On Energy-Aware Communication and Control Co-design in Wireless Networked Control Systems
Nicolas Cardoso de Castro, Carlos Canudas de Wit, Karl Henrik Johansson

To cite this version:
On Energy-Aware Communication and Control Co-design in Wireless Networked Control Systems *

Nicolas Cardoso de Castro *  
Carlos Canudas de Wit **

Karl Henrik Johansson ***

* INRIA Rhône-Alpes, NeCS Team, Grenoble, FRANCE (nicolas.cardosodecastro@inrialpes.fr)  
** CNRS, Department of Automatic Control, GIPSA-Lab, NeCS Team, Grenoble, FRANCE (carlos.canudas-de-wit@gipsa-lab.grenoble-inp.fr)  
*** ACCESS Linnaeus Centre, School of Electrical Engineering, Royal Institute of Technology (KTH), 100 44 Stockholm, SWEDEN (kallej@kth.se)

Abstract: Energy is a key resource in Networked Control Systems, in particular in applications concerning wireless networks. This paper reviews the multi-layer architecture of those systems in the light of their energy-use, and points out major contributions in the area of energy-management policies, layer per layer. This review of the literature is organized according to the layered communication architecture covering from bottom to top the Physical, Data Link, Network, and Application layers. We specifically focus on advances that concern energy-aware management in wireless communication and control co-design. It is argued that existing work is limited to single layer approaches, with a lack of design methods taking into account several layers.

1. INTRODUCTION

Networked Control Systems (NCS) are systems in which the sensors and the actuators communicate with the controller through a network (see Figure 1). We consider wireless networks where sensors usually have computational resources, a radio unit, a battery, and in some cases an energy-harvesting device. To save maintenance cost, these sensor nodes are designed to operate as autonomously as possible in term of energy consumption. Energy of these nodes is used to power the sensor electronics, the data processing and routing, and the radio transmission. Energy management is then a key issue to optimize the life-time operation of the overall wireless network. This issue is abundantly treated in the literature of Wireless Sensor Networks (WSN) [Raghunathan et al., 2002], but barely not in the framework of NCS. It is worth to mention that the radio unit consumes most of the effective power, some works reported up to 80% [Kimura and Latifi, 2005] of the total available power. Hence, there is room for added value in terms of intelligent use of the node energy at the radio communication stage, and on the information processing in view of reducing the cost per bit and its associated routing. Whereas reliability (i.e. successful packet delivery) is the main matter in WSN used for monitoring, latency (i.e. delayed packet delivery) is an important aspect in addition for NCS. Control laws are usually robust to some delays and packet losses. Thus, instead of maximizing the reliability, as in WSN, communication devices in NCS should balance reliability and latency to save energy while meeting the control requirements. This highlights that energy savings can be achieved through a co-design of control and communication.

The purpose of this paper is to review the multi-layer architecture of NCS in the light of their energy-use, and to point out major contributions in the area of energy-management policies, layer per layer. We specifically focus on advances that concern energy-aware management in communication and control co-design, but we argue also that there is a lack of integrated design methods that take into account the various layers in an overall systematic way suitable for energy-aware networked control. The review of the existing works reported in the literature is organized into a bottom to top 4-layers architecture, named here the NCS stack [Liu and Goldsmith, 2004]; Physical, Data Link, Network, and Application layers.

(1) **Physical layer** Performs the radio modulation of the digital data. The transmission power and the bit rate of the modulation can be controlled to save energy.

(2) **Data Link (MAC) layer** Defines how to use and share the transmission medium. Energy can be saved, by managing the activity modes of the radio unit, and by adapting the MAC protocol parameters to the channel conditions and the control requirements.

(3) **Network layer** Routes the data through the network. The parameters of the routing protocol can be adapted to trade transmission reliability/latency with its associated cost.

(4) **Application layer** Concerns the source encoding and decoding, and computes the control law. Encoding can reduce the amount of data to be sent or concentrate the information on low-cost codewords. The controller needs to be re-designed to tackle the variable delays and the information losses due to wireless communication.

---

* This work was supported by EU STREP project FeedNetBack FP7-ICT-2007-2.
2. PHYSICAL LAYER

The main role of the physical layer is to modulate the data into radio waves. Energy at this layer can be saved by controlling the transmission power or the data rate.

