**Author names:** Gaura, E. and Brusey, J.
**Title:** 'Bare necessities—Knowledge-driven WSN design
**Article & version:** Post-print version
**Original citation & hyperlink:**
Gaura, E. and Brusey, J. 'Bare necessities—Knowledge-driven WSN design' Sensors, 2011 IEEE (pp. 66-70). IEEE
http://dx.doi.org/10.1109/ICSENS.2011.6127213

# Bare necessities—Knowledge-driven WSN design

Elena I. Gaura, James Brusey and Ross Wilkins
Cogent Computing, Coventry University
Priory Lane, Coventry CV1 5FB
{e.gaura,j.brusey,wilkin24}@coventry.ac.uk

***Abstract*—The viability of wireless sensor applications often hinges on minimising power consumption whilst maximising the informational output. Although many low-level platform-oriented energy saving mechanisms have been developed, considerable savings are possible at application level. This work presents an approach to pushing the calculation of application-level state closer to the information source. The context in which this approach is evaluated is a residential building monitoring application. Combined with the Spanish Inquisition Protocol (SIP), this is shown, based on deployment data, to reduce the average transmission period for temperature data from once every 5 minutes to an average of once every 38 days for an allowed error threshold of 10% on any component of the application-level state. For combined sensing of temperature, relative humidity and CO2, the average transmission period drops to 13 days. This transmission reduction should considerably extend network life while having minimal effect on the usefulness of the information gathered. Most importantly, the underlying approach generalises to a wide variety of applications.**

***Index Terms*—Wireless Sensor Networks**

## I. Introduction

Tiny, low power, wireless computing platforms (motes) are moving from their research roots to encapsulation in useful products. As consumer products, they will shed their developer-oriented capabilities, provide processed information instead of raw data, and become highly tuned to the specific application. Central to this tuning is consideration of which information is both *relevant* and *unexpected*.

Research in the area of Wireless Sensor Networks (WSNs) is evolving. Initially, the focus was on the device: computation and communication were only just possible and tended to be unreliable. Next, the focus was on the network: careful design of communication strategies were needed to extend lifetime and thus to make the systems usable in the real world. These systems are now usable and focus is now moving towards the application. There is becoming less of a need to monitor metrics unrelated to the application, such as signal strength or network tree structure. Furthermore, there is greater trust in processing on the node. It is no longer assumed that every unprocessed sensor reading must be transmitted to a back-end store to allow checking of any processed output.

A key factor in WSN design is network longevity. The additional maintenance cost associated with replacing batteries is a significant incentive to optimise the design in this respect. Even so, it is rare for systems to achieve longer than a 1 year lifetime. For example, the Torre Aquila project [1] deployed a WSN to monitor the structural integrity of a heritage building to better plan maintenance. The deployment consisted of a number of node types measuring temperature, relative humidity, light, acceleration and fibre optic sensors all sampling at a rate of 10 minutes. Using two pairs of size C batteries, Ceriotti *et al*. estimate a node lifetime of one year. SensorScope [2] is a system for an indoor environmental monitoring network, built around the Telosb platform (measuring temperature, relative humidity and light) and uses the B-mac networking protocol. With the nodes sampling at 2 minutes, Schmid and Dubois estimate that the system will run for 61 days on a pair of AA batteries. The WISE-MUSE project [3] developed a WSN which monitors temperature, relative humidity and light in an art gallery for the preservation of collections. At a sampling rate of 10 minutes a life time of 2 months was achieved using a pair of AA batteries. The above three examples demonstrate the limited expected battery longevity of WSN systems monitoring simple measurands such as temperature and humidity and relatively low sample frequencies.

This paper builds on prior work [5] that developed the Spanish Inquisition Protocol (SIP), a generic, model-based approach to WSN transmission reduction. In SIP, little is assumed about the type or frequency of sensor measurement or the overall application. In contrast, the main contribution of this paper is to demonstrate how, building on the basic approach of SIP, one can take into account the information requirements of the application and that there are considerable performance benefits to be gained from doing so. As with SIP, this is a *timely* protocol in the sense that, unlike a simple reduction in transmission frequency, the approach presented here will transmit when a significant change occurs in the environment.

