

CRITICAL ISSUES IN ENERGY-CONSTRAINED SENSOR NETWORKS: SYNCHRONIZATION, SCHEDULING, AND ACQUISITION

Brian M. Sadler

Army Research Laboratory, Adelphi, MD 20783

ABSTRACT

Energy-constrained wireless sensor networks have conflicting requirements between the need to communicate, and the desire to avoid idle listening and thus save energy. Beneath this simple observation lies a complex design space that encompasses all layers of the radio through application. Broad design principles include the desire to exploit resources external to the network, such as beacons and (perhaps mobile) access points that do not have strong energy constraints. We consider synchronization, scheduled communications rendezvous, and packet acquisition. Synchronization accuracy worsens with prediction interval, saves energy by enabling scheduling, but costs energy to maintain. Acquisition is probabilistic due to noise, fading, and synchronization error, with operating points defining energy-performance trade-offs.

1. INTRODUCTION

The field of energy constrained sensor networks (EC-SN) is multi-disciplinary, with a broad variety of applications, sensing modalities, communication rates and ranges, and node densities; see [1] for a tutorial overview and extended reference list. EC-SNs have battery reliant nodes, which implies a finite lifetime, unless energy is available for recharging (e.g., energy harvesting) or new long-life battery technology succeeds (e.g., miniature atomic batteries). The interrelation of so many factors complicates the overall design, but three basic design principles appear clear.

1. *Signal processing viewpoint*: optimize distributed detection/estimation network tasks while minimizing use of communications.
2. *Communications viewpoint*: Support specific network goals while minimizing idle listening.
3. *Systems viewpoint*: Exploit external assets to the greatest extent possible.

The Moore's Law-type behavior of DSP hardware enables increasingly sophisticated signal processing (SP), or

Work partially supported by the DARPA Connectionless Networks program.

equivalently the same operations require less energy. However, perhaps more interesting is the lack of such rapid advance in analog components (RF front end, ADC), which are not exhibiting dramatic reductions in energy consumption. In addition, Shannon and Maxwell's theories reveal fundamental limits on energy required to reliably communicate, and these are not subject to advances in radio technology. (One can increase receiver sensitivity by reducing the thermal temperature, but at a significant energy cost.) The result is that operating the receiver produces a significant energy drain, even when no message is present, and it therefore becomes desirable to duty cycle the receiver, leading to design principle 2, and the caveat in 1.

Following principles 1 and 2 leads to several desirable network properties, often from both an SP and communications point of view. These include knowledge of node location (e.g., for network SP goals, for clustering and routing), and some level of network synchrony (e.g., time of event reporting, scheduling transmission/reception, cooperative relay). Node geolocation algorithms generally require active transmission or exploitation of an opportunistic source, and can take the form of a distributed algorithm or rely on external beacons [1, 2]. Our focus here is on obtaining and maintaining network synchrony to a desired precision, and its impact.

While distributed intra-network algorithms are available for node geolocation and network synchrony, these of course come with communications (and thus energy) costs. Clock drifts (and node motion or addition of new nodes) require synchronization (and geolocation) updates. This motivates the third design principle, that dramatic energy savings can be obtained when external assets are available which are not energy constrained (e.g., continuous power supply, connected to the internet, etc.). Primary examples include beacons for synchrony and geolocation. These could come from stationary or mobile access points, enabling offloading of many tasks from the sensor nodes, leading to simple and cheap designs. This is likely to be a key commercial paradigm, with access points or base-stations linked to fixed communications infrastructure. In military or ad hoc scenarios, introduction of a mobile access point simplifies both communications and signal processing [3].

2. SYNCHRONY

Duty cycling requires that each node maintains a clock, and communications rendezvous implies achieving some desired level of synchrony between nodes. So, in idle state, the node energy E_I may be dominated by the clock requirements, and at high duty cycle E_I dominates the node lifetime. Not surprisingly clock accuracy increases with energy [1], implying a tradeoff between the clock energy consumption per node and the update rate of a network synchronization algorithm. There are ensuing energy tradeoffs involving duty cycle, throughput delay, and node density.

