Abstract: Integrated Cellular Networks (ICNs) are normally constructed by adding ad hoc overlay on cellular networks to solve the latter’s flexibility and capacity expansion problems. In such networks, routing plays a critical role in finding a route to divert congested traffic from a congested cell to another less crowded cell. While much work has been conducted on routing protocols in ICNs, no dedicated work has been found for an important aspect of routing, namely source selection. The process of a source selection can be an algorithm which is designed for selecting a proper pseudo source to release its occupying channel to a blocked mobile user. Consequently, this pseudo source diverts its ongoing call to another cell by using a free channel in a neighbor cell via a relaying route. Based on an introduction of a representative ICN infrastructure, this paper proposes three source selection algorithms. Both numerical analysis and evaluation results are presented, which show the efficiency of the algorithms and their different ability in adapting to different network situations, such as traffic density and cell capacity.


1. INTRODUCTION

At present, cellular systems are widely deployed all around the world. However, as more and more mobile users join the pre-built and fixed cellular networks, current cellular system becomes very vulnerable to heavy and unbalanced traffic. Therefore, we need to modify current cellular systems by adding more flexibilities, such as dynamically adapting to instant network situations and balancing traffic between different cells [1]. As a result, some cells could be congested because of temporary events, such as traffic accidents and sports meetings. The demands in current cellular networks (CNs) stimulate people to integrate cellular networks with other networks to enhance the overall performance of cellular systems, such as iCAR [2] (Integrated Cellular and Ad Hoc Relaying System), MADF [3] (Mobile-Assisted Data Forwarding), and CACN (Converged Ad hoc and Cellular Network) [4], amongst many others. Meanwhile, the increasing popularity of multiple communication interfaces (e.g. cellular and IEEE 802.11) implemented in mobile devices eases the way of constructing integrated network architectures. A typical method, as it is for iCAR, MADF and CACN, is to use ad hoc relaying routes to divert blocked calls from congested cells to non-congested cells so as to solve hot cell problem. As such, routing protocols, which are to set up relaying routes from congested cells to non-congested or cold cells while accommodating and utilizing diverse physical interfaces and network infrastructures [5], play a critical role in ICN systems.

Routing in ICNs has its unique challenges. In mobile ad hoc networks, destinations are known to sources when route discovery is initiated. However, in ICNs such as the one illustrated in Figure 1, a call or data can be finally diverted to any suitable destination if the latter has free bandwidth to accommodate the request. As such a destination selection procedure is a necessity in ICNs. The feature of multiple destinations also indicates the existence of multiple relaying routes between a source and destination(s). Therefore, a route selection procedure is usually needed in ICN routing. If a source MH is unable to find a diversion station of sufficient bandwidth and within its transmission range, a source selection procedure is needed to select another source - we call this source as pseudo-source because it is not the source that initiates the calls. This pseudo source MH should have a TDS within its one-hop range and be able to find a relaying route to divert its own traffic. Then the pseudo source releases its occupied channel to the original source MH. This paper focuses particularly on the source selection procedure, which, to the best of our knowledge, has not been dedicated to by any research work. Here in this paper source selection means pseudo source selection, which is common for almost all ICN routing algorithms.

Obviously, a random selection of source node for a route discovery process is not reasonable in practice. In order to successfully establish relaying routes and increase the throughput of the home cell and the overall ICN system, proper algorithms are needed to find optimal source nodes that could divert calls and maximizing the overall system performance. This paper proposes three algorithms for SSP and investigates their performance under different network circumstances. The ones proposed in this paper have been integrated into the CACN systems [4]. However, its generality enables it to be applicable to other ICN systems such as iCAR, MADF and UCAN [6].

In our previous work, we proposed an extension of the existing integrated cellular and ad hoc network architecture iCAR by adding flexible access to traffic diversion stations. The corresponding routing algorithm was also designed, and numerical analysis and evaluation in terms of only overhead were also presented in [7]. Then in [4], we designed a new network architecture for the convergence of cellular network and ad hoc networks and proposed an adaptive routing protocol. In this paper we embark on a specific issue involved in routing of the above heterogeneous networks, namely, source selection procedure (SSP).

The rest of this paper is organized as follows. Section II presents a representative physical infrastructure of an ICN net-
work that accommodates features from both iCAR and MADF. Based on the characteristics of this ICN, Section III proposes and details three SSP’s. Section IV presents numerical analysis of the three SSP’s, mainly on the request rejection rate (RRR) of the ICN. According to the evaluation results given in Section V, the paper concludes in Section VI.

