Performance of the Sleep-Mode Mechanism of the New IEEE 802.16m Proposal for Correlated Downlink Traffic

Koen De Turck, Stijn De Vuyst, Dieter Fiems, Sabine Wittevrongel and Herwig Bruneel

Ghent University, Department of Telecommunications and Information Processing,
St. Pietersnieuwstraat 41, B-9000 Ghent, Belgium.

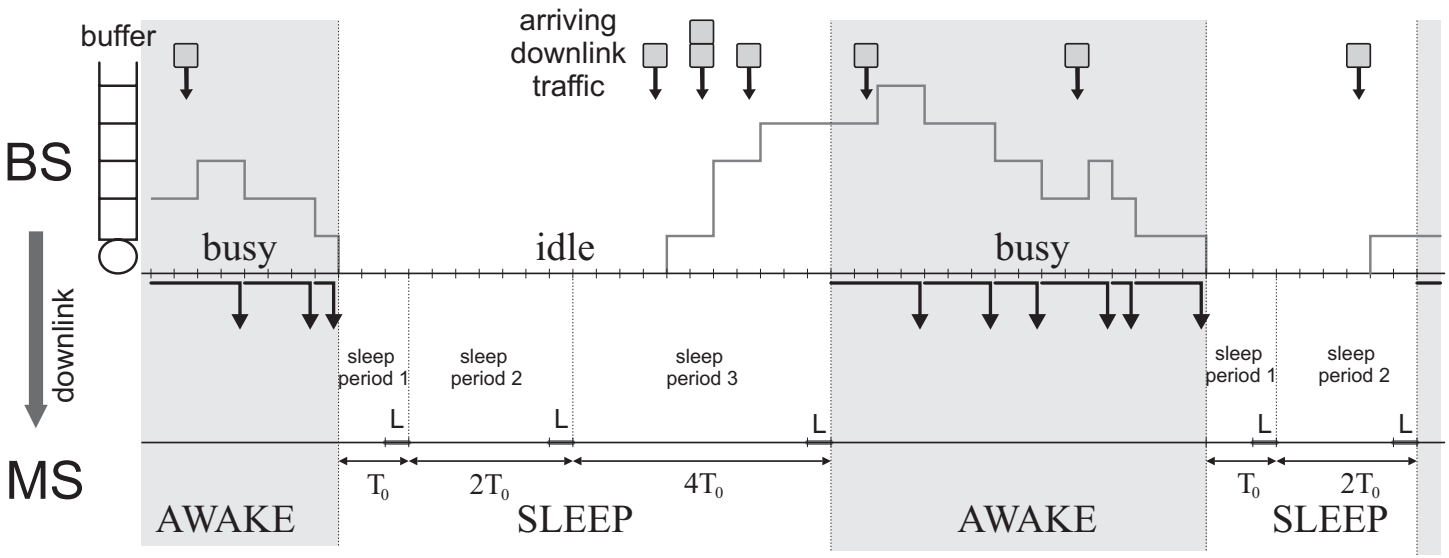
NETCOOP, Eindhoven — Nov 24, 2009

Outline

Why this Research?

- IEEE 802.16 (WiMAX) is an important emerging standard
- Energy-saving mechanisms are a hot topic
- Sleep mode is one of the main energy-saving measures included in the standard.
- Big players such as Samsung and Nokia have proposed new mechanisms that need performance evaluation.

