Erasure Coding for Decentralized Caching

Soheil Mohajer
University of Minnesota

Mohammad Ali Maddah-Ali
Bell Labs, Alcatel-Lucent

14 Feb. 2017, ITA, San Diego

Problem Formulation

- N files $X_1, X_2, X_3, \ldots, X_N, \quad |X_i| = F$
- K users, each with enough local storage to fetch M files
  - without knowledge of the users' request

$$Z_i = f_i(X_1, \ldots, X_N) \qquad\qquad |Z_i| \leq MF$$

  - coded / uncoded data to be cached
  - centrally decided / random placement
- Each user requests for one file in the library

$$\mathbf{d} = (d_1, d_2, \ldots, d_K) \in \{1, 2, \ldots, N\}^K$$

Problem Formulation

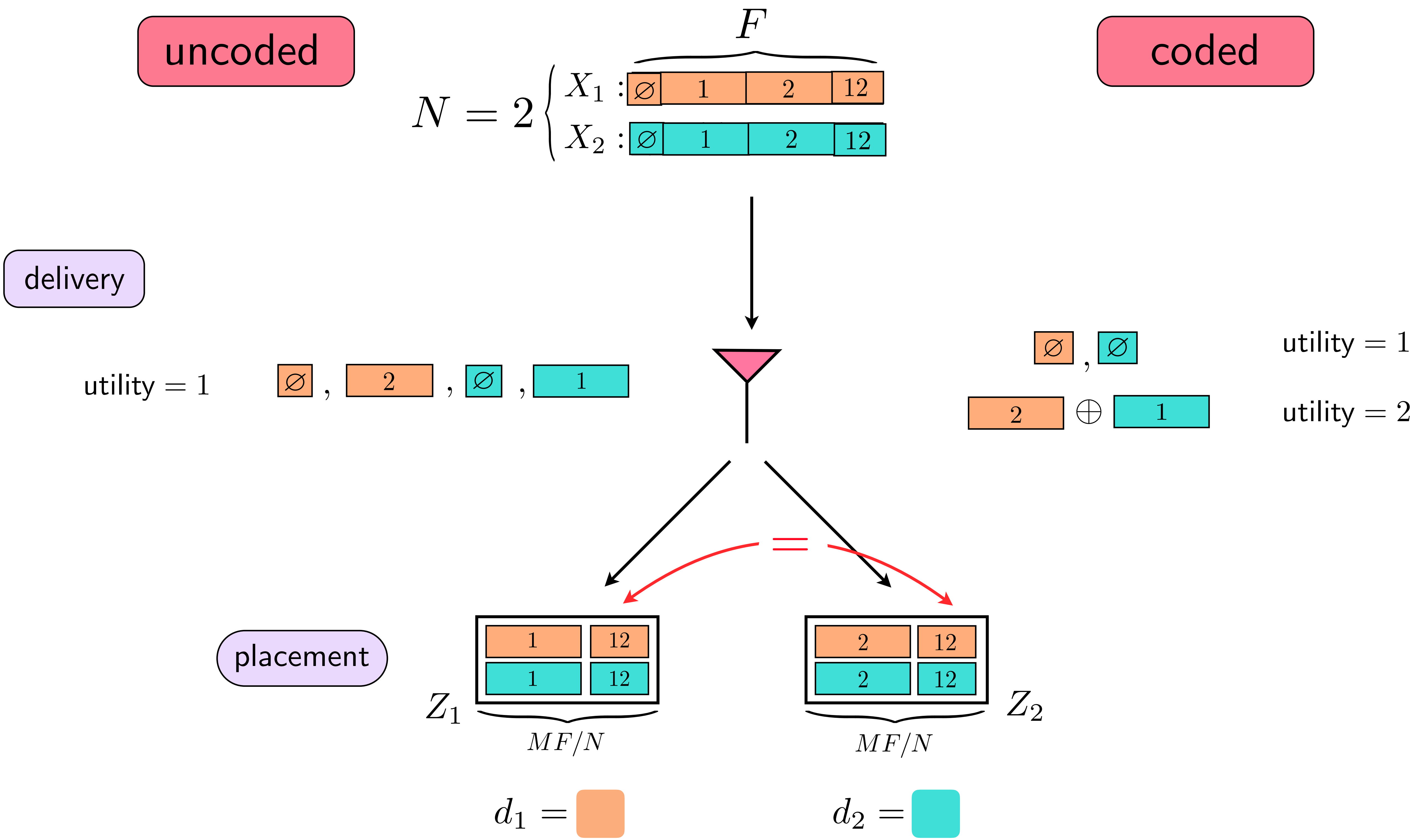
- Delivery

$$Y = g(X_1, \ldots, X_N; \mathbf{d}; Z_1, \ldots, Z_K)$$

such that: $(Y, Z_k) \rightsquigarrow X_{d_k} \qquad k = 1, 2, \ldots, K$

- Questions:
  - Optimal placement?
  - Minimum delivery rate $R = \frac{|Y|}{F}$

Uncoded vs. Coded Delivery



Centralized vs. Decentralized Placement

- Centralized: Placement is designed to maximize the utility

$$R_C(M) = K\left(1 - \frac{M}{N}\right)\min\left\{\frac{1}{1+KM/N}, \frac{N}{K}\right\}.$$

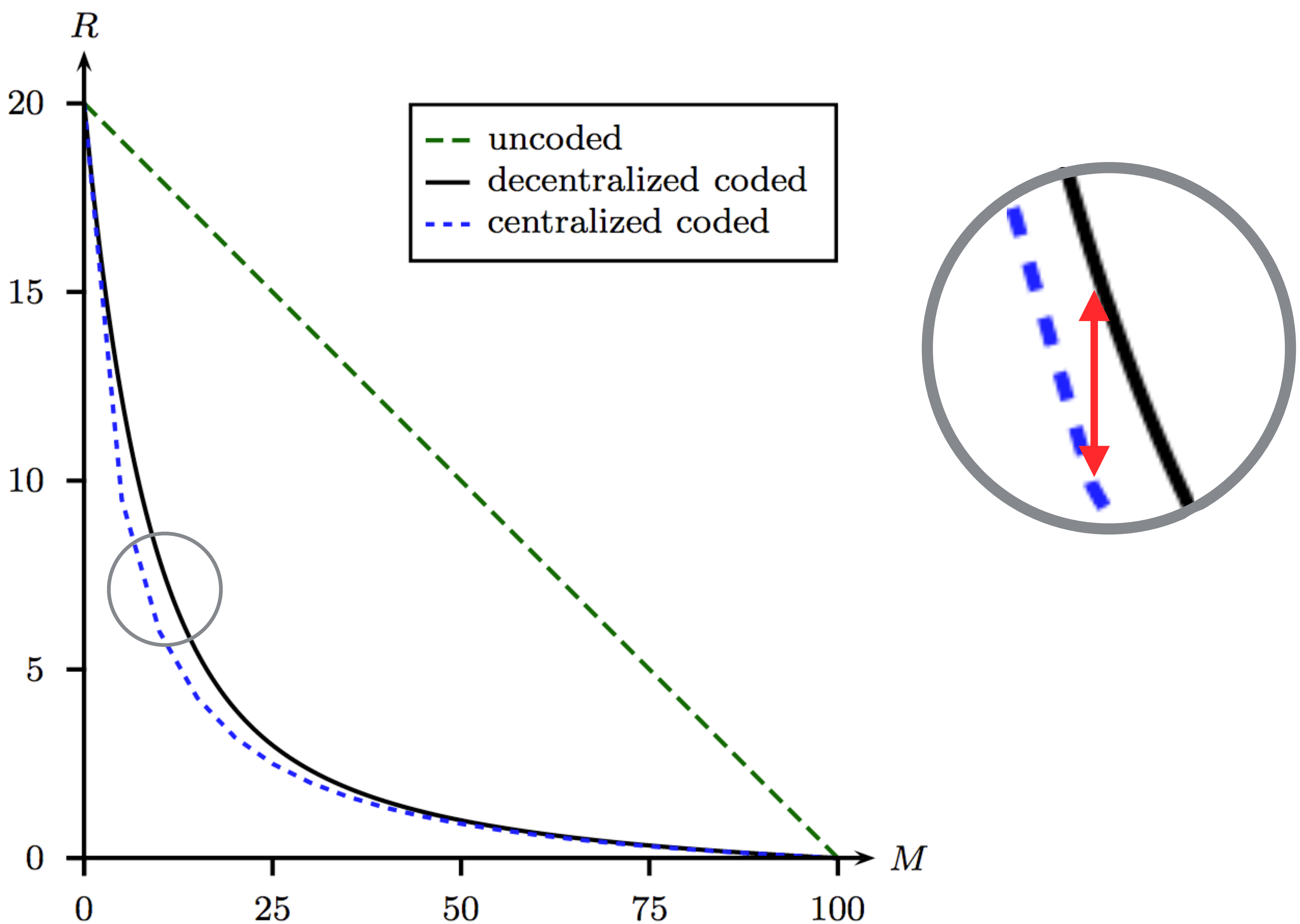
[Maddah-Ali & Niesen, Trans. Inform.'14]

