PACKET INTERLEAVING OVER LOSSY CHANNELS

Pierluigi Salvo Rossi [(1,4)], *Francesco Palmieri* [(2)], *Giulio Iannello* [(3)], *Athina P. Petropulu* [(4)]

[(1)] D.I.S. - Università di Napoli "Federico II", [(2)] D.I.I. - Seconda Università di Napoli
[(3)] Università Campus Bio-Medico di Roma, [(4)] E.C.E.D. - Drexel University

ABSTRACT

Real packet channels are characterized by losses and delays as they both affect the performance of applications. A large number of techniques have been proposed in the literature for coding and recovery of lost information. In this paper we introduce a design framework to analyze, the impact of packet interleaving at the transmitter on the Quality of Service (QoS). Preliminary results are presented for Markov-chain based losses and Gamma-distributed delays. The optimization furnishes optimal packet interleaving proportions depending on channel statistics and maximum allowed delay, which is typically application dependent.

1. INTRODUCTION

Transmission by packets, being Internet the global communication structure, is becoming progressively the main communication framework. Packet losses, due to congestion or else, characterize most communication links and introduce significant limitations to performing reliable real-time communications. It is well known that distributing information [3][4] over different packets can increase transmission reliability and some coding strategies (e.g. Multiple Description Coding (MDC) [2], FEC [5]) have been devised on the basis of this concept. The key idea is exploiting partial received information for partial source recovery. Interleaving can be used at the price of further delays. When some delay is allowed (depending on the specific application) and re-transmission is not possible, packet interleaving can significantly strengthen communications. However, proper choice and evaluation of interleaving strategies, must be based on knowledge of channel model.

In this paper we introduce a framework to analyze the impact on the overall performance of distributing source information. Packet losses are assumed to be modeled by a 2-state Markov chain and packet delays Gamma distributed [1]. The specific application determines the maximum allowed delay (τ_{max}) to be included into the total accounts of channel losses. Definition of a distortion measure allows to study the results of optimizing packet interleaving proportions showing how channel statistics affect total performance.

2. THE FRAMEWORK

Fig. 1 shows the reference scheme. A block of N_b bits is available at time nT. The channel is able to send blocks of N_b bits every T

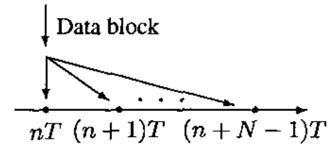


Fig. 1. Packet interleaving scheme.

seconds. Since the channel looses packets, it can be convenient to distribute the bits into N different packets.

When time is a precious resource, a large packet delay have the same effect of a loss on the application. With loss we denote that the packet cannot be delivered to the user (e.g. due to a buffer overflow, to errors obstructing decoding, etc.).

To formalize the problem, let us denote x_j the fractions of the N_b bits available at time nT and transmitted at time $(n+j)T$, being $0 \le x_j \le 1$ and $\sum_{j=0}^{N-1} x_j = 1$. $\mathbf{x} = [x_0, x_1, ..., x_{N-1}]^T$ is the *scrambling vector*. We assume a stationary behavior for the channel in order to refer our analysis to the N_b bits available at the generic instant nT. We say that the fraction of bits in a packet has been *erased* by the channel if the packet is lost or if it arrives too late. We assume that delays are Gamma(γ,ϑ) distributed, i.e. $f_\tau(t) = \frac{(t/\vartheta)^{\gamma-1} e^{-t/\vartheta}}{\vartheta\Gamma(\gamma)}$ and say $r_j(\tau_{max}) = Pr(\tau > \tau_{max} - jT)$. Note that $\{r_j(\tau_{max})\}$ is not-decreasing with respect to j, and is not-increasing with respect to τ_{max}.

