Abstract

An (n, M) vector code $\mathcal{C} \subseteq \mathbb{F}^n$ is a collection of M codewords where n elements (from the field $\mathbb{F}$) in each of the codewords are referred to as code blocks. Assuming that $\mathbb{F} \cong \mathbb{B}^\ell$, the code blocks are treated as ℓ-length vectors over the base field $\mathbb{B}$. Equivalently, the code is said to have the sub-packetization level ℓ. This paper addresses the problem of constructing MDS vector codes which enable exact reconstruction of each code block by downloading small amount of information from the remaining code blocks. The repair bandwidth of a code measures the information flow from the remaining code blocks during the reconstruction of a single code block. This problem naturally arises in the context of distributed storage systems as the node repair problem [Dimakis et al.]. Assuming that $M = |\mathbb{B}|^{k\ell}$, the repair bandwidth of an MDS vector code is lower bounded by $\big((n-1)/(n-k)\big) \cdot \ell$ symbols (over the base field $\mathbb{B}$) which is also referred to as the cut-set bound [Dimakis et al.]. For all values of n and k, the MDS vector codes that attain the cut-set bound with the sub-packetization level $\ell = (n-k)^{\lceil n/(n-k) \rceil}$ are known in the literature [Sasidharan et al., Ye and Barg]. This paper presents a construction for MDS vector codes which simultaneously ensures both small repair bandwidth and small sub-packetization level. The obtained codes have the smallest possible sub-packetization level $\ell = O(n-k)$ for an MDS vector code and the repair bandwidth which is at most twice the cut-set bound. The paper then generalizes this code construction so that the repair bandwidth of the obtained codes approach the cut-set bound at the cost of increased sub-packetization level. The constructions presented in this paper give MDS vector codes which are linear over the base field $\mathbb{B}$.