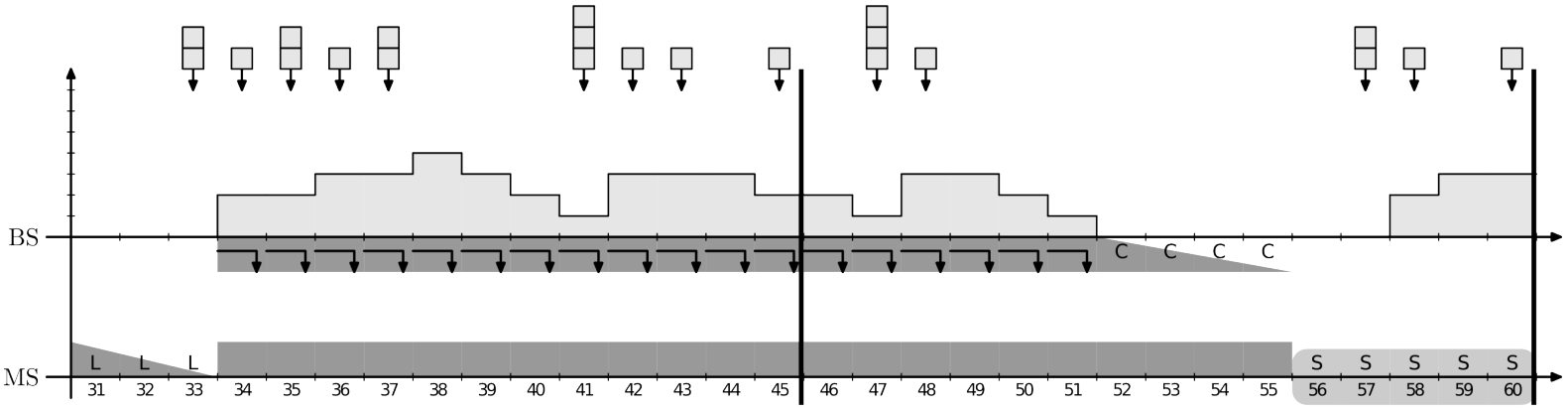
Sleep mode mechanisms in IEEE 802.16e



- There is a trade-off between packet delay and power consumption.
- Related to classic vacation models.

The new 'cyclic' sleep mode proposal

Sleep mode operation is organized in fixed cycles so that do not need to be renegotiated every time there is no traffic between BS and MS.



What are the performance characteristics of this new scheme?

Influence of other mobile stations is limited

It is important to note that it is the mobile station that goes into sleep mode, and that the BS has a dedicated buffer of each MS. The influence of other mobile stations is therefore only seen in the fact that the transmission rate may be increased when other MS are in sleep mode.
In this work, we concentrate on the single MS situation.

Modeling Assumptions

- We model the BS buffer as a discrete-time infinite buffer.
- A batch Markovian arrival process with a finite number of *phases* N governs the arrivals at the BS buffer. Let $\{varphi_i\}$ be the process of background states: $[\mathbf{A}(z)]_{ij} = \sum_k z^k a(k, j|i)$ $a(k, j|i) = \Pr[k\text{arrivals}, \varphi_1 = i|\varphi_0 = j]$.
- L, C and N denote the listening interval, the closedown interval and the cycle length.

Method

To determine the distribution of the buffer content by considering a Markov state space that consists of the buffer content u, the background state s and the position n within the cycle.
Embedded points:
Next step: detail all possible transitions from system state (u, n, s) of (u', n', s')

Transition probabilities (1)

The first case concerns the functioning of the system in active mode : when there are packets available to be sent ($u > 0$), then one of them is sent and the embedded Markov chain jumps to the next cycle position

From	To	Probability
(u, n, s)	$(u + k - 1, n + 1 \mod N, s')$	$[\mathbf{A}_k]_{ss'}$ if $u > 0$

Transition probabilities (2)

For the second case, we look at the closedown intervals, more specifically at closedown intervals that are interrupted because of a packet arrival. For this to happen, we must start from an empty buffer, then there must be a number of slots i in which no arrivals occur, followed by a slot during which there is at least one arrival. The next embedded instant is at the end of the slot thereafter.

From	To	Probability
$(0, n, s)$	$(k, n+i+2, s')$	$[z^{k+1}][\mathbf{A}_0^{i-1}(\mathbf{A}(z) - \mathbf{A}_0)\mathbf{A}(z)]_{ss'}$ if $0 < n < N - i - 2, 0 \leq i < C$

And so on. There are six cases in total.