When N packets are used to deliver the N_b bits we have to consider $M = 2^N$ different erasure configurations. Let us denote τ_{max} the maximum allowed delay for the N_b bits to be delivered. For the problem to be meaningful $\tau_{max} - (N-1)T > 0$, i.e. $N < \frac{\tau_{max}}{T} + 1$ [1]. Let $\mathbf{e}(i) = [e_0(i), e_1(i), ..., e_{N-1}(i)]^T$, with $i = 1, \ldots, M$, all the possible erasure configurations[2]. Let $\pi = [\pi, \ldots, \pi_M]^T$ be the probability vector of the erasure configurations where $\pi_i = Pr(\mathbf{e}(i))$ and $y(i)$ be the fraction of delivered bits if the erasure configuration $\mathbf{e}(i)$ occurs, i.e. $y(i) = \mathbf{x}^T\mathbf{e}(i) = \sum_{j=0}^{N-1} x_j e_j(i)$.

A quality function ($q(y)$), depending on delivered bits, characterizes our fidelity criterion. We assume that it is such that:

- $\frac{d}{dy}q(y) \ge 0 \quad , \quad 0 \le y \le 1,$
- $q(0) = 0, q(1) = 1.$

This work has supported by the Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) in the framework of the FIRB Project "Middleware for advanced services over large-scale, wired-wireless distributed systems (WEB-MINDS)" and the Office of Naval research under grant ONR-N00014-03-1-0123.

[1]The problem is interesting only if $\tau_{max} > T$, otherwise only one packet is allowed

[2]$e_j = 0$ (resp. $e_j = 1$) denotes that x_j has been erased (resp. not erased).

Let the QoS at the receiver be

$$QoS = E\{q(y)\} = \sum_{i=1}^{M} q(y(i))\pi_i = \sum_{i=1}^{M} \pi_i q\left(\sum_{j=0}^{N-1} e_j(i)x_j\right).$$

The problem is then the constrained following one,

$$\mathbf{x}_{opt} = arg \max_{\mathbf{x}\in\mathcal{X}} QoS \ ,$$

where $\mathcal{X} = \{\mathbf{x} \in \Re^N : \mathbf{o}^T\mathbf{x} = 1, 0 \le x_j \le 1, j = 0, \ldots, N-1\}$, and $\mathbf{o}$ denotes a vector of ones of length N.

The sign of the second derivative of the quality function distinguishes 2 cases,

- $\frac{d^2}{dt^2}q(t) \ge 0$, the problem has the trivial solution[3] $\mathbf{x}_{opt} = \mathbf{1}^{(0)}$,
- $\frac{d^2}{dt^2}q(t) \le 0$, then QoS is $\cap-$concave as linear combination with non-negative coefficients of $\cap-$concave functions, the only critical point is a maximum point, the solution has to be found into the set $\mathcal{X}$.

Let us consider the case when QoS is $\cap-$concave, it can be reasonable for applications like voice transmission. The optimum solution $\mathbf{x}_{opt}$ can be found by use of a gradient algorithm. Denoted $\dot{q}(y_0) = \frac{\partial}{\partial y}q(y)\Big|_{y=y_0}$, then

$$\frac{\partial}{\partial x_m}QoS = \sum_{i=1}^{M} \pi_i e_m(i)\dot{q}\left(\sum_{j=0}^{N-1} e_j(i)x_j\right).$$

For mathematical tractability we consider $q(y) = 1 - (y-1)^2$. Setting to zero all the partial derivative of QoS with respect to $\{x_0, \ldots, x_{N-1}\}$, it is easy to obtain the following equations system

$$\sum_{j=0}^{N-1} x_j \sum_{i=1}^{M} \pi_i e_m(i)e_j(i) = \sum_{i=1}^{M} \pi_i e_m(i) \ , \quad m = 0, \ldots, N-1 \ ,$$

that can be written as $\mathbf{S}\mathbf{x} = \mathbf{E}^T\pi$, where $\mathbf{S} = \mathbf{E}^T\mathbf{\Pi}\mathbf{E}$, $\mathbf{E} = \left[\mathbf{e}^T(1), \ldots, \mathbf{e}^T(M)\right]^T$, and $\mathbf{\Pi}$ is a diagonal matrix whose main diagonal is π. The solution of the constrained is obtained as solution of the following unconstrained problem (Langrange multipliers method) $\mathcal{R} = QoS + \lambda(\sum_{j=0}^{N-1} x_j - 1) - \sum_{j=0}^{N-1} \nu_j x_j$, (note that the constraints $x_j \le 1$ have been ignored as consequence of $\mathbf{o}^T\mathbf{x} = 1$, $x_j \ge 0$), that can be shown to be