II. A REPRESENTATIVE ICN INFRASTRUCTURE AND ITS ROUTING FEATURES

The success of a source selection algorithm is largely dependent on the physical structure of the ICN system concerned. One of the important design purposes of ICN’s such as iCAR, MADF and CACN is to realize load balancing between different cells so as to be adaptive to the fluctuation of traffic in each cell. In iCAR, a new type of hardware device called ARS (Ad-hoc Relaying Station) is introduced. ARS’s are deployed in managed locations [8] so that MH’s in a hot cell can utilize the bandwidth from a cold cell by accessing to ARS’s and through relaying routes, which are constructed only by ARS’s. Different from iCAR, MADF utilizes ad hoc interfaces implemented in MH’s to construct relaying routes without adding traffic diversion devices. Therefore, relaying routes in ICNs could be composed by both ARS’s and MH’s, or either of them. Then, we can construct an integrated network, as shown in Figure 1, with the characteristics of both iCAR and MADF. In CACN, as depicted in Figure 1, two air interfaces are utilized for the communication between nodes: C (Cellular) interface that operates at a cellular network frequency (in-band, e.g. 2G), and A (Ad-hoc) interface that operates at an ad-hoc network frequency (out-of-band, e.g. IEEE 802.11). Similar to the ARS’s in iCAR, a traffic diversion device called TDS (Traffic Diversion Station) is introduced. However, TDS’s are designed with more flexibility in that MH’s with or without A-interface are both allowed to access to TDS’s. The communication between TDS’s and MH’s through C-interface is at the cost of in-band frequencies, similar as the in-band "forwarding channel" in MADF. Although overloaded traffic can be diverted from one cell to another cell without the deployment of TDS’s, the total amount of diverted traffic in such a network could be limited compared with those with the support of TDS’s. For a more efficient use of out-of-band frequencies, TDS’s should prefer communicating through A-interface rather than through C-interface. Besides, relaying routes in the CACN systems are composed of both TDS’s and MH’s with A-interface, not just TDS’s. By this means, the number of TDS’s added in the network can be reduced, and relaying routes could be constructed more easily.

In CACN, if a MH in a hot cell is within the transmission range of a TDS, it can directly access to the TDS and makes a call by utilizing the bandwidth from a neighbour cell through a relaying route (like MH4 using MH4-TDS1-TDS2-BS2 in Figure 1). Due to the limited number and transmission range of TDS’s, MH’s within the area uncovered by any TDS’s are not able to directly divert their calling traffic through relaying routes (like MH1 in Figure 1). In this situation, the home BS (BS1) chooses a pseudo source (like MH2 in Figure 1) to release its occupied bandwidth for the use of the original source (MH1) just after the secondary source starts diverting traffic through a discovered relaying route. However, if most blocked calls are covered by TDS1, the bandwidth of TDS1 could be used out prior to other TDS’s because of a large number of MH’s accessing. Partially blocking TDS’s could dramatically deteriorate the performance of the network. In addition, a random selection of pseudo source does not consider the successful rate of relaying route discovery. This may lead to more calls being rejected. Hence, proper pseudo source selection algorithms are necessary for the selection of source nodes.

Whenever a MH in a congested cell sends a call and a transmission request to the home BS, the home BS choose a source node to divert its call according to a source selection algorithm. Then, this source node could be the original MH sending the call request, or another MH making an ongoing call. We call the source node sending the call request as "Original Source", and the source node releasing its occupied bandwidth to the original node as "Pseudo Source". The following sections describe the operation of three SSP’s that select proper pseudo sources according to different selection criteria and network conditions.

The operation of the proposed SSP’s in this paper are highly coupled with the physical architecture of the heterogeneous wireless network shown in Figure 1 and it takes advantage of these TDS’s. Therefore the presence of TDS’s is a necessity for these SSP’s. However, in terms of using ad hoc networks to divert cellular network traffic is concerned, there are two schools of thoughts as far as the presence of TDS’s or similar equipments is concerned. One school is represented by iCAR which advocates the introduction of a new type of devices that are dedicated for traffic diversion whereas the other, e.g., UCAN, adopts a purely ad hoc manner without the existence of any special device for traffic diversion. The ICN proposed in this paper falls into the category of the former.

III. PROPOSED ALGORITHMS FOR SOURCE SELECTION

This section discusses the process of reallocating the bandwidth released by the pseudo source and the operation of the three SSP’s proposed.

A. Bandwidth Reallocation

According to Section 2, a pseudo source is chosen by the home BS to release its occupied bandwidth without interrupting the present communication of the pseudo source. Figure 2
shows the main steps of bandwidth reallocation in a certain cell.

