In this section, we focus on the reduction of system energy consumption while ensuring the users' service requirements. We formulate the optimization problem for process-oriented user scheduling. Moreover, we proposed an efficient 3-step iterative algorithm with polynomial time complexity to solve the NP-hard problem.

A. Problem formulation

To fully utilize the slow time-varying characteristic of the large-scale channel fading, we divide the total service time into T time slots, each lasts Δt. The value Δt is carefully chosen so that $\beta_{i,j,t}$ remains constant in each time slot t. Thus, we make it possible to estimate $\beta_{i,j,t}$ for $\forall t \in \{1,...,T\}$ based on shipping lanes and timetable. With the large-scale channel fading known beforehand, we can further design and implement a process-oriented scheme foe user scheduling.

The total energy consumption of the system consists of cellular (BS) transmission part and D2D transmission part. Our objective is to minimize the system energy consumption by means of user scheduling in cellular transmission and D2D transmission. We further denote the transmission link from BS/user $i \in \{0,1,...,J\}$ ($i=0$ means BS, $i>0$ means user relay) to user $j \in \{1,...,J\}$ at time slot $t \in \{1,...,T\}$ by $i \to j@t$. We assume BS/user always use their max power $P_i^{\max}$ for transmission $i \to j@t$ and therefore the transmission speed is always at its maximum. For the transmission link $i \to j@t$, we denote the ratio of the used transmission time to time slot's duration by $\eta_{i,j,t} \in [0,1]$. $\eta_{i,j,t} = 0$ means no subcarrier is scheduled for the transmission $i \to j@t$, whereas $\eta_{i,j,t} \in (0,1]$ means a subcarrier is scheduled at time slot t and the transmission uses $\eta_{i,j,t}$ of the time slot's duration. By $C_{j,t}$ we denote the total data volume user j have by time slot t. Since the system has no D2D data reuse, user j' must have enough $C_{j',t}$ in order to transmit to j at t.

We formulate the optimization problem as

$$\min_{\mathbf{H} \in [0,1]^{(J+1) \times J \times T}} \left\{ \sum_{j=1}^{J} \left(\sum_{t=1}^{T} P_{0,j,t} \Delta t + \sum_{t=1}^{T} \sum_{j'=1}^{J} P_{j',j,t} \Delta t \right) \right\} \tag{2a}$$

$$s.t. \quad \left(\eta_{0,j,t} > 0\right) + \sum_{j' \neq j} \left(\eta_{j',j,t} > 0\right) + \sum_{j'' \neq j} \left(\eta_{j,j'',t} > 0\right) \leq 1 \tag{2b}$$

$$\sum_j \left(\sum_i \left(\eta_{i,j,t} > 0 \right) \right) \leq N \tag{2c}$$

$$r_{i,j,t}^{\max} = \log_2 \left(1 + \frac{P_i^{\max} \beta_{i,j,t} \left| h_0 \right|^2}{\sigma^2 + M} \right) \tag{2d}$$

$$r_{i,j,t} = r_{i,j,t}^{\max} \eta_{i,j,t} \quad , \quad P_{i,j,t} = P_i^{\max} \eta_{i,j,t} \tag{2e}$$

$$C_{j,t}\Big|_{t=0} = 0, C_{j,t}\Big|_{t=T} \geq C_j^{QoS} \tag{2f}$$

$$C_{j,t} = \sum_{\tau=1}^{t} \left(r_{0,j,\tau} \Delta t + \sum_{j'} r_{j',j,\tau} \Delta t - \sum_{j"} r_{j,j",\tau} \Delta t \right), C_{j,t} \geq 0 \tag{2g}$$

where $P_i^{\max} = \left\{ P_0^{\max}, \left\{ P_j^{\max} \right\} \right\}$ represents the maximum transmission power of BS and users, and $P_{i,j,t}$ $\mathbf{H} = \left\{ \eta_{i,j,t} \right\}^{(J+1)\times J\times T}$ since we have to consider transmissions from $(J+1)$ sources to J targets at T time slots $(J+1)\times J\times T$ dimension. Constraint in (2b) guarantees that users can only receive from one source since they have single antenna. Constraint in (2c) guarantees that at most N users can be severed simultaneously in the system, cellular or D2D, since there is only N subcarriers. In (2e) we create two denotations $r_{i,j,t}$ and $P_{i,j,t}$ for simplicity in (2a) and (2g). (2f) and (2g) make sure that the QoS constraint is met and relays can only transmit all they have at most.