- Decentralized: No pre-setting is required, random fetching

$$R_D(M) = K\left(1 - \frac{M}{N}\right)\min\left\{\frac{N}{KM}(1-(1-M/N)^K), \frac{N}{K}\right\}.$$

[Maddah-Ali & Niesen, Trans. Networking'15]

Centralized vs. Decentralized Caching



$$\frac{R_D(M)}{R_C(M)} \leq 12$$

numerically, 1.6

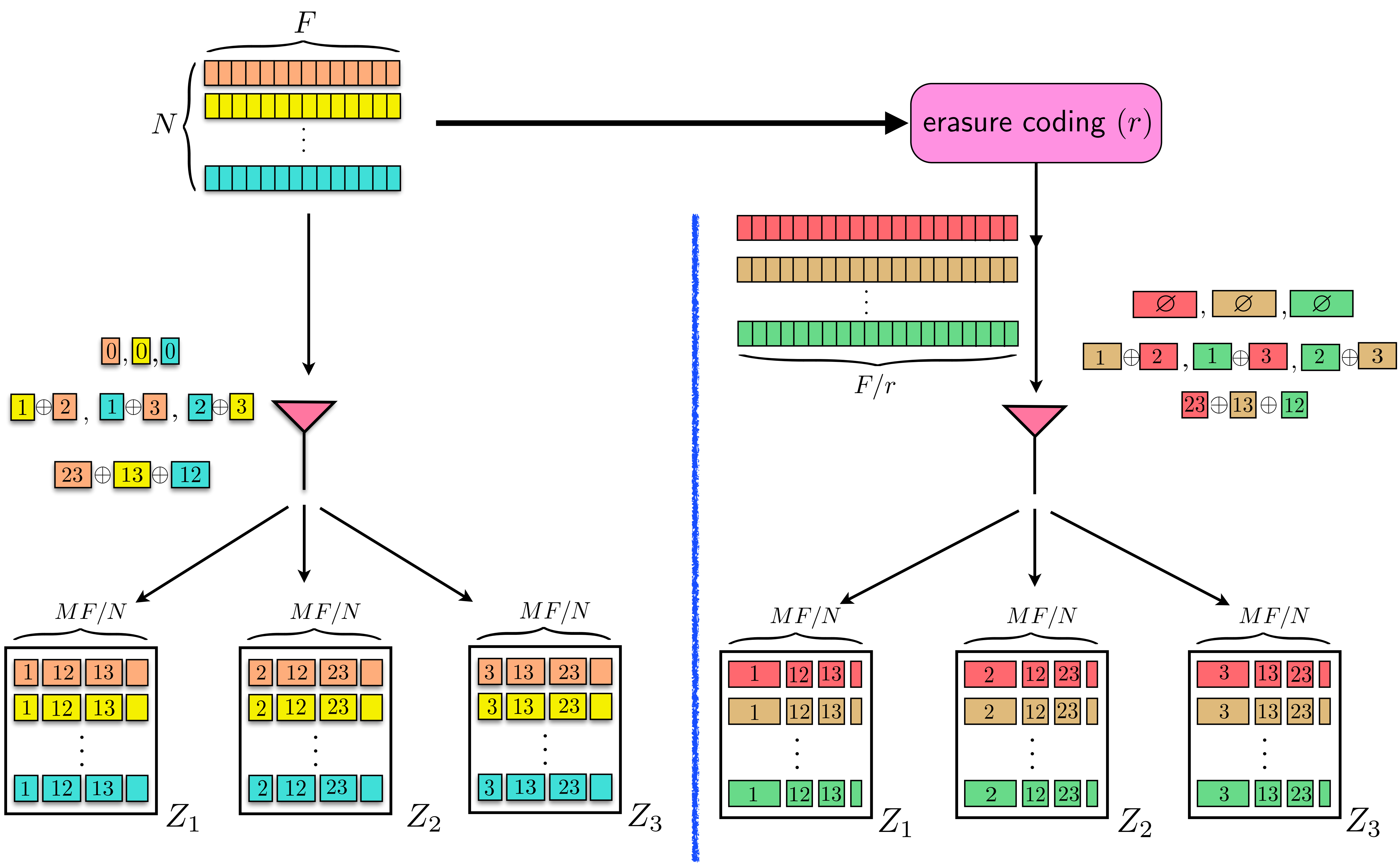
$K = 20$ users, $N = 100$ files

[Maddah-Ali & Niesen, Trans. Networking'15]

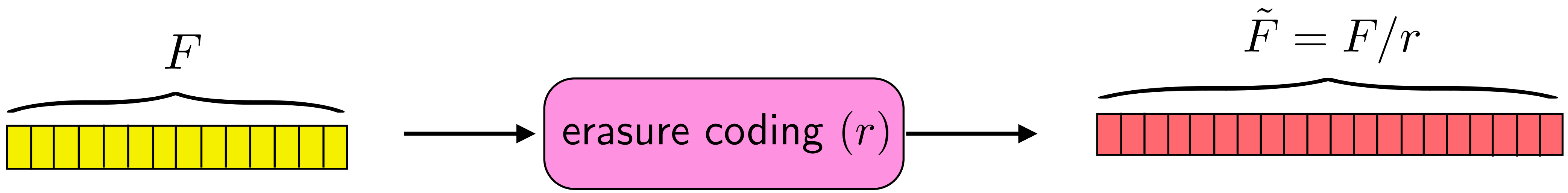
Uncoded vs. Coded Placement

- Uncoded: cache contents being union of pure packets from files
  - Several bounds with constant multiplicative gaps
    [Maddah-Ali et al, 15-16], [Ghasemi et al, 16], [Wang et al, 16], [Sengupta at al, 16], Ajaykrishnan et al, 15], [Tian, 15]
  - Proposed schemes are optimum when $K \leq N$
    [Yu et al, 16], [Kai et al, 16]
  - Delivery rate can be improved when $K \geq N$
    [Yu et al, 16], [Wan et al, 16]
- Coded placement: data can be coded before placement
  - Focus on joint coding (mixture) of files
  - Delivery rate can be improved! [Maddah-Ali et al, 15], [Tian & Chen.'16],…

Erasure Coding for Random Placement



Erasure Coding for Caching



- probability of caching each packet

$$q = \frac{MF}{N\tilde{F}} = \frac{Mr}{N}$$

- number of packets can be jointly encoded for u users

$$\beta_u = q^{u-1}(1-q)^{K-u+1}\frac{F}{r}$$

- number of desired packets already cached

$$\frac{M}{N}F = \frac{q}{r}F$$

Erasure Coding for Caching

- Total number of desired packets cached or received at user

$$\frac{q}{r}F + \sum_{u=1}^{K} \binom{K-1}{u-1} \lambda_u \beta_u \geq F$$

$\mathsf{pre-fetching}$ $\quad$ $\mathsf{delivery}$

- Total number of packets to be sent

$$R(r; \lambda_1, \ldots, \lambda_K) = \frac{1}{F} \sum_{u=1}^{K} \binom{K}{u} \lambda_u \beta_u$$