The paper continues, in Section II, with introducing the application of residential building monitoring and the associated information requirements. Section III presents the algorithmic approach, building on SIP. The performance of this system is further analysed in Section IV followed by concluding remarks in Section V.

## II. Towards Application-level state in buildings monitoring

Given the context of a specific application, raw sensor measurement data is often highly compressible. Although, while in initial stages of WSN development it is usually necessary to deploy additional sensors and gather much more data, once the key application "metrics" become established, fewer sensors and less frequent transmissions are sufficient. In this paper, this compact form is termed *application-level state* since it is both at an application-level (in terms of context) and referring to the condition of the monitored environment at a point in time.

A specific application used as a case study throughout this paper is that of residential building monitoring, which involved monitoring per room temperature, humidity, and air quality as well as energy usage for the whole house. The end-users of the data were social housing landlords. Their main aim was to be able to objectively assess home comfort versus energy expenditure. In consultation with the end-users, three key metrics were found that enabled them to understand the data and to make decisions based on it. These were:

- Per room exposure by band for temperature, relative humidity and air quality (see Fig. 2)
- Per house probability of "reasonable" comfort and expected comfort based on likely occupancy distribution,
- Overall energy usage per unit area per degree days.

Exposure bands primarily represent comfort but also hint at likely health impact. For example, cold can negatively affect the immune system; humidity can lead to toxic levels of mould growth; and poor air quality can lead to higher rates of lung-related illness. Prior work [4] explores the development of these metrics in more detail.

The second metric (comfort) combines probability of room occupancy with the probability of being comfortable in that room. Actual occupancy is not directly measured. Instead, an estimate of the occupancy distribution is made for day and night. For example, probability of being in a bedroom at night is high, while probability of being in the living room is lower. This relationship swaps over during daytime.

The third (energy) compares the overall (externally provided) energy usage (combining gas and electricity) with the number of heating (or cooling) degree days for the same period. Heating degree days (HDD) are defined as the integral over time of the difference between the external temperature and a base temperature. Typically 15.5 °C is used as the base temperature. Roughly speaking, the ratio of heat energy used to HDD is proportional to the specific heat loss of the building.

The number of unprocessed bits provided to obtain these three key measures from a year of monitoring temperature, humidity, and air quality (16 bits each) at 5 minute intervals in 11 locations around a house is $365 \times 288 \times 16 \times 3 \times 11 \approx 5.6 \times 10^7$ bits. Gas and electricity energy information is ignored (and is usually minimal).

After processing of the room-by-room measurement, we have a combination of exposure per room over 5 bands $10 \times 32 \times 5 = 1600$, plus comfort summary $32 \times 2 = 64$, plus degree day and energy summary $32 \times 2 = 64$, or a total of 1728 bits. If this summary information is transferred only once per year, the information reduction would be of the order of $1/32\,000$. A considerable saving is still possible if the information is transferred monthly or weekly.

Interestingly, the key metrics defined above tend to be stable over time. For example, the ratio of energy to degree day, being in rough correspondence to the building heat loss, will tend to remain the same season after season, year after year.

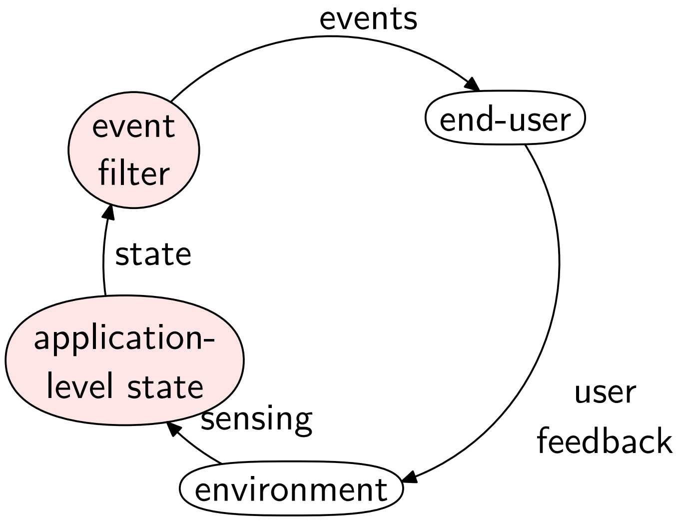


Figure 1. Summary of information flow. Pink shaded processes are performed per node.

