EAI/Springer Innovations in Communication and Computing

Pandian Vasant
Igor Litvinchev
José Antonio Marmolejo-Saucedo
Editors

Modeling, Simulation, and Optimization

Springer

EAI/Springer Innovations in Communication and Computing

Series editor
Imrich Chlamtac, CreateNet, Trento, Italy

Editor's Note

The impact of information technologies is creating a new world yet not fully understood. The extent and speed of economic, life style and social changes already perceived in everyday life is hard to estimate without understanding the technological driving forces behind it. This series presents contributed volumes featuring the latest research and development in the various information engineering technologies that play a key role in this process.

The range of topics, focusing primarily on communications and computing engineering include, but hardly limited to, wireless networks; mobile communication; design and learning; gaming; interaction; e-health and pervasive healthcare; energy management; smart grids; internet of things; cognitive radio networks; computation; cloud computing; ubiquitous connectivity, and in mode general smart living, smart cities, Internet of Things and more. The series publishes a combination of expanded papers selected from hosted and sponsored European Alliance for Innovation (EAI) conferences that present cutting edge, global research as well as provide new perspectives on traditional related engineering fields. This content, complemented with open calls for contribution of book titles and individual chapters, together maintain Springer's and EAI's high standards of academic excellence. The audience for the books consists of researchers, industry professionals, advanced level students as well as practitioners in related fields of activity include information and communication specialists, security experts, economists, urban planners, doctors, and in general representatives in all those walks of life affected ad contributing to the information revolution.

About EAI

EAI is a grassroots member organization initiated through cooperation between businesses, public, private and government organizations to address the global challenges of Europe's future competitiveness and link the European Research community with its counterparts around the globe. EAI tens of thousands of members on all continents together with its institutional members base consisting of some of the largest companies in the world, government organizations, educational institutions, strive to provide a research and innovation platform which recognizes excellence and links top ideas with markets through its innovation programs.

Throughs its open free membership model EAI promotes a new research and innovation culture based on collaboration, connectivity and excellent recognition by community.

More information about this series at http://www.springer.com/series/15427

Pandian Vasant • Igor Litvinchev
José Antonio Marmolejo-Saucedo
Editors

Modeling, Simulation, and Optimization

Springer

Editors
Pandian Vasant
Department of Fundamental and Applied Sciences
Universiti Teknologi PETRONAS
Tronoh, Malaysia

Igor Litvinchev
Universidad Autónoma de Nuevo León
San Nicolás de los Garza
Nuevo León, Mexico

José Antonio Marmolejo-Saucedo
Panamerican University
Mexico

ISSN 2522-8595 ISSN 2522-8609 (electronic)
EAI/Springer Innovations in Communication and Computing
ISBN 978-3-319-70541-5 ISBN 978-3-319-70542-2 (eBook)
https://doi.org/10.1007/978-3-319-70542-2

Library of Congress Control Number: 2017961571

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The first edition of the International Conference on Computer Science and Engineering (COMPSE) 2016 was held during November 11–12, 2016, at *Golden Sands Resort* in Penang, Malaysia. The objective of the international conference is to bring together experts and scientists in the research areas of computer science and optimization from all over the world to share their knowledge and experiences on the current research achievements in these fields. This conference provides a golden opportunity for the global research community to interact and share their novel findings and research discoveries among their colleagues and friends. The proceedings of COMPSE 2016 are published by EU Digital Library and indexed by DBLP, EI, Google Scholar, Scopus, and Thomson ISI.

For this edition, the program committee received over 100 submissions from 25 countries, and each paper was reviewed by at least four expert reviewers. The prominent technical committee has selected the best 44 papers for final presentation at the conference venue of Golden Sands Resort in Penang, Malaysia. The organizing committee would like to sincerely thank all the authors and reviewers for their wonderful job for this conference. The best and high-quality papers will be selected and reviewed by the international program committee in order to publish the extended version of the paper in the *EAI/Springer Innovations in Communications and Computing Book Series*, *Intelligent Decision Technologies: An International Journal* (IOS), the *Journal of Computational Science* (Elsevier), and the *Journal of Ambient Intelligence and Humanized Computing* (Springer).

COMPSE 2016 was organized by the European Alliance for Innovation (EAI). This conference would not have been organized without the strong support and help from the staff members of Golden Sands Resort and the organizing committee of COMPSE 2016. We would like to sincerely thank Prof. Imrich Chlamtac (University of Trento and Create-NET), Barbara Fertalova (EAI), and Lucia Zatkova (EAI) for their great help and support in organizing the conference. We also appreciate the fruitful guidance and help from Prof. Gerhard Wilhelm Weber (Middle East Technical University, Turkey), Prof. Rustem Popa ("Dunarea de Jos" University in

Galati, Romania), Prof. Goran Klepac (Raiffeisenbank Austria D.D., Croatia), Prof. Leopoldo Barron (Tecnológico de Monterrey, Mexico), Prof. Ivan Zelinka (VSB-TU Ostrava, Czech Republic), Dr. Jose Antonio Marmolejo (Universidad Anahuac Mexico Norte, Mexico), and Dr. Vo Hoang Duy (Ton Duc Thang University, Vietnam).

Mexico José Antonio Marmolejo-Saucedo
Tronoh, Malaysia Pandian Vasant
Nuevo León, Mexico Igor Litvinchev
November 2016

Contents

Chapter 1
Texture Synthesis and Design Based on Element Distribution Creation

Yan Gui, Yang Liu, and Feng Li

1.1 Introduction

Texture synthesis plays a key role in computer graphics and computer vision. There has been a plethora of research toward texture synthesis, of which example-based texture synthesis methods [1] has become the main threads. To a wide variety of textures ranged from stochastic to structured, we focus on deterministic textures that are formed by spatial repetition of texture elements. As shown in Fig. 1.1, the range from left to right are regular, near-regular, and non-regular deterministic textures. It would be very difficult to synthesize the repetitive elements adequately with most existing texture synthesis methods, such as local neighborhood matching-based methods [2–7] and optimization-based methods [8].

Potential solutions [9–15] have been explored through imitating the element distributions in input exemplars. For example, Dischler et al. [9] segmented an input exemplar into texture particles and generated 2D textures or textures on arbitrary surfaces by adding them according to sets of co-occurrences. Barla et al. [10] proposed a method to synthesize 2D arrangements of seed points and pasted input elements to those locations by local neighborhood matching. Ijiri et al. [11] synthesized 2D distributions by locally growing through rule-based heuristics. However, the above methods [9–11] cannot handle elements with complex shapes which are closely correlated with spatial distributions. Hurtut et al. [12] proposed a statistical model to learn spatial interactions between and within different categories. Passos et al. [13] presented an improved method for arrangement synthesis defined as 2D collection of elements, which provides control over local density of elements.

Y. Gui (✉) • Y. Liu • F. Li
Provincial Key Laboratory of Intelligent Processing of Big Data on Transportation, Changsha University of Science and Technology, Changsha, Hunan, P. R. China
e-mail: Guiyan122@163.com

P. Vasant et al. (eds.), *Modeling, Simulation, and Optimization*,
EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-319-70542-2_1

Fig. 1.1 Deterministic textures

Gui et al. [14] proposed a similar method for periodic pattern of texture analysis and synthesis based on texels distribution. Recently, Huang et al. [15] extend it to texture synthesis on arbitrary surfaces. However, these existing techniques mainly focus on texture reproduction, which maintains a visual similar to the original sample. It is observed that user manipulations are rarely provided over the fully automatic synthesis process, including the control of the positions and shapes of texture elements, and consistent transition among different texture elements, which can help synthesize a variety of textures.

In this paper, we first extract distributions of texture elements from a large number of deterministic texture samples by constructing their connectivity. And then these distributions can be divided into near-regular or non-regular categories through quantifying the neighborhood relationships between texture element and its adjacent texture elements in the constructed connectivity. We thus can expand an initial element arrangement in two different ways, by performing local or global growth of texture elements, in order to generate a new larger distribution of texture elements. By default, each texture element in distributions is represented by a discrete point, and the user may also optionally draw a shape contour for it. Once the distribution of texture elements is obtained, we arbitrarily use texture elements to replace each discrete point to synthesize the final textures, which the texture elements are extracted from the input deterministic textures. When pasting texture elements together, we apply a set of deformation operations, such as scaling, rotation, and thin-plate splines (TPS), to change the shapes of texture elements. Such deformation operations used in this paper are helpful, to avoid large overlapping and holes between texture elements, in order to ensure structure consistency. As shown further, our method can create a wide variety of textures as we attempt to grant users more and more control to the positions and shapes of texture elements.

1.2 Element Distribution Creation

1.2.1 Connectivity Construction and Analysis

Given a deterministic texture sample, the main task is to model the spatial neighborhood relationships among all complete texture elements. Note that the incomplete texture element that has the number of pixels on the border can be discarded. Each

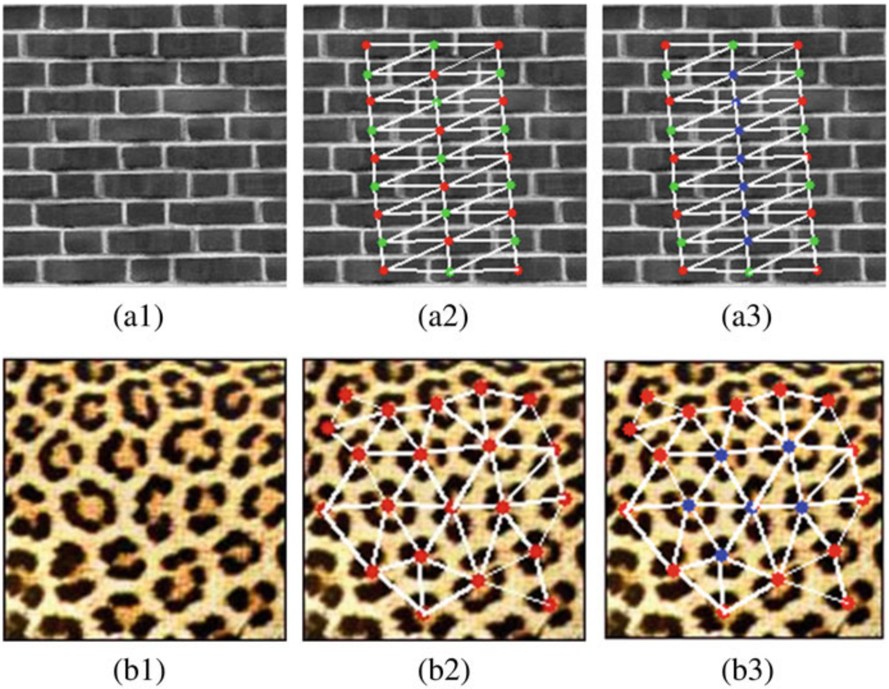

Fig. 1.2 Near-regular and non-regular connectivity

complete texture element in the texture sample can be represented by its bounding boxes, and the centers of these bounding boxes are used to define the element positions in the texture sample. According to the discrete positions information, the most suitable method used in [11, 14] is to extract the Delaunay triangulation in order to get a connectivity among all texture elements, which describes a spatial neighborhood relationships between each texture element and its all neighboring texture elements. The connectivity also can be called the distribution of texture elements. On the other hand, if there are texture elements with different classes in the texture sample, as shown in Fig. 1.2a2, b2, they are marked by using different colored points.

To the constructed connectivity, our focus is on exploring whether they are near-regular or non-regular, which are of primary importance to characterize the spatial arrangements. For this purpose, we first define a neighborhood that is composed of all neighboring texture elements of each texture element. In each neighborhood, edges are connected between two adjacent texture elements. The neighborhood can form a "ring shape" when the size of the neighborhood is equal to the number of edges. As Fig. 1.2a3, b3 shows, the texture elements which having neighborhood with ring shape are marked by blue. Based on these neighborhoods, if they have the same sizes and their ring shapes have more close appearance similarity that can be measured through the area of the neighborhood, the constructed connectivity is near-regular or regular (Fig. 1.2a2), else is supposed to be non-regular (Fig. 1.2b2). Indeed, the key for analyzing the constructed connectivity is to describe the discrete or compact structure information among texture elements, as texture elements in texture samples are independent to each other (Fig. 1.2b1), or define a partitioning of textures, with each texture element having a nonoverlapping but adjoining spatial extent (Fig. 1.2a1).

1.2.2 Local and Global Growth

By considering near-regular or non-regular connectivity, texture design and synthesis begin with it, which we call an initial elements arrangement, and expand it outward by placing the new positions, in order to reproduce a new larger elements distribution. Given a regular elements arrangement (Fig. 1.3a), we need to decompose it into two types of sub-models, including horizontal model (Fig. 1.3a1) and vertical model (Fig. 1.3a2), which can be used as the placement rules

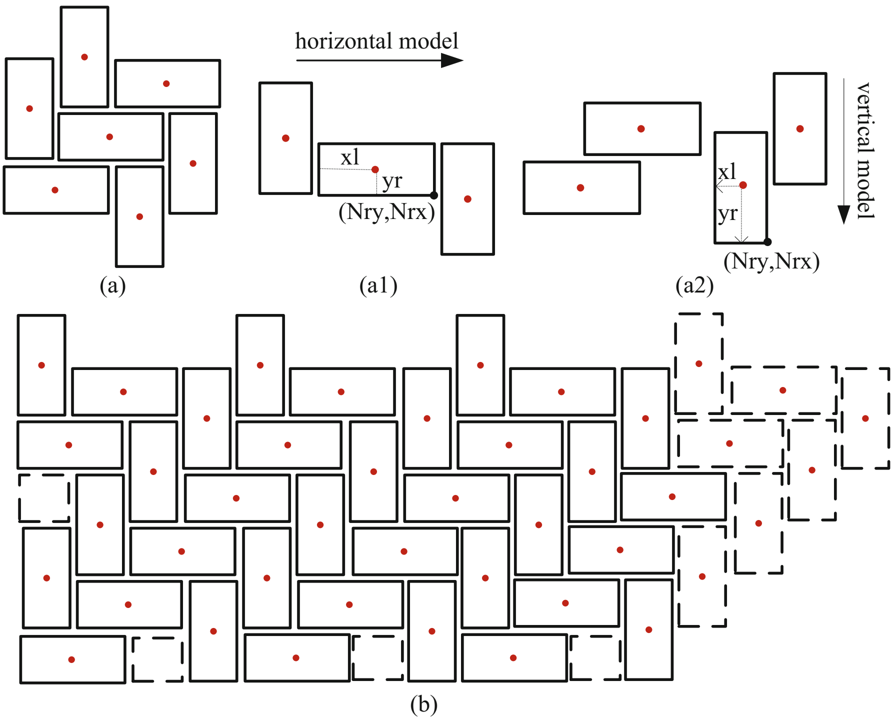

Fig. 1.3 Extension for regular elements distribution

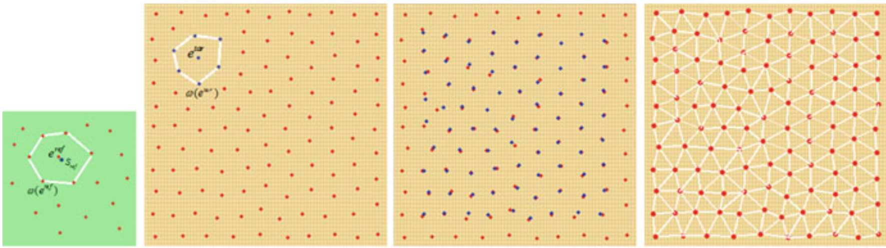

Fig. 1.4 Non-regular elements distribution adjustment

to guide the extension. More specifically, we expand it outward by placing the positions of the texture elements according to the decomposed horizontal model and vertical model, where the positions can be computed based on the relative positions of the placed texture elements. For instance, to perform extension by using horizontal model, the position of a new texture element (P_x, P_y) is computed as $P_x = Nrx + ov + xl$ and $P_y = Nry - yr + 1$, where (Nry, Nrx) is the coordinate of and lower right corner of the bounding box; xl and yr represent the shortest distance from the center to the bounding box (Fig. 1.3a1, a2); ov is a user-specified spacing between adjacent texels to avoid overlapping. Similarly, the position of a texture element in vertical direction can be computed based on vertical model. As shown in Fig. 1.3b, we reproduce the final element distribution through using the horizontal model and the vertical model alternatively. By default, each texture element in the synthesized element distributions is represented by a discrete point, and the user may also optionally draw a shape contour for texture elements. In addition, the class information for each new placed position can be recorded during the extension.