Optimization Problem

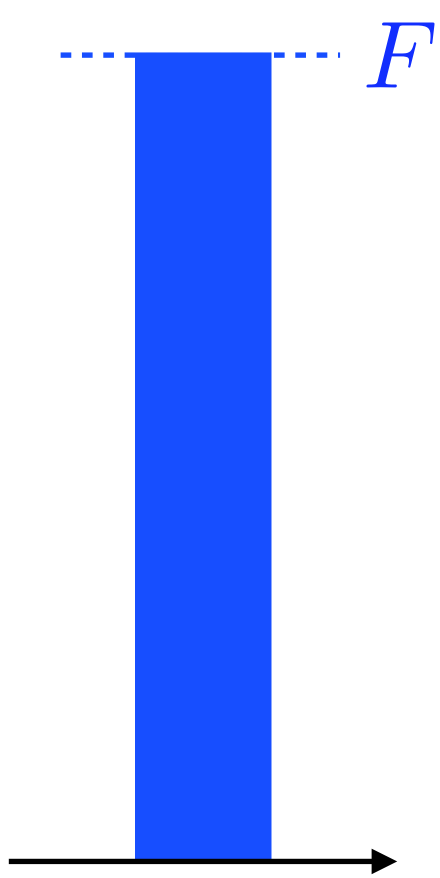
- Optimization over $\lambda_i's$

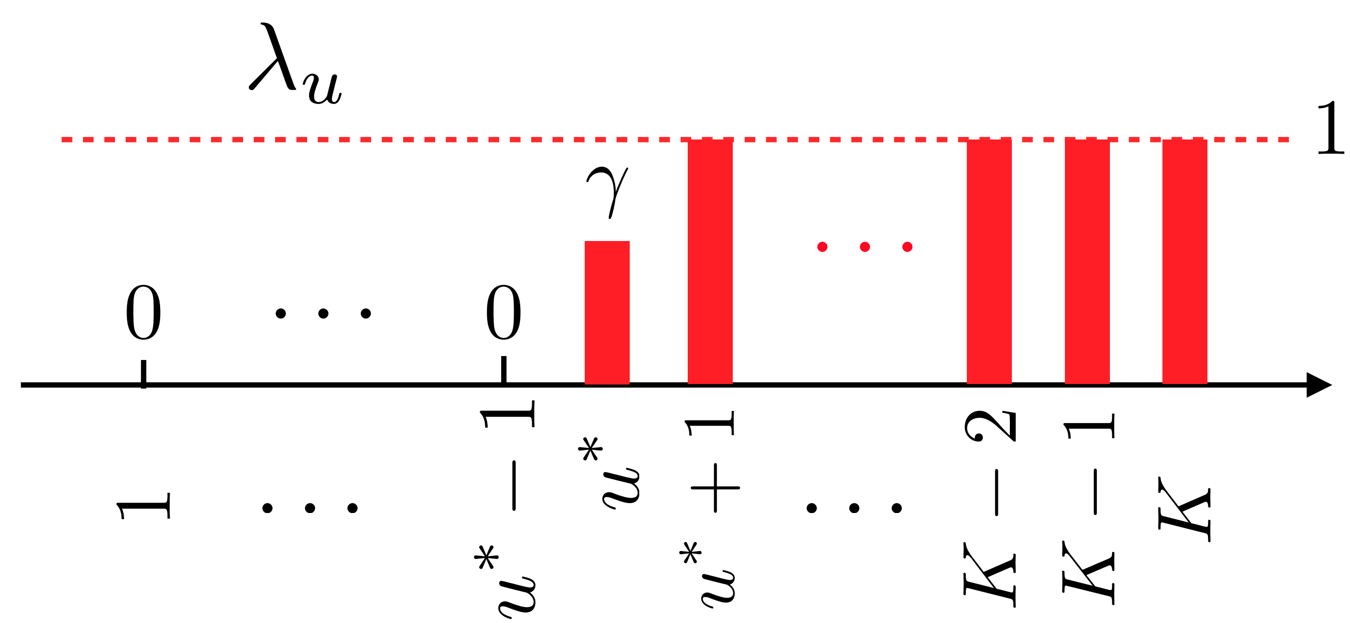
$$R(r) = \min_{\lambda_1,\dots,\lambda_K} \sum_{u=1}^{K} \binom{K}{u} \lambda_u \beta_u$$

$$s.t. \qquad \frac{q}{r}F + \sum_{u=1}^{K} \binom{K-1}{u-1} \lambda_u \beta_u \geq F$$

$$\lambda_u \in [0,1] \qquad u = 1, \dots, K$$

High-utility packets are preferable!





