4.1 Introduction

Data allocation problem in distributed database design is NP-hard [10, 57] and therefore optimal solutions cannot be found in fair period of time even for a small extent problems. To solve NP-hard problems in reasonable amount of time, one must has to remove the requirement that an algorithm must always generate a true optimal solution. This requirement will be replaced by the requirement that the algorithm seeks ‘close’ to the true optimal solution instead of true optimal solution of NP-hard problems in reasonable amount of time. Algorithms that seek such type of solutions of NP-hard problems are called approximation algorithms. In the last two decades, meta-heuristic algorithms have been flourishing for optimization problems. These algorithms are inspired by the behaviors of natural phenomena. These biologically inspired meta-heuristic algorithms have the capacity of find out optimal or near optimal solutions in a relatively large search space than the classical optimization approaches and it happened because of their domain, which is independent in nature.

In evolutionary computation research, genetic algorithm (GA) is one of the most popular techniques that belong to such category. Natural genetic variation and natural selection inspired operators are used by genetic algorithm [41]. Genetic algorithm is a population based optimization algorithms. Genetic algorithm has been successfully developed and applied to solve variety of real world problems.

Corcoran and Hale [35], March and Rho [69], Karlapalem et al. [51], Ahmad et al. [8], Loukopoulos and Ahmad [62], Rahmani et al. [80], Mamaghani et al. [68], and Tosun et al. [99] have used genetic algorithm in different ways to solve data allocation problem in distributed databases. Corcoran and Hale [35] and Ahmad et al. [8] have suggested genetic algorithm as an attractive way out for efficient and quality solution for data allocation problem in distributed database systems.

Recently, a new meta-heuristic algorithm named Biogeography-Based Optimization (BBO) is introduced by Simon [89]. BBO is a population based evolutionary
algorithms for global optimization. BBO is based on the theory of biogeography which is the study of the geographical distribution of biological organisms [89]. As all other evolutionary algorithms, BBO also uses operators to share information between solutions in each generation to improve the solution quality [65]. Thus, BBO can be used to solve many problems that have been solved by other evolutionary algorithms. BBO is being developed successfully and it is also being applied to solve problems in a variety of areas for example, optimal power allocation in wireless sensor networks [18], job shop scheduling problem [46], reconfiguration of distribution network [55], selection of the optimal electrochemical machining process parameters [73], economic and emission dispatch problems [85], optimal synthesis of thinned antenna arrays [94], feature subset selection [106] and optimization of objective functions [42, 65, 89]. Moreover Ma [65], Ma et al. [66], Simon [89], Simon et al. [90] and Simon et al. [93] have proved that BBO demonstrates good performance and outperform many other evolutionary optimization algorithms for example genetic algorithm (GA), differential evolution (DE), ant colony optimization (ACO), particle swarm optimization (PSO) etc. on various optimization problems.

Therefore, in this chapter it is attempted to find the optimal allocation of data in distributed database design using biogeography-based optimization (BBO) and simplified biogeography-based optimization (SBBO).

4.2 Biogeography-Based Optimization Technique

4.2.1 Biogeography

Biogeography is the study of geographical distribution of species in geographic spaces. MacArthur and Wilson [67] were the first to develop a mathematical model of biogeography. They have proposed “The Theory of Island Biogeography”. The theory of island biogeography suggests that the immigration and extinction determines the number and variation of species available on an island. The distance of an island influences the patterns immigration and emigration. The likelihood of immigration is far less at the islands that are more isolated as compare to those one which are less isolated.
The mathematical model on biogeography describes that the stability among the immigration of new species on an island and the emigration of already inhabited species is the determining factor to measure the rate of change in the patterns of species in terms of number and verity on an island [67]. The term “island” used here is descriptive rather than literary i.e. it is a habitat area which is geographical isolated from other habitat areas in terms of distance/space. From literary point of view, an island is isolated from other habitats by water. But islands can also be habitats those are isolated by wider range of desert, rivers, mountain ranges or other barriers.

