ENERNOC

EnerNOC Pty Ltd
Level 18, 535 Bourke Street
Melbourne, Victoria 3000
Australia

Tel: +61-3-8643-5900
www.enernoc.com.au
info@enernoc.com

Rob Jackson
Australian Energy Market Operator
Level 22, 530 Collins Street Melbourne VIC 3000

Submitted via email: rob.jackson@aemo.com.au

10 March 2017

Response from EnerNOC to AEMO's *Market Ancillary Services Specification Issues Paper*, published January 2017

Dear Rob,

EnerNOC is a global provider of energy intelligence software and demand-side management services, and is the largest independent demand response (DR) aggregator in the world. We work with commercial and industrial energy users to offer their demand-side flexibility into wholesale capacity, energy, and ancillary services markets, as well as demand response programs offered by retailers and regulated utilities in twelve countries around the world. Locally, EnerNOC is a market participant in the Wholesale Electricity Market (WEM), the National Electricity Market (NEM) and the New Zealand Electricity Market (NZEM). EnerNOC's regional head office for Asia-Pacific is located in Melbourne.

EnerNOC is grateful for the opportunity to provide comment on AEMO's Issues Paper. In the New Zealand Electricity Market, EnerNOC has been aggregating interruptible loads ("IL") and offering them into New Zealand's equivalents of the contingency Frequency Control Ancillary Services (FCAS) markets since 2009. Today, EnerNOC is the largest single provider of FCAS in New Zealand, where IL constitutes approximately 70% of the contingency FCAS supply mix. EnerNOC participates in similar contingency FCAS markets in Alberta, Ireland, the UK, and Germany.

In the NEM, EnerNOC has supported the *Ancillary Services Unbundling* rule change in its journey through consultation, and we are interested in ensuring it is implemented in a way that allows new technologies including IL to participate in the FCAS markets that AEMO oversees. In our view, IL and other distributed aggregated technologies can provide AEMO with valuable new frequency control tools and increase competition in the FCAS markets. However, they will only do this if market frameworks and service specifications do not unduly exclude them. It is through this lens that we offer these contributions to the questions posed in AEMO's Market Ancillary Services Specification ("MASS") Issues Paper.

1. ***What barriers to entry for new Market Participants and new technologies are contained in the current MASS and what options are available to overcome the barriers while maintaining the integrity of the markets?***

For a provider of aggregated IL, today's MASS presents no technical barriers to entry. The primary barrier to new participants is a practical barrier: that the MASS is difficult to read and understand. In order to reach the conclusion that there are no technical barriers, EnerNOC had to ask many clarifying questions to AEMO staff. The MASS could be made more accessible with more logically organised sections, less reliance on defined terms in the glossary, and by separating the descriptions of requirements for different types of controllers. It should be noted that the MASS has a companion document: the Market Ancillary Services Verification Tool (and its own guide document). This tool was immensely valuable to EnerNOC in understanding the calculations described in the MASS, and reduced the number of questions that required asking. It would be valuable if AEMO would continue to maintain and publish a verification tool alongside future iterations of the MASS.

In our view, the biggest potential barriers to entry for new-entrant FCAS providers lie not in the MASS itself, but rather in the various registration procedures that AEMO may implement to accommodate the Ancillary Services Unbundling rule change. The crucial aspect is how AEMO will handle aggregation. It seems likely that most new Market Ancillary Services Provider (MASP) applicants will be offering aggregations of something – i.e. distributed interruptible loads, or distributed networked batteries. At the time they register with AEMO to become a MASP, such aggregators are likely to have only a small quantity of FCAS available, from a handful of pilot sources. However, they will then substantially grow their quantities over time, as they add new sources to their portfolio. Aggregators may source their FCAS quantities from dozens, hundreds, or thousands of distributed sources. This differs from the types of new-participant applications AEMO typically receives: large power stations that register once, and whose maximum capability and maximum offered quantities do not change significantly over time.[1]

There are a few ways the new rule implementation could (unwittingly) go awry, and unduly deter new entrants from participating. For instance, AEMO could require that each time an aggregator adds a new source (to its existing registered dispatchable unit), dozens of pages of paperwork need to be filled out, a significant application fee is payable, a third party engineer must certify each technology installation, or any combination of these scenarios. Any such requirements would undermine the economics of FCAS provision from capable small-scale sources, and are likely to deter potential aggregators from applying to become a MASP.

There have been no indications that AEMO will put such requirements in place, but also no confirmation that it will not. Potential new market entrants are anxiously awaiting guidance on this topic, so that business plans can be finalised and technologies can be commercialised. EnerNOC is hopeful that AEMO will put reasonable and workable processes in place to implement the new rule, ideally involving a robust, one-time registration process for becoming a MASP,

[1] The Small Generation Aggregator framework seems the closest equivalent, although that does not have the same system security implications.

where the provider's technology and operational processes are demonstrated and certified once (in order to give AEMO the confidence it needs in the provider's ability to deliver FCAS reliably), then a simple and scalable process for adding new sources to an existing aggregation.

From the Issues Paper it is apparent that these registration-related considerations are outside the scope of this consultation and so we won't discuss them further, but we believe they are noteworthy because their resolution will materially impact the success of the Ancillary Services Unbundling rule change, by determining the number of new MASP applicants AEMO receives. If no new FCAS suppliers are attracted, there will have been little point in reforming the MASS to accommodate them.