Optimization Problem

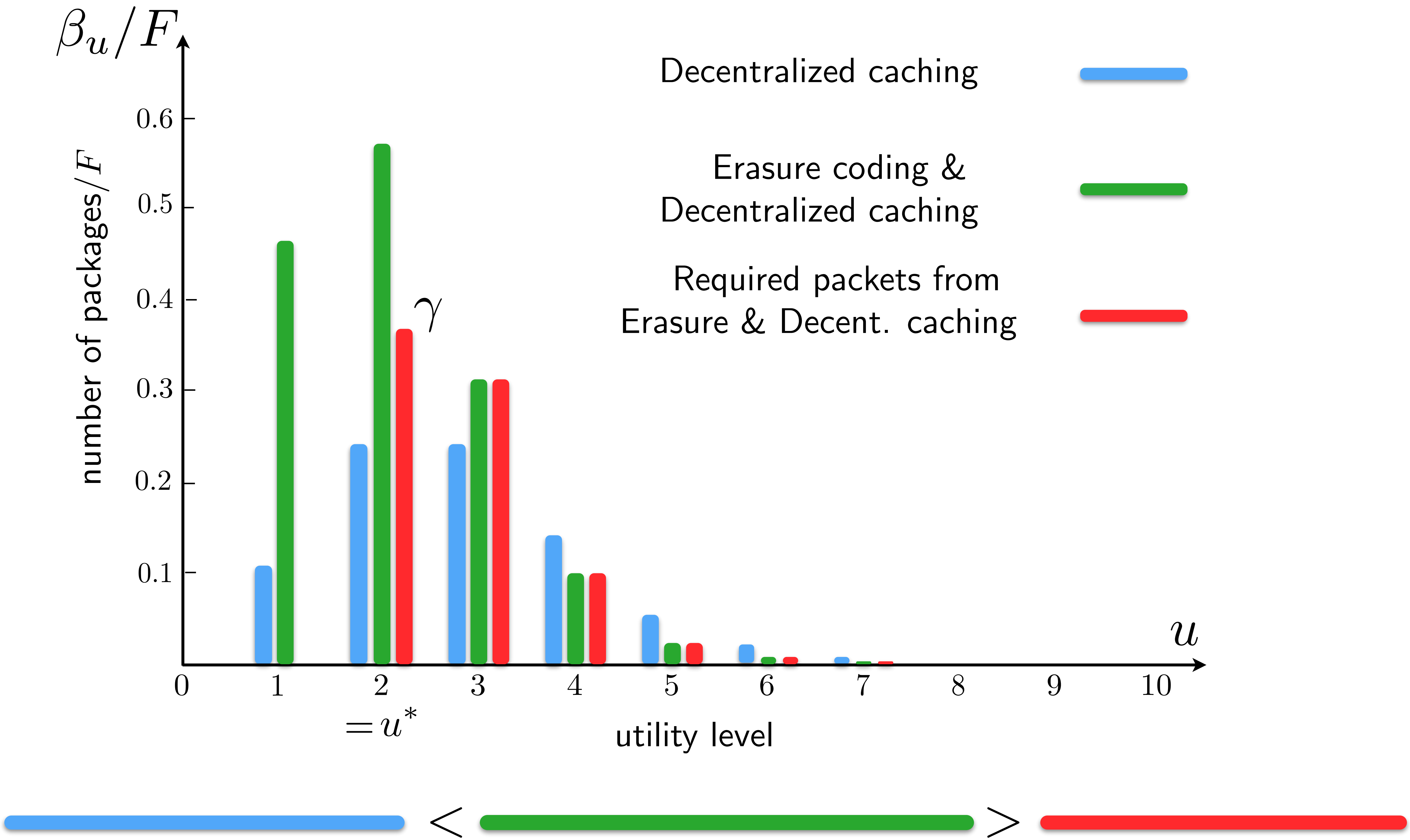
$$R(r) = \gamma \binom{K}{u^*} \beta_{u^*} + \sum_{u=u^*+1}^{K} \binom{K}{u} \beta_u$$