A geographical area suitable for survival and existence of biological species in terms of residence is having a high habitat suitability index (HSI). Vegetation diversification, patterns of rainfall, land mass and ranges of topographic features are the different attribute that are associated with HSI. The variables that characterize habitability are featured as suitable index variables (SIVs). As far independent or dependent nature of HSI and SIV is concerned, the SIVs can be regarded as habitat’s independent variables. Whereas HSI can be measured as a dependent variable [89]. It is also significant that the habitats having high HSI are densely populated and have many species. Habitats with low HSI are sparsely populated and have less number of species. Many species of high HSI habitats emigrate to the nearby habitats. This emigration accrues due to large number of species hosted by habitats having high HSI. Very few species immigrate to habitats with high HSI because they are lack of additional resources required for immigrating species. So, low immigration rate is resulted at the habitats having high HIS. Therefore, habitat with high HSI are more static than the habitat with low HSI. It can’t be construed that the emigrating species from a home habitat will result into the absolute vanishing of these species from their original habitat. Merely a small number of representatives of the emigrating species emigrate to the nearby habitat, so emigrating species are still presented in its home habitat.

In a habitat having low HSI has a small number of species. Therefore it has sufficient space to accommodate additional species. This results into high immigration of species to habitats having low HSI. On the contrary emigration rate at these habitats is low. At the same time, another factor which is quite significant in the context of the low HSI. This factor is related to the diversity of a habitat. The diversity of a habitat is will increased the suitability index proportionately because the low HSI may change to high
HSI with result of immigration of new species. In case the HSI of a habitat continues to be low after immigration of new species, then the species residing there will tend to died out. This will allow other species to immigrate there. So, habitat with low HSI are more dynamic than the habitat with high HSI.

Figure 4.1 describes a model of species abundance in a single habitat [67]. The emigration rate ($\mu$) and the immigration rate ($\lambda$) are functions of the number of species in a habitat.

In case there is no species present in the habitat then the immigration rate to the habitat is maximum ($I$) as shown Figure 4.1. As the number of species increases in the habitat, the immigration rate decreases. When the number of species in the habitat starts increasing, the immigration rate starts decreasing and becomes zero at a point when the number of species reaches to $S_{\text{max}}$, the highest permissible number supported by the habitat. On the other hand, when there is no species present in the habitat then the emigration rate is zero as evident from Figure 4.1. When the number of species in the habitat starts increasing, the emigration rate also starts increasing and becomes maximum ($E$) at a point when the number of species reaches to $S_{\text{max}}$, the highest permissible number supported by the habitat.

![Figure 4.1: Species Model of a Single Habitat [67]](image_url)
At \( S_0 \) when the number of species reaches equilibrium state, the rate of immigration and emigration are equal. The immigration and emigration curves shown in the Figure 4.1 as straight lines, but they might be more complex curves.

### 4.2.2 Biogeography-Based Optimization (BBO)

The biogeography-based optimization (BBO) is a newly developed population-based evolutionary technique. Biogeography-based optimization (BBO) is based on theory of biogeography. Simon [89] developed the biogeography-based optimization (BBO). BBO is primarily based on “*The Theory of Island Biogeography*” given by MacArthur and Wilson [67]. The mathematical model on biogeography describes that the stability among the immigration of new species on an island and the emigration of already inhabited species is the determining factor to measure the rate of change in the patterns of species in terms of number and verity on an island.

BBO works on a population of candidate solution to solve a global optimization problem. Candidate solutions are called habitats. Feasible solution to the problem undertaken is represented by habitat. Each solution feature of a habitat is called a suitability index variable (SIV) of that habitat. A quality solution is considered to be a habitat with a high habitat suitability index (HSI). The quality of a candidate solution is measured by its HSI. The “fitness” of each habitat is represented by its habitat suitability index (HSI). A low HSI habitat provides a poor solution.

A habitat which is accommodates a high number of species is represented by high HSI solutions whereas a habitat which is accommodates a less number of species is represented by low HSI solutions. Therefore HSI of a solution is determining index to figure out the number of species present in the solution. High HSI solutions share their features with other solutions in the population and low HSI solutions accepts shared features from other solutions [65, 89].

