ONLINE BAYESIAN ESTIMATION OF HIDDEN MARKOV MODELS WITH UNKNOWN TRANSITION MATRIX AND APPLICATIONS TO IEEE 802.11 NETWORKS

Tom Vercauteren

INRIA Sophia-Antipolis
Epidaure Research Group
06902 Sophia Antipolis FRANCE

Alberto Lopez Toledo and Xiaodong Wang

Columbia University
Department of Electrical Engineering
New York, NY 10027

ABSTRACT

We develop online Bayesian signal processing algorithms to estimate the state and parameters of an hidden Markov model (HMM) with unknown transition matrix. The first online estimator is based on the sequential Monte Carlo (SMC) technique and uses a set of sufficient statistics to carry the information about the transition matrix. A deterministic variant of the SMC estimator is then developed, which is simpler to implement and offers superior performance. Finally a novel approximate maximum a posteriori (MAP) algorithm is proposed. These algorithms offer a solution to the problem of estimating the number of competing terminals in an IEEE 802.11 network where better performance can be expected if the backoff parameters are adapted to the number of active users. Realistic simulations using the ns-2 network simulator are provided to demonstrate the excellent performance of the proposed estimators.

1. INTRODUCTION

We consider the following HMM with unknown parameters:

$$x_t \sim \mathcal{MC}(\boldsymbol{\pi}, \boldsymbol{A}), \quad y_t \sim \mathcal{B}(x_t), \tag{1}$$

where $\mathcal{MC}(\boldsymbol{\pi}, \boldsymbol{A})$ denotes a Markov chain with initial probability distribution $\boldsymbol{\pi}$ and transition probability matrix $\boldsymbol{A}$; x_t is the state realization of the Markov chain at time instant t; $\mathcal{B}(x)$ denotes the discrete probability distribution of the observations conditioned on the state realization; and y_t is the observation. For notational ease, x_t is assumed to take its values in the finite set $\mathcal{X} = [1, \cdots, N]$.

Denote the observation sequence up to time t as $\boldsymbol{y}_t \triangleq [y_1, y_2, \cdots, y_t]$ and the network state sequence up to time t as $\boldsymbol{x}_t \triangleq [x_1, x_2, \cdots, x_t]$. Let the model parameters be $\boldsymbol{\theta} = \{\boldsymbol{\pi}, \boldsymbol{A}\}$. Given the observations $\boldsymbol{y}_t$ at time t, we are interested in estimating the current state x_t. The computational complexity of the naive solutions to these problems grows exponentially with the time index t. The forward-backward procedure [1] provide a recursive algorithm to get a linear complexity growth but is not able to cope with unknown parameters. The expectation-maximization (EM) batch algorithm can be used it only converges to a local maximum of the likelihood function which can be quite different from the global one [1].

We resort to the Monte Carlo signal processing techniques to solve the above inference problems. While usual SMC methods are not well suited to parameter estimation, we show that the complete information about the transition matrix can be carried over through some sufficient statistics so that the algorithm developed in [2] can be adapted to our HMM problem. A deterministic variant of the SMC estimator is also developed, which is simpler to implement and offers superior performance. The idea of using a set of sufficient statistics to represent the information about the transition matrix is included in the deterministic sample filter setting proposed in [3]. The use of sufficient statistics is pushed one step further than in [2] because this information about the parameters is now integrated so that no Monte Carlo approximation needs to be performed. The exponential increase in complexity is avoided by discarding the tails of the posterior distributions. For some applications, this algorithms might still be somewhat too computationally demanding. Inspired from the deterministic sequential sampling, we develop an approximate MAP algorithm that trades accuracy for computational requirement. Both algorithms can be applied to any HMM with unknown transition probabilities (and unknown prior distribution) and these are main contributions of our work. The online algorithms lead to an approximation of the probability distribution function (or to an hard estimate for the approximate MAP algorithm) of the number of competing terminals at a specific time step given the entire set of observations.

