Discrete Event Simulation

Differentiated Repair Cost Analysis

October 2011

Contents

1 EXECUTIVE SUMMARY 3
2 BACKGROUND TO DISCRETE EVENT SIMULATION 4
3 APPROACH TO DISCRETE EVENT SIMULATION 5
3.1 Queuing Theory 5
3.2 Monte Carlo Simulation 5
3.3 Discrete Event Simulation 5
4 MODELLING DETAIL 7
4.1 Workforce Simulator 7
4.1.1 Inputs 8
4.1.2 Outputs 9
4.1.3 Model 9
4.2 The WDS Engine 9
4.2.1 Multi-agents model 10
4.2.2 Task Time Distribution in WDS 10
4.3 Chosen Geography 11
4.4 WDS Calibration & Accuracy 11
5 MODELLING RESULTS & MITIGATION 13
6 USAGE FACTORS 14

1 Executive Summary

This paper outlines the rationale behind Openreach's development of a bespoke Discrete Event Simulation (DES) model named the Workforce Dynamic Simulator (WDS); the principles of its operation and the detail of how scenarios are created within that model. It then describes how the model has been used to assess the cost differential between scenarios where all repair jobs are in line with the Wholesale Line Rental (WLR) Service Level Agreement (SLA), compared to where jobs are in line with the Local Loop Unbundling (LLU) SLA.

DES models are widely used and across industry sectors. This DES model is a development between Openreach's modelling specialists, the BT Innovate & Design research laboratories and Essex University. After three years of joint work, a tried and tested working model known as WDS was made available in 2011. Accuracy of the simulation to the real world has been tested and measured to be better than 98.5%. WDS simulates the allocation and scheduling of work over a three month period in a specific geography, using a rule-based allocation system that takes into account the specific attributes of individual jobs and Openreach engineers, including their locations and skills. That is, the WDS is used to model the fulfillment of jobs where it ensures that it gets the right person, with the right skills, to the right place and at the right time. Various "what-if" scenarios can be tested using these jobs and engineers and WDS provides quantitative measures on the execution of the jobs over the given time period. This model has been used by Openreach to explore the implications of a wide range of business scenarios.

In this specific instance, the WDS model has been used by Openreach to assess the potential additional resource commitments required to support all repair jobs in the timeframes for the LLU based repair SLA as opposed to WLR. The difference in repair requirements arises because the SLA for LLU guarantees a repair one day faster than WLR. Here the WDS is used to evaluate the hypothesis that LLU requires additional resource and quantifies the amount required. The WDS runs simulations based on real fault arrival times and locations, against real engineering skills, locations and availability. In this simulation, real jobs have been configured to attract either a WLR or an LLU service level.

It can be seen in section 5 that the service delivered with Openreach's existing engineering capability where repair demand is LLU is 9.5% adverse to that delivered in a WLR environment. This is driven by the fact that LLU (end of next working day, Monday to Saturday) is guaranteed to be completed one day earlier than WLR (end of next working day +1, Monday to Friday).
To mitigate this drop in performance section 5 runs additional scenarios where further resource is injected to restore LLU performance back to that seen in WLR. This additional resource equates to [✂]% or [✂] FTE.

Using the cost model for the LLU and WLR charge controls, this additional [✂]% resource requirement equates to an additional [✂] man hours [✂]. This equates to a 20% increase on the [✂] man hours required for repair jobs if they were only on the WLR service level.

Consequently, those products that are on the LLU service level should have an additional service level usage factor of 1.2 applied in the cost model to ensure that an appropriate share of repair costs are allocated to those products.

2 Background to Discrete Event Simulation

As the complexity of Openreach's business grows over time and with the increasing number of service levels, products and service level packages offered to customers, the need for a simulation model on which to perform a wide range of scenarios becomes increasingly important to support informed decision making for the business.

To ensure that we have the right tools and modelling capability, Openreach undertook a major piece of development to build a model that would test a number of known hypotheses and that could be used to scenario model other unknown factors. This development has been a major piece of work between Openreach's internal modelling specialists, our colleagues in the BT Innovate & Design research laboratories and Essex University. Following three years of collaboration, a tried and tested working model became available in 2011. This model can be, and now has been used to explore the implications of a wide range of business scenarios.

