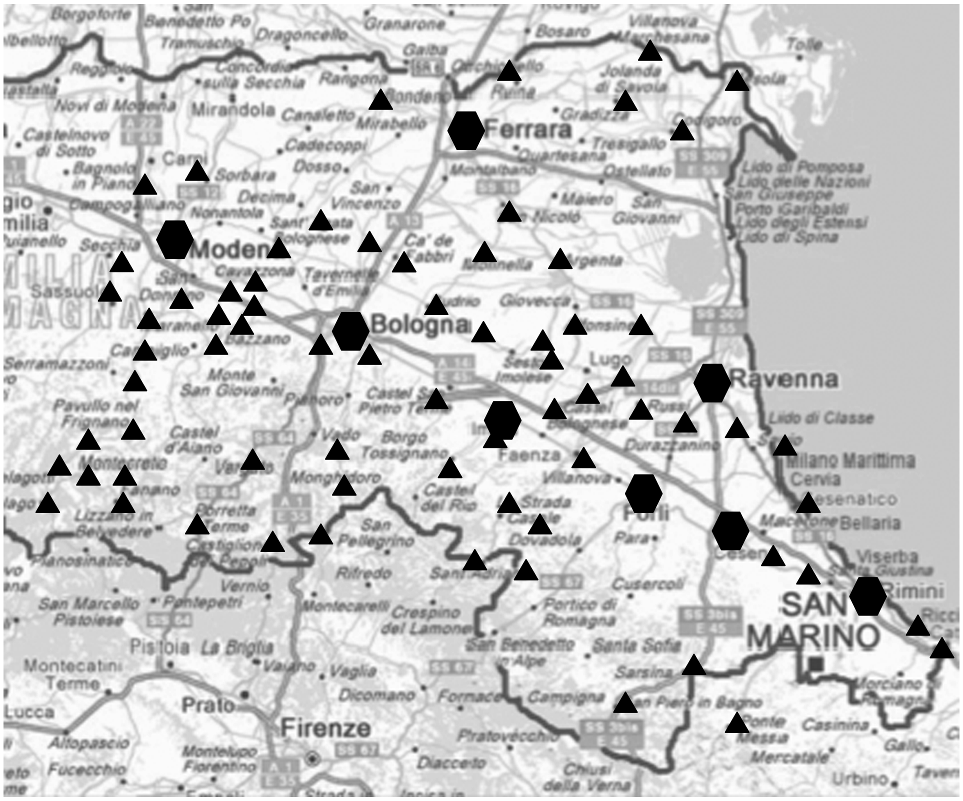


**Figure 3: Hera Comm's customer contact desks (CCDs) provide complete coverage of the eastern part of the Emilia-Romagna region in northern Italy. Hexagons represent the eight large CCDs and triangles represent the primary medium and small CCDs.**

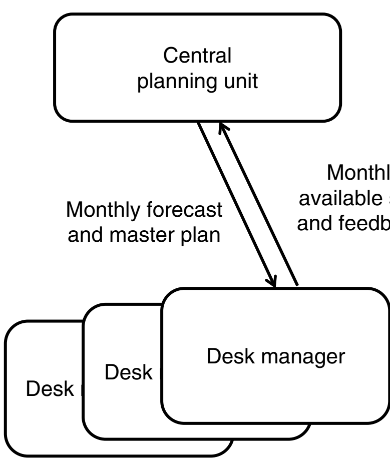


**Figure 4: A central planning unit interacts with the desk managers, performs medium-term planning, and controls CRM services.**

which are measured as the number of full-time equivalent (FTE) employees dedicated to CRM activities.

Desk managers provide feedback to the central planners about the forecasts and proposed schedules and provide information about local conditions that might influence the forecasts. They are responsible for operational management; they also fine-tune the staff schedules in response to variations in the user arrival process and availability of resources.

In such a demanding environment, the efficient assignment of personnel to front office desk activities represents a considerable competitive advantage for Hera. The time staff does not spend on front office activities can be used productively to perform the back office duties required to complete the customer service requests that need further investigation. Therefore, the challenge that desk managers face is finding the best trade-off between the efficacy (i.e., effectiveness)

RIGHTSLINK

Downloaded from informs.org by [70.83.24.169] on 27 October 2014, at 18:44 . For personal use only, all rights reserved.

required to achieve appropriate service levels (e.g., waiting times to access the desk) and the efficiency required to control the operational cost of the system. It is crucial to consider this trade-off in both the long-term design and the day-by-day management of the front office.

This paper describes an advanced decision support system, Sistema Previsionale Integrato Normalizzazione Tempi (SPRINT), an integrated forecast and time normalization system. The system provides complete management and optimization of the desk staff employees who deliver Hera's CRM services. SPRINT is the result of extensive collaboration between Hera Comm and Optit, an academic spinoff company of the University of Bologna. It includes both an accurate forecasting process to predict the arrival of users at CCDs and a two-phase staff-scheduling algorithm that incorporates a novel adaptive staffing technique. The processes developed employ a relatively simple structure that allows planners, who are not typically experts in optimization techniques, to easily understand and use the system. After more than two years of use, SPRINT has provided excellent results, in terms of both cost savings and level-of-service improvements of Hera's desk CRM services.

The paper is organized as follows: The *Background and Objectives* section introduces the various steps that led to the development of the SPRINT decision support system and the main objectives set by Hera's management. Subsequent sections describe in detail the overall approach and specific modules that comprise SPRINT—the demand forecasting module and the staff-scheduling optimizer (see Appendix A). In the *Implementation and Achieved Results* section, we discuss the benefits obtained through SPRINT since its deployment in 2011, and in *Lessons Learned and Future Developments*, we report on some lessons we learned in implementing this project and discuss future developments of the system.

## Background and Objectives

Starting in 2003 for the residential gas market and in 2007 for the electricity market, the Italian energy industry underwent a major deregulation process, as many other countries had previously. Under the supervision of national control agencies, several new companies, often created by the privatization of formerly public companies, entered the energy production and commercialization markets. Hera is an example of such a company; it is the aggregation of several former municipal agencies responsible for waste collection, water, gas, and energy production and distribution services. Since the beginning of the deregulation process, Hera has played an important role in this process, and it is now one of the major providers of gas and electricity in the Emilia-Romagna region, serving more than 2.5 million customers. At a national level, Hera is ranked fourth among gas providers and seventh among electricity providers.

The CRM service and the CCD network represent a major advantage for Hera in such a competitive market for three primary reasons: (1) such a comprehensive service is normally not available to competitors who usually manage postsales activities only through call centers or the Web; (2) Hera's residential, small-office, and home-office customers are accustomed to having superior desk services readily available to them; and (3) the quality of postsales customer care is crucial to acquiring and keeping customers. Hera's quality of service is regulated by SLAs that are measured by KPIs. The two most important SLA-related indicators are the mean waiting time (MWT) of users at a CCD and the percentage of users who must wait more than 40 minutes (PW40). Hera's CRM service has always been ranked among the best Italian services, and improving such a high-quality service in a rapidly expanding market requires significant economic and human resources.

The SPRINT project thus grew out of Hera's business requirements; the project's main objective was to enhance the quality of service of Hera's CRM system, particularly within the CCDs, without increasing costs. After some preliminary analysis, Hera management identified a way to improve both quality and efficiency by enhancing the process it uses to meet the demand and supply at CCDs. To achieve this objective, Hera Comm started a preliminary project, FAST, in 2007; through FAST, it introduced some procedural innovations (e.g., in CCD layout, desk opening and closing rules, customer-arrival forecasting and profiling, and desk staff training) based on office automation tools and standardized procedures. The experience it gained through FAST led to considerable quality

RIGHTSLINK

Downloaded from informs.org by [70.83.24.169] on 27 October 2014, at 18:44 . For personal use only, all rights reserved.

improvements and clearly showed the need to foster innovation by implementing more powerful and automated prediction and optimization techniques. In late 2009, therefore, Hera Comm initiated an ambitious project, SPRINT, in collaboration with Optit that had previously been involved in several projects related to service optimization at Hera and Hera Comm.

The objective of the SPRINT project was to design tools that achieve the following:

- Forecast the number of arrivals at the CCDs with high accuracy, and provide frequent forecast updates.
- Determine optimized scheduling and rostering of the personnel serving at each CCD, respecting the target values of SLA-related KPIs.
- Perform what-if analysis of scenarios in which the boundary conditions (e.g., the number of counters or staff personnel) can be varied.
- Organize and maintain all the information that planners require and improve their knowledge of the issues involved in CCD management.
- Define and monitor KPIs and objectives.

