Experiences of the Electricity System Operator Incentives Scheme in Great Britain

Guilherme Luiz Susteras, Graham Hathaway, Jeremy Caplin, Graham Taylor

For CIDEL Argentina 2010
International Congress on Electricity Distribution.

Abstract-- The Great Britain electricity market presents some peculiarities in that National Grid is one of the few (if not the only) for-profit large system operator. Such peculiarities lead the local regulator (Ofgem) to apply a set of incentive rules aimed at encouraging innovation and ingenuity in the tools utilized by National Grid to balance and operate the system. This paper presents an overview of the current market arrangements in Great Britain, an explanation of the incentivised cost components in the incentive scheme. Shortfalls in the current scheme format will be explored and suggestions to improve the effectiveness will be presented. Finally, the paper will touch upon the challenges facing future incentive schemes with the increasing penetration of renewable generation.

I. INTRODUCTION

National Grid Electricity Transmission owns and operates the high voltage electricity system in England and Wales and is the National Electricity Transmission System Operator in Great Britain (England, Wales and Scotland). It is important to note that National Grids business functions of ‘owning’ and ‘operating’ the transmission system are divided into two quite separate business operations and revenue streams. This paper concerns solely the second, the operation of efficient energy provision on a second by second basis.

The Great Britain electricity market presents some peculiarities in that National Grid is one of the few (if not the only) for-profit large system operator and, unlike most markets, National Grid is not responsible for fully dispatching the generators, but only for marginally balancing the system.

The greatest part of the task of balancing supply and demand of electricity which is carried by National Grid takes place through commercial relationships between generator and demand side parties, in a marketplace outside National Grid’s control. Of the 330 TWh of electricity transferred across the grid in 2009, just 4% require the balancing intervention of National Grid: 5 TWh of ‘offers’ to call on additional generation at times and 8 TWh of ‘bids’ to pull back generation at times. This balancing activity is the focus of this paper and represents circa £900m spend per year.

Such peculiarities lead the local regulator (Ofgem) to apply a set of incentive rules known as the Balancing Services Incentive Scheme (BSIS). These aim at encouraging innovation and ingenuity in the tools utilized by National Grid to balance and operate the system, above and beyond the requirements of its operating license, through the sharing of savings and overspends between the System Operator and the users [1].

A. Overview

Incentive schemes have been in place since 1994, but in its current shape BSIS has been in place since the introduction of the New Electricity Trading Arrangements (NETA) in 2001. Fig. 1 illustrates the evolution of the costs which National Grid has been incentivised since the introduction of the incentive scheme.

April 2005 saw the introduction of the British Electricity Trading and Transmission Arrangements (BETTA), which unified the electricity markets of Scotland with England and Wales. The costs to operate the system had then a significant change, mainly due to the need to manage the congested Anglo-Scottish interconnection.
Other changes to the market environment had impacts in the costs to operate the system, in particular the introduction of the Large Combustion Plant Directives (LCPD), an EU regulation aimed at reducing the emission of SOx and NOx, leading to restrictions in the operation of certain power stations; many of the affected units moving from ‘baseload’ to premium cherry-picking dispatch regimes.

Future changes are anticipated, especially considering the challenges to the country to meet its targets of energy produced form renewables sources. The increasing amount of wind generation and their characteristically intermittent and less predictable nature, allied with the retirement of flexible fossil fuelled generation, will bring increasing pressure to National Grid to ensure the maintenance of the current high levels of system security, whilst ensuring system operation costs are kept to their minimum.

II. CURRENT MARKET ARRANGEMENTS IN GREAT BRITAIN

In July 1998, the then Director General of Electricity Supply (DGES) proposed new arrangements for the wholesale trading of electricity in England and Wales.

In July 1999, the DGES published further proposals for implementing the new electricity trading arrangements in England and Wales. He also published a draft specification for the balancing mechanism and imbalance settlement.