To an initial elements arrangement with non-regular (Fig. 1.4a), we first need to provide a set of discrete points and then use Lloyd's method [10, 11] to obtain a random distribution of positions (Fig. 1.4b). Since the obtained distribution is close to uniformity because each point is located in the center of the Voronoi region, we further perform an adjustment in order to maintain a certain visual resemblance with the initial elements arrangement. Given are a neighborhood $\omega(e_{\text{ref}})$ taken from the initial elements arrangement D_{ref} and a neighborhood $\omega(e_{\text{tar}})$ obtained from the synthesized distribution D_{tar}, where e_{ref} and e_{tar} are the current selected position of the texels (as shown in Fig. 1.4a, b). The adjustment process falls into three stages.

Step 1 We find a matched e_{ref} from D_{ref} for the current selected e_{tar} in the D_{tar}. It is the one whose neighborhoods $w(e_{\text{ref}})$ has the most similar neighborhood condition to that of the selected position e_{tar}. We first sort position point $e_i^{\text{ref}} \in \omega\left(e^{\text{ref}}\right)$ and $e_j^{\text{tar}} \in \omega\left(e^{\text{tar}}\right)$ in counterclockwise order simply, and then the differences can be measured by using the following err function:

$$\mathrm{Err}\left(\omega\left(e^{\mathrm{ref}}\right), \omega\left(e^{\mathrm{tar}}\right)\right) = \sum_{i,j} w_1 \theta\left(e_i^{\mathrm{ref}}, e_j^{\mathrm{tar}}\right) + w_2 L\left(e_i^{\mathrm{ref}}, e_j^{\mathrm{tar}}\right), \quad i,j \in \{1, \ldots, N\} \tag{1.1}$$

where N is the number of texture elements in neighborhoods; $\theta\left(e_i^{\mathrm{ref}}, e_j^{\mathrm{tar}}\right) = \left|\theta_i^{\mathrm{ref}} - \theta_j^{\mathrm{tar}}\right|$ measures differences in angles. θ_i^{ref} is the angle between x-axis and edges $(e_i^{\mathrm{ref}}, e^{\mathrm{ref}})$, similarly to define θ_j^{tar}. $L\left(e_i^{\mathrm{ref}}, e_j^{\mathrm{tar}}\right) = \left|L_i^{\mathrm{ref}} - L_j^{\mathrm{tar}}\right|$ measures differences in length of edges. L_i^{ref} and L_j^{tar} are the length of edges $\left(e^{\mathrm{ref}}, e_i^{\mathrm{ref}}\right)$ and $\left(e^{\mathrm{tar}}, e_j^{\mathrm{tar}}\right)$, respectively. w_1 and w_2 are weights to balance the differences in angles and edge lengths. We can find the best matched e_{ref} when minimizing the error function (Eq. 1.1).

Step 2 We compute the reference shift vector S_{ref} of the corresponding e_{ref}, which is the distance from the position e_{ref} to the barycenter of the neighborhoods $w(e_{\mathrm{ref}})$.

Step 3 We translate the selected position e_{tar} by $S_{\mathrm{tar}} = (S_{\mathrm{ref}} A_{\mathrm{tar}})/(n A_{\mathrm{ref}})$, where A_{tar} and A_{ref} are the area of the neighborhoods $w(e_{\mathrm{tar}})$ and $w(e_{\mathrm{ref}})$, respectively, and n is the size of the neighborhoods $w(e_{\mathrm{tar}})$. The translated positions are marked by using blue (Fig. 1.4c).

When all available positions in D_{tar} have been translated through repeating Step 1 to Step 3, we can obtain the final elements distribution (Fig. 1.4d), having the stochastic property. On the other hand, we can control the density of texture elements in the synthesized elements distribution through increasing and decreasing the number of the initial discrete points set. Note that the class information in such elements distribution can be ignored.

1.3 Texture Elements Placement

In this phase, the main task is to replace each position with a texture element, which can be extracted from the given input texture sample, so that our texture design and synthesis can create a large variety of textures.

1.3.1 Texture Elements Extraction

Existing state-of-the art image cutout techniques, such as Lazy Snapping [16], GrabCut [17], and RepSnapping [18], can be used to extract individual texture elements. However, none of the above segmentation methods can work well for arbitrary texture samples. In this setting, input texture sample is represented by a graph $G = <P, E>$, with each pixel as one node $p \in P$ and pairwise adjacent pixels as edge $<p,q> \in E$. Based on appearance similarity between texture elements, their extraction is to optimize the following energy function:

$$E(f) = \sum_{p \in P} D_p\left(f_p\right) + \sum_{p<q \in N} V_{p,q}\left(f_p, f_q\right) + \sum_{i,j \in H} U'_{i,j}\left(f_i, f_j\right) \tag{1.2}$$

where the data item $D_p(f_p)$ measures the conformity of node p assigned label f_p, the smooth item $V_{p,q}(f_p, f_q)$ is charged for adjacent nodes with different labels, and the repetition energy item $U'_{i,j}\left(f_i, f_j\right)$ measures the labeling smoothness on the similar nodes; N is a neighborhood system, joining adjacent nodes in 4-neighbor (or 8-neighbor); H is an extended neighborhood systems based on the repetitive similarity. Unlike the optimization model in RepSnapping, the repetition energy item $U'_{i,j}\left(f_i, f_j\right)$ can be modified based on a robust appearance similarity δ among the repeated texels, which is defined as the following function:

$$U'_{i,j}\left(f_i, f_j\right) = \mu\, |\, f_i - f_j\, | \cdot \exp\left(-\beta \cdot \delta^2\,(i,j)\right)\,,\ < i,j > \in H \tag{1.3}$$

where μ is a trade-off weight and β is a constant. Inspired by the texture samples with their own pattern feature, the appearance similarity measurement $\delta(i,j)$ in Eq. (1.3) is developed by considering both colors and pattern features:

$$\delta\,(i,j) = \left\| c_i - c_j \right\| + \left\| T_i - T_j \right\| \tag{1.4}$$

where $||c_i - c_j||$ is used to measure the difference in color between nodes i and j; c_i and c_j are the average color of nodes i and j, respectively; $||T_i - T_j||$ describes the difference in pattern feature between nodes i and j; and T_i and T_j are the average feature vector of nodes i and j, respectively, which is computed by using Gabor wavelet transform [19] in multiple scales and orientations. And the values of these differences are normalized in Eq. (1.3). When minimizing the above energy function (Eq. 1.2), the texture elements can be extracted simultaneously by using max-flow min-cut algorithm. As illustrated in Fig. 1.5, only two (red) lines (Fig. 1.5 a1, b1) are drawn to indicate the texture element and another (blue) line to indicate the texture background; all texture elements in the texture samples can be extracted simultaneously (Fig. 1.5a2, b2). Here we only present a brief overview of this method, which are described in detail in [20].

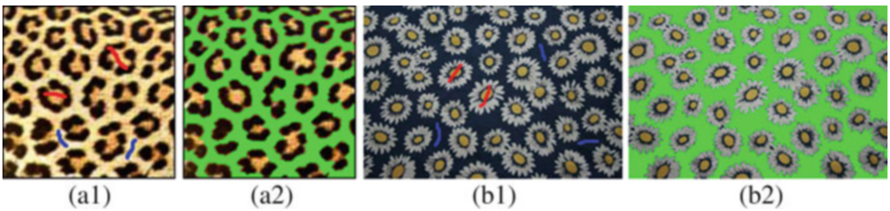

Fig. 1.5 Texture element extraction

1.3.2 Texture Element Deformation

Since the extracted texture elements keep their original shapes, large overlapping and holes between adjacent texture elements can be found when performing texture elements placement, which no guarantee that texture features or structures are continuous in final textures. In order to alleviate such effects, deformation operations are applied to change the shapes of texture elements, especially texture elements with quite irregular shapes. Generally, we observed that it is easy to control texture elements with regular shape, which mainly consists in changing the size and orientation of the texture element by using linear transformation operations, such as scaling and rotation, to keep the structure consistency. Indeed, our motivation is to transform the irregular shapes of texels by using TPS mapping [21]. Given a reference shape and a texture element to be deformation, we first need to sample contour points sets $\{(x_i, y_i)\}$ and $\left\{\left(x_i', y_i'\right)\right\}$, respectively (Fig. 1.6b), and then we determine the corresponding pairs of contour points between them, where i is used to index the sampled contour points.Therefore, the deformation between the reference shape and original counter can be defined as following:

$$x' = f_x\left(x, y\right) = a_1 + a_x x + a_y y + \sum_{i=1}^{n} \alpha_i \phi\left(\left\|(x_i, y_i) - \left(x, y\right)\right\|\right) \tag{1.5}$$

$$y' = f_y\left(x, y\right) = b_1 + b_x x + b_y y + \sum_{i=1}^{n} \beta_i \phi\left(\left\|(x_i, y_i) - \left(x, y\right)\right\|\right) \tag{1.6}$$

where (x, y) represents pixel in deformed texture element. (x', y') represents pixel in the texture element to be deformed. $\phi(r) = r^2 \log r$ is the radial basis function. $\|\cdot\|$ is the Euclidean distance. By using the following two additional constraints:

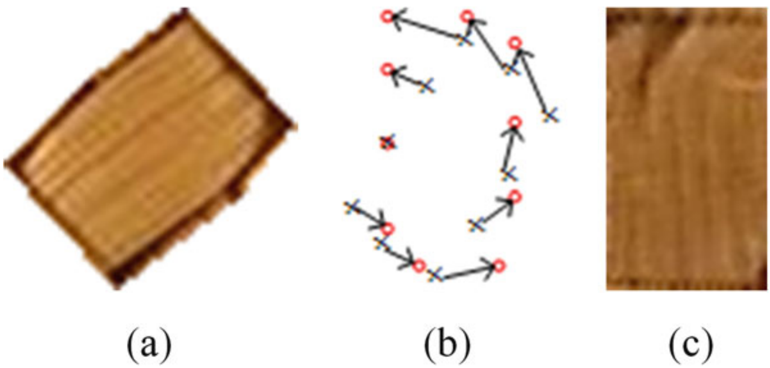

Fig. 1.6 Thin plate spline mapping

$$\sum \alpha_i = \sum \alpha_i x_i = \sum \alpha_i y_i = 0 \tag{1.7}$$

$$\sum \beta_i = \sum \beta_i x_i = \sum \beta_i y_i = 0 \tag{1.8}$$

The parameters a_1, a_x, a_y, α_i, b_1, b_x, b_y, β_i in Eqs. (1.5) and (1.6) can be determined by solving a linear system of equations. Thus, we can deform the region of the texture element according the mapping between $f_x(x, y)$ and $f_y(x, y)$. As Fig. X shows, when the corresponding pairs of contour points are constructed, the texture element (Fig. 1.6a) can be deformed to a rectangular region (Fig. 1.6c). The TPS mapping used in our method not only maintains the shape contour but also creates smooth and consistent warped content in the interior region of texture elements. Note that we use graph-cut-based segmentation method [5] to merge the conflicting regions, and we use example-based completion algorithm [22] to fill holes explicitly.

1.4 Experimental Results and Discussions

Figure 1.7 demonstrates the capability of our method for creating a large variety of textures from a small sample (Fig. 1.7a1–d1). If the initial elements arrangement conforms to element distribution of texture samples, the resulting textures have an appearance similar to the texture samples (Fig. 1.7a2, b2). Otherwise, the designed textures show various appearances (Fig. 1.7a3–a6 and b3–b6). For the last two examples, our presented method changes the elements density interactively, and the elements density increases gradually. In addition, we can interactively replace the initial elements arrangement (Fig. 1.7c6) and manipulate the scale or orientation of texture elements in order to generate various outputs (Fig. 1.7d5, d6).

In Fig. 1.8, we compare our approach with existing techniques of texels distribution-based [14], appearance-space synthesis [7], near-regular texture synthesis [6], graph-cut [5], image quilting [4], patch-based [3], and jump-map based [2]. The texture sample in Fig. 1.8a consists of oblique blocks, which have a near-regular structural layout but irregular color appearance in individual blocks. So far, we have not yet seen an existing texture synthesis algorithm that preserves structural regularity as well as structure continuity in the synthesized texture. The results synthesized by these existing techniques are illustrated in Fig. 1.8c–d. We can find that the quality of the texture produced with patch-based methods [3–7] is superior to that of pixel-based method [2]. However, structure misalignments still remain because there are no exact copies at the overlapping region of adjacent patches. Our interactive controllable technique consists in deforming each single texel into texel with a regular shape by using TPS firstly and then generating a new texture according to a user-provided freehand sketch (Fig. 1.8b1). By giving randomly

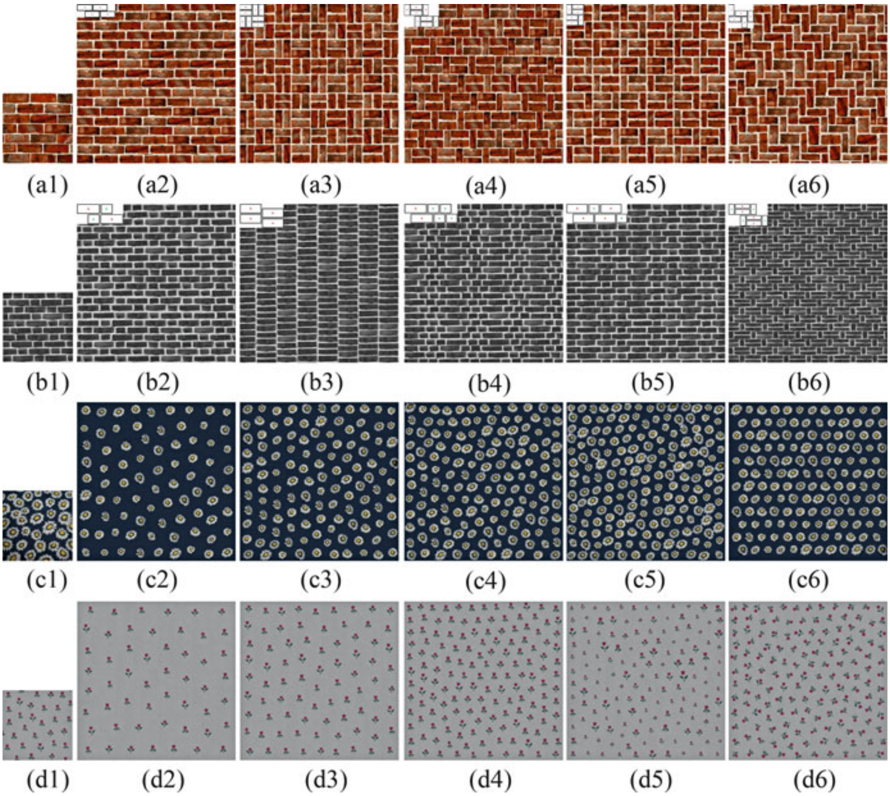

Fig. 1.7 More results of our texture design and synthesis

rotation, the final result can be obtained, as demonstrated in Fig. 1.8b2. From the designed textures, we can see that our method performs better than these existing techniques. The structural regularity and structure continuity are preserved in our designed results.