The main requisites established for SPRINT system design are twofold. First, give the planners access to forecasting and optimization tools in a user-friendly environment, and provide simple tools to guide them in effectively using the system to produce the desired results. Second, the system should integrate the processes required to plan, manage, and control CCDs. To this end, it must take into account the needs of central planners who, because they are responsible for long- and medium-term planning and have responsibility for operational management, must have full access to all components of the system. Desk managers must also use SPRINT; however, they should have access to only the limited components required for operational management.

Hera management defined the quantitative objectives of the SPRINT project as ambitious improvements in the five primary KPIs, which we show in the bulleted list next; the data in parentheses show the 2009 value for each objective; Hera provided us with these numbers for benchmarking purposes.

- Reduction of the user MWT to 13 minutes (previously 16 minutes).
- Reduction of the PW40 to six percent (previously nine percent).
- Increase (Hera did not quantify this increase) in the customer satisfaction index (CSI) for desk services (previously 72).
- Reduction of the yearly average of mean absolute percentage deviation (MAPD) of the forecast on user arrivals to 13 percent (previously 15–20 percent)
- Reduction of at least 30 percent of the backlog of back office (BBO) requests allocated to desk staff (previously 10,000).

The first four KPIs relate to SLAs (i.e., to service efficacy), whereas the last KPI measures the service efficiency in terms of resources saved at front office desks; these resources can be used to reduce the workload at the back office. Achieving these objectives would position Hera as an unparalleled leader with respect to the quality of CRM service among the Italian multiutilities, thus promoting loyalty among existing customers and attracting new ones.

The project quickly led to the development of prototypes for forecasting and optimization tools. In early 2010, we extensively tested these prototypes at three pilot CCDs—one large and two medium-size CCDs. The system's full-scale development started in fall 2010, and SPRINT became operational in February 2011. As we discuss in the *Implementation and Achieved Results* section, the results greatly surpassed the targets on all five KPIs, and provided relevant operational cost reductions. In 2011, as a result of SPRINT's contributions, Hera ranked first among Italian utilities for the quality of its CRM services with a mean waiting time of 10.5 minutes; the second-best result (A2A Milano) was 13.5 minutes, followed by Iren (Torino and Genova) with a mean waiting time of 20 minutes (Gruppo Hera 2012, p. 121).

## Approach to SPRINT

In this section, we introduce the approach we implemented in SPRINT. We first discuss the primary approaches that are discussed in the literature, and analyze the specific characteristics of Hera's staff-optimization problem. We then describe the overall structure of the SPRINT optimizer.

Workforce management applications attract considerable interest from the operations research (OR) community because of their economic relevance and the variety and difficulty of these optimization problems.

RIGHTSLINK

Downloaded from informs.org by [70.83.24.169] on 27 October 2014, at 18:44 . For personal use only, all rights reserved.

Ernst et al. (2004) present a general survey on this topic, whereas Blochliger (2004) provides a comprehensive tutorial for readers who are unfamiliar with the topic. The research group at the University of Bologna, who supported Optit in this project, has substantial experience in optimization applications for workforce management: Caprara et al. (1997) discuss an example of a railway crew management application.

The optimization of workforce management is typically separated into several phases that are solved sequentially to determine the assignment of employees to their daily shifts during a given time horizon. The time horizon spans several working days (e.g., a week or a month), each of which is divided into suitably small time slots that are time intervals of 15–60 minutes and represent the minimum time slot in which the daily work of an employee can be split. At Hera Comm, a time slot is 15 minutes; next, we identify the time slot either by its index or by its starting time, for example, time slot number 1 starts at 8:00 AM and time slot number 2 starts at 8:15 AM.

The service requests (or users) of various types represent the demand to be served. Each request has specific service times and some require specific employee skills.

Phase 1 is the demand forecast (DF) phase; in this phase, an evaluation of the service requests to be handled in the various time slots of the planning horizon must be determined. Subsequently, in Phase 2, the demand analysis (DA) phase, the forecast of service requests must be converted into the number of staff members required for each slot of the time horizon. Phase 3 is generally referred to as the staff-scheduling (SS) phase. In this phase, the set of daily shifts is constructed so that the demand is covered satisfactorily, whereas respecting the constraints and rules for the feasibility of the shifts (e.g., duration of the shift, number of working periods and rests, meal breaks). The final phase, Phase 4, deals with staff rostering (SR), in which the daily shifts are sequenced in lines of work (i.e., the rosters) covering the entire time horizon, and are possibly tailored to specific constraints and the daily availability of the employees.

## Characteristics of the Hera Comm Problem

The characteristics of the staff-management problem at Hera Comm permitted us to introduce a substantial simplification of the staff-management model; however, they made it impossible to use most of the existing forecasting and optimization approaches in the literature. For example, a SPRINT forecasting requirement is that it must provide the arrival rate of users at each CCD over a period of one or more months; however, as we discuss in the *Demand Forecasting Module* section, the best available forecasting processes developed for similar applications are effective over only short time horizons of one or two days.

A CCD includes several counters and offers customers a variety of services that are generally relevant to the primary user types—residential and business users; however, if special events or commercial actions suggest activation of dedicated counters, one or two additional user types can be considered. In 2007–2008, Hera Comm extensively reorganized their desk services, creating groups of counters, each serving only one user type. Although such a partition may sometimes create unfairness among users, because the staff is often able to serve different user types and the time slot is short (15 minutes), quickly adjusting imbalances in the service level is usually possible. Next, we refer only to user types that require a single service; this is the most frequent situation for residential customers, and the assumption is acceptable for business users if we consider all requests made by a given user as a single service.

Within a time horizon of one or more months, the total staff assigned to a CCD is fixed; hence, rosters are determined separately for each CCD. The CCDs are open from Monday to Friday, as is typical for multiutilities and other public services in Italy, and from 8 AM to 3 PM, a long opening interval in the Italian market. Therefore, we cannot adopt cyclic scheduling approaches, such as those developed by Bard et al. (2003), which are appropriate for a seven-day-week, 24-hour-per-day service. Staff members may not be available on specified days, for example, because of holidays or training, and they have different contract types; many work full time and some work part time, on some days only. The working shift for most employees is longer than the CCD's opening interval and, as we mention previously, they can generally serve different user types. The CCDs close at 3 PM. The few users still in the CCD at 3 PM are served by the available employees in one or two additional slots, always before 4 PM; thus, no staff overtime is generally required.

RIGHTSLINK

Downloaded from informs.org by [70.83.24.169] on 27 October 2014, at 18:44 . For personal use only, all rights reserved.

Finally, simplified meal-break rules are in force; the lunch period spans a subset of consecutive time slots, which start between 12 noon and 1:15 pm, and the employees are assigned to meal breaks as uniformly as possible during the lunch period. For example, if nine employees are working, three are assigned to each 30-minute meal break starting at 12 noon, 12:30 pm, and 1 pm, respectively. This, together with the other characteristics already mentioned , allows us to consider the personnel in an aggregated way. That is, we must know the total number of employees available on each working day of the planning horizon for each contract type, and determine how many employees are working at a counter in each time slot.

The SLA constraints to be considered relate to the MWT of the users and to the PW40. The target values for such KPIs are fixed to appropriate values for each CCD in accordance with the objectives in the SLAs for the SPRINT project discussed in the *Background and Objectives* section (e.g., 11 to 12 minutes for the MWT and five to six percent for the PW40, respectively). Additional constraints may be imposed to avoid opening or closing counters for too short a period. In particular, a counter should be closed for a sufficiently long period (e.g., one hour) to allow staff to handle back office duties without interruption.

Once we compute the aggregated schedule, determining the shifts to be assigned to individual employees is generally straightforward. Moreover, because the desk opening hours coincide with the working-shift duration of the vast majority of employees and assigning personnel to work on weekends or holidays is unnecessary, staff rostering is simple and does not require specific optimization tools. We then generate the service roster for the personnel at a CCD through the repeated solution of a daily optimization problem, which determines the number of counters to be open, the opening profiles, and the number of employees required to provide satisfactory service at a CCD for the predicted arrival rate.