In October 1999, the Office of Gas and Electricity Markets (Ofgem) and the Department of Trade and Industry (DTI) published a report confirming new arrangements for the wholesale trading of electricity. This report was produced after careful consideration of the comments received on the July 1999 proposals [2].

In March 2001 the UK government introduced the New Electricity Trading Arrangements (NETA). According to Ofgem, “the former flawed and much-criticised arrangements under the Electricity Pool meant that wholesale prices failed to reflect falling costs and increased competition. NETA created a market where electricity is traded like any other commodity through bi-lateral contracts, where prices are agreed between the two contracting parties, or on power exchanges”[3].

Later, in April 2005 the final significant market change was put in place with the unification of the Scottish market with that of England and Wales, meaning that suppliers and generators in the whole of Great Britain would now be able to operate under a single market structure.

These new arrangements meant that National Grid was no longer responsible for scheduling generation according to its demand forecast; instead, National Grid became responsible for balancing the system in real time, with the market (generators and suppliers) informing the system operator of their own schedules ahead of real time (called the gate closure). In practice, only generators inform their scheduled position, with National Grid producing its own demand forecast.

The information provided to National Grid by generators is called Physical Notification and it reflects the position each generating unit will take for each half-hour of the day. This information must be provided everyday at 11 o’clock in the morning for the next 24 hours and can be changed up to 1 hour ahead of real time (gate closure). Based on the information provided by generators and its own demand forecast, National Grid plans the operation of the system, considering its short term operating reserve requirements, primary and secondary frequency response requirements, potential congestions in the transmission system, etc.

The actual operation of the system is made by National Grid accepting bids (instructions for generators to reduce output) and offers (instructions for generators to increase output) in the Balancing Mechanism, performing over the counter (OTC) and exchange based trades and by using balancing services contracts (Fig 2).

![Fig. 2 Bid and Offer actions to balance market length](image)

III. MAIN INCENTIVISED COST COMPONENTS

The costs to which National Grid is exposed in its role as system operator and is incentivised to minimize under the BSIS can be divided into two main groups: energy related and constraints. The relative size of the components is illustrated at the end of this section (Fig. 5).

A. Energy related costs

Great Britain’s nominal operating frequency is 50Hz. In order to maintain the frequency within the statutory limits of 49.5 to 50.5, National Grid has to balance generation with demand on a second by second basis.

To ensure that the system isn’t operated outside of the statutory limits, National Grid sets a more restrictive operational frequency limit of 49.8 to 50.2Hz.

National Grid takes a number of different actions to ensure that the system frequency is maintained within the statutory and operational limits. The different actions are taken depending on the different timescales required for managing the system. Fig. 3 shows the four main actions taken to respond to system frequency fluctuations; margin and Short Term Operating Reserve (STOR), energy balancing, fast reserve and frequency response.
1) Operating Margin

The Operating Margin actions refer to the need to ensure there is enough synchronized generation to meet our Short Time Operating Reserve Requirement (STORR). This requirement is calculated so that the probability of demand not being met is only a total of one day in every 365 days. This operating margin is a balance between reducing the risk of demand disconnection and reducing the costs associated with operating margin actions.

Another component of Operating Margin is the downward regulation, which is provision of the capability of National Grid to reduce the amount of generation output on the system. In circumstances where demand is low and the majority of generation is operating inflexibly at, or near, its minimum stable output (i.e. the level at which it can not operate below), there may be insufficient available MW reduction capability (footroom) to allow the required level of negative reserve to be delivered. Actions have to be taken to exchange this inflexible generation with flexible generation – this is achieved by the de-synchronising of some of the units, allowing the output of other units to be increased above their minimum stable output.

2) Energy Balancing

Energy balance costs are those incurred by National Grid to correct for differences between the generation supplied by the market and the demand on the system. The following actions are taken to ensure that generation and demand are balanced:

- Buying and selling power in the Balancing Mechanism (BM) (otherwise known as accepting bids and offers)
- Pre-gate closure balancing transactions (PGBT)
- Trading outside the BM

The vast majority of energy related actions carried out by National Grid are in the form of bids and offers in the BM. For each settlement period, each generator submits prices at which National Grid can instruct them to vary their output during that period.