Figure 1.9 shows two examples in which the element distribution-based texture synthesis [14] fails to obtain an optimized stochastic distribution but which is handled well by our proposed method. In [14], they apply neighborhood comparison for generating a global distribution gradually. However, this method has the difficulty of expanding the distribution of positions infinitely because the error accumulates in each local growth step. As shown in Fig. 1.9a4, the synthesized distribution deviates greatly from the original distribution in samples, which cause large overlapping between adjacent texture elements (Fig. 1.9a3). Although the synthesized distribution is desired as shown in Fig. 1.9b4, the repetition in the resulting large textures (Fig. 1.9b3) is not satisfying because of the randomness. The advantage of our approach is that we obtain the new element distribution from user-provided initial elements arrangement by using Lloyd's method, controlling the

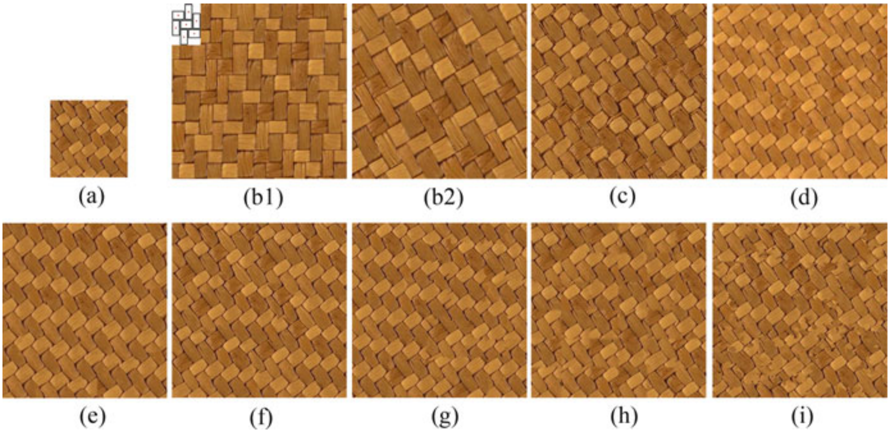

Fig. 1.8 Comparison with existing texture synthesis techniques [2–7, 14]

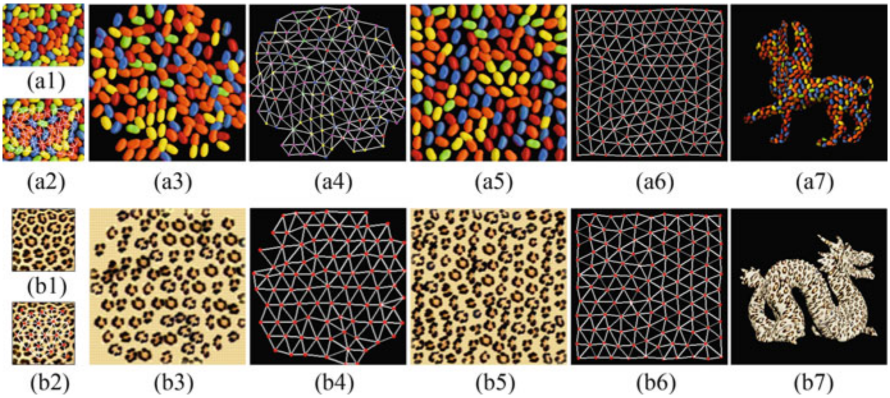

Fig. 1.9 A comparison with element distribution-based texture synthesis [14]

total number of texture elements interactively (Fig. 1.9a6, b6), so that our method produces high-quality synthesis results, as illustrated in Fig. 1.9a5, b5. In addition, we paste the synthesized textures onto arbitrary surfaces to add non-geometric details (Fig. 1.9a7, b7).

Figure 1.10 demonstrates another interesting application of our approach, namely, texture design from different texture samples. As shown in Fig. 1.10c–f, we can use texture elements extracted from different sample textures to create a variety of versatile textures. More specially, we also can provide different initial element arrangement to guide texture design, whereas other techniques usually cannot. This is demonstrated in Fig. 1.10e–g.

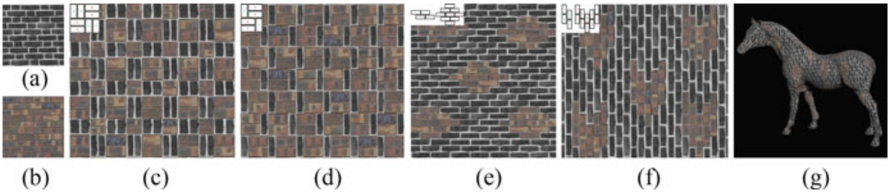

Fig. 1.10 Texture design from different texture samples

1.5 Conclusions

The main contribution of this paper is to introduce an interactive and controllable scheme for semantic texture elements recombination whose success relies on a key factor: a flexible texture synthesis and design procedure. More specifically, more informative element distributions are produced according to user's need and creation. By using all segmented texture elements and various element distributions, our technique can create a wide variety of textures, while the existing texture synthesis methods do not have such ability.

Acknowledgments The project was supported by the National Science Foundation of China (No. 61402053; No. 61602059) and by the Scientific Research Fund of Hunan Provincial Education Department (No. 16C0046; No. 16A008).

References

1. Wei LY, Lefebvre S, Kwatra V, Turk G (2009) State of the art in example-based texture synthesis. In: Eurographics 2009, state of the art report, EG-STAR
2. Zelinka S, Garland M (2002) Towards real-time texture synthesis with the jump map. In: EGRW '02. Proceedings of the 13th eurographics workshop on rendering, pp 99–104
3. Liang L, Liu C, Xu Y, Guo B, Shum HY (2001) Real-time texture synthesis by patch-based sampling. ACM Trans Graph 20(3):127–150
4. Efros A, Freeman W (2001) Image quilting for texture synthesis and transfer. In: SIGGRAPH 2001, computer graphics proceedings, pp 341–346
5. Kwatra V, Schiodl A, Essa I, Turk G, Bobick A (2003) Graphcut textures: image and video synthesis using graph cuts. ACM Trans Graph 22(3):277–286
6. Liu YX, Lin WC, Hays J (2004) Near-regular texture analysis and manipulation. In: SIGGRAPH '04: a SIGGRAPH 2004 papers, pp 368–376
7. Lefebvre S, Hoppe H (2006) Appearance-space texture synthesis. In: SIGGRAPH '06: ACM SIGGRAPH 2006 papers, pp 541–548
8. Kwatra V, Essa I, Bobick A, Kwatra N (2005) Texture optimization for example based synthesis. ACM Trans Graph 24(3):795–802
9. Dischler JM, Maritaud K, Levy B, Ghazanfarpour D (2002) Texture particles. Comput Graph Forum 21:401–410
10. Barla P, Breslav S, Thollot J, Sillion F, Markosian L (2006) Stroke pattern analysis and synthesis. Comput Graph Forum 25:663–671

11. Ijiri T, Mech R, Igarashi T, Miller GSP (2008) An example-based procedural system for element arrangement. Comput Graph Forum 27(2):429–436
12. Hurtut T, Landes PE, Thollot J, Gousseau Y, Drouillhet R, Coeurjolly JF (2009) Appearance-guided synthesis of element arrangements by example. In: Proceedings of the symposium on non-photorealistic animation and rendering, pp 51–60
13. Alves Dos Passos V, Walter M, Sousa M (2010) Sample-based synthesis of illustrative patterns. In: Computer graphics and applications (2010), PG 10, pp 109–116
14. Gui Y, Ma LZ (2010) Periodic pattern of texture analysis and synthesis based on texels distribution. Vis Comput 26(6–8):951–964
15. Huang J, Zhang L, Gui Y (2015) Surfaces texture synthesis based on texel distribution. J Chin Comput Syst (Chinese)
16. Li Y, Sun J, Tang CK, Shum HY (2004) Lazy snapping. ACM Trans Graph 23(3):303–308
17. Rother C, Kolmogorov V, Blake A (2004) Grabcut, interactive foreground extraction using iterated graph cuts. ACM Trans Graph 23(3):309–314
18. Huang H, Zhang L, Zhang HC (2011) RepSnapping: efficient image cutout for repeated scene elements. Comput Graph Forum 30(7):2059–2066
19. Manjunath BS, Ma WY (1996) Texture features for browsing and retrieval of image data. IEEE Trans Pattern Anal Mach Intell 18(8):837–842
20. Yang L, Li F, Gui Y (2015) Repeated texture elements extraction from texture images. J Front Comput Sci Technol 10(8):1154–1165
21. Belongie S, Malik J, Puzicha J (2002) Shape matching and object recognition using shape contexts. IEEE Trans Pattern Anal Mach Intell 24(4):509–522
22. Criminisi A, Prez P, Toyama K (2004) Region filling and object removal by exemplar-based image inpainting. IEEE Trans Image Process 13(1):1200–1212

Chapter 2
Selective Encryption Using Natural Language Processing for Text Data in Mobile Ad Hoc Network

Ajay Kushwaha, Hari Ram Sharma, and Asha Ambhaikar

2.1 Introduction

The world is moving toward wireless network nowadays, and thus ad hoc networks are also acquiring importance. An ad hoc network is defined as a wireless network in which all the nodes are able to communicate with each other directly without the need of a central access point. The performance of ad hoc network is good when less number of nodes are involved, but when the number of nodes increases, the performance gets affected and becomes difficult to manage. The mobile features make the nodes in the ad hoc network moving. As mobile ad hoc networks are widely used nowadays, the security requirements for the network are also increasing which can be provided by means of cryptography.

There may be two ways of keeping information secret: one is hiding the existence of the information and second is making the information unintelligible. Cryptography could be defined as the art and science of making the information secure from unintended audiences by encrypting it and thus making it unintelligible. Conversely, cryptanalysis is the art and science of decrypting the encrypted data. The plain text is converted to cipher text while performing encryption, and the cipher text is converted back to plain text in decryption. This cipher text is unintelligible to others while being transmitted in the network.

The encryption and decryption could be performed by the use of keys. There are two types of key-based encryption, symmetric and asymmetric algorithms. In case of symmetric algorithms, the key is same for both encryption and decryption, while asymmetric algorithms possess different keys. Symmetric algorithms may have stream ciphers and block ciphers. Stream ciphers encrypt single bit of plain

A. Kushwaha (✉) • H.R. Sharma • A. Ambhaikar
Rungta College of Engineering and Technology, Bhilai, Chhattisgarh, India
e-mail: ajay.kushwaha@rungta.ac.in; hrsharma44@gmail.com; dr.asha.ambhaikar@rungta.ac.in

P. Vasant et al. (eds.), *Modeling, Simulation, and Optimization*,
EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-319-70542-2_2

text at a time, whereas block ciphers encrypt a number of bits of plain text as a single unit. The key here is called secret key used in both sender's and receiver's end. One of the examples of symmetric algorithm is DES.

In asymmetric algorithm, the public key is available at both ends, while private key is available at only one side. When data is encrypted by public key, it can be decrypted by only private key and vice versa. The algorithm also called as public key cryptography provides the fit of authenticating the source as a means of digital signature. An example for asymmetric algorithms is RSA.

Natural language processing is related to the field of computer science, artificial intelligence, and computational linguistics concerned with computers and human languages. Natural language processing is related to the area of interaction between humans and computers. The biggest challenges in natural language processing are natural language understanding, word processing, information management, and enabling computers to obtain meaning from humans.

Selective encryption algorithms are popular in the current scenario due to the fact that they may reduce the overhead spent on data encryption/decryption and thus improve the efficiency of the network. This whole task is performed with the help of NLP. The approach removes the stop words from the messages and encrypts the significant data only, prior to sending over the network. The stop words are those words which are filtered out prior to or after from natural language text. They are common words which would to be of little value to the messages.

2.1.1 Challenges Faced by MANET

Distributed As the network is distributed, no central entity will be present, which makes the overall control over the network difficult [11].

Routing The changes in the network topology, protocols demanded as reactive rather than proactive, multicast routing, and multi-hop routes increase the challenge in routing. Routing in mobile condition results in link changes, increase in updates, and non-convergence of routing loop.

Security and Reliability This includes the need for different schemes of authentication and key management due to distributed environment reliability problem in wireless connection, etc.

Supporting Channel Access This includes no fixed base station because of distributed environment, difficulties in avoiding packet collisions, etc.

Dealing with Mobility Mobility affects signal transmission, multicasting, applications, routing, and channel access.

Power Management Techniques All wireless activities consume power which reduces the battery rapidly.

Location-Aided Routing This means, if the associated regions are known with the help of positioning information, routing process will be reduced as it will be spatially oriented.

Internetworking Internetworking here means connection between ad hoc network (i.e., MANET) and fixed networks. Dealing with such heterogeneous system is a challenge.

Frequent Network Partitions This disturbs the whole communication process and requires re-setup of network for further proceedings.

Quality of Service (QoS) Quality of services to be present in such dynamic environment is difficult to provide and does not possess fixed guarantee.

Utilizing Bandwidth Efficiently In wireless network, the bandwidth is limited, and utilizing it in the constantly changing environment is a challenge.

Changing Topology of the Network This includes challenges to routing protocols to be followed efficiently.

2.1.2 MANET Susceptibility to Attacks

1. Absence of centralized server for controlling and management of highly dynamic and big ad hoc network [11].
2. The scalability of network changing dynamically makes it difficult to maintain security.
3. As the nodes are supposed to be working in cooperation, any malicious node can take it as benefit for attacking.
4. Constantly changing network topology also incurs threats.
5. Lack of finite boundaries is in MANET.

The paper is organized in the following way: literature review is provided in Sect. 2.2. Section 2.3 gives the concept of selective encryption. Proposed method SSDE is introduced in Sect. 2.4, which is followed by the result analysis in Sect. 2.5. The paper is concluded in Sect. 2.6.

2.2 Literature Review

Yonglin et al. [1] present a probabilistic selective encryption algorithm which utilizes the advantages of the probabilistic methodology that aims to acquire additional uncertainty in text.

Matin et al. [2] examine the performance of the new cipher in MANET and wireless LAN networks and make a performance comparison with that of AES. In the paper, they focused on the security that is provided at the application level.

As the key size of the algorithm is larger, the time required to break an encryption scheme becomes so excessive that undesirable attacks are meaningless.

Shivendra and Aniruddha [3] propose and implement a combined approach for identification of a given unknown sample of cipher text. In the first part of the system, cipher text samples are generated randomly using different cipher algorithms. In the second part, the system analyzes the sample through (a) block length/stream detection, (b) entropy/reoccurrence analysis, (c) dictionary, and decision tree-based approach.

Zhou and Tang [4] proposed an implementation of a complete and practical RSA encrypt/decrypt solution based on the study of RSA public key algorithm. In addition, the encrypt procedure and code implementation are provided in detail.

Umaparvathi and Varughese [5] present a comparison of the most commonly used symmetric encryption algorithms AES (Rijndael), DES, 3DES, and Blowfish in terms of power consumption. A comparison has been conducted for those encryption algorithms at different data types like text, image, audio, and video. Experimental results are given to demonstrate the effectiveness of each algorithm.

Chang et al. [6] present a powerful and versatile security suite for the AODV (Ad hoc On-Demand Distance Vector) routing protocol. It offers coverage on common security aspects such as encryption and authentication, and it can be easily modified to work with any distance-vector-based routing for MANET (mobile ad hoc networks). The suite utilizes powerful authentication and user-adjustable encryptions based on digital certificate chaining and popular ciphers such as DES, AES, and RSA.

Nawneet et al. [7] concluded a comprehensive summary which discussed the vulnerabilities, challenges, and security attacks on ad hoc routing protocols which leads to difficulties in designing and development of a secure routing protocol and is a challenging task for researcher in an open and distributed communication environments.

The security mechanisms presented in this paper by Michiardi and Molva [8] are a practical response to specific problems that arise at a particular layer of the network stack. However, the proposed solutions only cover a subset of all possible threats and are difficult to integrate with each other.

Suresh et al. [9] study the major attack types that MANET faces and the security goals to be achieved. This paper gives out a brief survey of major security protocols with their relative comparison.

2.3 Concept of Selective Encryption

Selective encryption algorithms are popular in the current scenario due to the fact that they may reduce the overhead spent on data encryption/decryption and thus improve the efficiency of the network. In this section, we have proposed an algorithm for selective encryption and some commonly used techniques of the selective encryption.

The purpose of selective encryption algorithms is to encrypt only certain portions of the messages and provide trustworthy safety so as to secure the transmitted message confidentiality. Selective encryption is proficient to improve the scalability of data transmission and also reduces the processing time. NLP is used for selective encryption of messages. Natural Language Toolkit (NLTK) 3.0 with python 3.5 version is used for analyzing messages under this proposed method [10].

2.3.1 Steps of Selective Encryption While Processing Messages

1. Removing special characters from the messages like *$&?, etc.
2. Process of tokenization in which it extracts all the words present in the messages.
3. Dropping stop words (common words) and collecting significant data (keywords) from the messages.
4. All key words are encrypted, and the rest of the common words are sent as it is onto the network.

In order to understand the steps of selective encryption, a code along with an example is given in Fig. 2.1a, b.