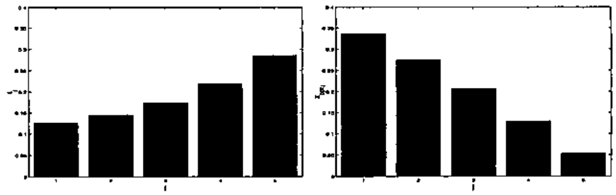
$$\begin{aligned} \mathbf{x} \ &= \ \mathbf{S}^{-1}\mathbf{E}^T\pi + \frac{1}{2}\mathbf{S}^{-1}\sum_{j\in J_0} \nu_j \mathbf{1}^{(j)} + \\ &- \frac{\mathbf{o}^T\mathbf{S}^{-1}\mathbf{E}^T\pi + \frac{1}{2}\mathbf{o}^T\mathbf{S}^{-1}\sum_{j\in J_0}\nu_j\mathbf{1}^{(j)} - 1}{\mathbf{o}^T\mathbf{S}^{-1}\mathbf{o}}\mathbf{S}^{-1}\mathbf{o} \ , \end{aligned}$$

where $J_0 = \{0 \le j \le N-1 : x_j = 0\}$ is the set of active nonlinear constraints.

3. INDEPENDENT LOSSES

In our first approach let us assume that losses and delays are independent identically distributed (iid), i.e $\pi_i = \prod_{j=0}^{N-1} Pr(e_j(i))$,with

[3]The solution is found at the boundary of $\mathcal{X}$. The property of non-decreasing trend of r_j forces the solution to be $\mathbf{x}_{opt} = \mathbf{1}^{(0)}$, where $\mathbf{1}^{(j)}$ denotes a vector of zeros whose j-th element is 1.



(a) Erasure vector ($\mathbf{p}$). (b) Scrambling vector ($\mathbf{x}_{opt}$).

Fig. 2. Almost-water-filling solution ($N = 5$, $g = 5$, $\vartheta = 30$, $T = 25$, $p = 0.1$, $\tau_{max} = 300$).

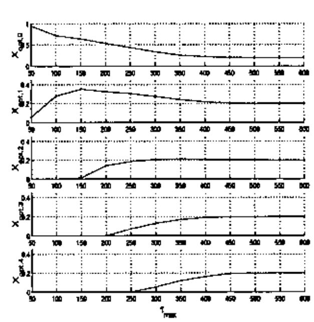


Fig. 3. Components of the optimum scrambling vector with respect to maximum allowed delay, ($N = 5$, $\gamma = 5$, $\vartheta = 30$, $T = 25$, $p = 0.1$).

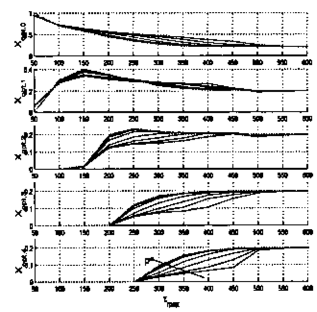


Fig. 4. Components of the optimum scrambling vector with respect to maximum allowed delay and average loss probability, ($N = 5$, $\gamma = 5$, $\vartheta = 30$, $T = 25$).

$$\begin{cases} Pr(e_j(i) = 0) & = p + (1-p)r_j(\tau_{max}) \ , \\ Pr(e_j(i) = 1) & = 1 - Pr(e_j(i) = 0) = (1-p)(1 - r_j(\tau_{max})) \end{cases} ,$$

and $p = Pr(\text{generic packet is lost})$.