If a MH in congested cell is trying to make a call, it sends a Source Node Request Packet (SNREQ) to the home BS (see step 1 in Figure 2). After receiving a SNREQ, the home BS broadcasts a MH List Request Packet (MLREQ) to all TDS’s within the home cell (see step 2 in Figure 2). Once a TDS receives a MLREQ, it broadcasts a Neighbor Discovery Request Packet (NDREQ) (see step 3 in Figure 2). After receiving Route Request Packets (RREQ) (see step 8 in Figure 2), the TDS return a list of MH’s to the home BS by sending a MH List Reply Packet (MLREP) (see step 4 in Figure 2). Then, the TDS return a list of MH’s to all MH’s within its coverage (see step 3 in Figure 2), which respond Neighbor Reply Packets (NREP) to the TDS (see step 4 in Figure 2). Then, the TDS return a list of MH’s to the home BS by sending a MH List Reply Packet (MLREP) (see step 5 in Figure 2), which also contains the bandwidth status of the TDS. After receiving the MLREP’s from all the TDS’s upper bounded to a timeout, the home BS applies a rational SSP to analyze the information included in MLREP’s so as to choose a proper source node (see step 6 in Figure 2). Following this, the home BS sends a Source Node Reply Packet (SNREP) to the decided pseudo source if the source node is not the original source (see step 7 in Figure 2) so that the pseudo source starts a route discovery process by broadcasting Route Request Packets (RREQ) (see step 8 in Figure 2). After receiving Route Reply Packets (RREP) (see step 9 in Figure 2), the pseudo source releases its occupied bandwidth and starts diverting its calling traffic through a relaying route (see step 10 in Figure 2). Finally, the home BS reallocates the released bandwidth to the original source (see step 11 in Figure 2).

B. Three Algorithms for Source Selection Procedure

As mentioned above, SSP’s are designed to run in BS’s. The objective of SSP is to choose a source node to divert its call. At the same time, SSP should also consider the network situation to improve the overall performance of the system. In other words, source nodes chosen by the home BS should have the most possibility of discovering a relaying route, and a large number of source nodes will not partially block TDS’s. Additionally, the operation of SSP depends on the information included in MLREP packets. The following is an example of MLREP packet mainly containing three fields: <TDS ID, MH ID List, Bandwidth Info>. In the MLREP packet, “TDS ID” refers to the address of a TDS in the home BS. "MH ID List" includes a list of the addresses of MH’s within the transmission range of the TDS. "Bandwidth Info" is the available free bandwidth (or channel) in the TDS.

After receiving MLREP packets, the home BS gets a big list of MH’s. In fact, a source node should satisfy the requirements that a source node is an MH within the transmission range of the home cell (relevant information can be obtained by checking the register information of MH’s), and a source node should be covered by at least one TDS. Hence, the home BS chooses all the MH’s, which fulfill the basic requirements, to build up a table of available source nodes. Then, the home BS chooses a final source node to divert its traffic according to a source selection algorithm. As shown in Table 1, MHi is the address of an available source node. TDSij is the address of a reachable TDS of MHi (assuming the RRR of each TDS is identical at the moment). P

\[ P_{MHi} = \prod_{j=1}^{N_i} P_{ij} \]  

(1)

Thus, the source node chosen by SSP1 is the MH with minimum RRR, namely \( min(P_{1MHi}, ..., P_{NMHi}) \) (assuming the number of available MH’s in Table 1 is \( n \)). To simplify the notation, we assume that the RRR of each TDS is identical at \( P \). Then, the RRR of MHi is:

\[ P_{MHi} = \prod_{j=1}^{N_i} P_{ij} = (P)^{N_i} \]  

(2)

Thus, SSP1 just chooses a source node with the maximum number of reachable TDS’s (namely \( max(N_1, ..., N_n) \)), because \( P \leq 1 \). As shown in Figure 3, MH1 can only broadcast RREQ.

<table>
<thead>
<tr>
<th>MH ID</th>
<th>Home TDS ID</th>
<th>Bandwidth Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>MHi</td>
<td>TDSij</td>
<td>Bij</td>
</tr>
<tr>
<td>MHi</td>
<td>TDSij</td>
<td>Bij</td>
</tr>
<tr>
<td>MHi</td>
<td>TDSij</td>
<td>Bij</td>
</tr>
<tr>
<td>MHi</td>
<td>TDSij</td>
<td>Bij</td>
</tr>
</tbody>
</table>

Table 1. Available Pseudo Sources

Fig. 2. Operation of Bandwidth Reallocation.
to TDS1, but MH2 can broadcast RREQ’s to TDS2, TDS3 and TDS4. MH2 stands a better chance to find a relaying route.