If the market is long (generation in excess of demand), National Grid instructs generation to decrease their output by accepting bid prices. Conversely, if the market is short (demand in excess of generation), National Grid instructs generating units to increase their output by accepting offer prices. For energy balancing purposes, these actions are performed in cost order and feed into the processes undertaken to calculate system imbalance prices.

PGBTs are a tool used by National Grid to instruct units earlier than one hour ahead of real time. Still ahead of real time, National Grid can also utilize trades through two different types of forward energy products: Power Exchange Trades and Over the Counter Trades.

Power Exchanges are electronic trade matching systems where participants enter prices at which they are prepared to buy or sell electricity. The exchange automatically matches like prices/volumes and the trade is done. This is an anonymous process with the Power Exchange effectively becoming the counterparty to the trade.

Over the Counter Trades are bilateral contracts negotiated between counterparties and can be tailored to suit the requirements of the individual parties.

3) Fast Reserve

Fast Reserve is used to control frequency changes that might arise from sudden changes in generation or demand, such as incidents involving generation disconnection or rapid demand changes resulting from TV pickups (TV pickup is a typical phenomenon in the British system where sudden increases in demand are observed during commercial breaks or at the end of TV programs).

Fast Reserve delivers active power according to certain criteria through an increased output from generation or a reduction in consumption from demand sources, following receipt of an electronic despatch instruction from National Grid.

4) Frequency Response

National Grid must maintain the continuously changing system frequency within the statutory limits, as defined in the National Electricity Transmission System Security and Quality of Supply Standards (NETSSQSS)[4]. To assist with this, National Grid procures frequency response from units, which can be categorised as either dynamic response or non-dynamic response.

National Grid procures three different types of balancing services to assist with frequency control:

- Mandatory Frequency Response (MFR),
- Firm Frequency Response (FFR),
- Frequency Control by Demand Management (FCDM)

All generators bound by the requirements of the Grid Code [5] are required to have the capability to provide MFR. MFR is an automatic change in active power output in response to a system frequency change.

FCDM provides frequency response through the interruption of demand. The electricity demand is automatically interrupted when the system frequency falls to below a trigger threshold. The demand customers who provide the service are prepared for their demands to be interrupted for up to 30 minute duration. Historic statistical trends have
shown that interruptions are likely to occur between approximately ten to thirty times per annum.

B. Reactive Power

National Grid manages the voltage of the GB system, to meet Transmission Licence requirements for secure and stable power transmission and to ensure quality of supply to customers. Voltages are largely determined by the flows of reactive power on the system. National Grid ensures that reactive power is provided on a local basis to meet the constantly varying needs of the system so that there are sufficient reactive power reserves available to meet contingencies, such as generation plant losses and circuit trips.

To assist with controlling reactive power flows, National Grid procures reactive power as a balancing service. It is obligatory for generators that are party to the Grid Code to have the capability to provide reactive power. These synchronous generators can be controlled to absorb or generate reactive power depending on the excitation. National Grid instructs these generators as to the level of reactive power that should be generated or absorbed to keep the system voltages within acceptable limit.

Reactive power is procured via the reactive power market, the arrangements of which are defined in section 4 of the CUSC [6]. The reactive power service is either procured via market agreements, or default payment arrangements.

C. Constraints

A constraint occurs when the capacity of transmission assets is exceeded so that not all of the required generation can be transmitted to other parts of the network, or an area of demand cannot be supplied with all of the required generation.

