Predicate Signaling in Distributed Sensor Networks

Agostino Capponi
Department of Computer Science
California Institute of Technology
acapponi@cs.caltech.edu

K. Mani Chandy
Department of Computer Science
California Institute of Technology
mani@cs.caltech.edu

Ibrahim Fatkullin
Department of Mathematics
The University of Arizona
ibrahim@math.arizona.edu

ABSTRACT

Nodes in a sensor network can generate messages periodically, or when anomalies are detected, or when queried by other nodes. In this paper we propose a strategy called predicate signaling that generalizes these schemes by generating messages when specified predicates - that can deal with both time and anomalies - hold. We show how power consumption, message generation rates and estimation errors can be controlled by choosing predicates appropriately. We compare predicate signaling with other schemes. We derive formulas based on stochastic differential equations to estimate performance measures in predicate signaling. We analyze measurement data, and we compare simulations based on measured data with results predicted by our theory.

I. INTRODUCTION

A. Problem Overview

In this paper we study mechanisms for reducing communication in distributed systems that obtain information from sensors and other sources of data, estimate the system state by processing this information, and execute responses appropriate to the estimated state. Many applications require low communication rates for reasons such as energy constraints, secrecy, reduced electromagnetic interference and limited bandwidth. We also discuss the concomitant effects of reducing communication on computation and bandwidth requirements.

Specifications: Distributed systems that respond to conditions in the environment can be specified by a set of rules of the form: **when** a predicate P begins to hold **then** execute action e, where P is a predicate on the history of global states of the system [11], [4]. The system initiates action e when the value of predicate P changes from false to true (the system may have another action e' that is executed when $\neg P$ changes value from false to true). The specification also includes quality of service and accuracy requirements; these aspects are discussed later. Examples of systems that respond to conditions include those that detect and intercept intruders, detect and warn when a tsunami is likely to hit, and detect and respond to situations that require control-law changes in multi-agent control systems. Systems specified by when-then rules are called *sense and respond* systems [11].

Errors: Each node has access only to its own local state. A node *estimates* the global state by fusing local state information sent to it by other nodes. Nodes can exactly compute only certain types of predicates on global states [12]; estimates of other predicates may be incorrect. Consider a node w responsible for initiating action e when the value of a predicate P changes from false to true. A *false positive* arises if node w's estimator for P has value true while P has value false. Let d be the delay from the point in time at which P begins to hold to the point in time at which node w initiates an action. During the delay interval, the estimated value for P is false while P is true; thus, *a false negative* holds for the period of the interval. Design specifications (discussed later) include maximum acceptable rates of error.

Next, we discuss different schemes by which nodes in sensor networks communicate with each other, and then propose a common framework - predicate signaling - that unifies these schemes and allows for systematic analysis of tradeoffs between them.

B. Common Signaling Schemes

Time-Driven Signaling Consider a node w of a distributed system that executes an action e when its estimate of a predicate P becomes true. Typically P is a function of the local states of other nodes. In time-driven signaling, nodes send local state information that node w needs to estimate P periodically. The optimal period is determined by trading off costs of erroneous estimates against the costs of more frequent messages and computation. While time-driven signaling is appropriate for applications such as data gathering, it is inappropriate if systems are required to respond to only rare events, such as tsunamis, in which case P rarely becomes true. For example, if a system turns on backup power generators when a brownout is imminent, and if this situation arises only when temperatures exceed 100 degrees, there is little value in sending temperature measurements every minute while the temperature is below 95 degrees.

Anomaly-Based Signaling A goal for a communication strategy is: *nodes should communicate only when they have to*. Anomaly-based signaling helps in coming closer to this goal by adopting the scheme discussed next.

Predictive models that track measurements exist for many applications including weather, power grids, stock markets, intruder behavior, and virus propagation. The model that predicts a parameter — such as amount of rainfall — most accurately, may be different under different global conditions such as when a hurricane is approaching or when there is

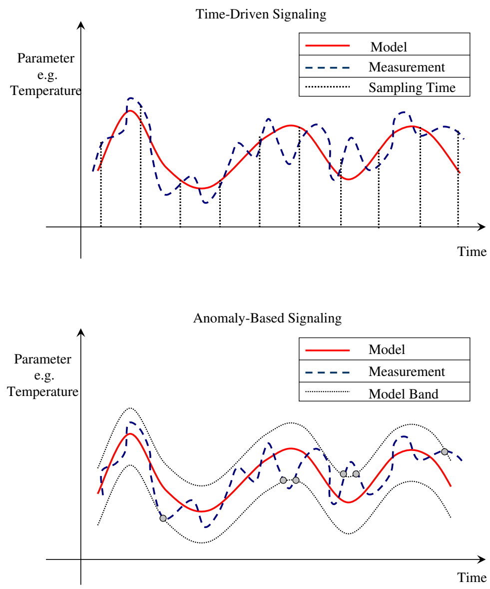


Fig. 1. Time Driven Signaling and Anomaly-Based signaling schemes

high pressure over nearby deserts. Accurate predictive models can be used to reduce the volume of communication between nodes by using a protocol described in the next couple of paragraphs.

Consider communication from a node v to a node w. Some system parameters are observable by both nodes; for example, both nodes may have local clocks that give the same time (to a sufficient degree of accuracy), and thus time is observable by both nodes. Parameters that are at least partially observable by multiple nodes are not limited to time; for instance multiple nodes may be able to observe signals sent by a satellite. Some parameters that are observable by node v may not be observable by w; for instance, if nodes v and w are far apart, w may not be able to observe the temperature at v. For brevity we use the term v's *local* parameters for parameters that v can observe but w cannot. The communication strategy uses models that predict v's local parameters given the parameters observable by both v and w. Associated with communication from node v to node w is a set of such predictive models.

At any given time, node v uses one of the models, say model M, in its set. Node v sends messages to w only when: (1) the values that v measures of its local parameters deviate by more than a specified threshold from the values predicted by the model M that v is using currently, or (2) v changes the model M that it is using. These messages, which we call signals, include the current measured values of v's local parameters and the model that v uses from that point onwards until the next message indicating a model change. Node w estimates v's local parameters from (1) the model M that v is currently employing, (2) the parameters that both v and w can observe, and (3) the messages that w has received.

The rate at which messages are sent decreases with the predictive accuracy of the model. An advantage of anomaly-based signaling is that *the absence of signals conveys information*: namely the information that measurements match the current model. By contrast, the absence of signals never conveys information in (traditional) models of distributed systems - see theories of process knowledge and learning in traditional distributed systems [10], [6], [13] - because these models do not deal with real time. However, the problem of estimation becomes difficult when measurements are noisy and this issue is discussed later.

An illustration of the two signaling techniques is shown in figure 1. The first figure shows time-driven signaling where values are measured and signals are sent periodically. The next figure shows a model that predicts a local parameter, say temperature, of a node v, as a function of time - a parameter that is observable by another node w. Associated with this model is a tolerance band: a message is sent by v to w only when v's measurement of its temperature crosses this tolerance band. This signal includes the value of v's measurement and the model that v starts to use from that point onwards. Node w estimates v's temperature given v's current model, the current time (observable by both nodes), and messages received by w from v.