The typical user arrival rate is high (i.e., 20–60 arrivals per hour), as Figure 5 shows, and the average service time of a user at a counter varies from 10 to 22 minutes. Thus, it is comparable to the duration of a time slot, which is 15 minutes. As a result, the system is highly saturated and cannot be optimized using traditional

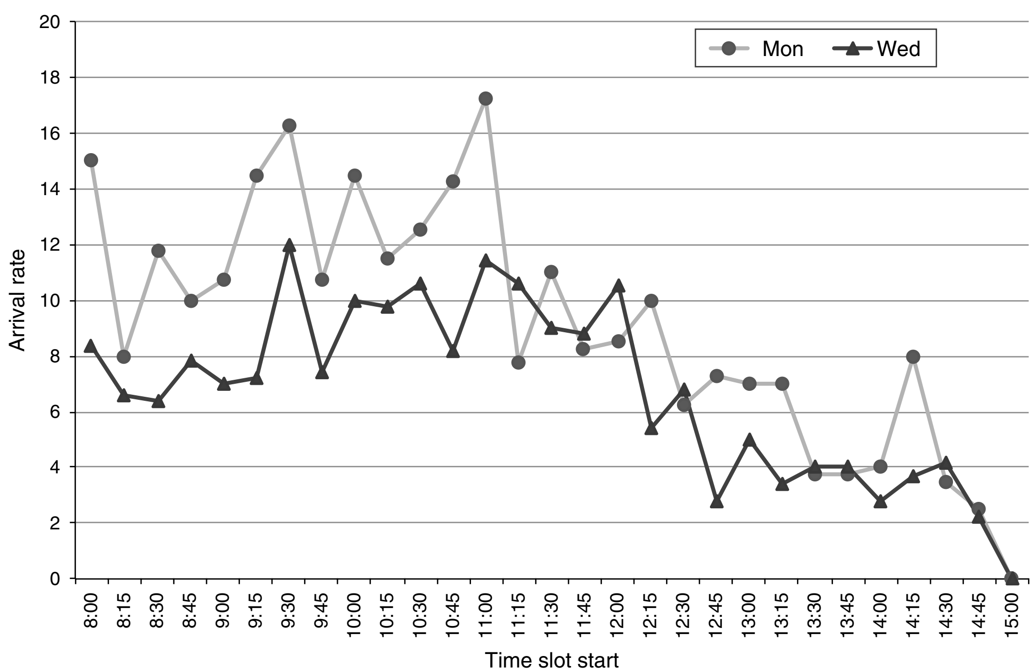


**Figure 5: The distribution of user arrival rate by time slot at a large Hera Comm CCD on two weekdays shows that between 20 and 60 users per hour may typically request service. Time slots are identified by their start times.**

RIGHTSLINK

Downloaded from informs.org by [70.83.24.169] on 27 October 2014, at 18:44 . For personal use only, all rights reserved.

approaches, such as the static independent period-by-period (SIPP) process that Green et al. (2001) discuss; the SIPP process determines the staffing required to reach an adequate level of service separately for each period. Then, it optimizes the schedule to minimize the overall staff cost.

In situations in which a CCD is highly congested, where staffing cannot be determined independently for each period, introducing a level of integration between the DA and SS phases (i.e., Phases 2 and 3) is appropriate if we are to manage the stochastic nature of the problem. A common way to achieve such integration is to adopt a two-phase approach based on two interacting algorithms. The first is a schedule generator (SG), which produces a scheduling solution based on the current staffing requirements; the second is a schedule evaluator (SE) that estimates the cost and feasibility of the current solution. For other examples of two-phase processes used to solve various applications, see Atlason et al. (2008) and Ingolfsson et al. (2010) (call centers), Kabak et al. (2008) (retail sector), and van Dijk and van der Sluis (2006) (airport check-in). The two modules interact iteratively by redefining the staffing requirements whenever the service levels are not met or the cost measured in terms of FTE employees is excessive, until a satisfactory solution is found (see Figure 6).

**Structure of the SPRINT Optimizer**

The overall algorithmic approach adopted in SPRINT determines, separately for each CCD, the arrival forecast for the entire planning horizon using the approach we describe in the *Demand Forecasting Module* section. The daily shifts for the available staff are then computed for each CCD and each day of the horizon by using the two-phase iterative algorithm illustrated in Figure 6. The algorithm consists of the following steps.

(1) The DA module determines the required staffing for each time slot using an innovative adaptive staffing approach. Adaptive staffing takes into account the congestion of the queue system, thus indirectly the constraints on the SLAs, through the value of a system parameter *beta* (see the *Data Analysis Module and Adaptive Staffing* section).

(2a) The SG module computes, through a dedicated integer linear programming (ILP) model, the shifts that minimize the staff required to serve the users (see the *Staff-Scheduling Optimizer* section). The optimization model does not explicitly consider the (nonlinear) constraints on the SLAs: thus, the solution determined may be infeasible.

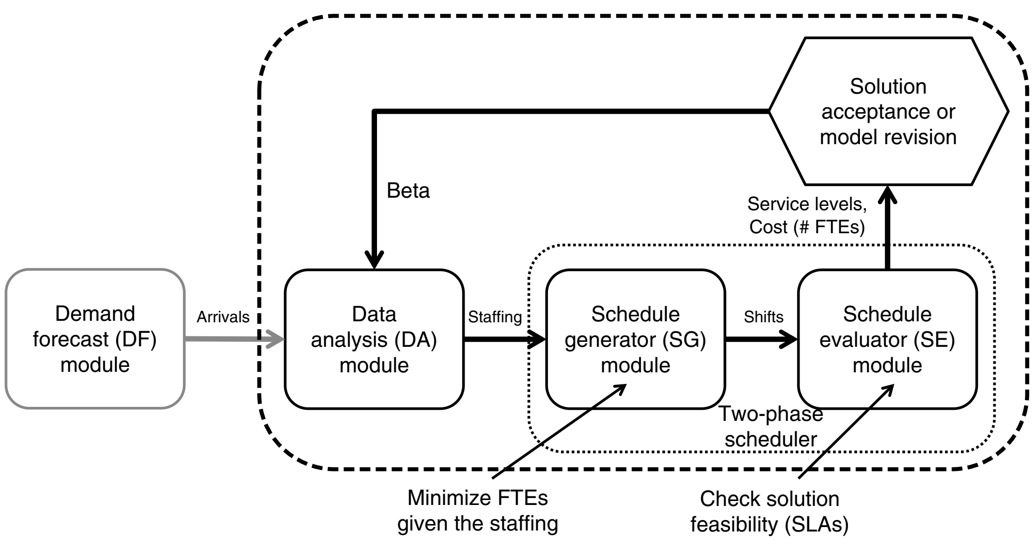


**Figure 6: The SPRINT optimizer is organized into several modules and provides optimized staff scheduling, given the forecast of the demand to be served. At its core is a two-phase scheduler that iteratively generates the schedules and evaluates their performance until it finds a satisfactory schedule.**

Downloaded from informs.org by [70.83.24.169] on 27 October 2014, at 18:44 . For personal use only, all rights reserved.

(2b) The SE module computes the average values of the SLA-related KPIs (i.e., the MWT and PW40) by using a fast engine that simulates the working day at the CCD based on the current staff schedule.

(3) The overall feasibility of the schedule is evaluated by comparing its KPIs with the target values. If the current solution is infeasible, the current value of parameter *beta* for the adaptive staffing is updated to increase the staffing in the next iteration. Otherwise, *beta* is updated to reduce the staffing. The algorithm stops if it either finds a satisfactory solution or reaches a maximum number of iterations; otherwise, it returns to step 1.

We initialize the previous algorithm with a starting value for the parameter *beta* and return either the best feasible solution found or the least infeasible one. This latter case may sometimes occur during days with an exceptionally high arrival rate or when a CCD is particularly understaffed; however, the infeasibility may be offset in the next days by decreasing the target values for the SLAs at that CCD.

In the following sections , we describe the main modules of the SPRINT optimizer; Appendix A provides additional technical details.

## Demand Forecasting Module

Demand forecasting is a crucial step in the SPRINT approach. Effective staff scheduling clearly relies on accurate and detailed information about the future demand to be met.

As we discuss previously, the planning processes involve mainly medium-term functionalities. Hence, the forecast must be performed for periods of one or more consecutive months; for each CCD and each day of the future planning horizon, both the arrival process and the service-time distribution must be determined.

Two types of input data are available at Hera Comm to feed the forecasting model. The main time series is the detailed data about each user who arrives. This data source has been available since 2008–2009; for each CCD, we can use it to derive the type of user, arrival-time distribution, service duration, rejection rate, and opening interval of each counter. In particular, the data set that includes time-related data about the total number of arrivals per day and per CCD is the main source of input information for the forecast. A second relevant set of data includes the number of bills sent out for each CCD each day.