In case of selective encryption algorithms, there is involvement of uncertainty in the message encryption process while determining the uncertain pattern of encrypted messages. Thus, uncertainty may enhance the security of data transmission, since all messages are assumed to have equal importance. Thus, uncertainty becomes one of the principle factors while designing a selective encryption-based cryptosystem. Usually, the more is the uncertainty involved, the more is the cryptosystem effective.

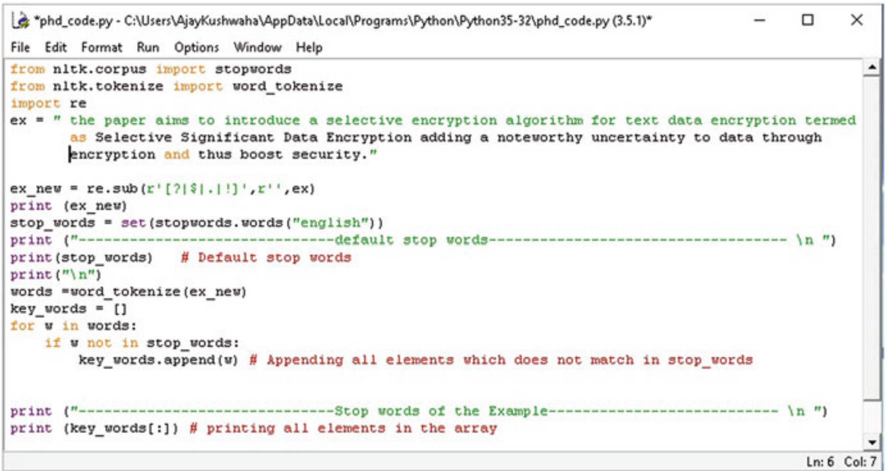

Fig. 2.1 (**a**) Python code along with NLTK for word processing. (**b**) Example of word processing

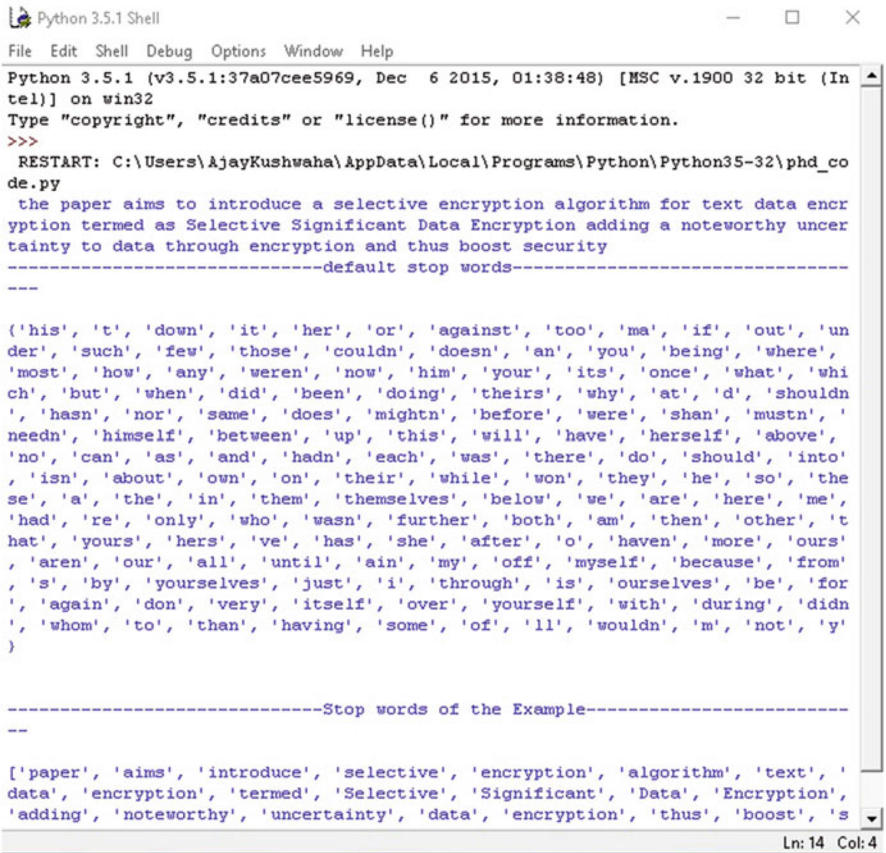

Fig. 2.1 (continued)

At the present time, selective encryption algorithms are mainly applied in the energy-aware environments or large-scale data transmission, such as wireless sensor networks (WSNs), mobile ad hoc networks (MANETs), multimedia communications, etc. In a WSN, each device uses battery as its power supply and, therefore, has inhibited computational ability, so it is difficult for a sensor to spend too much computational cost on data encryption and decryption. Under these circumstances, the design of a selective encryption algorithm with less processing time but with comparatively high security level is enormously significant. Multimedia communication often requires real-time data transmission, so large amount of audio and video data need to be transmitted securely. If all multimedia data are encrypted, this will create large amount of overhead, so multimedia data is difficult to transmit timely and the quality of communication cannot be guaranteed.

2.3.2 Full Data Encryption

In full data encryption, whole data that is to be sent over the network is encrypted before transmitted to the receiver side.

2.3.3 Toss-a-Coin Method

This method is a form of selective encryption, in which the whole message which is to be transmitted is divided into two groups – even and odd – and from the starting of the message, each odd word belongs to odd group and each even word belongs to the even group. The uncertainty involved here is which group will be encrypted, i.e., even or odd is not known. As only one group is encrypted, it makes the encryption selective. Now which group will be encrypted is decided by tossing a coin. Here only 50% of data is encrypted; thus, not much data is reduced, and also involvement of uncertainly is less.

2.4 Selective Significant Data Encryption (SSDE)

The approach selects the significant data there in the message and encrypts them prior to sending over the network. Significant data implies the keyword that holds the meaning of entire message. Excluding significant ones, the rest commonly used words like articles, pronouns, conjunctions, prepositions, and interjections are sent without encoding. The flowchart of the proposed method (i.e., SSDE) is given in Fig. 2.2 which shows the execution of SSDE algorithm.

Proposed algorithm
Step1. Input messages
Step2. Remove special characters from the messages like *$&?, etc.
Step3. Process of tokenization in which it extracts all the words present in the messages
Step4. Dropping stop words (common words) and collecting significant data (keywords) from the messages using NLTK package in python
Step5. Encrypt significant data (keywords) using Blowfish algorithm, and send the message to network
Step6. Send the stop words to network without encryption

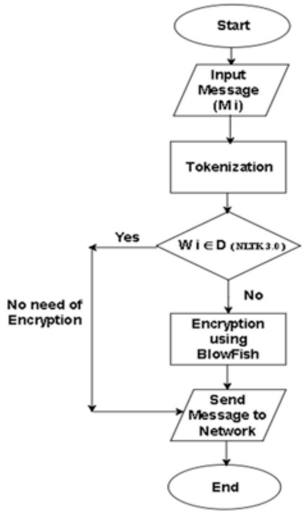

Fig. 2.2 Flow chart of selective significant data encryption

Table 2.1 Performance metrics

Performance metrics	
Encryption time	Time taken to encrypt plain text into cipher text
Decryption time	Time taken to decrypt cipher text into plain text
End-to-end delay	Time taken to transfer packets from source to destination
Battery consumption	Power consumed during transmission
Throughput	How much encrypted data can be transferred from one location to another in a given amount of time
Residual battery	Remaining amount of battery power after transmission

2.5 Result Analysis

In order to observe the characteristics of SSDE, we carried out an extensive set of experiments within a wireless environment. The experimental setup is done using Red Hat 6.0 32-bit operating system and NS 2.34, with Intel(R) core(TM) i3 processor, CPU M 480 at 2.67 GHz and 2.66 GHz, and 4 GB installed RAM. In this work, the proposed method SSDE is compared with commonly used techniques, that is, full data encryption and toss-a-coin method. Figure 2.3 illustrates how nodes are communicating to each other in ad hoc network. Each experiment is run for 50 ns of simulation time. During the simulation experiment, the compared systems are all run under the identical scenario. The performance metrics for evaluating the SSDE are encryption time, decryption time, battery consumption, end-to-end delay, and throughput as shown in table 2.1.

The symmetric key encryption algorithms AES, DES, and Blowfish are implemented using Red Hat 6.0 32-bit operating system and NS 2.34 with packet size for text data to be 512 bytes.

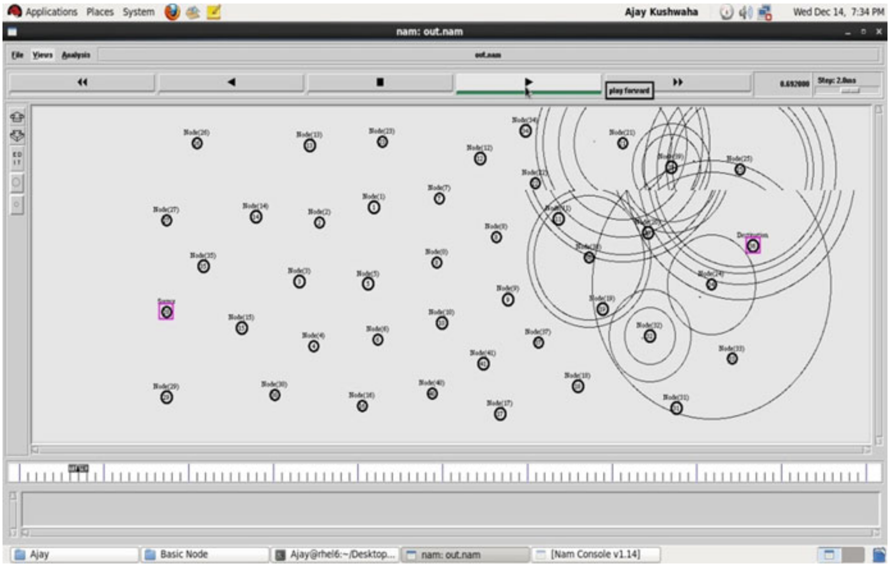

Fig. 2.3 Basic mobile nodes in MANET

Table 2.2 Encryption time (ms)

Encryption algorithm	Text (512 bytes)
DES	434
AES	214
Blowfish	262

Table 2.3 Decryption time (ms)

Decryption algorithm	Text (512 bytes)
DES	451
AES	221
Blowfish	253

Table 2.4 Throughputs (kbps)

Throughput	Text (512 bytes)
DES	28.13
AES	5.27
Blowfish	35.20

Table 2.2 shows the encryption time in milliseconds of DES, AES, and Blowfish encryption algorithms for text file.

Table 2.3 shows the decryption time in milliseconds of DES, AES, and Blowfish encryption algorithms for text file.

Table 2.4 shows the throughput in kbps of DES, AES, and Blowfish encryption algorithms for text file.

Table 2.5 shows the battery consumption in joules of DES, AES, and Blowfish encryption algorithms for text file.

Table 2.5 Battery consumption (joules)

Battery consumption	Text (512 bytes)
DES	83.087
AES	72.087
Blowfish	85.544

Table 2.6 End-to-end delay (ms)

End-to-end delay	Text (512 bytes)
DES	48.7766
AES	49.1367
Blowfish	48.7349

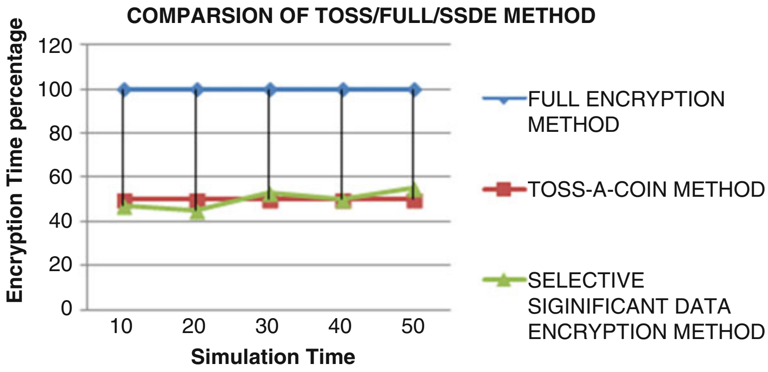

Fig. 2.4 Encryption time percentage vs. simulation time

Table 2.6 shows the end-to-end delay of DES, AES, and Blowfish encryption algorithms for text and image files.

The experimental result shows that Blowfish performs better when all performance metrics are considered. So, we have used Blowfish algorithm for our proposed method.

As stated earlier, two approaches are used as the comparable models with our proposed system. The first approach encrypts all messages without leaving any text unencrypted and thus termed as full data encryption. In the second approach, half of the data is encrypted and is termed as toss-a-coin method.

Figures 2.4 and 2.5 represent the comparison of encryption, decryption, and simulation time based on three approaches. Figure 2.4 shows that both toss-a-coin and SSDE have an obvious lower encryption time than full decryption. This advantage is because of selective encryption which reduced the overhead. In Fig. 2.5, the decryption time in full data encryption is more as compared to both toss-a-coin and SSDE, and thus selective encryption is superior to full data encryption for utilization of resources in the network.

Figure 2.6 compares the battery consumption of toss-a-coin, full data encryption, and SSDE, respectively. Figure 2.6 displays that SSDE has lower battery consumption than full data encryption method and more than toss-a-coin. As it is difficult

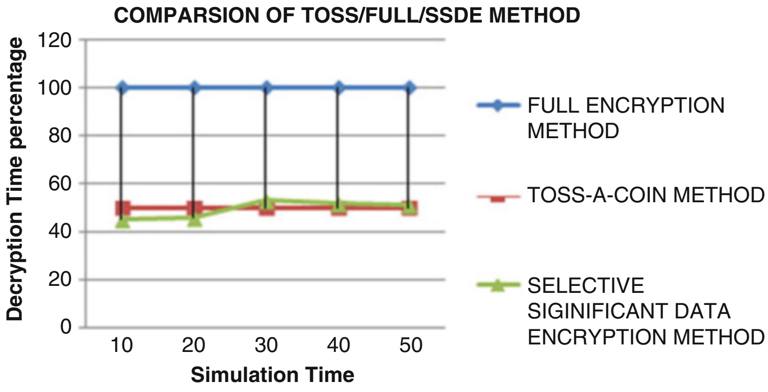

Fig. 2.5 Decryption time percentage vs. simulation time

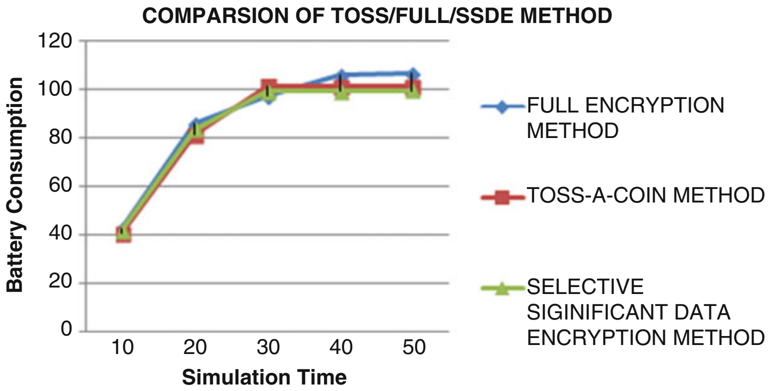

Fig. 2.6 Battery consumption vs. simulation time

to identify what parts of messages are encrypted in SSDE thus gives an added advantage. It is evident that SSDE is more efficient and time-saving when compared with full data encryption and toss-a-coin method in all aspects like encryption, security, etc.

2.6 Conclusion

This paper introduces a better solution for data encryption in wireless networks. The approach is based on selective encryption, which is one of the most promising solutions nowadays to reduce cost of data protection as well as providing sufficient uncertainty for reliability and improved data security. The performance of the method is evaluated based on the extensive set of experiments. The results demonstrate the effectiveness of SSDE over other methods in wireless networks. Thus, the provided solution gives a feasible solution for secure wireless communication in mobile ad hoc network. This method can be used in social chatting apps, military security, corporate world communication, and government activities involving text data encryption. This method can be used for text data only. In the future, this method can be extended for other file formats (i.e., audio, video, etc.).