3 Approach to Discrete Event Simulation

To deliver this model there were three potential approaches that could have been used. These are Queuing Theory, Monte Carlo Simulation, and DES. Openreach considers that DES is the most appropriate modelling technique to use to simulate the real world of its business, as it can take into account actual geographic data including job locations, engineering skill levels and rostering, to mimic the complexities of the environment in which it operates.

3.1 Queuing Theory

Queuing theory is the mathematical study of waiting lines, or queues. The theory enables mathematical analysis of several related processes, including arriving at the (back of the) queue, waiting in the queue (essentially a storage process), and being served at the front of the queue. The theory permits the derivation and calculation of several performance measures including the average waiting time in the queue or the system, the expected number waiting or receiving service, and the probability of encountering the system in certain states, such as empty, full, having an available server or having to wait a certain time to be served.

Reasons for not using this methodology

Queuing theory is simplistic in its assumptions as it is a form of theoretical mathematics, e.g. task time distributions are very constrained to forms that may not reflect operational reality (such as Gaussian[1] distributions when Gamma[2] may be better). Additionally Queuing Theory cannot use real data that varies by individual engineers, e.g. actual job locations, actual engineer start locations, actual Preferred Working Areas (PWAs), skills, and rosters. Finally it cannot mimic the way that a work delivery process performs Real Time Priority job allocation as that is too specific (albeit crucial to this topic that centers on Service Levels) and not based on a mathematical equation.

3.2 Monte Carlo Simulation

Monte Carlo methods are a class of computational algorithms that rely on repeated random sampling to compute their results. Monte Carlo methods are often used in simulating physical and mathematical systems. These methods are most suited to calculation by a computer and tend to be used when it is not feasible to compute an exact result with a deterministic algorithm. They are used to model phenomena with significant uncertainty in inputs, such as the calculation of risk in business.

Reasons for not using this methodology

Monte Carlo is essentially a stripped-down, basic version of DES that makes more assumptions and simplifications. Typically Monte Carlo would be too simplified to use actual geographic data, e.g. actual job locations, actual engineer start locations, and actual PWAs. Similarly it would ignore variations by individual engineers, e.g. skills and rosters. Finally it would not have the capability to mimic the complexities of the way that a work allocation system performs Real Time Priority job allocation.

3.3 Discrete Event Simulation

In DES, the operation of a system is represented as a chronological sequence of events. Each event occurs at a moment in time and each event marks a change of state in the system. The simulation maintains at least one list of simulation events. This is sometimes called the 'pending event set' because it lists events that are pending as a result of previously simulated events but have yet to be simulated themselves. An event is described by the time at which it occurs and a type (for example, appointed job), indicating the code that will be used to simulate that event.

---

1 The graph of a Gaussian distribution is a characteristic symmetric "bell curve" shape that quickly falls off towards plus/minus infinity.

2 The gamma distribution is a two-parameter family of continuous probability distributions. It is non linear and has scale and shape parameters.

When events are instantaneous, activities that extend over time are modelled as sequences of events. Some simulation frameworks allow the time of an event to be specified as an interval, giving the start time and the end time of each event. The best simulation framework has a multi-agent simulation engine supporting an interval-based event model.

Reasons for using this methodology

DES is capable of realistically modelling a process comprised of a complex chain of events with multiple queues, distinctive transition probabilities, geographical characteristics and density functions. DES is used in many industries including manufacturing, healthcare, public sector, supply chain & logistics and energy as it is a recognised and accepted modelling application. The National Research Council has recognised the benefits of DES to model real world complexities, and has stated "[w]hen a company begins to confront increasing complexity, whether in products, production processes, or both, they should consider the benefits of discrete event simulation". [3]

Openreach evaluated a number of "off the shelf" DES packages however they were all found to be lacking the necessary features and/or functionality. Openreach therefore decided to commission its own bespoke Discrete Event Simulator.

