Given a set of files $\mathcal{F} = \{F_1, ..., F_n\}$ with equal size s, where s is the number of packets in F_i, and we generate $s_g = (1+\theta)s$ 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_g 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 n different kinds of commodities with n source and destination pairs of (sv_i, dv_i). Assume that all commodity can be fully fractionally delivered alone, which means each commodity can always be delivered in the form of splittable packets in the absence of other commodities.

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

Let $\lambda \in [0,1], \lambda << \theta$ be the threshold percentage for each file to be successfully decoded, which means if $s_r = (1+\lambda)s$ packets can be delivered to destination, file F_i can be treated as successfully delivered. We use a variable f_i to represent whether or not F_i is successfully delivered. For the sake of simplicity, we rescale each file and edge capacity in the network by s_r so that this problem thus becomes an unit flow multi-commodity problem.

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

Algorithm 1 Random Rounding Algorithm for Formulation 1

Input:

Directed Connection Graph $G(V, E)$;
File set $F = \{F_1, ..., F_n\}$;
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: OPT_{ALG}

1: Change the last constraint to be $0 \leq f_i \leq 1$;
2: Then it is relaxed to an LP, solve this LP and get optimal solution $\tilde{f}_i$;
3: With probability $\tilde{f}_i$, set $f_i = 1$, otherwise set it to 0;
4: Scale up the fractional flow $\tilde{f_{i,e}}$ from the LP solution on edge e for commodity i by $\frac{1}{\tilde{f}_i}$, i.e., $f_{i,e} = \tilde{f_{i,e}} \times \frac{1}{\tilde{f}_i}$;

Let OPT_i be the optimal solution of the IP in Formulation 1, and let OPT_f be the total amount commodities delivered by solving the linear relaxation LP of IP in Formulation 1 where the variables f_i are relaxed to assume any value in $[0, 1]$. It is obvious to see that $OPT_f = \sum \tilde{f}_i$. Define the solution from the above algorithm as ALG, and total amount of commodities delivered by ALG as OPT_{ALG}.

We use Chernoff Bound over continuous random variables to bound the probability of achieving a fraction of the optimal solution.

Fact 1 (Chernoff-Bound) . Let $X = \sum_{i=1}^{n} X_i$ be a sum of n independent random variables $X_i \in [0, 1], 1 \leq i \leq n$. Then $P(X < (1 - \epsilon) \cdot E(X)) \leq exp(-\epsilon^2 \cdot E(X)/2)$ holds for $0 < \epsilon < 1$.

Since we have already scaled down the flow by s_r, thus the variables are between 0 and 1 so that we can apply the above Chernoff Bound.

Claim 2. $Pr[OPT_{ALG} < (1 - \epsilon) \cdot OPT_f] \leq e^{-\epsilon^2 \cdot OPT_f/2}$

Proof: For each commodity i, the expectation of f_i is $E(f_i) = 1 \cdot \tilde{f}_i + 0 \cdot (1 - \tilde{f}_i) = \tilde{f}_i$. Recall that $OPT_f = \sum \tilde{f}_i$, and let $OPT_{ALG} = \sum f_i$.

$$\begin{aligned} Pr[OPT_{ALG} < (1 - \epsilon) \cdot OPT_f] &= Pr[\sum f_i < (1 - \epsilon) \cdot OPT_f] \\ &\leq e^{-\epsilon^2 \cdot OPT_f/2} \end{aligned} \tag{2}$$ $\square$

Since we also assume that all commodity can be fully fractionally delivered alone, therefore, we get:

$$OPT_f \geq 1 \tag{3}$$

Since we have proved that the optimal fractional solution is an upper bound of optimal solution from ILP problem, by taking $\epsilon = 2/3$ in the Chernoff Bound, we get,

Theorem 3. The probability of achieving less than 1/3 of the profit of an optimal solution is upper bounded by $e^{-2/15} \approx 0.87$.

