CRM Design

Discussion paper

Dieter Jong
EIYA CONSULT SPRL | CHEMIN DE LA TRAVERSE 10, 1320 NODEBAIS

Table of Contents

1 Purpose of the document

The purpose of this document is to provide a profound discussion paper on the design of a capacity remuneration market in the context of Belgium. While it is impossible to discuss every aspect of such a CRM in the energy market, it aims at providing an initial set of thoughts and considerations to be taken into account. It is an open invitation to other authors to correct, add, delete, contribute,… in any way possible to converge to an optimal understanding and design of the envisaged Belgian CRM.

Opinions in this paper are personal and should not be used outside the framework of this discussion document. This document does not pretend to be a final position of any of the authors, but it is merely a snapshot in an ongoing train of thoughts regarding the subject.

Any party or author that wishes to use parts of the document to support his public opinion shall be free to do so, but shall not refer to other authors nor quote those participating to this exercise.

2 Executive summary of the design principles

- Penalties for unavailability should be capped to the annual auction revenue of a unit/bid
- Portfolio bids should have an incentive to remain as small as possible
- A paid-as-cleared mechanism is necessary
- Existing capacity is obliged to offer at 0 euro/MW
- Shutdown announcement for any generation asset is to be provided at least 1 year in advance to ensure capacity firmness
- Shutdown announcements are to lead to a non-reversable process of closure to avoid manipulation of the offer curve
- Demand should appear in the demand curve for capacity, not in the offer curve
- The demand curve should be build up bottom-up taking into account the consumers active in the grid both for their maximum price and their volume contribution. They should be able to make multiple bids in terms of volume and price (stacking).
- Only for those consumers who do not explicitly make a bid, an implicit bid is to be made based on the VOLL.
- The VOLL should be sufficiently high.
- YMR and MMR volumes are to be based on their maximum grid capacity, de-rated using an historically based peak synchronicity factor.
- Consumers are to be technically limited to the maximum peak demand they are registered for in the CRM in times of scarcity.
- Reliability options are to be avoided. If applied, only 1 strike price should be applicable, related to the most expensive marginal production cost, and defined ex-ante of the CRM auction.

- Possible uncertainty on the strike price should lead to a higher strike price, rather than a lower one.
- No reliability options on demand, but a penalty scheme to oblige their power limitation in a moment of scarcity to the cleared capacity.
- Penalties on the demand side for overconsumption during moments of scarcity should mount up to 1000 euro/kW.
- The demand in the CRM has to equal the total electricity consumption of Belgium. Not only the Elia grid load.
- The scarcity indicator should be the real time available generation margin.

3 Design should include a phase-out pathway

A CRM is an alternative way of organizing an energy market, as there are multiple other possibilities and hybrid solutions. As we do not know what the future will bring, we have to make sure that the design of any market model is future proof to the best of our abilities. But equally important, is that we should consider that whatever we put into place, it might not be desirable in the future anymore. An important feature of any market is that the design should take into account the potential phase-out of the market model itself. Especially with respect to the CRM in Belgium, which will be granted a temporary authorisation for maximum 10 years from Europe.

Launching markets is a lot more easy than terminating them. As a lot of investors and traders will take positions in a market, inevitably, any shut down scenario is going to 'steal' assets or expected revenue from some players. A practical example in a CRM is an industrial investing in his capability to shut down temporarily his processes. If the CRM stops auctioning at some point in time, he looses the return on his investment. One can give many more examples of this loss of revenue, which is not purely linked to the loss of the auction revenue, but also on the effect of the rest of the market functioning including the loss of expectations.

The easiest way of shutting down a market scheme is to terminate it when the value of the market drops to nearly zero. In such a scenario, there is no longer any direct loss or opportunity loss from participants or stakeholders. As the market value is defined as the product of the price and the volume traded, this can be achieved by either the volume or the price of the market that drops to zero. As such, the design of the market should take this possibility into account, and should facilitate it. In a market wide CRM scheme as envisaged in Belgium, the volume of the market can only drop to 0 if society would stop using electricity. As this is unlikely to happen in the near future (next 15 years), the focus should be on the price formation.

