

## IMPACT OF REALISTIC COMMUNICATIONS FOR FAST-ACTING DEMAND SIDE MANAGEMENT

P. Dambrauskas, M.H. Syed, S.M. Blair, J.M. Irvine, I.F. Abdulhadi, G.M. Burt  
University of Strathclyde – UK  
[paulius.dambrauskas@strath.ac.uk](mailto:paulius.dambrauskas@strath.ac.uk)

D.E.M. Bondy  
DTU – Denmark  
[bondy@elektro.dtu.dk](mailto:bondy@elektro.dtu.dk)

### ABSTRACT

*The rising penetration of intermittent energy resources is increasing the need for more diverse electrical energy resources that are able to support ancillary services. Demand side management (DSM) has a significant potential to fulfil this role but several challenges are still impeding the wide-scale integration of DSM. One of the major challenges is ensuring the performance of the networks that enable communications between control centres and the end DSM resources. This paper presents an analysis of all communications networks that typically participate in the activation of DSM, and provides an estimate for the overall latency that these networks incur. The most significant sources of delay from each of the components of the communications network are identified which allows the most critical aspects to be determined. This analysis therefore offers a detailed evaluation of the performance of DSM resources in the scope of providing real-time ancillary services. It is shown that, using available communications technologies, DSM can be used to provide primary frequency support services. In some cases, Neighbourhood Area Networks (NANs) may add significant delay, requiring careful choice of the technologies deployed.*

### INTRODUCTION

Maintaining the frequency of an electrical network is the responsibility of the system operator. The decommissioning of large conventional generators and the rapidly increasing penetration of non-synchronous, highly-intermittent renewable generation within networks is leading to a decrease in system inertia, making the task of maintaining the network frequency even more difficult and expensive. To cope with the intermittency introduced by renewable energy resources and to counter the decline in system inertia, faster acting and cheaper – yet reliable – balancing resources are sought by system operators [1]. The provision of frequency balancing services using demand side resources is an attractive solution. To aid and encourage demand side participation in frequency balancing services, aggregation models are being proposed by system operators. This has helped increase the participation of demand side resources, which is evident from the existence of 22 demand side aggregators within Great Britain [2]. However, the aggregators have a preference of resources greater than 250 kW thereby

rendering a large volume of highly-distributed flexibility available within the premises of domestic customers unutilized. The preference of aggregators towards larger resources is justifiable due to the lack of confidence in the ability of highly-distributed domestic devices to deliver time-critical ancillary services. A number of field trials [3-4] conducted throughout Europe demonstrate that – as a proof of concept – domestic devices, controlled and operated by the principles of transactive energy [5], can be utilized for the provision of time-critical ancillary services. As per the transactive energy principles, an aggregator runs an electronic market, where a price signal issued by the aggregator is used as the control signal for the domestic devices to deliver their flexibility.

Frequency balancing services typically do not allow for participation of aggregators that employ a transactive energy market to control their portfolio of flexible devices. This is because it is unclear to which of the available frequency balancing services the aggregator's flexible resources can contribute. In order to participate in a particular frequency balancing service, an aggregator needs to adhere to the technical specifications of the balancing service, which usually comprise a fixed minimum volume and a time within which the response needs to be provided. With the aggregation model proposed and promoted by the system operators, the accumulation of a minimum volume of reserves is no longer an issue. It is difficult for an aggregator to commit to a certain frequency service provision without assessing the capability of its portfolio to adhere to the response time requirements of the service.

The time within which an aggregator with a portfolio of highly-distributed flexible resources can provide response is the time it takes for the price signal (i.e. the control signal) generated by the aggregator to reach the devices. This time, or latency, is highly dependent upon the communications infrastructure utilized.

Characterizing the performance of communications for its various applications within power systems is a topic of active research. In particular, a thorough comparison of various wireless technologies in terms of their aptness to be utilized for monitoring and controlling distributed renewable energy resources has been presented in [6]. Analysis of the performance of one or more wireless technologies within a wide area network (WAN), neighbourhood area network (NAN) and home area network (HAN) for demand side management (DSM) applications is presented in [7-9]. However, any DSM

application that uses communications will utilize more than one network type and therefore analysis of individual networks separately is not sufficient to determine the actual system latency and thus its ability to reliably provide the stipulated ancillary services.