Let us call $\mathbf{p}(\tau_{max}) = [p_0(\tau_{max}), p_1(\tau_{max}), \ldots, p_{N-1}(\tau_{max})]^T$ the *erasure vector* where $p_j(\tau_{max}) = Pr(e_j = 0)$. The sequence of probabilities p_j is also not decreasing with respect to j since $p_{j+1} = p + (1-p)r_{j+1} \ge p + (1-p)r_j = p_j$.

Computer simulations show the optimum solution have an *almost-water-filling* behavior, where the scrambling vector ($\mathbf{x}$) and the erasure vector ($\mathbf{p}$) play the role of the power resource and the noise respectively. Fig. 2 shows an example where $N = 5$, $\gamma = 5$, $\vartheta = 30$, $T = 25$.

Computer simulations also confirm (when $\frac{1}{p_{N-1}} \ge N - 1$) the validity of the following approximation[4],

$$x_j \approx \frac{\frac{1}{p_j} - (N-1)}{\sum_{k=0}^{N-1}\frac{1}{p_k} - N(N-1)} \ . \tag{1}$$

Eq. (1) clearly shows the almost-water-filling behavior of the optimum solution, as well as the asymptotic optimum solution ($\tau_{max} \to +\infty$) $x_\infty = [1/N, \ldots, 1/N]^T = \mathbf{x}_{flat}$.

It can be interesting to study the trend of the optimum solution with respect to the average loss of the channel and the maximum allowed delay of the application. Fig. 3 shows the optimum scrambling vector ($\mathbf{x}_{opt}$) for a given channel with respect to maximum allowed delay (τ_{max}). It can be noted how increasing of the maximum allowed delay has the effect of decreasing the first compo-

[4]For $N = 2$ this is the exact solution.

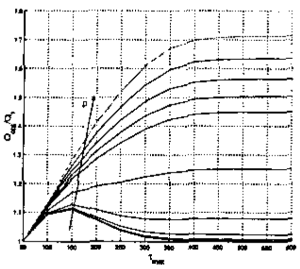


Fig. 5. Advantage of interleaving with respect to maximum allowed delay and average loss probability, ($N = 5$, $\gamma = 5$, $\vartheta = 30, T = 25$).

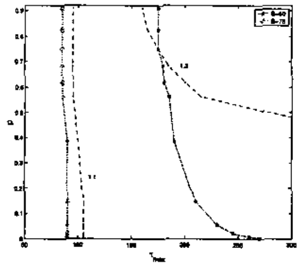
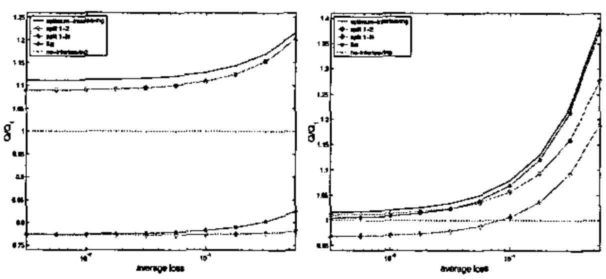


Fig. 6. Region for convenience of interleaving in the (τ_{max}, p)-plane with curves of advantage of interleaving (dashed line), ($N = 5$, $\gamma = 5$, $\vartheta = 30$, $T = 25$).



(a) $\tau_{max} = 150$. (b) $\tau_{max} = 300$.

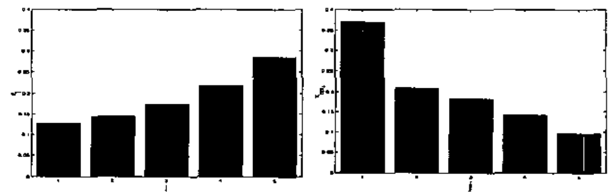
Fig. 7. Advantage of interleaving in terms of QoS ratio for some simple interleaving configurations ($N = 5, \gamma = 5, \vartheta = 30, T = 25$).

nent of the scrambling vector and increasing the remaining ones, i.e. induce more scrambling in the optimum solution. Furthermore it can be noted how the asymptotic solution is $\mathbf{x}_{flat}$. Fig. 4 shows the optimum scrambling vector ($\mathbf{x}_{opt}$) for a given delay distribution with respect to maximum allowed delay (τ_{max}) and average loss probability (p). It can be noted how also increasing of average loss probability has the effect of inducing more scrambling in the optimum solution. In all cases the asymptotic solution is $\mathbf{x}_\infty = \mathbf{x}_{flat}$.