[3] National Research Council, *Information Technology for Manufacturing*, Washington, D.C., National Academy Press, 1995. The National Research Council is a public organisation. Its duty is to perform, promote, spread, transfer and improve research activities.

4 Modelling Detail

4.1 Workforce Simulator

In a service industry the scheduling of work consists of allocating the geographically distributed service parameters. In the telecoms industry these parameters are the individual tasks generated by CPs and End Users (jobs) and the resources (engineers) to fulfill the service obligations on each and every one of those jobs. In the simplest form this is about getting the right person, with the right skills, to the right place and at the right time.

The complexity of this holistic view grows exponentially as the individual characteristics of the jobs and the people are taken into account. Each of these two components have both “static” and “dynamic” characteristics. Static factors include Service Levels, importance scores,[4] product, network components and appointment times. When each of these static characteristics is taken into account, there are in excess of 20,000 permutations. Once account is taken of the dynamic components, e.g. travel speed, time remaining to commitment times, real time priority, job arrival times etc, the number of permutations is increased by a number of magnitudes.

To reduce the complexity of this scheduling scenario, it is decomposed into sub-components - each one for a geographical area. The UK is broken down hierarchically into logical areas called domains. These domains are then further decomposed into PWAs[5] in which the individual exchanges are located. It is in these domains and PWAs that jobs are located in, and to which engineers are assigned each day. These working areas determine the geographical zone and types of jobs (e.g. Field, Frames, Provision and Repair) the engineers can realistically cover.

The way to handle this real-time dynamic workforce scheduling scenario is through a pre-emptive scheduling system which computes a work and route plan for each job and each engineer, and a work allocation system, which assigns jobs to engineers following the schedule. If an event were to occur that causes disturbance (e.g. a very high priority job arrives), and this makes the schedule no longer “executable”, the simulator would be required to assess the impact of these disturbances and re-schedule the new profile of work.

Our WDS is a simulation environment that enables simulation of these static and dynamic characteristics. WDS computes an estimated schedule, and then runs the travel and execution of jobs by engineers over a period of 3 months (91 days), using a rule-based allocation system.

WDS also provides quantitative measures on the execution of the jobs over the given time period and the chosen geographical domain. WDS is decomposed in to three main parts, the inputs, the model, and the outputs as shown below in Figure 1.

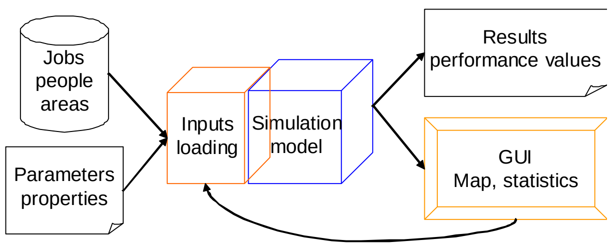

Figure 1 Inputs and outputs of the system

---

[4] Each job has a unique importance score that is applicable to its Service Level and product.

[5] A PWA is a selection of exchange boundaries in which an engineer will be allocated jobs. Typically a PWA will have a diameter of approximately 15km.

4.1.1 Inputs

WDS includes an inputs loader where files in CSV format can be loaded. The content of these data files determines the tasks (jobs), the resource (people) and the PWAs that the engineers will work in. Examples showing the characteristics of all these attributes are shown below.

Task Files

Domain	Leeds & West Yorkshire
Job Number	ABC123
Skill Code	LLU Frames provision
Task Type	FLTP04
Task Category	Provision
Easting's	61553
Northing's	24519
Post code	LS1 1BA
Zone Code	North East
Created Date	14/08/2011 08:52:00
Appointment Start	17/08/2011 13:00:00
Appointment End	17/08/2011 17:00:00
Attend by time	17/08/2011 15:30:00
Commitment time	17/08/2011 23:59:59
Estimated duration	01:30
Estimated travel	00:17
Importance score	200

> In this example the job is appointed for the 17th August and the engineer must attend no later than 15:30 on that day.