Biogeography-based optimization (BBO) [89] assumes that each solution (habitat) has a symmetric species curve with \( E = I \), but HSI of a solution determines the \( S \) (number of species) value of that solution. Figure 4.2 represents the emigration and immigration curves of two candidate solutions to some problem. \( S_1 \) represents a low HSI solution and \( S_2 \) represents a high HSI solution. The comparison of \( S_1 \) and \( S_2 \) depicts that a small number of species represented by \( S_1 \) with low HSI habitat, whereas the number of
species is high in $S_2$ with high HSI habitat. Consequently, $S_1$ is representing higher immigration rate in comparison to the immigration rate of $S_2$. At the same time $S_1$ is representing low emigration rate as compare to $S_2$ [89].

In BBO, $\lambda_k$ and $\mu_k$ are, respectively, representing the immigration rate and the emigration rate of the $k^{th}$ candidate solution (habitat). The immigration rate and emigration rate are functions of number of species in the habitat [89]. The immigration rate ($\lambda_k$) and emigration rate ($\mu_k$) can be calculated as follows [89]:

$$\lambda_k = I \left(1 - \frac{K}{n}\right)$$  \hspace{1cm} 4.1

$$\mu_k = E \left(\frac{K}{n}\right)$$  \hspace{1cm} 4.2
Where, \( I \) is the maximum possible immigration rate; \( E \) is the maximum possible emigration rate; \( K \) is the number of species of the \( k \)th individual and \( n \) is the maximum number of species.

In biogeography-based optimization, there are two operators: migration and mutation [89].

### 4.2.2.1 Migration

Migration is a probabilistic operator that improves the quality of a habitat. The immigration and emigration rate of each solution are used to probabilistically share the information between habitats [65]. For each habitat \( H_i \), its immigration rate \( \lambda_i \) is used to probabilistically make a decision whether to immigrate or not. If \( H_i \) is selected for immigration, then the emigrating habitat \( H_e \) is selected probabilistically based on the emigration rate \( \mu_e \). Migration is represented as [89]:

\[
H_i (SIV) \leftarrow H_e (SIV)
\]

Where, \( SIV \) is a solution feature of the candidate solution to the optimization problem.

Select \( H_i \) with probability based on \( \lambda_i \);

**If** \( H_i \) is selected

Select \( H_e \) with probability based on \( \mu_e \);

**If** \( H_e \) is selected

Randomly select an SIV from \( H_e \);

Replace a random SIV in \( H_i \) with one selected from \( H_e \);

**End if**

**End if**

**Figure 4.3: Pseudo-code for Migration [64]**
4.2.2.2 Mutation

Mutation is a probabilistic operator that randomly modifies a habitat’s SIV [89]. A randomly generated SIV replaces a selected SIV in the solution \( H_i \) according to a mutation probability, which is predefined. The main reason of mutation is to increase diversity of the population. Mutation is useful for poor solution as well good solution. For low HSI solutions, mutation gives them an opportunity of enhancing the quality of solutions, and for high HSI solutions, mutation is capable of making them better. Mutation rate \((m)\) is calculated by using following equation [89]:

\[
m_i = m_{\text{max}} \left( 1 - \frac{P_i}{P_{\text{max}}} \right)
\]

Where, \( m_{\text{max}} \) is the user defined maximum mutation probability, \( P_{\text{max}} \) maximum probability and \( P_i \) is the solution probability.

![Figure 4.4: Pseudo-code for Mutation [64]](image)

4.2.3 BBO Algorithm

The basic structure of biogeography-based optimization algorithm is given below [89, 64]:

**Step 1:** Initialize the BBO parameters:

- Maximum Immigration rate \((I)\) = Maximum Emigration rate \((E)\)
- Maximum Mutation rate
- Elitism Parameter
Size of the Population

Termination Condition

**Step 2:** Generate a random set of habitats. Each habitat represents a potential solution to the given problem.

**Step 3:** Evaluate habitats and compute corresponding HSI value of each habitat.

**Step 4:** Calculate the immigration rate ($\lambda$) and emigration rate ($\mu$) for each habitat according to HSI of each habitat by using Equation 4.1 and Equation 4.2 respectively.