2. SUFFICIENT STATISTICS

A well-known strategy for Bayesian inference is to choose the prior distributions with a suitable form so that the posteriors belong to the same functional family as the priors. The priors and posteriors are then said to be conjugate. The

choice of the functional family depends on the likelihood. Here it can be seen that the discrete states x_t are drawn from multinomial distributions. For this kind of likelihood functions it is well known that the Dirichlet distribution provide conjugate priors. If $\boldsymbol{u}$ has a multivariate Dirichlet distribution $\boldsymbol{u} \sim \mathcal{D}(\gamma_1, \cdots, \gamma_N)$ with $\gamma_i > 0$, then,

$$p(\boldsymbol{u}) = \frac{\Gamma\left(\sum_{i=1}^N \gamma_i\right)}{\prod_{i=1}^N \Gamma(\gamma_i)} \prod_{i=1}^N u_i^{\gamma_i - 1}, \qquad (2)$$

with $\boldsymbol{u} = [u_1, \cdots, u_N]$, $u_i \geq 0$, $\sum_{i=1}^N u_i = 1$ and $\Gamma(\cdot)$ is the Gamma function. We will assume multivariate Dirichlet priors for both the initial probability vector $\boldsymbol{\pi}$ and the ith row of the transition matrix $\boldsymbol{a}_i$,

$$\boldsymbol{\pi} \sim \mathcal{D}(\rho_{1,0}, \rho_{2,0}, \cdots, \rho_{N,0}) \qquad (3)$$
$$\boldsymbol{a}_i \sim \mathcal{D}(\alpha_{i,1,0}, \alpha_{i,2,0}, \cdots, \alpha_{i,N,0}) \quad i = 1, \cdots, N. \quad (4)$$

With this choice the conditional posterior distributions is also Dirichlet. Let's denote $\rho_{i,t-1}$ and $\alpha_{i,j,t-1}$ the parameters of these distributions at time $t-1$

$$p(\boldsymbol{\pi}|\boldsymbol{x}_{t-1}, \boldsymbol{y}_{t-1}) = \mathcal{D}\left(\boldsymbol{\pi}; \rho_{1,t-1}, \cdots, \rho_{N,t-1}\right), \qquad (5)$$
$$p(\boldsymbol{a}_i|\boldsymbol{x}_{t-1}, \boldsymbol{y}_{t-1}) = \mathcal{D}\left(\boldsymbol{a}_i; \alpha_{i,1,t-1}, \cdots, \alpha_{i,N,t-1}\right). \quad (6)$$

The posterior distribution of the parameters therefore only depends on the set $\boldsymbol{T}_{t-1} = \{\alpha_{i,j,t-1}, \rho_{m,t-1}\}_{(i,j,m) \in [1,N]^3}$ of sufficient statistics. It is then easy to show that

$$p(\boldsymbol{\pi}|\boldsymbol{x}_t, \boldsymbol{y}_t) = p(\boldsymbol{\pi}|\boldsymbol{x}_{t-1}, \boldsymbol{y}_{t-1}), \qquad (7)$$
$$p(\boldsymbol{a}_i|\boldsymbol{x}_t, \boldsymbol{y}_t) = p(\boldsymbol{a}_i|\boldsymbol{x}_t) \propto p(x_t|x_{t-1}, \boldsymbol{a}_i)p(\boldsymbol{a}_i|\boldsymbol{x}_{t-1})$$
$$= \mathcal{D}(\boldsymbol{a}_i; \alpha_{i,1,t}, \cdots, \alpha_{i,N,t}), \qquad (8)$$

with $\alpha_{i,j,t} = \alpha_{i,j,t-1} + \mathbb{I}(x_{t-1} - i)\mathbb{I}(x_t - j)$; $\mathbb{I}(\cdot)$ being an indicator such that $\mathbb{I}(x) = 1$ if $x = 0$, and $\mathbb{I}(x) = 0$ if $x \neq 0$. $\boldsymbol{T}_t$ is then easily updated.

