Methods for Allocating Access in the Congestion Management Model

Creative Energy Consulting

August 2021

Introduction

This paper examines and analyses methods for access allocation in the ESB's proposed Congestion Management Model (CMM)

There are basically two elements to the CMM:

1. Dispatched generation is paid at its local price (LMP). This creates a settlement surplus: the difference between the price retailers pay (RRP) and the price generators get paid (LMP)
2. This settlement surplus is returned to generators.

LMPs are already produced by the NEM dispatch engine (NEMDE) and so it is simply a question of applying these to dispatched quantities. However, there are several options for allocating the settlement surplus between generators, each with their own strengths and weaknesses. This note describes and discusses those options.

In this note, losses are ignored for simplicity, so in the absence of any congestion all LMPs are equal to RRP and there is no surplus to allocate. So only congested situations need to be considered.

This paper is presented in three main chapters:

- *theory and concepts*: how the key elements of settlement surplus, access, transmission constraints and nodal prices mathematically relate to each other;
- *allocation methods*: description of four methods for allocating access;
- *other issue*s: some miscellaneous issues that arise in the implementation of the access allocation methods

The paper does not attempt to formally assess or rank the four methods. Any ranking would reflect the various objectives of the CMM and their prioritisation.

Theory and Concepts

Constraint Formulation

Congestion occurs when a transmission constraint in dispatch becomes binding, which means that dispatch must be adjusted to ensure the associated transmission element is not overloaded. For simplicity, we will consider the situation where only a single constraint is binding in dispatch[1].

A transmission constraint takes the form:

$$\alpha_A \times Q_A + \alpha_B \times Q_B + \alpha_C \times Q_C \leq FGX \qquad (1)$$

where:

Q_A is the quantity of generation dispatched from generator *A,* etc

α_A is the *participation* of generator *A* in dispatch etc; a fixed number, usually between 0 and 1[2]

FGX is the flowgate capacity, which reflects the capacity of the associated transmission element or the transmission network more generally

[1] multiple binding constraints are discussed in the final chapter of this report
[2] treatment of negative participation factors is discussed later in this note

Figure 1, below, illustrates how such a constraint arises, using the example of three equal-sized generators (120MW each), with different participation factors:

- the blue generator has 75% participation
- the red generator has 100% participation
- the green generator has 25% participation

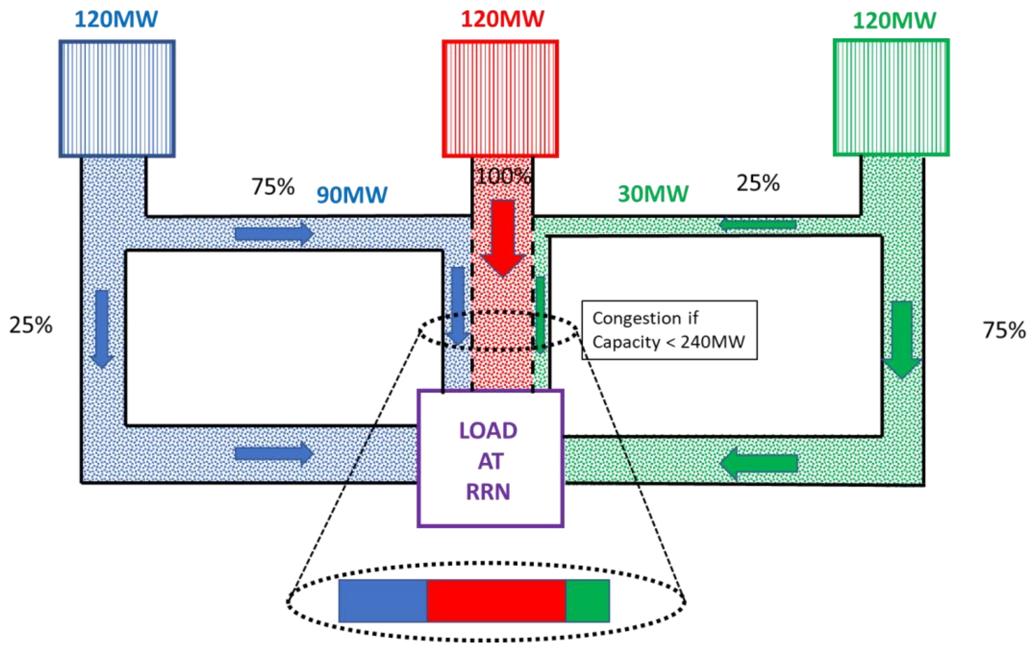

Figure 1: Diagrammatic representation of power flow through a flowgate

Figure 1 illustrate this as a system of pipes between the generators and the regional reference node (RRN). Generation output is represented by the three shaded boxes at the top of the diagram. The level of dispatch is represented by the amount of each box that is shaded; in this cases, full shading means the generators are fully dispatched.

The pipes mean that some electricity flows through the transmission element (or flowgate) that is the subject of the constraint equation, whilst some bypasses it. This is illustrative: power flows on a transmission grid are rather more complex than this[3].

The magnified strip at the bottom of the picture shows how much each generator contributes to the flow through the flowgate. Although each generator is dispatched at the same level, the red generator has the highest contribution, because it also has the highest participation factor.