In some cases, an anomalous situation may be representable as a composition of simpler anomalous situations. For example a power brownout is likely when temperatures are very high **and** when power lines are saturated. In such cases, a predicate P is expressed as a composition, such as $P = Q \wedge R$ of predicates Q and R. Different nodes estimate the different predicates P, Q and R, and send signals when the estimates of their predicates change; for instance, the node that estimates P does so using the the most recent estimates of Q and R that it receives.

Anomaly-based signaling is an example of collaborative signaling processing [9]. Both the sender and the receiver collaborate on using the same set of models to predict unobservable values from observables.

A potential criticism is that accurate models may require a lot of computation, and so what is saved in communication may be lost in computation. In many applications, such as weather, the model maps time (which is assumed to be the only observable parameter) to the unobservables. Such models can be represented by table lookups where values in tables are precomputed. Furthermore, if the model maps time to an unobservable value then we know which value in the table will be looked up next — indeed it is the next value in the table. In this case table lookup can be implemented on an inexpensive medium such as tape.

An alternate scheme that is not reported in this paper is to use another simple model: assume that the last value of the signal received is the best predictor of the future. Initial experiments with weather data suggests that this works well. An advantage of such a model is that no storage or computation is required: a sensor generates a message only when the measured value deviates from the value of the last signal generated.

Query-Based Signaling In some applications, a sensor network sends messages when it receives queries. An important performance metric is the delay between the query and the response [2], [8]. The network behaves as a database containing the current and historical values of system parameters. In the context of sense and respond systems, a controller generates queries for the sensor network when a predicate P, to which the controller is required to respond, is likely to change value from false to true. Query-based signaling can be integrated into the predicate signaling framework by making the response e (to a predicate becoming true) the generation of a query.

C. Predicate Signaling

Predicate signaling is a generalization of the signaling schemes, discussed above, that allows designers to make design choices that tradeoff relative advantages of different signaling schemes within a common framework. Moreover, predicate signaling is a natural framework for sense and respond systems specified by when-then rules; when a predicate P becomes true then execute an action e where the action for a sensor or fusion node is the generation of a signal. Next we show how predicate signaling generalizes other schemes.

Periodic signaling with a period D is a special case of predicate signaling in which the predicate is to send messages every D time units.

$$\exists \mathrm{integerk} :: \mathrm{t} = \mathrm{k} \times \mathrm{D}$$

Anomaly-based signaling is a special case in which the predicate is to send a message when an anomaly occurs. Query-based signaling is incorporated into predicate signaling in the following way: The signal sent when a predicate becomes true is a query to the sensor network to send additional information. The optimal policy for a node v, that is responsible for initiating a response when a predicate P becomes true, may be to request additional information if v has received no messages for such a long time that its estimate of P is likely to be inaccurate.

Predicate signaling is not better or worse than other schemes; it is a unified framework for systematic analysis of tradeoffs.

Problem Setting for Predicate Signaling The *history* of a system at a point T in time specifies the (global) system state at each time t from the instant ($t = 0$) of system initiation to the time $t = T$. The history is a function from time t where $0 \leq t \leq T$ to system states.

We are given the following:

1) A set of pairs (P, e), where P is a predicate on the history of global states of a system and e is an action.
2) For each pair (P, e) the net benefits of initiating action e with delay d after P becomes true.
3) For each pair (P, e), the costs of executing action e while P is false.
4) Constraints on energy, message bandwidth and computing capacity.

The problem is to maximize the net benefit per unit time subject to the given constraints. A definition of net benefit is the difference between the benefits of executing appropriate actions and the costs of executing inappropriate actions. In this short paper we have space to investigate only a very small part of the overall problem: namely we give analytical and simulation results of how the estimation error and the message rate depends on the adopted signaling scheme.

II. Design Issues

Next we list some of the relative advantages of different signaling schemes and then discuss how predicate signaling allows designers to make tradeoffs between them within a unified framework.

1) **Reduction in average load**: If predictive models are accurate then communication rates are lower in anomaly-based signaling than in time-driven signaling for the same degree of error. Also, the rate of generation of queries in query-based signaling can be reduced by querying for additional measurements only when local measurements deviate from the current model.
2) **Reduction in energy**: Lower communication rates consume less energy. Anomaly-based signaling conserves energy for the times when it is needed: times when reality doesn't fit any model.
3) **Increase in load variance**: The loads on communication and computational resources are more stable in time-driven signaling systems than in anomaly-based signaling systems which may have long periods of inactivity punctuated with bursts of frenetic activity.
4) **Greater consequence of lost messages**: A single message loss can be catastrophic in anomaly-based signaling. Consider the consequence of the loss of a message from a node v to a node w where the message contains (1) the information that v's estimate of a predicate P has become true and (2) the new model that v starts to use from that point onwards. Then node w will estimate the value of P incorrectly and use the wrong (old) model in its estimation. There will be no subsequent message from v while v's measurements match v's current model. Node w will continue to use the wrong model until it receives a subsequent message from v. This incorrect use of the wrong model can persist for an arbitrarily long time. By contrast, in periodic signaling, if a single message is lost, the error can be repaired by the message sent in the next period. Since the cost of lost messages is high in anomaly-based signaling more computationally expensive communication protocols are used than for time-driven signaling; for instance, senders resend messages repeatedly until an acknowledgment is received.
5) **Computational impact**: In many cases anomaly-based signaling requires less computation, averaged over time, than time-driven signaling.

By suitably combining predicates from time-driven and anomaly-based signaling, we can choose design points in between time-driven and anomaly-based signaling and thus make tradeoffs between them. For instance we can reduce communication and computational requirements while limiting the consequences of lost messages and load variance. Also, by generating queries for additional measurement data only when certain predicates hold, and specifying these predicates appropriately, designers can control message generation rates while limiting errors.

III. Filter for Predicate Signaling

Predicates are estimated from measurements, and the measurements may have error; this error is usually called *measurement noise*. Predicate estimation is a problem in filtering [14]: estimating true values from noisy measurements. A huge body of literature exists on filtering with time-driven signaling, beginning with the Kalman filter [14]. There are substantial differences in the algorithms for filtering when signals are generated on a time-driven basis and when signals are generated only when a predicate begins to hold. To illustrate the differences between the algorithms we review, in a few lines, the essential ideas of the Kalman filter and then discuss the new challenges introduced by predicate signaling.

Kalman Filter and New Challenges In the simple Kalman filter, we assume that the *a priori* probability density of the parameter of interest and the noise of the measuring device are normal. Applying recursively Bayes' Law, the normal *posteriori* probability density can be efficiently derived, and the mean and variance of the *posteriori* density depends on the means and variances of the *a priori* density and the measurement noise. The new problem introduced by predicate signaling is *conditioning*: the algorithm estimates the distribution of the true value *given* that measured values satisfied a specified model ($\neg P$ holds) between signals. This conditioning leads us to a different technique for estimation and different algorithms.

