



**Description:** Two-level cache hierarchy tree consisting of  $n_l$  leaves. The random variable  $d_i$  encodes the random delay for cache  $i$  to download the content from its parent. The server is assumed to contain all objects. The request rate of object  $i$  at leaf  $j$  is denoted  $\lambda_{i,j}$ .