References

1. Boukerche A, Mokdad L, Ren Y (2011) Performance analysis of a selective encryption algorithm for wireless ad hoc networks. In: Wireless communications and networking conference, IEEE, pp 1038–1043
2. Matin MA, Hossain MM, Islam MF, Islam MN, Hossain MM (2009) Performance evaluation of symmetric encryption algorithm in MANET and WLAN. In: International conference for Technical Postgraduates (TECHPOS), IEEE, Kuala Lumpur, pp. 1–4
3. Mishra S, Bhattacharjya A (2013) Pattern analysis of cipher text: a combined approach. In: International Conference on Recent Trends in Information Technology (ICRTIT), IEEE, pp 393–398
4. Zhou X, Tang X (2011) Research and mplementation of RSA algorithm for encryption and decryption. In: International Conference on Strategic Technology (IFOST), IEEE, pp 1118–1121
5. Umaparvathi M, Varughese DK (2010) Evaluation of symmetric encryption algorithms for MANETs. In: International Conference on Computational Intelligence and Computing Research (ICCIC), IEEE, Coimbatore, pp 1–3
6. Chang JT, Gundala S, Moh T-S, Moh M (2009) VESS – a versatile extensible security suite for MANET routing. In: Pacific Rim conference on communications, computers and signal processing, IEEE, Victoria, pp 944–950
7. Raj N, Bharti P, Thakur S (2015) Vulnerabilities, challenges and threats in securing mobile ad-hoc network. In: Fifth international conference on communication systems and network technologies, IEEE, pp 771–775
8. Michiardi P, Molva R (2004) Ad hoc networks security. Moblie Ad Hoc Networking. Wiley-IEEE Press, Hoboken 12:329–354
9. Kumar S, Pruthi G, Yadav A, Singla M (2012) Security protocols in MANETs. In: Second international conference on advanced computing & communication technologies, IEEE, pp 530–534
10. Bird S, Loper E, Klein E (2009) Natural language processing with python. O'Reilly Media Inc, United States of America
11. Parmar V, Goyal RRP (2011) MANET : vulnerabilities, challenges, attacks, application. Int J Comput Eng Manag 11:32–37

Chapter 3
A Novel Hybrid Artificial Bee Colony with Monarch Butterfly Optimization for Global Optimization Problems

Waheed Ali H.M. Ghanem and Aman Jantan

3.1 Introduction

There are a lot of problems in the real world that involve a set of potential solutions, from which the one with the best quality is termed as the optimal solution, and the method of searching for such a solution is known as mathematical optimization. The quality of solutions is represented by the ability to maximize or minimize a certain function, called the objective function, while the pool of possible solutions that can satisfy the required objective is called the search space. One can traverse all possible solutions, examine the result of the objective function in each case, and select the best solution. However, many real problems are intractable using this exhaustive search strategy. In these problems, the search space expands exponentially with the input size, and exact optimization algorithms are impractical. The historical alternative in such situations is to resort to heuristics, similar to simple rules of thumb that humans would utilize in a search process. Heuristic algorithms implement such heuristics to explore the otherwise prohibitively large search space, but they do not guarantee finding the actual optimal solution, since not all areas of the space are examined. However, a close solution to the optimal is returned, which is "good enough" for the problem at hand.

W.A.H. M. Ghanem (✉)
School of Computer Science, Universiti Sains Malaysia, Gelugor, Penang, Malaysia

Faculty of Education-Saber, University of Aden, Aden, Yemen

Faculty of Engineering, University of Aden, Aden, Yemen
e-mail: waheed.ghanem@gmail.com

A. Jantan
School of Computer Science, Universiti Sains Malaysia, Gelugor, Penang, Malaysia
e-mail: aman@cs.usm.my

P. Vasant et al. (eds.), *Modeling, Simulation, and Optimization*,
EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-319-70542-2_3

The next step would be to generalize those heuristics in higher-level algorithmic frameworks that are problem independent and that provide strategies to develop heuristic optimization algorithms. The latter are known as metaheuristics [1]. Early metaheuristics were based on the concept of evolution, where the best solutions among a set of candidate solutions are selected in successive iterations, and new solution are generated by applying genetic operators such as crossover and mutation to the parent solutions.

Similar to and including evolutionary algorithms, many metaheuristics were based on a metaphor, inspired by some physical or biological processes. Many recent metaheuristics mimic the biological swarms in performing their activities, in particular, the important tasks of foraging, preying, and migration. Popular examples of developed metaheuristic algorithms in this category include particle swarm optimization (PSO) [2], which is inspired by the movement of swarms of birds or fishes; ant colony optimization (ACO) [3, 4], which is inspired by the foraging behavior of ants, where ants looking for food sources in parallel employ the concept of pheromone to indicate the quality of the found solutions; and artificial bee colony (ABC) algorithm, inspired by the intelligent foraging behavior of honeybees [5, 6].

The idea of deriving metaheuristics from natural-based metaphors proved so appealing that much more of such algorithms have been and continue to be developed. A few more examples include cuckoo search (CS) [7, 8], biogeography-based optimization (BBO) [9], animal migration optimization (AMO) [10], chicken swarm optimization (CSO) [11], grey wolf optimization (GWO) [12], krill herd (KH) [13], and monarch butterfly optimization (MBO) [14], which is inspired by the migration behavior of monarch butterfly. The bat algorithm (BA) [15] also belongs to the metaheuristics that are based on animal behavior, inspired by the echolocation behavior of bats in nature. On the other hand, several metaphor-based metaheuristics are derived from physical phenomena such as simulated annealing (SA) [16] which is inspired by the annealing process of a crystalline solid.

The aforementioned metaheuristics are classified as stochastic optimization techniques. To avoid searching the whole solution space, they include a randomization component to explore new solution areas. Though these random operators are essential, they can introduce two types of problems. First, if the randomization is too strong, the metaheuristic algorithm might keep moving between candidate solutions, loosely examining each localized region and failing to exploit promising solutions and find the best solution. Second, if the search process is too localized, exploiting the first found good solutions very well but failing to explore more regions, the algorithm might indeed miss the real optimal solution (called the global optimum) and trap into some local optima.

The perfect balance between exploitation and exploration is essential to all metaheuristics. In fact, it is whether and how this balance is achieved that distinguishes most metaheuristics from each other and forms a source of new attempts to improve existing algorithms, possibly by hybridizing ideas from more than one metaheuristic strategy [18].

In this paper, we follow this path and introduce a new hybrid metaheuristic that augments the popular ABC algorithm with a feature from the MBO algorithm so as to make the correct balance between randomization of local search and global search.

The rest of this article is organized as follows. Section 3.2 describes the proposed HAM method, while Sect. 3.3 explains the setup of experimental evaluation. Section 3.4 presents and discusses the obtained results, and finally Sect. 3.5 concludes the paper.

3.2 Proposed HAM Algorithm

This section introduces the HAM algorithm, which is based on the standard artificial bee colony algorithm [5, 6] and monarch butterfly optimization algorithm [14]. The ABC algorithm was proposed by Karaboga for optimizing numerical problems in 2005, and several developments were based on this algorithm [19–21]. The MBO algorithm was proposed by Gai-Ge, Suash, and Zhihua in 2015. It is a new nature-inspired metaheuristic optimization algorithm that works by simplifying and idealizing the migration behavior of monarch butterfly individuals between two distinct lands, namely, northern USA (Land1) and southern Canada (Land2). For more details about the two algorithms, please refer to [5, 14].

The exploitation and exploration concepts are undoubtedly considered exceedingly important characteristics in metaheuristic algorithms. In fact, the best metaheuristic algorithm is the one that strikes a balance between these two mechanisms, as a consequence of enhancing the solving of (low- and high-dimensional) optimization problems. The mechanism of exploitation is based on current knowledge to seek better solutions, while the mechanism of exploration is based on searching the entire area of the problem for an optimal solution.

Particularly, by analyzing the standard MBO algorithm, it could be noticed it has an effective capability of exploring the search space; nevertheless, it does possess a weak ability to exploit the search space due to the intermittent use of Levy flight by the updating operators which in turn drives the algorithm to large random steps or moves. On the other hand, the ABC algorithm has the capability of exploring the search space, as well as it has a decent capability in finding the local optima through the employee and onlooker phases. So these two phases in ABC algorithm classed as a local search process. The ABC algorithm is mostly dependent on selecting the solutions that improve the local search. While the global search in the ABC algorithm is implemented through the scout phase, which leads to reducing the speed convergence during the search process.

The main idea of the hybrid proposed algorithm is based on two ameliorations: firstly, the main objective in modifying the MBO algorithm improves the exploitation versus exploration balance, by modifying the butterfly adjusting operator in order to increase the search diversity and balance the insufficiency of ABC algorithm in global search efficacy. Algorithm 1 shows the amended version of the operator. The second enhancement is done by replacing the modified butterfly adjusting operator of MBO with the employee phase of ABC algorithm. The enhanced operator is called the "employee bee adjusting operator," and the resulting modified phase is called the "employee bee adjusting phase."

The main objective of the employee bee adjusting phase is to update all the solutions in the bee population, whereas each solution is a D-dimensional vector. While the initialization phase is used to define all the variables that would be defined in the standard ABC algorithm and assign them suitable values. Although the HAM algorithm is essentially founded on all the parameters of the original ABC algorithm, it uses three new control variables: *limit1*, *limit2*, and the maximum walk step variable $S_{\max}$; these three variables are used in the employee bee adjusting phase.

Algorithm 1: Employee bee adjusting phase

```
Begin
For i = 1 to SN do
Calculate the walk step dx by Equation (1);
Calculate the weighting factor by Equation (2);
For j = 1 to D do
If rand ≥ limit1 then
Generate the j^th element by Equation (3);
Else
Randomly select a food Source (r) by Equation (4);
If rand < limit2 then
Generate the j^th element by Equation (5);
Else
Generate the j^th element by Equation (6);
If rand < BAR then
Generate the j^th element by Equation (7);
End if
End if
End if
End for j
Evaluate the fitness value of the candidate solution x_i.
Apply a greedy selection process between x_i and x_best
If solution x_i does not improve, trial_i = trial_i + 1,
Otherwise trial_i = 0.
End for i
End
```

In Algorithm 1 above, each employee bee of the employee bee adjusting phase has been assigned to an independent food source whereby it generates a new solution either by a new mutation operators or through Levy flight. The mutation operators are based on the two control variables: *limit1* and *limit2*. The focal point of *limt1* and *limit2* is to fine-tune the exportation versus exploitation through improving the global search diversity. As shown in Algorithm 1, the first step is to use the Levy flight to compute a walking step "*dx*" for the i^{th} bee by Eq. 3.1, and then it uses the Eq. 3.2 to compute the weighting factor "∝," where *t* is the current generation and $S_{\max}$ represents the max walk step that a bee individual can move in one step. Then, the Algorithm 1 uses Eq. 3.3 to update the solution element, when (rand ≥ *limit1*), for each element *j* of the D dimensions.

$$dx_k = \text{levy}\left(x_j^t\right) \tag{3.1}$$

$$\propto \ = S_{\max}/t^2 \tag{3.2}$$

$$x_{i,j}^{t+1} = x_{\text{best},j}^t \tag{3.3}$$

$x_{i,j}^{t+1}$ represents the location of the solution *i,* the j^{th} element of solution x_i at generation $t + 1$, while $x_{\text{best},j}^t$ represents the best location among the food sources, the j^{th} element of x_{best} at generation *t*, so far with respect to the i^{th} bee. On the other hand, if (rand < *limit1*) then another set of updates are performed. First, a random food source (equivalent to a random solution or bee) is selected from the current population using Eq. 3.4. Then, depending on whether a randomly generated value is smaller than *limit2*, Eq. 3.5 is used to update the solution elements, as follows:

$$r = \text{round}\left(\left(\text{SN}^* \text{ rand}\right) + 0.5\right) \tag{3.4}$$

$$x_{i,j}^{t+1} = x_{r,j}^t + 0.5^*\text{rand}^* \left(x_{\text{worst},j}^t - x_{r2,j}^t - x_{\text{best},j}^t\right) \tag{3.5}$$

$x_{i,j}^{t+1}$ represents the location of the solution *i*, the j^{th} element of solution x_i at generation $t + 1$. $x_{\text{best},j}^t$ represents the best location among the food sources, the j^{th} element of x_{best} at generation *t*. And $x_{\text{worst},j}^t$ signifies the worst location among the food sources, the j^{th} element of x_{worst} at generation *t*. $x_{r,j}^t$ represents the location of the solution *r* calculated by Eq. 3.4, the j^{th} element of x_r at generation *t*. The *t* in Eq. 3.5 is the current generation number.

On the other hand, if the randomly generated value was bigger than *limit2*, the solution elements are updated by Eq. 3.6, where $x_{i,j}^{t+1}$ is the j^{th} element of solution x_i at generation $t + 1$, which represents the location of the solution *i*; $x_{\text{best},j}^t$ is the j^{th} element of x_{best} at generation *t*, which represents the best location among the food sources so far; $x_{\text{worst},j}^t$ is the j^{th} element of x_{worst} at generation *t*, which represents the worst location among the food sources so far, while $x_{r,j}^t$ is the j^{th} element of x_r at generation *t*, which represents the location of the solution *r* calculated by Eq. 3.4.

$$x_{i,j}^{t+1} = x_{r,j}^t + 0.5^*\text{rand}^* \left(x_{\text{best},j}^t - x_{r3,j}^t - x_{\text{worst},j}^t\right) \tag{3.6}$$

The HAM algorithm also used the Levy flight function but with a smaller probability of execution to reduce its impact on the exploitation process. Assuming the execution path has already passed the tests of limit1 and limit2 control variables, then the algorithm performs another random check against the BAR parameter, right after the update by Eq. 3.6 to further change the value of $x_{i,j}^{t+1}$ occasionally by the amount $\propto \times (dx_k - 0.5)$, as per Eq. 3.7.

$$x_{i,j}^{t+1} = x_{i,j}^{t+1} + \propto \times (dx_k - 0.5) \tag{3.7}$$

Finally, the employee bee adjusting phase tests the limits of the newly created solution to make sure it is within the permissible limits for the optimization problem, and after that it evaluates the fitness value that is produced by the new solution and uses the greedy selection process between the best and the new solutions to select the better one. In the case that the resulted solutions are not improved, then a trial counter is increased by one. The HAM algorithm relies on the implementation of the original ABC algorithm without any change, which can be found in [22].

3.3 Experimental Evaluation

In this section, we lay out the experimental setup through which we have evaluated the proposed algorithm, HAM.

3.3.1 General Setup

3.3.1.1 Hardware and Software Implementation

All the experiments have been conducted on a laptop with an Intel Core i5 2.4 GHz processor and RAM 8 GB. The proposed HAM algorithm is a software implementation based on the implementation of ABC and MBO. The software implementation tests were carried out in MATLAB R2009b (V7.9.0.529) on a windows 7 box.

3.3.1.2 Parameters

For fair comparison purposes, we set all common control parameters to the same values. This includes mounting the dimensionality of search space to 10 and the population size to 50 for all methods. And here below, we present the parameters for all methods in this work.

The control variables that have been set across all experiments of HAM algorithm are as follows: *limit1* is set to 0.8, *limit2* is set to 0.5, migration period Per_i is set to 1.2, migration ratio p is set to 0.4167, and finally $S_{\max}$ is set to 1.0. Moreover, the ABC algorithm has had the following parameter settings: limit was set to 100, and the colony size was set to 100, employed bees = 50 and onlooker bees = 50. Finally,

Table 3.1 Benchmark functions used for evaluating the proposed algorithm

No.	Name	Equation	Low	Up
1	Sphere	$f(x) = \sum_{i=1}^{n} x_i^2$	−100	100
2	Schwefel 2.22	$f(x) = \sum_{i=1}^{n} \lvert x_i \rvert + \prod_{i=1}^{n} \lvert x_i \rvert$	−1.28	1.28
3	Schwefel 1.2	$f(x) = \sum_{i=1}^{n} \left(\sum_{j=1}^{i} x_j \right)^2$	−5.12	5.12
4	Schwefel 2.21	$f(x) = \sum_{i=1}^{n} \lvert x_i \rvert + \prod_{i=1}^{n} \lvert x_i \rvert$	−600	600
5	Schwefel 2.26	$f(x) = -418.983 \sum_{i=1}^{n} \left[x_i \sin \left(\sqrt{\lvert x_i \rvert} \right) \right]$	−50	50
6	Rosenbrock	$f(x) = \sum_{i=1}^{n-1} \left[100 \sqrt{\lvert x_i - x_i^2 \rvert} + (1 - x_i)^2 \right]$	−100	100
7	Step	$f(x) = \sum_{i=1}^{n} \lfloor \lvert x_i \rvert \rfloor$	0	3.1416
8	Quartic	$f(x) = \sum_{i=1}^{n} i x_i^4 + \text{rand}\,[0, 1)$	−5	10

the variables for MBO algorithm were fixed as follows: the butterfly adjusting rate *BAR* is set to 0.4167, max step $S_{\max}$ is set to 1, the migration period $\text{Per}_i = 1.2$, and the migration ratio $p = 0.4167$, as per the setup in the original work of MBO [14].