## COMMUNICATIONS PERFORMANCE

When information is sent from a point A to point B, the information may pass through multiple nodes within various networks before reaching point B. There is a delay associated with each of the nodes as shown in Figure 1. The sum of the various delays at each of the nodes constitutes the total latency of communications between point A and point B.

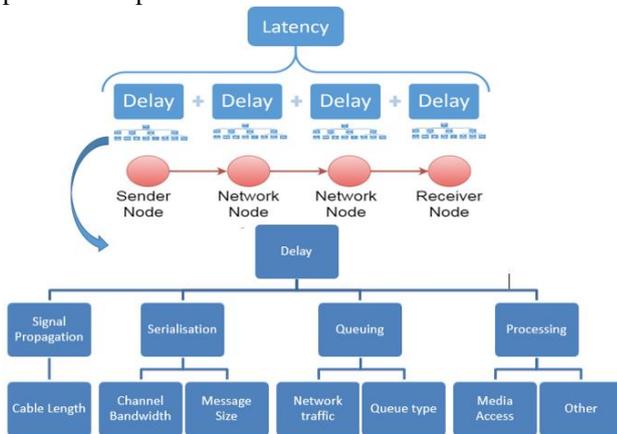


Figure 1: Main sources of delay

This paper assumes that a transactive energy market is deployed by the aggregator to control its portfolio, and characterises the performance of communications for such a system. The relevant types of communications networks involved in real-time control of DSM are:

- Core network,
- Wide area network (WAN),
- Neighbourhood area network (NAN), and
- Home area network (HAN).

The core network connects the transmission system operator (TSO) or distribution network operator (DNO) to the aggregators and the aggregator to the WAN. The WAN passes the information to the NAN as illustrated in Figure 2.

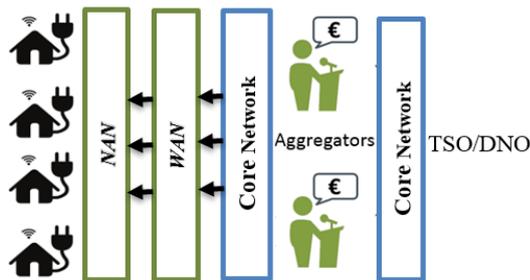


Figure 2: Communications networks for control of DSM resources

The NAN distributes data among individual premises through HANs as shown in Figure 2 and Figure 3. In the

following sections, these networks will be analysed in detail and the associated latency will be characterised.

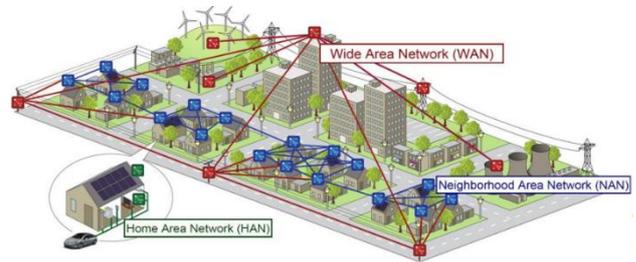


Figure 3: Urban communications network [10]

## Core network communications

A core network is a long-range and high-performance communications network, typically implemented using fibre, which aggregates all non-local communications traffic and facilitates the transfer to other cities, countries and continents. In the context of a power grid, control centres and most of the high voltage substations have a private fibre-based communications system as their core network. Latencies associated with such networks are small and under normal operation and do not go beyond 10 ms within countries the size of the UK or 30 ms across Europe [11].

In an uncongested core network, propagation delay is the largest contributor to the latency, which will be numerically demonstrated later in this section. Propagation delay is the time it takes for a signal to travel from the transmitter to the receiver, and depends only on the distance between them. Other contributors, such as transfer delays, which result from the packet size/line capacity ratio, and processing delays that are inherent in the network's switches and routers, are less significant.

Queuing delays depend on the ratio between the rate of incoming network traffic and the capacity of the outgoing line. A large burst of data could temporarily increase the queuing time significantly and in extreme cases could lead to data loss or a retransmission which would add more delay. However, high capacity core networks rarely operate at a load level where queuing delay is significant. Therefore, the queuing delay contribution to the latency of the core network under normal operating conditions can be assumed to be negligible.