Computation

We notice that the Markov chain is skip-free to the left in u. This means that we can use Neuts' matrix-analytic method.

$$\mathbf{P} = \begin{pmatrix} \boldsymbol{\mathcal{B}}_0 & \boldsymbol{\mathcal{B}}_1 & \boldsymbol{\mathcal{B}}_2 & \cdots & & \\ \boldsymbol{\mathcal{A}}_0 & \boldsymbol{\mathcal{A}}_1 & \boldsymbol{\mathcal{A}}_2 & \cdots & & \\ & \boldsymbol{\mathcal{A}}_0 & \boldsymbol{\mathcal{A}}_1 & \boldsymbol{\mathcal{A}}_2 & \cdots & \\ & & \boldsymbol{\mathcal{A}}_0 & \boldsymbol{\mathcal{A}}_1 & \boldsymbol{\mathcal{A}}_2 & \cdots \\ & & & \ddots & \ddots & \ddots \end{pmatrix}. \tag{1}$$

We find the block matrices $\{\boldsymbol{\mathcal{A}}_k\}$ and $\{\boldsymbol{\mathcal{B}}_k\}$ using the tables in the previous section.
However, the blocks are quite big: MN, meaning that matrix multiplications cost M^3N^3 operations.

Efficient computational scheme

The matrices $\{\mathcal{A}_k\}$ have a block circulant structure:

$$\mathcal{A}_k = \begin{pmatrix} 0 & \mathbf{A}_k & 0 & \cdots & 0 \\ 0 & 0 & \mathbf{A}_k & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & 0 & \mathbf{A}_k \\ \mathbf{A}_k & 0 & \cdots & 0 & 0 \end{pmatrix} \tag{2}$$

A matrix is block circulant iff there exists a block row vector $\mathbf{a}$ such that $[\mathcal{A}_k]_{ij} = [\mathbf{a}]_{i-j \mod N}$.

Properties of block-circulant matrices

- Block-circulant matrices are closed under addition and multiplication.
- Link with DFT: efficient product of two block circulant matrices with first block row vectors equal to $\mathbf{m}_1$ and $\mathbf{m}_2$: its first block row vector is: $\mathrm{IDFT}(\mathrm{DFT}(\mathbf{m}_1) * \mathrm{DFT}(\mathbf{m}_2))$.

Efficient computational scheme

Proposition. If the matrices $\{\mathcal{A}_k\}$ are block-circulant, then so is the fundamental matrix G.
Indeed, entry $[\mathbf{G}]_{ij}$ denotes the probability that, starting from level n in phase i, the process returns to level $n-1$, phase j in finite time. This probability depends merely on the difference (modulo N) between the cycle positions.

New algorithm for the matrix $\mathbf{G}$

Therefore, we can replace the customary iteration step by something much faster:

$$\hat{\mathbf{g}} = \sum_{k=0}^{\infty} \hat{\mathbf{a}}_k * \hat{\mathbf{g}}^{*k}$$

with $\hat{\mathbf{g}}$ the DFT transform of the first block row of $\mathbf{G}$, and $\hat{\mathbf{a}}_k$ is the DFT of $[0, \mathbf{A}_k, 0, \cdots, 0]$.
The first block row vector of $\mathbf{G}$ denotes the distribution of the length of sub-busy periods *modulo* N.

Computing performance measures

- From the distribution at the embedded points, we find the distribution at random points (details in the paper).
- Via Little's law, we can relate mean buffer content to mean delay.
- From the distribution at any type of instant, we can also derive the average power consumption.

Power consumption

We found these values for WiFi devices (no WiMAX data yet).