3. THE ONLINE SMC ESTIMATOR

In the usual SMC methodology with known parameters $\boldsymbol{\theta}$, the posterior distributions are approximated by a set of random samples and weights,

$$\hat{p}_{\boldsymbol{\theta}}(\boldsymbol{x}_t|\boldsymbol{y}_t) = \frac{1}{W_t} \sum_{k=1}^K w_t^{(k)} \mathbb{I}\left(\boldsymbol{x}_t - \boldsymbol{x}_t^{(k)}\right), \qquad (9)$$

where $W_t = \sum_{k=1}^K w_t^{(k)}$. One single couple $(\boldsymbol{x}_t^{(k)}, w_t^{(k)})$ is commonly referred as a *particle* and the set is said properly weighted if the approximation is unbiased [4]. The sequential algorithms use the old streams $\boldsymbol{x}_{t-1}^{(k)}$ and draw x_t to obtain a recursive approximation of $p_{\boldsymbol{\theta}}(\boldsymbol{x}_t|\boldsymbol{y}_t)$. In order to avoid the degeneracy of such an algorithm, it is necessary to insert a resampling step consisting in eliminating the unlikely samples, and multiplying the very likely ones.

If the parameters $\boldsymbol{\theta}$ are unknown, the usual approach is to include them into the state vector. Because of the static evolution of the parameters, the space of parameters is only explored during initialization which is obviously inefficient. Several works have proposed better algorithms for dealing with static parameters within an SMC framework [2, 4]. In this work, we make use of the approach developed in [2]. In the model under consideration, the posterior distribution of $\boldsymbol{\theta} = \{\boldsymbol{\pi}, \boldsymbol{A}\}$ given $\boldsymbol{x}_t$ and $\boldsymbol{y}_t$ has is defined by Dirichlet distributions depending on some sufficient statistics $\boldsymbol{T}_t = \boldsymbol{T}_t(\boldsymbol{x}_t, \boldsymbol{y}_t)$ that can easily be updated (7)-(8).

Suppose a set of properly weighted samples with respect to $p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{t-1})$ is available at time $(t-1)$,

$$\hat{p}(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{t-1}) = \frac{1}{W_{t-1}} \sum_{k=1}^K w_{t-1}^{(k)} \mathbb{I}\left(\boldsymbol{x}_{t-1} - \boldsymbol{x}_{t-1}^{(k)}\right). \quad (10)$$

The main idea is to get a Monte Carlo approximation of $p(\boldsymbol{x}_t, \boldsymbol{\theta}|\boldsymbol{y}_t)$ from (10) and from the set of sufficient statistics $\left\{\boldsymbol{T}_t^{(k)}\right\}_{k=1}^K = \left\{\boldsymbol{T}_t(\boldsymbol{x}_t^{(k)}, \boldsymbol{y}_t)\right\}_{k=1}^K$. The approximation of the marginal distribution $p(\boldsymbol{x}_t|\boldsymbol{y}_t)$ is then simply obtained by discarding the samples $\boldsymbol{\theta}^{(k)}$. Therefore, only the samples $\boldsymbol{x}_t^{(k)}$ and the corresponding sufficient statistics $\boldsymbol{T}_t^{(k)}$ are stored, but samples for $\boldsymbol{\theta}$ are drawn jointly to $x_t^{(k)}$ to simplify the computations. Specifically this approach is based on the following identity:

$$p(\boldsymbol{x}_t, \boldsymbol{\theta}|\boldsymbol{y}_t) \propto p(\boldsymbol{x}_t, \boldsymbol{\theta}, y_t|\boldsymbol{y}_{t-1}) \qquad (11)$$
$$\propto p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{t-1})p(\boldsymbol{\theta}|\boldsymbol{T}_{t-1})p(x_t|x_{t-1}, \boldsymbol{\theta})p(y_t|x_t, \boldsymbol{\theta}).$$