The best existing approaches proposed to forecast the arrival process in similar applications, such as call centers, are based on univariate time-series models (Taylor 2008). According to Taylor, however, these approaches are empirically competitive with simpler ones, such as deseasonalized historical averages (HAs), only for short-term predictions of one or two days. Prior to starting this project, Hera Comm had been using HAs to forecast the arrivals. However, it considered the HA quality to be insufficient for effective planning.

When compared with the observed arrivals at Hera Comm CCDs, HAs generally showed a MAPD that was 15–20 percent larger than the observed values. For example, during the first five months of SPRINT operations, the MAPD of the mean value for HAs was as high as 16.4 percent (see Figure 9). In addition, the analysis of the existing historical data showed the strong influence of the billing process on the arrival rate in a day, because information and complaints about bills are a major reason that users access the CCDs.

For all these reasons, we adopted a hybrid algorithm that uses an M5 model tree (Quinlan 1992) to predict the daily user arrivals, coupled with a simple top-down approach to separate this total into forecasts for each time slot and user type. The M5 model tree in Quinlan (1992) is a popular approach that derives the specific mathematical model from a general framework and is therefore best suited to predict the behavior of complex systems and data mining; see Witten et al. (2005) for an introduction to model trees. The M5 model tree combines classification and regression features in a simple and effective way; a tree-structured regression is built on the assumption that functional dependency between input and output values is not constant in the domain of the explanatory variables (e.g., the day of the week or the information on the billing process), but can be considered as such in smaller subdomains. Therefore, the M5 model tree represents the division of the input domain into subdomains; we develop a specific multivariate regression model for each subdomain. Figure 7 illustrates the input domain partitioning and model-tree construction. Here, different models are constructed depending on the weekday and on billing activity level, and the result on the $y$ axis is

RIGHTSLINK

Downloaded from informs.org by [70.83.24.169] on 27 October 2014, at 18:44 . For personal use only, all rights reserved.

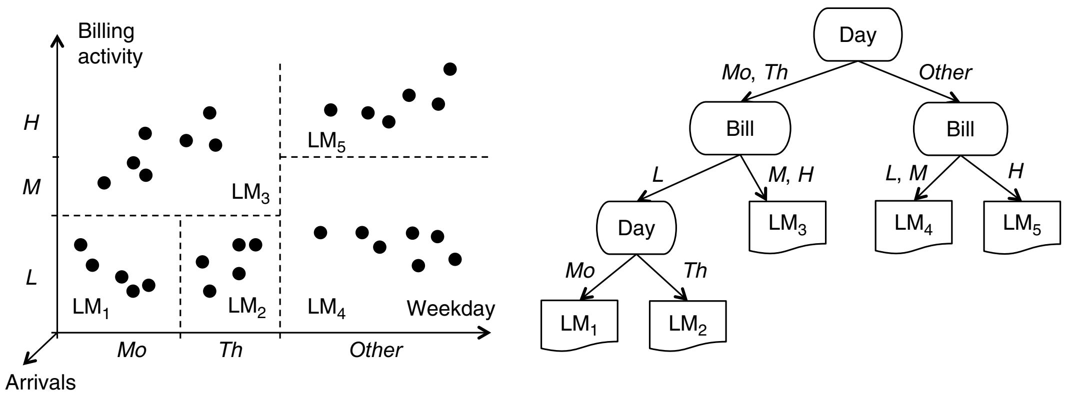


**Figure 7: Model trees partition the input space into subdomains (right) and construct a specific model for each subdomain (left). The leaves of the tree correspond to the linear models $LM_1$–$LM_5$ associated with these subdomains.**

the number of total arrivals in the day, which is later disaggregated by type of user and time slot, as we describe next. In practical applications, these models proved to be as accurate as competing approaches (e.g., artificial neural networks), but could be trained more quickly and provide a structural representation of the rules of the model that practitioners can understand easily. Bhattacharya and Solomatine (2005) show a water sector application. Central planners and desk managers easily understanding the model turned out to be as crucial as the forecast accuracy in promoting their acceptance and use of SPRINT.

Once the forecast of the total arrivals has been determined for each day of the planning horizon, we divide it into forecasts for each time slot and user type based on a top-down approach. For each site, we use the previous month's average distribution of arrivals by time slot and user type on specific days of the week (see the example in Figure 5). As our previous experience with the hybrid top-down approach has shown, it produced much better results than directly estimating the arrivals per time slot and user type through an M5 model. Using the same data source, we also compute the distribution of service time per user type at each CCD (see Figure 8).

The quality of the forecast obtained with the hybrid model has consistently proved to be very good. Figure 9 summarizes the results we obtained for the first five months of 2011, after the SPRINT system became operational. This figure relates to the total of the eight large CCDs that serviced more than 120,000 users during these five months. For each month, the figure shows the average percentage ratio of the absolute deviation between the forecast and the observed value (MAPD) for our system, denoted as hybrid forecast (HF), both with and without information on the number of bills issued. The same information is also reported for a benchmark forecast process, marked as HA, based on deseasonalized historical average. Finally, the chart clearly shows the positive impact of including in the model information on the number of bills issued, which further reduce the forecast's MAPD by about 6.5 percent. The quality of the results is also confirmed for the period from March 2011 to April 2013, as Figure 10

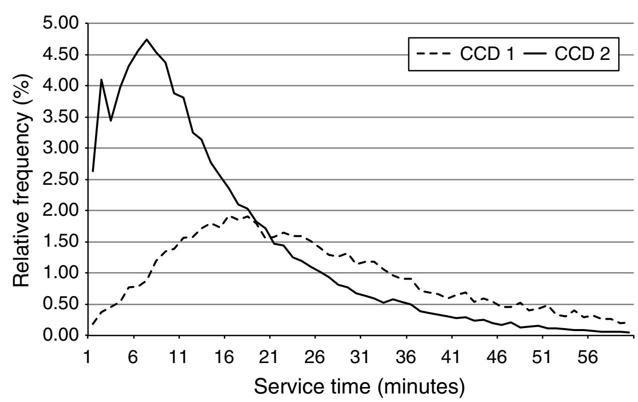


**Figure 8: The distribution of CCD service times shows that the average is close to 15 minutes, which is the duration of the planning time slot. The figure shows two CCDs in which CCD2 is twice as large as CCD1.**

RIGHTSLINK

Downloaded from informs.org by [70.83.24.169] on 27 October 2014, at 18:44 . For personal use only, all rights reserved.

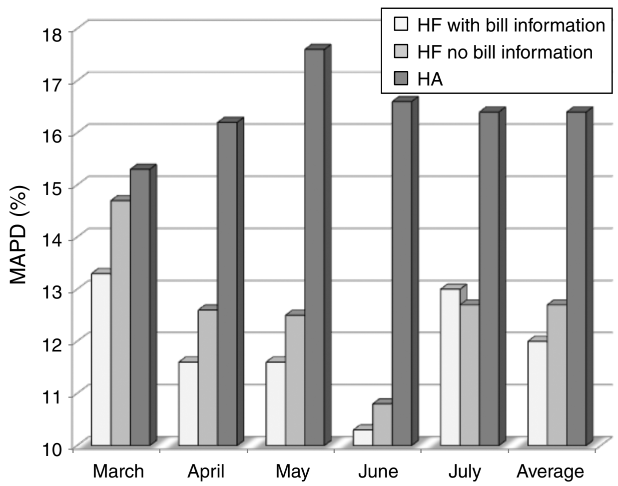


**Figure 9: During the first five months of SPRINT operations in 2011, the performance of the hybrid forecast (HF), both with and without information on the number of bills, was better than that of the historical average (HA) that Hera Comm used previously.**

illustrates; MAPDs are almost always between 10 and 15 percent, with an average of 12.3 percent; thus, they surpass the HA competing approaches by more than 30 percent.

The results of our forecasting approach had a positive impact on Hera Comm's planning process. It allowed the company to reliably predict arrivals; the MAPD error objective was approximately 12.3 percent—below the very ambitious objective of 13 percent that we defined at the beginning of the project. We can attribute part of this success to the system's user friendliness and its ability to generate the yearly forecast for a CCD in less than five minutes of computing time.