3.3.2 Benchmark Function

This paper uses a set of eight test functions for global numerical optimization. These functions are listed in Table 3.1 alongside their respective equations and properties**.**

3.4 Results

Table 3.2 lists the optimization results when applying the eight optimization test functions to ABC, MBO, and our HAM methods. The listed values are the optimal values of the objective function achieved by each algorithm after iterating over 50 generations. The *mean* values in the table are averaged over 20 runs (each run constitutes 50 iterations) and listed along the *standard deviation*. The *min* values, however, are the best results achieved by each algorithm at all. By the "best result," we mean the closest result to the actual optimal value of the function.

It is evident from Table 3.2 that the HAM algorithm can reach a better optimum on average, at least with respect to the set of benchmark functions used in the experiments (HAM has better average results in the case of seven out of eight test functions). For ease of recognition, the best average result is marked with bold font and shaded in a gray cell. The *min* values are bold font to identify the absolute best minimum achieved for each function. Note that this value is meaningful because it happened that the minimum achieved values by the algorithms for the selected benchmark functions are closest to the real optimum. With respect to the set of test functions used in our evaluations, HAM could achieve the best result in six out of eight cases.

Table 3.2 The *min*, *mean*, and *standard deviation* of test function values found by ABC, MBO, and the proposed HAM algorithms, averaged over 20 experimental runs. The dimensions set to 10

	ABC			MBO			HAM		
No	Best	Mean	Std. dev	Best	Mean	Std. dev	Best	Mean	Std. dev
1	4.13E−04	1.01E−02	9.37E−03	5.14E−04	8.67E−01	2.73E + 00	3.57E−05	9.37E−05	6.24E−05
2	1.07E−01	2.64E−01	1.21E−01	3.79E−02	2.18E+00	4.08E+00	7.14E−03	1.70E−02	8.60E−03
3	6.66E+02	2.11E+03	1.00E+03	7.61E−03	4.18E+03	3.20E+03	6.72E−03	2.97E+00	1.16E+01
4	9.89E+00	2.14E+01	7.13E+00	2.35E−02	1.68E+01	1.57E+01	5.84E−03	1.22E−02	4.64E−03
5	2.45E+02	6.47E+02	1.89E+02	1.29E−04	1.22E+03	6.79E+02	1.13E+03	1.55E+03	1.85E+02
6	2.31E+01	1.78E+02	2.33E+02	8.91E+00	7.14E+04	1.54E+05	8.20E+00	8.56E+00	1.24E−01
7	3.44E+00	5.30E+00	1.85E+00	6.52E−06	1.29E+03	2.43E+03	2.51E+00	2.52E+00	6.70E−03
8	2.02E+00	2.97E+00	4.55E−01	1.83E+00	3.31E+00	1.34E+00	1.54E+00	2.19E+00	2.43E−01

On another perspective, we also graphed the optimization process of each algorithm (for each benchmark function) as the value of the so-far best solution versus the current iteration, which shows the search path in terms of selected best solution per iteration. The curve of this kind is expected to decline overall at a slope that reflects the convergence speed of the algorithm (there is no degradation during the process of any included metaheuristic algorithm, as the best solution is either improved or kept unchanged at all iterations). Therefore, these graphs can be called the convergence plots of the algorithms. Because of the large number of plots, we include hereby representative samples of the convergence plots in Fig. 3.1, which compares the convergence of HAM with the two most related metaheuristic techniques: ABC and MBO**.**

Figures 3.1a–d shows that the HAM algorithm enjoys not only a superior overall performance in terms of the quality of the found optimal solution but also a faster convergence especially in the earlier stages. Although the starting points of the algorithms are close to each other in the plots of the four testing functions in the figure, the proposed HAM method does not trap into a quick local optimum, unlike the original ABC and MBO algorithms, for example.

3.5 Conclusion

The proposed algorithm in the article, HAM, is founded on two metaheuristics algorithms, which are the artificial bee colony and monarch butterfly optimization algorithms. It is worth to mention that HAM is the first novelty hybrid algorithm born of these two algorithms. In addition HAM is composite of three phases: employee bee adjusting phase, onlooker phase, and scout phase. The initial phase is a modified version of the adjusting operator in MBO algorithm, while the second and last (onlooker and scout) phases are identical to those original equivalents found in ABC algorithm.

The crux of HAM development aims at finding a higher convergence speed and best optimal solutions than its predecessors by enhancing diversification of MBO that has been used to augment good intensification ability of ABC. In future works, we intend to utilize HAM algorithm as a neural network trainer as well as to extend the method for solving multi-objective optimization problems to serve many other various purposes.

Acknowledgments This work has been funded by Universiti Sains Malaysia, APEX (308/AIPS/ 415401), and also supported by the Fundamental Research Grant Scheme (FRGS) 203/PKOMP/6711426].

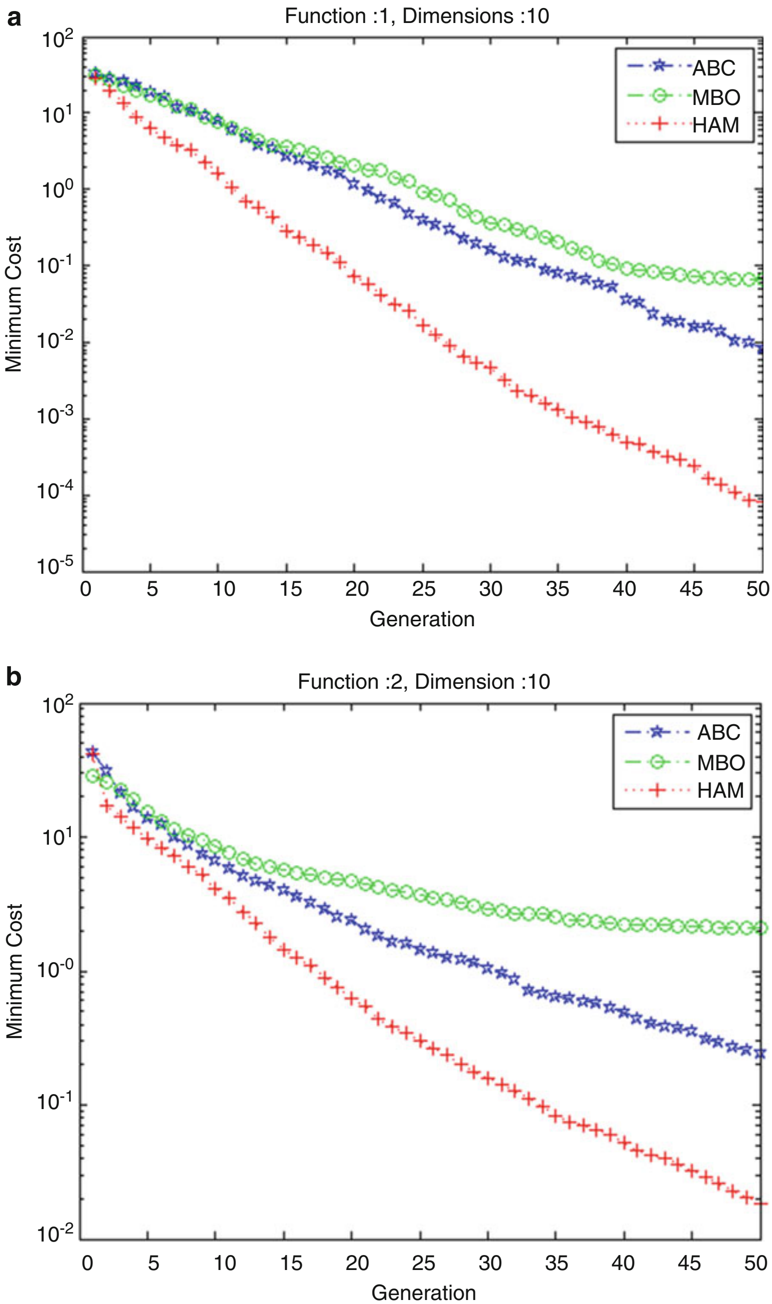

Fig. 3.1 Performance of ABC, MBO, and HAM algorithms for (**a**) F1, (**b**) F2, (**c**) F3, and (**d**) F4 benchmark functions

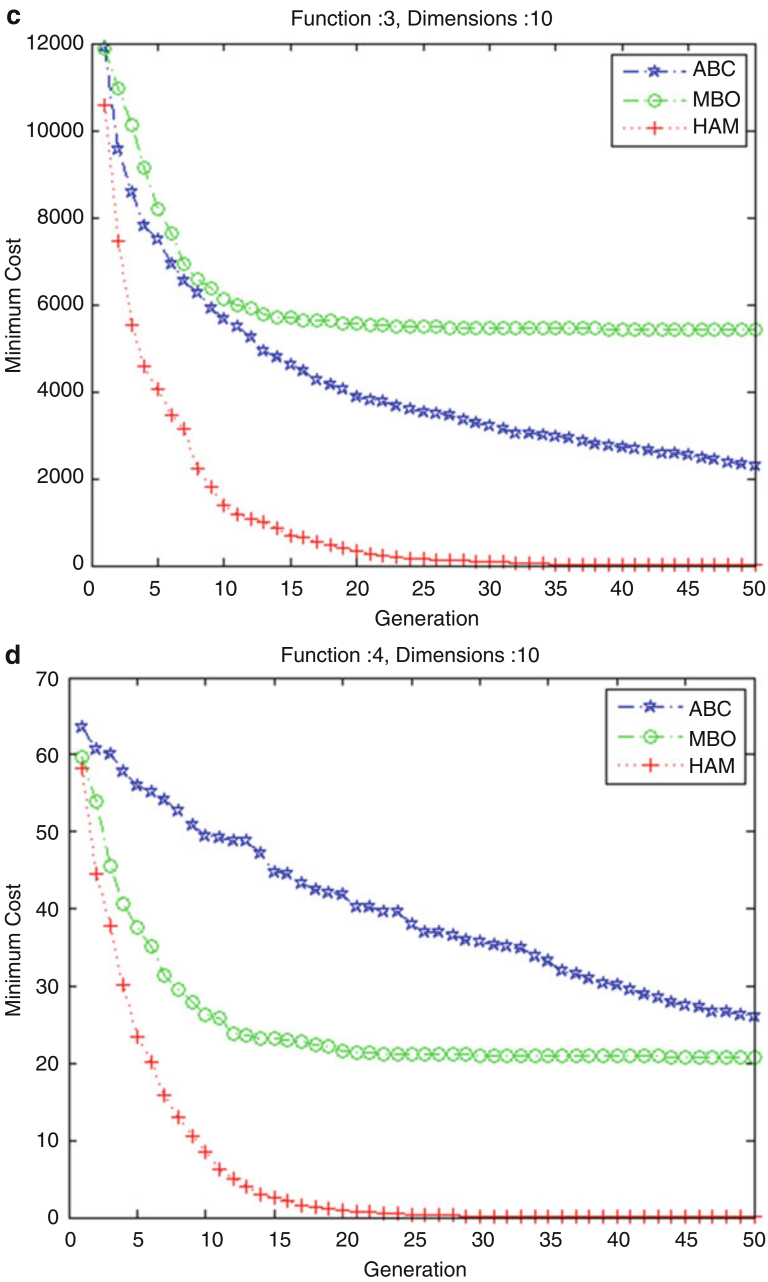

Fig. 3.1 (continued)

References

1. Sörensen K, Glover FW (2013) Metaheuristics. In: Encyclopedia of operations research and management science. Springer, New York, pp 960–970
2. Eberhart RC, Kennedy J (1995) A new optimizer using particle swarm theory. In: Proceedings of the sixth international symposium on micro machine and human science, New York
3. Dorigo M, Birattari M, Stützle T (2006) Ant colony optimization. IEEE Comput Intell Mag 1(4):28–39
4. Dorigo M, Maniezzo V, Colorni A (1996) Ant system: optimization by a colony of cooperating agents. IEEE Trans Syst Man Cybern B Cybern 26(1):29–41
5. Karaboga D (2005) An idea based on honey bee swarm for numerical optimization. Technical report-tr06, Erciyes University, Engineering Faculty, Computer Engineering Department
6. Karaboga D, Basturk B (2007) A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. J Glob Optim 39(3):459–471
7. Yang X-S (2010) Nature-inspired metaheuristic algorithms. Luniver Press, Frome
8. Yang X-S, Deb S (2009) Cuckoo search via Lévy flights. In: Nature & biologically inspired computing, 2009. NaBIC 2009. World congress on, IEEE
9. Simon D (2008) Biogeography-based optimization. IEEE Trans Evol Comput 12(6):702–713
10. Li X, Zhang J, Yin M (2014) Animal migration optimization: an optimization algorithm inspired by animal migration behavior. Neural Comput Applic 24(7–8):1867–1877
11. Meng X et al (2014) A new bio-inspired algorithm: chicken swarm optimization. In: Advances in swarm intelligence. Springer, New York, pp 86–94
12. Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. Adv Eng Softw 69:46–61
13. Gandomi AH, Alavi AH (2012) Krill herd: a new bio-inspired optimization algorithm. Commun Nonlinear Sci Numer Simul 17(12):4831–4845
14. Wang G-G, Deb S, Cui Z (2015) Monarch butterfly optimization. Neural Comput Applic 28(3):1–20
15. Yang X-S (2010) A new metaheuristic bat-inspired algorithm. In: Nature inspired cooperative strategies for optimization (NICSO 2010). Springer, Berlin, pp 65–74
16. Kirkpatrick S, Vecchi MP (1983) Optimization by simulated annealing. Science 220(4598):671–680
17. Gandomi AH, Yang XS, Talatahari S, Alavi AH (2013) Metaheuristic application in structures and infrastructures. Elsevier, Waltham, Mass
18. Črepinšek M, Liu S-H, Mernik M (2013) Exploration and exploitation in evolutionary algorithms: a survey. ACM Comput Surv (CSUR) 45(3):35
19. Ghanem, Waheed Ali HM, Jantan A (2016) Novel multi-objective artificial bee Colony optimization for wrapper based feature selection in intrusion detection. Int J Adv Soft Comput Appls 8(1):70–81
20. Karaboga D, Ozturk C (2009) Neural networks training by artificial bee colony algorithm on pattern classification. Neural Network World 19(3):279
21. Ghanem WAHM, Jantan A (2014) Using hybrid artificial bee colony algorithm and particle swarm optimization for training feed-forward neural networks. J Theoret Appl Inf Technol 3:67
22. Bolaji ALA, khader AT, Al-Betar MA, Awadallah MA (2013) Artificial bee colony algorithm, its variants and applications: a survey. J Theoret Appl Inf Technol 47(2):434–459

Chapter 4
Domain Model Definition for Domain-Specific Rule Generation Using Variability Model

Neel Mani, Markus Helfert, Claus Pahl, Shastri L. Nimmagadda, and Pandian Vasant

4.1 Introduction

We are primarily concerned with the utilization of a conceptual domain model for rule generation, specifically to define a domain-specific rule language (DSRL) [1, 2] syntax, its grammar for business process model and domain constraint management. We present a conceptual approach for outlining a DSRL for process constraints [3]. The domain-specific content model (DSCM) definition needs to consider two challenges. The first relates to the knowledge transfer from domain concept to conceptual model, where model inaccuracies and defects may have been translated because of misunderstandings, model errors, human errors or inherent semantic mismatches (e.g. between classes). The other problem relates to inconsistency, redundancy and incorrectness resulting from multiple views and abstractions. A domain-specific approach provides a dedicated solution for a defined set of

N. Mani (✉) • M. Helfert
School of Computing, ADAPT Centre for Digital Content Technology, Dublin City University, Dublin, Ireland
e-mail: neel.mani@computing.dcu.ie; markus.helfert@computing.dcu.ie

C. Pahl
Faculty of Computer Science, Free University of Bozen-Bolzano, Bolzano, Italy
e-mail: Claus.Pahl@unibz.it

S.L. Nimmagadda
School of Information Systems, Curtin Business School (CBS), Perth, WA, Australia
e-mail: shastri.nimmagadda@curtin.edu.au

P. Vasant
Department of Fundamental and Applied Sciences, Universiti Teknologi PETRONAS, Tronoh, Malaysia
e-mail: pvasant@gmail.com

P. Vasant et al. (eds.), *Modeling, Simulation, and Optimization*,
EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-319-70542-2_4

problems. To address the problem, we follow a domain model language approach for developing a DSRL and expressing abstract syntax and its grammar in BNF [4] grammar. A domain-specific language (DSLs) [5–7] refers to an approach for solving insufficient models by capturing the domain knowledge in a domain-specific environment. In the case of semantic mismatches/defects, a systematic DSL development approach provides the domain expert or an analyst with a problem domain at a higher level of abstraction.