Based on the importance sampling paradigm, a Monte Carlo approximation of (11) can be obtained by keeping the past simulated streams $\left\{\boldsymbol{x}_{t-1}^{(k)}, w_{t-1}^{(k)}\right\}_{k=1}^K$ unmodified and drawing $(\boldsymbol{\theta}^{(k)}, x_t^{(k)})$ from a proposal distribution

$$q(\boldsymbol{\theta}, x_t|\boldsymbol{x}_{t-1}^{(k)}, \boldsymbol{y}_t) = q_1(\boldsymbol{\theta}|\boldsymbol{x}_{t-1}^{(k)}, \boldsymbol{y}_t) \cdot q_2(x_t|\boldsymbol{x}_{t-1}^{(k)}, \boldsymbol{y}_t, \boldsymbol{\theta}).$$

The weights are updated according to the usual rule:

$$w_t^{(k)} \propto w_{t-1}^{(k)} \frac{p\left(\boldsymbol{\theta}^{(k)}|\boldsymbol{T}_{t-1}^{(k)}\right) p\left(x_t^{(k)}|x_{t-1}^{(k)}, \boldsymbol{\theta}^{(k)}\right) p\left(y_t|x_t^{(k)}, \boldsymbol{\theta}^{(k)}\right)}{q_1\left(\boldsymbol{\theta}^{(k)}|\boldsymbol{x}_{t-1}^{(k)}, \boldsymbol{y}_t\right) q_2\left(x_t^{(k)}|\boldsymbol{x}_{t-1}^{(k)}, \boldsymbol{y}_t, \boldsymbol{\theta}^{(k)}\right)}.$$
$$(12)$$

The sufficient statistics are then updated and the samples for $\boldsymbol{\theta}$ discarded. Resampling can be performed as usual and the estimation of $\boldsymbol{\theta}$ is done through the use of Rao-blackwellization as follows [4],

$$\mathbb{E}\{\boldsymbol{\theta}|\boldsymbol{y}_t\} \approx \frac{1}{W_t} \sum_{k=1}^K w_t^{(k)} \mathbb{E}\left\{\boldsymbol{\theta}|\boldsymbol{T}_t^{(k)}\right\}. \qquad (13)$$

In this setting, the use of the optimal proposal distributions $q_1(\boldsymbol{A}|\boldsymbol{x}_{t-1}^{(k)}, \boldsymbol{y}_t) = \prod_{i=1}^N p(\boldsymbol{a}_i|\boldsymbol{x}_{t-1}^{(k)}, \boldsymbol{y}_t)$ for the transition matrix and $q_2(x_t|\boldsymbol{x}_{t-1}^{(k)}, \boldsymbol{y}_t, \boldsymbol{\theta}) = p(x_t|\boldsymbol{x}_{t-1}^{(k)}, \boldsymbol{y}_t, \boldsymbol{\theta})$ for

the state realization, lead to a weight update formula that does not depend on the actual values sampled at time t. It is therefore possible to compute the weights before sampling. It very attractive since we are now able to perform resampling before sampling and thus lowering the loss of diversity occurring during the usual resampling scheme. The same idea appears in the auxiliary sample filter [4] where a part of the weight can be computed before sampling and the rest is roughly estimated. The approximate weights are then used to perform the resampling as a prior step. The major difference here is that the complete weights can be pre-computed and that we advocate to use resampling only if necessary as is commonly done in the usual SMC procedure.

4. DETERMINISTIC SEQUENTIAL SAMPLING

The online SMC estimator randomly generates samples according to $p(x_t|\boldsymbol{x}_{t-1}^{(k)}, \boldsymbol{y}_t)$ for $x_t \in \mathcal{X}$. Consequently, some information is distorted if the number of Monte Carlo samples is not sufficiently large. Furthermore, the use of the optimal sampling distribution implicitly led us to consider all possible extensions of a sample. This was possible because x_t can only take values from the finite set $\mathcal{X}$. An alternative deterministic approach, developed in [3], and extended here to the case of unknown parameters and Markov state processes, consists of explicitly considering all K_{ext} possible extensions of the K samples and then perform a selection step so as to avoid the exponential increase of the number of samples and keep a constant number K of them. Another idea in this context is that there is no point in keeping different samples representing the same path. Therefore the selection step should not rely on the usual resampling scheme. The simplest idea kept in this paper is to select the K most likely samples at each time step. Strictly speaking, we then loose the properly weighted characteristic by cutting out the tails of the p.d.f. during the selection step. To avoid this, a more sophisticated scheme is developed in [5], some of the most likely samples are kept and resampling without replacement is performed on the remaining ones.