B. Algorithm

The problem in (1) is a discrete non-convex optimization problem and is NP-hard. Therefore, conventional methods for solving linear or convex optimization problems are no longer applicable. We decompose the problem into three simpler sub-problems. First, we ignore the subcarrier constraint and only consider the cellular transmission. Second, we use an iterative algorithm to make sure the cellular transmission uses no more than N subcarriers and get a suboptimal solution for the cellular-only system. Last, we use another iterative algorithm to substitute part of the cellular transmission links for D2D transmission link sets for less energy consumption. Each of the substitution D2D link sets consists of exact one cellular link $0 \rightarrow j'@t_1$ and one D2D relay transmission link $j' \rightarrow j@t_2$. Two links (one cellular and one D2D) in the substitution set must use less energy combined than the original cellular-only link for improvement energy-wise. This results in that the transmission time of the two links in the substitution set must be less than time slot duration Δt.

For the first two sub-problems, we consider a cellular-only system. We fix $i=0$ since users can only receive data from BS.

$$\min_{\mathbf{H_0}\in[0,1]^{J\times T}}\left\{\sum_{j=1}^{J}\left(\sum_{t=1}^{T}P_{0,j,t}\Delta t\right)\right\} \tag{3a}$$

s.t.

$$\sum_{j}\left(\eta_{0,j,t}>0\right)\le N \tag{3b}$$

$$r_{0,j,t}^{\max}=\log_2\left(1+\frac{P_{0,j,t}\beta_{0,j,t}\left|h_0\right|^2}{\sigma^2+M}\right) \tag{3c}$$

$$r_{0,j,t}=r_{0,j,t}^{\max}\eta_{0,j,t}\ ,\ P_{0,j,t}=P_0^{\max}\eta_{0,j,t} \tag{3d}$$

$$C_{j,t}\Big|_{t=0}=0, C_{j,t}\Big|_{t=T}\ge C_j^{QoS} \tag{3e}$$

$$C_{j,t}=\sum_{\tau=1}^{t}\left(r_{0,j,\tau}\Delta t\right), C_{j,t}\ge 0 \tag{3f}$$

$P_0^{\max}$ represents the maximum transmission power of BS. Here $\mathbf{H_0}=\left\{\eta_{0,j,t}\right\}^{J\times T}$ since the optimization is currently in a $J\times T$ subspace (only one source BS) in the first two cellular-only sub problems. Constraint in (1b) is not necessary here since users can only receive data from one source, namely BS.

a. Sub-problem 1 - optimal user scheduling for cellular system regardless of subcarrier count

For the first sub-problem, we optimize P with constraint (3c)-(3f), ignoring the subcarrier constraint in (3b). This means we assume that the BS is omnipotent: that is, BS can serve infinite number of users. In this case, the optimization variables of different users in no longer correlated, and the optimal solution of this problem can be obtained by scheduling each user separately. The problem can be reduced to $\min_{\mathrm{H}\in[0,1]^T}\left\{\sum_{t=1}^{T}P_{0,j,t}\Delta t\right\}$. Note that $r_{0,j,t}$ is a monotone increasing function of $\beta_{0,j,t}$, therefore we can obtain the optimal solution for each user by assigning time slots with best CSI.

Define $\mathbf{S}$ the set of chosen transmission link at a specific time slot, i.e., $(i,j,t)\in\mathbf{S}$ if $i\to j@t$ is a chosen transmission link at time slot t. We propose Algorithm 1 to solve the first sub-problem.