**Step 5:** Probabilistically choose a non-elite immigration habitat ($H_i$) based on the immigration rate ($\lambda_i$).

**Step 6:** If $H_i$ is selected then select emigrating habitat ($H_j$) based on the emigration rate ($\mu_j$).

**Step 7:** If $H_j$ is selected then randomly select a SIV from $H_j$ and $H_i$:

$$H_i \text{(SIV)} \leftrightarrow H_j \text{(SIV)}$$

**Step 8:** Probabilistically perform mutation based on the mutation rate using Equation 4.4.

**Step 9:** If the termination condition is not met go to Step 3.

Else, terminate.
Pseudo-code of BBO is given in Figure 4.5

Create a random set of habitat (population) \( H_1, H_2, \ldots, H_n \);

Compute corresponding HSI values of each habitat;

While the termination condition is not satisfied do

Compute immigration rate (\( \lambda_i \)) and emigration rate (\( \mu_i \)) for each habitat based on HSI;

Select \( H_i \) with probability based on \( \lambda_i \);

If \( H_i \) is selected

Select \( H_e \) with probability based on \( \mu_e \);

If \( H_e \) is selected

Randomly select an SIV from \( H_e \);

Replace a random SIV in \( H_i \) with one selected from \( H_e \);

End if

End if

Select \( H_i \) (SIV) based on mutation probability \( m_i \);

If \( H_i \) (SIV) is selected

Replace \( H_i \) (SIV) with a randomly generated SIV;

End if

Recalculate HSI of each habitat;

End while

Figure 4.5: Pseudo-code for BBO [65]
4.3 BBO’s Comparison with Other Evolutionary Algorithms

BBO is a population-based optimization algorithm [89]. As all other evolutionary algorithms, BBO also uses operators to share information between solutions in each generation to improve the solution quality [65]. Thus, BBO can be used to solve many problems that have been solved by other evolutionary algorithms. BBO has some characteristics that make it different from other evolutionary algorithms.

- BBO does not involve reproduction or the generation of “children” rather it modifies the solution using migration. This makes BBO different from reproduction strategies such as GA and other evolutionary strategies [65, 89].
- BBO is also different from ant colony optimization (ACO), because BBO maintains its set of solutions from one generation to the next. ACO, on the other hand generates a new set solution after every generation [65, 89].
- BBO is contrasted with particle swarm optimization (PSO) and differential evolution (DE). PSO changes the solutions indirectly by another variable called velocities of the solutions. DE is not biologically motivated and it changes the solutions based on the differences between other solutions. On the other hand, BBO solutions are changed directly by migration of SIVs between different habitats [89].

<table>
<thead>
<tr>
<th>Table 4.1: BBO vs. GA [46]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BBO</strong></td>
</tr>
<tr>
<td>Population Based</td>
</tr>
<tr>
<td>Habitat</td>
</tr>
<tr>
<td>SIV</td>
</tr>
<tr>
<td>Habitat consisted of SIV</td>
</tr>
<tr>
<td>Migration operator (No reproduction)</td>
</tr>
<tr>
<td>Mutation operator</td>
</tr>
<tr>
<td>Goodness of the solution is characterized by high HIS</td>
</tr>
<tr>
<td>No solution is discarded from the initial population during iteration rather solution are modified</td>
</tr>
</tbody>
</table>
4.4 Framework for Data Allocation Using BBO

4.4.1 Habitats Representation

4.4.1.1 Non-replicated Allocation

Each habitat is defined by a vector with SIV, where each SIV represents the allocation of a fragment to a site. The numbers of SIVs in a habitat depend on the number of fragments that has to be allocated. Each habitat represents a candidate solution i.e. data allocation schema. The \(i^{th}\) individual habitat \((H_i)\) can be defined as follows:

\[
H_i = [\text{SIV}_1, \text{SIV}_2, \text{SIV}_3, \ldots, \text{SIV}_m]
\]

Where, \(m\) is the total number of fragments and \(\text{SIV}_k \in S\), the set of all the sites in the communication network.