The dependence of the advantage of interleaving is showed in Fig 5 by the ratio between $QoS(\mathbf{x}_{opt})$ and $QoS(\mathbf{1}^0)$ (the QoS for the interleaving and non-interleaving cases respectively). It can be noted how the advantage of interleaving is increasing with the loss probability and with the maximum allowed delay.

High loss probability and high maximum allowed delay induce more scrambling in the optimum solution and increase the advantage of the interleaving.

On the basis of this consideration it can be interesting to divide the (τ_{max}, p)-plane, characterizing both the channel and the application, into regions where it can be more or less convenient to use packet interleaving. Because of the basic solution for packet transmission is $\mathbf{x} = \mathbf{1}^{(0)}$, i.e. the information is all transmitted in the first packet, we consider how $x_{opt,0}$ decreases with respect to p and τ_{max}. Fig. 6 shows 2 curves delimiting 3 regions. The $o - green - line$ corresponds to the (τ_{max}, p)-pairs when $x_{opt,0} = 0.75$. The $* - red - line$ corresponds to the (τ_{max}, p)-pairs when



(a) Erasure vector ($\mathbf{p}$). (b) Scrambling vector ($\mathbf{x}_{opt}$).

Fig. 8. Effect of memory on the almost-water-filling solution ($N = 5$, $g = 5$, $\vartheta = 30$, $T = 25$, $p = 0.1$, $\tau_{max} = 300$, $\rho = 10$).

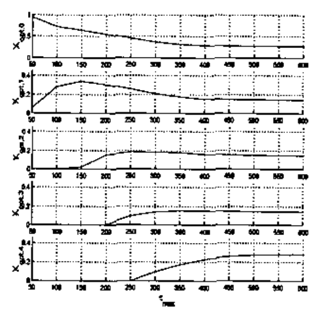


Fig. 9. Components of the optimum scrambling vector ($\mathbf{x}_{opt}$) with respect to maximum allowed delay(τ_{max}), ($N = 5$, $\gamma = 5$, $\vartheta = 30$, $T = 25$, $\psi_B = 0.1, \rho = 10$).

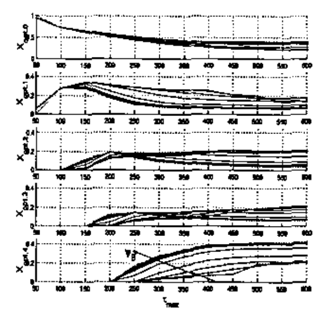


Fig. 10. Components of the optimum scrambling vector with respect to maximum allowed delay and average loss probability, ($N = 5, \gamma = 5, \vartheta = 30$, $T = 25, \rho = 10$).

$x_{opt,0} = 0.50$. So the region on the left of the $o - green - line$ can be considered a region where interleaving can be ignored, the region between the $o - green - line$ and the $* - red - line$ can be considered a region where interleaving can be appropriate, the region on the right of the $* - red - line$ can be considered as a region when interleaving has to be considered.

However, on the basis of Fig. 5, we said that the advantage of interleaving is function of the loss probability and the maximum allowed delay, so it is important to consider also how convenient is the optimum solution with respect to the absence of interleaving. Fig. 6 shows also two more curves corresponding to the (τ_{max}, p)-pairs when $\frac{QoS(\mathbf{x}_{opt})}{QoS(\mathbf{1}^0)} = constant$. On their basis we can say that some (τ_{max}, p)-pairs belonging to region where interleaving is convenient can be considered as belonging to the region where interleaving can be ignored, as the effective advantage that interleaving gives is lower than expected.