## Data Analysis Module and Adaptive Staffing

The staffing requirements for a given working day of a CCD are determined through an innovative approach that we term adaptive staffing. This approach extends other well-known staffing approaches from the literature, such as LagMax (Green et al. 2003)that define the actual staffing in a time slot as a function of the user arrival rate in a subset of slots around the current one. Once the staffing requirement for each time slot has been determined, the staff shifts can be finalized using either manual or optimization techniques, which must explicitly consider the behavior of a congested queuing system.

How we define the staffing in our approach is particularly suitable for the staff-scheduling problems we attempt to solve through SPRINT (i.e., the average value of service time for the users is similar to the

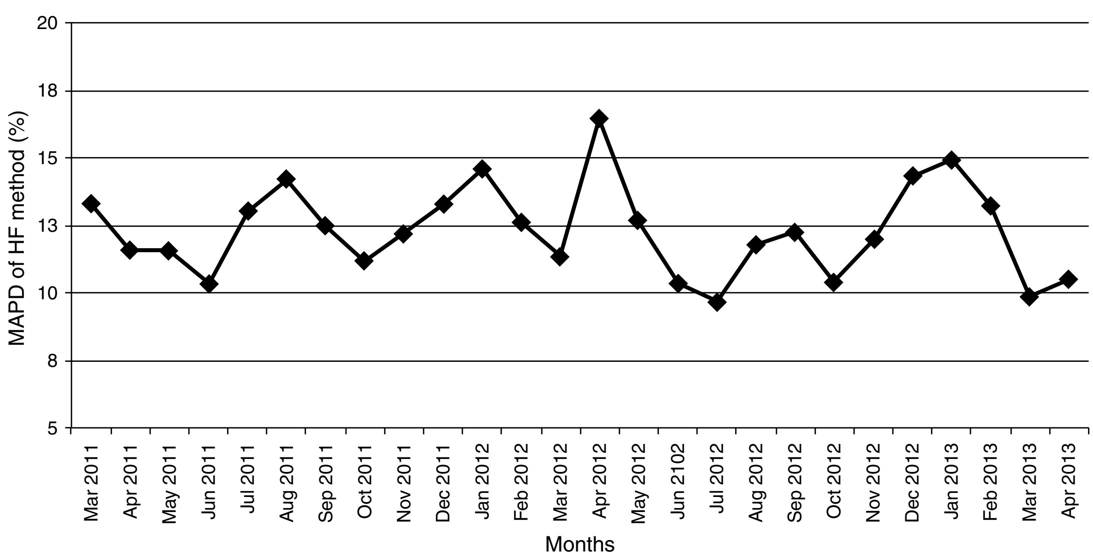


**Figure 10: During the first two years of SPRINT's operations, the forecast's MAPD with respect to observed values remained fairly stable at around 13 percent.**

Downloaded from informs.org by [70.83.24.169] on 27 October 2014, at 18:44 . For personal use only, all rights reserved.

slot duration). As Figure 8 shows, the average service times in our case are typically between 10 and 22 minutes, and our planning slot is 15 minutes. In this environment, alternative staffing rules based on the approaches described in the literature generate poorer results.

In the adaptive staffing approach, we anticipate arrival peaks by defining the staffing requirement in the current slot as a function of a weighted sum of the arrivals in the current and subsequent slots; that is, we aggregate all the users of a given type arriving in a time slot, and impose the constraint that the available resources, expressed in terms of total working minutes at the counters, are sufficient to serve a fraction of the total service time of such users. The remaining fraction of users is implicitly assumed to be served in subsequent time slots, with the users that arrive during that time slot. In our case, the service times are relatively short with respect to the maximum time a user may stay in the system. Hence, it is acceptable to postpone the start of the service to time slots that follow the customer's arrival slot, thus allowing for a reduction in the number of required open counters in a given time slot. Appendix A shows the specific formulas we used to define the staffing and the weights for each time slot.

As previously discussed, the staffing behavior is influenced by a single global parameter, *beta*, whose values vary between 0 and 1, which allows us to control the distribution of service for the users who arrive during the ensuing time slots. That is, by varying *beta*, we can impose either an under- or oversizing of the service to compensate for the linearization of the queuing effect and the other approximations of the staff-scheduling model in our two-phase approach. As Figure 11 illustrates, the role of *beta* on model behavior is as follows: the smaller the value of *beta*, the larger the percentage of user service time covered in a time slot. For example, if *beta* equals 1.0, the staffing is determined so that about 50 percent of the user service is completed during the time slot in which these users arrive. The service percentage then gradually increases in the subsequent one, two, or three time slots, reaching 80, 100, or 110 percent, respectively; with this *beta* value, the users are likely to be served within at most three time slots after their arrival. However, when *beta* equals 0.0, the staffing number is larger and the service

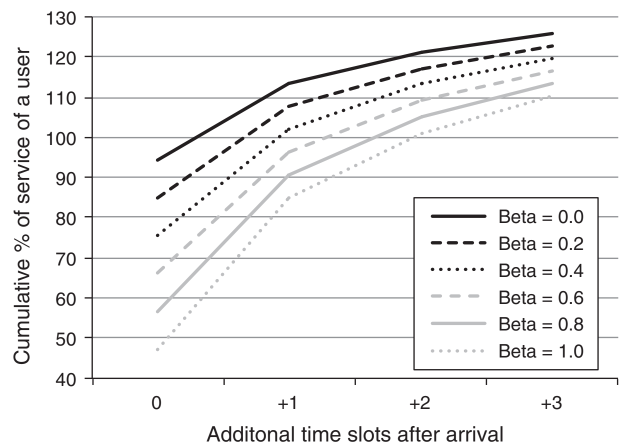


**Figure 11: Adaptive staffing permits a planner to use a single parameter, *beta*, to control the percentage of service time of a user arriving at a time slot that is processed within a small set of subsequent slots.**

for all users starts during either their arrival time slot or the next one. Note that, because the aforementioned linearization of the queuing effect is inherent in the two-phase approach, the percentage values may be larger than 100 to increase the probability that staffing is sufficient to serve the users within two or three time slots.

The simple approximation technique we implemented within the adaptive staffing module rapidly produces high-quality schedules; in addition, a single parameter allows users who are not expert in mathematical optimization to determine the best compromise between required staffing and desired service quality.

The experience we gained during more than two years of testing SPRINT showed us that adaptive staffing is a key element in the successful application of the two-phase optimization approach. In other situations (e.g., with much shorter service times), the benefit with respect to staffing techniques, such as SIPP or LagMax, is generally smaller. However, we conducted preliminary testing on a benchmark set containing instances characterized by increased average service times and different average arrival rates. For each instance, we ran the two-phase algorithm by using the adaptive and LagMax approaches to determine the staffing requirement during each time slot. We consistently observed that (1) in moderately congested conditions, both approaches determine feasible solutions, but the adaptive solution uses fewer

RIGHTSLINK

Downloaded from informs.org by [70.83.24.169] on 27 October 2014, at 18:44 . For personal use only, all rights reserved.

staff resources, (2) in more congested situations, adaptive staffing finds feasible solutions, whereas LagMax almost always violates the SLA's target values, and (3) in heavily congested situations, neither process finds feasible solutions, but the adaptive approach produces much smaller SLA violations.

## Staff-Scheduling Optimizer

We adopted the two-phase approach (see Figure 6) to determine staff rostering for a given planning horizon, which spans one or more months, by repeatedly solving daily problems for each CCD.

The daily staff-management optimization problem requires minimizing the resources needed to service the users arriving on a given day, while respecting the imposed target values for SLA-related KPIs (i.e., the MWT and PW40) and operational constraints (e.g., a counter's minimum or maximum opening hours). Because the adaptive staffing mechanism only implicitly models the SLAs depending on the current value of the parameter *beta*, we check these constraints through the fast simulator of the SE module. The model optimization is iterated with a new *beta* value, until a satisfactory solution is found, in terms of SLA target values and resources used. Hence, the main objective of the scheduling optimizer is to minimize the resources used to service the customers under the current staffing process. We achieve this by maximizing the number of time slots in which staff members are not used to provide front office duties at the CCD; thus, they are available to perform back office operations. The objective function also incorporates two penalty terms. The first is used to reduce the possible unserved user demand, whenever a feasible solution is not found because of insufficient staff availability. The second reduces the number of counter openings.

We formulate the resulting problem as an ILP model, which includes specific constraints on the available staff in each time slot, the minimum opening and closing time of each counter, and staffing. Appendix A shows the ILP model. For size instances to be solved within SPRINT, the optimal integer solution can be obtained quickly using modern ILP solvers. For example, by using the well-known public domain solver GLPK with cuts and feasibility pump enabled, an instance with 32 time slots of 15 minutes each, four user types, and 11 counters may be solved in a few seconds on a standard personal computer. The use of professional solvers, such as IBM CPLEX, Gurobi, or Fico Xpress, may provide solutions in much shorter run times.