One problem of the selection is that SSP1 could only choose MH’s within a specific cross area with most reachable TDS’s. As shown in Figure 3, SSP1 could only choose source nodes within the cross area covered by TDS2, TDS3 and TDS4, but not choose source nodes covered by TDS1. The result is that the bandwidth of TDS’s covering the specific cross area is firstly consumed. Then, these TDS’s could become congested but other TDS’s still have lots of free bandwidth unused. Therefore, we designed other SSP’s to solve this problem.

SSP2 tries to balance the bandwidth consumption amongst TDS’s when choosing source nodes. Using the information in Table 1, the home BS can get the bandwidth status of all available pseudo sources. \(B_{MHi}\) refers to the free bandwidth of the reachable TDSij of MH. Then, the average bandwidth of all reachable TDS’s of an available source node is:

\[
B_{MHi} = \frac{\sum_{j=1}^{N_i} B_{ij}}{N_i} \tag{3}
\]

To achieve the balance of TDS bandwidth consumption, SSP2 chooses pseudo sources with maximum \(B_{MHi}\), namely \(\max(B_{MHI1}, ..., B_{MHI\text{max}})\). As a result, source nodes will consume the bandwidth of TDS’s in average, and avoid partially congesting TDS’s. However, such a selection of source nodes may lead to a relatively higher RRR, because SSP2 is designed not to find a source node with a minimum RRR but to choose a source node with a maximum average bandwidth of reachable TDS’s. Then, the instantaneous RRR in SSP2 is determined by the RRR of current source nodes with a random number of reachable TDS’s. Thus, compared with SSP1, SSP2 can avoid partially blocking TDS’s by balancing the consumed bandwidth of TDS’s, but the RRR in SSP2 could be higher due to the unpredictable number of reachable TDS’s of source nodes.

SSP3 takes both the number and the average free bandwidth of reachable TDS’s into consideration. Therefore, SSP3 can achieve a balance of the bandwidth consumption of TDS’s without highly increasing the RRR of the system during the selection of source nodes. In SSP3, node MH has two weights for selecting pseudo sources, namely \((P_{MHI1}, B_{MHI1})\) as mentioned in SSP1 and SSP2. The combined weight calculated by SSP3 is

\[
W_i = B_{MHI}(1 - PMHi) (1 - PMHi) \text{ refers to the possibility of successfully discovering a relaying route). Also, the weight (Wi) can be named as the average free bandwidth of reachable TDS’s of MH in probability, compared with \(B_{MHI}\). According to formula (2) and (3) shown above, the combined weight is:

\[
W_i = \frac{\sum_{j=1}^{N_i} B_{ij}}{N_i} [1 - (P)^{N_i}] \tag{4}
\]

SSP3 chooses pseudo sources with the maximum \(W_i\), namely \(\max(W_i, ..., W_n)\). Because \(P\) in the above formula is the average RRR of TDS’s, the calculation of \(P\) depends on the statistical results calculated by BS’s or TDS’s. To reduce the amount of calculation in BS’s or TDS’s, the combined weight of a MH can be simplified as following. For the MH’s given in Table 1, both the number of reachable TDS’s \((N_i)\) and the average free bandwidth of reachable TDS’s \((B_{MHI})\) are sorted in order. Then, instead of \((P_{MHI}, B_{MHI})\), the weight of \(MHi\) in SSP3 is \((N_T, N_B)\) \(N_T\) refers to the order of MH in terms of the number of reachable TDS’s, and \(N_B\) refers to the order of MH in terms of the average free bandwidth of reachable TDS’s. Then, the combined weight of \(MHi\) in SSP3 is simplified as:

\[
W_i = N_T N_B \tag{5}
\]

SSP3 could just choose pseudo sources with the simplified minimum \(W_i\).

SSP-LA (Source Selection Procedure with Location Aid) is devised to select source nodes following SSP1, SSP2 or SSP3. After applying SSP1, SSP2 or SSP3 to the selection of source nodes, a final source node could still not be specified because some MH’s in Table 1 may get the same value of weight during the calculation of SSP. Therefore, after SSP1, SSP2 or SSP3, the mobility and location information of MH’s could be taken into consideration for the further selection of source node. The main idea of SSP-LA is to choose MH’s with the most possibility of moving out of the home cell during call time, because MH’s will automatically release its occupied bandwidth (either the bandwidth from the home BS or the bandwidth from TDS’s) and use the bandwidth from the adjacent BS’s as they move to the neighbor cells. To simplify the analysis, we assume that the shape of the home cell is a circle with a transmission radius at \(R\) (instead of a hexagon with a centre-to-vertex distance at \(R\)).
radius at \( r \), which indicates that the MH moves under a random direction model. As shown in Figure 4, \( L \) indicates the distance between the home BS and a MH.