Figure 1 shows that, with all generators at full output, 240MW flows through the flowgate. If the flowgate capacity is less than 240MW, there will be congestion and one or more generators will have to be dispatched back below their full output to avoid the flowgate being overloaded.

[3] for example, some of the output from the red generator would, in an electricity network, flow through the blue and green diversion channels

Single Binding Constraint

When the constraint binds, the two sides of the constraint inequality become equal:

$$\alpha_A \times Q_A + \alpha_B \times Q_B + \alpha_C \times Q_C = FGX$$

In the example in figure 1, above, the constraint would bind if FGX falls below 240MW. Various constrained dispatch scenarios are presented in the next chapter.

In this situation, the dispatch engine calculates a *shadow price* for the binding constraint which will be referred to as the flowgate price (FGP). Because of the way that LMPs relate to the flowgate price, a settlement surplus is created in LMP settlement, which is determined by the formula:

settlement surplus$ = FGP x FGX

Note that this is a $/hr figure, so one twelfth of this will be generated in each dispatch interval. For simplicity, the 1/12 will be ignored in this discussion.

One way to think about the flowgate price is as a tariff on generation that flows through the flowgate. The tariff revenue is simply the product of the tariff rate (FGP) and the flow volume (FGX). When generators are settled at LMP, they effectively pay this tariff, which gives rise to the tariff revenue – the settlement surplus.

Geometrically, this surplus can be represented as the area of a rectangle as shown in figure 2. Allocating this surplus then means dividing up this area between generators.

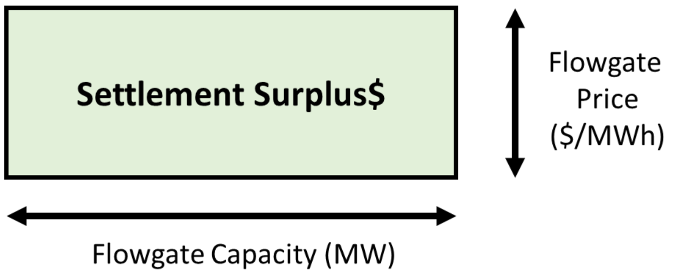

Figure 2: Geometric representation of the settlement surplus

Who should receive rebates?

The first question is which generators should receive a share of this settlement surplus. The congestion will only affect those generators that participate in the constraint: ie that appear on the LHS of the inequality and have a non-zero α. The congestion will impact on their LMP and also on their dispatch level. Generators not participating will continue to have LMP = RRP and have their dispatch unaffected. Thus, the surplus should only be allocated to *participating* generators.

If a generator is unavailable, it will also obviously not be impacted by congestion. Therefore, the surplus should only be allocated to *available* participating generators.

The second question is whether a participating generator is *actually* impacted by the congestion. Consider a generator who is out-of-merit (OOM): ie they have an offer price higher than RRP. These generators will not be dispatched with or without congestion, so the congestion has no impact on them. Therefore, OOM generators should not be allocated any of the surplus[4]. The surplus should only be allocated to *in-merit* available participating generators.

[4] the issue of how to identify OOM generators is discussed in the final chapter of this report

Flowgate entitlements

Figure 3 below illustrates one way of allocating the surplus between the three generators *A, B* and *C.*

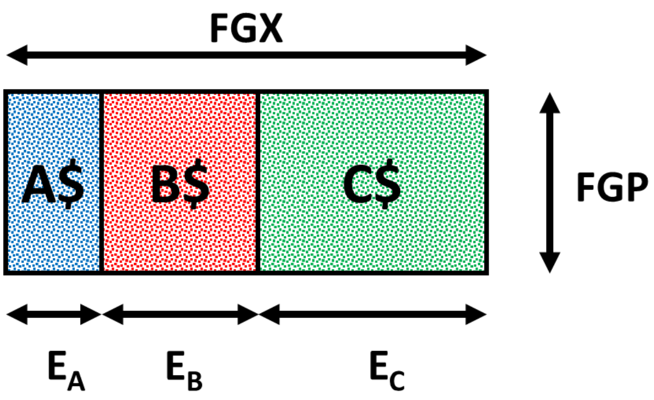

Figure 3: geometric representation of the allocation of the settlement surplus

For simplicity, the area is subdivided into rectangles, with widths equal to E_A, E_B, E_C. So, for example, generator *A* receives a rebate equal to:

gen A rebate$ = E_A x FGP

The *E* numbers are referred to as (flowgate) *entitlements*. Comparing figures 2 and 3, it will be seen that we must have:

$E_A + E_B + E_C = FGX$

Thus, the problem of allocating the surplus is equivalent to allocating the flowgate capacity between generator entitlements. This simplifies the problem, as the magnitude of the FGP then becomes irrelevant. There are several possible methods for allocating FGX, which are discussed in the next chapter of this paper.