2. ***What options exist for determining the total change in power flow from aggregated loads?***

In our view, for an aggregated IL unit using a simple switching controller and supplying contingency FCAS, the current measurement and verification (M&V) in today's MASS is straightforward and appropriate – representing a simple "load prior to excursion, minus load after excursion" calculation. Said otherwise, values "FA", "SA" and "DA" are all appropriate references for measuring the "prior" value. This is similar to the method Transpower employs in New Zealand to measure the response of their 'Fast Raise' equivalent service.

3. ***Do you agree with the approach to determining the performance of variable generation and if not, how should it be determined?***

No response.

4. ***Other than high speed recorders, what options exist for verifying the performance of the plant while maintaining the integrity of the services?***

High speed recorders are an essential component of verifying performance of the plant or aggregated unit. EnerNOC can't envision a way of verifying Fast Raise performance with due accuracy and integrity without employing a high speed recorder.

50ms resolution is probably higher than is actually required to verify the delivery of the services currently specified in the MASS. It seems likely that working from, say, 100 ms resolution data would not increase measurement errors significantly.

While it is convenient and pleasant to work with ample data, it is important to bear in mind that each element of the service specification imposes costs on providers. These costs are ultimately borne by consumers. Worse, if the per-site costs become too high, they will form a barrier to participation by smaller sites. This will reduce competition, which could greatly increase clearing price of the services.

In general, requirements should be as loose as can be made to work adequately. Beyond that point, there should be some consideration of the trade-off between the benefits of reduced measurement errors and the greater costs borne by consumers.

If AEMO intended only to procure the services currently specified in the MASS, then a 50 ms resolution could be considered an over-specification. However, such high-resolution data does provide a degree of future-proofing, as it could be used to verify the delivery of much faster services.

5. ***Do you agree that the principle underlying the MASS should be related to the control of power system frequency, and not just the delivery of defined an amount of energy? What other principles do you believe are required?***

In general, we agree. The purpose of the MASS should be to allow AEMO to procure the services it needs to control frequency[2], and do so in a technology-neutral and least-cost manner. That said, FCAS suppliers need to design their operations around creating offers that are reasonable, deliverable, and compliant with participant obligations under the NER. The MASS is the document that will inform how FCAS suppliers decide on an offer quantity, in MW. As such, it is imperative that the MASS describe a clear framework against which suppliers can create offers and be confident that the offers are compliant with the MASS – regardless of real-time power system frequency conditions, and how AEMO may decide to employ schemes to control frequency.

6. ***Given these principles, what is the most appropriate performance measure for regulation services?***

EnerNOC is supportive of AEMO's initiative to improve the MASS by putting more rigour around how the regulation services are verified, including verifying real-time response to AEMO's AGC signals. This will bring the regulation services in line with the contingency services, which are verified using a clear and robust methodology following each event. Given that regulation costs are recovered from participants using a robust methodology ("causer pays") it seems appropriate to have a similarly robust methodology to ensure that enabled regulation providers are indeed providing the services that the market is paying for. The magnitude of regulation costs recovered ($32m in 2015 and $68m in 2016[3]) is such that it seems appropriate to include a robust definition of the service in the MASS, including a verification mechanism.

7. ***What limitations exist to inhibit plant enabled to provide one of the contingency services from handing-over smoothly to other services following an event and how can these limitations be addressed?***

EnerNOC is concerned that if implemented too prescriptively, the proposed "description of the expected transition response" in section 3.2.4 of the Issues Paper would preclude most IL from participating in the FCAS markets.

In practice, very few loads can be controlled in a proportional fashion to provide a linear ramp up or down. A far wider range of loads can be switched off rapidly by a switching controller.

[2] To within the operating band defined in the Frequency Operating Standards overseen by the AEMC's Reliability Panel.
[3] Source: AEMO A/S Payments Summary Reports, accessed via AEMO website 9 January 2017.

The current MASS (and equivalent documents in all the other markets in which EnerNOC operates) allow for contingency frequency response services to be provided by loads being switched off and on by switching controllers. The implementation of the Ancillary Services Unbundling rule should allow large-scale provision of IL at low cost, so long as switched loads can participate. Requiring any form of ramping would limit participation to those rare loads capable of providing regulation services, resulting in much smaller volumes and higher costs.

Further, a switching controller connected to a load can respond and deliver its FCAS quantity very quickly. The response time depends on the load, but many can achieve less than 1 second. Since the purpose of the Fast Raise service is to arrest the falling frequency, faster responses are more valuable and should be encouraged. Today's MASS does this by instructing suppliers to offer "twice the time average of the response between zero and six seconds"[4]. The Issues Paper notes that AEMO proposes to include a description of the expected transition response[5]. Mandating a *linear* response would be a mistake, as not only would it preclude participation by most loads, but it would also remove the incentive for enabled Fast Raise suppliers to respond as quickly as possible. It would thus hinder AEMO's ability to *arrest* falling frequency.