A. Representation of Noise as Stochastic Processes

We begin by studying a model of noise suitable for a filter employable with anomaly-based signaling; later we extend the filtering algorithm to predicate signaling.

Model Noise: We define the noise of a model M, at a time t, to be $f(t)-g(t)$: the difference between the true value $f(t)$ of a parameter at time t and the value $g(t)$ predicted for that parameter by model M at t. We also use the term process noise for model noise.

Measurement Noise: We define measurement noise at time t to be $m(t)-f(t)$: the difference between the measured value $m(t)$ of a parameter and its true value $f(t)$ at time t.

A signal is generated when the measurement deviates from the current model by more than a given threshold, i.e., when $m(t)-g(t)>\sigma$ where σ is the threshold. Thus a signal is generated when the measured value (model noise plus measurement noise) exceeds the threshold.

Our problem is to estimate the true value $f(t)$ of a parameter at time t given the signals received up to that point in time. This problem is equivalent to estimating the model noise $f(t)-g(t)$ at time t, since the true value $f(t)$ can be obtained by summing the model noise and the model prediction $g(t)$.

Since we assume that measurements are being taken continuously, a reasonable mathematical abstraction for measurement noise and model noise is that they are continuous stochastic processes. A simple choice for this abstraction is the Ornstein-Uhlenbeck (OU) process [7] or the Wiener process. An OU process can be specified by a stochastic differential equation:

$$dX_t = -\alpha X_t + \beta dW_t \tag{1}$$

where W_t is an independent Wiener process, and α and β are non-negative (usually positive) constants. The value of an OU process, at any time, is a normal random variable, with mean 0 (zero) and with a finite variance provided α is positive. The value of a Wiener process, at any time, is a normal random variable, with mean 0 (zero) but its variance increases without bound.

We represent model noise by an OU process because we expect to use accurate models and so the variance of model noise does not increase without bound over time. We could represent measurement noise either as an OU process or as a Wiener process; for this analysis we use the simpler Wiener process. This assumption is valid provided predicates are used to ensure that the time between signals is not too large.

Let $X_t^{(1)}$ be the measurement noise, and let $X_t^{(2)}$ be the model noise. The measurement and model noise processes are described by the following pair of stochastic differential equations:

$$\begin{cases} dX_t^{(1)} &= adW_t^{(1)} & \text{(measurement noise)} \\ dX_t^{(2)} &= -\alpha X_t^{(2)}dt + \beta dW_t^{(2)} & \text{(model deviation)} \end{cases} \tag{2}$$

where $W_t^{(1)}$ and $W_t^{(2)}$ are independent Wiener processes, a, α and β are parameters. Under the mapping of time $a^2 t \to t$, we can rewrite the system of equation (2) as

$$\begin{cases} dX_t^{(1)} &= dW_t^{(1)} & \text{(measurement noise)} \\ dX_t^{(2)} &= -\kappa X_t^{(2)}dt + \epsilon dW_t^{(2)} & \text{(model deviation)} \end{cases} \tag{3}$$

where $\kappa = \frac{\alpha}{a^2}$ and $\epsilon = \frac{\beta}{a}$.

The solution of the system of differential equations (3) is a diffusion process on $\mathbb{R}^2$ with generator

$$\widehat{\mathcal{L}} := -\kappa x_2 \frac{\partial}{\partial x_2} + \frac{1}{2}\left(\frac{\partial^2}{\partial x_1 \partial x_1} + \frac{\partial^2}{\partial x_2 \partial x_2}\right) \tag{4}$$

B. Conditional probabilities given predicates

Initially restrict attention to a simple predicate: the difference between measured and modeled values exceeds a constant. We call this constant σ. Thus, the predicate is:

$$\exists k \in \mathbb{N} : |x_t^{(1)} + x_t^{(2)}| \geq k\sigma$$

We introduce the following mathematical model. The plane $\mathbb{R}^2$ is split into (overlapping) domains $\Omega_k = \{(x_1, x_2) : |x_1 + x_2 - k\sigma| < \sigma\}$, $k \in \mathbb{Z}$. Observe that the boundary $\partial\Omega_k$ consists of two disconnected components (parallel lines) which belong to Ω_{k-1} and Ω_{k+1}; we denote them by Γ_{k-1} and Γ_{k+1} respectively. An illustrative example, including few domains is presented in Figure 2.

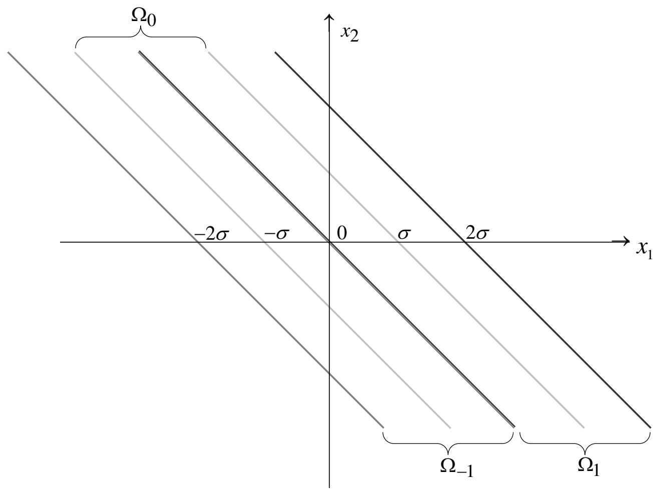


Fig. 2. Domains Ω_{-1}, Ω_0 and Ω_1

We solve the following problem: given an increasing sequence $0 = t_0 < t_1 < \ldots < t_n$ of hitting times and a sequence n_i ($n_{i+1} = n_i \pm 1$), reconstruct the probability distribution for $\boldsymbol{X}_t$ conditional on the event $\{\tau_i = t_i, k_i = n_i, i = 1 \ldots n, \tau_{n+1} > t\}$. Our approach is to employ the Fokker-Planck equation. Since $\boldsymbol{X}_t \in \Omega_{k_i}$ for $t \in (t_i, t_{i+1})$, the probability density $p_t(x)$ that $\boldsymbol{X}_t$ is found at $\boldsymbol{x}$ satisfies the Fokker-Planck equation:

$$\begin{cases} \partial_t p_t(\boldsymbol{x}) & = & \widehat{\mathcal{L}}^* p_t(\boldsymbol{x}) \\ p_t(\boldsymbol{x}) & = & 0 \qquad\qquad \text{if } \boldsymbol{x} \in \partial\Omega_{K_i} \end{cases} \tag{5}$$