Access

Entitlements are a useful concept for understanding methods for allocating the surplus. However, they don't in themselves address the key issue for generators, which is how much *access* to the RRP any particular allocation method will give them. Access refers to the quantity on which a generator is paid RRP under the CMM, as expressed by the equation:

Payment$ = A x RRP + (Q – A) x LMP (2)

Where:

Payment$ is the aggregate of the LMP payment and the allocation from the settlement surplus

A is the access quantity in MW

Q is the dispatch quantity in MW

Thus, each generator is paid RRP on its access, plus LMP on any difference between access and dispatch. Under the current market design, where access is set *equal* to dispatch, equation 2 simplifies to the familiar:

Payment$ = Q x RRP

In the case of a single constraint binding in dispatch[5]. there is a straightforward relationship[6] between access and entitlement:

$$A = E / \alpha$$

Or

$$E = \alpha \times A$$

Putting that into our entitlement allocation equation, we get:

$$FGX = E_A + E_B + E_C = \alpha_A \times A_a + \alpha_B \times A_B + \alpha_C \times A_C$$

This is similar to the NEMDE constraint on dispatch when it is binding:

$$\alpha_A \times Q_a + \alpha_B \times Q_B + \alpha_C \times Q_C = FGX$$

Thus, for any entitlement allocation, it would be feasible to dispatch each generator to its associated access level without violating the constraint[7]. Or, conversely, any feasible dispatch can be converted into an entitlement allocation[8]. In dispatch, the contribution of a generator to the LHS of the constraint equation[9] is referred to as flowgate *usage*. Thus, in our dispatch analogy, the flowgate entitlement is set equal to the flowgate usage arising in the feasible dispatch on which it is based.

These relationships mean that there are two fundamental approaches to deriving an access/entitlement allocation:

- find a feasible dispatch and infer entitlements from this ($E = \alpha \times Q$); or
- allocate the entitlements directly and then infer access from this ($A = E/\alpha$)

Both approaches are used in the allocation methods discussed in the next chapter.

Dispatch Efficiency Dividend

As discussed, each of the access allocation options reflects a feasible dispatch. Thus, in principle, generation could be dispatched identically to its allocated access (Q = A) and paid RRP.

$$Pay\$ = A \times RRP$$

However, the problem with this approach is that this access-based dispatch will generally be inefficient and, of course, one of the aims of the CMM is to allow a *more* efficient dispatch. Dispatch inefficiency is reflected in the total operating cost of the generation that is dispatched. If dispatch efficiency can be improved then, other things being equal[10], generator profitability is improved.

Under CMM, this improved profitability is paid out as an *efficiency dividend,* as shown below. The dividend received by each generator is non-negative, and the aggregate of these payouts equals the cost savings from the more efficient actual dispatch compared to the access dispatch. The dividend

---

[5] multiple binding constraints are discussed in the final chapter of this note
[6] this relationship arises because of algebraic relationships between LMP, alpha and FGP.
[7] note that this might lead to a generator having access in excess of its availability, so strictly speaking this would not be a feasible dispatch
[8] note this is true for a single binding constraint. Access and dispatch where there are multiple binding constraints are considered further in the final chapter.
[9] equation 1, above
[10] specifically that RRP is unchanged

is paid out not as increased revenue but increased *profitability*: the difference between revenue and operating cost. The

Recall that the payment to a generator under CMM is given by the equation:

Pay$ = A x RRP + (Q-A) x LMP

Operating cost can be expressed as:

Cost$ = Q x C

Where:

C is the short-run marginal cost of generation[11]

So the profit[12] is the difference between revenue and cost:

Actual Profit$ = A x RRP + (Q-A) x LMP – (Q x C)

= A x (RRP – C) + (Q-A) x (LMP – C)

In the case of access dispatch, Q= A and so:

Access Profit$ = A x (RRP – C)

The efficiency dividend is the difference between actual profit and access profit:

Dividend$ = (Q-A) x (LMP – C)

And:

Actual Profit$ = Access Profit$ + Efficiency Dividend$

The dividend payment is non-negative because of the relationships between dispatch quantity, LMP and C, assuming that generators bid at cost (BID = C). This is shown in table 1 below.

Situation	**LMP vs C**	**A vs Q**	**Sign of (Q-A) x (LMP – C)**
Fully Dispatched	LMP > C	Q ≥ A	Non-negative
Partially Dispatched	LMP = C	Unknown	Zero
Not dispatched	LMP < C	Q ≤ A	Non-negative

Table 1: sign of dividend under different dispatch situations

Now, to first order, the *aggregate* profit from efficient dispatch does not depend upon the access allocation method[13]. The allocation method simply changes the relativity of the two components on the RHS. If the access dispatch is quite efficient – ie fairly similar to actual dispatch – the aggregate efficiency dividend (the difference in cost between these two dispatches) will be relatively small,

[11] for simplicity of exposition operating costs are assumed to be proportional to output, but this argument can be generalised to non-linear cost structures
[12] this is the short-run - or operating - profit; the difference between spot market revenue and operating costs
[13] to second order, it *could* have an effect, given that it may affect pricing power and so the extent to which generators bid strategically away from cost and therefore reduce dispatch efficiency

meaning that the dividends payable to individual generators will also be small[14]. This is an important factor that will be considered in examining the allocation options.

Conclusions

The settlement surplus arising on a flowgate is the product of the flowgate capacity and flowgate price. Allocating the settlement surplus is equivalent to allocating the flowgate capacity between entitlements. This, in turn, is equivalent to finding a feasible dispatch and defining entitlements based on that dispatch.