Another characteristic of aggregated IL is that it can't typically restore linearly over a given period, as a generating unit might. The interruption of a load is achieved via automated signal to a PLC, or in some cases, a circuit breaker. However, the restoration of a load must typically be carried out manually, often with the involvement of on-site staff, due to safety considerations[6]. This may mean that it takes longer than one, four, or five minutes for an IL aggregation to return to its pre-contingent load level. Example IL restore processes and relating timings from EnerNOC's IL customers in New Zealand are provided in Appendix A. Requiring IL to restore in an exact linear ramp would preclude IL technology from participating entirely, and deprive the NEM of a valuable tool for arresting frequency.

Figure 1 below shows a typical example of how 100 MW of physical aggregated IL would respond to a frequency excursion, if enabled for all three services[7]. Although each individual load is switched off almost instantaneously, since the response times vary between loads, the aggregate effect is more ramp-like. At the restore, the slope is much gentler, because response times vary

[4] Issues Paper p10. The effect of this M&V is that any technology that can deliver its physical quantity faster than a linear ramp, can offer a quantity in excess of its physical quantity. In the example in Figure 1, this means the FCAS supplier would be credited with > 100 MW of Fast Raise delivery (the yellow 'area under the curve'), despite supplying exactly 100 MW of physical load shed.

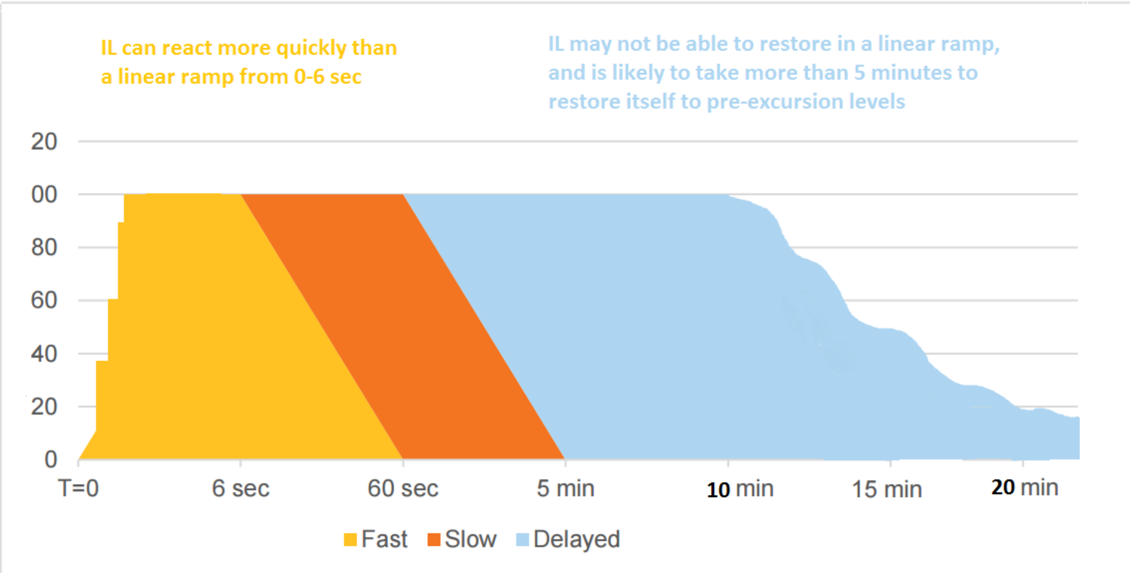
[5] Issues Paper, p12

[6] e.g. it would be unsafe to remotely restart many types of interrupted industrial process, as it is imperative to ensure that on-site staff are clear of machinery, etc.

[7] Note that IL is load decrease, rather than generation increase. However, for consistency with the diagrams in the Issues Paper, we have depicted IL in the inverse direction, as if it were a generation increase. Also note the diagram reflects the response of a single aggregated IL unit that is enabled for *all three contingency raise services. simultaneously*. This is consistent with Figure I in the Issues Paper. As intended in the MASS, an aggregated IL unit would begin its ramp back to its normal level based on the earlier of 1) the expiry of the timeframe of the slowest enabled service or 2) *frequency recovery* to 49.9 Hz. As such, this diagram (and presumably also the diagrams in the Issues Paper) depict an excursion where the frequency remained below 49.9 Hz for at least ten minutes.

considerably due to the manual involvement, and because different pieces of plant start up in a variety of ways.

Figure 1: Example profile of a single aggregated IL unit employing switching controllers, and enabled for all three services, during a situation where frequency remains below 49.9 Hz for at least ten minutes:



A real-world example of the restoration process of EnerNOC's New Zealand IL portfolio is provided in Appendix B.

There are ways that AEMO can work with IL providers to ensure a "smoother" restore to pre-contingent load levels, avoiding unmanageable frequency oscillations[8]. In New Zealand, the system operator manages individual IL providers to restore load in a controlled fashion. AEMO could request that IL providers restore loads in blocks, or similarly restore in an intentionally staggered fashion, so that AEMO can re-dispatch the system in parallel with the restoration of significant IL load. IL aggregators will know the restore capabilities of their constituent loads, and may have flexibility in orchestrating a restore process.

In ensuring the MASS efficiently maintains the power system security, arresting frequency fall should be the primary objective, whereas ensuring a precise restoration to pre-contingent levels should be a secondary consideration. The costs of failing to arrest frequency decline are extreme (triggering UFLS, or a cascade failure), whereas the costs of an imprecise transition back to pre-contingent levels are relatively minor (perhaps involving increased need for regulating FCAS for a handful of dispatch intervals). Mandating an unnecessarily prescriptive restoration profile would exclude aggregated IL from participating in the FCAS markets, and sacrifice the former objective for the sake of the latter.