People Files

Absence Start Time	17/08/2011 10:00:00
Absence End Time	17/08/2011 12:00:00
Absence Type	Training
Lunch Duration	00:40
Lunch Start Time	17/08/2011 12:00:00
Sign on Time	17/08/2011 07:00:00
Sign off Time	17/08/2011 18:30:00
PIN	MY02CRE
Start Easting's	61550
Start Northing's	24530
Finish Location	Seacroft

> In this example this particular engineer is only available until 10:00 and after 12:40

PWA and Exchange Files

In addition to Task and People data WDS holds all the engineers skills and preferred work areas. A pictorial example of tasks, people and exchanges is shown at a particular point in time as the schedule changes. Jobs are shown as black circles; engineers are shown as red squares.

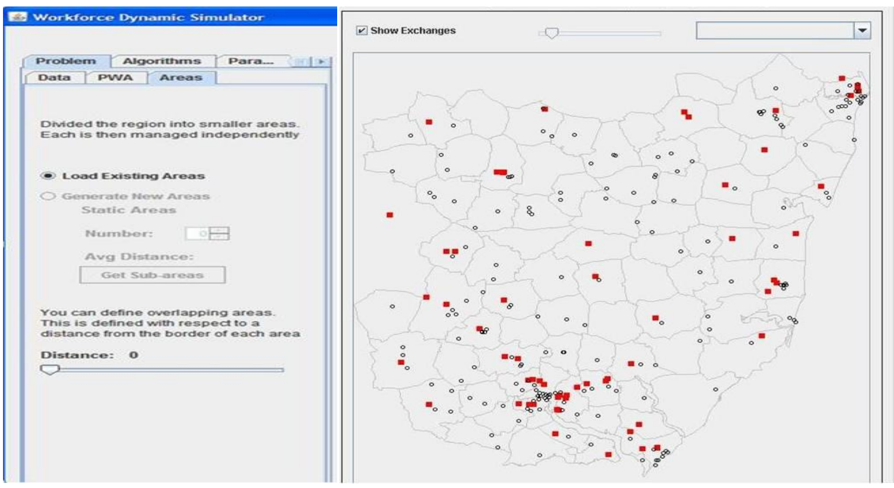

Figure 2 Jobs and engineer locations

It can be seen from Figure 2 that the model has to simulate the best scheduling algorithm to ensure that all the jobs in this geography are covered within the service level targets by the available resource.

4.1.2 Outputs

The simulation results are available for the volume of successful jobs, failed jobs and completed jobs per day and are expressed as the service level percentage success. A Graphical User Interface contains the map of the investigated geographical domain on which the user can follow the journey of people and change of jobs status as the simulation progresses. In a typical geographical simulation over the 91 day period the model will output the result for approximately [✂] jobs.

4.1.3 Model

This includes the multi-agent systems, the implementation of the main concept, the simulation mechanism, data loading and the simulation results reporting.

4.2 The WDS Engine

The WDS engine simulates the process of work scheduling, task allocation to people and the processing of tasks by field force over a working period. The engine simulates each working day, using the set of tasks available in the work demand (at the beginning of the day and incoming during the day). The engine continues the simulation until all the tasks are dealt with, which may take several (simulated) days. Performance measures are computed for each day. **Figure 3** outlines a typical scenario simulated in WDS.