By analyzing Figure 4, the probability of \( MHi \) moving outside the home cell (\( P_{Mi} \)) equals to the area of the \( MHi \)'s moving field outside the home cell (\( A_{out} \)) divided by the overall area of the \( MHi \)'s moving field (\( A_m \)), namely:

\[
P_{Mi} = \frac{A_{out}}{A_m} \tag{6}
\]

Also, the moving radius of \( MHi \) (\( r \)) is the speed of \( MHi \) (\( Si \)) multiplied by the average call time (\( T_c \)). In fact, \( T_c \) is a statistical value calculated by BS's [9]. For example, \( T_c \) could be the average value of last \( N_{call} \) calls, namely:

\[
T_c = \frac{\sum_{i=1}^{N_{call}} T_i}{N_{call}} \tag{7}
\]

The speed of \( MHi \) could also be calculated as the result of the moving distance of \( MHi \) during a period from \( t_1 \) to \( t_2 \) (\( D \)) divided by the moving period (\( t_2 - t_1 \)) [10], namely:

\[
S_i = \frac{D}{t_2 - t_1} = \frac{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}{t_2 - t_1} \tag{8}
\]

In formula (8), \((x_1, y_1)\) is the location of \( MHi \) at \( t_1 \), and \((x_2, y_2)\) is the location of \( MHi \) at \( t_2 \). The location information of a certain MH can be obtained by many ways, such as [11, 12] and [13]. Hence, \( L \) could also be computed according to the locations of MH's and BS's. According to formula (6) and (7), we have:

\[
A_m = \pi r^2 = \pi S_i^2 T_c^2 \tag{9}
\]

In terms of \( A_{out} \), it can equal to the result of (\( A_m \)) minus the area of the \( MHi \)'s moving field inside the home cell (\( A_{in} \)), namely:

\[
A_{out} = A_m - A_{in} \tag{10}
\]

As shown in Figure 4,

\[
\theta_1 = \arccos\left(\frac{R^2 + L^2 - r^2}{2RL}\right) \tag{11}
\]

\[
\theta_2 = \arccos\left(\frac{r^2 + L^2 - R^2}{2rL}\right) \tag{12}
\]

Then,

\[
A_{in} = \left[\frac{2\theta_1 \pi R^2}{2\pi} - \frac{R \sin \theta_1 R \cos \theta_1}{2}\right] + \left[\frac{2\theta_2 \pi r^2}{2\pi} - \frac{r \sin \theta_2 r \cos \theta_2}{2}\right]
\]

\[
= \theta_1 R^2 + \theta_2 r^2 - \left(2 \sin \theta_1 \cos \theta_1 + r^2 \sin \theta_2 \cos \theta_2\right) \tag{13}
\]

Finally, we have:

\[
P_{Mi} = 1 - \frac{A_{in}}{\pi S_i^2 T_c^2} \tag{14}
\]

According to formula given above, the probability of \( MHi \) moving out (\( P_{Mi} \)) can be calculated. Thus, MH's with maximum \( P_{Mi} \) are chosen as source nodes. If the number of MH's with maximum \( P_{Mi} \) is still more than one, MH's with the furthest distance from the home BS are chosen as pseudo sources. Alternatively, we could also randomly choose one of the MH's left after all selections.

### IV. NUMERICAL ANALYSIS OF THE PROPOSED ALGORITHMS

This section evaluates the performance of proposed SSP’s in terms of the average request rejection rate (RRR) of the whole system and the signaling overhead introduced by SSP’s. To simplify the analysis, we assume that MH’s randomly choose one of their reachable TDS’s as its first hop to divert traffic, and each TDS has the same request rejection rate under the same network environment. Also, whatever temporary events (or overloaded traffic) happen inside or outside the area covered by TDS’s, traffic diversion services are offered to MH’s only after applying SSP’s.

#### A. Request Rejection Rate Analysis

Table 2 shows the parameters used for RRR analysis. According to the parameters given in Table 2, we have:

\[
T_{Ti} = N_a T_T \tag{15}
\]

SSP1 focuses on finding source nodes with minimum number of reachable TDS’s, but may cause that diversion traffic floods some TDS’s and lets the bandwidth of other TDS’s unused. As shown in Figure 5, TDS1, TDS2 and TDS3 have the highest priority of being chosen to divert traffic, because they cover MH’s with 3 reachable TDS’s (such as MH1 in Figure 5). Following them, TDS4, TDS5, TDS6 and TDS7 have the same priority. At last, TDS8 has the lowest priority, because it does not have a cross-covered field. Hence, TDS1, TDS2 and TDS3 consume their traffic firstly, and TDS8 consumes its traffic lastly.