In order for a price to drop to 0 in an auction you should have an offer curve that starts at 0. So the design of the market should ensure that there are offers that have to bid in at 0 euro. Why would a producer place a bid at 0 euro? If his marginal cost to provide the capacity is 0, then logically, he would be bidding at 0 euro if he gets paid the market clearing price. This is the economic theory about marginal clearing prices in auction designs in general. But the above sentence contains a couple of important hypothesis:

1. The marginal cost question: a CRM obliges a producer to be available when a period of scarcity presents itself. As you can expect energy prices to be rather high in such a time, the producer has a natural interest to comply with this request. As such, his marginal cost would be 0.
   But if the CRM imposes any penalties on the producer to be there, there is a small chance that he will be in maintenance or technically unavailable at the time of need, and his marginal cost would then become the risk that he has to pay a penalty. As such, a penalty would prevent a 0 marginal cost offer if the penalty is not linked to the return of an auction. The latter has been the case with many ancillary service products in the energy market for instance. If the maximum penalty is equal to the annual revenue generated by the auction, the producer doesn't risk to loose anything more than the opportunity loss coming from the auction itself. As such, with such a penalty scheme, he would still have a 0 marginal cost to bid in.
   But a CRM cannot really function without minimum penalties, as this would introduce an uncertainty on the security of supply that it tries to achieve. If the market would clear at 0, this would reduce the maximum penalty to 0, meaning that no producer has any interest to respect the received capacity obligation. One could argue that if the market clears at 0, that there is an abundancy of supply anyway to meet the demand. But still, the theoretical risk remains, and becomes all the more relevant in a small market with a high concentration on certain big units. A 1 GW unit in Belgium for instance would represent about 5-10% of the adequacy balance, so its closure might tip the scale significantly even in a 0 price clearing scenario. The more decentralised the market is, the less likely a significant % of the installed capacity closure will occur. This is also related to the envisaged portfolio bidding that is common practice in the Belgian market system. Portfolios should be encouraged to remain as small as possible, specifically for this purpose.
   Additionally, if the CRM interferes with the possibility of the producer to capture any profit in the energy only market, the marginal cost would be quantified as the loss of potential revenue (opportunity cost). This is the case in Belgium, as the reliability options scheme introduces the 'loss' of high spot prices that have to be paid back above the strike price. So one should balance the introduction of reliability options against the 0 marginal cost offer possibility. But reliability options, although merely defined in the high level CRM law today, seem to be the preferred option in Belgium. Therefore, we can conclude that in the current scheme of the BE envisaged CRM, it is unlikely that there will be a natural offer curve starting at 0 euro. The only option to achieve this part of the curve, is to oblige producers to bid in at 0 euro.
2. The market clearing price: a producer will only bid his marginal costs, if he assumes that any market price that he shall receive and that comes out of the auction has to be at least equal to his bid. In a paid as bid scheme, the producer will always include

a (positive) margin that he hopes to make if he is selected. So naturally, he won't be offering at 0 euro. Therefore, a paid as cleared is a necessity for an offer curve to start at 0.

How do you oblige a producer to bid in at 0 euro from a legal perspective without nationalising the asset and without imposing a 'cost' or a 'risk' on the producer that is unfair? In most CRMs, they obtain this result by obliging existing capacity to bid in at 0 euro, under the condition that a paid-as-cleared scheme is applied. As such, they are assured to receive the same contribution as the winning bid. The winning bid can either be set by a new project, which means they'll get a contribution that is equal to what a new project would need to get to ensure financial profitability. If the producing asset is so inefficient that it cannot survive with this support in the market, there is a risk that the asset will close down to limit his costs. But as it closes down, it creates an additional shortage in the market which might lack this peak capacity. There are a couple of solutions to this problem.