The volume of a constraint refers to the amount of generation that the transmission capacity is exceeded by i.e:

\[
\text{Volume of constraint} = \text{Volume of generation} - \text{Demand} - \text{Capacity of transmission system}
\]

As shown in Fig. 4 below, as generation and demand vary, the volume of constrained generation also changes. The green line indicates the system boundary capability. This can be an intact system capability, i.e. all equipment in service, or with circuits out of service. When the excess generation (i.e. generation – demand) exceeds the boundary capability, shown by the red line, the constraint is considered ‘active’ and action is required to ensure that the boundary capability is not exceeded.

A finite capability exists to transfer power in either direction across a boundary. Where constraints require that the transfer out of an area is reduced, by reducing generation or increasing local demand, these are termed “export” constraints. Circumstances where generation within the local group needs to be increased, or demand reduced, are termed “import” constraints. The example shown in Fig. 4 is an export constraint.

National Grid procures balancing services to manage system flows and alleviate constraints. The three categories that constraints are identified as are ‘intact’, ‘planned’ and ‘fault’ constraints.

- Intact constraints: where all transmission equipment is available for use (in service) and limitations of transmission capacity exist, the resultant constraint is referred to as ‘intact’. Reinforcement works to the transmission system or permanent changes to the generation and demand portfolio within the group are required to provide a lasting solution to such constraints.

- Planned: removing equipment from service for construction work or maintenance increases the power flowing on other system routes. This can intensify existing system issues and can result in additional constraints or create wholly new ones. The majority of constraint costs are typically due to planned outage work on the system.

- Faults: equipment failure or damage requires an outage to repair or replace the equipment. Fault outages, by their nature, cannot be forecast and can not be subject to normal planning practices and so can lead to high constraint costs.

When any constraint is managed by limiting or increasing the output of a generator, costs are incurred. Other actions can be taken to alleviate constraints such as intertripping, forward trading or bi-lateral contracts to change or limit generator output.

Intertrip services are required as an automatic control arrangement where generation may be reduced or disconnected following a system fault event to relieve localised network overloads, maintain system stability, manage system voltages and/or ensure quick restoration of the transmission system. There are two types of intertrip service, commercial intertrips, and system to generator operational intertrips.

System to generator operational intertripping can be used to alleviate constraints, which generally exist because of the system conditions occurring at the time that a generator signs
their transmission connection agreement, through the CUSC. Costs associated with system to generator operational intertripping include a capability fee and a tripping fee. The capability fee includes the cost of maintenance of the intertrip equipment and the training of staff. The tripping fee includes the costs associated with the contract energy imbalance that occurs when an intertrip trips due to fault.

Commercial intertrips are used to alleviate wider system issues, including constraints; they may be agreed at the time of a connection agreement or negotiated on an ad-hoc basis. The fees associated with commercial intertrips vary depending on the agreements made, but such fees can include an availability fee, and/or arming fee and a tripping fee if the intertripping facility is actually used.

The relative size of the scheme component is illustrated below (Fig. 5):

![Key components of BSIS scheme by value](image)

**Fig. 5 Relative size of BSIS cost components**

**IV. INCETIVISATION IN ACTION**

The Incentive Scheme is designed as a pain or gain system; National Grid is allowed to retain a share of any value created, but it must bear a share of any costs, should the targets not be met. The benefits of the savings achieved by National Grid are realized by the consumers through reductions in the Balancing Services Use of System (BSUoS) charges that are passed through to them.

As Fig. 6 illustrates, a cap is applied to the amount of profit National Grid can retain from the BSIS outturn and a collar is applied to the amount of loss that National Grid can be subject to. A ‘dead band’ range is agreed, where there is no pain or gain share of the BSIS outturns for National Grid.

For example, if the upside sharing factor has been agreed at 25%, then for every £1 under the deadband threshold the BSIS outturn is, National Grid would be allowed to retain 25p. Conversely if, for example, the downside sharing factor has been agreed at 30%, then for every £1 the BSIS outturns above the deadband threshold then National Grid will have to pay 30p towards BSUoS costs.

Prior to the commencement of each BSIS, the sharing factors and dead band are agreed between National Grid and Ofgem. The industry is consulted on the scheme design of each BSIS, via a consultation process led by National Grid [7]. All of the responses are carefully considered prior to the development and implementation of the final scheme.