DSL is a promising solution for raising the level of abstraction that is easier to understand or directly represent and analyse, thus, attenuating the technical skills required to develop and implement domain concepts into complex system development. Furthermore, DSLs are either textual or graphical language targeted to specific problem domains by increasing the level of automation, e.g. through rule and code generation or directly model interpretation (transform or translate), as a bridge, filling the significant gap between modelling and implementation. An increase in effectiveness (to improve the quality) and efficiency of system process is aimed at rather than general-purpose languages that associated with software problems. Behavioural inconsistencies of properties can be checked by formal defect detection methods and dedicated tools. However, formal methods may face complexity, semantic correspondence and traceability problems. Several actions are needed for implementation of any software system. These are from a high-level design to low-level execution. The enterprises typically have a high level of legacy model with various designs in a domain or process model. Automatic code generation [8–11] is a well-known approach for getting the execution code of a system from a given abstract model. The rule is an extended version of code since code requires compiling and building, but the rule is always configurable. Rule generation is an approach by which we transform the higher-level design model as input and the lower level of execution code as output. It manages the above-mentioned constraint.

We structure the modelling and DSL principles in Sect. 4.1. In Sect. 4.2, we discuss the state of the art and related work. We give an overview of the global intelligent content processing in a feature-oriented DSL perspective, which offers the domain model and language definition in Sect. 4.3. Then, we describe the ontology-based conceptual domain model in Sect. 4.4. The description of the domain model and language expressed in terms of abstract and concrete syntax are given in Sect. 4.5. As a part of DSRL, the general design and language are presented in Sect. 4.6. Section 4.7 provides details regarding implementation of a principal architecture of DSRL generation and how it translates from the domain model to the DSRL. We discuss analysis and evaluation of the rule in Sect. 4.8. Finally, we conclude our work with future scope. Throughout the investigation, we consider the concrete implementation as a software tool. However, a full integration of all model aspects is not aimed at, and the implementation discussion is meant to be symbolic. The objective is to outline the principles of a systematic approach towards a domain model used, as source model, and the domain-specific rule language, as a target for content processes.

4.2 Related Work

The web application development is described as a combination of processes, techniques and from which web engineering professionals make a suitable model. The web engineering is used for some automatic web application methodologies such as UWE [12], WebML [13] and Web-DSL [14] approaches. The design and development of web applications provide mainly conceptual models [15], focusing on content, navigation and presentation models as the most relevant researchers expressed in [16, 17]. Now, the model-driven approach for dynamic web application, based on MVC and server, is described by Distante et al. [18]. However, these methods do not consider the user requirement on the variability model. To simplify our description, we have considered the user requirement, and according to the need of the user, the user can select the feature and customize the enterprise application at the dynamic environment.

A process modelling language provides syntax and semantics to precisely define and specify business process requirements and service composition. Several graph and rule-based languages have been emerged for business process modelling and development, which rely on formal backgrounds. They are Business Process Modeling Notation (BPMN) [19], Business Process Execution Language (BPEL)/WS-BPEL, UML Activity Diagram Extensions [20], Event-Driven Process Chains (EPC) [21], Yet Another Workflow Language (YAWL) [22], WebSphere FlowMark Definition Language (FDL) [23], XML Process Definition Language (XPDL) [24], Java BPM Process Definition Language (jPDL) [25] and Integration Definition for Function Modeling (IDEF3) [26]. These languages focus on a different level of abstraction ranging from business to technical levels and have their weaknesses and strengths for business process modelling and execution. Mili et al. [27] survey the major business process modelling languages and provide a brief comparison of the languages, as well as guidelines to select such a language. In [28], Recker et al. present an overview of different business process modelling techniques. Among the existing languages, BPMN and BPEL are widely accepted as de facto standards for business process design and execution, respectively.

Currently, there is no such type of methodology or process of development for creating a rule-based system in a web application (semantic-based). Diouf et al. [29, 30] propose a process which merges UML models and domain ontologies for business rule generation. The solution used for semantic web has ontologies and UML to apply to the MDA approach for generating or extracting the rules from high level of models. Although the proposed combination of UML and semantic-based ontologies is for extracting the set of rules in target rule engine, they only generate the first level of the abstraction of the rules.

Our approach provides the systematic domain-specific rule generation using variability model. The case study uses intelligent content processing. Intelligent content is digital that provides a platform for users to create, curate and consume the content in dynamic manner to satisfy individual requirements. The content is stored, exchanged and processed by a dynamic service architecture, and data are exchanged, annotated with metadata via web resources.

4.3 Business Process Models and Constraints

We use the intelligent content (IC) [31] processing as a case study in our application. The global intelligent content (GIC) refers to digital content that allows users to create, curate and consume content in a way that fulfils dynamic and individual requirements relating to information discovery, context, task design and language. The content is processed, stored and exchanged by a web architecture, and the data are revised, annotated with metadata through web resources. The content is delivered from creators to consumers. The content follows a certain path that consists of different stages such as extraction and segmentation, named entity recognition, machine translation, quality estimation and post-editing. Each stage, in the process, comes with its challenges and complexities.

The target of the rule language (DSRL) is an extendable process model notation for content processing. Rules are applied at processing stages in the process mode.

The process model that describes activities remains at the core. It consists of many activities and sub-activities of reference for the system and corresponds to the properties for describing the possible activities of the process. The set of activities constitutes a process referred to as the extension of the process, and individual activities in the extension are referred as instances. The constraints may be applied at states of the process to determine its continuing behaviour depending on the current situation. The rules combine a condition (constraint) on a resulting action. The target of our rule language (DSRL) is a standard business process notation (as shown in Fig. 4.1).

The current example is a part of digital content (processing) process model as shown in Fig. 4.1, a sample process for the rule composition of business processes and domain constraints that conduct this process. The machine language activity translates the source text into the target language. The translated text quality decides whether further post-editing activity is required. Usually, these constraints are domain-specific, e.g. referring to domain objects and their properties, respectively.

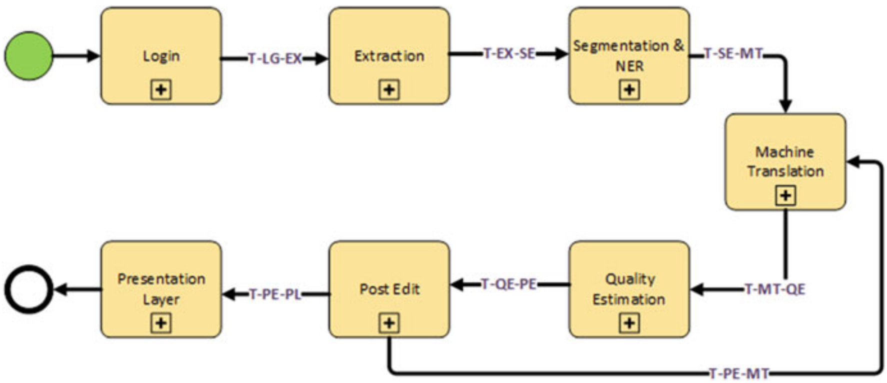

Fig. 4.1 Process model of global intelligent content

4.4 Ontology-Based Conceptual Domain Model

We outline the basics of conceptual domain modelling, the DSRL context and its application in the intelligent content context. The domain conceptual models (DCMs) are in the analysis phase of application development, supporting improved understanding and interacting with specific domains. They support capturing the requirements of the problem domain and in ontology engineering. A DCM is a basis for the formalized ontology. There are several tools, terminologies, techniques and methodologies used for conceptual modelling, but DCMs help better understanding, representing and communicating a problem situation in specific domains. We utilize the conceptual domain model to derive at a domain-specific rule language.

A conceptual model can define concepts concerning a domain model for a DSL, as shown in the class model in Fig. 4.2, which is its extended version described in [3]. A modelling language is UML-based language defined for a particular domain, defining relevant concepts as well as a relation (intra or inter-model) with a metamodel. The metamodel consists of the concrete syntax, abstract syntax and static semantics of the DSL. The abstract syntax defines modelling element, such as classes, nodes, association, aggregation and generalization, and relationships between the modelling elements [3].

A DSRL reuses domain model elements or define new modelling element depending on the domain concept and its relations with other elements. Modelling of elements is done with two fundamental types: concepts and relations. The concepts are based on domain concepts such as entity, state, action, location, risk, menu and relations which are used to connect elements. The relations are also of two types:

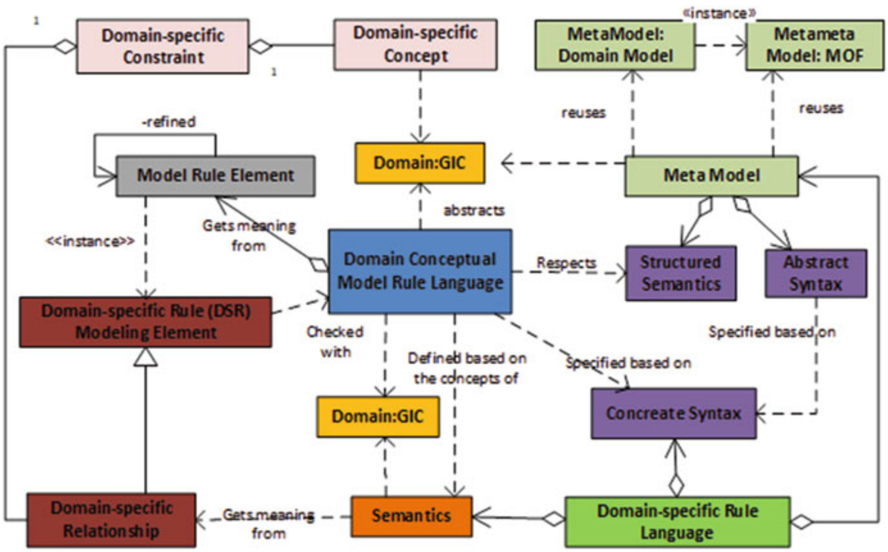

Fig. 4.2 A domain model-based DSRL concept formalization

generalization and association. Aggregation is a special type of an association. The association has domain-specific relations like conditional flow, multiplicity and aggregation. Furthermore, relations may have properties such as symmetry, reflexivity, equivalence, transitivity and partial order.

4.5 Language Definition of a Domain Model

Domain model serves as the very basis of all types of business applications that run on a domain, both individual and enterprise applications. The objective is to define the language for domain model and recognizes the internal data structures or schema used in it. It is easy to transform or translate the graphical domain model into textual rule language for a particular domain. In this scenario, the objective data structure refers to the storability of a domain model in a vulnerable environment, as in the case with a rule language. This is because the target of a language for mapping the translated domain model's knowledge into XML schema of DSRL follows the rule paradigm.

4.5.1 Language Description

Metamodelling is used to accomplish specifications for the abstract syntax. We introduce the domain model language by analysing its syntax definition (Fig. 4.3 shows in EBNF notation). The language with its basic notions and their relations is defined with structural constraints (for instance, to express containment relations

1	*Domain*	*::=*	*<Domain model>*	*Domain definition*
2	*Concept*	*::=*	*<Concept>*	*Concept definition*
3	*Class*	*::=*	*<Attributes>,< Operation>, < Receptions>, <Template Parameters>,< Component>, <Constraints>, <Tagged Values>*	*Class Definition*
4	*<Relations>*	*::=*	*<Association>|<DirectedAssociation>|<Reflexive Association>|<Multiplicity>|<Aggregation>|<Co mposition>|<Inheritance/Generalization>|<Reali zation>*	*Class relationships*
5	*<Association>*	*::=*	*'→' | '✱' | '::'*	*Structural relationship between objects (classes) of different type*
6	*<Type>*	*::=*	*<BuiltinType>|<UCase Ident >| <EnumType>*	*Domain model type concept type or extended type enumeration type list type*
7	*<PrimitiveTypes >*		*<String>,<Integer>,<Boolean>, ...,<Date>*	*Domain model primitive (built-in) types*

Fig. 4.3 Syntax definition of domain model language

or type correctness for associations), multiplicities, precise mathematical definition and implicit relationships (such as inheritance, refinement). The visual appearance of the domain-specific language is accomplished by syntax specifications, which is done by assigning visual symbols to those language elements that are to be represented on diagrams.

4.5.2 Syntax

For describing the language in general, the rule language checks various kinds of activities. The primary requirement is to specify the concept of the syntax (i.e. abstract and or concrete syntax) and develop its grammar. The semantics is designed to define the meaning of the language. The activities are completed by concepting, designing and developing systematic domain-specific rule language systems; defining the functions and its parameters, priorities or precedence of operators and its values; and naming internal and external convention system. The syntaxes are expressed with certain rules, conforming to BNF or EBNF grammars that can be processed by rules or process engine to transform or generate the set of rules as an output. The generated rules follow the abstract syntax and grammar to describe the domain concepts and domain models because both the artefacts (abstract syntax and grammar) are reflected in the concrete syntax.

4.5.3 Abstract Syntax

The abstract syntax refers to a data structure that contains only the core values set in a rule language, with semantically relevant data contained therein. It excludes all the notation details like keywords, symbols, sizes, white space or positions, comments and colour attributes of graphical notations. The abstract syntax may be considered as more structurally defined by the grammar and metamodel, representing the structure of the domain. The BNF may be regarded as the standard form for expressing the grammar of rule language, and some type describes how to recognize the physical set of rules. Analysis and downstream processing of rule language are the main usages of abstract syntax. Users interact with a stream of characters, and a parser compiles the abstract syntax by using a grammar and mapping rules.

For example, we do process activities in our case domain (Global Digital Content): Extraction and Machine Translation (MT). The list of the process model, event and condition are the following:

List of Process

```
<Process-ModelList>::= <gic:Extraction>|
<gic:MachineTranslation>
```

List of Event

```
<EventList> ::= {
gic:Text -->SourceTextInput,
gic:Text -->SourceTextEnd,
gic:Text -->SourceTextSegmentation,
gic:Text -->SourceParsing,
gic:Text -->MTSourceStart,
gic:Text -->MTTargetEnd,
gic:Text -->TargetTextQARating,
gic:Text -->TargetTextPostEditing,
}
```

List of conditions

```
<ConditionList>::= <gic:Extraction.Condition>|
            <gic:MachineTranslation.Condition>
<gic:Extraction.Condition>::=
     IF(<gic:Text.Length::=<L)|
     IF(<Source.Language::==Language_List>)
     IF(<Target.Language::==Language_List>)
     IF(<SingleLangugeDetection((gic:Text)::==  True
                                                  |False>)
     IF(<MultiLanguageText(gic:Text)::== True|False>)
<gic:MachineTranslation.Condition>::=
      IF (<gic:Translation
(Source.Lang, TargetLang,gic:Text)  ::= True|False>)|
    IF (<gic:Translation.Memory ::=
        <TM)(Mem Underflow)|
    IF (<gic:Translation.Memory ::= >TM)
        (Mem Overflow)|
    IF(<gic:Translation(gic:TxtSource,Source.Lang)>
gic:Translation(gic:TxtTarget,Target.Lang)>)
```

Where L is length of text and TM is the specific memory size.