Either, a solution is to do a multi-round auction process. If the prices are too low, the production plants that are no longer sustainable, will have to announce their closure. In a 2nd or 3rd fixing, the market cancels them from the offer curve, and they fix the price again. Off course, the call to withdraw the plant has to be followed by an actual termination. The unit cannot remain in the market, as it is no longer profitable, to prevent the withdrawal of production capacity in the offer curve just to boost prices (which would be market manipulation). Forcing a shut down is legally difficult, but without this condition, it is unlikely that no manipulation will occur.

Another solution is to allow existing capacity to indicate a price, but these bids are not taken into account to set the price during the auction process. It is to some extent, an automated multi-round fixing in 1 cycle. With the same conditions that would apply.

Another possibility is to store the closed plants in a strategic reserve until the next CRM auction, which means society will take over the fixed costs of keeping the plant operational for a maximum of 1 year. Society would prefer this option, which is paying the fixed running costs of 1 or a couple of assets, instead of withdrawing them from the offer curve and pay a higher clearing price to all capacity (marketwide CRM scheme).

A particular situation is applicable to Belgium, where closures of major plants have to be announced more than 1 year ahead. As such, they cannot close during the envisaged period covered by the auction. It is however questionable how this criterium will be used towards all participating technologies that are not major centralised production units, as they don't have the same shut down procedure. The design of the CRM should take into account for all technologies that will participate how to tackle this issue. An extension of the procedure currently applied to the big power plants is most likely not realistic on other technologies, and but perhaps desirable. Note that, we are talking about the situation in which there is over-supply and a risk mitigation to prevent massive closures, which are not envisaged anyway in a decentralised market from many assets at the same time. A distinction is to be made between demand and generation technologies. Demand that leaves the market, by direct consequence, respects its obligations under the CRM. So the problem does not exist.

4 Difference between generation and demand

4.1 Demand or offer curve participation?

With regards to the distinction between generation and demand, the question arises whether demand should participate in the offer curve or in the demand curve. The difference has a rather big impact on multiple parts of the design of the CRM.

If demand has to participate in the offer curve, than the first question that surfaces is how the demand curve will take this into account. Because the main argument for this strategy is to simplify the calculation of the peak demand curve down to a top down analysis of the peak demand that was either historically seen or modelled (and a model should be verified by the actual results on a regular basis). If the demand side was historically not participating to the peak demand in the moments with the lowest remaining available margin in the system, then it is not taken into account in the demand curve in such an approach. Therefore, it should not be capable of offering any demand reduction in the offer curve, as it would not reduce the peak capacity. On the other hand, if accidently they would consume during the peak, because they don't have any obligation to avoid these moments, they could pose a problem to the system health.

If the demand was historically participating to the peak consumption, then it can lower the peak by reducing its own contribution. The same reasoning applies: by how much can it really reduce the peak demand, and by how much is there a liability that if a part is left without any obligation, that it might be there at the exact wrong moment in time.

An average way of taking this into account is to use some kind of derating factor that would vary between 100% if the demand is always fully activated in times of scarcity, or 0% if the demand is always switched off at those times. This allows some kind of alignment between the demand curve and how it reflects this capacity, and the active reduction the demand side could engage itself to.

But the approach of a de-rating factor is a kind of averaging methodology. Moments of scarcity are rare historically, and therefore most designs rely on modelling the grid (not enough relevant data points on an identical 'society' framework). But models are build on historical data and situations also, and therefore, again, apply a kind of averaging methodology. The model will detect 'on average' how the demand was present in modelled moments of scarcity. But then there is the statistician who drowned crossing a stream with an average depth of six inches.

Although a joke, it demonstrates perfectly the problem that we encounter while using derating factors to cope with the demand potential. The lower the derating factor, the less the demand will receive an obligation to remain idle during a moment of scarcity. Which in return increases the risk that it might be activated at the wrong moment.

