Given a set of files $\mathcal{F} = \{F_1, ..., F_f\}$ with equal size s, where s is the number of packets in F_i, and we generate $s' = (1+\theta)s$ packets where $\theta \in [0,1]$. We represent each file $F_i \in \mathcal{F}$ as a node $F_i =< s_i, d_i, Generate_i >$ where s_i, d_i denote the source and destination for that file and $Generate_i$ is the time when the file is generated. For each node $F_i =< s_i, d_i, Generate_i >$, we add a virtual source node sv_i having an edge with capacity s' into file node F_i. And we add a virtual destination node dv_i having an incoming edge of infinite capacity from all the connection nodes which contain d_i as destinations. In this way we can model this problem as multi-commodity flow problem with f different kinds of commodities with f source and destination pairs of (sv_i, dv_i).

Formulation 1: Maximize the total number of files that can be successfully decoded.

Let $\lambda \in [0,1]$ be the threshold percentage for each file to be successfully decoded, which means if $\lambda s'$ of s' packets can be delivered to destination, file F_i can be treated as successfully delivered. We use a variable x_i to represent whether or not F_i is successfully delivered. This problem thus becomes an Integer Linear Programming (ILP).

$$
\begin{aligned}
\text{Input: Directed Connection Graph } & G(V,E), \text{File set } \mathcal{F} \text{ with } f \text{ files} \\
\text{Source-sink pair } & (sv_i, dv_i) \text{ and threshold } \lambda,\ \forall F_i \in F \\
\text{Capacity } & c(u,v) \quad \forall (u,v) \in E \\
\text{Variable: Flow } & f^i_{(u,v)} \quad \forall F_i \in F, (u,v) \in E \\
& x_i \quad \forall F_i \in F \\
\text{Maximize } & \sum_{i=1}^{f} x_i \\
\text{Subject to } & \sum_{(sv_i,v)\in E} f^i_{(sv_i,v)} = f_i, && \forall F_i \in F \\
& \sum_{(u,v)\in E} f^i_{(u,v)} = \sum_{(v,u)\in E} f^i_{v,u}, && \forall F_i \in F, \forall v \in V - \{sv_i, dv_i\} \\
& \sum_{i=1}^{f} f^i_{(u,v)} \le c_{(u,v)}, && \forall (u,v) \in E \\
& f^i_{(u,v)} \ge 0, && \forall F_i \in F, \forall (u,v) \in E \\
& f^i \le s', && \forall F_i \in F \\
& f^i \ge x_i \lambda s', && \forall F_i \in F \\
& x_i \in \{0,1\}, && \forall x_i
\end{aligned} \tag{1}
$$

Let OPT_f be the total amount commodities delivered by solving the linear relaxation LP of IP in Formulation 1 where the variables x_i are relaxed to assume any value in $[0,1]$. It is obvious to see that $OPT_f = \sum \tilde{x_i} \times (1+\epsilon)$, and let $\beta(\epsilon) = (1+\epsilon)ln(1+\epsilon) - \epsilon$. Define the solution from the above algorithm as ALG, and total amount of commodities delivered by ALG as OPT_{ALG}.

Claim 1. $Pr[OPT_{ALG} < \frac{1-\epsilon}{1+\epsilon} \cdot OPT_f] \le e^{-\beta(-\epsilon)\cdot OPT_f/(1+\epsilon)}$

Proof: For each commodity i, the expectation of x_i is $E(x_i) = 1 \cdot \tilde{x_i} + 0 \cdot (1 - \tilde{x_i}) = \tilde{x_i}$. Let $\mu = \sum \tilde{x_i} = \frac{OPT_f}{1+\epsilon}$, and $OPT_{ALG} = \sum x_i$.