where γ and u^* are uniquely determined from

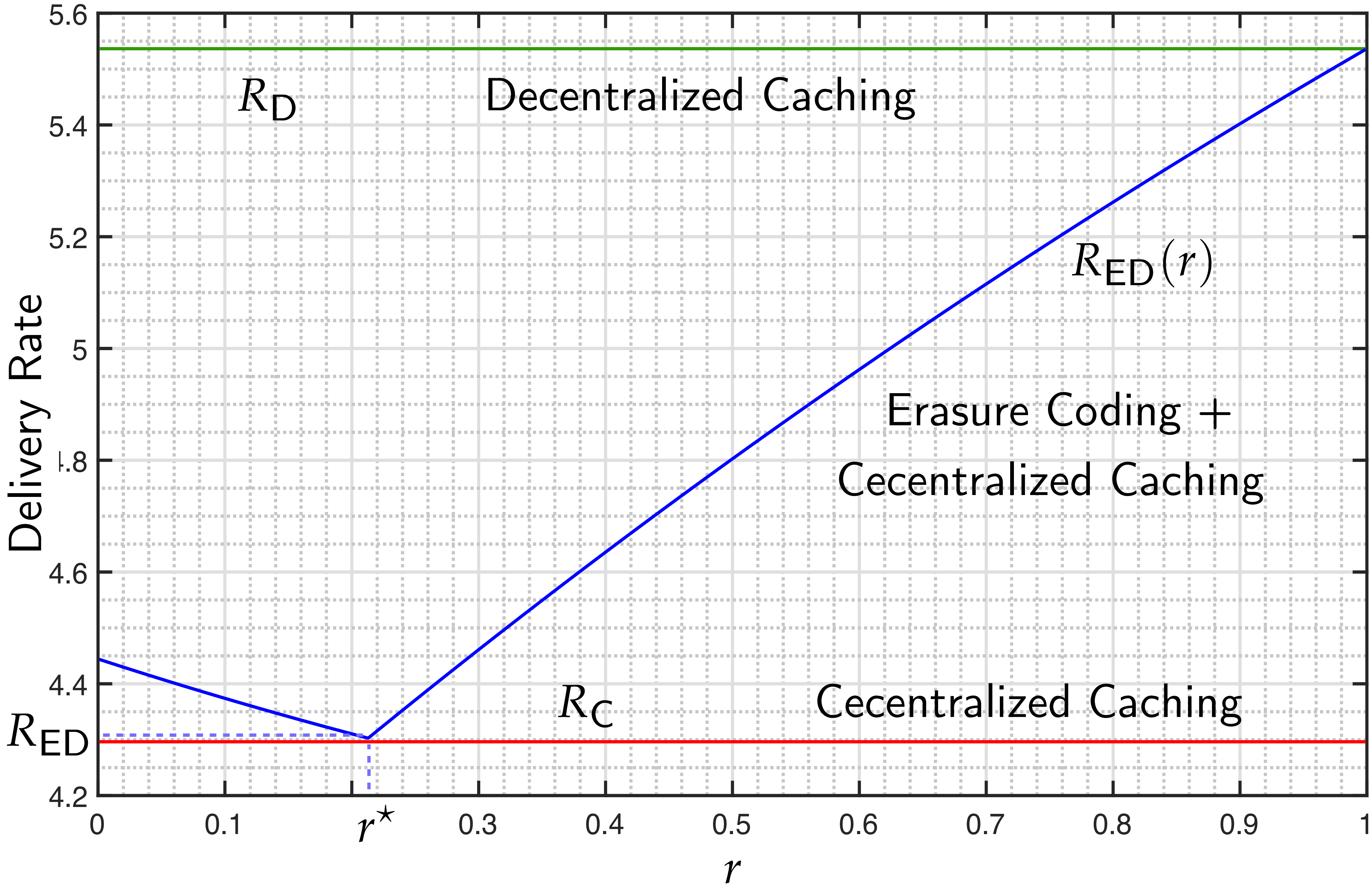
$$\frac{q}{r} F + \gamma \binom{K-1}{u^*-1} \beta_{u^*} + \sum_{u=u^*+1}^{K} \binom{K-1}{u-1} \beta_u = F$$