2.1 Power control

The transmission power can be adapted as a function of the transmitter-to-receiver distance, or the channel conditions. Arbitrarily large power will induce interferences with other communication links and will have unmoderated energy consumption. The transmission power is generally related to the communication reliability. Power control can rely on the receiver observations which are fed back to the transmitter, or on the channel conditions probed by the transmitter.

Power control based on the receiver consists in, increasing the transmission power at the transmitter side when packets are not well received, or decreasing the power in case of success. The feedback loop is achieved through a predictive controller in [Quevedo and Ahlen, 2008], or simple algorithms in [Zurita Ares et al., 2007] (see Figure 2) and in [Correia et al., 2007]. In both [Quevedo and Ahlen, 2008] and [Zurita Ares et al., 2007], a mechanism is used to detect the errors in the received packets which are related to the bit error rate. The receiver sends back that error metric to the transmitter which adjusts its power, whereas in [Correia et al., 2007] the decision is based on acknowledgements of the receiver.

In the approach using the channel conditions, each node controls its transmission power based on channel measurements such as the interference power, the received signal strength indication [Fink et al., 2009], the packet error rate [Zurita Ares et al., 2007], the noise power and the signal attenuation [Correia et al., 2007]. Authors in [Zurita Ares et al., 2007, Correia et al., 2007] propose algorithms to compute the minimum transmission power ensuring correct reception, given a channel model and channel measurements. Although, in this approach, the knowledge of the receiver is also fed back to the transmitter to determine the channel state, it is different from power control based on the receiver which does not rely on channel model.

Power control needs extra communication between the receiver and the transmitter. It is shown in [Johansson et al., 2007] that in a WSN where nodes are not too far from each other (up to 100 m to 1 km, depending on the set up), a scheme with a power control mechanism consumes more energy than a scheme with a fixed power amplitude plus a given margin ensuring correct bit error rate when the channel deteriorates. The short distances between nodes make the consumption dominated by radio circuitry instead of the transmission energy. The energy saved by decreasing the transmission power under good channel conditions is wasted by the extra communication used to decide the power amplitude. Power control should then only be considered in the case of distant nodes or energy-efficient radio units where the transmission power dominates the unit consumption.

2.2 Bit rate control

Power control is mainly used to regulate transmission quality under packet loss and signal interference. An alternative to regulate transmission reliability without increasing the transmission power, at the price of increasing the transmission latency, is the bit rate control, sometimes referred to as Dynamic Modulation Scaling (DMS).

DMS is a methodology that consists in switching the modulation characteristics (under constant power), to improve the signal reception. Figure 3 shows two modulation strategies: the 8-Phase Shift Keying (8-PSK), which codes 8 words by changing the signal phase; and the 16-Quadrature Amplitude Modulation (16-QAM), which codes 16 words by changing the phase and the amplitude. Switching between these two types of modulation changes the number of bits sent in a period of time (the size of the constellation).

Switching is in general performed via a scheduler [Schurgers et al., 2003] that selects the best modulation scheme as a function of the packet rate load. When the demand is low, the scheduler uses low-bit modulation scheme that uses less energy per period. However, some interesting works, such as [Zafer and Modiano, 2006], consider the relationship between the bit rate, the energy consumption and the latency, to derive energy-aware time-constrained switching policies.

One major difficulty of DMS is that the receiver must be aware of a modulation change, which implies communications between the transmitter and the receiver before a switching.

In terms of control, adapting the modulation can be viewed as receiving measurements with variable delays or with variable quantization level.

3. DATA LINK (MAC) LAYER

The Data Link (MAC) layer defines how to access the medium, shared among all the nodes. An important amount of energy is wasted by nodes in the idle-listening state (the node is in the receiving mode without data to receive). Radio units typically save energy by switching between different activity modes (idle, off, in addition to receiving, transmitting) when no communication is needed. This energy management implies that some of the node components are switched off at odd moments. This concerns: the frequency synthesizer, the crystal oscillator, and the voltage regulator in the radio unit [Brownfield et al., 2006], but also the sensing unit, the processor, and the memories. Switching between modes is both time and energy consuming. A trade-off has to be made between node awareness and energy consumption.