There are two standards that define communications for energy resources within customer premises: OpenADR and IEEE 2030.5 (Smart Energy Profile 2.0). Both standards define HTTP as a top layer protocol, and an overview of HTTP headers in this application is given in [9]. In this paper, it is assumed that the packet size is 500 bytes which consists of:

- Data: 100 bytes
- HTTP headers: 300 bytes
- TCP headers: 30 bytes
- IPv6 headers: 40 bytes
- Data Link overhead: 30 bytes

The communications between the power utility and the WAN, as shown in Figure 2, is assumed to be provided by a single dedicated Optical Carrier 3 (OC-3c)/Synchronous Transport Module 1 (STM1) circuit with 149.76 Mbit/s capacity, based on Synchronous Digital Hierarchy (SDH)/Synchronous Optical Networking (SONET) technology. In order to estimate the latency for the core network it is necessary to know the distance between the TSO/DNO and the WAN via the aggregator. For this example, the scale of Scotland is used and the full distance of 500 km is assumed. Using Equation 1, this results in 2.5 ms of propagation delay from the signal traveling in the fibres.

$$T_{Propagation} = \frac{Distance}{\frac{2}{3} * Speed\ of\ light} \quad (1)$$

Assuming that there is a need for a switch to retransmit the data on average every 50 km [12], the data will have to be serialised and deserialised independently for each hop in the network (i.e. ten times in this example) because the delay from propagation is larger than from serialisation. Using Equation 2, this contributes another 0.534 ms.

$$T_{Serialisation} = \frac{Packet\ size}{Capacity\ of\ the\ communications\ technology} \quad (2)$$

Internal processing within each switch introduces 10-30  $\mu$ s delay [13], which under the assumption of ten hops and 25  $\mu$ s each will contribute 0.25 ms of latency.

With communications traffic arriving at a switch based on a Poisson process, and assuming the time taken to forward packets out of a switch follows an exponential distribution [14], the average queuing delay can be estimated using Equation 3. This equation is based on queuing theory and the "M/M/1 model" using on Kendall's notation [14]. If the DSM scheme is assumed to share the communications network with other applications, the added queuing delay would be 0.134 ms at each point of intersection. This assumes that the communication flow shares the core network with other traffic which utilises 80% (119.8 Mbit/s) of the line bandwidth with constant periodic background traffic (such as from Phasor Measurement Units or IEC 61850 Sampled Values).

$$T_{Queuing} = \frac{1}{Line\ Capacity - Incoming\ traffic\ (\frac{Mbits}{s})} \quad (3)$$

Taking into account all non-negligible delay sources (i.e. propagation, serialisation and switch processing) for the core network shows that the main sources of delay depend on parameters which tend to remain constant. The core network communications latency is therefore estimated to be 3.284 ms with relatively small variation over time [15].

### **Wide Area Network communications**

Long-Term Evolution (LTE) is the fourth generation (4G) cellular technology, and LTE networks can offer a practical and high-performance solution for WANs in many cases. LTE is expected to be increasingly adopted by electrical utilities [16].

For a DSM application, LTE WANs receive the activation signal from the aggregator via the core network, which is multicasted to the NAN coordinators as shown in Figure 3. The physical distance for the communications in WAN networks is relatively small compared to the core network and therefore the propagation delay is negligible. LTE networks utilise a licenced radio spectrum which is limited and has to be actively managed and split into bandwidth channels to provide services to individual users. The capacity of an LTE channel varies depending on the bandwidth of the channel (1.4 MHz-20 MHz), the modulation scheme (QPSK, 16QAM or 64QAM), antenna configuration (SISO, 2x2 MIMO, or 4x4 MIMO), and other parameters. Therefore, the available channel capacity can vary between 1 Mbit/s and 300 Mbit/s. Serialisation given that the sender and the receiver are in a relatively close proximity will happen essentially in parallel; based on Equation 2, assuming a 1 Mbit/s capacity, it will take 4 ms to transfer data from the LTE base-station to the NAN coordinator. Before the transfer can begin, the base-station is required to schedule the bandwidth for a channel and inform the receiving device to prepare for a connection, which can take approximately 15 ms to complete [17].