Intra-network synchronization schemes require a message passing protocol, and often rely on message round-trip delay estimation, e.g., see [4, 5]. Synchronization to a single master node is possible, or early-late adjustment can be used to converge to and track a global mean. In the following, we assume an external clock is available, and consider issues of accuracy and the implications for duty cycling and acquisition.

Two clocks running at constant rates can be related simply as $y_i = a + bx_i$, where (x_i, y_i) are the i th simultaneous clock readings, a is the offset, and b is the skew between them. Suppose GPS or other accurate clock is available via broadcast, e.g., from anchored nodes. We model this as

$$y_i = a + bx_i + \epsilon_i, \quad 1 \leq i \leq n, \quad (1)$$

where y_i is the broadcast time subject to error ϵ_i , and x_i is the local clock reading. We assume $\epsilon_i \sim N(0, \sigma_i^2)$, and measurements are uncorrelated ($E[\epsilon_i \epsilon_j] = 0, i \neq j$). The error ϵ_i arises primarily due to variability in the node message processing time, as well as propagation time. Any quantization or other error is absorbed into ϵ_i . We can refine this model to include multiple sources of uncertainty, e.g., on x_i as well as y_i , although it appears that processing time uncertainty will dominate. Experimental evidence exists to support the Gaussian assumption, e.g., see [6], although if ϵ_i is non-Gaussian then robust versions of what follows may readily be applied. Skew b and offset a are, in general, time varying, e.g., due to temperature change and aging. Aging effects are rather slow, and temperature changes can be monitored. In the following we assume a and b constant, keeping in mind that some drift is inevitable and so estimates must be updated.

Motivated by the desire to use a low energy device, the local clock reading x_i is assumed to have less, perhaps much less, accuracy than y_i . We receive n values of y_i and find estimates \hat{a} and \hat{b} , which then may be used to predict future values of the broadcast clock based on the local clock. The problem is in standard linear regression form, and so we can apply many well known useful results including bounds and confidence intervals, distribution, and maximum likelihood estimates via least squares (LS), e.g., see [7].

Denote the vector model as

$$y = X\beta + \epsilon \quad (2)$$

where $\beta = [a, b]'$, $X = [\mathbf{1}_n, x]$, $\mathbf{1}_n = [1, \dots, 1]'$, $y = [y_1, \dots, y_n]'$, and similar for x and ϵ . We assume $\sigma_i^2 = \sigma^2 \forall i$, but note that weighted LS can be used with changing variance. The resulting LS (ML) estimate is

$$\hat{\beta} = (X'X)^{-1}X'y. \quad (3)$$

The estimate $\hat{\beta} = [\hat{a}, \hat{b}]'$ is efficient with variance given by the diagonal of $(X'X)^{-1}\sigma^2$. If σ^2 is unknown then it can be estimated via

$$s^2 = \frac{1}{n-2} \left\{ \sum_1^n y_i^2 - n\bar{y}^2 - b \left[\sum_1^n (y_i - \bar{y})(x_i - \bar{x}) \right] \right\}, \quad (4)$$

where \bar{x} and \bar{y} are the sample means of the elements of x and y , respectively.

Let x_j denote local clock times at which we desire to estimate the corresponding broadcast clock values. Generally, suppose we have m values of x_j in a vector $\tilde{x} = [x_1, \dots, x_m]'$. Then, broadcast clock estimates \hat{y} can be obtained from linear interpolation via

$$\hat{y} = \tilde{X}\hat{\beta}, \quad (5)$$

with $\tilde{X} = [\mathbf{1}_m, \tilde{x}]$.

Suppose that n observations are given. Then, at a specific future time x_0 on our local clock (T seconds beyond the n th observation, say), the prediction of the broadcast clock value y_0 can be written as

$$\hat{y}_0 = \hat{a} + \hat{b}x_0 = \bar{y} + \hat{b}(x_0 - \bar{x}). \quad (6)$$

What accuracy is expected in the prediction for a given n observations? How large can T be for a given desired accuracy? These questions can be answered using confidence intervals on \hat{y}_0 .