Here $\widehat{\mathcal{L}}^* = \frac{1}{2}\left(\frac{\partial^2}{\partial x_1 \partial x_1} + \epsilon^2 \frac{\partial^2}{\partial x_2 \partial x_2}\right) + \kappa \frac{\partial}{\partial x_2} x_2$ is the adjoint of the generator $\widehat{\mathcal{L}}$; the initial condition $p_{t_i}(\boldsymbol{x}) = q_i(\boldsymbol{x})\delta_{\partial\Omega_{k_{i-1}}}(\boldsymbol{x})$ is the single layer density on $\partial\Omega_{k_{i-1}} \subset \Omega_{k_i}$ corresponding to the distribution of exit locations from the previous domain $\partial\Omega_{k_{i-1}}$. Let λ_n and $\phi_n(x_1, x_2 - k\sigma)$ be the eigenvalues and the eigenfunctions of $\widehat{\mathcal{L}}^*$ in the strip Ω_k with zero boundary conditions (observe that k only appears through translation of x_1). Let us label the eigenvalues as $0 \geq \lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_n$. Let φ_n be a set of functions such that for any pair (i, j) it holds that

$$\int_{\Omega_{k_i}} \phi_i(x_1, x_2 - k\sigma)\varphi_j(x_1, x_2 - k\sigma) d\boldsymbol{x} = \delta_{i,j}$$

In other words, $(\varphi_n, \phi_n)_{n \in \mathbb{N}}$ form a bi-orthogonal set. Furthermore, assume that $\phi_0(\boldsymbol{x})$ is normalized to have integral one. Using the orthogonality we can write any solution of equation (5) as

$$p_t(x_1, x_2) = \sum_{n=0}^{\infty} p_n^{(i)} e^{\lambda_n (t - t_i)} \phi_n(x_1, x_2 - k_i\sigma) \tag{6}$$

where the coefficients $p_n^{(i)}$ are related to the initial conditions $q_i(\boldsymbol{x})$ as

$$p_n^{(i)} = \int_{\Omega_{k_i}} p_{t_i}(\boldsymbol{x})\varphi_n(x_1, x_2 - k_i\sigma) d\boldsymbol{x} \tag{7}$$

The equality above follows by expanding p_{t_i} using the definition given in equation (6) and then using the fact that $(\varphi_n, \phi_n)_{n \in \mathbb{N}}$ form a bi-orthogonal set. Since $p_{t_i}(\boldsymbol{x}) = q_i(\boldsymbol{x})\delta_{\partial\Omega_{k_{i-1}}}$ we can rewrite the integral as

$$p_n^{(i)} = \int_{\partial\Omega_{k_{i-1}}} q_i(\boldsymbol{x})\varphi_n(x_1, x_2 - k_i\sigma) ds, \tag{8}$$

We can finally compute $p_m^{(i+1)}$ using the flux operator $\widehat{\mathcal{F}}$ (see Appendix A) and obtain:

$$\begin{cases} p_m^{(i+1)} & = & \sum_{n=0}^{\infty} p_n^{(i)} e^{\lambda_n (t_{i+1} - t_i)} T_{m,n} \\ T_{m,n} & = & \int_\Gamma \widehat{\mathcal{F}} \phi_n(x_1, x_2) \boldsymbol{n}(\boldsymbol{x}) \varphi_m(x_1, x_2 \pm \sigma) ds \end{cases} \tag{9}$$

where $\Gamma = \{\boldsymbol{x} : x_1 + x_2 = \sigma\}$: the "$\pm$" sign corresponds to $k_{i+1} = k_i \mp 1$. Thus the whole problem is effectively reduced to the computation of the *transfer matrix* $T_{m,n}$ and of the eigenvalues $\lambda_{m,n}$.

IV. Theory and Measurements

In this section we evaluate predicate signaling in the setting of habitat monitoring. This problem has received much attention from the sensor network community [1], [3]. We fit our models to measurements of weather data extracted from the historical Integrated Surface Hourly database. The data gives parameters such as temperature and pressure at hourly intervals at different locations over 5 years. We use a single *simple predictive model* to illustrate a point: if there are benefits with using even a single simple model then there surely are benefits with using a set of accurate models, where the model most appropriate for each point in time is employed. the predicted parameter is the average over the 5 years. For instance, the predicted pressure at December 25, 9 AM is the average of the measured pressure at December 25, 9 AM, over the 5 years. The benefits of predictive signaling improve with the quality of the predictive models, and so if there appear to be benefits with even a single simple model then there surely are benefits with a set of sophisticated models.

Next we show that the Ornstein-Uhlenbeck and Wiener processes are satisfactory models for the data, and then we present experimental results concerning communication requirements and estimation accuracy.

A. Fitting of Data

We estimated the parameters of the pair of stochastic differential equations (2) that fits data at a given location and for a given parameter (such as temperature or pressure), as follows. For the purposes of fitting the model to data we assume that there is no measurement noise. We first estimate the parameters κ and ϵ of equation (3) that specify the stochastic process for model noise. We do so by computing

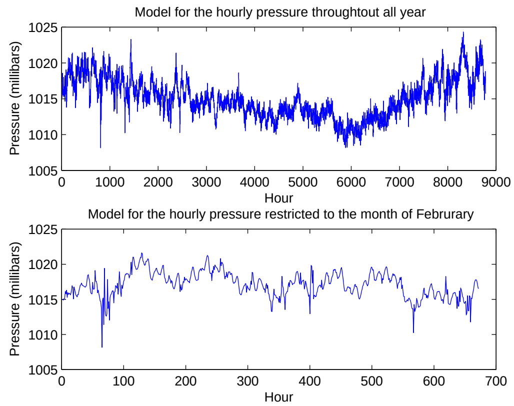


Fig. 3. Model of the hourly pressure for the location of S.Diego North Island

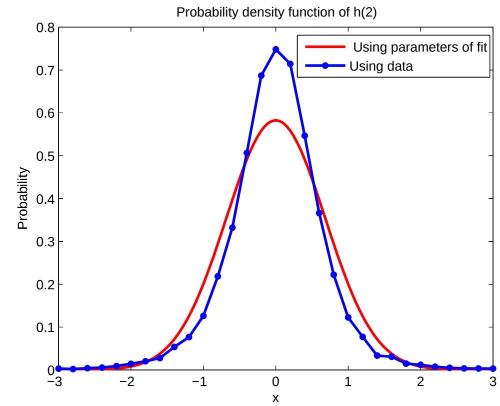


Fig. 5. Estimated probability density function of the random variable $h(2)$ calculated using sea-level pressure observations reported by sensors located at S.Diego North Island. The gaussian probability density function of $h(2)$ computed using the process parameters $\epsilon = 0.4972$ and $\kappa = 0.027$ has mean 0 and variance 0.4783

the distribution of the difference $h(\tau)$ between model noise values at two times separated by a duration τ:

$$h(\tau) = X^{(2)}_{t+\tau} - X^{(2)}_t$$

We compute the correlation function $c(\tau)$ for the parameter from the measured data. Then we compute the values of κ and ϵ so that $h(\tau)$ best fits $c(\tau)$.

Due to space limitations we only report results obtained using pressure data collected at a single location (San Diego North Island). All other experiments in this paper refer to this data source. For the other locations that we tested we determined that the quality of the fit is approximately as for this location with parameters: $\epsilon \in [0.2, 0.7]$, $\kappa \in [0.01, 0.1]$.