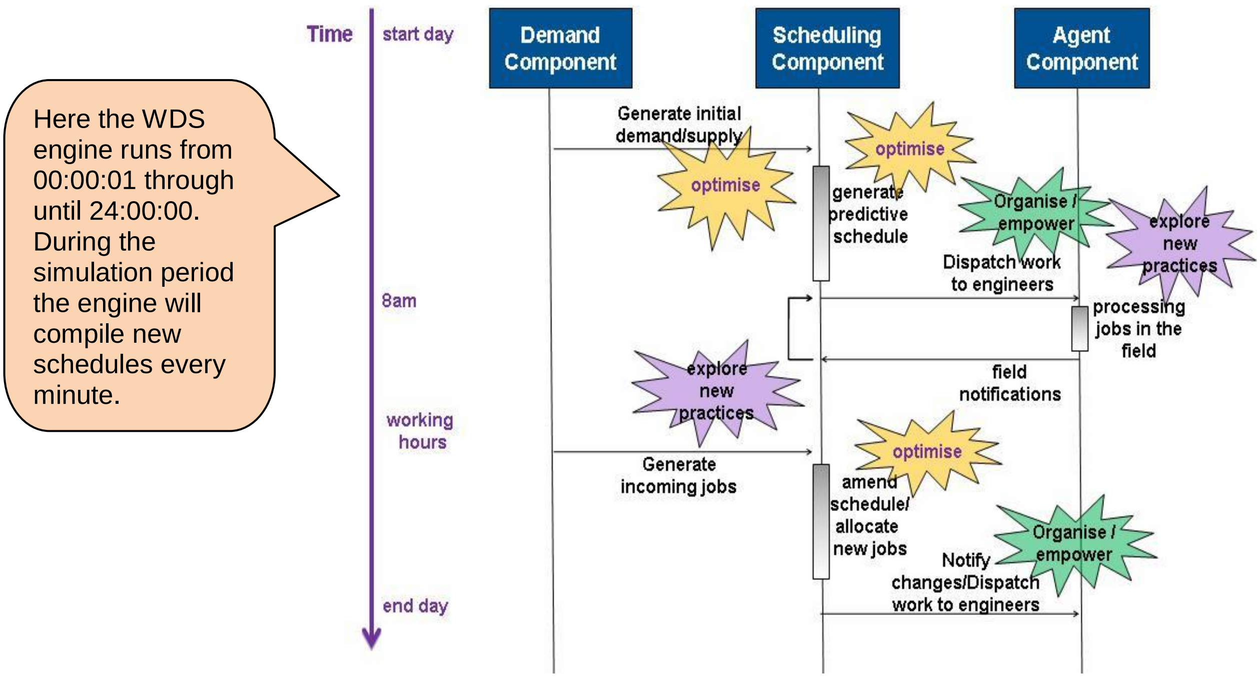

Figure 3 Sequence diagram of a typical scenario simulated in WDS

The Scheduling component creates a schedule that is passed to the work allocation dynamic agent, and also creates the next iterative dynamic schedule which is then passed to the allocation agent while the jobs are being performed during the working day. It provides the tasks sequence to the dispatcher. **Figure 4** below provides a diagram of the overall simulation environment.