The variance of the prediction in (6) is given by

$$\text{var}(\hat{y}_0) = \sigma^2 \left\{ 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_i (x_i - \bar{x})^2} \right\}. \quad (7)$$

If σ^2 is unknown, then s^2 may be used in this expression. The $100(1 - \alpha)\%$ confidence interval around \hat{y}_0 is given by

$$\hat{y}_0 \pm t(\nu, 1 - \alpha/2) \text{var}(\hat{y}_0)^{1/2}, \quad (8)$$

where $\nu = n - 2$ degrees of freedom, and $t(\cdot, \cdot)$ is the t -distribution. For example, for a given future local time x_0 , and $\alpha = 0.05$, then 95% of our random observations of y_0 are expected to lie in the interval defined by (8).

Example: In this example we consider prediction (extrapolation), in order to make a communications rendezvous

as accurately as possible. To motivate, suppose a 1 msec slot is used, then we might wish to obtain future timing accuracy to a fraction of a msec to avoid excess wasted receiver on-time and detection processing. The prediction is made using (6).

Suppose $n = 100$ observations are available, for two cases. (1) every second in $[0, 99]$ secs, and (2) every 10 seconds in $[0, 990]$ sec. Figure 1 plots $t(\nu, 1 - \alpha/2) \text{var}(\hat{y}_0)^{1/2}$ against prediction time for case 1 (top) and case 2 (bottom), with prediction time of zero corresponding to the last observation time (99 or 990 secs). Two measurement error variance cases are shown with standard deviations 3.2×10^{-5} and 10^{-4} ($\sigma^2 = 10^{-8}$ and $\sigma^2 = 10^{-9}$).

The solid curve in the bottom panel shows the best performance with longer observation time (100 measurements in 990 secs or 16.5 minutes) and smaller error variance. The solid line indicates that with confidence of 95%, prediction out to 3,000 seconds (50 minutes) will be within ± 0.1 msec of the actual reading of clock y , and under ± 0.2 msec at 6,000 seconds, which might be reasonable for a 1 msec slot time. The worst performance case is shown in the top curve of the top panel, based on 100 measurements in 100 seconds, where prediction error after 500 secs (8.3 minutes) exceeds ± 0.4 msec. \square

In situations with prolonged network inactivity, coarse synchrony can be maintained over quiescent periods. Then, upon initiation of network traffic, a tighter synchrony can be reestablished. Transmission duration can be adaptively extended to ensure high probability of detection at the receiver, with the transmission duration required growing with prediction time T .

The analysis of this section can be more generally applied. One interesting case that is directly covered is when a duty-cycled GPS receiver (y) is incorporated within a node, along with a low power clock (x) that is running continuously. In this case σ^2 will be small, because both x_i and y_i are available within the same node.

3. ACQUISITION

When the receiver awakens for communications rendezvous, it must determine if a packet is present and if so demodulate it, subject to uncertainties in timing, carrier frequency, and phase. The oscillator employed for communications may in many cases be the same device as the local clock, in which case both timing and carrier offset are related. Alternatively, a second oscillator may be spun up just prior to rendezvous. To study the acquisition problem and find guidelines on desirable accuracy of timing and frequency synchronization between transmitter and receiver prior to rendezvous, consider a QPSK-DS (direct sequence) system, employing a non-coherent correlator detector, e.g., see Viterbi [8]. The non-coherent approach enables detection and timing acqui-