As previously discussed, given the approximations generated by the adaptive staffing technique, we cannot guarantee that the counter-opening profiles obtained by the ILP model are completely feasible with respect to the desired SLA target values. Therefore, we use a fast simulation model to evaluate MWT and PW40. The simulation model assumes a nonhomogeneous Poisson process for the arrivals and an Erlang distribution for the service time. The parameters of these distributions are derived from the historical data of the specific CCD. To achieve satisfactory stability of the results, we run the simulation model for 1,500 repetitions, which require a few seconds of computing time.

Evaluating the FTE-employee savings achieved by using SPRINT at CCDs is difficult to quantify, because the historical data available are insufficient to reconstruct reliable benchmark scenarios; however, we performed individual analyses for the largest CCDs during the model's validation in 2010. These analyses consistently showed that the SPRINT two-phase optimizer produces schedules that, when compared to manually generated schedules, (1) drastically improve the service quality with the same number of resources (i.e., personnel), or (2) make available more resources for back office functions, while maintaining the same service level solutions.

The optimizer and the simulator interact, as Figure 6 depicts, where the control parameter *beta* is used at each iteration to alter the optimizer behavior. The first tentative solution is calculated by using an initial intermediate value for *beta* parameter (generally equal to 0.5). Then a binary search is performed to calculate the maximum *beta* value that corresponds to a solution whose MWT and PW40 values are smaller than the target values, stopping whenever the difference between consecutive parameter values is below a given threshold. For example, when the threshold equals 0.1, at most four or five iterations are performed. Thus, only a few seconds are required to optimize the daily work of a CCD. Clearly, the user can run a single optimization-simulation with a prespecified *beta* value. Note that whenever the arrival process makes it

RIGHTSLINK

Downloaded from informs.org by [70.83.24.169] on 27 October 2014, at 18:44 . For personal use only, all rights reserved.

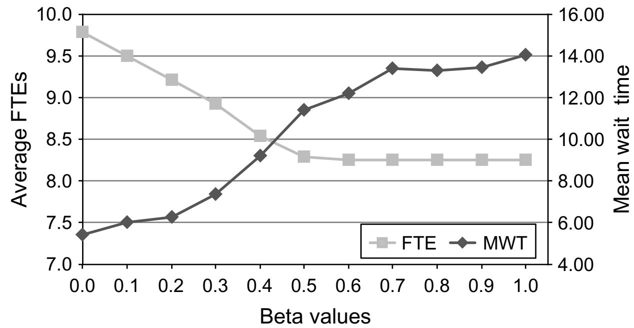


**Figure 12: With adaptive staffing, by changing the value of *beta*, we can easily determine the appropriate compromise between service quality, measured by the MWT, and the required FTE employees.**

impossible to achieve a feasible solution with respect to the desired SLA target values, the binary search determines the solution that has the least number of violations.

Figure 12 shows the typical relationship between the value of parameter *beta* and that of the SLA target, as evaluated by the simulator. For *beta* values ranging from 0.0 to 1.0, the figure reports the values of one SLA-related KPI, the MWT, and corresponding values of resource consumption, measured in FTEs. In this case, the target value for the MWT is 13 minutes. The best compromise solution is therefore found when *beta* equals 0.6.

### Proactive Optimization Strategy

During 2012, after one year of SPRINT usage, we conducted some preliminary analyses to determine the impact of using the staff scheduler to improve performance. In particular, we examined the results obtained by some desk managers who began to use the staff scheduler when they detected a considerable deviation between the observed and forecast arrival rates in the first few hours after opening a desk. We observed that this proactive strategy permitted them to take quick action in the face of a substantial demand increase (e.g., because of accidental billing errors or a new commercial campaign or product) or decrease (e.g., because of weather conditions, or unforeseen variations in public transportation services) by opening and closing some counters.

As we describe in the next section, the use of such a proactive strategy greatly helped in maintaining stable performance of the desk CRM services, even in presence of a considerable increase in demand and the activities desk personnel were required to perform.

## Implementation and Results Achieved

We developed the SPRINT software solution using a Google Web Toolkit and Enterprise Java Beans 3 framework, and delivered it as a software-as-a-service (SaaS) solution; both central planners and desk managers (see Figure 4) share a multiple-user Web interface (see Figure 13). The central planners load via FTP massive amounts of input data, such as time series of user arrivals, service times, and number of bills. Appendix B provides more details on the structure of the software system.

In this section, we outline the results we obtained since we deployed SPRINT in February 2011. For the first nine months, we used SPRINT for only the eight large CCDs operating in major towns; in November 2011, we began to gradually implement it in the 14 medium-size CCDs. It now operates at 22 CCDs that handle more than 75 percent of the total yearly user contacts and 85 percent of the service requests.

After two years of SPRINT operations, Hera Comm management performed an extensive evaluation of SPRINT's impact on the CRM process; this analysis showed that several quantitative and qualitative benefits resulted from SPRINT usage.

The quantitative impacts of SPRINT are best measured at an aggregate level. Table 1 shows that over this two-year period, Hera Comm surpassed all its initial primary KPI targets. All SLA-related KPIs (i.e., MWT, PW40, and CSI) improved substantially during the first year. We also saw an impressive increase in the effectiveness of the resource usage, represented by a substantial reduction in BBO as a result of an increase in time available for back office duties by the desk staff. It is worth noting that such achievements were obtained despite a large increase in customer contacts and a staff reduction of four employees. As we state previously, these results allowed Hera to rank first for quality of service among Italian utilities in 2011.

By 2012, the planning and operations staff had become familiar with the SPRINT tools; we attribute this familiarity with the good business results achieved because the desk staff was able to provide more support

RIGHTSLINK

Downloaded from informs.org by [70.83.24.169] on 27 October 2014, at 18:44 . For personal use only, all rights reserved.


**Figure 13: SPRINT has a simple front-end interface that displays all relevant information about the staff-scheduling solution.**

| | 2009 (pre-SPRINT) | Mar 2011–Dec 2011 | | Jan 2012–May 2013 | |
|---|---|---|---|---|---|
| Average monthly user contacts at the large CCDs only | 20,500 | 23,700 | +16% | 25,900 | +26% |
| MWT: Mean waiting time (min) | 16.00 | 10.32 | −36% | 10.37 | −35% |
| PW40: % of users waiting > 40 min | 9.00 | 4.60 | −49% | 4.60 | −49% |
| CSI: Customer satisfaction index | 72 | 78 | +8% | 81 | +13% |
| Staff available | 193 | 189 | −2% | 188 | −3% |
| BBO: Backlog of back office allocated to desk staff | 10,000 | 1,500 | −85% | 600 | −94% |
| No. of monthly sales activities by desk staff | 417 | 1,420 | +241% | 3,360 | +706% |

**Table 1: During the first two years of SPRINT usage, CCD contacts increased progressively. In addition, all KPIs related to service, quality, and cost improved considerably.**

to sales activities. Performance and service indicators remained substantially the same, despite an additional 10 percent increase in demand; the result was an approximate tenfold increase in the number sales actions performed monthly by desk staff in 2012 relative to 2009. This indicates SPRINT's capability to effectively support strategic business activities, while controlling resources and KPIs.

Putting the savings in an appropriate context is difficult because an analysis of the economic impact of these savings and the revenue generated is currently in progress. Clearly, a reduction of three percent in FTE desk staff employees, in conjunction with a demand increase of more than 25 percent, although Hera maintained a better level of service than that of its competitors, is noteworthy. Prior to the SPRINT implementation, maintaining a constant quality of service required resource increases at desk staff proportional to that of the demand increases. In addition, the high quality of CRM service has been instrumental in Hera's consistent growth in the highly competitive Italian energy market.

Downloaded from informs.org by [70.83.24.169] on 27 October 2014, at 18:44 . For personal use only, all rights reserved.