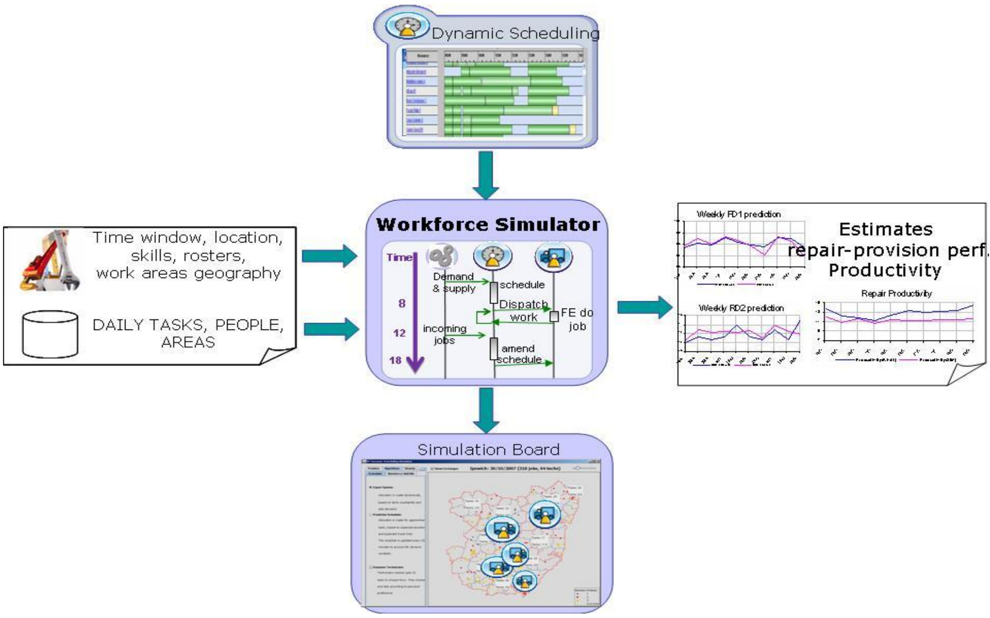

Figure 4 Overall Simulation environment

4.2.1 *Multi-agents model*

The interaction between field engineers and job management systems is modelled as a negotiation between three types of agents: Job Generation, Scheduler and Technician. The simulation consists of generating jobs, reporting them to the scheduler agent which is then responsible to allocate them to available technicians.

The Job Generation Agent generates jobs based either on historical, artificial or mixed (historical and artificial) data. A job can be either reported at the beginning of the day or, in order to simulate incoming jobs, or at any time during the day. This agent is responsible for feeding jobs into the simulation at the time corresponding to the time the job was reported. This agent simulates the entry of jobs at their real report time so that it deliberately impacts on the way that work is scheduled to engineers.

The Scheduler Agent receives incoming jobs from the job generation agent. It then allocates them to available technicians. The agent is linked to an interface of a scheduling algorithm which is notified of any event which might affect the scheduling (i.e., new jobs, availability of technicians, delays etc.).

The Technician Agent models the behavior of the technician, i.e. receiving a job from the scheduler agent, travelling to the job location and performing the job, then waiting for the next job. The technician agent reports the completion of each of these steps to the scheduler agent (e.g., arrival at job location, job starts, job ends). The technician agent is inactive at the appropriate time to simulate a lunch break or absence leave.

The model re-evaluates the actual remaining time of travel and work constantly in real time. It uses distribution functions to generate a deviation on estimated job duration and travel time. This models dynamic events occurring during the working day and night and simulates the uncertainty of the real world.

4.2.2 *Task Time Distribution in WDS*

The task estimated duration that is used is the estimated duration specified in the task data files. During the simulation run, the actual task duration, that is the simulated time actually spent by the technician to perform the task, is evaluated as estimated duration plus a deviation computed from a statistical distribution. This distribution is applied dynamically while a job is performed by a technician. The distribution used does not change according to task properties but returns a new value each time it is invoked (randomly) so it depends on the time of the simulation when the task is considered to be started or assigned to a technician. Intuitively, the empirical distribution should mimic closely events that occur in the real world.

4.3 Chosen Geography

As DES is the chosen method of effectively simulating reality as closely as possible, careful consideration has to be made on what to simulate, where to simulate and over what period of time the simulation should run. The following explains the rationale for adopting our chosen route.

In a telecoms environment we need to be able to conduct scenarios over a period of time that not only reflects how our business responds to natural variations over time, but also takes into account the behaviors of our CPs and their End Users. Typically the provision demands placed on our business by day of week, time of day and also taking into account appointments where the End User has chosen a date well into the future. From an operational perspective we also need to model the agreed engineering rostered attendance patterns. Taking all these into account the most appropriate period of time to conduct a DES over is 13 weeks or specifically 91 days.

Simulating all the visit activities at a UK level over that period of time is totally unmanageable as the model would have to run the scheduling engine for [✂] jobs over a total area of 250,000 square kilometers. Taking into account the individual job, people and exchange attributes this task becomes impossible in terms of processing overheads.[6] Our taskforce work allocation system that is in use across the UK splits into domains, of which there are 73 in total which vary according to a number of factors, typically urban rural mix and population density. The way that WDS models scenarios is based on the current structure in which we select one domain.

To achieve this the UK mainland was evaluated at a domain level against the key elements that WDS uses in its simulations and careful selection of the most suitable Operations Area was made. The Leeds & West Yorkshire domain was chosen as being representative. The key attribute that was taken into account when making this decision was engineering travel time.