Furthermore, any significant change in the value may be of importance. Has a refurbishment improved the insulation? Are new tenants adjusting the heating system parameters? Identifying when key metrics change can be insightful.

Given that transmission of bits is the main energy cost for wireless nodes, the above analysis suggests that performing at least some of the processing on the node will substantially extend the battery life and this is explored in the next section.

## III. Bare necessities as a node-based, on-line algorithm

The overall method proposed here is summarised in Fig. 1. Each node senses the environment and converts the measured values into application-level state. Following the SIP algorithm, an event filter checks for changes in the state beyond some threshold compared to the last transmitted state. When a significant event is detected, a packet is transmitted to the database. Note the calculation of state is online and thus before the full time series is available. An example online calculation for exposure bands is given below.

### A. *Encoding exposure bands*

Exposure bands for temperature, relative humidity and air quality can be thought of as a discrete form of probability distribution. For each room, when a measurement of temperature, relative humidity and air quality is made, a per-band count is incremented if the measurement falls into that band. Let $b_T(i,t), b_H(i,t), b_A(i,t) \rightarrow \{0,1\}$ be predicate functions for temperature, humidity and air quality, respectively, giving 1 if the measurement at time $t$ is in the $i$th band and 0 otherwise. For a finite deployment period (say 1 year) involving $k$ time intervals, the probability that the band is $i$ is simply the average,

$$\frac{1}{k} \sum_{0 \leq t < k} b(i,t).$$

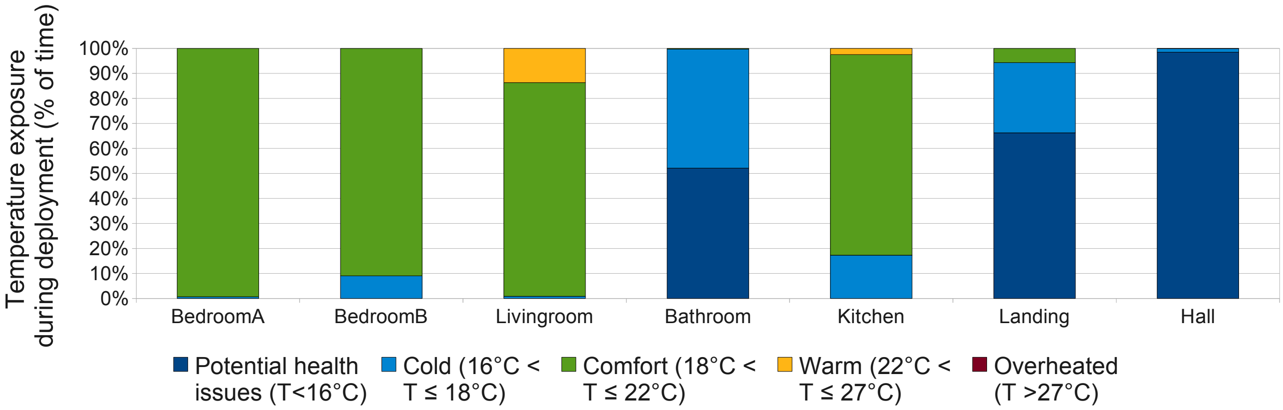


Figure 2. Sample exposure metric for room temperature shown as a stacked bar chart based on monitoring for a two week period.

Table I
BN TEMPERATURE INFORMATION PACKETS TRANSMITTED ON AVERAGE FOR 9 ROOMS.

| | Average Transmissions | | | | |
|---|---|---|---|---|---|
| Half-life | One Week | Two Weeks | One Month | Six Months | Year |
| One Day | 3.33 (± 3) | 8.33 (± 5) | 13.44 (± 6) | 63 (± 13) | 208.22 (± 12.5) |
| One Week | 3.33 (± 3) | 6.22 (± 4) | 7 (± 4.5) | 13.44 (± 6) | 37.56 (± 10.5) |
| One Month | 3.33 (± 3) | 5.78 (± 4) | 6.22 (± 4) | 8.33 (± 5) | 15.11 (± 6.5) |
| Six Months | 3.33 (± 3) | 5.78 (± 4) | 6.11 (± 4) | 7.56 (± 4.5) | 9.44 (± 5) |