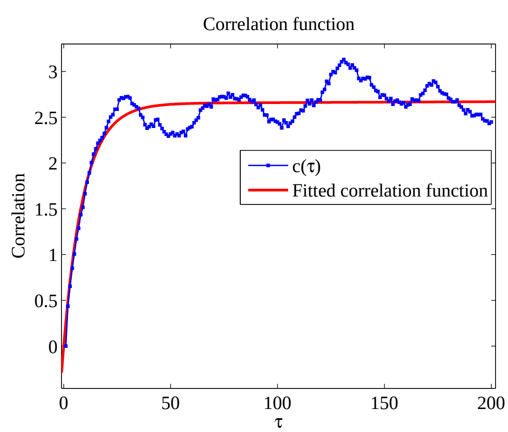


Fig. 4. Correlation function obtained using hourly sea-level pressure measurements produced by sensors located at S.Diego North Island. The fitted parameters are $\epsilon = 0.4972, \kappa = 0.027$

Next consider estimating the parameter a which specifies the degree of measurement noise. In the data, pressure measurements are produced by an instrument which consists of redundant digital pressure transducers. Brownian noise is usually the dominant noise component though flicker noise and thermal noise also contribute for this transducer. The accuracy of the sensor is ± 0.02 inches of mercury, while the resolution is 0.003 inches of mercury for measurement and 0.005 inches of mercury for reporting. Due to high accuracy of the sensor, the contribution of measurement noise is small, and we have chosen the parameter a to be 0.2.

Figure 4 compares the correlation between measurements T units apart with predictions from a model. Figure 5 compares the probability density of the difference between model noise 2 units apart with predictions from a model. The figures suggest that the model fit is satisfactory.

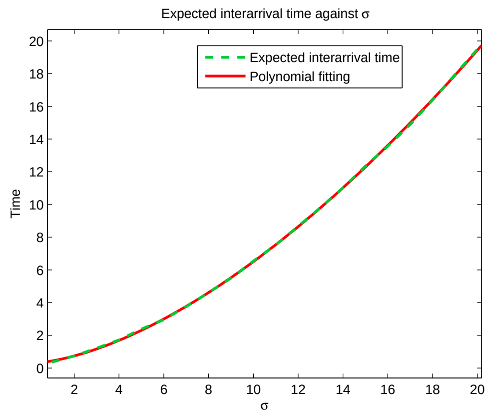
B. Message generation rates

We expect the rate at which messages are generated to decrease the larger the value of the threshold. We conducted the following experiment to estimate message rates. In the experiment, a message was generated when the model given above (average over the 5 years) deviated from the measurement by a value greater than the threshold. The model was run over all points - every hour of every day - over the 5 years. The average and maximum times between messages was computed for each value of the threshold. The graphs are shown in figures 6(a) and 6(b) .

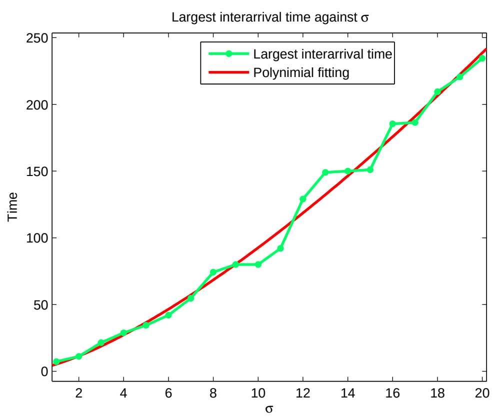
As expected, average and maximum message inter-arrival times increase with the threshold. A fitting with the polynomial $ax^{\alpha}+b$ was carried out, and the coefficients of the polynomials are shown.

C. The distribution of the estimator given asynchronous messages

Next we evaluate how accurately the distribution of the estimator can be computed. At the instant a signal is received giving the value of the parameter, the estimator distribution is determined by the measurement noise. Later, the distribution is influenced by both measurement noise and model noise; the variance for both types of noise is monotonic non-decreasing with time while there are no signals. The estimator distribution converges to the equilibrium distribution as the time after the



(a) Average time between two consecutive messages. The parameters of the polynomial fitting are $a = 0.1485, b = 0.284, \alpha = 1.622$



(b) Largest time between two consecutive messages. The parameters of the polynomial fitting are $a = 3.782, b = 1.587, \alpha = 1.381$

Fig. 6. Inter-arrival time of messages

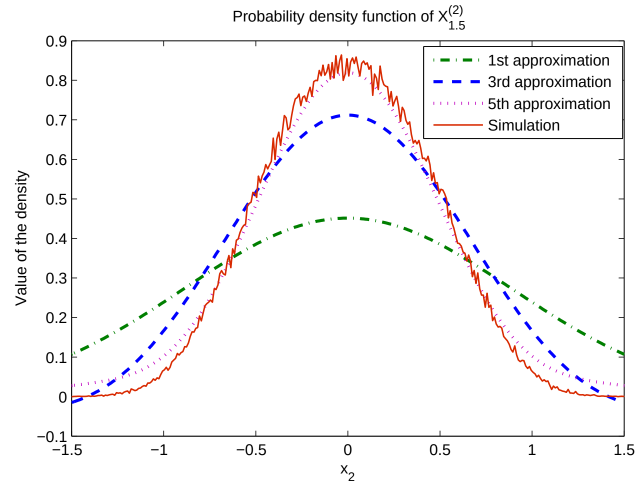


Fig. 7. Probability density function of $X^{(2)}_{1.5}$

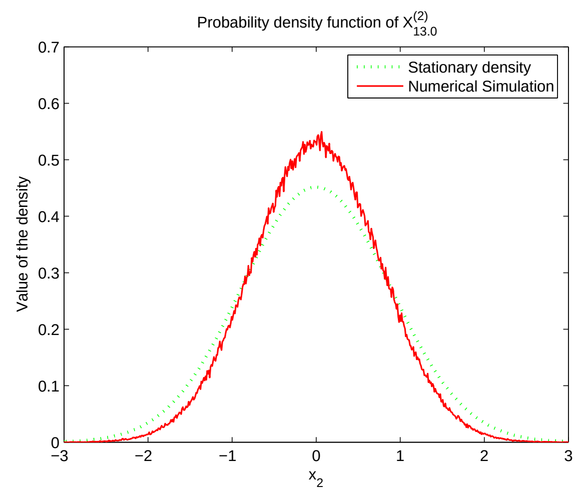


Fig. 8. Probability density function of $X^{(2)}_{13}$

signal is received becomes large; this distribution is obtained by using only the smallest eigenvalue of the Fokker-Planck operator. The estimator distribution a short time after a signal is received can only be computed with using several of the smaller eigenvalues.