![Diagram of modulation constellations](image-url)
Scheduling policies are used to determine when, how long, and how deep a node should sleep. The work in [Brownfield et al., 2006] relies on a TDMA-based MAC protocol, which ensures medium sharing, and moreover is periodic and allows to schedule sleep time as shown in Figure 4. The scheduling policy depicted in this figure can be improved by adding modes with intermediate sleep depth to balance energy savings and reactivity when the sleep periods are too short to turn off completely the node. Sleep periods can be extended by adapting the data slot to the real time needed for transmission [Brownfield et al., 2006]. The major drawback of TDMA-based scheduling is the scalability since a slot must be assigned to each node, the latency increases with the number of nodes.

3.2 MAC protocol tuning

The MAC protocol parameters can be adapted to balance reliability and latency in order to meet the control requirements and to minimize the energy consumption. The protocols considered in this section cover both the Physical and the Data Link (MAC) layers, and sometimes also the Network (routing) layer. As concluded in [Liu and Goldsmith, 2003], there is no protocol dedicated to NCS.

The authors in [Fischione et al., 2009a] have derived analytical models of the delay, the reliability and the energy consumption in unslotted 802.15.4 protocol, see Figure 5. From these models, the sleep time $R_t$, and the listening time $R_l$ of the CSMA-based MAC protocol can be chosen to minimize the energy consumption under delay and reliability constraints. This scheme is more energy-efficient than X-MAC, it uses CSMA, which is scalable whereas TDMA is not, and it lets the control application adapt the latency and the reliability according to the control requirements.

Original works in [Park et al., 2008, Di Marco et al., 2010] derive protocols suitable for control applications. They both aims to minimize the network consumption under latency and reliability constraints. The network topology is based on clustering in both works but [Park et al., 2008] uses CSMA-based MAC protocol whereas [Di Marco et al., 2010] uses a hybrid TDMA/CSMA one. The protocol variables (degrees of freedom) are the wake-up rate and the clustering in [Park et al., 2008], and the slot duration, the access probability, and the wake up probability in [Di Marco et al., 2010].

4. NETWORK LAYER

The Network (routing) layer is in charge of choosing a path (i.e. a series of relay nodes) in the network to transmit packets between the transmitter and the receiver, in an energy-efficient manner. The usual metric at this layer is the network life-time, i.e. the time between the commissioning and the moment when a node first runs out of battery.

4.1 Energy-efficient routing

Literature is well furnished about routing protocols for WSN. A comprehensive study can be found in [Al-Karaki and Kamal, 2004] where the authors distinguish between protocols based on the network structure from the ones based on protocol operation. In general terms, the former use the topology of the network and the amount of power available at each node to choose a route, see Figure 6, whereas the latter have some parameters which allow to dynamically control the routing policy or the amount of data transmitted.

Some interesting protocols based on protocol operation are proposed in [Al-Karaki and Kamal, 2004]. However these protocols do not expose their tunable parameters to the application layer, meaning that the energy cannot be balanced according to the control performances in the framework of NCS. Among all the protocols surveyed by [Al-Karaki and Kamal, 2004], only query based routing protocols offer tunable parameters to the application layer. In this approach, the controller node is the source of all the traffic in the network, since it requests measurements from the sensor nodes or it runs in open-loop. However, the delay on the measurements is doubled, and the controller has to waste energy to detect unpredictable disturbances.

Breath and TrEEND protocols, already discussed in section 3.2, also cover the Network layer. In spite of all these achievements, there are not yet standard fully dedicated protocols that accommodate all the desired requirements from NCS.
Fig. 6. Network topology from [Akyildiz et al., 2002] considering the power available (PA) at each node, and the energy cost $\alpha_i$ to send one packet over a link $i$. The transmitter $T$ should choose a route to send data to the sink $S$, between:

Route 1: $T$-$B$-$A$-$S$, $PA_{tot} = 4$, $PA_{min} = 2$, $\alpha_{tot} = 3$.

Route 2: $T$-$C$-$B$-$A$-$S$, $PA_{tot} = 6$, $PA_{min} = 2$, $\alpha_{tot} = 6$.