One could argue that the derating factor can only be used to determine the volume that the demand side can offer, while not determining the capacity obligation that is applied to the site. In practice, this means that a 10 MW peak consumer, with a derating factor of 70%, will only be allowed to offer 7 MW of volume in the offer curve. So he will only be rewarded for 7 MW, instead of 10. But he might consume, and therefore will pay, for 10 MW. This would introduce a structural overpayment of capacity, without the possibility for a consumer to recover his costs even if he disconnects fully from the grid. If applied market wide, this means consumers will pay for the theoretical peak they can get, if they would simultaneously spike. Which in practice will not occur, and therefore all consumers will receive a 'synchronicity' discount from the overfinancing of the system. Although this will solve the issue related to the transfer from consumers to producers in general, it will create distortions on the local level. Because demand that always spikes when the scarcity occurs will get 'free capacity' due to this effect, and vice versa for the demand that usually disconnects anyway at those moments.

Derating factors are usually applied to model the unforeseen technical failures that can occur. Like the N-1 design criterium that is fairly well known to electricity grid design modelling. But not to 'average out' the potential, for which the above shows that it would not be a very robust method. Note that on the long run, if demand has always shut down in moments of scarcity, due to the existence of a capacity market, it raises the question on how the derating factor will evolve over time. Because as the demand is never present during moments of scarcity, any model will eventually evolve to a derating factor of 0%, whereas once the demand is no longer incentivized to respects its obligation, it might popup for 100% at the moment of scarcity. It creates a vicious circle over the long term.

If derating factors are applied on every individual site, to take into account their participation to the synchronous demand peak, then one can ask himself the question: why don't we take this information into account in the demand curve. And if we do, then why raise the demand curve by a demand calculation of what could happen at any price, only to see the counter weight of that volume in the offer curve at a limited price? Wouldn't it be more elegant then to take the volume into account in the demand curve at a limited price right from the start and leave the offer curve unchanged?

If demand can participate in the demand curve, it requires a bottom up approach to determine the peak demand instead of a top down approach. This complicates the calculation of the demand side modelling, but as is shown above, this is inevitable anyway if the correct incentives are to be provided.

A bottom up approach would require a close cooperation between the transmission system operator and the distribution grid operators to analyse the contribution of every single consumer to the peak demand. For most quarter-hourly metered volumes, this comes down to a correlation analysis of individual time series compared to the national peak demand.

Although it concerns a lot of data, it is perfectly feasible on a scale like Belgium, as this kind of analysis have been performed already with respect to grid fee simulations recently in Flanders for instance by Fluvius.

For annual or monthly metered volumes, this contribution is not measured, and therefore requires a calculation method. Typically, low voltage networks are designed while taking into account a synchronicity factor of 20%. So the grid operator assumes when designing the grid, that a 10 kW connection would be able to supply 5 10 kW consumers. That doesn't mean that 1 consumer does not peak at 10 kW at some point in time. But that if he does so at the moment of scarcity, that 4 other consumers are most likely not going to consume anything at that moment. As the connection limit of a cable imposes a hard limit on the maximum power consumption that can be supplied through it, this is to some extent similar to the design issue the CRM faces. As such the 20% could be a good indication. Applied to the low voltage network, this means that every low voltage consumer participates to the CRM with about 20% of his maximum connection capacity.

But the 20% can also be calculated and verified historically to avoid any over- or underdesign issues from the distribution grid planning to be introduced into the CRM design. If we consider the historic peak demand occurrences, and we subtract the quarter-hourly time series from it, we remain with the low voltage 'unmetered' peak demand. This result divided by the connection capacity of the underlying 'unmetered' connection points will result in a percentage that could be applied instead of the 20% if needed. In conclusion, even unmetered volumes like YMR and MMR can be taken into account in a bottom up approach of the peak demand determination. Note that digital meters will facilitate the process and limit the uncertainty even further in the future.