[8] Note that, in contrast to the way IL is *delivered* in binary fashion over a sub-second timeframe, IL aggregations with switching controllers will not *restore* loads in 'binary' fashion, which might cause a 'shock to the system' or an oscillation. Rather, IL aggregations will restore in a gradual fashion over a timeframe of minutes to hours.

Avoiding over prescription: a useful case study from New Zealand

The most cost-effective approach is to minimise the requirements, so that the service is required to be just good enough to solve to problem (arresting, stabilising, restoring frequency). This will provide a better outcome for consumers than specifying an awe-inducingly superb service that can only be provided by a handful of providers at great cost. New Zealand provides a useful example here: it was noted that some providers of their equivalents to FCAS contingency raise services tended to over-deliver substantially in certain circumstances. There was concern that this over-delivery could lead to the frequency overshooting, such that an under-frequency event would be immediately followed by an over-frequency event. This could cause the System Operator to breach its obligation to control frequency below 52.0 Hz, and potentially lead to system collapse (if the over-frequency recovery went far enough to cause generators to trip off). It was intuitively obvious that such a problem could occur, and would be very serious, and could be avoided by imposing restrictions on the amount of over-delivery allowed. However, such restrictions would have increased the cost of providing the services, and may have prevented some types of suppliers from providing the services. A study was performed to verify that this risk was genuine.[9] The results surprised everyone: the anticipated problems wouldn't occur unless 25% of the system load was participating in providing the IL service – an implausibly high figure, and far below the quantities of IL the System Operator had contracted into the system. The report concluded that "over-frequency due to IL over-provision is not currently an issue nor is it likely to become an issue in the foreseeable future", and so the restrictions were not imposed and over-prescription costs were avoided.[10]

We suspect that the need for smooth ramps between services may be similar: something that seems like an obviously good idea, if costs and competitive impact were ignored, but which may not actually be necessary to achieve the desired outcome of arresting, stabilising, and restoring frequency.

It may be that frequency control in parts of the NEM is more challenging than in New Zealand, such that tighter specifications can be shown to be necessary. However, we consider that requiring defined ramps in response, or tight upper bounds on delivered quantities, would constitute an over-prescription of the services, and would result in fewer providers being able to provide the services, at a higher cost, to the detriment of consumers.

During further consultation, it would be useful if AEMO would elaborate for participants on the value of having "orderly" transitions. As noted, many (newer) contingency FCAS technologies can respond very quickly, in a binary fashion, to contribute to frequency arrest. However, after that

---

[9] Transpower, System Operator TASC 045 Report: *Investigation into over-provision of interruptible load*, 1 May 2014, available at https://www.transpower.co.nz/sites/default/files/bulk-upload/documents/TASC%20035%20Report.pdf

[10] Importantly, the New Zealand System Operator modeled IL under a range of scenarios, including 'worse case' scenarios with low inertia and low demand. At the time of modeling, the System Operator had contracts for 335 MW of "Fast Raise" equivalent service ("FIR"), and 790 MW of "Slow Raise" equivalent service ("SIR"). Each scenario modeled these quantities of IL responding, even if not enabled to provide the service (at the 1 second mark for FIR, and between the 2-5 second mark for SIR). In all modeled scenarios, it was determined that any frequency over-recovery would remain well below 52 Hz.

time, such technologies are unlikely to restore themselves to pre-contingent levels in a sudden, binary fashion (see Appendix A for an indicative restore profile for an IL unit). As a result, the risks of post-contingency frequency oscillations or "shocks to the system" are likely to be low – and needn't necessitate making the MASS' new service-handover descriptions so prescriptive that such technologies decline to participate. In the amended MASS, it may suffice to stipulate that any aggregated unit (be it IL, batteries, or otherwise) is prohibited from 'switching back' in binary fashion, and must stagger its restore by at least some minimum amount.

This does mean that some responses would tend to continue beyond the expected period. However, proposal in the Issues Paper that slower services should not be delivered if the frequency has already recovered to the normal operating band (or above the provider's allocated *frequency deviation setting)*[11] would prevent the scenario shown in Figure 3 of the Issues Paper from occurring. In Figure 3, the switched unit enabled for Delayed Raise (in blue) would only initiate its response at the 60 second mark if the frequency remained below the unit's *frequency deviation setting* at that time. This would indicate that, despite the over-delivery of the yellow Fast Raise unit and the orange Slow Raise unit (by continuing to respond and failing to initiate handover at the desired time), frequency was still low enough to require the contribution of the blue Delayed Raise unit, and the actions of the three units together were assisting AEMO in restoring frequency to the normal operating band. As such, we consider that the actions of the three units in Figure 3 are consistent with AEMO's principle that controlling system frequency should take precedence over simply registering and assessing each unit's energy response.