Allocation Methods

Overview

As discussed above, there are two approaches to allocating entitlements and therefore the settlement surplus:

- directly allocate the FGX amongst entitlements, or
- find a feasible dispatch and deduce entitlements by multiplying dispatch quantities by participation factors.

Four allocation methods are described in this chapter, three of which start from a feasible dispatch. Each method is illustrated using our earlier example of three equal-sized (in terms of availability) generators, constrained behind a single congested flowgate. Generalising to differently-sized generators is straightforward. Generalising to multiple binding constraints is less obvious, and is discussed in the next chapter.

In the example, it is assumed that each generator has a positive participation factor. Options for dealing with negative participation factors are also discussed in the next chapter.

Pro Rata Options

These two methods, *pro rata entitlement* and *pro rata access*, take the same idea – of sharing equally – but approach the allocation from opposite ends. In each case the *pro rating* is with respect to the in-merit availability, for reasons discussion above. In the example of equal-sized generators, this means equal allocations.

With *pro rata entitlement*, the FGX is just divided up to give each generator an allocation in proportion to its size[15]. This is illustrated in figure 4, below, for our three equal-size generators. In this figure – and in subsequent figures – FGX is reduced to 96MW, thus causing congestion.

[14] recalling that these are non-negative, so it is not possible for large dividend components to net out to a small aggregate amount.
[15] ie its in-merit availability

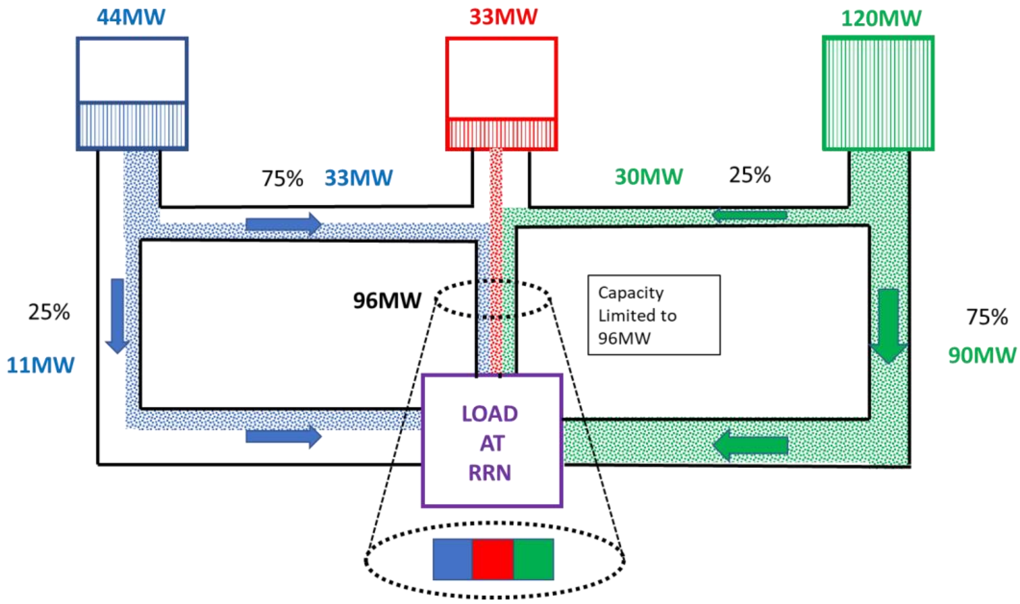

Figure 4: Pro Rata Entitlement Method

In figure 4, and in similar ones below, the "water level" in the box representing each generator is its dispatch level, which is analogous to its access in the corresponding CMM allocation. The width of its flow through the flowgate is the flowgate usage, corresponding to its entitlement in the CMM allocation. This translates directly to the size of its rebate, which is illustrated in the coloured rectangle at the bottom of the figure.

Because access is the entitlement divided by the participation factor, a generator with a lower participation factor enjoys a higher level of access than a similar-sized generator with a higher participation factor.

With a simple *pro rata* approach, a generator with a very low participation factor might receive access in excess of its availability; that is to say, it is better off than if there were no congestion. Such an outcome seems unnecessary and also counterproductive – given that this excess access allocated to one generator means that other generators receive less access. To avoid this, access is *capped* at availability, with the excess then shared between the remaining generators. So, in the example in figure 4, the green generator is allocated slightly *less* than a third of the FGX, to avoid it being overcompensated: allocated 30MW instead of 32MW. The other generators get slightly more entitlement as a result: 33MW each instead of 32MW each.