Proof. By taking $\epsilon = 2/3$, Equation 2 becomes:

$$Pr[OPT_{ALG} < \frac{1}{3} \cdot OPT_f] \leq e^{-(\frac{2}{3})^2 \cdot OPT_f/2} \tag{4}$$

$$= e^{-2 \cdot OPT_f/9} \tag{5}$$

By Equation 3, we know that the minimum value of OPT_f is 1, by taking $OPT_f = 1$, we get the upper bound of $e^{-\epsilon^2 \cdot OPT_f/2}$. Therefore,

$$Pr[OPT_{ALG} < \frac{1}{3} \cdot OPT_f] \leq e^{-2/9} \tag{6}$$

Since $OPT_i \leq OPT_f$, thus we get,

$$Pr[OPT_{ALG} < \frac{1}{3} \cdot OPT_i] \leq e^{-2/9} \qquad \square$$

Claim 4 Given a single edge, choose ϵ' such that $\max \frac{\tilde{f_{i,e}}}{\tilde{f_{i,e}}} \leq \epsilon'$ .The probability that ALG exceeds the edge capacity constraint by a factor of $\gamma = (1 + \epsilon' \cdot \sqrt{2 \log |V| \Delta_F})$ is bounded by $|V|^{-4}$

We analyze the probability that our algorithm violates capacity constraints by a certain factor, we employ Hoeffding's Inequality

Fact 5 (Hoeffding's Inequality) **[?]** . Let $\{X_i\}$ be independent random variables, s.t. $X_i \in [a_i, b_i]$, then $Pr(\sum_i X_i - E(\sum_i X_i) \geq t) \leq exp(-2t^2/\sum_i (b_i - a_i)^2)$ holds.

Proof: Fix an edge $e \in E$, for commodity i, with probability $1 - \tilde{f_i}$, the flow on edge e for commodity i is set to 0, i.e., $f_{i,e} = 0$, with probability $\tilde{f_i}$, the flow on edge e for commodity i is set to $\tilde{f_{i,e}} \cdot \frac{1}{\tilde{f_i}}$. Let Δ_F be the number of files going through edge e.

Then the expectation of $f_{i,e}$ is

$$E(f_{i,e}) = \tilde{f_{i,e}} \cdot \frac{1}{\tilde{f_i}} \cdot \tilde{f_i} + 0 \cdot (1 - \tilde{f_i}) = \tilde{f_{i,e}} \tag{7}$$

Let F_e denotes the flow on edge e by ALG, then $F_e = \sum_{i, f_{i,e} \neq 0} f_{i,e}$ and the expectation of F_e is

$$E[F_e] = \sum_{i, f_{i,e} \neq 0} \tilde{f_{i,e}} \cdot \frac{1}{\tilde{f_i}} \cdot \tilde{f_i} = \sum_{i, f_{i,e} \neq 0} \tilde{f_{i,e}} \tag{8}$$

Let $t = \epsilon' \cdot \sqrt{2 \log |V| \Delta_F}$, by applying Hoeffding's Inequality, we get,

$$\begin{aligned} Pr[F_e - E(F_e) \geq t] &\leq exp(\frac{-2t^2}{\sum_i (\frac{\tilde{f_{i,e}}}{\tilde{f_{i,e}}})^2}) && (9) \\ &\leq exp(\frac{-2 \cdot \epsilon'^2 \cdot \log |V| \cdot \Delta_F}{\epsilon'^2 \cdot \Delta_F}) && (10) \\ &= |V|^{-4} && (11) \end{aligned}$$

Given Equation 8, and let $\gamma = (1 + \epsilon' \cdot \sqrt{2 \log |V| \Delta_F})/(1 + \epsilon)$,

$$\begin{aligned} & Pr[F_e - \tilde{c_e} \geq \epsilon' \cdot \sqrt{2 \log |V| \Delta_F}] && (12) \\ = \; & Pr[F_e \geq (1 + \epsilon' \cdot \sqrt{2 \log |V| \Delta_F})] && (13) \\ = \; & Pr[F_e \geq \gamma \cdot c_e] && (14) \\ \leq \; & |V|^{-4} && (15) \end{aligned}$$

There are at most $|V|^2$ edges, by applying union bound over all edges using Claim 4, we can get that the probability that ALG exceeds any of the edge capacity constraints by a factor of $\gamma = (1 + \epsilon' \cdot \sqrt{2 \log |V| \Delta_F})$ is upper bounded by $|V|^{-2}$.