8. ***What should AEMO consider when drafting of a detailed description of the transition requirements from one contingency service to the next?***

See response to #7

9. ***What limitations exist to inhibit the ability of plant to resume standard operations in a timely manner following the recovery of local frequency?***

See response to #7

10. ***In your opinion, what is the most practical and efficient method of measuring the performance of generating units or loads registered to provide regulating raise or regulating lower services?***

After testing and commissioning, there should be ongoing monitoring of the accuracy with which enabled units meet their AGC targets. Further, a time limit for meeting the AGC target should be defined and enforced. Regulation *payers* are assessed payments based on their causer-pays factors, calculated every 4 seconds. It seems reasonable that regulation *providers* should be assessed via a similar methodology, to ensure the market is receiving the quantities of regulation it is paying for. The method employed by PJM appears reasonable for use in the NEM.

[11] Issues Paper, p.13.

11. ***What are the limitations to your plant's ability to perform accurately and in a timely manner to AGC signals for regulation services? And if so how can these be overcome?***

N/A

12. ***Do you agree with the principles for contingency verification and if not what principles should apply?***

We agree with the general principles in section 3.3.2. A re-wording into plainer English would certainly be valuable. We do, however, have some reservations:

1) An Aggregated IL unit may not be able to restore to pre-contingent levels in a smooth ramp, nor on a precise timeline (see response to question #7).

2) As mentioned in our response to question #7, service requirements should be as loose as can be made to work. Imposing limits on over-delivery would preclude participation by some types of load. The tighter the limits, the more loads would be precluded, and the greater the scale advantage to the largest aggregator. If it is actually necessary to do this to avoid causing over-frequency problems, so be it. But analysis of New Zealand's system found that it was not.

3) In the unique situation where an aggregated IL unit is enabled for Fast Raise and Delayed Raise but NOT Slow Raise, it would be impossible for the unit to avoid delivering a Slow Raise quantity. This is because IL cannot ramp up and down within a 10 minute period the way a generating unit might be able to. If enabled in this configuration, a generator might ramp up from 0-6 seconds, ramp back down from 6-60 seconds, then (if the frequency remains below provider's allocated *frequency deviation setting)* ramp back up from 60-300 seconds. IL cannot do this, based on the way loads are restored in manual fashion. If there is a way for a participant to submit offers such that they will never find their unit enabled in such configuration, we'd be glad for AEMO to provide such guidance.

Detailed comments on the use of the 'reference frequency trace':
The utilisation of a 'reference frequency trace' is necessary for suppliers to be able to determine whether their delivered quantities were compliant with their enabled quantities. However, this is only so because of the way today's MASS determines the start time of an excursion. Under today's MASS, the 'clock starts' counting to 6 seconds as soon as the frequency leaves the NOFB (i.e. 49.85 Hz), rather than when the frequency reaches an enabled provider's assigned *frequency deviation setting*. The effect of this is that, depending on the RoCoF of the excursion, the amount of time an enabled provider has to deliver their response (within the 6 second window) will vary. In order to assess whether the delivered quantity is compliant with the enabled quantity, it is thus necessary to prepare offers and calculate compliance against a standard scenario. The standard scenario described in today's MASS is a reference frequency trace with a linear decline of 0.125 Hz/sec, where T=0 is at 49.85 Hz.

The effect of this methodology is that two FCAS suppliers with identical physical response capabilities, but different *frequency deviation settings* will have different timeframes within which to deliver their Fast Raise service, and will thus have to submit different offers. Providers with lower *frequency deviation settings* will have to make smaller offers than identical providers with higher settings. Figure 2 below details the range of possible scenarios for Fast Raise.

Figure 2: How the assigned *frequency deviation setting* will determine offered quantities

Provider's allocated *frequency deviation setting*	*Setting*'s distance below NOFB (49.85 Hz, when the 'clock starts')	Using 'reference frequency trace' with ramp of 0.125 Hz/sec, how many seconds should provider expect frequency to take to travel between NOFB and *setting?*	Total # seconds provider should plan to be able to capture in the Fast Raise window (out of a max of 6 sec)	Example: A unit with a switching controller can increase output by 10 MW over 6 seconds, following a linear ramp. What quantity of Fast Raise should it offer?
Units: Hz	Units: Hz	Units: Sec	Units: Sec	Units: MW
49.80	0.05	0.40	5.60	9.33
49.75	0.10	0.80	5.20	8.67
49.70	0.15	1.20	4.80	8.00
49.65	0.20	1.60	4.40	7.33
49.60	0.25	2.00	4.00	6.67

In terms of simplifying contingency FCAS verification in the MASS, the simplest change AEMO could make would be to make the 'clock start' when the frequency reaches each provider's allocated *frequency deviation setting,* rather than when the frequency leaves the NOFB. This would greatly reduce complexity, remove the need to employ a reference frequency trace, and ensure that all suppliers are able to create the same apples-to-apples offers reflective of their plant's physical contributions over 6 seconds.

13. What amendments are required to the FCASVT to better represent the performance of your plant?

The FCASVT is a useful tool for participants and we are glad that AEMO intends to continue to maintain it. The FCASVT is designed to model generation increases and decreases, which means that IL data must be inverted to negative values prior to loading. While this isn't the best visual representation of IL plant, it's a minor cosmetic inconvenience and needn't require re-design of the FCASVT.