Under the *pro rata access* approach, each generator is allocated access in proportion to its size. Recall that this access allocation must represent a feasible dispatch. Dispatching all generators at their full availability would overload the constraint (which is why there is congestion in the first place). Thus, dispatch – and access – of each generator has to be scaled back until the dispatch is feasible. This is illustrated in figure 5, below

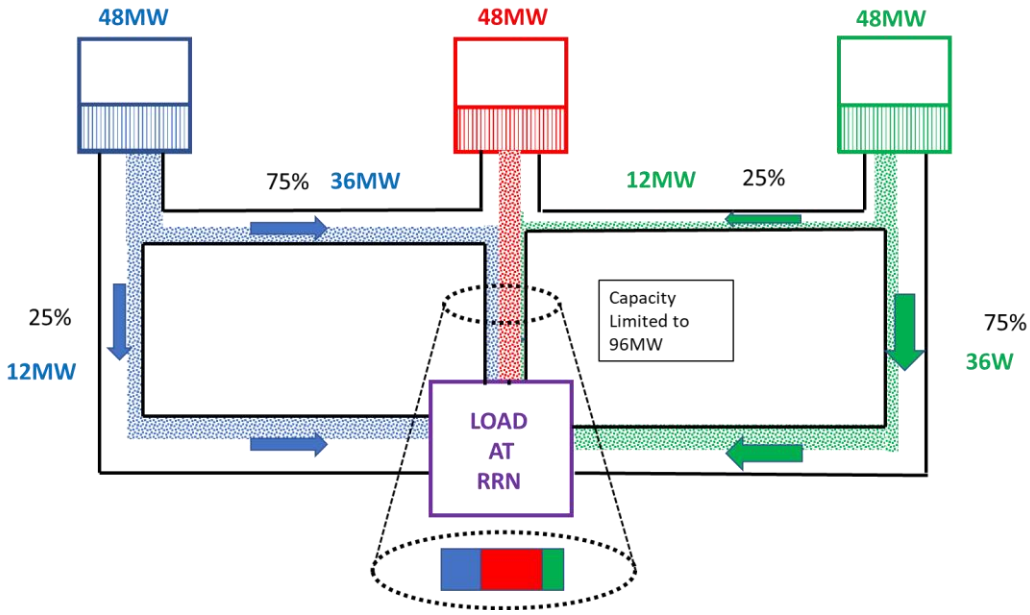

Figure 5: Pro Rata Access Method

It turns out that each generator can be dispatched at 48MW, causing the constraint to bind but not violate. So each generator receives 48MW of access.

Because entitlement is the product of participation and access, generators with high participation get a disproportionately high entitlement – and so share of the settlement surplus – under this approach.

In summary:

- under pro rata entitlements, the generators get equal[16] entitlements, but access is inversely proportional to participation; so low participation generators get more access.
- under pro rata access, the generators get equal access, but entitlements are proportional to participation, so high participation generators get higher entitlements.

Thus, there is no way to treat the generators equally on *both* measures: access and entitlements. In choosing between these two approaches, it would have to be decided which measure of "fairness" is preferred.

Winner Takes All

There are two options based on an efficient dispatch, depending upon whether generator operating costs are known. The winner takes all (WTA) is used where costs are *not* known. The inferred efficient dispatch (IED) method where costs *are* known (or, at least, estimated) is discussed in the next section.

Since, under WTA, costs are not known, access is based on the efficient dispatch that would be calculated if every generator submitted identical offer prices. Because, in this scenario, NEMDE considers that all the constrained generators have the same cost, it attempts to maximise *aggregate* generation (in order to offset the more expensive generation at the RRN) without overloading the constraint. This is done by dispatching generators in order of participation factor, starting from the

[16] or roughly equal, allowing for some capping of access at availability

lowest. This is referred to as *winner takes all*, because it likely leads to some generators getting fully dispatched (and so receiving maximum access) and others fully constrained (and so receiving zero access).

The WTA allocation for our three-generator example is presented in figure 6, below.

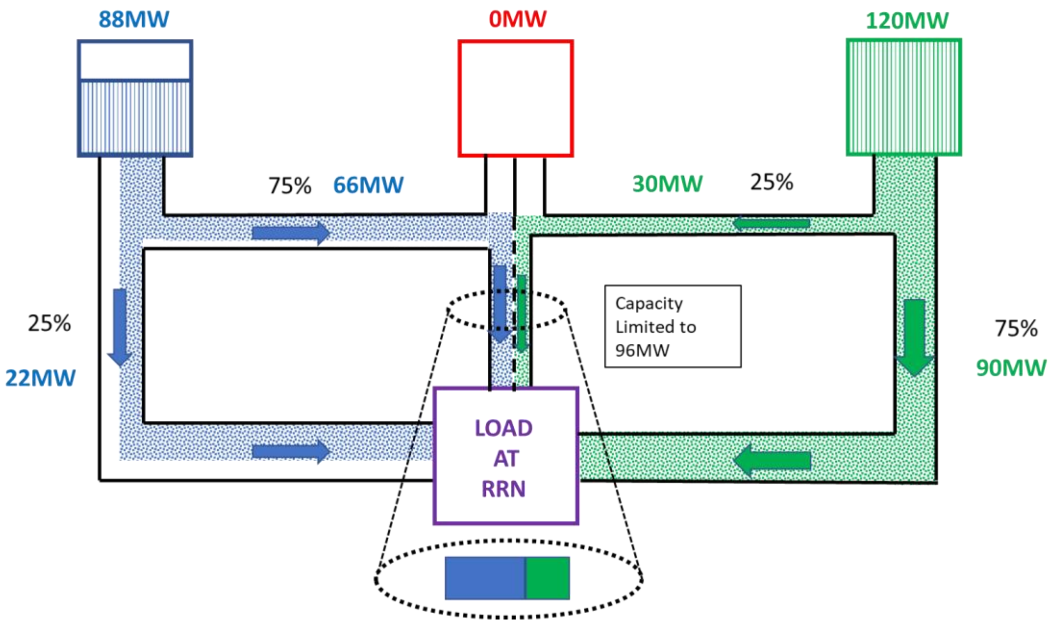

Figure 6: Winner Takes All method

It will be seen that the highest-participation (red) generator receives zero access. The lowest-participation (green) generator receives full. The third (blue) generator receives partial access, which corresponds to being the marginal generator in the associated dispatch. In general, there is just one marginal generator behind a constraint[17]. So, under WTA, one generator receives partial access and the remainder receive either full or zero access.