Algorithm 1 optimal user scheduling for cellular system regardless of subcarrier count

for each user j :

while $C_{j,T} \geq C_{j,QoS}$ not met :

find $(0, j, t) = \arg\max\left\{r_{0,j,t}^{\max}\right\}$

set $\eta_{0,j,t} = 1$

update $\mathbf{S_1} = \mathbf{S_1} \cup \{(0, j, t)\}$, update $C_{j,t}$, update $P_{0,j,t}$.

b. Sub-problem 2 - suboptimal user scheduling for cellular system

The solution $\mathbf{S_1}$ returned by Algorithm 1 is not a feasible one for the cellular-only system in (3a)-(3f) since (3b) has not been taken into account. We design an effective method to approach the suboptimal feasible solution iteratively.

As $\mathbf{S_1}$ is the optimal solution for (3c)-(3f), the original problem in (3a)-(3f) is equivalent to minimizing the energy consumption gap between $\mathbf{S_1}$ and the result $\mathbf{S_2}$ in sub-problem 2, and the second sub-problem can be expressed as

$$\min_{\mathbf{H_0}\in[0,1]^{J\times T}} \left\{ \sum_{t=1}^{T}\sum_{j=1}^{J}\left(\underset{(0,j,t')\in\mathbf{S_2}}{P_0^{\max}\Delta t_{0,j,t'}} - \underset{(0,j,t)\in\mathbf{S_1}}{P_0^{\max}\Delta t_{0,j,t}} \right) \right\} \tag{4a}$$

s.t.

$$\sum_j \left(\eta_{0,j,t} > 0\right) \leq N \tag{4b}$$

$$r_{0,j,t}^{\max} = \log_2\left(1 + \frac{P_{0,j,t}\beta_{0,j,t}\left|h_0\right|^2}{\sigma^2 + M}\right) \tag{4c}$$

$$r_{0,j,t} = r_{0,j,t}^{\max}\eta_{0,j,t} \quad , \quad P_{0,j,t} = P_0^{\max}\eta_{0,j,t} \tag{3d}$$

$$C_{j,t}\Big|_{t=0} = 0, C_{j,t}\Big|_{t=T} \geq C_j^{QoS} \tag{4e}$$

$$C_{j,t} = \sum_{\tau=1}^{t}\left(r_{0,j,\tau}\Delta t\right), C_{j,t} \geq 0 \;\; \Delta P_{0,j,t} = \underset{(0,j,t)\in\mathbf{S_2}}{P_{0,j,t}} - P_{0,j,t} \tag{4f}$$

Note that solving this sub-problem is a process of adjusting the user scheduling result in $\mathbf{S_1}$. We propose an iterative method as shown in Algorithm 2.

Algorithm 2 suboptimal user scheduling for cellular system

initialize $\mathbf{S_2} = \mathbf{S_1}$

while $\forall t, \sum_{j} \left(\eta_{0,j,t} > 0 \right) \leq N$ not met :

find $\left(0, j^*, t^*\right) = \arg \min_{\substack{(0,j,t)\in \mathbf{S_1} \\ (0,j',t')\notin \mathbf{S_2}}} \left\{ r_{0,j,t} - r_{0,j',t'} \right\}$, where $\sum_{j'} \eta_{0,j',t'} \leq N-1,\ \sum_{j} \delta_{0,j,t} > N$

set $\mathbf{S_2} = \mathbf{S_2} \setminus \left\{ \left(i, j^*, t^*\right) \right\},\ \eta_{i,j^*,t^*} = 0$

while $C_{j,T} \geq C_{j,QoS}$ not met :

find $\left(0, j, t\right) = \arg \max_{(0,j,t)\notin \mathbf{S_2}} \left\{ r_{0,j,t} \right\}$, where $\sum_{j} \delta_{0,j,t} \leq N-1$

set $\eta_{0,j,t} = 1$

update $\mathbf{S_2} = \mathbf{S_2} \cup \left\{ \left(0, j, t\right) \right\}$, update $C_{j,t}$, update $P_{0,j,t}$.