There is issue related to the FCASVT that could be improved: at present, the MASS is silent on how providers of aggregated services should time-align and combine the data from their multiple high-speed recorders, prior to loading them into the FCASVT.[12] One complication is that the clocks on the many distributed high-speed recorders are unlikely to be synchronised sufficiently accurately

[12] We don't think it would be worthwhile to try to extend the FCASVT to accept data from each element in an aggregated resource, as large aggregations produce more data than can sensibly be handled in a spreadsheet.

to be directly comparable.[13] EnerNOC's recommendation is that the MASS instruct participants to time-align each meter's logged recordings to the time the frequency excursion was detected[14] and then sum the MW figures prior to loading them into the FCASVT.

14. How could the response of a large number of small scale installations, such as batteries at households, be verified in response to a local frequency disturbance?

These types of sources should be able to be verified using the principles specified in 3.3.2 of the Issues Paper, provided the same requirement for high-speed data recordings is applied to small scale installations, and the MASS provides guidance as to how to time-align individual traces before loading the aggregated dataset into the FCASVT.

15. What barriers exist to aggregated generation or loads with switching controllers being configured to provide a staggered response rather than have all units with the same settings and how can these be overcome?

In principle, AEMO's proposed approach of staggering *frequency deviation settings* within a single aggregated dispatchable unit and using settings beyond today's 0.05 Hz step changes is philosophically reasonable and technologically feasible.

The challenge posed by the proposed framework would be in verification. At present, verification is straightforward: if the frequency deviation setting is reached, the unit should deliver its enabled response; otherwise it shouldn't.

Under the proposed framework, where a dispatchable unit has a range of frequency deviation settings, AEMO will not always know how much response to expect from the unit:

- If the frequency doesn't fall as far as the first frequency deviation setting, then no response is expected.
- If the frequency falls as far as the last frequency deviation setting, then the full enabled response is expected.
- If the frequency falls to somewhere in between, then some intermediate level of response is expected.

Distributing the settings would make the size of the response to intermediate frequencies progressive and roughly proportional to frequency. However, when preparing and submitting its offers, the aggregator will be monitoring and forecasting the total amount of response available from the all the resources in the dispatchable unit, without regard to how the response is split up between sources with different frequency deviation settings.

---

[13] In principle, they could be synchronised to each other very accurately, by using GPS clocks. However, this would be prohibitively costly and greatly complicate installations. The approach taken in other markets is to have time-of-day clocks that are only as accurate as in typical meters, but use the high-resolution frequency measurements to achieve accurate alignment.

[14] Aligning either on the first 50 ms frequency measurement that falls outside the NOFB, or on the first that falls at or below the participant's *frequency deviation setting.*

As a result, the distribution of the response between frequency deviation settings will vary, so in a single 'intermediate excursion' (in which the frequency falls to an intermediate value within the unit's range), it will not be possible to verify that the response is of the right size, because the correct size is undefined: it's just somewhere between zero and the full enabled quantity. This is not necessarily a problem – it is just something of a departure from previous practice: traditionally, delivering a quantity less than enabled quantity would indicate non-compliance with a dispatch instruction, but in this situation it would be the desired outcome. Appendix C provides an illustration of this challenge.

An alternative would be to require each aggregator to have one dispatchable unit per frequency deviation setting per region, rather than just one unit per region: that way, AEMO would know exactly which resources should respond to each event, given the lowest frequency reached, making verification straightforward. However, we advise very strongly against this, as it would undermine the principles of aggregation by switching the aggregation hierarchy from the 'dispatchable unit' level to the 'frequency deviation setting' level.

An aggregator achieves reliable response by aggregating across many dissimilar customers. Smaller aggregations are less predictable and less reliable. In addition, the administrative complexity of monitoring and forecasting performance and making appropriate offers is incurred on a per-aggregation basis. Managing a dozen aggregations requires a dozen times as much effort and overhead costs as managing one aggregation.

Moving from each MASP having one dispatchable unit per region to each MASP having one dispatchable unit per frequency setting per region would greatly reduce the accuracy of offers and increase costs, and may well render participation infeasible for many prospective new entrants.

Another verification option would be to assess compliance only when a unit's complete range of settings was reached during an excursion. This preserves the principle of aggregation at the 'dispatchable unit level' and would pose no problems for aggregators, but may be unsatisfactory to AEMO in that AEMO may not be able to assess compliance in all instances – only following 'large' excursions that reach at least 49.6 Hz and thus trigger the unit's entire range of allocated frequency deviation settings.

How 'staggered response' might be implemented:
One method AEMO could employ to achieve the 'staggered response' is to rely on each aggregator to distribute their constituent sources across an assigned range of frequencies, attempting to achieve some sort of "even" distribution across the range.

Upsides to this method:

- It could be relatively easily accommodated by aggregators, and would not undermine the use of aggregation.
- It allows aggregators some flexibility to assign their most tolerant sources higher frequency settings, and their most sensitive sources lower settings, or to rotate sources between frequency settings to share the pain more evenly.
- It achieves AEMO's desired effect of providing a roughly proportional response.

Downsides to this method:

- It would be difficult to verify performance in ‘intermediate excursions’ whose lowest frequencies fall within the range of trigger thresholds: problems with the frequency allocation process might only become apparent after multiple events.
- An aggregator’s portfolio might be “chunky” in that it includes some constituent sources that contribute large quantities, and others that contribute small quantities, and thus “evenly” distributing constituent sources across a range of frequencies may be impossible.
- The constituent sources within an aggregator’s offer will change throughout the day, as some sources come in/out of the aggregate offer, and/or individual source’s available quantities fluctuate. The effect would be that the quantities assigned to each setting vary considerably throughout the day. Appendix D illustrates this challenge.