The effect of the incentive scheme is that National Grid maintains a continuous tight watch on the costs of operating the system; making tactical decisions to optimize the system while minimizing cost on a minute-by-minute basis. This is particularly true for the scheduling of plant and the scheduling of engineering works which could incur constraint costs.

**V. CURRENT SCHEME ISSUES**

The current BSIS incentive scheme is a ‘fully bundled’ scheme, which means that National Grid’s profit or loss depend on the overall performance of the components that form BSIS. This framework has been in place since 2001 but it is not without flaws.

**A. Scheme Structure and Adjustments**

To OFGEM the BSIS scheme is a convenient one-stop-shop solution under which National Grid has responsibility for managing the grid. However, standing back one sees that only circa one-third of the costs embraced by the BSIS scheme are actually within National Grid’s direct control. The remaining two-thirds are determined by external forces, such as generators running regimes and market fuel prices. As such the BSIS target agreed at the beginning of each year between OFGEM and National Grid comprises roughly 1/3 National Grid’s budget forecast and 2/3 National Grid’s forecast of external market effects.

To help mitigate against external drivers the current incentive scheme contains adjustment factors for power price and market length [Net Imbalance Adjustment, NIA]. For 2010/11 the introduction of Reactive Imbalance Adjustment is proposed. These adjustment mechanisms are agreed with OFGEM as attempts to model and cancel out the effects of the external factors. For 2009 the value of the NIA adjustment was circa £300m. This is deducted from the full scheme cost of circa £900m before the incentive rules are applied and the profit/loss sharing calculated.

The incentive scheme has traditionally followed a similar format year on year, with a total target cost and sharing factors.
agreed annually. Even accounting for the adjustments (above) a major problem with this approach is the extreme volatility of the total cost of factors outside National Grid’s control. The proposal for the 09/10 scheme indicated only an 80% probability of the outturn lying within a £300m range about the scheme target figure. There was a 43% chance of the agreed scheme hitting either the profit cap or collar.

Further, drivers behind constraint costs have a number of differences from those underlying the management of energy and reactive power; particularly in relation to the seasonal and yearly impact on costs of planned and unplanned outages. This creates a different risk profile for constraints when compared with the other BSIS components.

In practice, for National Grid to win, it implies beating its own forecast through the introduction of cost saving initiatives (which will eventually become part of the baseline forecast and, therefore, any profit can be only realized in a single year) and/or benefiting from some windfall in the year.

B. Alternatives

An alternative structure might be a periodic review by the regulator of actions taken by National Grid to reduce costs. This would be an analogous to performance management schemes adopted by many companies these days, with a bonus being allocated to National Grid where the company has met or exceeded pre-agreed objectives. Such a scheme would place a high administrative burden on the regulator in reviewing National Grid’s actions, and would necessitate a relationship of trust between both parties.

A variant on the performance management model would be a post event system where National Grid identifies actions that have resulted in savings, and receive a share of the agreed value of the savings, or where the regulator identifies sub-optimal actions and fine National Grid a share of the additional costs they incurred. Again, this places an administrative burden on the regulator in closely monitoring the market. It removes the problem of windfall gains or losses that characterise the current incentive scheme, and allows longer term investment by National Grid, where an action delivering a saving over several years could receive a share of the saving over a longer period.

C. Unbundling the Incentive Scheme

Bundling constraints and energy into a single scheme and determining suitable incentive parameters tends to lead to a scheme that does not accurately reflect the relative risk profiles of the set of costs but rather a compromise between them.

One solution may be to un-bundle the scheme into separate components, each with their own incentivisation. However this risks confusing and conflicting messages being sent, together with an additional administrative burden which could add another level of complexity to an already complex scheme.