From Bayes theorem we have

$$p(\boldsymbol{x}_t|\boldsymbol{y}_t) \propto p(y_t|\boldsymbol{x}_t, \boldsymbol{y}_{t-1})p(x_t|\boldsymbol{x}_{t-1}, \boldsymbol{y}_{t-1})p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{t-1}) \quad (14)$$

with $p(x_t|\boldsymbol{x}_{t-1}, \boldsymbol{y}_{t-1}) = \int p_{\boldsymbol{\theta}}(x_t|x_{t-1})p(\boldsymbol{\theta}|\boldsymbol{T}_{t-1})\mathrm{d}\boldsymbol{\theta}$ and $p(y_t|\boldsymbol{x}_t, \boldsymbol{y}_{t-1}) = \mathcal{B}(y_t; x_t)$. Thanks to the Dirichlet prior, the integral can be computed analytically. Based on the facts that $p_{\boldsymbol{\theta}}(x_t|x_{t-1}) = a_{x_{t-1}, x_t}$ and that $p(\boldsymbol{a}_i|\boldsymbol{T}_{t-1}) = \mathcal{D}\left(\boldsymbol{a}_i; \alpha_{i,1,t-1}, \cdots, \alpha_{i,N,t-1}\right)$ we can show that

$$p(x_t = i|\boldsymbol{x}_{t-1}, \boldsymbol{y}_{t-1}) = \mathbb{E}_{p(\boldsymbol{\theta}|\boldsymbol{T}_{t-1})}\left\{a_{x_{t-1}, i}\right\}$$
$$= \frac{\alpha_{x_{t-1}, i, t-1}}{\sum_{j=1}^N \alpha_{x_{t-1}, j, t-1}}. \quad (15)$$

The recursion (14) can thus be computed analytically. Sup-

pose a set of weighted samples containing no duplicate and representing $p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{t-1})$ is available at time $(t-1)$. The state transition distribution can be written as

$$p(x_t|\boldsymbol{x}_{t-1}, \boldsymbol{y}_{t-1}) = \sum_{i=1}^N p(x_t = i|\boldsymbol{x}_{t-1}, \boldsymbol{y}_{t-1})\mathbb{I}(x_t - i).$$

and thus $p(\boldsymbol{x}_t|\boldsymbol{y}_t)$ can be approximated by

$$\hat{p}^{ext}(\boldsymbol{x}_t|\boldsymbol{y}_t) = \frac{1}{W_t^{ext}} \sum_{k=1}^K \sum_{i=1}^N w_t^{(k,i)} \mathbb{I}\left(\boldsymbol{x}_t - \left[\boldsymbol{x}_{t-1}^{(k)}, i\right]\right), \quad (16)$$

where the weight update is given by

$$w_t^{(k,i)} \propto w_{t-1}^{(k)} \mathcal{B}(y_t; i) \alpha_{x_{t-1}^{(k)}, i, t-1}^{(k)} / \sum_{j=1}^N \alpha_{x_{t-1}^{(k)}, j, t-1}^{(k)}. \quad (17)$$

The selection step is then performed to retain a fixed number of samples. As one can see such a procedure has many benefits since the parameters are analytically integrated out, no computationally expensive random sampling has to be done and no computation needs to be done twice.