Note that, whilst the green generator has the higher access (getting full access), it has the lower entitlement. And vice versa for the blue generator.

One special case is a radial constraint, in which all generators have the same participation factor. In this case, because generators have identical cost and participation, NEMDE must treat them all equally, which in practice means an outcome identical to the pro rata access method.

WTA dispatch is a familiar dispatch outcome under the current market design, because where generators are constrained, they are incentivised to rebid down to the market price floor, in order to maximise their dispatch and revenue. Thus, they have (or at least appear to NEMDE to have) identical costs, and WTA dispatch is the result.

This similarity between the WTA allocation and existing dispatch outcomes can be considered both a strength and a weakness. Its strength is that it minimises the impact on existing generators, who of course are accustomed to this access allocation and have developed their trading systems and business models accordingly. On the other hand, part of the motivation for introducing the CMM is to avoid these kind of dispatch outcomes, whereas the WTA allocation in a sense simply perpetuates them.

[17] It is possible to have two or more marginal generators if they have identical participation factors.

The efficiency dividend discussed earlier is pertinent to the WTA allocation. With the WTA access dispatch being identical (other things being equal) to what is seen currently, the efficiency dividend formulation demonstrates that no generator is worse off under CMM[18] and many generators will be better off: sharing the benefits of improved dispatch efficiency.

Inferred Efficient Dispatch

WTA assumes all generators have the same cost. In the *inferred efficient dispatch* method, costs are taken into account, dispatching low-cost generators in preference to high-cost generators in the access dispatch, other things being equal. Of course, the costs used in the allocation can't just be based on bids, because this is what we have today and, as discussed, just collapses into a WTA outcome with "disorderly" bidding. Instead, costs must be *inferred* (hence the name of the method) in another way, independent of bids. Methods for inferring costs are discussed further in a section below.

An IED outcome is presented in figure 7 below. Note that previous methods have not required knowledge of generator costs, so these are now added to the picture.

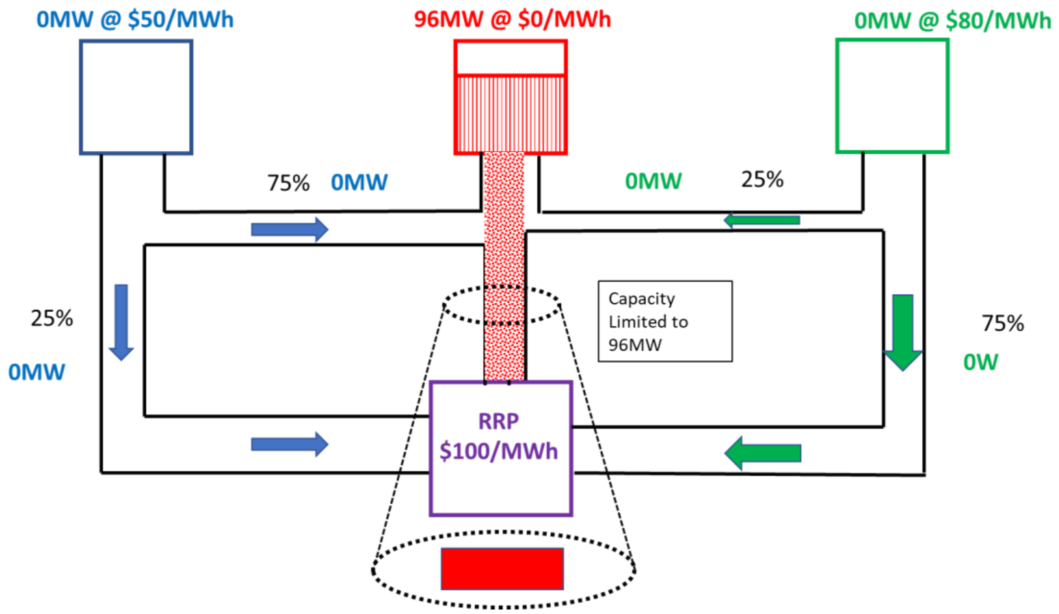

Figure 7: Inferred Efficient Dispatch method with RRP = $100/MWh

This allocation is based on an RRP of $100/MWh. Unlike the other allocation methods, the efficient dispatch will depend directly upon the level of RRP[19]. If RRP were rise to $1000/MWh, a different dispatch is efficient, as shown in figure 8, below.

[18] with the WTA allocation; noting the assumption that RRP is unchanged.

[19] noting that the level of RRP is *indirectly* relevant to the other methods, because it can affect the level of in-merit availability.