If we are capable of defining the individual volume contribution of a consuming connection point, then the question remains whether it would be feasible to collect a price indication from each of these connection points to establish a price based demand curve that could replace the top down analysis. If all consumers participate by contributing their maximum price that they are willing to pay for their synchronous peak demand, then the demand curve can be constructed as such. It would result in an ideal situation in which no consumer would pay more than needed to secure adequate capacity for his needs. If the price of the auction would go too high, consumers would have chosen themselves to be limited or even disconnected in those moments of scarcity as they would consider the cost too high for the resulting benefits from it.

But it is highly unlikely that all consumers will explicitly state a price, and therefore the question remains what to do for those who don't. Actually, that question is the same as in the top down analysis. The top down analysis assumes that all consumers have a certain value of loss load, their willingness to pay for the availability of power. Therefore, we have already set the price for those consumption points that do not explicitly want to specify a price: they agree to the VOLL. The only design feature that should be respected is that the VOLL should ideally be higher than the actual willingness to pay, and as such does not create a free option for those consumers who would have a higher willingness to pay but are capped by the VOLL. A high VOLL would also create an incentive to consumers to become active in their explicit bidding.

The top down demand curve analysis takes a secondary objective into account: the willingness to pay is also depending on the frequency of the loss of load. This is expressed through the LOLE (loss of load expectation). Indeed, it is fair to say that a consumer is willing to pay more for the final kW that powers his essential applications then he is willing to pay for the optional kW to power his nice-to-have maybe applications. In order to cope with that characteristic, a consumer should even be allowed to express multiple price thresholds for different parts of his peak power consumption. And for those consumers not explicitly stating a price on a part of their peak power needs, the VOLL would apply on those kW in the demand curve.

It would solve the derating factor issue, as the consumer will only be paying for the capacity that he asked for, and will be technically limited to that capacity in moments of scarcity.

4.2 Reliability options for design

It would also solve the reliability option problem for the demand side. Because a reliability option imposes on a generator that he pays back the spike prices above a certain threshold. The reasoning behind it, is that a consumer already paid for the availability of sufficient capacity, and therefore, a market price well above the marginal generation costs should in theory be impossible, as sufficient availability of competition would cap the market to the marginal cost of the last available generation asset. Therefore, any market price above the highest marginal cost of each generation unit participating to the CRM should be impossible, or relates to some kind of wind fall profit that should flow back to the consumer via a reliability option scheme. It is called a wind fall profit, as in an energy only market, the peak generator would have to recover his fixed and capital costs through the occurrences of price spikes in times of scarcity. But as the CRM has the goal of eliminating structural scarcity, these price spikes would be less high and less likely. If they however do occur, due to international import and export trade for instance, the consumer risks to pay 2 times: once for the CRM, and once for the price spikes that shouldn't have occurred.

Note that in this point of view, it would be unrealistic to put various strike prices out there. There is only 1 marginal cost price that would be the highest of the market wide CRM supported fleet, and that one is the highest possible market price without wind fall profits. All lower prices that could occur in the market, would have occurred anyway, with or without a CRM, and are therefore no windfall profit for the generators. Simply put, if you are able to construct a generator with a marginal cost of 1 euro/MWh, you will construct it anyway (with or without a CRM), because your business model is expecting higher prices on a significant amount of hours which gives you the return on investment. The introduction of a CRM would provide you with an additional revenue stream, but it would also, in an isolated market, cap the market prices in times of scarcity to the marginal cost of the unit necessary to supply the final MW, which is a loss of opportunity. The windfall profit of a CRM would then not be the difference between the market price and the 1 euro/MWh marginal cost of the generator, but it would be the difference between the expected revenue from price spikes in an unconstrained market and the constrained market via the CRM reliability options. As such, this would be: (expected price spike level – marginal cost of the most expensive unit/the real market price) * the chance/frequency of moments of scarcity that would have triggered the most expensive unit to run. Note that this formula is equal for any participating generator which is selected through the CRM and has a marginal production cost that is below the highest marginal production cost. As such, it would be unfair to introduce different strike price levels for different participating units, as this would create an additional opportunity loss for generators, that they will try to recover in their CRM bid.