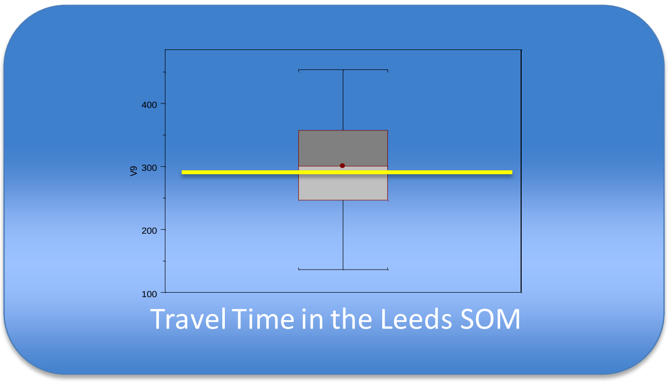

Figure 5 Leeds domain is representative of the national average for engineering travelling time

Out of the 73 UK geographies selected Leeds & West Yorkshire is midpoint at position 36 and is only 2% lower than the UK average.

4.4 WDS Calibration & Accuracy

With the range of parameters that the model uses, and the attributes of the jobs and the people, the model requires calibration. Calibration is essentially a measure of how well the model

6 To ensure the accuracy of WDS simulations it is essential that the scheduling engine runs in real time – in our design - every minute. This can only be achieved at a domain level.

simulates reality. There are a number of stages that have taken place as part of the model development and testing to ensure that the model performs to the highest possible accuracy in terms of the service levels that are output.

The model has a number of user definable parameters that can be used to maximize that accuracy. Shown here is the accuracy as measured in our base scenario. The simulated result shown here is what the model simulates against a set of base (actual) task and people files over the 91 day period. The actual overall result is what has been experienced in reality by the end users, with the same jobs and people, over the same 91 day period, and delivered via our taskforce work allocation systems.

This shows that there is a very small modeled variance of +1.1% between what actually happens and what has been modelled. The service measure used is not purely a measure of whether or not the jobs were completed, but were they completed on the right day, in the right timescales and did the engineer arrive within the appointment slot? The results shown below are a measure that takes all these factors into account and is expressed as the percentage of jobs that were successful against the overall number of jobs in the time period.

Sim Overall	Act Overall	Results Var
98.6%	97.5%	1.1%

Daily weekday performance is achieved as follows between the actual and the simulated results -

- Mean daily variance Simulated to actual =0.92%
- Max daily variance Simulated to actual =3.54%

To achieve this level of accuracy the model produces in excess of 130,000 scheduling iterations. For the Leeds & West Yorkshire domain [✂] jobs over 91 days with 5 simulation runs (to increase robustness) requires over 4 billion computations to support the number of scheduling iterations.

5 Modelling Results & Mitigation

WDS was used in this particular case to run two specific scenarios regarding Service Levels for WLR and LLU within the Leeds & West Yorkshire domain using real engineer and job data. For scenario 1 the commit times on all Repair jobs were modified so that they were in line with WLR SLA, i.e. “End Of Next Working Day +1”.The model was then run for 5 simulations over the 91 day period with [✂] jobs and resulted in an overall service level of 98.3%, as measured using BT’s Right First Time criteria. For scenario 2 the same jobs were then modified so that they were in line with LLU SLA, i.e. “End Of Next Working Day”. The model was again run for 5 simulations for the same 91 day period and the service level fell to 88.8%

Modelled Scenarios		
		Service Level
Scenario 1	WLR Base	98.3%
Scenario 2	LLU Base	88.8%

To restore the service levels in an LLU environment back to those achieved in the WLR scenario it is necessary to incrementally add more resource. This additional resource has been chosen with random locations and roster patterns that effectively mimic those that would occur in reality. The skills of these people are also complementary to those occurring in reality. The graph below shows how the service level recovers as the additional resource is injected.

It can be seen that the relationship between service and additional resource is non-linear so a number of scenarios had to be completed so that an accurate view of the resource requirements could be established.