However, given that the house will change over time and that the end-user is more interested in current behaviour over past behaviour, a decay can be applied to older measurements based on an exponential decay constant $0 < \gamma < 1$, giving,

$$B_k(i) = \frac{1}{\alpha_k} \sum_{0 \le t \le k} b(i,t)\,\gamma^{k-t},$$

where the normalising value $\alpha_k$ is chosen such that $\sum_i B_k(i) = 1$. Note that the decay half-life is $t_{1/2} = T \ln 2/(1-\gamma)$ where $T$ is the sensing period.

A recursive estimate at time $k$ can be obtained by maintaining an intermediate sum,

$$B^-(i) \leftarrow \gamma B^-(i) + b(i,k)$$

for all $i$, thus giving the distribution,

$$B_k(i) \leftarrow B^-(i) / \sum_i B^-(i)\,.$$

The resulting vector can then be used as the application-level state, which is considered eventful if any element changes by some threshold (e.g., 10%). The above algorithm is referred to here as the BN algorithm and is summarised in Algorithm 1.

**Algorithm 1** Online BN algorithm for estimating exposure band distribution $B$.

1) (*update band count*)
   $B^-(i) \leftarrow \gamma B^-(i) + b(i,k)$,
   for each measurand and for all $i$.
   The predicate function $b(i,k)$ gives 1 if the current reading $k$ is in band $i$ and zero otherwise. The update decays the current count estimate by decay constant $\gamma$ and then increments the currently active band. The decay half-life is $t_{1/2} = T \ln 2/(1-\gamma)$ where $T$ is the sensing period.
2) (*update distribution*)
   $B(i) \leftarrow B^-(i) / \sum_i B^-(i)$,
   for each measurand and for all $i$.
   This converts the counts to a distribution that sums to 1.
3) (*event detect*)
   if, for any $i$, $|B(i) - B'(i)| > \varepsilon$ then
   a) transmit $B$ and
   b) update last transmitted state $B' \leftarrow B$

## IV. RESULTS

An example temperature exposure metric graph is shown in Fig. 2. In this graph, rooms that are comfortable for the two week sample period are clearly distinct from those that are too cold or too hot. Longer sampling periods are required to build a more accurate estimate of the likelihood of exposure to the different temperature bands and a full year (at least) is required to obtain an estimate for all seasons.

Figure 3 shows how a single temperature band ("comfort") evolves over a year for a single room. The standard temperature exposure (Std. T. exp.) percentages for this band are calculated on a non-overlapping two week window. During Summer months, the "comfort" band drops away as the "warm" band takes over. The smoothed estimate of this band (BN) lags the standard two week estimate but provides a better long term estimate for the band. The "BN reconstructed" line shows what the sink will estimate as the value for this band. In this example, a message is transmitted (shown as a vertical