Route 3: $T$-$D$-$S$, $PA_{tot} = 3$, $PA_{min} = 3$, $\alpha_{tot} = 4$.

Route 4: $T$-$F$-$E$-$S$, $PA_{tot} = 5$, $PA_{min} = 1$, $\alpha_{tot} = 6$.

The route is chosen according to the routing scheme. Maximum $PA_{tot}$ route is the 2nd. Minimum energy route is the 1st. Maximum hop route is the 3rd.

4.2 Network coding

In the routing protocols discussed previously, the relay nodes just forward the packets they receive. Under the network coding paradigm, the nodes can perform some operations on the data from the packets they receive before forwarding them. Energy-aware network coding aims to limit the amount of data traveling in the network. It is achieved by dropping redundant information, or by compressing correlated data from neighboring nodes. Compressing data is also called aggregation or fusion.

An interesting work proposed by [Abdelzaher et al., 2004] controls the degree of data aggregation by using the traffic load as a feedback variable in order to meet some end-to-end delay constraints.

There are only few contributions, as [Abdelzaher et al., 2004], which are using network coding in NCS. However, there are plenty of publications about distributed estimation, which can be seen as network coding in the sense that the nodes perform local computations on the measurements and send only their own results. Distributed estimation is shown to be very performant in [Fischione et al., 2009b], giving lower estimation error variance than distributed Lagrangian message passing.

Network coding is not only used to reduce traffic in a network, but also to recover from network failures. In the case of failures such as channel fading or shadowing, cooperative diversity, [Laneman, 2004], can be used either to improve robustness to fading without increasing the power consumption, or to save energy for equivalent communication performances. This technique exploits spatial diversity to emulate an antenna array, increasing signal redundancy. This technique can be used to save further energy and it can be used in practice in algorithms with limited computational resources as NCS or WSN.

Analysis of networks using network coding is not trivial. Authors in [Koetter and Méard, 2001] present an algebraic framework to study the capacity issue in such networks. This work provides relationships between systems of polynomial equations and the solutions to network problems. These tools permit to solve problems such as feasibility of given set of connections in a network or recovery from non-ergodic link failures. This approach is promising since algebra is a well furnished part of mathematics.

5. APPLICATION LAYER: QUANTIZATION AND SOURCE CODING

The application layer hosts the executive part of the sensing and the control, that we have divided into two categories; the sensor node that does the signal sensing and processing, and the control node that decodes the received information and computes the control action. In this section we present works related with the sensor node, and in particular aspects in relation with quantization and energy-aware source coding. The application layer at the control node is treated in Section 6.

5.1 Quantization and rate constraints

In the sensor node, the measurements from the (analog) real world are turned to digital values by the process of quantiza-
tion, which introduces a loss of information dependent to the quantizer specifications. Moreover the amount of traffic in a network is bounded, introducing data rate constraints. These both limitations can compromise the stability of a system, and cannot be ignored in the analysis and the design of a NCS.

Quantization is often included in the process of source coding, which aims at compressing data measured by the sensor to make the transmissions more efficient by reducing the total amount of bits and, in some cases, the cost per bit associated to the transmission. Source coding is discussed in next subsection. A comprehensive study about control under data rate constraints can be found in [Nair et al., 2007]. Authors show that there is a trade-off between data rate and control performances and that there exists a minimum rate (related to the system instability) needed to achieve stability. They investigate coding strategies with finite memory coder and show that asymptotic stability cannot be achieved.

5.2 Uniform energy coding

Uniform energy coding assumes that the transmission cost is uniformly distributed among all the codewords, or equivalently that bits “0” and “1” are associated to an equal energy transmission cost. In this context, the coding strategy consists in trying to minimize the total number of bits that the codewords will use without any other consideration about the nature of the bits value themselves. Differential coding encodes differences (or variations) of a signal rather than its absolute value. Delta modulation is one of the simplest form of differential coding. It uses a 2-level (1-bit) quantizer and a dynamic predictor implemented on both transmitter and receiver sides. The node sends “0” if the difference between the measure and the prediction is positive, and “1” otherwise. This reduces the amount of data transmitted to only one bit per sample. This scheme exhibits a trade-off between stability and precision tuned via the predictor’s gain. The closed-loop stability of this scheme in the context of NCS has been studied in [Canudas De Wit et al., 2009].