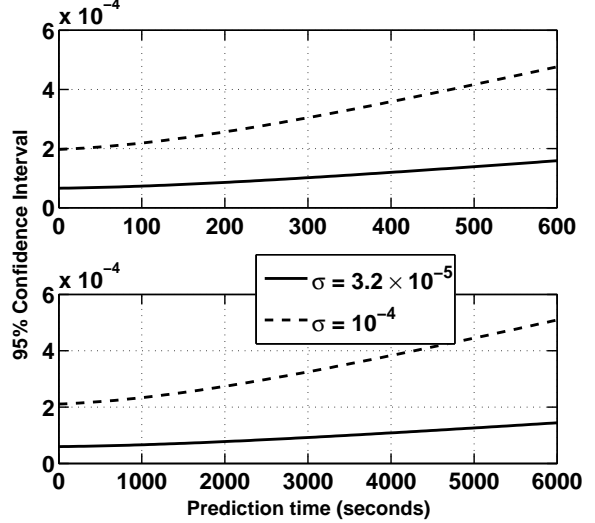


Fig. 1. Confidence intervals for time prediction.

sition (chip synch) prior to carrier frequency and phase estimation, so long as the carrier offset is not too large as quantified below. For example, a training preamble or superimposed training sequence might be employed.

Let f_c be the nominal carrier frequency, and local reference denoted $\cos 2\pi(f_c + \Delta f)t + \phi$, with Δf and ϕ unknown to the receiver. From the non-coherent correlator we obtain $\gamma = \gamma_I^2 + \gamma_Q^2$, where γ_I and γ_Q are the outputs of the I and Q channels, respectively. For the noise-free case

$$E^2[\gamma_I] + E^2[\gamma_Q] = N^2 E_c \cdot D(\Delta f), \quad (9)$$

where

$$D(\Delta f) = \left[\frac{\sin(\pi N T_c \Delta f)}{\pi N T_c \Delta f} \right]^2. \quad (10)$$

Here, the spreading sequence has N chips, each of duration T_c seconds and energy E_c , so γ is a measure of received energy. $D(\Delta f) \leq 1$ is a degradation factor caused by frequency mismatch Δf , and γ does not depend on ϕ . To maximize $D(\Delta f)$, we want $N T_c \Delta f$ as small as possible. To keep the detector degradation constant, with $\Delta f \neq 0$, an increase in sequence length N requires a corresponding increase in bandwidth $1/T_c$. This is because, as Δf grows, the degradation worsens proportional to the total sequence time $N T_c$.

We find $\text{sinc}^2(.44) \approx 1/2$, so we desire $\Delta f \leq .44/N T_c$ for degradation ≤ 3 dB. Serious degradation occurs when $\Delta f \leq 1/N T_c$. If Δf is too large, then the receiver must search over frequency offset by computing more correlations, i.e., by searching the (non-coherent) ambiguity surface. For relatively small N (of order 10^3 or less, say), all correlation lags may be computed simultaneously by employing the FFT, and searching over Δf requires additional DSP energy via multiplication and IFFT operations.

Implicit in the preceeding is the assumption that when the receiver is activated a packet will be arriving, which will not always be the case. So, we next consider detection as characterized by probability of detection P_D for a given probability of false alarm P_F . A false alarm results in further unnecessary receiver processing (e.g., demodulation, decoding) before rejection of the signal present hypothesis, hence energy wasted. Missed detection requires retransmission (depending on network quality of service) and so will trigger more significant energy expenditure.

Assume Δf is small, in a slow Rayleigh fading channel, i.e., the fading is constant over NT_c but independent from realization to realization. The latter occurs even in static networks in the presence of moving scatterers (e.g., moving trees, cars, people). For the non-coherent QPSK-DS system,

$$P_F = e^{-\theta/V}, \quad (11)$$

and

$$P_D = e^{-\theta/V_f} \quad (12)$$

where $V = N\sigma^2$, σ^2 is the additive white Gaussian noise variance, and θ is the detection threshold set to establish P_F . Multi-user interference can be readily incorporated if a Gaussian assumption is employed. $V_f = V(1 + \bar{\mu})$, where

$$\bar{\mu} = \frac{N\bar{E}_c R^2(\tau)}{\sigma^2} \quad (13)$$

is the average SNR with degradation due to small timing error τ ,

$$R(\tau) = \frac{\sin \pi \tau / T_c}{\pi \tau / T_c}, \quad (14)$$

and \bar{E}_c is the average energy per chip of the Rayleigh faded signal. Small τ implies the timing is known to within one chip; in practice many lags must be computed, perhaps via FFT, with the correlation peak having the above characteristics.