$$R^* = \min_{0<r\leq 1} R(r)$$

Skewed Histogram vs. Cut-off

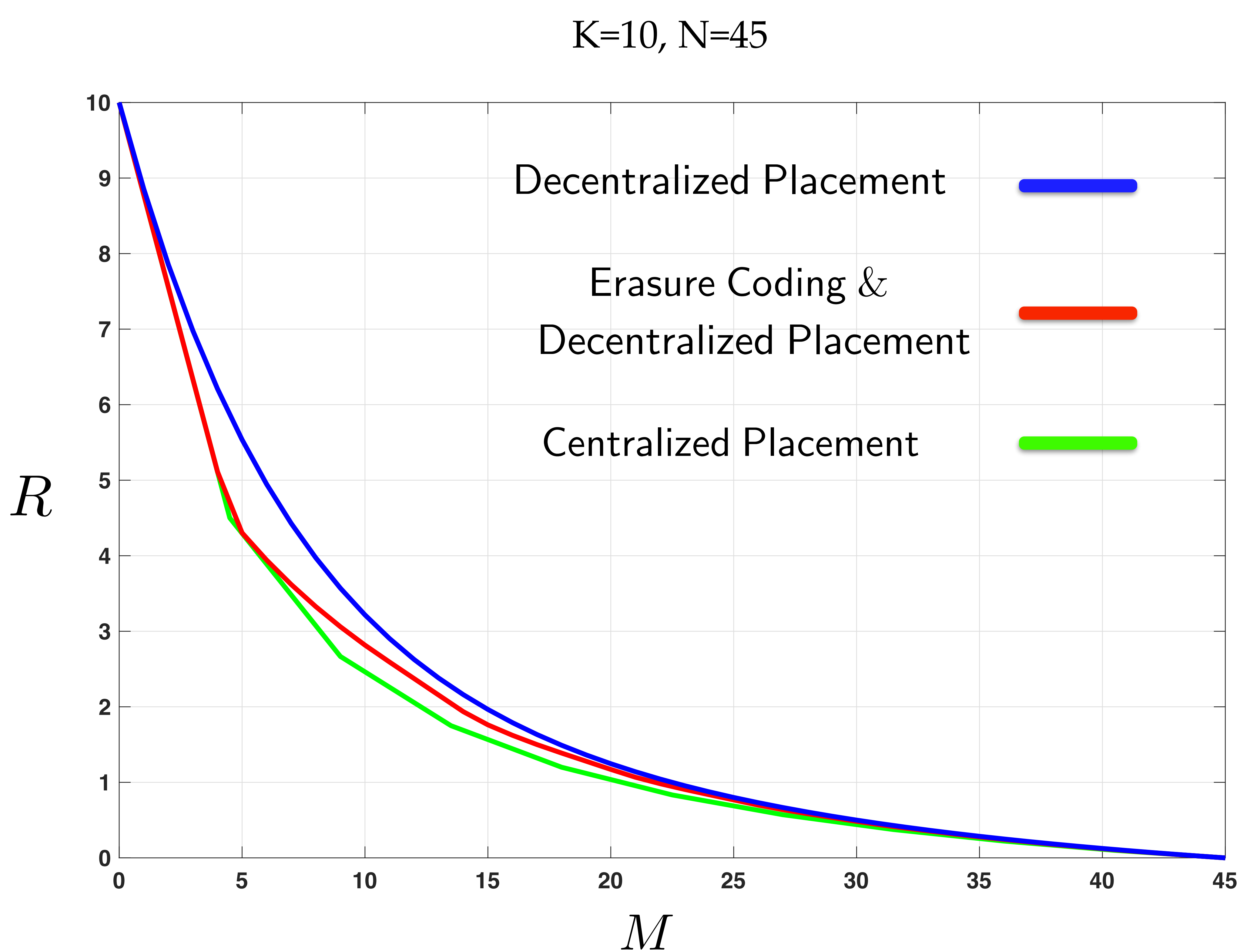


It works!



K=10, N=45, M=5

Cache-Rate Trade-off



A few Remarks

- Erasure pre-coding can improve the delivery rate of decentralized caching
- It is low-complexity compared to file combination
- Asymptotically meets the centralized caching for small cache size

Questions!