As the highest marginal cost would only be known after the clearing of the CRM, but it does constitute an important parameter in the cost recovery bid of a generator (opportunity loss), one could argue that the strike price of the reliability option has to be defined up front based on an estimated marginal cost for the final production unit. If the strike price is below the marginal costs, the participant to the CRM risks that if he has to run during moments of scarcity at a loss as he will have to payback the difference between the market price and the strike price. This would increase the CRM auction price, equal to the risk assessment of the participant as to how many hours he potentially has to run at a loss. It would also create an incentive for the participant to try to avoid being available during moments of scarcity to avoid these running losses, which relates to the potential penalties he will have to pay for non-availability (see above).

If the strike price is above the marginal cost, it would create a windfall profit for the producer in case he has the highest marginal cost, and he therefore would have set the highest possible price in the energy only market with an adequate balance between offer and demand (as is insured by the CRM). Similar to the price increase in the CRM that a low strike price has, the opportunity gains will lower the CRM bid price of a participant in case the strike price would be above his marginal cost. He will assess how many hours he is supposed to run in a worst case scenario, and will subtract these 'revenues' from his bid price in the CRM. Note that he will have an incentive to be available, and as such, there is no relation to the penalty for non-availability.

The CRM and energy only market act as communicating vessels in this respect. But as Europe puts the main focus of the market model on the energy only market, and a design focus to limit the CRM revenues to the bare minimum, one could argue that a higher strike price, above the marginal costs, is most likely the preferred option. But it shouldn't be too high to avoid significant wind fall profits. It is however rather likely that in Belgium, our marginal generator unit in times of scarcity would be a kerosene fuelled turbine, for which the strike price could be rather well estimated up front.

The problem starts when demand is allowed to bid into the offer curve. Demand can have extremely high marginal costs to shut down their once-every-10-years applications. For a pure demand side oriented technology, that actually shuts down demand (shifts demand in time), one could argue that a strike price is unnecessary. As there will be no windfall profit for this unit. They won't 'make money', as they will shut down and 'not buy', as opposed to a generator who would 'sell' in the market. It is a cost avoidance scheme, instead of an occasional revenue surplus.

However, if demand is shifting to other fuel sources than electricity, the demand reduction becomes economically similar to the use of a generation asset on site, behind the meter, to reduce the demand for power in those moments. To ensure a level playing field with the generators, and a fair contribution to the other consumers, a strike price should be applied to the demand side offerings as well in this case. For the exact same reasoning: if they are the most expensive marginal cost to provide the final MW in times of scarcity, than any price above this marginal cost would be an excessive market price which would create a windfall profit for the entire market.

But as these prices could be very high, one could be tempted to solve the issue by applying different strike prices for different offers in the CRM. As said above, this is not in line with the primary objective of a reliability option and would increase the cost and importance of the CRM to the detriment of the energy only market.

If demand is participating however in the demand curve, the problem of the reliability option does not apply. The demand side would have evaluated how much the availability of power would be worth to them, and as of which amount they would consider alternatives (either a shut down, a shift in time, or a fuel shift). This individual evaluation of how many private investment would be needed to enable these activations, would result in their maximum bid price in the CRM auction. If the clearing price of the CRM is below their maximum bid, they would be considered as a consumer during times of scarcity, and will therefore not be subject to any power limitations, but they would pay the CRM cost to those technologies that provide them the opportunity at a lower cost than their own individual cost assessment.

If the clearing price of the CRM is above their maximum bid, they wouldn't have to pay for the CRM, as they are considered 'disconnected' at a moment of scarcity. They should make the investment of being able to shut down/shift, and they should ensure that in times of scarcity, they disappear from the demand side in the market. If they are however active, they should face penalties, to make sure that they don't have an incentive to bid too low in the CRM compared to the necessary investment they have to make. In such a case, a strike price or reliability option is not applicable either.