c. Sub-problem 3 - suboptimal user scheduling for D2D underlying cellular system

After the first two problems, we have already claimed an approximation of the optimal solution for the cellular-only system in a $J \times T$ subspace. In sub-problem 3, we change part of the cellular transmission links into D2D transmission links for better energy efficiency, and optimize in a $\left(J+1\right) \times J \times T$ subspace. Given that $\mathbf{S_2}$ is only based on constraint (3a)-(3f), the original problem in (1a)-(1g) is equivalent to maximizing the energy consumption reduction between $\mathbf{S_3}$ and the result $\mathbf{S_2}$ in sub-problem 2, and the third sub-problem can be expressed as

$$E_{0,i',t_1}^{\text{temp}} = P_0^{\max} \left(\frac{r_0^{\min}}{r_{0,i',t_1}} \Delta\tau \right)$$

$$E_{i',j,t_2}^{\text{temp}} = P_{i'}^{\max} \left(\frac{r_0^{\min}}{r_{i',j,t_2}} \Delta\tau \right)$$

$$\max_{\mathbf{H}\in[0,1]^{(J+1)\times J\times T}} \left\{ \sum_{t=1}^{T} \sum_{j=1}^{J} \left(\underset{(0,j,t)\in \mathbf{S_2}}{P_0^{\max} \Delta t_{0,j,t}} - \sum_{i=0}^{J} \underset{(i,j,t)\in \mathbf{S_3}}{P_i^{\max} \Delta t_{i,j,t}} \right) \right\} \tag{5a}$$

s.t.
$$\left(\eta_{0,j,t} > 0 \right) + \sum_{j' \neq j} \left(\eta_{j',j,t} > 0 \right) + \sum_{j'' \neq j} \left(\eta_{j,j'',t} > 0 \right) \leq 1 \tag{5b}$$

$$\sum_{j} \left(\sum_{i} \left(\eta_{i,j,t} > 0 \right) \right) \leq N \tag{5c}$$

$$r_{i,j,t}^{\max} = \log_2 \left(1 + \frac{P_{i,j,t} \beta_{i,j,t} \left| h_0 \right|^2}{\sigma^2 + M} \right) \tag{5d}$$

$$r_{i,j,t} = r_{i,j,t}^{\max} \eta_{i,j,t} \quad , \quad P_{i,j,t} = P_i^{\max} \eta_{i,j,t} \tag{5e}$$

$$C_{j,t}\big|_{t=0}=0, C_{j,t}\big|_{t=T}\geq C_j^{QoS} \tag{5f}$$

$$C_{j,t}=\sum_{\tau=1}^{t}\left(r_{0,j,\tau}\Delta t+\sum_{j'} r_{j',j,\tau}\Delta t-\sum_{j"} r_{j,j",\tau}\Delta t\right), C_{j,t}\geq 0 \tag{5g}$$

Here $\mathbf{H}=\{\eta_{i,j,t}\}^{(J+1)\times J\times T}$ since the optimization is now in a $(J+1)\times J\times T$ subspace since there are $(J+1)$ sources (BS/users), J targets and T time slots. For simplicity, by $0\rightarrow j@t_0$ we denote the cellular transmission link in $\mathbf{S_2}$ that is to be replaced by a set of D2D links in sub-problem 3. Whereas the substitution set of D2D links in this paper consist of exact one cellular link $0\rightarrow j'@t_1$ and one D2D relay transmission link $j'\rightarrow j@t_2$. Two links (one cellular and one D2D) in the substitution set must use less energy combined than the original cellular one. Accordingly, the transmission time of the two links in the substitution set must be less than time slot duration Δt, and therefore $\eta_{0,j',t_1}\in(0,1)$ and $\eta_{j',j,t_2}\in(0,1)$. We propose another iterative method as shown in Algorithm 3.