Habitats are encoded in a binary representation. For example in the case of 3 sites and 5 fragments, a habitat is represented by 5 SIVs of 2 bit each, one SIV for each fragment. If fragment 1 is allocated to site 3; fragment 2 is allocated to site 1; fragment 3 is allocated to site 3; fragment 4 is allocated to site 2; fragment 5 is allocated to site 3, then the habitat for these allocated fragments will be \{11 01 11 10 11\}.

4.4.1.2 Replicated Allocation

Each habitat is defined by a vector with SIV. Each SIV is represented by a set of \(n\) bits, where \(n\) is the total number of site in the communication network. The numbers of SIVs in a habitat depend on the number of fragments that has to be allocated. Each habitat represents a candidate solution i.e. data allocation schema. The \(i^{th}\) individual habitat \((H_i)\) can be defined as follows:

\[
H_i = [\text{SIV}_1, \text{SIV}_2, \text{SIV}_3, \ldots, \text{SIV}_m]
\]

where, \(m\) is the total number of fragments and \(\text{SIV}_k\) is a bit structure representing the decision variables \(X_{kj}\). The decision variable \(X_{kj}\) has value 1 if the fragment \(F_k\) is allocated to a site \(S_j\); otherwise 0.

For example in case of 5 sites and 4 fragments, if first fragment is assigned to site 2 and site 4 then the value of related SIV will be \{01010\}, second fragment is assigned to site 1, site 3 and site 4 then the value of related SIV will be \{10110\}, third fragment is assigned to site 2, site 3 and site 4 then the value of related SIV will be \{01110\} and
fourth fragment is assigned to all the sites then the value of related SIV will be \{11111\}. A binary string is formed by concatenating the assignment values of all the data fragments i.e. \{01010 10110 01110 11111\}. A potential solution (habitat) to the data allocation problem is represented by the binary string.

**4.4.2 Habitat Evaluation**

In case non-replicated allocation, habitat suitability index (HSI) of each habitat \(H_k\) is calculated using Equation 3.5 and in case of replicated allocation, habitat suitability index (HSI) of each habitat \(H_k\) is calculated using Equation 3.14. Habitats generating less cost are having high HSI and habitats generating high cost are having low HSI. The corresponding immigration rate \(\lambda_k\) and emigration rate \(\mu_k\) of each habitat \(H_k\) is calculated using Equation 4.1 and Equation 4.2 respectively. For each generation, all the habitats are evaluated and the habitat which minimizes the cost function will be preferred as a solution to the data allocation problem.

**4.4.3 Migration**

Let \(N\) be the size of population that is the total no of habitats are \(N\) and \(H_k\) is the \(k^{th}\) individual in the population. At each generation, the immigration habitat \(H_i\) is selected with a probability proportional to the immigration rates \(\lambda_i\). Roulette-wheel selection method is used for selecting immigration habitat:

\[
\text{Probability that } H_i \text{ is selected for immigration} = \frac{\lambda_i}{\sum_{i=1}^{N} \lambda_i}
\]

The emigration habitat \(H_e\) is selected with a probability proportional to the emigration rates \(\mu_e\). Roulette-wheel selection method is used for selecting emigration habitat [92]:

\[
\text{Probability that } H_e \text{ is selected for emigration} = \frac{\mu_e}{\sum_{e=1}^{N} \mu_e}
\]

Single immigration-based BBO [91] approach is used to probabilistically share information between different habitats of the population during the migration process.
4.4.4 Mutation

To simplify the mutation process, mutation is implemented in a standard way as it is done in other evolutionary algorithms [92]. The probability of mutation is defined as a constant \( p_m \in [0,1] \). At each generation, a uniformly distributed random number (\( r \)) is generated, where \( r \in [0,1] \). If \( r < p_m \) then a randomly selected SIV from \( H_i \) is replaced with a randomly generated SIV.

4.4.5 Constraints Handling

The process of initial population of habitats is randomly generated which satisfy the various constraints of data allocation problem. In case of non-replicated allocation, constraints are defined by Equation 3.2. In case of replicated allocation, constraints are defined by Equation 3.9 and Equation 3.10.

During each generation of the algorithm, a constraint checking procedure is run to check the modified habitat in order to assure that the respective constraints of the data allocation problem are satisfied. The modified habitat will be introduced into the population of the next generation only if it is feasible and satisfies the respective constraints.