How high should this demand penalty be to ensure compliance of the demand site with the auction results? The penalty should be at least equal to the peak demand of the site (the actual peak demand, not the requested one in the CRM tender) at the market clearing price of the CRM. If the penalty is lower than this value, it would incentivise the demand site to offer a low volume at a low price. As such, it would either set the price to this low value, or it would in any case lower the total market value of the CRM with its lower volume demand. But in the end, if it would consume more, it would not pay more than the (artificially) low result of the CRM. As this creates a structural shortage in the market (under dimensioning of the CRM) this is not desirable.

If the penalty is too high, this would create a risk for the demand site if they would not clear in the CRM. If they would put their price below the clearing price, they would get an additional obligation compared to the situation in which they would set a demand price above the clearing price. With a potentially high risk on this additional obligation, the demand site would receive an incentive to overprice their demand, in order to avoid the penalty risk. The overpricing would come on top of the necessary investment that would be needed in their individual case to limit their power usage in times of scarcity and will reflect as such the technology risk that the infrastructure would not respond to the activation signal to limit the power usage. But if the infrastructure would not respond, it would introduce a structural shortage risk on a system wide level. As such, one could argue that the penalty should remain sufficiently high at least.

Sufficiently high relates thus to the technology risk of the application needed for the consumer to limit his power consumption at a moment of scarcity. We can assume that the most expensive way of shifting demand to another moment in time, is a battery. Batteries currently can be bought for less than 1000 euro/kW with a 2 hour back-up. As such, the penalty per kW could be 1000 euro. As such, the consumer, in the event he is unable to sufficiently guarantee his power limitation, will have at least the incentive to build a battery in order to avoid the penalty resulting from the CRM.

5 Identifying a moment of scarcity

As is shown above, many of the risks related to the bidding in a CRM are related to the potential cost/opportunity loss times the chance of occurrence. The latter is the amount of moment of scarcity that can be expected in the next year, covered by the CRM. In order to identify this amount, a clear definition of a moment of scarcity is needed.

A moment of scarcity, in respect of the CRM, is a situation in which a structural shortage of generation capacity to meet the maximum demand risks to destabilize the grid. As electricity cannot be stored, an overconsumption on the grid shall start to lower the frequency of the grid, which in turn shall activate safety procedures at the generation units resulting in a potential total black-out. The CRM tries to avoid these moments of structural shortage. It is important to align on this definition, as it clearly means that a CRM, in the Belgian context, is not a price regulation tool.

A moment of structural shortage therefore is the balance between two variables:

1. The demand
2. The generation

The demand can be defined as the total load on the Elia grid, meaning the total demand on all substations at 30 kV or higher. If this definition of the demand is used, it means that the CRM should act accordingly, and disregard the underlying voltage levels, both in terms of generation assets as well as in terms of demand. As such, the total peak demand on the concerned substations should be measured, modelled and covered by generation assets directly connected to this voltage level or above. This would however raise issues as to how to translate the CRM to the final consumer invoice, as the contribution of each substation inevitably will look at the underlying assets. Knowing that a large part of the peak demand is also generated at the level below the 30 kV, it would make no sense to treat this part of the grid as a black box that cannot participate in the auction.

Therefore, demand can be defined as the total electricity consumption in Belgium at any point in time. If defined as such, the CRM should cover the total demand as well in terms of volume, but any grid connected asset, consumer or producer, should be allowed to contribute. As such, any consumer or producer shall contribute to the CRM whatever he or she is willing to pay for their share in the total demand.