Fig. 7 shows the QoS for 5 configurations,

- *optimum-interleaving*: $\mathbf{x} = \mathbf{x}_{opt}$,
- *no-interleaving*: $\mathbf{x} = \mathbf{1}^{(0)}$,
- *split 1-2*: $\mathbf{x} = \left[\frac{1}{2}, \frac{1}{2}, 0, \ldots, 0\right]^T$,
- *split 1-N*: $\mathbf{x} = \left[\frac{1}{2}, 0, \ldots, 0, \frac{1}{2}\right]^T$,
- *flat-interleaving*: $\mathbf{x} = \mathbf{x}_{flat}$.

They give idea of how interleaving can improve performance.

4. MARKOV-CHAIN LOSSES

Let us assume now that delays are iid, but losses are correlated. We consider the case where loss phenomenon evolves according to a 2−state Markov chain where *Good* and *Bad* denote absence and presence of loss respectively and

$$\begin{cases} p = Pr(current\ state\ is\ Bad \mid previous\ state\ was\ Good) \\ q = Pr(current\ state\ is\ Good \mid previous\ state\ was\ Bad) \end{cases},$$

and being in *Good* state delays are Gamma distributed with parameters $\{\gamma, \vartheta\}$. Due to the memory of the channel we have,

$$\pi_i = Pr(e_0(i)) \prod_{j=1}^{N-1} Pr(e_j(i)|e_{j-1}(i)) \ .$$

Denoting $\psi_G = \frac{q}{p+q}$ and $\psi_B = \frac{p}{p+q}$ the steady-state probabilities of the Markov chain, it is easy to obtain

$$\begin{cases} Pr(e_j(i) = 0) = \psi_G r_j + \psi_B \\ Pr(e_j(i) = 1) = \psi_G (1 - r_j) \\ Pr(e_j(i) = 0|e_{j-1}(i) = 0) = \frac{\psi_G r_{j-1}(p+(1-p)r_j)+\psi_B(1-q+qr_j)}{\psi_G r_{j-1}+\psi_B} \\ Pr(e_j(i) = 1|e_{j-1}(i) = 0) = \frac{\psi_G r_{j-1}(1-p)(1-r_j)+\psi_B q(1-r_j)}{\psi_G r_{j-1}+\psi_B} \\ Pr(e_j(i) = 0|e_{j-1}(i) = 1) = (1-p)r_j + p \\ Pr(e_j(i) = 1|e_{j-1}(i) = 1) = (1-p)(1-r_j) \end{cases}$$

where the average loss probability is ψ_B.

It is common to consider as measure of the channel memory (ρ) the ratio between the conditional loss probabilities, $\rho = \frac{1-q}{p}$. Note that $\rho = 1$ means that the Markov chain reduces to a Bernoulli model (iid losses), while $\rho > 1$ means that a loss is more likely given that the previous state was *Bad* than *Good*, i.e. presence of memory (the more ρ, the more the memory).

Repeating the same step considered for the iid case we obtain the following. The effect of memory ($\rho > 1$) can be seen in

- the central component of the optimum solution $(\mathbf{x}_{opt1}, \ldots, \mathbf{x}_{optN-2})$ are partially reversed in the external component $(\mathbf{x}_{opt0}, \mathbf{x}_{optN-1})$ with respect to the case of no memory. This phenomenon is more accentuated the more ρ, for large ρ the optimum solution has not an almost-water-filling behavior.
- $\mathbf{x}_\infty \neq \mathbf{x}_{flat}$, the asymptotic optimum solution is not the flat-interleaving, it can be noted that $x_{\infty 0} \approx x_{\infty N-1} > x_{\infty 1} \approx \ldots \approx x_{\infty N-2}$, and for $\rho >> 1$ it happens $\mathbf{x}_\infty \approx \left[\frac{1}{N}, 0, \ldots, 0, \frac{1}{N}\right]$,
- increasing average loss probability and/or maximum allowed delay has a reduced effect in the advantage of interleaving with respect to the case of no memory,

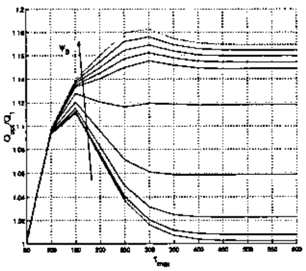


Fig. 11. Advantage of interleaving with respect to maximum allowed delay and average loss probability, ($N = 5$, $\gamma = 5$, $\vartheta = 30$, $T = 25$, $\rho = 10$).