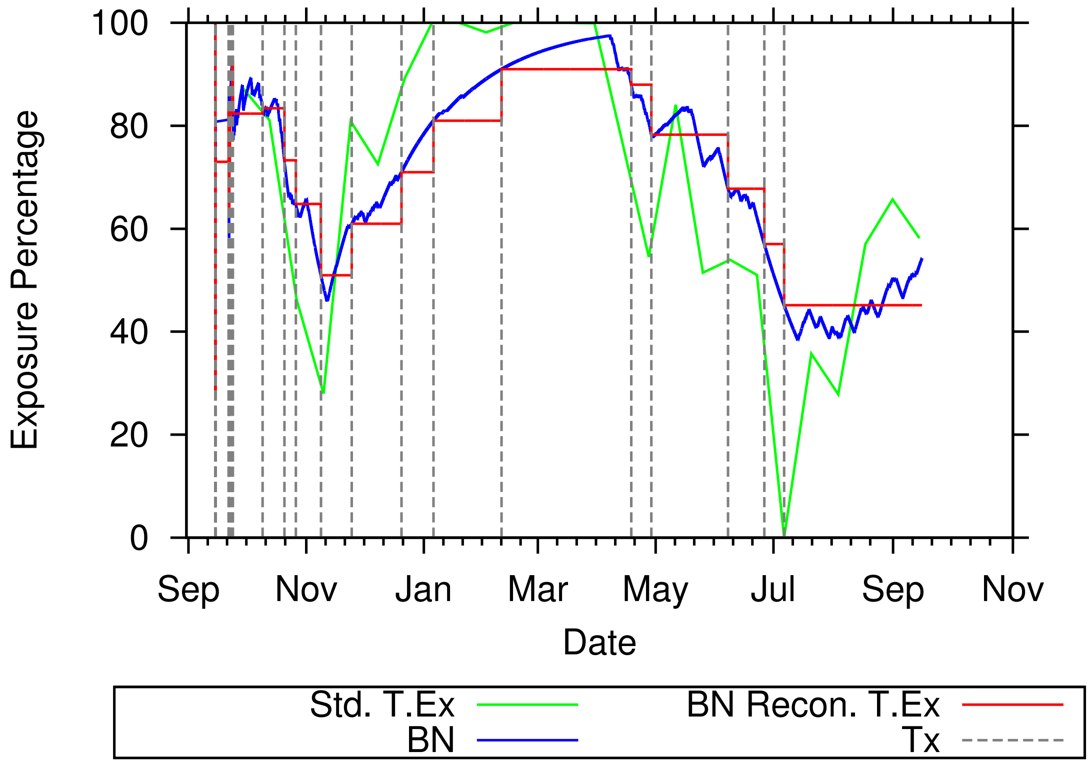


Figure 3. Annual evolution for a single band showing how the BN algorithm reduces transmissions.

Table II
RMSE OF TEMPERATURE BN FOR A 1 YEAR DEPLOYMENT

| Half-life | RMS error in band estimate (%) |
|---|---|
| One Day | 19 ± 1 |
| One Week | 13 ± 3 |
| One Month | 12 ± 4 |
| Six Months | 9.3 ± 9 |

Table III
COMPARING THE PERFORMANCE OF BN ($t_{1/2} = 1$ MONTH) WITH SIP FOR ONE YEAR OF TEMPERATURE DATA.

| | Transmissions | % of raw | RMSE in band % |
|---|---|---|---|
| Raw | 102236 | 100.000% | n/a |
| BN | 15 ± 6.5 | 0.015% | 12 ± 4 |
| Linear SIP | 2900 ± 700 | 2.809% | 0.9 ± 0.2 |

dashed line) when at least one band varies from the sink's estimate by more than 10%.

As shown in Table I, the performance of BN in terms of reducing packets depends on the half-life smoothing parameter and the deployment period. These results are based on a 5 minute sensing cycle (or 288 per day). For longer deployment periods, a larger half-life reduces transmissions more. The error margins shown are based on assuming a Poisson distribution and estimating the 5%- and 95%-iles.

It might be expected for the error to increase with larger half-life smoothing values. However, it was found that the RMSE tended to decrease and this is shown in Table II. The RMSE is calculated based on the difference between temperature exposure band percentages for each two-week non-overlapping window and the reconstructed BN estimate at the same time point. This approach to calculating the error might be slightly unfair to the BN approach as the BN algorithm aims to give an estimate of the long-term band percentage rather than the short term one.

Since the BN algorithm is essentially a derivative of SIP, it is interesting to compare the relative performance of the two algorithms. The performance is shown in Table III and it is clear from this that BN gives a considerable saving in terms of total number of transmissions both over a simplistic sense-and-send approach and the more sophisticated SIP algorithm. In this case, SIP performance was estimated based on a linear model of temperature and assuming that only temperature and rate of change of temperature were transmitted.

A natural extension to BN is to support additional sensing modalities, such as humidity and CO2. Table IV shows that performance is slightly worse than the single modality case but the overall benefit is still considerable. Alongside this, the error, as shown in Table V is no worse and possibly slightly better in some cases. A slight reduction in error might be expected since the algorithm may need to transmit more often for one modality yielding a slightly better estimate for the other modalities at the sink.