Because the current diversion traffic in the home cell is \( T_c - T_{MAX} \), then:

\[
\frac{T_c - T_{MAX}}{T_T} = \frac{1}{N_a} \sum_{i=1}^{N_a} T_T \]

\[
N_a = \alpha \tag{16}
\]

Considering the average RRR of TDS’s (\( P \)), \( P \) is actually a function of the number of current available TDS’s in a cell (\( N_C \)), namely \( P = f(N_C) \). Then, assuming:
In SSP1, due to the partial use of TDS’s, NC decreases with the rise of diversion traffic. Then, assuming:

$$N_C = \frac{N_T}{\alpha} \left( N_T - \frac{T_C - T_{MAX}}{T_T} \right)$$  \hspace{1cm} (17)

In formula (16) and (17), $\alpha$ and $\beta$ are constants related to the distribution density and bandwidth of TDS’s, and $P_{\text{default}}$ is the default value of $P$ without the effect of diversion traffic. Thus, according to formula (2), the RRR in a certain cell by applying SSP1 ($P_{R1}$) is:

$$P_{R1} = (P_{\text{default}} \frac{\beta N_T}{N_C})^\alpha$$  \hspace{1cm} (18)

**SSP2** aims at balancing diversion traffic among TDS’s. According to formula (3), SSP2 chooses source nodes which have the maximum average bandwidth of reachable TDS’s. In other words, SSP2 selects pseudo sources by considering only the bandwidth effects of TDS’s. Hence, every TDS has the same opportunity of being chosen to divert a call. In fact, the instant RRR of SSP2 fluctuates between the maximum value of RRR ($P^N$) and the minimum value of RRR ($P^1$). According to the parameters shown in Table 2, the average RRR in a certain cell by applying SSP2 ($P_{R2}$) is:

$$P_{R2} = \sum_{i=1}^{N} \left( \frac{N_i}{N_T} P^i \right)$$  \hspace{1cm} (19)

In formula (19), the number of current available TDS’s ($N_C$) does not change like the way in SSP1. Actually, $N_C$ starts to decrease only as every TDS in a certain cell nearly exhausts its own bandwidth, because the bandwidth consumption of each TDS during traffic diversion is balanced by SSP2. Additionally, the change of $P$ in SSP2 also depends on formula (16).

**SSP3** takes both the RRR and bandwidth of TDS’s into consideration. According to formula (3), (4) and (5), SSP3 uses a combined weight ($W_i$) related to the average free bandwidth of reachable TDS’s ($B_{MHi}$) and the RRR of each available pseudo source ($P_{MHi}$)) to choose source nodes. Then, SSP3 chooses source nodes with maximum $W_i$. By this means, source nodes will not partially congest TDS’s, but still keep a relatively low RRR. In SSP3, the number of current available TDS’s ($N_C$) changes like the way in SSP2. Hence, the RRR of SSP3 ($P_{R3}$) is mainly related to the number of reachable TDS’s. The probability of consuming the bandwidth in the area covered by $i$ TDS’s is:

$$P_T = \frac{i}{\sum_{i=1}^{N} i}$$  \hspace{1cm} (20)

According to the assumptions given above, TDS’s, which covers the same field, have the same probability of being chosen to divert traffic. Hence, we assume that TDS’s covering the same field have the same free bandwidth ($T_{Ti}$). According to formula (2), the RRR of a MH drops with the increase of the number of reachable TDS’s of the MH. Thus, the MH with maximum $W_i$ is also a MH with maximum value of $iT_{Ti}$. Moreover, assuming the current source node has $d$ reachable TDS’s, and it has not start diverting traffic, then:

If $nT_{Ti} = \max(T_{T1},...,NT_{TN})$, then $\delta = n$

Thus, according to formula (1), the current RRR in SSP3 is:

$$P_{R3} = (P)^\delta$$  \hspace{1cm} (21)

Considering formula (19), we have:

$$P_{R2} = P \leq \sum_{i=1}^{N} \left( \frac{N_i}{N_T} P^i \right) \leq P^N$$  \hspace{1cm} (22)

In terms of formula (18), changes from $N$ to 1 as the traffic of the system rises, and $P$ drops with the decrease of NC (formula (16)). Therefore, $P_{R1}$ could be smaller than $P_{R2}$ at a low level of system traffic, and higher than $P_{R2}$ at a high level of system traffic because of the higher TDS’s block rate caused by a partial selection of pseudo sources in SSP1. In SSP3, because (formula (21)) fluctuates between $N$ and 1, $P_{R3}$ is an intermediate value between $P_{R1}$ and $P_{R2}$ in general. Hence, if a network has a low overloaded traffic, or the number and the bandwidth of TDS’s is enough, we can use SSP1 to choose pseudo sources. By contrast, SSP2 is suitable for networks with limited number of TDSs and high overloaded traffic. SSP3 could be applied to networks with intermediate overloaded traffic.