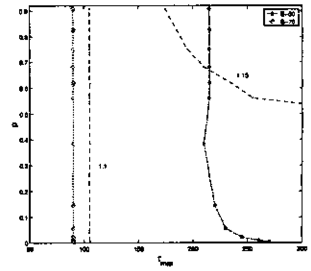
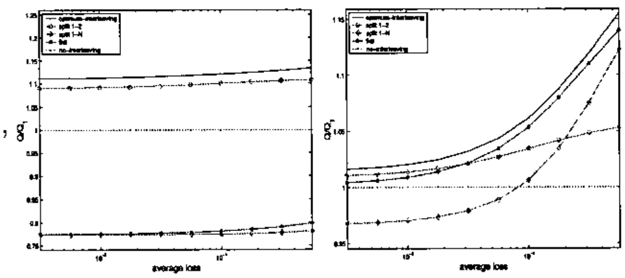


Fig. 12. Region for convenience of interleaving in the (τ_{max}, ψ_B)-plane with curves of advantage of interleaving (dashed line), ($N = 5, \gamma = 5$, $\vartheta = 30, T = 25, \rho = 10$).



(a) $\tau_{max} = 150$.

(b) $\tau_{max} = 300$.

Fig. 13. Advantage of interleaving in terms of QoS ratio for some simple interleaving configurations ($N = 5, \gamma = 5, \vartheta = 30, T = 25, \rho = 10$).

5. CONCLUSIONS

In this paper we have considered the effectiveness to distribute source information over different packet basing on an analysis framework appropriately introduced to take into account losses and delays statistics of the end-to-end packet channel. The optimum solution also depends on the fidelity criterion for the specific application. The case of independent and iid losses and delays leads to an almost-water-filling solution for the packet interleaving proportions. The presence of memory changes the solution making the first and the last components become prevalent. The influence of losses and tolerated delay has been considered to distinguish scenario in which interleaving can be more or less effective, as well as some simple interleaving configurations are considered to give idea of the advantage. Future work will be focused on multiple paths scenario, i.e. find the optimum transmission strategies when source information can be distributed over different packets and different Internet paths.

6. REFERENCES

[1] V. Paxson, *End-to-End Internet Packet Dynamics*. IEEE Trans. on Networking, Vol. 7(3), pp. 277–292, June 1999.

[2] V.K. Goyal, and J. Kovacevic, *Generalized Multiple Description Coding with Correlating Transforms*. IEEE Trans. on Information Theory, Vol. 47(6), pp. 2199–2224, Sept. 2001.

[3] J.N. Laneman, E. Martinian, G.W. Wornell, J.G. Apostolopoulos, S.J. Wee *Comparing Application- and Phisical-Layer Approaches to Diversity on Wireless Channels*. Procs. of IEEE ICC, pp. 2678–2682, May 2003.

[4] E. Martinian, C.-E.W. Sundberg, *Decreasing Distortion Using Low Delay Codes for Bursty Packet Loss Channels*. IEEE Trans. on Multimedia, Vol. 5(3), pp. 285–292, Sept. 2003.

[5] T. Nguyen, A. Zakhor, *Path Diversity with Forward Error Correction (PDF) System for Packet Switched Networks*. Procs. of IEEE INFOCOM, pp. 663–672, Apr. 2003.

[6] P. Salvo Rossi, G. Romano, F. Palmieri, and G. Iannello, *Interleaving for Packet Channels*. Procs. of CISS, pp. 1560–1564, Mar. 2004.