Algorithm 3 suboptimal user scheduling for D2D underlying cellular system

initialize $\mathbf{S_3}=\mathbf{S_2}$

for all user j :

set $\mathbf{R}=\phi$ as temporary set for all possible d2d link sets to user j , $r_0^{\min}=\min_{(0,j,t)\in\mathbf{S_2}}\{r_{0,j,t}^{\max}\eta_{0,j,t}\}$

for all time slot t_2 :

for all relay $j'\neq j$:

if j & j'& the system free $@t_2$ and $r_{i',j,t_2}^{\max}\geq r_0^{\min}$:

for all time slot t_1 where j'& the system free $@t_1$ and $t_1\neq t_2$ and $r_{0,j',t_1}^{\max}\geq r_0^{\min}$:

$$\textbf{if } \begin{cases} C_{j',t_2-1} \geq r_0^{\min}\Delta t, & \text{if } t_1 > t_2 \\ C_{j',t_2-1} + r_{0,j',t_1}^{\max}\Delta t \geq r_0^{\min}\Delta t, & \text{else} \end{cases} :$$

$$\text{set} \quad \mathbf{R} = \mathbf{R} \cup \left\{\left[\left(0, j', t_1\right), \left(j', j, t_2\right)\right]\right\}$$

while $\mathbf{R} \neq \phi$:

update $r_0^{\min} = \min_{(0,j,t)\in\mathbf{S}_2} \left\{r_{0,j,t}^{\max}\eta_{0,j,t}\right\}$

if $\eta_{0,j',t_1} + \frac{r_0^{\min}}{r_{0,j',t_1}^{\max}} \leq 1$ and $\eta_{j',j,t_2} + \frac{r_0^{\min}}{r_{j',j,t_2}^{\max}} \leq 1$ where $\left[\left(0, j', t_1\right), \left(j', j, t_2\right)\right] \in R$

and $j \,\&\, j'\&$ the system free $@t_2$ and $j'\&$ the system free $@t_1$:

$$\Delta P = P_{0,j,t_0} - \left(P_{0,j',t_1} + P_{j',j,t_2}\right)$$

else:

$$\Delta P = -\infty$$

if $\max\left\{\Delta P\right\} > 0$:

$$\left(\left(i, j, t_0\right), \left(i, j', t_1\right), \left(j', j, t_2\right)\right) = \arg\max\left\{\Delta P\right\}$$

$$\text{set} \quad \eta_{0,j,t_0} = 0, \eta_{0,j',t_1} += \frac{r_0^{\min}}{r_{0,j',t_1}^{\max}}, \eta_{j',j,t_2} += \frac{r_0^{\min}}{r_{j',j,t_2}^{\max}}$$

update $S_3 \leftarrow S_3 \setminus \left\{\left(i, j, t_0\right)\right\}, S_3 \leftarrow S_3 \cup \left\{\left(\left(i, j', t_1\right), \left(j', j, t_2\right)\right)\right\}$, update $C_{j,t}, C_{j',t}$, update $P_{0,j,t_0}, P_{0,j',t_1}, P_{j',j,t_2}$

$$\text{set} \quad R \leftarrow R \setminus \left\{\left(\left(i, j', t_1\right), \left(j', j, t_2\right)\right)\right\}$$

else :

break

Here $j \,\&\, j'\&$ the system free $@t_2$ means that

$$\sum_{i\neq j}\left(\eta_{i,j,t_2} > 0\right) + \sum_{j^*\neq j}\left(\eta_{j,j^*,t_2} > 0\right) \leq 1$$

and

$$\sum_{i\neq j'}\left(\eta_{i,j',t_2} > 0\right) + \sum_{j^{**}\neq j'}\left(\eta_{j',j^{**},t_2} > 0\right) \leq 1$$

and $$\sum_j \sum_i \left(\eta_{i,j,t_2} > 0\right) \le N .$$

And j'& the system free $@t_1$ means that

$$\left(\eta_{0,j',t_1} > 0\right) + \sum_{j^* \ne j'} \left(\eta_{j^*,j',t_1} > 0\right) + \sum_{j^{**} \ne j'} \left(\eta_{j',j^{**},t_1} > 0\right) \le 1$$

and $$\sum_{j^*} \left(\sum_{i^*} \left(\eta_{i^*,j^*,t_1} > 0\right) \right) \le N$$

Therefore the system constraint in (2b) and (2c) is met.