We compare the reconstruction of the probability density function obtained using the formula (6) with the parameters $a = 0.2, \epsilon = 0.4972, \kappa = 0.027$ with the probability density function computed by means of numerical simulations. For our experiments we consider the time differences to be respectively 1.5 and 15.0 and fix $t_0 = 0$. Clearly, when $t = 1.5$, a larger number of eigenvalues contribute to the reconstruction of the probability function, while for $t = 13.0$ all eigenvalues except $\lambda_{1,0}$ decay and the probability density function in (6) converges to the stationary distribution given by the normalized eigenfunction $\phi_{1,0}$ corresponding with the smallest eigenvalue $\lambda_{1,0}$.

The results are shown in Figures 7 and 8.

D. Decrease in estimation confidence with message absence

An advantage of predicate signaling is that the absence of messages conveys the information that measurements fit the model. Though the absence of signals conveys information, the presence of signals conveys even more information. The estimate for an unobservable parameter, at any point in time, is a random variable whose variance increases with the time since the last signal was received. A question that arises in predicate signaling is how rapidly do confidence intervals for estimates grow? If (say the 95%) confidence interval gets large very rapidly then a node fusing information from multiple locations is likely to make erroneous estimates if it hasn't received signals for some time. The rate at which the confidence interval increases is one of the factors that determines the time-driven aspect of predicate signaling. Even if measurements fit the model, signals may need to be generated merely so that confidence intervals of estimates are reasonably small.

Of course, while no signal is generated the measured value

	λ	p
$(1,0)$	-8.7909	1.2732
$(1,2)$	-10.1409	-0.3183
$(1,4)$	-11.4909	0.0398
$(1,6)$	-12.8409	-0.0033
$(1,8)$	-14.1909	0.00020723

TABLE I
VALUES OF THE FIVE LARGEST EIGENVALUES AND CORRESPONDING $p^{(0)}$ OBTAINED USING FORMULA (6)

falls within the specified threshold. So, one approach to maintaining small confidence intervals is to make the threshold small. But a consequence of small thresholds, as we saw in the previous experiments, is that messages are generated more frequently. Thus, designers have to tradeoff message frequency with model accuracy. The next set of performance measures that we studied deal with this issue.

We have taken the variance computed using the probability density function in equation (6) as a measure of the estimation error. The mean squared error is minimized by the expectation of the signal given the measurements and the minimum expected loss is the variance.

Figure 9 shows the variance for different values of σ as a function of time assuming the process to be all the time within the initial domain Ω_0.

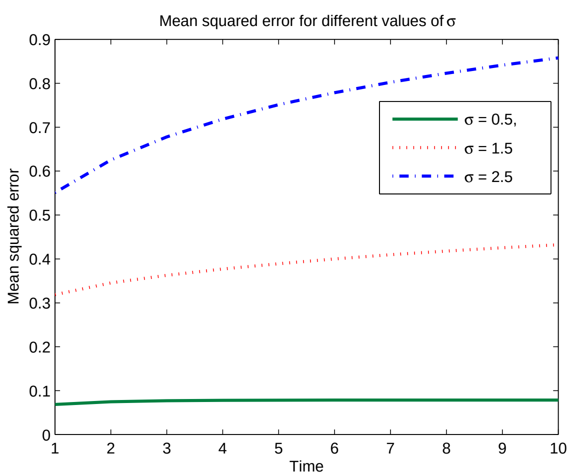


Fig. 9. Expected loss functions for different values of σ under the assumption that the process is always within the first domain

Finally consider a case where a message is lost in a system that uses a protocol such as UDP in which (unlike TCP) messages are not resent until an acknowledgment is received. Assume that the next message sent after a message is lost is indeed delivered to its destination. This second message repairs (at least some of) the damage of the lost message because each message that is delivered tells the receiver the model that the sender is currently using. What are the average and worst-case durations between the sending of two successive messages? The average time is proportional to $\frac{\sigma^2}{1+\epsilon^2}$ as shown in Appendix D. Experiments on the measured data shows that the worst-case time is indeed many hundreds of time steps 6(b).

V. CONCLUSIONS

The graph of average signal inter-arrival time shows the benefits of using model-based or anomaly-based signaling. We are sure that more accurate models will give significantly reduced message generation rates. Even for the simple model, the maximum time between signals is very large. Thus, the loss of a single message can result in the wrong model being used in estimation for a long time. This suggests that pure anomaly-based signaling is inadequate for even simple models. Some combination of the common signaling schemes should be used.

The increase in standard deviation (and thus confidence intervals) of estimators with time since the last signal suggests that there are benefits to incorporating query-based signaling: when the confidence interval gets unacceptably large, a query is sent asking for more recent measurements.

The relationships between the performance measures suggest systematic ways for designers to make tradeoffs between message rates and accuracy. Next, we discuss some of the weaknesses of predicate signaling and suggest ways of ameliorating the consequences of these weaknesses.

Sense and respond systems are useful when the rare event — the earthquake, the intruder, the virus — occurs. The model that represents the rare situation may not be as accurate as models that represent typical behavior. Thus, when the rare event occurs, the frequency of messages required to obtain satisfactory accuracy in estimation may be extremely high. This condition holds whether the signaling scheme is time-driven, anomaly-based or query-based: if we cannot model what is going on we have to measure frequently. A criticism of predicate signaling is that worst-case message rates may be as bad as for time-driven signaling, and thus bandwidth requirements may be just as big.

Predicate signaling does have advantages even in this case. Firstly energy can be conserved during the long periods in which models accurately represent measurements. The conserved energy can be used when truly required: when actual behavior doesn't fit into expected norms. Secondly, even though worst-case bandwidths are high, the bandwidth can be used almost all the time for other applications. The bandwidth needs to be reserved for signaling only for the very rare periods in which models don't represent measurement.

Anomaly-based signaling assumes either that models can be computed rapidly or that results are precomputed and placed in tables so that the execution of a model can be reduced to a table lookup. If the model maps a large number of observable parameters to an unobservable parameter then the cost of table lookup is significant. Initial experiments suggest the adequacy of simple models that make predictions based on the most recent measurements or on recent measurements combined with time.

Appendix A: Fokker-Planck equations

Consider a diffusion process on $\mathbb{R}^d$ with generator $\hat{\mathcal{L}}$; the latter has a general form

$$\hat{\mathcal{L}}_{\boldsymbol{x}} = \sum_{i=1}^{d} b_i(\boldsymbol{x})\partial_{x_i} + \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}(\boldsymbol{x})\partial^2_{x_i x_j}. \quad (10)$$

The probability density $p_t(\boldsymbol{x}|\boldsymbol{y})$ conditioned on that the particle which start at $\boldsymbol{y}$ at time 0 be located at $\boldsymbol{x}$ at time t satisfies the Fokker-Planck equation

$$\partial_t p_t(\boldsymbol{x}|\boldsymbol{y}) = \hat{\mathcal{L}}^*_{\boldsymbol{x}} p_t(\boldsymbol{x}|\boldsymbol{y}). \quad (11)$$