An additional [✂] of resource was required to maintain the service levels seen in scenario 1 with the tighter constraints of scenario 2. This equates to an additional [✂] FTE

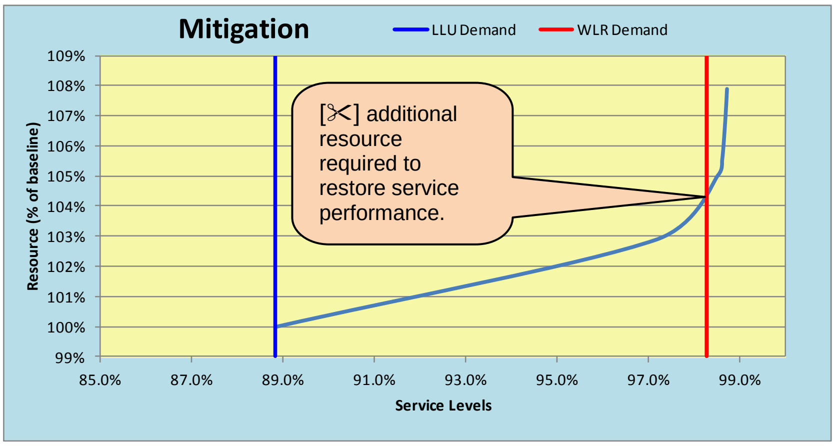

Figure 6 Additional resource required for different service levels

6 Usage Factors

In its consultation document for the forthcoming LLU and WLR charge controls, Ofcom proposes to allocate the same repair costs to LLU and WLR even though LLU is guaranteed to be repaired a day faster than WLR.[7] The WDS modeling undertaken above demonstrates that LLU repair requires a higher resource commitment and so its costs are higher than for WLR.

To reflect the additional costs for LLU repair in the cost modelling, Openreach has used the output of the WDS model to estimate the additional costs associated with repairing jobs to the LLU SLA level.

The WDS showed that an additional [✂] of resource was required to maintain the same service levels observed if all repair jobs were fulfilled to the LLU SLA rather than the WLR SLA. The additional [✂] resource requirement approximately equates to an additional [✂] man hours [✂]. From the LLU and WLR charge control cost model, the additional man hours represents a 20% increase on the [✂] man hours required for all repair jobs if they were only on the standard WLR service level.

Therefore those products that are on the LLU service level should have an additional service level usage factor of 1.2 applied to ensure that an appropriate share of repair costs are allocated to those products. To achieve this, the service level usage factor should be multiplied with the Fault Rate usage factors proposed in Ofcom's consultation for the LLU and WLR charge controls to provide a combined usage factor to allocate repair costs in the cost model. This will ensure that repair costs reflect the level of resource required given the SLA associated with the product.

The table below shows the resultant combined usage factor and Openreach's estimate of the impact by product on a unit basis by applying the combined usage factor instead of only the Fault rate usage factor proposed by Ofcom in its consultation. That is, an increase in repair costs for the MPF and SMPF Rental products is shown given that faster repair time, and an offsetting reduction in repair costs for WLR Basic Rental.[8] The overall repair costs remain the same – it is only the allocations which change as a consequence of the additional service level factor.

Product	**Service Level**	**Fault incidence**		**Service level factor**		**Usage factor**	**£ per line impact estimate**
WLR Basic Rental	SL1	1.00	x	1.00	=	1.00	[✂]
WLR Premium Rental	SL2	1.00	x	1.20	=	1.20	[✂]
MPF Rental	SL2	1.04	x	1.20	=	1.25	[✂]
SMPF Rental	SL2	0.15	x	1.20	=	0.18	[✂]

Figure 7 Openreach estimate of impact of Service Level usage factor on the repair costs[9]

7 Ofcom, *Charge control review for LLU and WLR services*, Consultation, 31 March 2011.

8 Note that WLR Premium price is not impacted by proposed service level factor in the LLU and WLR charge control model is it is not subject to charge control or cost orientation obligations.

9 SL1 equates to 'WLR' in the modeling described above, while SL2 equates to 'LLU'.