Example: Figure 2 plots P_F vs P_D , for $N = 100$, $\sigma^2 = 1$, and average SNR $\bar{\mu}$ varied through \bar{E}_c . For constant $\bar{\mu}$ the ROC curve is generated by varying θ . Here $\tau = 0$, so timing degradation is not included. The effect of fading is apparent, because high average SNR is needed to achieve high P_D with low P_F . Potentially much more energy is wasted for a missed detection (requiring retransmission) than a false alarm, so a relatively high P_F may need to be tolerated. \square

4. CONCLUSIONS

Synchronization of wirelessly connected nodes comes with a relative accuracy that worsens as the prediction interval grows. Synchronization saves energy by enabling receiver duty cycling, but costs energy to maintain.

Packet acquisition is probabilistic due to noise, fading, and synchronization error, with performance described by the receiver operating curve. While both types of detection

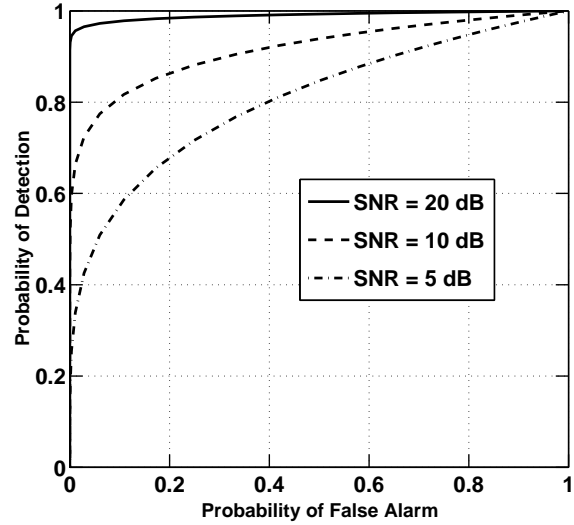


Fig. 2. Receiver operating curves (ROC) in Rayleigh fading.

error incur energy loss, missed detection may be much more costly than detector false alarm.

Battery driven networks that are fully stand-alone require significant resources for geolocation, synchronization, and other basic tasks that must be performed in order for the network to function. When available, external resources can relieve the network in terms of node design and complexity, as well as energy, and therefore greatly extend the useful network lifetime.

5. REFERENCES

- [1] B. M. Sadler, "Fundamentals of energy-constrained sensor network systems," *IEEE AES Magazine*, tutorial supplement, 2005.
- [2] B. M. Sadler, R. J. Kozick, L. Tong, "Multi-modal sensor localization using a mobile access point," *Proc. ICASSP'05*.
- [3] P. Venkatasubramanian, S. Adireddy, L. Tong, "Sensor networks with mobile access: optimal random access and coding," *IEEE Jnl. Selected Areas in Comm.*, August 2004.
- [4] S. Ganeriwal, R. Kumar, M. B. Srivastava, "Timing-sync protocol for sensor networks" *ACM Conf. on Embedded Networked Sensor Systems (SENSYS 2003)*.
- [5] F. Sivrikaya, B. Yener, "Time synchronization in sensor networks: a survey," *IEEE Network*, July/August 2004.
- [6] J. Elson, L. Girod, D. Estrin, "Fine-grained network time synchronization using reference broadcasts," *Proc. Fifth Symp. on Oper. Syst. Design and Impl. (OSDI'02)*, Dec. 2002.
- [7] R. G. Miller, *Simultaneous Statistical Inference*, 2nd ed. (Springer, 1981).
- [8] A. J. Viterbi, *CDMA: Principles of Spread Spectrum Communication* (Wesley, 1995).