Can we measure the total electricity demand? For the moment, we cannot. There is a large part of the low voltage network that is not measured, as has been discussed before. But we do measure in real time the load on the 30 kV and above substations, as is already published on the Elia website in near real time. We could look at the bottom up calculation of the demand on the lower voltage level, and correlate this demand during its highest peaks with the available meter readings on the substations. As such, we can define a contribution factor, that could show us how far we are from the peak demand for which the CRM was designed at any moment in time. A small example: let's say during one of our highest peak demands, the load on the 30 kV network mounted up to 5 GW. Then we know that as long as the load remains below 5 GW, that we are not yet near a moment of scarcity for which the CRM was designed. As such, we could look at the real time metering of the 30 kV infeed, to determine how far, relatively in percentage, we are from 5 GW of consumption. If we reach a 95% threshold, we could define this as a moment of scarcity. Similar, we could use the real time metering of Elia on the higher voltage posts to define the relative distance to a maximum peak demand.

This definition however does not take into account that there might be a lot of demand that is willing to shut down to avoid a scarcity issue in rare occasions. As such, the total demand could be higher than the volume auctioned in the CRM. For instance, if in the future a lot of wind would be available at some point in time, we have a lot more generation capacity available, and as such, demand can be allowed to consume without any limitation. However, if the wind is not available, an even lower level of demand could raise a moment of scarcity. If we would start to shut down demand as of a 97% threshold from the maximum peak demand that was auctioned in the CRM, we might be shutting down demand in cases where there is an abundancy of renewable energy available to cope with the demand.

Other CRM markets look at the available generation margin. This looks at the difference between the available generation and the actual demand. As such, a high availability of generation allows a demand that is well above the volume auctioned in the CRM. But in a moment of scarcity because of low generation availability, demand would have to be reduced to the level that was auctioned in the CRM. The European network codes oblige nearly all generation units to provide an availability schedule to the TSO. As such, the TSO has a clear overview of what the available generation capacity is down to the level of 1 MW units. There is off course some generation capacity below that level, namely solar and potentially small scale wind turbines, but we can assume that those have a similar availability as the utility scale renewables that are scheduled in accordance with the network codes. In practice, moments of scarcity will be moments without a lot of renewable power, so we can assume that all generation units that are not explicitly scheduled to the TSO are not available during moments of scarcity. In conclusion the TSO will have a full overview of the available generation capacity in real time after the application of the network codes.

As the TSO also has a view on the total demand, upscaling it from its substation metering as discussed above, it could calculate the difference between the current level of demand, and the available generation margin left in the system. This would be a dynamic indicator of the potential scarcity in the network.

The main question that remains, is how to tackle the import/export with neighbouring countries. In the design of the CRM, import has to be taken into account on the offer curve. The way the import is taken into account in the design of the CRM, should be reflected in the determination of the real time generation availability margin. As the import/export is measured on a continuous basis, this should be feasible. One could argue that the contribution of the import to the availability margin should be limited to a positive impact. As otherwise, the margin would decline due to an export position of the country. As in the design of the CRM, export will not be taken into account, it seems logical to apply a minimum contribution of the import to the available generation margin of 0 in situations of export.

Elia proposes to use the day-ahead price signal as a scarcity indicator. History has shown that although prices are above average during moments of scarcity, the actual highest price spikes are not related to scarcity at all. A clear example is the price spike on Belpex due to a malfunctioning of the market coupling which lead to an isolated fixing with explicit cross-border schedules. As this was a very rare and unusual situation, prices spiked in Belgium, without any scarcity issue. Most price spikes that mount up to > 1000 euro/MWh are the result of a market failure (bid typo, IT problem,…) rather than a scarcity problem on the market. As such, the day-ahead price is a bad indicator for scarcity, unless these extreme price spikes related to market malfunctioning are filtered out. But filtering them out is rather difficult. Especially in the extremely rare situation where a price spike might accidentally occur at the same moment as a scarcity situation. It will be a rather subjective exercise to determine to what extent each price spike is related to a market malfunctioning and to what extent it is related to a scarcity issue. In order for a day-ahead price to be used as an indicator to trigger the availability of demand and generation units, a clear set of rules should be established. Which most likely will be based on the proposed available generation margin to begin with.

Note as well that the spot prices are set by the European market interactions, and that a spike in Belgium can occur because of a malfunctioning in another country. Which makes an objective assessment from a Belgian perspective even more difficult.