5.3 Non-uniform energy coding

Non-uniform energy coding assumes that the transmission cost is not uniformly distributed among all the codewords, i.e. some codewords (or bits) are more costly than others. These differences can be captured by assuming some particular probability distribution, or by changing the distribution via the well known process of transform coding.
Minimum Energy (ME) coding combines the On/Off Keying (OOK) modulation with a particular energy-efficient coding method. OOK modulation emits a fixed signal with a given power to transmit a high bit (i.e., "1") and no signal to transmit a low bit (i.e., "0"). The goal of ME coding is to produce codewords formed by the minimum number of high-cost bits (or equivalently with the maximum numbers of "0"). The base work on ME coding [Erin and Asada, 1999] assumes that the source symbol probabilities are known. For a given dimension of alphabet, codewords with fewest high bits are assigned to the symbols with the highest probabilities and inversely for codewords with fewest low bits.

Entropy Coding aims to reduce further the average amount of bits to be sent. It is assumed that a certain event (i.e., stand-still event) has a higher probability to occur. Entropy coding is often associated to event-based detector, and a variable length encoding process. The event-based detector has the ability to quantify and to differentiate stand-still signal events from the changes in the source (level crossing detector), see Figure 7. Coding is effectuated by defining a 3-valued alphabet for the minimum bit case, and (2L + 1)-valued alphabet for a general case with a precision inversely proportional to L ∈ Z+. Entropy coding improves the energy savings by assigning a probability distribution to the events. The mean transmission energy can be substantially improved for systems where the stand-still events may have high probability to occur. An example of an entropy coding mechanism is the run-length encoding as show in Table 1. This results in transmissions inherently asynchronous as the transmission period depends on the signal events.

![Figure 7. Illustration of the 3-levels coding scheme, from [Camadas De Wit and Jaglin, 09]. The signal detection levels are spaced by the quantum Δ. Uniform samples are taken every time T_s triggered by a local synchronous clock. The level detector produces a signal s_n (identified by "01" or "10") whenever a level crossing takes place, and a "00" if the signal remains within the level. The parameter Δ defines a trade-off between stability and precision.](image)

6. APPLICATION LAYER: ASYNCHRONOUS CONTROL

The distributed nature of sensor and control nodes in the context of NCS makes each of these components to operate with their own clock and to react asynchronously to events. This forces the control node to perform a synchronization between nodes, or to operate under asynchronous updating. Operating under asynchronous sampling, i.e. the samples are received aperiodically, is more involved but it is also more energy-efficient, and it is likely to perform better with bad channel conditions than a synchronous sampling which temporary loses the synchronization or often loses packets.

6.1 Split sensing and control

In this approach, sensing and control computing are done in separated nodes. The sensor node only measures and processes the sensed data to be transmitted to the control node, in charge of updating the control law. Energy-efficient asynchronous control relies on reducing the communication load needed to update the control law. Energy-efficiency sensing schemes must be implemented in the sensor node to decide when to send new measurements to ensure closed-loop stability and given control performances.

Zero Order Holder (ZOH) based controller is the simplest approach to deal with aperiodic samplings due to event-based quantizer used in the sensor node. The controller keeps applying the same control input until a new measurement is received and used to update the control input [Rabi and Baras, 2007]. The ZOH can also be employed to hold the received information rather than the control input.

Time-varying delay can be used to model the asynchronous arrivals, at time t_k, of measurements at the control node [Fridman et al., 2004]. The time-varying delay is then modeled as:

\[
\tau(t) = t - t_k, \quad 0 \leq \tau(t) \leq h
\]

where h is the maximum time interval between two measures. The asynchronous arrival of the sensed information can thus be modeled as: \(x(t_k) = x(t - \tau(t))\), and then a static feedback K can be designed to cope with this delayed state and to stabilize the resulting closed-loop system. A limitation of this approach is due to long stand-still signals (in particular when system is at its equilibrium) in event-based sensing, as they will introduce too long h leading to a very conservative control design.