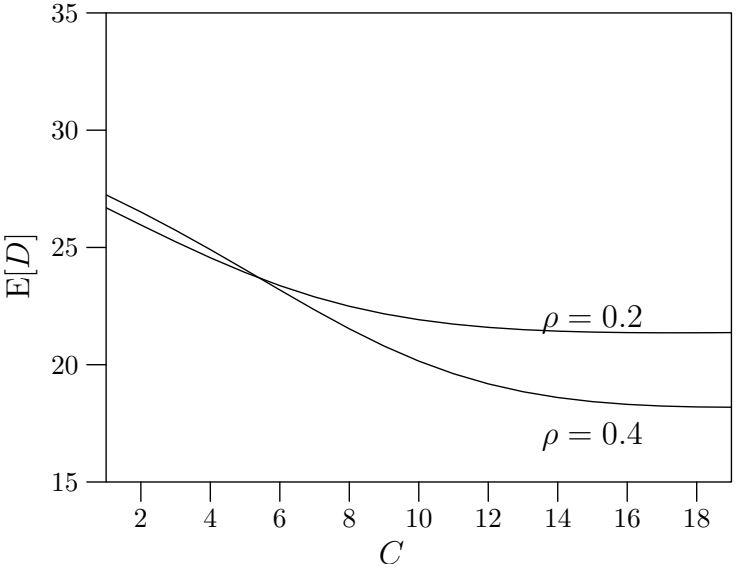
Table: Power consumption parameters

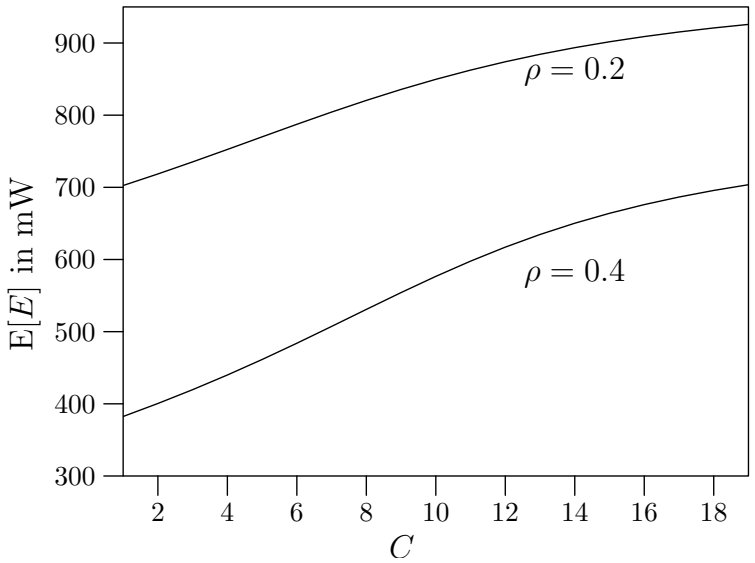
Parameter	Value	Description
P_s	0.045 W	Sleep mode power
P_ℓ	1.15 W	Listening mode power
P_r	1.40 W	Receive mode power
P_t	1.65 W	Transmit mode power

The analysis gives us the probabilities of each type of event, so we can easily find the mean power consumption.
Simpler metric for power consumption: Pr[sleep], which is maximally $1 - \rho$.

Some Numerical Results

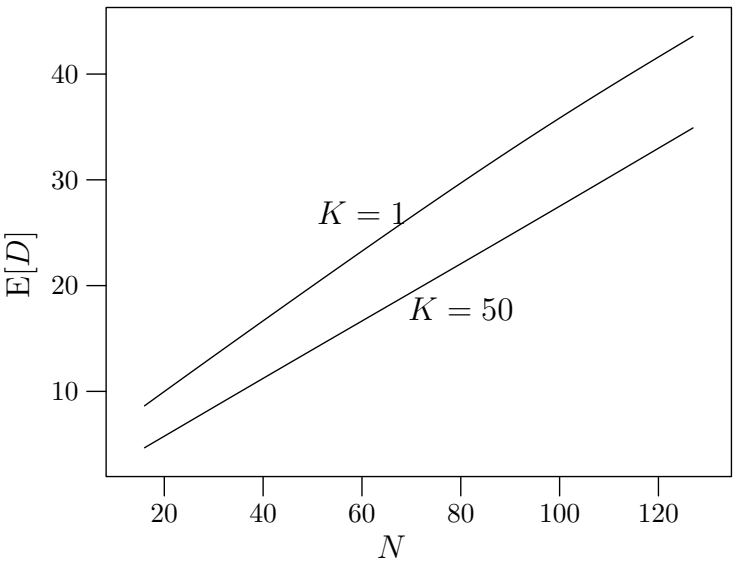
We assume an On-Off arrival source, with correlation factor K. and Pr[ON] $= \sigma$.
Influence of the length C of the closedown interval on the performance measures, for loads $\rho = 0.2$ and $\rho = 0.4$, and $N = 64$, $L = 1$ and $\sigma = 0.2$.