Here $\hat{\mathcal{L}}^*$ is the adjoint of $\hat{\mathcal{L}}$:

$$\hat{\mathcal{L}}^*_{\boldsymbol{x}} = -\sum_{i=1}^{d} \partial_{x_i} b_i(\boldsymbol{x}) + \frac{1}{2}\sum_{i,j=1}^{d} \partial^2_{x_i x_j} a_{ij}(\boldsymbol{x}). \quad (12)$$

Let Ω be an open subset of $\mathbb{R}^2$ with smooth boundary $\partial\Omega$. Let us add the initial condition to equation 11 that the process has not touched the boundary $\partial\Omega$ up to time t. Furthermore, let $\varrho_t(\boldsymbol{x}|\boldsymbol{y})$ be as p_t, except conditional on the no-touching event. Formally

$$\varrho(\boldsymbol{x}|\boldsymbol{y}) = p_t(\boldsymbol{x}|\boldsymbol{y})\left[\int_\Omega p_t(\boldsymbol{x}|\boldsymbol{y})\right]^{-1} \quad (13)$$

Since p_t satisfies equation (11) with the boundary condition $p_t(\boldsymbol{x}|\boldsymbol{y}) = 0$ for $x \in \partial\Omega$, we can rewrite it as

$$p_t(\boldsymbol{x}|\boldsymbol{y}) = \sum_{k=0}^{\infty} p_k(\boldsymbol{y}) e^{\lambda_k t}\phi_k(\boldsymbol{x}) \quad (14)$$

where λ_k and ϕ_k are respectively the eigenvalues and eigenvectors of $\widehat{\mathcal{L}}^*$ in Ω with zero boundary conditions (we set $0 \geq \lambda_0 \geq \lambda_1 \ldots \geq \lambda_n$). As shown in Appendix B, $\phi_0(\boldsymbol{x})$ is sign-definite; assume it is positive and normalized to have integral over Ω be equal to unity. Observe that if ever the stationary $\varrho(\boldsymbol{x}|\boldsymbol{y})$ exists, then $\phi_0(\boldsymbol{x}) = \lim_{t\to\infty} \varrho_t(\boldsymbol{x}|\boldsymbol{y})$ (for any $\boldsymbol{y} \in \Omega$), thus the condition of sign-definiteness is equivalent to existence of the stationary conditional (on non-touching) distribution.

Suppose we know at time t that our measured model deviation touched the boundary for the first time; what is the distribution of its location on $\partial\Omega$? Since $\mathcal{L}^* = \nabla_x \cdot \widehat{\mathcal{F}}$, where $\widehat{\mathcal{F}}$ is the flux operator and using Gauss theorem we obtain that the answer follows from the formula

$$\begin{cases} \int_\Omega \widehat{\mathcal{L}}^* p_t(\boldsymbol{x}) d\boldsymbol{x} & = \int_{\partial\Omega} \widehat{\mathcal{F}} p_t(\boldsymbol{x}) \cdot d\boldsymbol{s}(\boldsymbol{x}) \\ \widehat{\mathcal{F}} p_t(\boldsymbol{x}) & := \frac{1}{2}\partial_{\boldsymbol{x}} \cdot [\hat{a}(\boldsymbol{x}) p_t(\boldsymbol{x})] - \boldsymbol{b} \cdot p_t(\boldsymbol{x}) \end{cases} \quad (15)$$

On the boundary $\partial\Omega$, $p_t(\boldsymbol{x}) = 0$, thus $\widehat{\mathcal{F}} p_t(\boldsymbol{x}) = \frac{1}{2}\partial_{\boldsymbol{x}} \cdot [\hat{a}(\boldsymbol{x}) p_t(\boldsymbol{x})]$ on $\partial\Omega$. Here $\hat{a}$ is a matrix whose ij–th component is a_{ij}. Thus given a probability density $\varrho(\boldsymbol{x})$, $\widehat{\mathcal{F}}\varrho \cdot d\boldsymbol{s}$ is the probability flow through the element ds. This implies that the single layer density given by $(\widehat{\mathcal{F}}\varrho \cdot \boldsymbol{n})\delta_{\partial\Omega}(\boldsymbol{x})$ (not normalized as written) corresponds to the escape distribution on $\partial\Omega$.

Appendix B 0-order approximate operator

We find all eigenvalues and eigenfunctions for the following operator

$$\hat{\mathcal{L}}^* = \hat{\mathcal{L}}^*_1 + \hat{\mathcal{L}}^*_2, \quad \hat{\mathcal{L}}^*_1 := \frac{1}{2}\partial^2_{x_1 x_1} + \kappa\partial_{x_1} x_1, \quad \hat{\mathcal{L}}^*_2 := \frac{1}{2}\partial^2_{x_2 x_2}$$

in the domain $\Omega := \mathbb{R} \times [-L, L]$. In order to find the eigenvalues ν_n and the eigenfunctions $\xi_n(x_1)$ of $\hat{\mathcal{L}}^*_1$, we observe that the greatest (smaller by the absolute value) eigenvalue ν_0 and the corresponding eigenfunction $\xi_n(x_1)$ are given by

$$\nu_0 = 0, \quad \xi_0(x_1) = \sqrt{\frac{\kappa}{\pi}} e^{-\kappa x_1^2}, \quad (16)$$

Notice that the eigenfunction ξ_0 is normalized to have total integral unity. Setting $\xi_n(x_1) = \xi_0(x_1) h_N(x_1)$ we obtain equations for $h_n(x_1)$:

$$\partial^2_{x_1 x_1} h_n(x_1) - 2\kappa x_1 \partial_{x_1} h_n(x_1) = 2\nu_n h_n(x_1). \quad (17)$$

This immediately implies that $h_n(x_1)$ are expressed using Hermite polynomials as

$$h_n(x_1) = H_n(\sqrt{\kappa} x_1), \quad (18)$$

thus we obtain

$$\nu_n = -\kappa n, \quad \xi_n(x_1) = \sqrt{\frac{\kappa}{\pi}} e^{-\kappa x_1^2} H_n(\sqrt{\kappa} x_1), \quad n = 0, 1, 2, \ldots \quad (19)$$

The operator $\hat{\mathcal{L}}^*_2$ has the eigenvalues $\mu_m = -\pi^2 m^2/8L^2$ and the eigenvectors

$$\begin{cases} \psi_m(x_2) & = \frac{\pi}{4L}\cos\frac{\pi m x_2}{2L} \ (m = 1, 3, \ldots) \\ \psi_m(x_2) & = \frac{\pi}{4L}\sin\frac{\pi m x_2}{2L} \ (m = 2, 4, \ldots) \end{cases} \quad (20)$$