EnerNOC is in favour of implementing this method or something similar to it, rather than a blanket prohibition on the use of switching controllers, provided the method gives AEMO requisite confidence and certainty.

16. ***What limits exist in switching controllers on potential range of frequency settings and can this be adjusted?***

See comments in #15

17. ***Do you agree with the proposed principles for the allocation of switching controller frequency settings? If not what principles should apply?***

See comments in #15

Thank you for the opportunity to contribute to the Amendment of the Market Ancillary Service Specification (MASS). Please do not hesitate to contact me if you have any queries.

Regards,

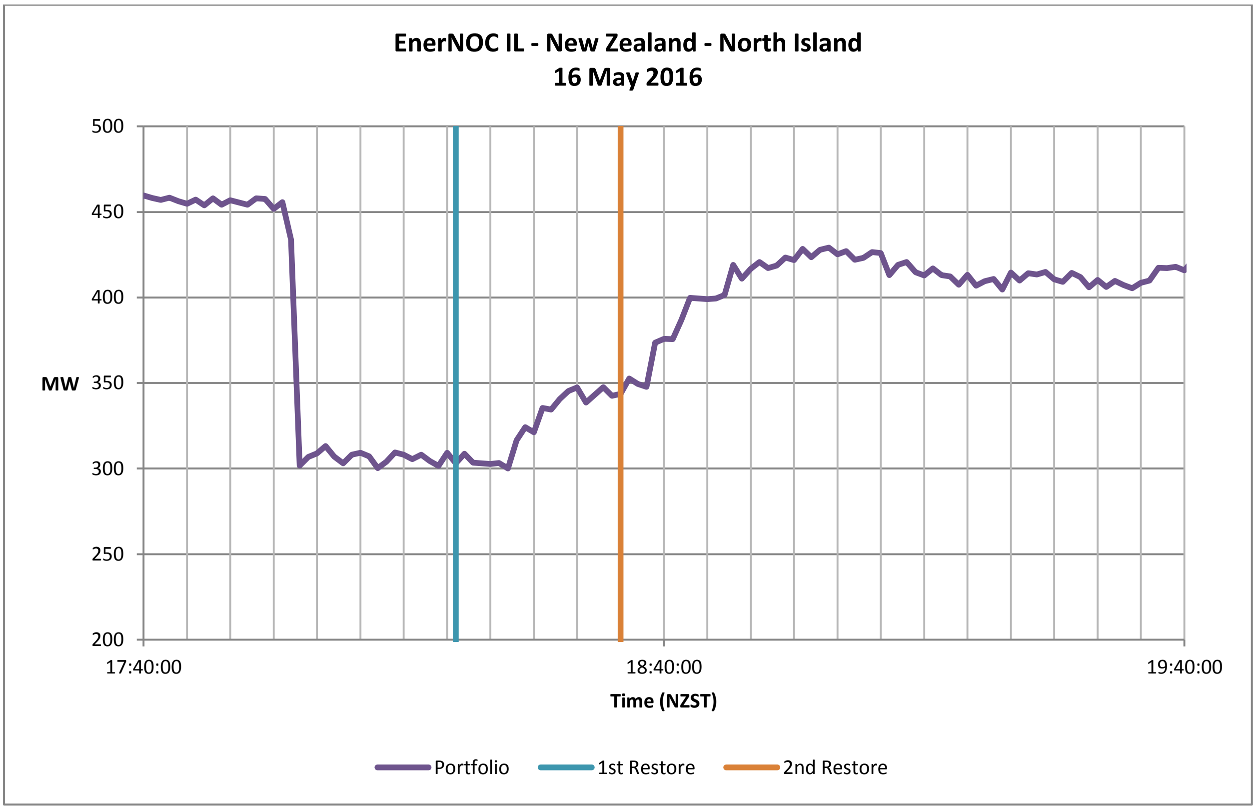
Matt Grover
Manager, Market Development
mgrover@enernoc.com | 03 8643 5907

Appendix A: Indicative IL restore processes, and related timings.

Example #	Plant/Process Type	IL Restore Process: Description	IL Restore Process: Approx. Duration
1	Melters (steel/glass)	Check input systems, check process bath conditions. If bath of molten material has cooled start up may be gradual. This typically means a "soak period" with ~10-20% power input for 1-3 hours to heat material to temperature before restoring power fully.	5-10 minutes
2	Pulp Milling	Check upstream feed systems, check filter screens for blockages, check load in effluent oxidation systems, attempt start of refiners/grinders. When operation is stable ramp from 20% to 100% as process allows.	10-30 minutes
3	Air Compression	Checks of safety systems and start-up of large electrical motors.	30-60 minutes

Appendix B: Restoring an IL unit post-contingency

This chart shows EnerNOC's Fast Raise IL portfolio response to an under-frequency excursion in New Zealand on 16 May 2016. The excursion was caused by the trip of a CCGT, and the frequency reached 49.18 Hz.