Wireless communications by nature must deal with interference, multipath fading and environmental conditions. LTE copes with this through a combination of link adaptation to vary the transmission rate (modulation and coding) depending on the channel capacity and two types of Automatic Repeat Request (ARQ) to retransmit data which is received with errors. Such retransmissions and rate reductions will impact latency. For a conventional mobile phone user, the LTE WAN latency is approximately 20 ms [17], but for DSM applications the signal needs to be transmitted to a significantly larger number of devices which will increase the latency of the WAN because of media access and the limited spectrum resource available for the base-station.

A WAN LTE network is simulated and analysed for the application of demand response in [7]. This publication shows that latency of an LTE network depends significantly on processing and queuing with a small additional delay from serialisation and negligible delay from propagation. The overall latency is 40 $\pm$ 20 ms depending on the number of devices receiving the message and on the channel allocation scheme.

### **Neighbourhood Area Network communications**

ZigBee is a low power communications technology used for home automation and for connecting smart meters, distribution automation devices and data aggregators in low voltage distribution networks using multi-hop mesh networks. The coverage of the technology can vary between 10 m and 7 km (line of sight) [6].

A NAN receives data from the WAN network and distributes it to all houses (or other properties) within the neighbourhood ZigBee network. Such networks vary in size and topology, both of which directly influence the

performance of the network.

ZigBee technology is based on the IEEE 802.15.4 physical and data link layers which define the capacity of 250 kbits/s using one of the 16 channels in the unlicensed 2.4 GHz industrial, scientific, and medical (ISM) radio band. It has a maximum packet size of 127 bytes and maximum payload of 102 bytes (or 81 bytes if encryption is added). [18]. Physical distances in NANs are smaller than WANs and thus propagation delay is negligible. However, the capacity of ZigBee technology (using the 2.4 GHz ISM band) is four times smaller than the minimum capacity of LTE used for WAN communications. This means that a significant amount of delay will result from serialisation, queuing and media access. Given the limited capacity, the effect of communications traffic and the increase in the size of the network is also significant on the performance of the network.

A ZigBee-based NAN has been modelled in [8], which includes analysis of networks with 10, 15 and 20 nodes configured in various topologies. The resulting average network latencies range from 57.5 ms (in a randomly distributed 10-node topology) to 583.1 ms (in a 20 node “U” shaped topology). The latencies for individual nodes depend on the number of hops required and how many nodes share the same route to the ZigBee coordinator. These results did not take into account packet loss and radio interference which would further increase the overall latencies. Reference [19] shows an IEEE 802.15.4 network which takes into account interference and the resulting end-to-end latency is 600-800 ms when using advanced data priority aware media access.

Similarly, reference [20] shows delays of 800 ms per hop using a random scheduling method. Furthermore, a non-linear increase in latency per hop from 400 ms to 1500 ms is shown when increasing the number of hops from 1 to 6, using a more advanced media access strategy.

Considering the DSM use case, a single DSM message would have 500 bytes. This can be reduced to 430 bytes by discarding the IPv6 and WAN data link headers that are not required in the NAN. These 430 byte packets would be split into five parts and encapsulated into ZigBee frames. For a message to be received, all of the parts would have to be received by the domestic smart meter (which is assumed to have the capability of a local controller for the DSM resources). Based on [8] [19] [20] an estimate for the latency in a NAN network can be approximated as  $1200 \pm 900$  ms for the first frame and another  $300 \pm 200$  ms for consecutive frames depending on the size (up to 5 hops) and topology of the network, and media access strategy.

### Home Area Network communications

The HAN is assumed to be delivered by a ZigBee network within the home, which connects e.g. a heat pump/boiler, EV and potentially a solar-battery system to the smart meter. The ISM band used by ZigBee has many other domestic uses: Wi-Fi, Bluetooth, cordless phones, baby monitors, and emissions from microwave ovens. This

means that there is a large amount of potential interference for ZigBee and that media access is the most significant source of delay in this network.

In [9], a ZigBee-based HAN is analysed for the application of demand response. The two main parameters in media access that define the performance of this network have been identified as the probability of a successful Clear Channel Assessment (CCA) and the media access strategy in this case - the number of Guaranteed Time Slots (GTS). The maximum latency with no interference theoretically can vary between 43 ms and 3350 ms depending on the number of GTS assigned (1 to 7 GTS). Assuming 5 or more GTS and the probability of CCA greater than 0.6, would bound the upper latency limit to ~300 ms.