5. APPROXIMATE MAP ALGORITHM

For HMM with known parameters, the Viterbi algorithm provides a recursive solution to get the best state sequence estimation in terms of the maximum a posteriori (MAP) [1]. When the parameters are unknown, the most common procedure is to use an EM algorithm which only converges to some local maximum of the a posteriori density but above all it is a batch procedure and thus can not be used in our setting. In this section another approach, based on the use of sufficient statistics developed above, is taken. An approximate MAP algorithm is presented whose computational load and memory need are equivalent to a usual Viterbi algorithm. We are interested in recursively maximizing $p(\boldsymbol{x}_t|\boldsymbol{y}_t)$ with respect to $\boldsymbol{x}_t$. In order to do that, the Viterbi algorithm uses $\delta_t(i) = \max_{\boldsymbol{x}_{t-1}|x_t=i} p(\boldsymbol{x}_t|\boldsymbol{y}_t)$. From (14) we have

$$\delta_t(i) = p(y_t|x_t = i) \max_{x_{t-1}|x_t=i} \max_{\boldsymbol{x}_{t-2}|x_{t-1}, x_t=i} [p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{t-1})p(x_t|\boldsymbol{x}_{t-1}, \boldsymbol{y}_{t-1})], \quad (18)$$

that can recursively be computed if the transition matrix is known by taking $p(x_t|\boldsymbol{x}_{t-1}, \boldsymbol{y}_{t-1}) = a_{x_{t-1}, x_t}$ out of the inner max so that $\delta_t(i) = p(y_t|x_t = i) \max_j[\delta_{t-1}(j) \cdot a_{j,i}]$. The estimate $\tilde{\boldsymbol{x}}_t$ of $\boldsymbol{x}_t$ at time t is then given by $\max_i[\delta_t(i)]$. When the transition matrix is unknown, even if the probability of any path can be analytically computed, such a recursion cannot directly be used because $p(x_t|\boldsymbol{x}_{t-1}, \boldsymbol{y}_{t-1})$ depends on $\boldsymbol{x}_{t-2}$. However if we make the approximation that

$p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{t-1})p(x_t|\boldsymbol{x}_{t-1},\boldsymbol{y}_{t-1})$ is maximized when only the second term $p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{t-1})$ is maximized,

$$
\max_{\boldsymbol{x}_{t-2}|x_t=i,x_{t-1}=j}[p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{t-1})p(x_t|\boldsymbol{x}_{t-1},\boldsymbol{y}_{t-1})] =
$$
$$
p(\boldsymbol{x}_{t-2}^*,x_{t-1}=j|\boldsymbol{y}_{t-1})p(x_t=i|\boldsymbol{x}_{t-2}^*,x_{t-1},\boldsymbol{y}_{t-1}) \quad (19)
$$

where $\boldsymbol{x}_{t-2}^* = \arg\max_{\boldsymbol{x}_{t-2}|x_t=i,x_{t-1}=j} p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{t-1})$, we can then derive an approximate MAP algorithm. Our simulations showed that (19) provides rather good results given the low complexity of the resulting algorithm. Our approximation $\hat{\delta}_t(i)$ of $\delta_t(i)$ can be computed recursively as

$$
\hat{\delta}_t(i) = p(y_t|x_t=i)\max_j\left[\hat{\delta}_{t-1}(j)\cdot\frac{\alpha_{j,i,t-1}^{(j)}}{\sum_{k=1}^N \alpha_{j,k,t-1}^{(j)}}\right],
$$
$$(20)$$

where the sufficient statistics $\boldsymbol{T}_{t-1}^{(j)}$ correspond to the retained path ending with $x_{t-1}=j$. Our estimate of $\boldsymbol{x}_t$ at time t is the state that maximizes $\hat{\delta}_t(i)$.