During the first two years of SPRINT usage, Hera's planning process improved rapidly as a result of the availability of accurate forecasts that could be generated within in a short time. The staff scheduling solutions compare favorably to previous manually generated schedules in terms of speed and because they provide a good balance between the effectiveness and efficiency of the solutions. Finally, both the forecast and staff optimizer are used widely for scenario analysis to support strategic decisions, such as sizing resources and performing sensitivity analysis on factors outside the control of the CCD; these include increases or decreases in demand, impacts of commercial campaigns, and modifications to the number and structure of CCDs. Another qualitative benefit is the achievement of greater effectiveness of the human resources dedicated to planning activities. The greater efficiency achieved by using the tools that automate repetitive and time-consuming tasks increases staff ability to concentrate on value-added tasks, such as monitoring and control activities and strategic decision making. Prior to the SPRINT implementation, planners did not have a consistent management strategy on which to base their planning activities; each was free to adopt his (her) own strategy, which differed across the CCDs. In addition, when we monitored SPRINT's use among desk managers, we saw that the percentage of days in which they use SPRINT to monitor or replan services averages far above 90 percent. Finally, SPRINT allowed Hera to more fully exploit management data used previously for management summary reports.

## Lessons Learned and Future Developments

We have described SPRINT, a state-of-the-art decision support system that optimizes scheduling and rostering of customer contact desk personnel for a large multiutility company.

SPRINT represents the final step of an extensive organizational effort that Hera Comm began in 2007, two years before this project began. From 2007 to 2009, management focused heavily on promoting the widespread and uniform use of existing data to support planning activities across its CCDs (e.g., through a simple demand forecast process and the design of specific planning policies). This effort led to significant improvements in service quality and resource usage. For example, between 2006 and 2009, the mean waiting time at CCDs decreased from 24 to 16 minutes.

The need for additional improvements in its approaches motivated Hera Comm to develop the SPRINT system in conjunction with Optit. Therefore, we strongly believe that the systematic use of the tools within SPRINT is responsible for most of the improvements obtained with the SPRINT system for two primary reasons. First, during the project, SPRINT was implemented in large CCDs that Hera Comm incorporated in 2009. The KPI improvement observed at this CCD was much larger than at other CCDs. We see this as an indirect measure of (1) a structured planning approach, and (2) specific tools implemented in SPRINT. As Table 1 illustrates, the use of such tools allowed Hera Comm to maintain high quality of service, even as the total demand increased considerably; in contrast, using Hera's previous planning processes, performance generally followed the demand trend.

One key factor in SPRINT's success is its considerable user friendliness; it uses simple interfaces in which all key information is displayed. A second factor is the stability of its performance over the time of the forecast and optimization algorithms. As a result, SPRINT has become a widely used and accepted support tool by both central planners and desk managers. SPRINT demonstrated the capability of operations research tools to effectively support business strategic actions while controlling resources and KPIs.

The SPRINT system and the core forecasting and optimization tools are being continually improved. In the near future, our development activities will focus on the full implementation of the proactive rescheduling policy tested in 2012, and with a refinement of short-term forecast by using univariate techniques from the literature. In addition, we will partially redesign the data management and reporting components to reduce the time required for several time-consuming processes within the forecast and reporting systems. For the staff-scheduling model, we are developing an overall stochastic optimization framework for constructing more robust solutions with respect to the arrival process.

Finally, Optit generalized and incorporated the core elements of SPRINT into an innovative product for desk

RIGHTSLINK

Downloaded from informs.org by [70.83.24.169] on 27 October 2014, at 18:44 . For personal use only, all rights reserved.

optimization, SPOT, which another Italian multiutility company adopted in early 2012.

## Acknowledgments
The SPRINT project is the result of the efforts of a large team of people at Hera Comm and Optit. The success achieved is largely the result of their enthusiastic and highly professional commitment. In particular, we thank Valerio Vannini and Giulia Biancardi at Hera Comm, and Fabrizio Antonioli, Fabio Lombardi, and Mattia Manfroni at Optit. We also thank the anonymous referees and the editor for their valuable comments.

## Appendix A. Adaptive Staffing and the ILP Model for Staff Scheduling
Table A.1 includes the main input parameter and decision variables used in the adaptive staffing process and in the optimization model for staff scheduling. In our approach, we measure time in discretized time slots of 15 minutes. To simplify the notation, we identify slots by their integer index value. For example, slot 1 corresponds to the 8:00–8:15 AM time interval, slot 2 is 8:15–8:30 AM, and slot 32 is the 3:45–4:00 PM interval.

| Parameter | Description |
|---|---|
| $T$ | Set of time slots in a day ($T = \{1, \ldots, 32\}$). |
| $D$ | Duration of time slot ($D = 15$ minutes). |
| $S$ | Set of user types ($\|S\| = \{2, 3\}$). |
| $I$ | Set of counters at a CCD ($5 \leq \|I\| \leq 20$). |
| $I_s$ | Subset of counters for user type $s \in S$. |
| $N_{st}$ | No. of users arriving in slot $t \in T$ for service $s \in S$. |
| $T_s$ | Average service time for type $s \in S$. |
| $A_s$ | Target average waiting time for type $s \in S$. |
| *Beta* | Parameter for adaptive staffing tuning. |
| $Q_s(t_1, t_2)$ | Work demand of type $s \in S$ (in minutes) in slot interval $\{t_1, \ldots, t_2\}$ with $t_1, t_2 \in T$. |
| $B_t$ | No. of available staff members in slot $t \in T$. |
| $K^O$ | Minimum no. of slots a counter must stay open. |
| $K^C$ | Minimum no. of slots a counter must stay closed. |
| $M$ | Penalty for each minute of unserved demand. |
| $\epsilon$ | Penalty for each counter opening (in minutes). |

| Decision variable | Description |
|---|---|
| $g_t$ | No. of unassigned staff in slot $t \in T$. |
| $y_{it}$ | 1 if counter $i \in I$ starts an open period in slot $t \in T$, 0 otherwise |
| $x_{it}$ | 1 if counter $i \in I$ is open in slot $t \in T$, 0 otherwise. |
| $d_{st}$ | Unserved work demand (in minutes) of type $s \in S$ in time slot $t \in T$. |

**Table A.1: The table shows the main parameters and decision variables used in the model. Typical values for large CCDs are reported in parentheses.**

The work demand $Q_s(t_1, t_2)$ for user type $s$ in a subperiod $[t_1, t_2] \subseteq T$ is given by

$$Q_s(t_1, t_2) = \sum_{t=t_1}^{t_2} N_{st} T_s f_s(t_2 - t + 1), \tag{A1}$$

where $f_s(\Delta)$ is the weight associated with a subset of $\Delta$ consecutive slots. We define it as the fraction of service time of the users who arrive during the first slot and who must be served within the set of slots in the subset. As Figure 11 illustrates, the weight increases with the number of slots $\Delta$ and decreases with the value of parameter *beta* so that the staffing is sufficient (i.e., has enough working minutes) to service all customers arriving in a time slot within a few subsequent slots. More precisely, the weight is defined also taking into account the target waiting time $A_s$, the actual service time $T_s$, and the slot duration $D$ according to the following formula:

$$f_s(\Delta) = w_s\left(1 - \frac{beta \cdot A_s + T_s}{2\Delta \cdot D + T_s}\right), \tag{A2}$$

where the numerator of the fraction takes into account the total time the user remains in the CCD. The denominator is instead an enlarged duration of the $\Delta$ slots because user arrivals are distributed during a slot and are not all concentrated at its starting time of the time slot. In addition, $w_s$ is a distortion coefficient, also dependent on *beta*; its role is to increase the overstaffing effect associated with smaller values of the parameter. Note that such coefficients drastically increase when the target waiting time is small with respect to the slot duration.

$$w_s = \begin{cases} -\ln\left(\dfrac{2beta - 1}{4}\dfrac{D}{D + A_s} + \dfrac{1}{e}\right), & \text{if } beta < \dfrac{1}{2}; \\ 1, & \text{if } beta \geq \dfrac{1}{2}. \end{cases} \tag{A3}$$

*Objective Function*: The hierarchic objective function (A4) calls for the maximization of the number of unused personnel who may hence be used to perform back office duties. The second term imposes the minimization of the unserved demand, where $M$ is a very large weight. The last term is used to minimize, through the penalty $\epsilon$ expressed in minutes, the total number of counter openings to favor more compact schedules.

$$\max\left\{D\sum_{t\in T} g_t - M\sum_{s\in S,\, t\in T} d_{st} - \epsilon \sum_{i\in I,\, t\in T} y_{it}\right\}. \tag{A4}$$