4.4.6 BBO Algorithm for Data Allocation

The algorithm of the proposed allocation of data fragments in a static distributed database environment is given below:

**Input:** Database Information, Applications Information, Sites Information and Network Information

**Output:** Non-replicated/Replicated Fragment Allocation Schema

**Method:**

**Step 1:** Initialize the BBO parameters:

\[
\text{Maximum Immigration rate } (I) = \text{Maximum Emigration rate } (E)
\]
Maximum Mutation rate (MR)
Elitism Parameter
Size of the Population
Termination Condition

**Step 2:** Generate a random set of habitats equal to the size of the population satisfying the different constraints (Equation 3.2 in case of non-replicated allocation, and Equation 3.9 and Equation 3.10 in case of replicated allocation)

**Step 3:** Evaluate habitats and compute corresponding HSI value of each habitat (In case of non-replicated allocation, HSI of each habitat \( H_k \) is calculated using Equation 3.5 and in case of replicated allocation, HSI of each habitat \( H_k \) is calculated using Equation 3.14)

**Step 4:** Calculate the immigration rate \( \lambda \) and emigration rate \( \mu \) for each habitat according to respective HSI using Equation 4.1 and Equation 4.2 respectively

**Step 5:** Select a non-elite immigration habitat \( H_i \) using roulette-wheel selection based on the immigration rates \( \lambda_i \) and randomly select a SIV of \( H_i \)

**Step 6:** Select emigrating habitat \( H_e \) using roulette-wheel selection based on the emigration rates \( \mu_e \) and randomly select a SIV from \( H_e \)

**Step 7:** \( H_i \text{(SIV)} \leftrightarrow H_e \text{(SIV)} \)

**Step 8:** Generate \( r \in [0,1] \);
Figure 4.6: Data Allocation Framework using BBO

1. Initialize Information related to Database, Applications, Sites and Communication Network
2. Set BBO Parameters
3. Generate a Random Set of Habitats
4. Evaluate Habitats and Compute corresponding HSI value of each Habitat
5. Calculate Immigration Rate and Emigration Rate of each Habitat
6. Select a non-elite Immigrating Habitat \( H_i \) using Roulette-Wheel Selection and Pick a random SIV from \( H_i \)
7. Select Emigrating Habitat \( H_e \) using Roulette-Wheel Selection and Pick a random SIV from \( H_e \)
8. \( H_i (\text{SIV}) \rightarrow H_e (\text{SIV}) \)
9. Generate a random number \( r \) \( 0 \leq r \leq 1 \)
10. Is \( r \) less than Maximum Mutation Rate?
    - No
    - Yes
11. Replace randomly selected SIV from \( H_i \) with randomly generated SIV
12. Calculate HSI of \( H_i \) and Generate New population from modified Habitat \( (H_i) \)
13. Is Predefined Criterion Satisfied?
    - No
    - Yes
14. Fragments Allocation Schema
Step 9: If $r < MR$ (Maximum mutation rate) then a randomly selected SIV from $H_i$ is replaced with a randomly generated SIV

Step 10: Calculate HSI of $H_i$. Verify the feasibility of the solution and check the corresponding constraints

Step 11: If the modified habitat ($H_i$) is feasible and satisfies the respective constraints then the modified habitat will be replaced with its new version in the population of the next generation

Step 12: If the termination condition is not met go to Step 3

Else terminate

4.5 Simplified Biogeography-Based Optimization Technique

Simplified biogeography-based optimization (SBBO) is a simplified version of the BBO [91]. Simplified biogeography-based optimization (SBBO) always uses the best solution (habitat) from the population as the emigration habitat and any other randomly chosen solution (habitat) from the population is selected as the immigration habitat [91]. The immigrating habitat is selected from a uniform probability distribution. The migration curves of the SBBO are shown in Figure 4.7. All the solutions (habitats) in