Algorithm 1 Random Rounding Algorithm for Formulation 1

Input:

Directed Connection Graph $G(V, E)$;

File set $F = \{F_1, ..., F_f\}$;

Source-sink pair (sv_i, dv_i) and threshold λ associated for each $F_i \in F$;

Capacity $c(u, v),\ \forall (u, v) \in E$;

Output:

1: Divide all edge capacities by $1 + \epsilon$, where $0 < \epsilon < 1$, i.e., $\tilde{c_e} = \frac{c_e}{(1+\epsilon)}$;

2: Change the last constraint to be $0 \le x_i \le 1$;

3: Then it is relaxed to an LP, solve this LP and get optimal solution $\tilde{x_i}$;

4: With probability $\tilde{x_i}$, set $x_i = 1$, otherwise set it to 0;

5: Scale up the fractional flow $\tilde{f_{j,e}}$ from the LP solution on edge e for commodity j by $\tilde{x_j}$, i.e., $f_{j,e} = \tilde{f_{j,e}} \times \frac{1}{\tilde{x_j}}$;

$$
\begin{aligned}
Pr[OPT_{ALG} < \frac{1-\epsilon}{1+\epsilon} \cdot OPT_f] &= Pr[\sum x_i < (1-\epsilon) \cdot \mu] \\
&\le e^{-\beta(-\epsilon)\mu} \\
&\le e^{-\beta(-\epsilon) \cdot OPT_f / (1+\epsilon)}
\end{aligned} \tag{2}
$$

□

If $OPT_f \ge \alpha \cdot \ln f$, then, $Pr[OPT_{ALG} \ge \frac{1-\epsilon}{1+\epsilon} \cdot OPT_f]$ is at least $1 - \frac{1}{poly(f)}$.

If $OPT_f < \alpha \cdot \ln f$, then, $OPT_{ALG} < \alpha \cdot \ln f$, even if we get 0 file, we are off from OPT_f by no more than $\alpha \cdot \ln f$ additive factor. Putting it all together, with probability $1 - \frac{1}{poly(f)}$, we get at least $\max\{0, (1-\epsilon)/(1+\epsilon) \cdot OPT_f - \alpha \ln f\}$.

Claim 2. $Pr[ALG\ does\ not\ satisfy\ edge\ capacity\ contraints] \le \frac{1}{|E|}$

Proof: Fix an edge $e \in E$, for commodity j, with probability $1 - \tilde{x_j}$, the flow on edge e for commodity j is set to 0, i.e., $f_{j,e} = 0$, with probability $\tilde{x_j}$, the flow on edge e for commodity j is set to $\tilde{f_{j,e}} \cdot \frac{1}{\tilde{x_j}}$. Then the expectation of $f_{j,e}$ is $E(f_{j,e}) = \tilde{f_{j,e}} \cdot \frac{1}{\tilde{x_j}} \cdot \tilde{x_j} + 0 \cdot (1 - \tilde{x_j}) = \tilde{f_{j,e}}$. Let F_e denotes the flow on edge e by ALG, then $F_e = \sum_{j, f_{j,e} \neq 0} f_{j,e}$ and the expectation of F_e is $E[F_e] = \sum_{j, f_{j,e} \neq 0} \tilde{f_{j,e}} \cdot \frac{1}{\tilde{x_j}} \cdot \tilde{x_j} = \sum_{j, f_{j,e} \neq 0} \tilde{f_{j,e}} \le \tilde{c_e}$.

Let $X_j = \frac{f_{j,e}}{d}$, then, $F(e) \ge (1+\epsilon) \cdot \tilde{c_e}$ iff $\sum_{j, f_{j,e} \neq 0} X_j \ge (1+\epsilon) \cdot \frac{\tilde{c_e}}{d}$. Therefore, $Pr[F(e) \ge (1+\epsilon) \cdot \tilde{c_e}] \le e^{-\beta(\epsilon) \cdot \tilde{c_e}/d}$. By scaling up the capacities by ϵ, i.e., $c_e = (1+\epsilon) \cdot \tilde{c_e}$, and $\beta(\epsilon) \ge \frac{2\epsilon^2}{4.2+\epsilon}$, and by assumption $\tilde{c_e}/d \ge \frac{4.2+\epsilon}{\epsilon^2} \cdot ln|E|$, we have $Pr[F(e) \ge c_e] \le \frac{1}{|E|^2}$. By applying union bound over all edges, we can get $Pr[ALG\ does\ not\ satisfy\ edge\ capacity\ contraints] \le \frac{1}{|E|}$.