Another solution would be to un-bundle the scheme and jettison those parts which cannot be directly influenced by National Grid. This would remove the distraction of external forces and enable more focused, incentivised direct management of National Grid functions (few other companies would countenance incentivisation where the proportion of external risk is so large). However, this implies OFGEM picking up the management of the jettisoned parts. It may also remove National Grid’s overarching view of the UK electricity sector.

D. Multi Year Incentive Scheme

The current Incentive Scheme runs for a period of one year. The process of agreeing a target for the incentive scheme between National Grid and Ofgem repeats over a year cycle and carried a high administrative burden. Further, the incentivisation can be seen as erosive insofar as any gains made in one year form the base negotiating position in the next, so tending to favour short-term rather than long term solutions.

A longer term scheme would provide some certainty to National Grid on the longer term cost targets, enabling decisions to be made for investments with increased certainty, such as the consideration of investments in cost reduction tools or resources with a longer than one year payback.

VI. CHALLENGES FOR FUTURE INCENTIVE SCHEMES

The UK electricity sector is poised for considerable change, both in terms of generation / demand mix and the commercial market structure philosophy. On 3 February 2010 OFGEM published the findings of its ‘Project Discovery’ study which came to the dramatic conclusion that the existing UK market mechanisms could not “…deliver both security of supply and environmental objectives at affordable prices longer term, given the nature and scale of challenges facing the GB market.” [8]. It recommended a ‘more interventionist’ strategy, with policy proposals ranging in scale; from step-change reform, through new market mechanisms for renewables, to a complete metamorphosis of the electricity sector which would re-define the roles of energy buying and system balancing.

A. Impact of renewables

Over the same time period the British Government is driving for 20% of all energy to be produced from renewable sources by year 2020. This corresponds to 35% of electricity from renewables. The initiative is promoted through the government ‘Renewables Obligation’ in place since 2002 as managed by OFGEM to encourage growth of the sector [9]. Under this scheme renewables get a subsidy that can be greater per MW than the value of electricity they produce.

In reality the bulk of future renewables will be wind power, there being few other alternatives. However, once the typical load factor of UK windfarms is taken into account (circa 30%), some 70GW of wind would be needed to meet the government target. This is greater than the GB winter peak demand circa of 60 GW. It implies that system balancing services will have to accommodate massive swings in wind output, which will take precedence over the running of
conventional plant. It also implies dynamic transient commercial agreements while new commercial mechanisms are put in place to cope with the displacement of conventional plant, and whose cost per MW is likely to escalate dramatically as a result. This may indicate that the National Grid’s system balancing role will become a greater task, but at the same time be even more dominated by the ‘virtual’ markets of renewables (& carbon) trading rather than the physical movement of electrons.

B. Effects of Smart Technology

In parallel with above National Grid’s balancing service is likely to be impacted by greater penetration of ‘Smart’ technologies; smart metering, smart networks, together with increasing embedded (downstream, bio) generation, micro CHP technologies. These initiatives, coupled with dynamic pricing of electricity, are intended to make demand and supply more self-balancing downstream and external to the Grid. This could lessen the role of National Grid’s system balancing activity.

C. Outlook

With the above in mind the outlook for the BSIS scheme is uncertain. Yet the Balancing Services function of National Grid must continue to face the challenge of managing the system under some sort of incentive, while securing stable prices for contracts & tenders (e.g. for Response) in a market of changing commercial dynamics, changing plant mix, and in the face of increasing expenditure on elements such as Margin, Footroom and Constraints while new network capacity is built.

VII. CONCLUSIONS

This paper has presented an overview of the energy balancing service provided by National Grid in the UK power system. From this it can be seen that the incentive mechanism “BSIS Scheme” has led to the development of a comprehensive set of commercial and technical tools in order to balance and operate the system. The extent to which this scheme will continue to serve as a practical incentive when faced with the challenges of a transforming industry is open to wider debate.

VIII. DISCLAIMERS

The views expressed in this paper are those of the authors alone and not those of the National Grid PLC.

IX. REFERENCES