Notes & Takeaways:

- 'Portfolio' plot uses 1-minute interval data. Gridlines are drawn at 5-minute intervals.
- Approximately 150 MW of Fast Raise (equivalent) was delivered
- In New Zealand, IL must continue to respond until given clearance by the System Operator. Once the frequency is stable, the System Operator coordinates with IL providers to bring loads back in stages, in tandem with the re-dispatch of supply-side resources. This contrasts with the NEM, where the MASS indicates that FCAS providers are always free to restore after 10 minutes.
- In this instance, EnerNOC received clearance to restore one block of load at T+20min (blue plot). EnerNOC received clearance to restore the remainder of load at T+37 minutes (orange plot).
- The first loads began returning to service approximately 7 minutes after initial clearance, and the IL portfolio had returned to nearly pre-contingent load levels approximately 35 minutes after initial clearance.
- The restoration process follows a roughly linear ramp over 35 minutes, at an average of approximately 3.75 MW / Min.

Appendix C: Effects of staggering frequency settings within an aggregated dispatchable unit (DUID)

	Scenario 1 - Today's MASS **1 DUID, with 1 *frequency deviation setting*** **DUID's offer fully enabled by AEMO**			
	Source #	Aggregator's Expected MW	Actual Delivered MW	Frequency Setting (Hz)
	1	1.0	0.75	49.80
	2	1.0	0.75	49.80
	3	1.0	0.75	49.80
	4	1.0	1.00	49.80
	5	1.0	1.00	49.80
	6	1.0	1.00	49.80
	7	1.0	1.00	49.80
	8	1.0	1.25	49.80
	9	1.0	1.25	49.80
	10	1.0	1.25	49.80
DUID's offer to AEMO		**10.0**		
DUID's delivered MW			10.0	
DUID's compliance with AEMO's DI			100%	

	Scenario 2 - Future MASS **1 DUID, with range of *frequency deviation settings*** **DUID's offer fully enabled by AEMO**			
	Source #	Aggregator's Expected MW	Actual Delivered MW	Frequency Setting (Hz)
	1	1.0	0.75	49.80
	2	1.0	0.75	49.78
	3	1.0	0.75	49.76
	4	1.0	1.00	49.74
	5	1.0	1.00	49.72
	6	1.0	1.00	49.70
	7	1.0	1.00	49.68
	8	1.0	1.25	49.66
	9	1.0	1.25	49.64
	10	1.0	1.25	49.62
DUID's offer to AEMO		**10.0**		
A - If excursion to 49.75 Hz				
Aggregator's expected MW		3.0		
DUID's delivered MW			2.3	
DUID's compliance with aggregator expectation			75%	
DUID's compliance with AEMO's DI			23%	
B - If excursion to 49.65 Hz				
Aggregator's expected MW		8.0		
DUID's delivered MW			7.5	
DUID's compliance with aggregator expectation			94%	
DUID's compliance with AEMO's DI			80%	
C - If excursion to 49.6 Hz				
Aggregator's expected MW		10.0		
DUID's delivered MW			10.0	
DUID's compliance with aggregator expectation			100%	
DUID's compliance with AEMO's DI			100%	

Appendix D: Example of complexity involved with frequency setting allocation under 'staggered response' approach.

In this example, an aggregator has 1 dispatchable unit (DUID) comprised of 10 individual sources. The max enablement capability of the dispatchable unit is 25.0 MW[15]. The aggregator has "evenly" allocated the individual sources across five frequency settings, targeting 5.0 MW in each setting. However, throughout the day (trading intervals A-B-C) the quantities available at each setting vary, and the relative distribution across the range varies from the target of 20%.

			Actual MW Available				
Source #	**Max Cap (on file with AEMO)**	**Freq Setting**	**Trading interval A**		**Trading interval B**		**Trading interval C**
1	1.0	Setting 1	-		0.8		-
2	1.0	Setting 3	0.7		0.9		0.7
3	1.0	Setting 4	1.0		-		0.9
4	2.0	Setting 1	1.9		2.0		-
5	2.0	Setting 1	2.0		1.9		-
6	2.0	Setting 2	-		2.0		2.0
7	3.0	Setting 2	2.5		3.0		3.0
8	4.0	Setting 3	4.0		2.0		3.8
9	4.0	Setting 4	3.2		3.2		2.9
10	5.0	Setting 5	4.7		3.8		5.0
DUID Max Cap	**25.0**						
DUID Offer Quantity			**20.0**		**19.6**		**18.3**

Frequency Setting	**MW**	**%**	**MW**	**%**	**MW**	**%**	**MW**	**%**
Setting 1	5.0	20%	3.9	20%	4.7	24%	-	0%
Setting 2	5.0	20%	2.5	13%	5.0	26%	5.0	27%
Setting 3	5.0	20%	4.7	24%	2.9	15%	4.5	25%
Setting 4	5.0	20%	4.2	21%	3.2	16%	3.8	21%
Setting 5	5.0	20%	4.7	24%	3.8	19%	5.0	27%
Sum	**25.0**		**20.0**		**19.6**		**18.3**	

[15] EnerNOC's understanding is that AEMO plans to implement some sort of registration process whereby each time an aggregator adds a new source to an existing aggregated dispatchable unit, the new source is documented and/or approved, and the standing data of the disputable unit (it's breakpoints and upper enablement limit in its FCAS trapezium) is revised to reflect the new, larger quantity available.