## IMPACT ON DSM PERFORMANCE

### Analysis of DSM for fast frequency response

The performance of the communications networks that connect a control centre to the end devices has been analysed in the previous section. The results show that the latency expected for DSM applications is between 1-4.5 seconds. The contribution to the delay from each network type and the main influences to the delay have been summarised in Figure 4.

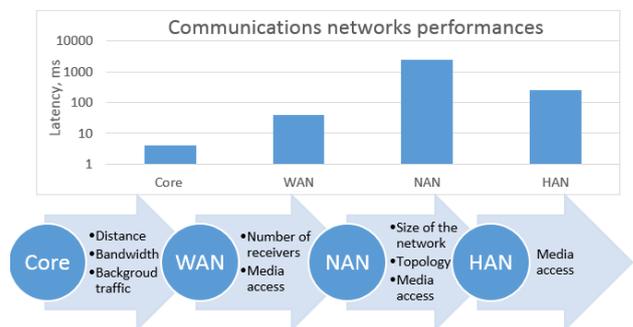


Figure 4: Communication network performance for DSM application and main influences to the latency

For conventional primary frequency response services, the participating device (or aggregator) needs to respond within 10 seconds of the frequency deviation. Based on the analysis conducted in this paper, it is evident that an aggregator that deploys a transactive energy market to control its portfolio can participate in primary frequency response services. However, newer ancillary services such as enhanced frequency response in Great Britain [1] expects a device to respond in less than 1 second. Based on the assumed communications infrastructure, it would not be possible for an aggregator deploying a transactive energy market to control its portfolio to participate in such services. However, because 70-91% of the delay is due to the NAN, this can be mitigated by using a higher performance communications technology in the NAN compared to that assumed above. Also, some of the aspects of communications networks like the bursty traffic and its implication on queuing delay are difficult to estimate thus detailed communications emulation would be needed to more accurately estimate the performance.

### Scalability in future power system architectures

As has been highlighted previously, the delay is dependent upon the number of devices that will be utilizing the same communication infrastructure. With the number of devices capable of providing flexibility set to increase rapidly in the future, a centralized approach to frequency management might not be feasible.

Acknowledging this issue, ELECTRA IRP (a European FP7 project) [21] has proposed a distributed architecture for controlling future power systems, called the Web of Cells. The Web of Cells concept involves the division of the present day power system into smaller, more automated, regions or cells. Each cell is responsible for addressing any voltage or frequency deviation arising within it, utilizing the most local resources as possible. The performance models of communications systems under development by the authors will be used to more fully characterise the impact of communications performance on new frequency and voltage controls. This will support the better and quantitative understanding of the benefit of such decentralised concepts.

### CONCLUSIONS AND FUTURE WORK

This paper provides a detailed and realistic view of all the communications networks required to deliver a real-time DSM scheme. Each network type has been analysed and the most important aspects, in terms of latency, have been identified. The analysis enables the overall communications latency for a DSM application to be estimated, and establishes that it is possible for resources in domestic customer premises to participate in primary frequency response.

Future work will develop an end-to-end simulation to incorporate various types of communications traffic that could be utilising the same communications networks, characterise instantaneous latencies and investigate the effect of varying latency on the capability of containing frequency and restoring balance to the power system using DSM. This will be demonstrated using future power system architectures, utilising novel decentralised control methods. This work will be further expanded to incorporate delays associated with DSM operation, sensors, control algorithms and cyber security provisions.

### Acknowledgments

The work in this paper has been supported by RTDS Technologies Inc., Energy Technology Partnership Scotland and the European Commission under the FP7 project ELECTRA (grant no: 609687). Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the European Commission.

### REFERENCES

[1] Wall, P, Shams, N, Terzija, V, Hamidi, V, Grant, C, Wilson, D, Norris, S, Maleka, K, Booth, C, Hong, Q & Roscoe, A 2016, Smart frequency control for the future GB power system. in IEEE PES

ISGT Europe. IEEE.