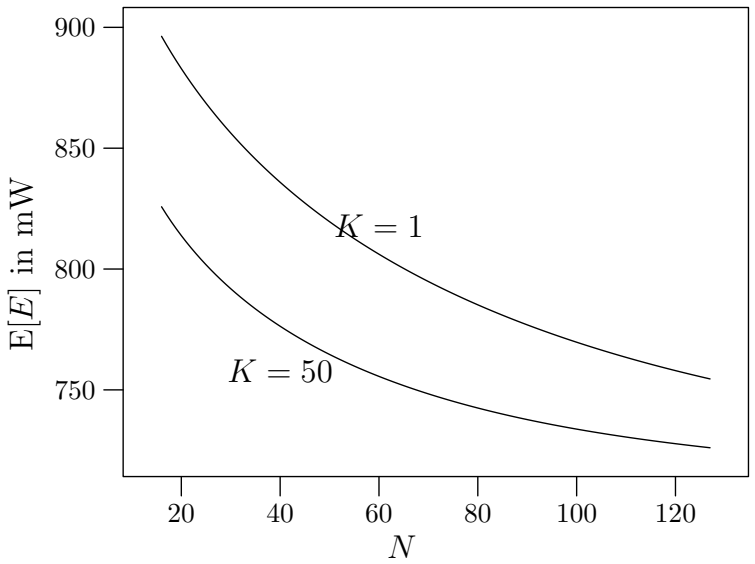




Some Numerical Results(2)

Mean delay and energy consumption versus the cycle length for different levels of correlation, where $L = 1, C = 8, \rho = 0.2, \sigma = 0.2$.





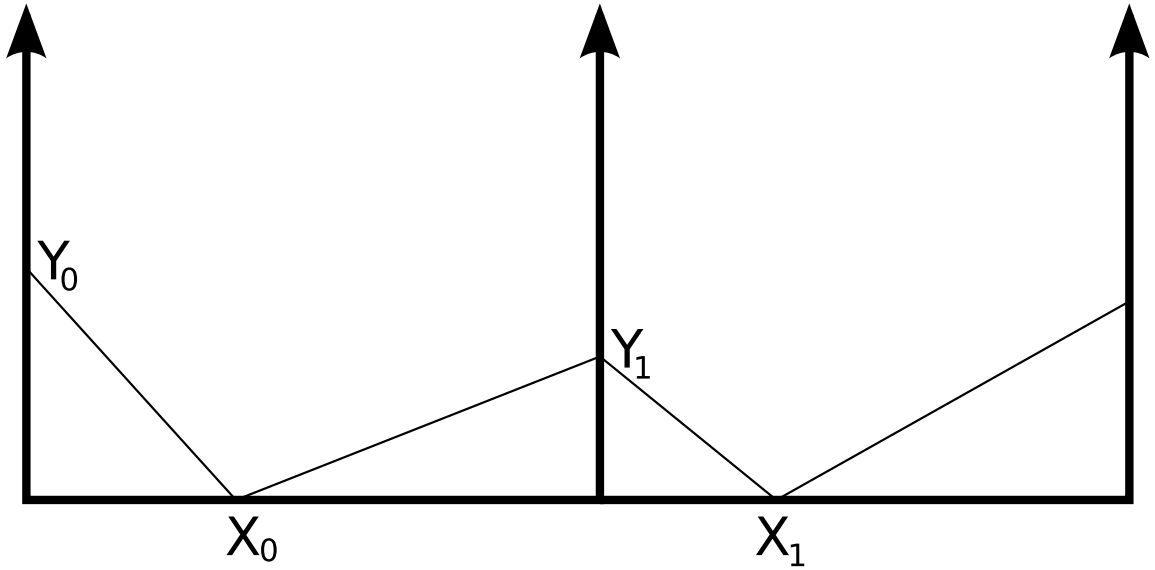
Approximation

Although the described numerical procedures are fairly efficient, there is room for a simple approximative scheme.
The most popular approximation tools are not readily usable:

- Using Poisson arrivals instead of D-BMAP does not really lead to simple expressions.
- When using heavy traffic or large deviations the effect of sleepmode is not seen, as the interesting things happen at the boundary.