## V. CONCLUSIONS

This paper presents an extension to the Spanish Inquisition Protocol (SIP) that focuses the WSN system on transmitting

Table IV
AVERAGE INFORMATION PACKETS TRANSMITTED FOR MULTI-MODAL (TEMPERATURE, HUMIDITY AND CO2) BN

| | Average Sends | | | | |
|---|---|---|---|---|---|
| Half-life | One Week | Two Weeks | One Month | Six Months | Year |
| One Day | 5.11 (± 4) | 19.67 (± 7.5) | 39.56 (± 10.5) | 193.44 (± 12) | 556.44 (± 20) |
| One Week | 5.11 (± 4) | 14.11 (± 6.5) | 17.33 (± 7) | 45.22 (± 11) | 105.89 (± 17) |
| One Month | 5.11 (± 4) | 13 (± 6) | 15.11 (± 7) | 24.11 (± 8) | 45.22 (± 11) |
| Six Months | 5.13 (± 4) | 12.75 (± 6) | 14.5 (± 6.5) | 20.25 (± 7) | 26.75 (± 8.5) |

Table V
RMSE FOR MULTI-MODAL BN

| | RMS error in band estimate (%) | | |
|---|---|---|---|
| Half-life | Temperature | Humidity | CO2 |
| One Day | 17 ± 1 | 19± 2 | 15 ± 10 |
| One Week | 9.1 ± 1 | 10± 1 | 11± 10 |
| One Month | 9.2 ± 1 | 10 ± 3 | 13± 6 |
| Six Months | 10 ± 5 | 12 ± 4 | 14± 5 |

the "bare necessities"—the few bits of information that are needed for the end-user to gain an understanding of the phenomena under study. As with SIP, a key benefit of the approach is that it is *timely*—transmissions occur when the environment changes significantly. The resulting performance gain is significant. While SIP is remarkable in reducing transmitted packets to about 3%, BN improves this further by a factor of over 100 times (to about 0.02%). As with SIP, BN imposes some penalties, such as the need for calculation to occur on the node and a slight loss in the accuracy of the resulting information, however in the context of the application, these penalties are slight. Furthermore, the saving in terms of transmission reduction is multiplied in a multi-hop network due to the need to forward packets for other nodes.

While the BN algorithm is an important contribution in itself, the overall aim of this paper is to demonstrate a general approach to building application-specific knowledge-driven systems that derive *application-level state* at the node. If there is a key roadblock to wider adoption of this approach, it is that end-users and developers still do not have sufficient trust in WSN systems to allow them to go beyond sense-and-send. When WSN hardware and software become more proven and trustworthy, knowledge-driven WSN design will not only improve performance for existing applications but also enable many new ones that were previously unfeasible.

## ACKNOWLEDGEMENTS

The authors would like to thank the Orbit Housing Group for supporting this work through both funding and expertise.

## REFERENCES

[1] M. Ceriotti, L. Mottola, G. P. Picco, A. L. Murphy, S. Guna, M. Corrà, M. Pozzi, D. Zonta, and P. Zanon, "Monitoring heritage buildings with wireless sensor networks: The torre aquila deployment," in *In Proc. of the 8th ACM/IEEE Int. Conf. on Information Processing in Sensor Networks (IPSN). Best Paper Award*, 2009.

[2] T. Schmid, H. Dubois-ferrière, and M. Vetterli, "Sensorscope: Experiences with a wireless building monitoring sensor network," in *In Proc. First Workshop on Real-World Wireless Sensor Networks (REALWSN05)*, 2005.

[3] L. M. R. Peralta, L. M. P. L. de Brito, B. A. T. Gouveia, D. J. G. de Sousa, and C. D. S. Alves, "The wise-muse project: Environmental monitoring and controlling of museums based on wireless sensors networks," in *Electronic Journal of Structural Engineering*, 2009.

[4] E. I. Gaura, J. Brusey, R. Wilkins, and J. Barnham, "Inferring knowledge from building monitoring systems: The case for wireless sensing in residential buildings," in *Proc. Conf. Clean Technology*. NSTI, June 2011.

[5] D. Goldsmith and J. Brusey, "The spanish inquisition protocol: Model based transmission reduction for wireless sensor networks," in *Proc. IEEE Sensors 2010*, 2010.