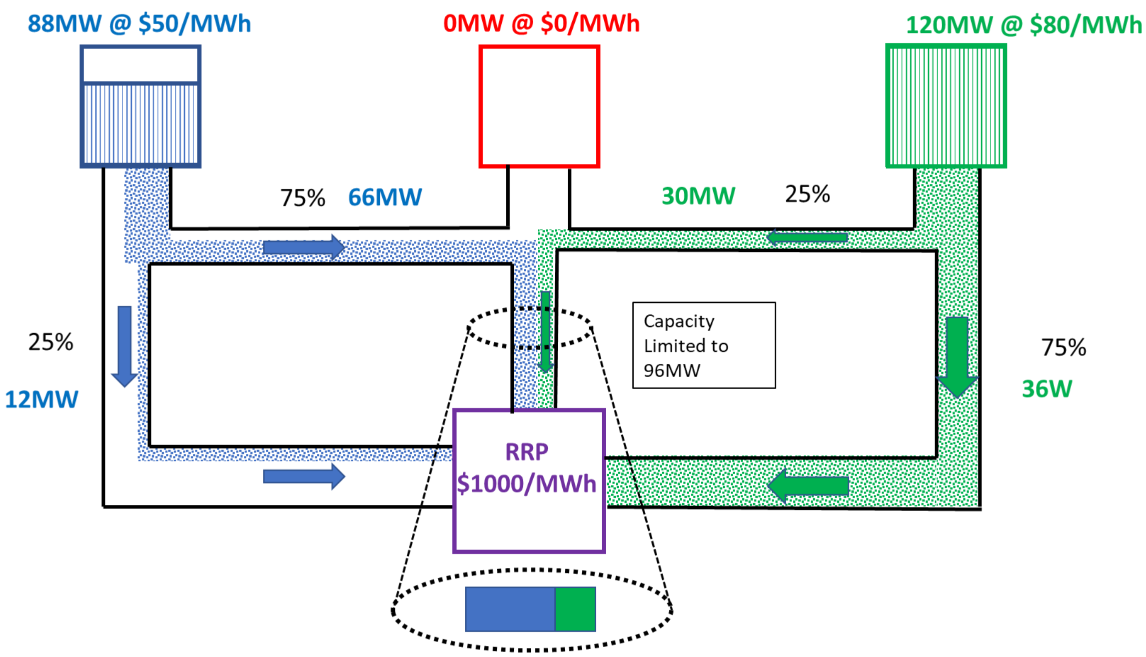

Figure 8: Inferred Efficient Dispatch method with RRP = $1000/MWh

Note that, for the lower RRP, the efficient dispatch selects the lowest cost generator (red), even though it has a high participation factor. This can be thought of as a kind of extension of the method of removing OOM generators. In this case, an OOM generator would have a bid above $100, but in the efficient dispatch, the generators with bids at $80 and $50 are also taken out of dispatch.

For the higher RRP, the efficient dispatch is identical to the WTA dispatch shown in figure 6. In general, for very high RRPs, the efficient dispatch will converge towards the WTA dispatch.

Whilst this dependence of the allocation on RRP appears complex, it simply reflects what would happen under a real-life efficient dispatch. Indeed, if costs are inferred accurately and generators bid at cost, the inferred and actual dispatches will be identical, meaning a simple payment of RRP on the dispatched quantity, with no side payments based on LMP. And, framed in this way, it might be argued that it is those allocation methods which are *independent* of RRP that are really more complex, given that the difference between access and actual dispatch – and the associated efficiency dividends – will vary with RRP.

Of course, given the difficulties in estimating costs accurately, and the likelihood that generators will commonly bid away from cost for strategic reasons[20], it is unlikely that access and actual dispatches will exactly match. Nevertheless, out of all of the methods, this seems the one likely to minimise exposure to LMP, which is a point in its favour. It is a kind of “have your cake and eat it” scenario: have the cake of efficient dispatch whilst eating the cake of minimal LMP exposure.

Conclusions

There are myriad ways of allocating access and a few of the more promising candidates have been described here. Whilst the strengths and weakness of these methods have not been explored in detail, it is fair to say that the preferred option will be selected depending on the prioritisation of different objectives, such as simplicity, risk, transitional impacts etc.

[20] eg to cover forward contract positions

Other Issues

Multiple Binding Constraints

The discussion above assumes that a single constraint is binding across the NEM. In practice, it will be common for two or more constraints to bind concurrently. Therefore, the allocation methods need to be generalised to accommodate this situation.

Conceptually, there is no difficulty with this. If two constraints, say, bind, a settlement surplus is created on each constraint. For example, if constraints 1 and 2 bind then:

settlement surplus\$ = FGP_1 x FGX_1 + FGP_2 x FGX_2

The allocation method can then be applied to each constraint separately: generators participating in constraint *1* will get a share of the surplus arising on that constraint; generators participating in constraint *2* will get a share of that surplus. A generator participating in both constraints will get two payments, which are simply added together.