### B. Signaling Overhead Analysis

In this subsection, we only analyze the signalling overhead brought by SSP’s over the whole system. As illustrated in Figure 2, there are 12 main steps for a successful traffic diversion service. For the analysis of routing overhead (see step 8 and 9 in Figure 2), we demonstrate the details in [7]. By applying SSP’s to the system, signalling overhead are brought to Original Source, Pseudo Source, Home BS, Neighbour MH and TDS, as
shown in Figure 2 and Figure 6. In this analysis, we calculate the signalling overhead needed for one successful bandwidth reallocation, and the amount of signalling overheads introduced to different nodes are computed separately.

Considering both Figure 2 and Figure 6, original source sends one SNREQ and receives one Bandwidth Re-allocation packet to and from Home BS (see steps 1 and 11 in Figure 2). Pseudo source (we choose MH2 in Figure 6 as a pseudo source) receives one SNREP and replies one Bandwidth Release packet after route discovery (see steps 7 and 10 in Figure 2). HOME BS sends NTDS number of MLREQ packets to TDS's and receives NTDSB number of packets to and from TDS's. Then, the overall number of signalling packets which HOME BS needs to process (see steps 1, 2, 3, 4, 5, 7, 10 and 11 in Figure 2) is:

\[ N_{SB} = 4 + N_{TS} + N_{TDSB} \]  

A TDS with enough free bandwidth needs to send the following number of NDREQ to all MH within its transmission area through ad hoc interface (see step 3 in Figure 2), and receives the same number of NDREP from its neighbour MH's (see step 4 in Figure 2):

\[ \frac{N_{MH}}{3\sqrt{2}} \pi R_{TDS}^2 = \frac{2\pi N_{MH} R_{TDS}^2}{3\sqrt{3} R_{BS}^2} \]  

Thus, the overall number of signalling packets which a TDS needs to process (see step 2, 3 and 4 in Figure 2) is:

\[ N_{STDS} = 1 + \frac{2\pi N_{MH} R_{TDS}^2}{3\sqrt{3} R_{BS}^2} \]  

Hence, the signalling overheads which HOME BS and TDS need to process are mainly related to the number of TDS's and the density of MH's. With the increase of diversion calls, the signalling overheads will obviously go up. With the bandwidth balancing in SSP2 and SSP3, the number of TDS's does not reduce quickly with the increase of diversion traffic, compared with SSP1. However, more TDS's will bring more signaling overheads to the network.

Considering the overall signalling overheads in the Home Cell (see steps 1, 2, 3, 4, 5, 7, 10 and 11 in Figure 2), we have:

\[ N_{overall} = 4 + N_{TDS} + N_{TDSB}(1 + \frac{2\pi N_{MH} R_{TDS}^2}{3\sqrt{3} R_{BS}^2}) \]  

V. ALGORITHM EVALUATION AND RESULT DISCUSSION

In this section, we evaluate the three SSP’s in terms of the average RRR of the overall network and the signalling overhead in Home BS and TDS’s, by plugging in reasonable values of parameters in formula (15) through (26). The numerical model is built by using MATLAB. The bandwidth and coordinates for both ‘Home BS’ and ‘TDS’ are pre-allocated. Each MH is deployed in random position, and the required bandwidth of every call is identical. The general cell topology is similar to Figure 5. There are 8 TDS’s deployed in the Home BS. The maximum traffic that a BS can carry without traffic diversion is \( T_{MAX} = 100 \) Erlangs. The maximum traffic that a TDS can burden is \( T_r = 10 \) Erlangs. The average traffic of each call is 1 Erlang. The average RRR of TDS’s is \( P_r = 0.5 \). The average number of MH’s within the transmission range of a TDS’s is 15. Because the RRR in SSP3 (\( P_{RR} \)) fluctuates between a maximum value and a minimum value, we only compute the average value of \( P_{RR} \) during the evaluation. Considering the signalling overhead analysis, we calculate the average signalling overheads required for one successful bandwidth re-allocation at each certain traffic level, and the amount of signalling overheads are only related to the number of TDS’s and MH’s without the consideration of collision. Because SSP’s which run in Home BS are centralized algorithms, the signalling overheads introduced by SSP’s are used to collect the network information for the decision of pseudo sources. Hence, the signalling overheads in SSP2 and SSP3 are identical due to the same method used for collecting the network information.