Communication-controlled scheme aims to control the transmission policy to send as few messages as possible between the sensor and the control nodes. A predictor (or an estimator) is typically implemented in the sensor and control nodes, and the sensor node only sends measurements if the prediction error crosses a given threshold, as shown in Figure 8 with an MPC controller (from [Bernardini and Temporad, 2008]). Authors in [Imer and Basar, 2005] derive an optimal estimator over a finite horizon of N samples where the sensor is only allowed to send \(M < N\) measurements to the estimator. The decisions of the sensor are based on the knowledge of the process to be estimated and the knowledge of the controller. The work in [Xu and Hespanha, 2006] provides design and analysis of communication logic in NCS. The authors investigate the stability of the system and the trade-off between communication cost and control performances. They propose stochastic and deterministic logics to limit the amount of data transmitted.

![Figure 8. Illustration of split sensing and control approach. The sensor transmits measurements asynchronously to save communication energy. The MPC controller uses local predictions when no measurement is received.](image)

<table>
<thead>
<tr>
<th>Input</th>
<th>Period</th>
<th>Output</th>
<th>Input</th>
<th>Period</th>
<th>Output</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>0</td>
<td>001</td>
<td>00</td>
<td>0000001</td>
<td>3T_s</td>
</tr>
<tr>
<td>00</td>
<td>2T_s</td>
<td>011</td>
<td>0</td>
<td>000000</td>
<td>3T_s</td>
</tr>
<tr>
<td>01</td>
<td>00001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Run-Length Encoding, from [Camadas De Wit and Jaglin, 09]. T_s is the sampling time.
In the cooperative sensing and control approach, the sensor node processes the measurements but also computes parameters for the control law. It pre-computes some parameters that will be used to generate the control law, see Figure 9. The time-profile of the control evolution is here pre-constrained to some specific profiles such as series of impulses, piece-wise constant, or linear growing. The sensor node sends parameters describing the time-profiles.

The cooperative sensing and control approach has been introduced in [Aström and Bernhardsson, 2002], even though the problem is not explicitly addressed in the framework of NCS, using impulses-shaped time-profile for the control. The parameters provided by the sensor node are the time instant to apply the impulses, and their magnitudes. The asynchronous sampling shows a better output variance than the periodic one. Another application, using a PI-control structure, was introduced in [Rabi and Johansson, 2008]. On the sensor node, two level crossing detectors are implemented. The first, called the P-part, is triggered by the regulation error while the second, the I-part, is triggered by the integral of the error. The sensor node sends time intervals for updating the control waveform which is made of an impulse plus a piece-wise constant value. The trade-off between control performances and communication consumption can be balanced through the thresholds used at the sensor side. However the authors do not propose specific rules to tune the gains. And they show that sustained oscillations can result from bad tuning of the gains and the thresholds. Literature is not well furnished about this approach and open issues, such as multiple sensor nodes or real channel conditions, are to be investigated.

7. CONCLUSION

We have reviewed the four layers of the NCS stack: Physical, Data Link, Network and Application layers, emphasizing on energy consumption. Energy-efficient techniques exist at each layer but they are rarely suitable for the NCS framework. It appears to be a gap between the application layer, mostly treated by the control community, and the ones below, addressed by the communication community. Even though control applications consider real channel conditions (packet loss, delay, limited data rate), they mainly use only one sensor node and one control node, avoiding medium sharing and routing issues. On the other hand, MAC and routing protocols often control their energy consumption themselves, leaving few or no degrees of freedom to the application layer. This highlights the need for protocols dedicated to NCS, allowing the control application to change the latency and the reliability to meet the control requirements under energy constraints, as proposed in [Park et al., 2008, Fischione et al., 2009a, Di Marco et al., 2010].

Most part of the node energy is consumed by the radio unit, with important waste in idle-listening state. An open issue we want to address in our future work is the co-design of activity mode management and asynchronous control, to improve further energy savings.

REFERENCES

C. Fischione, S. Coten Egeren, P. Park, K.H. Johansson, and A. S. Vincentelli. MAC analytical modeling and optimization in unslotted IEEE 802.15.4 WSN. In Communications Conference on Sensor, Mesh and Ad Hoc Communications and Networks, 2009a.