However, some new complexity is introduced when thinking about access. A generator participating in two binding constraints will get two respective entitlement allocations, each with (in general) a different level of implied access. But there can only be one access quantity. So which is it? The answer is that the *effective* access quantity will be a weighted average of the implied access quantities arising from each constraint, with the weighting factor being the product of the participation factor and FGP. For example, with two binding constraints:

effective access = (FGP_1 x α_1 x A_1 + FGP_2 x α_2 x A_2) / (α_1 x FGP_1 + α_2 x FGP_2)

Another twist is that the access quantities do not represent a feasible dispatch. Obviously, a generator can only be dispatched to a single, particular quantity, so having different access quantities for different constraints breaks that analogue. Furthermore, the effective access will also not generally reflect a feasible dispatch.

In principle, the two methods based on a dispatch approach could be adapted to represent a feasible dispatch. For example, the WTA outcomes arising from an efficient dispatch of generators with identical offer prices is seen in today's market, and could be replicated. Similarly, an efficient dispatch could be calculated using inferred generator costs, for the IED method. But these options are more complicated, as they rely on the use of NEMDE – or similar dispatch engine – to co-optimise generator output across multiple binding constraints.

In any case, it is likely that most generators only participate in a single binding constraint, for any dispatch interval, even if there are other constraints – with *other* participating generators – binding concurrently. Even where a generator does participate in multiple, concurrently-binding constraints, one of these constraints is likely to dominate in terms of FGP and/or participation factor. Therefore, it is doubtful whether such more sophisticated methods of handling multiple constraints would be justified. Instead, the simple approach of independent allocations for each binding constraint should be used.

Negative Participation Factors

The above discussion assumes that generation participation factors are positive. In reality, they can be – and commonly are – negative. A generator with a negative participation factor in a constraint is referred to as a *flowgate support generator*. The "support" – in the sense that its output helps to relieve congestion - is specific to the constraint, given that it might have positive participation factors in other constraints.

There are two different ways to deal with flowgate support generators in the allocation method. The simplest is to allocate zero entitlements to them. This means they are simply paid LMP[21]. The LMP they receive will be higher than RRP, due to their negative participation in the binding constraint. Because they help to relieve congestion, they will also generally not be constrained off[22]. So, they would receive, in a sense, full access to the RRP, plus a bit extra from the excess of LMP over RRP.

Alternatively, the generator can be paid RRP on its output. As with the current market design, this effectively means that the generator receives access equal to its dispatched output. In entitlement terms, it is allocated an entitlement equal to:

$$E = \alpha \times Q$$

Since alpha is negative, the entitlement is also negative. This means the entitlement of other, positive participation generators, can be increased. So, whichever allocation method is used, it would operate on this enhanced availability of entitlements.

There are pros and cons with each approach. However, whichever is chosen, it is easily accommodated into the allocation methods described above

Estimating Generator Costs

The methods described above rely on estimates of generator costs, in two ways. Firstly, all methods seek to exclude out-of-merit generators: those with operating costs higher than RRP. Secondly, the IED method calculates a dispatch based on estimated operating costs.

There are two basic approaches to estimating costs. The first is the familiar "bottom up" approach of understanding the particular generation technology and then making estimates of the component input factors: eg heat rate, fuel cost etc. The second is an *inference-based* approach, where costs are inferred from historical patterns of generator behaviour. For example, if – historically – a generator has always run when prices (LMP or RRP, depending upon the market design) are above $100/MWh, but rarely runs when prices are below this, it can be inferred that its costs are around $100/MWh[23]. Or, at least, it only runs when prices are above $100/MWh; possibly its costs are below $100/MWh, but there is some "economic withholding" for strategic reasons. Either way, $100/MWh would be a suitable price point for deciding whether the generator is out-of-merit, and as an input into the efficient dispatch calculation. Indeed, a method that reflects strategic behaviour is arguably a better approach.

The problem with both approaches is that costs can vary. This is particularly the case for energy-constrained generators (hydro and storage), whose offer prices will depend upon storage status and on expectations of future prices. For short-duration storage (eg batteries), these prices can vary quite rapidly, so it will be difficult to make statistically valid inferences from historical behaviour. In this case, it might be suitable to infer dispatch quantities rather than offer prices. For example, a 4-

[21] assuming that there is only the single binding constraint

[22] although they could potentially still be constrained off if they positively participate in another binding constraint

[23] and note that this assessment does not need to refer to its offer price. It might be offering at $100/MWh continuously. But it might, instead, offer at the market price floor whenever prices are above $100/MWh and at the market price cap whenever prices are below. Both lead to the same dispatch outcomes and so the same inferred costs. But an approach based on, say, average historical offer price would come up with a very different answer. Particularly when offer prices are manipulated simply to make the generator look more attractive in the IED and thus likely to be allocated higher levels of access.

hour storage scheme might aim to discharge daily between 4pm and 8pm and this could be incorporated directly into the inferred efficient dispatch

Notwithstanding these inevitable inaccuracies in the cost estimates, it may be preferable to utilise these rather than just ignore cost. For example, an efficient dispatch method using approximate cost assumptions may be superior to a WTA method which simply assumes (implicitly) that all generators have the same cost.

Conclusions

The methods described in the previous chapter can easily be generalised to accommodate multiple binding constraints and negative participation factors. However, the relationship between access and a feasible dispatch may then become less clear.

Generators costs used for determining out-of-merit generators, and also specifically for the IED method, can be inferred based on historical behaviour or estimated using conventional bottom-up calculations.