Scaling by cycle length

Let the cycle length N approach infinity, and scale time and space accordingly:



Some calculations show that

$$\mathrm{E}\,X = \rho \quad \text{and} \quad \mathrm{E}\,Y = \lambda(1-\rho).$$

From which we have the crude approximations:

$$\mathrm{E}\,W \approx \frac{1}{2}(1-\rho)N \quad \text{and} \quad \Pr[\text{sleep}] \approx 1-\rho.$$

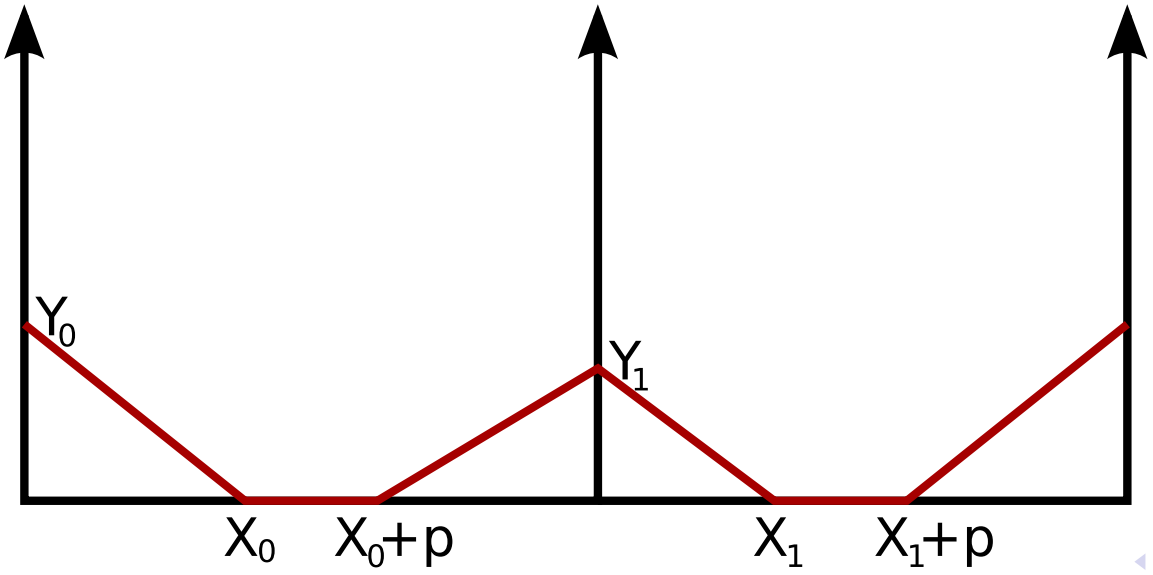
Scaling by cycle length(2) : refinement

Note that closedown effects get filtered away with this crude scaling.

The 'total closedown period' T_C consists of a sequence of idle and busy periods until an idle period is larger than C. We would like to have a scaling such that $N \to \infty$ and $\frac{\mathrm{E}\, T_C}{N} \to p$. Now, for Poisson arrivals, some simple calculations show:

$$\mathrm{E}\, T_C \approx \gamma e^{\lambda C}.$$

Hence, we choose $C = \frac{1}{\lambda}(\log(pN) - \log \gamma)$.



Scaling by cycle length(3)

We find easy formulas:

$$\Pr[\text{sleep}] \approx (1-\rho)(1-p)$$

$$\mathrm{E}\, W \approx \frac{1}{2}(1-\rho)(1-p)^2 N.$$

This suggests that -under this scaling- closedowns are not optimal! Indeed, for a given delay constraint, it is better to choose N short enough and $p = 0$, so that Pr[sleep] stays optimal.
Possible improvements of this scaling:

- Arrival processes that are not smooth at cycle scale.
- Tails instead of averages.
- Multiple mobile stations.

Conclusions

We analyzed the new cyclic WiMAX sleep mode mechanism. We found an exact numerical procedure for the buffer performance and for the power consumption. We also showed a crude but promising 'scaled' model.
Any questions?