Appendix A
Proof of Theorem 2

According to the proposed communication protocol, the BS transmission is given higher priority than source transmission due to the real-time of patients' data. Therefore, source can be selected to transmit the data only when all $m \to d$ links are outage, and there would be four cases.

Case 1: The selected BS is m^* and the packet number in its cache is 0, i.e., $\mathcal{D} = m^* \wedge l^* = 0$. The elements of $\mathbb{G}_{l^*}^{PNI}$ is divided into two categories, one belongs to the subset $\mathcal{G}_{l^*}^{sm^*}$ (i.e., $sm \in \mathcal{G}_{l^*}^{sm^*}$) whose channel quality satisfies the condition (3), the other one belongs to the subset $\mathcal{G}_{l^*}^{sm^*}$ but not to the set $\mathbb{G}_{l^*}^{PNI}$ (i.e., $s\hat{m} \notin \mathcal{G}_{l^*}^{sm^*}$ and $s\hat{m} \in \mathbb{G}_{l^*}^{PNI}$), and theirs whose channel quality dose not satisfy the condition (3). Note that for different sets of $\mathcal{G}_{l^*}^{sm^*}$, the probabilities that $s \to m^*$ link is selected for transmission leading to the state transition from S_i to S_j have different values. Therefore, we need to find out all possible subsets of $\mathbb{G}_{l^*}^{PNI}$ that each contains the link $s \to m^*$. Overall, when $\mathcal{D} = m^* \wedge l^* = 0$, the probability that the source transmission leads to the connected state S_j can be formulated as

$$\begin{aligned}
A_{i,j}^{i.n.d} &= \sum_{\mathcal{G}_{l^*}^{sm^*}} \frac{1}{|\mathcal{G}_{l^*}^{sm^*}|} \mathbb{P}\Big(g_{sm} \geq \theta_s, \forall sm \in \mathcal{G}_{l^*}^{sm}\Big) \mathbb{P}\Big(g_{s\hat{m}} < \theta_s, \forall s\hat{m} \notin \mathcal{G}_{l^*}^{sm} s\hat{m} \in \mathbb{G}_{l^*}^{PNI}\Big) \prod_{d=1}^{\tau-1} \Theta_{id}^{DSI} \\
&= \sum_{\mathcal{G}_{l^*}^{sm^*}} \frac{1}{|\mathcal{G}_{l^*}^{sm^*}|} \underbrace{\prod_{sm \in \mathcal{G}_{l^*}^{sm^*}} \mathbb{P}(g_{sm} \geq \theta_s) \prod_{\substack{s\hat{m} \notin \mathcal{G}_{l^*}^{sm^*} \\ s\hat{m} \in \mathbb{G}_{l^*}^{PNI}}} \mathbb{P}(g_{s\hat{m}} \geq \theta_s)}_{ⓐ} \prod_{d=1}^{\tau-1} \Theta_{id}^{DSI}
\end{aligned} \tag{1}$$

where the term $1/|\mathcal{G}_{l^*}^{sm^*}|$ is due to the fact that we select one of them uniformly when $\mathcal{G}_{l^*}^{sm^*}$ consists of multiple elements, the term ⓐ is $\mathcal{P}_1(\mathcal{G}_{l^*}^{sm^*})$ and can be derived as (31) by some numerical calculations and simplification, $\prod_{d=1}^{\tau-1} \Theta_{id}^{DSI}$ is the probability that all $m \to d$ links are outage, which is the same as that in Lemma 1.

Case 2: The selected BS is m^* and the packet number in its cache is $0 < l^* < \tau - 1$, i.e., $\mathcal{D} = m^* \wedge 0 < l^* < \tau - 1$. Based on Algorithm 1, the fewer packets the BS owns, the higher priority it is selected as the receiver. Thus, only when the quality of these $s \to m$ links does not satisfy the condition (3) where the involved BSs own fewer packets than BS m^*, i.e., $\psi(U_m^i) \in [0, l^* - 1]$. Thus, when $\mathcal{D} = m^* \wedge 0 < l^* < \tau - 1$, the probability that the source transmission leads to the connected state S_j can be formulated as

$$
\begin{aligned}
A_{i,j}^{i.n.d} &= \sum_{\mathcal{G}_{l^*}^{sm^*}} \frac{1}{|\mathcal{G}_{l^*}^{sm^*}|} \prod_{sm \in \mathcal{G}_{l^*}^{sm^*}} \mathbb{P}(g_{sm} \geq \theta_s) \prod_{\substack{s\hat{m} \notin \mathcal{G}_{l^*}^{sm^*} \\ s\hat{m} \in \mathbb{G}_{l^*}^{PNI}}} \mathbb{P}(g_{s\hat{m}} \geq \theta_s) \\
&\quad \times \mathbb{P}\Big(g_{sm} < \theta_s, \forall sm \in \mathbb{G}_l^{PNI} \wedge 0 \leq l \leq l^* - 1\Big) \prod_{d=1}^{\tau-1} \Theta_{id}^{DSI} \\
. &= \sum_{\mathcal{G}_{l^*}^{sm^*}} \frac{1}{|\mathcal{G}_{l^*}^{sm^*}|} \underbrace{\prod_{sm \in \mathcal{G}_{l^*}^{sm^*}} \mathbb{P}(g_{sm} \geq \theta_s) \prod_{\substack{s\hat{m} \notin \mathcal{G}_{l^*}^{sm^*} \\ s\hat{m} \in \mathbb{G}_{l^*}^{PNI}}} \mathbb{P}(g_{s\hat{m}} \geq \theta_s) \prod_{l=0}^{l^*-1} \mathbb{P}(g_{sm} < \theta_s)}_{ⓑ} \prod_{d=1}^{\tau-1} \Theta_{id}^{DSI}.
\end{aligned} \tag{2}
$$

where the term ⓑ is $\mathcal{P}_2(\mathcal{G}_{l^*}^{sm^*})$ and we can obtain (32) by do some numerical calculations.