6. NUMBER OF COMPETING TERMINALS IN IEEE 802.11 NETWORKS

The performance of the IEEE 802.11 DCF is very sensitive to the number of users competing to access the wireless channel. The problem of estimating the number of competing terminals in an IEEE 802.11 wireless network with an extended Kalman filter (EKF) has been addressed recently in [6]. The estimation is based on the observed collision probabilities in the shared wireless channel. It is shown in [7] that if the nodes are in saturating regime and the system reaches a steady state then the collision probability p_c can be expressed as a function of the number of competing terminals x as $p_c = h_{W,m}(x)$ where W and m are the exponential backoff parameters. Each terminal can measure the packet collision probability by monitoring the channel activity. By counting the number of experienced collisions C_{coll}, as well as the number of observed busy slots C_{busy} within one *observation slot* we get a measurement y_t of the collision probability that follows a binomial distribution $y_t \sim \mathcal{P}(B, h_{W,m}(x))$. With this knowledge, IEEE 802.11 nodes can adapt their window parameters to optimize the total network throughput.

The EKF approach implicitly uses a linear Gaussian dynamic model departing from the discrete nature of the variables of interest, and the underlying model of IEEE 802.11 networks corresponds then to an HMM with unknown transition matrix. We applied the algorithms developed in this paper to a more realistic model, as our simulations do not strictly make use of the Markovian assumption. Indeed our ns-2 simulation uses a exponential On-Off activation process in continuous time which is not Markovian anymore in the irregular discrete timescale developed in [7]. The active
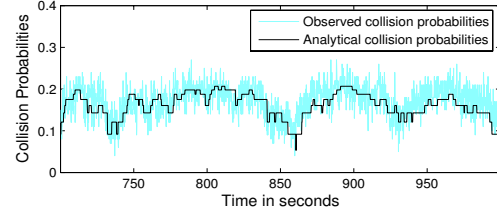

Fig. 1. Observed collision probabilities.

nodes are in saturating mode and the departures happen according to a realistic IEEE 802.11 protocol. The terminals used $B = 100$ for estimating the collision probability, resulting in noisy estimates of the probability of collision as shown in Fig. 1 (for $W = 32$, $m = 5$). Note that increasing B would result in better estimates but, on the other side, would increase the delay as the nodes need to wait more slots to update their estimates.

Estimator	model-based	ns-2 (W,m)		
	W=32, m=5	(16,6)	(32,5)	(64,4)
Gibbs Sampler	0.50	2.2621	1.7616	2.1621
SMC	0.63	2.8542	3.0417	2.5929
Deterministic	0.53	2.4434	2.903	2.2015
EKF+CUMSUM	3.2	3.4396	3.2909	2.8487
Approximate MAP	0.55	2.8822	2.8033	2.0066

Table 1. MSE obtained in our simulations.

Table 1 shows the results of a 300 seconds simulations with the standard 1Mbps IEEE 802.11 implementation in ns-2 2.27 for common values of W and m, as well as for data extracted from the Bianchi's model. Both model-based and ns-2 results show that our estimators outperform the EKF based estimator of [6] and approach that of a Batch estimator based on the Gibbs sampler.

7. REFERENCES

[1] L R Rabiner, "A tutorial on hidden Markov models and selected application in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–285, Feb. 1989.

[2] G Storvik, "Particle filters for state-space models with the presence of unknown static parameters," *IEEE Trans. Signal Processing*, vol. 50, no. 2, pp. 281–289, Feb. 2002.

[3] Elena Punskaya, *Sequential Monte Carlo Methods for Digital Communications*, Ph.D. thesis, University of Cambridge, 2003.

[4] A Doucet, N de Freitas, and N Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, 2001.

[5] Paul Fearnhead and Peter Clifford, "On-line inference for hidden Markov models via particle filters," *J. R. Statist. Soc. (B)*, vol. 65, no. 4, pp. 887–899, 2003.

[6] Giuseppe Bianchi and Ilenia Tinnirello, "Kalman fitler estimation of the number of competing termnals in an IEEE 802.11 network," in *Proc. Infocom 2003*, San Francisco, CA, Mar. 2003, vol. 2, pp. 844–852.

[7] Giuseppe Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Select. Areas Commun.*, vol. 18, no. 3, pp. 535–547, Mar. 2000.