[2] National Grid, 'Demand Side Response', Available [Online]: <http://www2.nationalgrid.com/UK/Services/Balancing-services/Demand-Side-Response/>

[3] R. Kamphuis, B. Roossien, M. Eijgelaar, H. De Heer, J. van de Velde and A. Van Den Noort, 2013, "Real-Time trade dispatch of a Commercial VPP with Residential Customers in the PowerMatching City SmartGrid Living Lab", Proceedings of 22nd International Conference on Electricity Distribution, CIRED, Stockholm.

[4] K. Kok et al., "Dynamic pricing by scalable energy management systems — Field experiences and simulation results using PowerMatcher," IEEE Power and Energy Society General Meeting, San Diego, CA, 2012, pp. 1-8.

[5] K. Kok and S. Widergren, "A Society of Devices: Integrating Intelligent Distributed Resources with Transactive Energy," in IEEE Power and Energy Magazine, vol. 14, no. 3, pp. 34-45, May-June 2016.

[6] Zahurul, S., et al. "Future strategic plan analysis for integrating distributed renewable generation to smart grid through wireless sensor network: Malaysia prospect." Renewable and Sustainable Energy Reviews 53.C (2016): 978-992.

[7] A. Roy, H. Kim, N. Saxena and R. Kandoori, "LTE multicast communication for demand response in smart grids," IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), New Delhi, 2014, pp. 1-6.

[8] Mohammad Shams Yazdi, "Optimal resource allocation for QoS supports in smart grid neighborhood area network, Msc Thesis, Ryerson University available [Online]: <http://www.ee.ryerson.ca/~yhe/my%20papers/Thesis/Msc%20Mohammad%20Shams%20Yazdi%202013.docx>

[9] V. Kounev and D. Tipper, "Advanced metering and demand response communication performance in Zigbee based HANs," INFOCOM, Proceedings IEEE, Turin, 2013, pp. 3405-3410.

[10] Fluid Mesh, Urban communications networks Available [Online]: <http://www.fluidmesh.com>

[11] Verizon, 'IP Latency Statistics' Available [Online]: <http://www.verizonenterprise.com/about/network/latency/>

[12] N. Massa, 'Fiber Optic Telecommunication', Springfield Technical Community College Springfield, Massachusetts, Available [Online]: <https://spie.org/Documents/Publications/00%20STEP%20Module%202008.pdf>

[13] Cisco, 'Design Best Practices for Latency Optimization', white paper, 2007 Available [Online]: [https://www.cisco.com/application/pdf/en/us/guest/netso/ns407/c654/ccmigration\\_09186a008091d542.pdf](https://www.cisco.com/application/pdf/en/us/guest/netso/ns407/c654/ccmigration_09186a008091d542.pdf)

[14] Teknomo, Kardi, Queuing Theory Tutorial Available [Online]: <http://people.revoledu.com/kardi/tutorial/Queueing/>

[15] S. M. Blair et al., "Validating secure and reliable IP/MPLS communications for current differential protection," 13th International Conference on Development in Power System Protection 2016 (DPSP), Edinburgh, 2016, pp. 1-6.

[16] Nokia, 'Field Area Network for Power Utilities', Available [Online]: <https://networks.nokia.com/power-utilities/field-area-network>

[17] UNIK, 'Latency in LTE Systems, Available [Online]: [http://cwi.unik.no/images/Latency\\_in\\_LTE\\_comments.pdf](http://cwi.unik.no/images/Latency_in_LTE_comments.pdf)

[18] M. Franceschinis, C. Pastrone, M. A. Spirito and C. Borean, "On the performance of ZigBee Pro and ZigBee IP in IEEE 802.15.4 networks," 2013 IEEE 9th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), Lyon, 2013, pp. 83-88.

[19] I. S. Al-Anbagi, M. Erol-Kantarci and H. T. Mouftah, "Fairness in delay-aware cross layer data transmission scheme for wireless sensor networks," 26th Biennial Symposium on Communications (QBSC), Kingston, ON, 2012, pp. 146-149.

[20] I. Hosni, F. Théoleyre and N. Hamdi, "Localized scheduling for end-to-end delay constrained Low Power Lossy networks with 6TiSCH," IEEE Symposium on Computers and Communication (ISCC), Messina, 2016, pp. 507-512.

[21] ELECTRA IRP, <http://www.electrairp.eu/>