![Figure 4.7: Migration Curves of SBBO in a Population of n Habitats [91]](image-url)
an n-habitat population have an equal probability of immigration except for the fittest solution (habitat). The fittest solution (habitat) has a 100% probability of emigration and a zero probability of immigration [91].

```
Find the fittest solution. Call this solution $H_e$;
Randomly select an SIV from $H_e$;
Select the immigrating habitat ($H_i$) from a uniform probability distribution;
Replace a random SIV in $H_i$ with one selected from $H_e$ i.e.

$$H_i(SIV) \leftarrow H_e(SIV)$$
```

Figure 4.8: Pseudo-code for SBBO Migration [91]

4.6 Framework for Data Allocation Using SBBO

4.6.1 Habitats Representation

Representation of habitat is same as discussed in Section 4.3.1 for both non-replicated and replicated allocation.

4.6.2 Habitat Evaluation

In case non-replicated allocation, habitat suitability index (HSI) of each habitat ($H_k$) is calculated using Equation 3.5 and in case of replicated allocation, habitat suitability index (HSI) of each habitat ($H_k$) is calculated using Equation 3.14. Habitat generating minimum cost for allocation will be the fittest solution (habitat). For each generation, all the habitats are evaluated and the habitat which minimizes the cost function will be preferred as a solution to the data allocation problem.

4.6.3 Migration

Let $N$ be the size of population that is the total no of habitats are $N$ and $H_k$ is the $k^{th}$ individual in the population. At each generation, the fittest habitat ($H_e$) is selected as an
emigration habitat and any other randomly chosen habitat from the population is
selected as the immigration habitat.

Single immigration-based BBO [91] approach is used to probabilistically share
information between different habitats of the population during the migration process.

4.6.4 Mutation

Mutation process is same as explained in Section 4.3.4.

4.6.5 Constraints Handling

Method of handling different constraint during each generation is also same as
explained in Section 4.3.5.

4.6.6 SBBO Algorithm for Data Allocation

The algorithm of the proposed allocation of data fragments in a static distributed
database environment is given below:

**Input:** Database Information, Applications Information, Sites Information and
Network Information

**Output:** Non-replicated/Replicated Fragment Allocation Schema

**Method:**

**Step 1:** Initialize the SBBO parameters:

- Maximum Mutation rate (MR)
- Size of the Population
- Termination Condition
Step 2: Generate a random set of habitats equal to the size of the population satisfying the different constraints (Equation 3.2 in case of non-replicated allocation, and Equation 3.9 and Equation 3.10 in case of replicated allocation)

Step 3: Evaluate habitats and compute corresponding HSI value of each habitat (In case of non-replicated allocation, HSI of each habitat (Hk) is calculated using Equation 3.5 and in case of replicated allocation, HSI of each habitat (Hk) is calculated using Equation 3.14)

Step 4: Find the fittest solution. Call this solution $H_e$

Step 5: Randomly select an SIV from $H_e$

Step 6: Select the immigrating habitat ($H_i$) from a uniform probability distribution

Step 7: Replace a random SIV in $H_i$ with one selected from $H_e$ i.e.

$$H_i(\text{SIV}) \leftarrow H_e(\text{SIV})$$

Step 8: Generate $r \in [0,1]$;

Step 9: If $r < \text{MR}$ (Maximum mutation rate) then a randomly selected SIV from $H_i$ is replaced with a randomly generated SIV;

Step 10: Calculate HSI of $H_i$. Verify the feasibility of the solution and check the corresponding constraints;

Step 11: If the modified habitat ($H_i$) is feasible and satisfies the respective constraints then the modified habitat will be replaced with its new version in the population of the next generation;

Step 12: If the termination condition is not met go to Step 3
Else terminate;
Figure 4.9: Data Allocation Framework using SBBO
4.7 Summary

This Chapter starts with a general outline of the biogeography-based optimization. After the general outline of BBO, migration and mutation operators used in the biogeography-based optimization are described in detail. Biogeography-based optimization is also compared with other evolutionary algorithm. Framework for data allocation using biogeography-based optimization (BBO) has been proposed. Finally, a simplified version of biogeography-based optimization known as simplified biogeography-based optimization (SBBO) is explained and another framework for data allocation using simplified biogeography-based optimization (SBBO) has been proposed.