Since $\hat{\mathcal{L}}^*[\xi_n(x_1)\psi_m(x_2)] = (\nu_n + \mu_m)\xi(x_1)\psi_m(x_2)$ we conclude that $\hat{\mathcal{L}}^*$ has the following eigenvalues and eigenvectors:

$$\begin{cases} \lambda_{m,n} = -\frac{\pi^2 m^2}{8L^2} - \kappa n \\ \phi_{m,n}(x) = \frac{\sqrt{\kappa\pi}}{4L} e^{-\kappa x_1^2} H_n(\sqrt{\kappa} x_1)\cos\frac{\pi m x_2}{2L} \end{cases} \quad (21)$$

where $n = 0, 1, 2, \ldots$, $m = 1, 3, 5, \ldots$. If $m = 2, 4, 6, \ldots$, "cos" function is changed to "sin". Observe that $\phi_{1,0}(x)$ is positive with total integral over Ω equal to unity. Since Hermite polynomials are a set of orthogonal polynomials in $\mathbb{R}$ the functions $\varphi_{m,n}$ are defined as

$$\varphi_{m,n} = \frac{4}{\pi 2^n n!} H_n(\sqrt{\kappa} x_1)\cos\frac{\pi m x_2}{2L}, \quad (22)$$

as before, "sin" has to be taken if m is even. The normalizing coefficient is chosen to have

$$\int_\Omega \phi_{m,n}(\boldsymbol{x})\varphi_{m',n'}(\boldsymbol{x})d\boldsymbol{x} = \delta_{m,m'}\delta_{n,n'}$$

Appendix C Eigenfunctions in the diagonal strip

It has been shown in Appendix A that our probability density conditioned on that the domain has not been touched is the solution of the eigenvalue problem formulated in equation (11). Thus we need to find the eigenvalues and eigenfunctions of the Fokker-Planck operator

$$\hat{\mathcal{L}}^* = \frac{1}{2}[\partial^2_{x_1x_1}] + \epsilon^2\partial^2_{x_2x_2} + \kappa\partial_{x_2}x_2$$

in a strip $\Omega := \{(x_1, x_2) : |x_1 + x_2| < \sigma\}$ (zero boundary conditions). Introducing

$$y_1 = \frac{x_1 + x_2}{\sqrt{1+\epsilon^2}}, \quad y_2 = \frac{\epsilon x_1 - x_2/\epsilon}{\sqrt{1+\epsilon^2}}, \tag{23}$$

we obtain

$$\hat{\mathcal{L}}^* = \frac{1}{2}[\partial^2_{y_1y_1}+\partial^2_{y_2y_2}]+\frac{\kappa}{1+\epsilon^2}[\partial_{y_2}y_2-\epsilon(y_1\partial_{y_2}+y_2\partial_{y_1})+\epsilon^2\partial_{y_2}y_2],$$

and the domain becomes $\Omega = \{(y_1, y_2) : |y_1| < \frac{\sigma}{\sqrt{1+\epsilon^2}}\}$. Regular perturbation theory may now be applied to this problem. For the lowest order approximation we may use the results of Appendix B to get

$$\begin{cases} \lambda_{m,n} & \approx & -\frac{\pi^2 m^2}{8L^2} - \kappa n \\ \phi_{m,n}(x) & \approx & \frac{\sqrt{\kappa\pi}}{4L}e^{-\kappa y_2^2}H_n(\sqrt{\kappa}y_2)\cos\frac{\pi m(x_1+x_2)}{2\sigma} \end{cases}$$

(”sin” has to be taken if m is even)

Appendix D

The objective of this section is to derive an analytic formula for the expected value of the time between two consecutive exit times (i.e. the time when the boundary of the current domain is touched). This is obtained by using the Dynkin formula [5] which states:

Let $f \in \mathbb{L}^2(\Omega)$. Suppose τ is a stopping time such that $E[\tau] < \infty$. Then

$$E[f(X^x_\tau)] = f(x) + E[\int_0^\tau \hat{\mathcal{L}}f(X^x_s)ds] \tag{24}$$

where X^x_t is a process starting at x at time 0

We want to chose a function f such that

(1) $f(X^x_\tau)$ is easy to compute and it does not depend on τ, e.g. a function f which is identically zero on the boundary of the domain
(2) $\hat{\mathcal{L}}f$ is easy to compute and the result is a simple function of τ, e.g. $\hat{\mathcal{L}}f = 1$, so that the the integral in the equation (24) is just τ.

Applying Dynkin's formula for such a choice of f we would obtain $0 = f(x) + E[\tau]$, thus $E[\tau] = -f(x)$. This gives the expectation of the hitting time τ. Furthermore, it implies that f is the solution of the differential equation $\hat{\mathcal{L}}f = 1$ with Dirichlet (zero) boundary conditions. It can be easily verified that $f(x_1, x_2) = \frac{(x_1+x_2)^2}{1+\epsilon^2} - \frac{\sigma^2}{1+\epsilon^2}$ solves the equation. Under the assumption that the process starts at $(0, 0)$ at time 0, then the expected value of τ equals $\frac{\sigma^2}{1+\epsilon^2}$

References

[1] D.Estrin L.Girod M.Hamilton A.Cerpa, J.Elson and J.Zhao. Habitat monitoring: application driver for wireless communications technology.

[2] K.Ramchandran A.G.Dimakis, V.Prabhakaran. Ubiquitous access to distributed data in large-scale sensor networks through decentralized erasure codes. In *Proceedings of the 4-th International Symposium on Information Processing in Sensor Networks*, pages 111–117, 2005.

[3] A.Mainwaring, J. Polastre, R. Szewczyk, D. Culler, and J. Anderson. Wireless sensor networks for habitat monitoring. In *Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Application, WSNA02, Atlanta, USA*, pages 88–97. ACM, 2002.

[4] E. Karpilovsky D. Zimmerman B. Aydemir, K. M. Chandy. Event-driven architectures for distributed crisis management. In *Proceedings of the 15th IASTED International Conference on Parallel and Distributed Computing and Systems*, 2003.

[5] B.Oksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer, sixth edition, 2003.

[6] K.M. Chandy and Jayadev Misra. How processes learn. *Distributed Computing*, 1(1), January 1986.

[7] C.W.Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*, volume 13 of *Springer Series in Synergetics*. Springer, third edition, 2004.

[8] F.Zhao and L.Guibas. *Wireless sensor networks: an information processing approach*. Morgan Kaufamann, first edition, 2004.

[9] J.Liu L.Guibas F.Zhao, J.Liu and J.Reich. Collaborative signal and information processing: An information directed approach. *Proceedings of the IEEE*, 91(8):1199–1209, 2003.

[10] Yoram Moses Joseph Y. Halpern, Ronald Fagin and Moshe Y. Vardi. *Reasoning About Knowledge*. MIT Press, 1995.

[11] K.M.Chandy. Sense and respond systems. *Proceedings of the 31st Annual International Conference of the Association of System Performance Professionals*, December 2005.

[12] K.M.Chandy and L.Lamport. Distributed snapshots: Determining global states of distributed systems. *ACM Transactions on Computing Systems*, 3(1):63–75, February 1985.

[13] Yoram Moses. Knowledge and communication: a tutorial. In *Proceedings of the 4th conference on Theoretical aspects of reasoning about knowledge*, pages 1–14, 1992.

[14] A.H. Sayed T.Kailath and B.Hassibi. *Linear Estimation*. Prentice-Hall, 2000.