*Constraints*: The model is subject to the constraints listed as follows:

$$x_{it} - y_{it} - [x_{i(t-1)}]_{t>1} \leq 0 \quad \forall\, i \in I,\ t \in T, \tag{A5}$$

$$g_t + \sum_{i\in I} x_{it} = B_t \quad \forall\, t \in T, \tag{A6}$$

RIGHTSLINK

$$x_{it} - \sum_{p=\max(1,\, t-K^O+1)}^{t} y_{ip} \geq 0 \quad \forall\, i \in I,\, t \in T,\, t > 1, \tag{A7}$$

$$\sum_{p=t-K^C}^{t-1} x_{ip} + K^C y_{it} \leq K^C \quad \forall\, i \in I,\, t \in T,\, t > K^C, \tag{A8}$$

$$\sum_{i\in I_s}\left(D\sum_{t=t_1}^{t_2} x_{it}\right) + (d_{st_2} - [d_{s(t_1-1)}]_{t_1>1}) \geq Q_s(t_1, t_2)$$
$$\forall\, t_1, t_2 \in T,\, t_1 \leq t_2,\, |t_2 - t_1| \leq 4,\, s \in S, \tag{A9}$$

$$x_{it}, y_{it} \in \{0, 1\} \quad \forall\, i \in I,\, t \in T, \tag{A10}$$

$$g_t \geq 0 \quad \forall\, t \in T, \tag{A11}$$

$$d_{st} \geq 0 \quad \forall\, s \in S,\, t \in T. \tag{A12}$$

For a given time slot, constraints (A5) relate to $x$ and $y$ of each counter, whereas constraints (A6) impose that, at most, the available number of employees is used. Constraints (A7) stipulate that a counter must stay open for at least $K^O$ slots following its opening. Similarly, constraints (A8) impose that when a counter is closed, it must remain closed for at least $K^C$ slots. The adaptive staffing constraints (A9) require that for each set of up to four consecutive slots, the number of open counters should meet the corresponding work demand. Note that if the available staff $B_t$ is insufficient, a portion of the demand may be left unserved; this is penalized by the objective function. Finally, constraints (A10)–(A12) define the type of the decision variables.

The previous model can easily be modified to incorporate additional constraints, such as those requiring a minimum number of open counters in certain time slots.

## Appendix B. The SPRINT Software System

The logical architecture of the SPRINT system has three layers. The first layer is responsible for input activities. This includes configuration data and parameters of all CCDs (e.g., number and type of counters, opening hours, meal-break rules, number of employees, type of contracts), import procedures for historical data on user arrivals and number of bills, input of exogenous events that may affect forecasting (e.g., strikes, bad weather conditions). Desk managers can also input short-term data as immediate variations of the forecast for arrivals because of exogenous events, including the unavailability of specific personnel because of illness or service duties.

The second level includes the system's core modules: the forecasting module, optimizer, and simulator. Central planners have full access to these modules and can run forecasts and medium- and long-term optimizations and provide them to the desk managers who can only modify the short-term parameters and run single-day reoptimizations.

The third layer addresses output functionalities that include reports on the results provided by the various modules and specific monitoring of the system's KPIs (e.g., the SLAs and SPRINT usage by desk managers). SPRINT effectively integrates with the planning processes, particularly those that address medium- and short-term planning, which require that the system provide quick responses and be easy to use to enable users to react to unforeseen situations.

Each month for each CCD, the system automatically produces a forecast for the next three months and determines a proposal for staff scheduling and rostering for the entire month, with a granularity of 15-minute time slots. The central planner may introduce modifications because of considerations, such as the knowledge of a future commercial action that may require increased staffing in certain areas, and publish them to the desk managers.

The desk manager receives both the forecast and proposed plans so that he (she) can determine the necessary resources. Day-by-day, the desk manager can access the front end for short-term planning (see Figure 13), where daily arrivals, scheduling, and available resources are visible, together with all related information and parameters. By using this front end, the current conditions (e.g., the arrival rate or the available resources) are easy to modify and the user can very quickly obtain a reoptimized solution.

## References

Atlason J, Epelman MA, Henderson SG (2008) Optimizing call center staffing using simulation and analytic center cutting-plane methods. *Management Sci.* 54(2):295–309.

Bard J, Binici C, de Silva AH (2003) Staff scheduling at the United States Postal Service. *Comput. Oper. Res.* 30(5):745–771.

Bhattacharya B, Solomatine DP (2005) Neural networks and M5 model trees in modelling water level-discharge relationship. *Neurocomputing* 63(January):381–396.

Blochliger I (2004) Modeling staff scheduling problems: A tutorial. *Eur. J. Oper. Res.* 158(3):533–543.

Caprara A, Fischetti M, Toth P, Vigo D, Guida PL (1997) Algorithms for railway crew management. *Math. Programming* 79(1–3): 125–141.

Ernst AT, Jiang H, Krishnamoorthy M, Sier D (2004) Staff scheduling and rostering: A review of applications, methods and models. *Eur. J. Oper. Res.* 153(1):3–27.

Green LV, Kolesar PJ, Soares J (2001) Improving the SIPP approach for staffing service systems that have cyclic demands. *Oper. Res.* 49(4):549–564.

Green LV, Kolesar PJ, Soares J (2003) An improved heuristic for staffing telephone call centers with limited operating hours. *Production Oper. Management* 12(1):46–61.

Gruppo Hera (2012) Sustainability report. Gruppo Hera, Bologna, Italy.

Gruppo Hera (2013) Benchmark 2013. Accessed June 4, 2013, http://www.gruppohera.it/gruppo/investor_relations/peers/.

Ingolfsson A, Campello F, Wub X, Cabral E (2010) Combining integer programming and the randomization nd nd method to schedule employees. *Eur. J. Oper. Res.* 202(1):153–163.

Kabak O, Ulengin F, Aktas E, Onsel S, Topcu YI (2008) Efficient shift scheduling in the retail sector through two stage optimization. *Eur. J. Oper. Res.* 184(1):76–90.

Quinlan RJ (1992) Learning with continuous classes. Adams A, Sterling L, eds. *Proc. Fifth Australian Joint Conf. Artificial Intelligence* (World Scientific, Singapore), 343–348.

Downloaded from informs.org by [70.83.24.169] on 27 October 2014, at 18:44 . For personal use only, all rights reserved.

Downloaded from informs.org by [70.83.24.169] on 27 October 2014, at 18:44 . For personal use only, all rights reserved.

Taylor JW (2008) A comparison of univariate time series methods for forecasting intraday arrivals at a call center. *Management Sci.* 54(2):253–265.

van Dijk NM, van der Sluis E (2006) Check-in computation and optimization by simulation and IP in combination. *Eur. J. Oper. Res.* 171(3):1152–1168.

Witten IH, Frank E, Hall MA (2005) *Data Mining*, 3rd ed. (Morgan Kaufmann Publishers, Burlington, MA).

**Daniele Vigo** is professor of operations research at the Department of Electrical, Electronic and Information Engineering of the University of Bologna. His research interests are in the design and implementation of innovative algorithms for hard combinatorial problems arising is several application fields, such as logistics, energy production, and personnel management. He supervised numerous applied projects in collaboration with important national and international companies and research centers. In 2007 he founded Optit srl, an accredited spinoff company of the University of Bologna specialized in the development of decision support systems based on state-of-the-art operations research methods.

**Claudio Caremi** is chief operations officer at Optit srl. He has more than 10 years of business and management consulting experience in project management, change management, and operations improvement in a broad range of industries (utilities, retail, services).

**Angelo Gordini** is a senior java developer with 10 years of IT experience on Web development and system integration, and five years of experience at Optit in decision support systems development, team management, data analysis, and MILP modeling for several applications, such as front desk services, energy production and distribution, and waste logistics.

**Sandro Bosso** is director of the Consumer Market Division at Hera Comm and chief executive officer at AcegasAps-Service. He has been responsible for new services development and innovation for consumer market at Telecom Italia headquarter (Rome). He was also involved in various international projects in partnership with Telecom Italia Research Center (CSELT).

**Giuseppe D'Aleo**, after an experience at ENI group in the field of computer training, was for several years the head of a local front-end infrastructure in a utility company. In 2006 he started working at Hera Comm, where he was involved in F.A.S.T. project that laid the foundation of SPRINT. He is currently manager of the Complaint Department for Consumer Market Division and continues working on special projects for the optimization of the customer management activities, particularly related to Six-Sigma lean organization.

**Beatrice Beleggia** is responsible for planning and reporting of Consumer Market Division at Hera Comm. Her main focus is planning and monitoring costs and performance of the resources involved in the management of consumer market (front and back office). She has previously worked as a consultant in the utilities sector mainly dealing with restructuring processes and development of new organizational models.

RIGHTSLINK