According to Figure 7, the average RRR in SSP2 keeps stable (at 0.25 in this case), because source nodes are chosen only according to the bandwidth status of TDS’s, not according to the number of reachable TDS’s. Supposing the average RRR of TDS’ is 0.5 as designed above and all neighbor cells have
enough free bandwidth to support diversion traffic, the request rejection rate of a call is only related to the number of its reachable TDS’s. If a TDS have some free bandwidth unused, it can still divert traffics from MH’s. When the diversion traffic is low, the RRR in SSP1 and SSP3 stays at a low level (at 0.13 in this case), compared with that in SSP2. However, as the diversion traffic in the home cell goes up, the average RRR in SSP3 jumps to an intermediate stage (at around 0.23 in Figure 7) because the bandwidth balancing amongst TDS’s causes that SSP2 choose source nodes with less number of reachable TDS’s. In SSP1, with the increase of the diversion traffic, more and more TDS’s are blocked due to the partial selection of TDS’s. The congestion of TDS’s causes a high average RRR of TDS’s. Thus, the RRR in SSP1 increases rapidly with increase of the diversion traffic (from 30 Erlangs to 60 Erlangs). In addition, the RRR in both SSP2 and SSP3 only rises sharply as the diversion traffic increases to a very high level (at around 72 Erlangs in this case), because the balance of the bandwidth consumption of TDS’s result in that only very high diversion traffic can congest TDS’s, and more TDS’s start to become congested at this traffic level. As more TDS’s become congested, the average RRR increases accordingly. Hence, if TDS’s have limited available bandwidth, SSP2 or SSP3 could be used to choose pseudo sources to avoid congesting TDS’s. If the diversion traffic in a network is not high, SSP1 can be adapted to achieve a low RRR of the system.

In respect of the signalling overhead, the number of signalling messages which Home BS in SSP1 needs to process decreases as the diversion traffic increases (as shown in Figure 8). This is because more diversion traffic from the Home BS could cause more TDS’s being congested due to the lack of bandwidth balance amongst TDS’s in SSP1, and more congested TDS’s lead to more dropping of signalling messages and as such less signalling messages being eventually transmitted. However, because both SSP2 and SSP3 can balance the consumption of bandwidth amongst TDS’s, the number of signalling messages decreases only when diversion traffic goes very high (at around 72 Erlangs in Figure 8). As mentioned above, as signalling overheads are only related to the number of TDS’s or MH’s, the overheads only drop when the number of available TDS’s or MH’s goes down. Then, the signalling messages which Home BS in SSP2 or SSP3 needs to process decreases only when the amount of the diversion traffic is very high. A similar trend happens as analyzing the signalling overheads brought to TDS’s. The number of signalling messages which a TDS needs to process is more than those which a Home BS needs to process, because a TDS needs to broadcast and receive signalling messages from all its neighbor MH’s. In normal cases, the number of neighbor MH’s are more the number of TDS’s. To achieve a small amount of signaling overheads, SSP1 can be applied, but this may cause a higher request rejection rate, compared with SSP2 and SSP3.

Figure 9 gives the overall signalling overheads introduced by SSP’s. The whole trend of each curves are similar to those in Figure 8. The reasons of how the signalling overheads changes are similar to the explanations of Figure 8.

VI. CONCLUSION

This paper focuses on the SSP part of routing protocols in ICNs. Three SSP algorithms are designed to run in BS’s to choose pseudo sources to divert their calling traffic. SSP’s proposed in this paper aim to choose source nodes that have the most possibility of successfully discovering relaying routes.
Moreover, the selection of source nodes in SSP’s also tries to balance the bandwidth of TDS’s, which is used for traffic diversion. The evaluation of these algorithms has been presented and some guidelines as to when to use which algorithm have also been discussed. According to the number and bandwidth of TDS’s, network planners can choose a reasonable SSP to achieve a relatively low call block rate. Alternatively, by estimating the amount of overloaded traffic, planners can choose a SSP to reduce the number of TDS’s deployed in each cell.

In this paper the design of SSP’s is mainly driven by the need of reducing the request rejection rate of a cellular network system. Though the SSP’s proposed in this paper are not designed specifically for data transmission, the principle adopted by the algorithms is equally applicable to data transmission. Considering the guarantee of data transmission, we need to specify the bandwidth required by the original source, and then the released bandwidth of pseudo sources (which could be more than one) should be able to satisfy the bandwidth requirement.

REFERENCES