Case 3: The selected partner pair is (m^*, k) and the packet number in its cache is $l^* = 0$, i.e., $\mathcal{D} = (m^*, k) \wedge l^* = 0$. Notice that we would select one idle BS as the cooperative jammer only when the quality of all $s \to m$ links does not satisfy the condition (3). So, when the BSs m^* and k are selected as the partner pair, there have three-level meanings: 1) the channel gains of these $s \to m$ links where the involved BSs have the non-empty caches, i.e., $\psi(U_m^i) \geq 1$, must satisfy the condition $\theta_s > g_{sm}$; 2) the channel gains of $s \to m^*$ and $m^* \to k$ links must satisfy the condition $\theta_s > g^{sm^*} \geq g_{km^*} \Xi_{sk}^{m^*}$ and $g_{km^*}\beta_{m^*} \leq \min_{m \neq m^*, k}\{g_{mm^*}\beta_{m^*}\}$; 3) for other BSs which own the empty cache, i.e., $\bar{m} \in \mathbb{G}_{l^*}^{PNI} \wedge \bar{m} \neq m^*$, we have $\theta_s > g_{s\bar{m}}$ and $g_{km^*}\beta_{m^*} < \min_{\hat{m} \neq \bar{m}}\{g_{\hat{m}\bar{m}}\beta_{\bar{m}}\}$. Thus, when $\mathcal{D} = (m^*, k) \wedge l^* = 0$, the probability that the source transmission leads to the connected state S_j can be formulated as

$$
\begin{aligned}
A_{i,j}^{i.n.d} &= \prod_{l=1}^{\tau-1} \mathbb{P}(g_s^m < \theta_s) \prod_{d=1}^{\tau-1} \Theta_{id}^{DSI} \; \mathbb{P}\Big(g_s^m < \theta_s, \bar{m} \in \mathbb{G}_{l^*}^{PNI} \wedge m \neq m^*\Big) \\
&\quad \underbrace{\mathbb{P}\left(\Big(\theta_s > g_{sm^*} \geq g_{km^*} \Xi_{sk}^{m^*}\Big) \wedge g_k^{m^*} \beta_{m^*} \leq \min\left\{ \min_{\substack{\bar{m} \in \mathbb{G}_{l^*}^{PNI}/m^* \\ \hat{m} \neq \bar{m}}} \{g_{\hat{m}\bar{m}}\beta_{\bar{m}}\}, \min_{m \neq m^*, k} \{g_{mm^*}\beta_{m^*}\} \right\} \right)}_{ⓒ}.
\end{aligned} \tag{3}
$$

where the term ⓒ is $\mathcal{P}_{km^*}^1$ and can be derived as (34) by some numerical calculations and simplification.

Case 4: The selected partner pair is (m^*, k) and the packet number in its cache is $0 < l^* < \tau - 1$, i.e., $\mathcal{D} = (m^*, k) \wedge 0 < l^* < \tau - 1$. The difference between this case and Case 3 is that all

$s \rightarrow m$ links are outage where the number of the packets in the cache of the involved BS m is less than l^*. Similar to the Case 3, the probability that the source transmission leads to the connected state S_j can be formulated as

$$A_{i,j}^{i.n.d} = \prod_{l=1}^{\tau-1} \mathbb{P}(g_{sm} < \theta_s) \prod_{d=1}^{\tau-1} \Theta_{id}^{DSI} \, \mathbb{P}\Big(g_{sm} < \theta_s, \bar{m} \in \mathbb{G}_{l^*}^{PNI} \wedge m \neq m^*\Big) \prod_{l=0}^{l*-1} \Theta_{il}^{PNI} \cdot \underbrace{\mathbb{P}\left(\left(\theta_s > g_{sm^*} \geq g_{km^*} \Xi_{sk}^{m^*}\right) \wedge g_k^{m^*} \beta_{m^*} \leq \min\left\{ \min_{\substack{\bar{m} \in \mathbb{G}_{l^*}^{PNI}/m^* \\ \hat{m} \neq \bar{m}}} \{g_{\hat{m}\bar{m}} \beta_{\bar{m}}\}, \min_{m \neq m^*, k} \{g_{mm^*} \beta_{m^*}\} \right\}\right)}_{d}. \quad (4)$$

where the expression of the term d is same as the term 3 in Case 3.

Overall, the state transition caused by the source transmission from state S_i to state S_j under the proposed communication protocol only have the above four cases. Therefore, according to the formulas (1)-(4), we can derive $A_{i.j}$ as the result (30).

Remark 1: Similarly, the readers can prove other theorems and corollaries based on the above proof process. Importantly, different from the cache states, since each packet has a unique arrival time, the BS can be uniquely represented by the delay information of the oldest packet in its cache queue.