4.5.4 Concrete Syntax

Rule languages use textual concrete syntax, which implies that a stream of characters expresses the programme syntax. The modelling languages traditionally have used graphical notations and primarily in modelling languages. Though textual domain-specific languages (and mostly failed graphic-based general-purpose languages) have been in use for a long time only recently, the textual syntax has found a prominent use for domain-specific modelling. Textual, concrete syntax form has been traditionally used to store programmes, and this character stream is transformed using scanners and parsers into an abstract syntax tree for further processing by the programming languages. In the modelling languages, editors have found a major usage, as it directly manipulates the abstract syntax and uses projection to render the concrete syntax in the form of diagrams.

The concrete syntax of DSLs is expected to be textual by default. If good tool support is available, the textual support has been found to be adequate for comprehensive and complex software systems. The programmers write lesser code in DSL as compared to a GPL for expressing the same functionality – because the available abstractions are quite similar to the domain. An additional language module suitable for the domain is defined easily by the programmers.

4.6 Rule Language Definition

Now we go back to the full rule definition. The DSRL grammar [2] is defined as follows. We start with a generic skeleton and then map the globic domain model (gic).

```
<DSRL Rules> ::= <EventsList>
                <RulesList>
                <ProcessModelList>
<EventLists> ::= <Event> | <Event> <EventLists>
<Event> ::= EVENT <EventName> IF <Expression> |
            EVENT <EventName> is INTERN or EXTERN
<RulesList> ::= <Rule> | <Rule> <RuleList>
<Rule> ::= ON<EventName>
           IF<Condition>DO<ActionList>
<ActionList> ::= <ActionName> |
                 <ActionName>,<ActionList>
<ProcessModelList> ::= <ProcessModel> |
                       <ProcessModel>,
                       <ProcessModelList>
<ProcessModel> ::= ProcessModel <ProcessModelName>
               ::= <ProcessModelName>
                   [TRANSITION_(SEQUENCIAL
                          (DISCARDDELAY))],
                   [TRANSITION_PARALLEL
                      (DISCARD|DELAY)]
                   [INPUTS(<InputList>)]
                   [OUTPUTS(<OutputList>)]
```

TRANSITION_SEQUENCIAL and TRANSITION_PARALLEL denoted as transitions of a process model.

The description of DSRL contains lists of events, condition, an action of the rules and process model states. An event can be an internal or external (for rules generated as an action, it may be the INTERNAL or EXTERNAL term) or generated when the expression should have been satisfied by the condition or becomes true. An event name activates with ON syntax, which is a Boolean expression to determine the conditions that apply and the list of actions that should be performed when event and condition are matched or true (preceded by DO syntax). The process models contain the state name. A certain policy is decided in the process model when sequential, and parallel actions are performed or sent in that state. An action is an executable programme or set of computation decussation. The action provides methods or function invocation, creating, modifying, updating, communicating or destroying an object. DISCARD allows discarding the instructions, and DELAY allows delaying the instructions but one.

For example, the gic:Extraction is used in an event on Text Input by user as a source data.

```
EVENT IF TextInput_ON
EVENT gic:TextInput::BOOL IF TextInput_Get
EVENT gic:TextInput::BOOL IF TextInput_ON
ON presence
IF (gic:SourceLang::EN) DO
(
     ON presence
      IF (gic:TextLength <X) DO
gic:Translate(Text)
ELSE

 Notification to user (Text LENGTH LESS   THAN X)
)
ELSE
Notification to user(Source language is invalid)
```

4.7 Implementation of Principle Architecture of DSRL Generator

The principle architecture of a domain-specific rule (DSR) is the automated model to text generator on accessible domain models, extracting information from them and translating it into output in a specific target syntax. This process follows the concept of model-driven architecture which depends on the metamodel. The modelling language with its concepts, the source syntax, semantics and its rules are required by the domain-specific framework and target environment. We present the process architecture of a domain model translation and the target rule environment in Fig. 4.4.

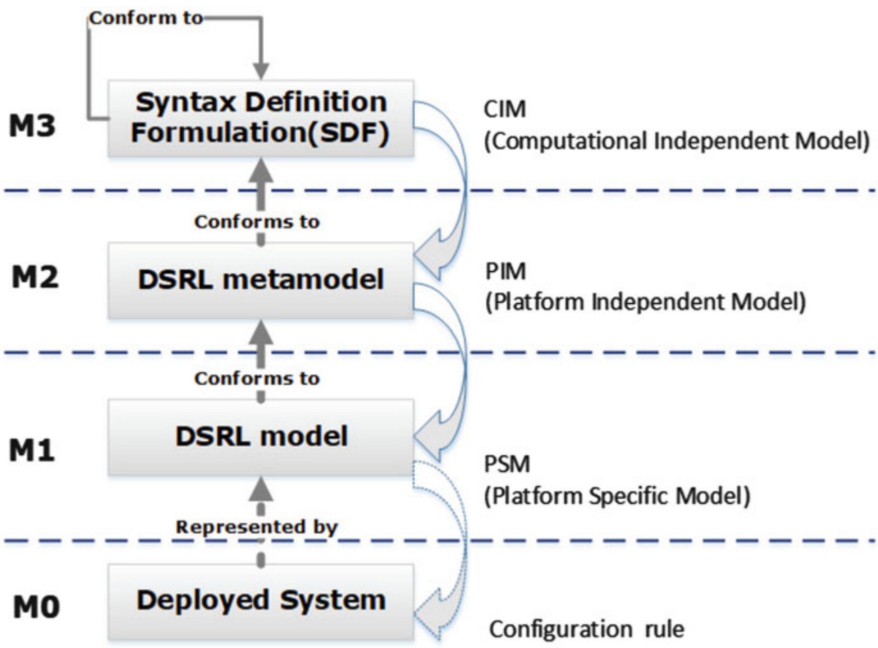

Fig. 4.4 MDA organization view of models approach and artefacts of DSRL generator

4.7.1 Architecture

The architecture of the DSRL generator follows the MDA as four-level model organization, presented by Bézivin [32] as illustrated in Fig. 4.4. At the top level, the M3 is the Syntax Definition Formalism (SDF) metametamodel which is the grammar of the SDF. This level is also known as Computational Independent Model (CIM) or metametamodel as defined (and thus conforms to) itself [33]. A self-representation of the BNF notation takes some lines. This notation allows a defining infinity of well-formed grammars. A given grammar allows description of the infinity in syntactically correct DSR configuration.

At the M2 level, we describe the DSRL metamodel, i.e. the grammar of DSRL with ECA as defined in SDF, and this level is called Platform Independent Model (PIM). The metamodel conforms to the metametamodel at level M3.

At the M1 level, we describe DSRL models for configuration applications. It is known as platform-specific model (PSM) consisting of entity and definitions. The model conforms to the metamodel at level M2. The bottom level is called M0; we define the configuration of BPM customization consisting of DSR and XML rules, which represent the models at the M1 level.

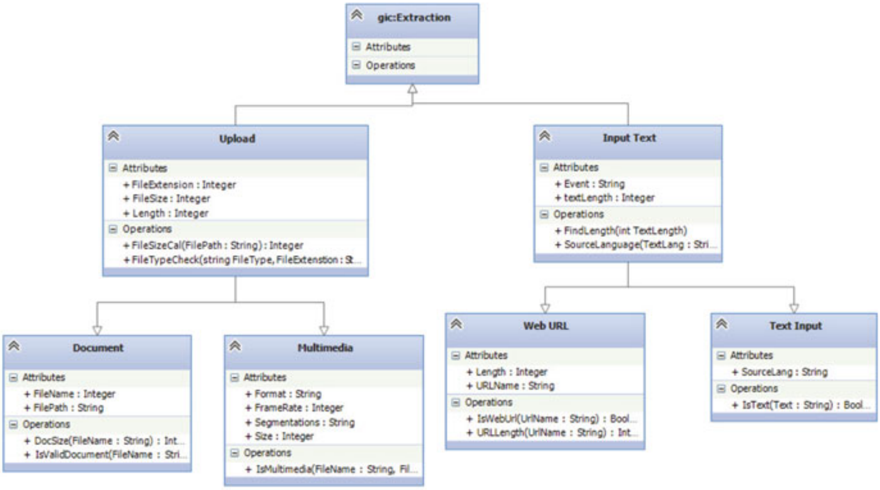

Fig. 4.5 Source metamodel of gic:Extraction as example DSRL used for mapping

4.7.2 Mappings Domain Model and Domain-Specific Rule Language

A mapping is description of mapping rule definitions, generation, configuration and execution of order specification. Each mapping rule specifies what target model fragment is created for the given DSR. The mapping rule body contains one or more class of the domain model occurrences (Sect. 4.4) with all attribute and operational value set. Expressions for attribute, functional and operational setting are based on a specific source metamodel; Fig. 4.5 shows the corresponding domain model of gic:Extraction subtype of digital content process used as a source metamodel to describe the DSRL conceptualization as illustrated in Fig. 4.2 (Sect. 4.3). Although the given source metamodel is completely translated into graphical model to text rule by using the grammar or language definition of source and target metamodel as given in Sects. 4.4 and 4.5, it is sufficient to show all basic mapping constructs.

4.7.3 Domain Model Translation into DSR

A rule generation is made automatic such that domain models are accessed in a way to extract information and translate it into output in a specific syntax based on feature model, as selected by the domain user. This process model is guided by the metamodel, the modelling language with its high level of concepts, syntactical and semantics rules. The input required by the user (selection of the feature model) needs a domain to process the domain model, target environment as rule generation and configuration. We present the process of domain model translation and target rule environment in Fig. 4.2.

In an example of a rule generation from a domain model, we propose an approach: the domain model would be translated into a domain-specific rule language through a model transformation or translation, and then the DSRL metametamodel would be synthesized into DSR text by means of a rule generator. It is advantageous to have syntactical and semantic domain translation achieved by a graphical model to text model translation. It is a dedicated technology, because rule generators deal with the abstract and concrete syntaxes of the target language (in Sect. 4.3) directly. The entire process separates two distinct tasks (translation and synthesis) that are performed using appropriate tools.

For translation, the models need to be expressed in a modelling language (e.g. UML for design models, and programming languages for source models). A metamodel expresses the modelling languages' syntax and semantics by themselves. For example, the syntax of the domain metamodel has feature notations expressed using class diagrams, whereas its semantics is described by well-defined rules (expressed as OCL constraints) and a mixture of natural languages [34]. Based on the language in which the source and target models of a translations (grammar of model change) or transformation are expressed, a distinction is made between endogenous and exogenous transformations. The endogenous transformations are expressed between models in the same languages (when the grammar and structure are same).

The exogenous transformations are conversions made between models and expressed using different languages (the grammar and syntax are different), which is also known as translation. Essentially the same distinction was proposed in [35] but ported to a model transformation setting. We use the exogenous transformation that taxonomy and graphical model to text rule, whereas the term translation is used for an exogenous transformation (Fig. 4.6).

4.8 Generated Rule Analysis and Evaluation

In rule generation evaluation, we validate the type of generated output concerning the correctness, completeness, output effectiveness and efficiency. Our primary goal is to have a proof of fully functional and operational correctness and completeness of the rule for its feature requirement selected by the domain user. Our rule evaluation consists of the following:

- Validation of rule generation concerning under- and over-generation.
  - Under-generation – We define under-generation as missing instance (e.g. events, actions, etc.) at the time of generation or after generation.
  - Over-generation – This is identified as an added information regarding syntax and semantics (functional and operational information).
- Evaluation of syntactical and semantical correctness of generated rule fulfils our goal. The above results imply that if one knows, for example, how to

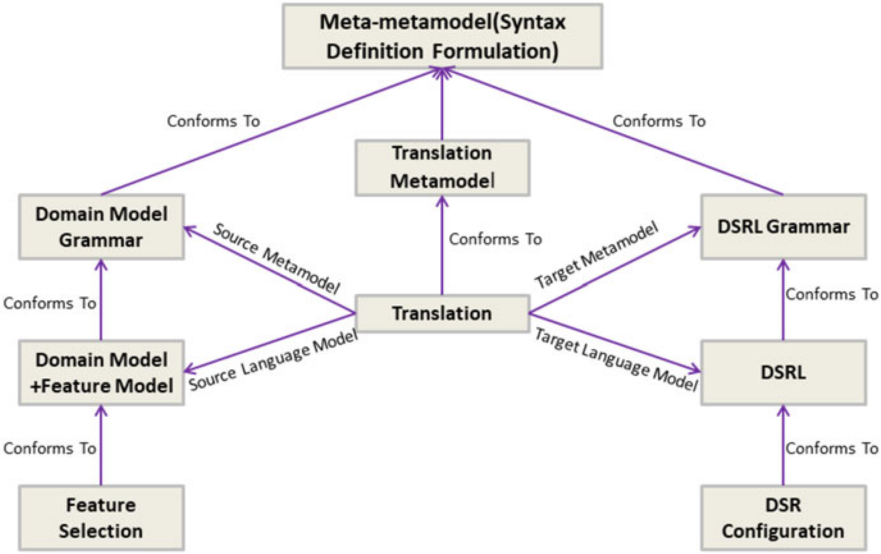

Fig. 4.6 Domain model to DSR translations

formulate partial correctness of a given deterministic algorithm in predicate mathematics, the formulation of many other properties of the algorithm in predicate mathematics could have been straightforward. As a matter of fact, partial correctness has already been formulated in predicate mathematics and manual rule templates of many feature deterministic algorithms.

  – Syntactical correctness means correct use of keys, functions, values and grammatical rules.
  – Semantical correctness is important for functional and operational point of view; here, we validate the correct order of generated rule and grammatical sequence.
  – Grammatical correctness means the generated rules follow SDF grammar which is used in M3 level of MDA model as shown in Fig. 4.4.
  – Comparing automated and handwritten rules.

- Evaluation of completeness of generated rule.

  – Completeness: Most rule languages are not designed as targets for rule generation, because of lack of functional and operational parameters in programme fragments. Major challenging part in the rule sets has two constraints: deadlock situation and live-lock situation.

    · Identification of rule deadlock.
    · Identification of rule live lock.

It also needs to be identified if rules of the application require any additional set of rules to function as desired.

4.9 Discussion

In this paper, we have proposed a syntax definition for domain model language for rule generation and presented a DSR generation development for domain model through MDA approach using domain variability. We have presented a novel approach of handling knowledge transfer from domain concept (domain model) to conceptually configurable rule language, avoiding the inaccuracies and misunderstanding, model error, human error or semantic mismatches during translation of graphical abstract model to text rule. We have added adaptivity to the domain model. We provide a conceptual view of domain-specific rule generation and manage the domain model and variability model using MDA. It helps in managing frequent changes of the business process along with variability schema of a set of structured variation mechanisms for the specification. The domain user can generate the DSRs and configure domain constraint in a dynamic environment. They can generate and configure DSRs without knowing any technical and programming skill. The novelty of our approach is a variability modelling usage as a systematic approach to transforming (generate) domain-specific rule from domain models.

We plan to extend this approach in combination with our existing work on business process model customization based on user requirement (feature model, domain model and process models) so that a complete development life cycle for the customization and configuration of the business process model is supported. We explore further research that focuses on how to define the DSRL concerning abstract and concrete syntactical description with grammar formation across different domains, converting conceptual models into generic domain-specific rule language which are applicable in other domains. So far, it is a model for text translation but has the potential to serve as a system that learns from existing rules and domain models, driven by the feature model approach with automatic constraint configuration that is resultant to an automated DSRL generation.

Acknowledgements This research is supported by Science Foundation Ireland (SFI) as a part of the ADAPT Centre for Digital Content Technology at Dublin City University (Grant No: 13/RC/2106) and EAI COMPSE 2016, Penang, Malaysia.

References

1. Mani N, Pahl C (2015) Controlled variability management for business process model constraints. In: ICSEA 2015, the tenth international conference on software engineering advances. IARIA XPS Press
2. Mani N, Helfert M, Pahl C (2016) Business process model customisation using domain-driven controlled variability management and rule generation. Int J Adv Softw 9(3–4):179–190
3. Tanrıöver ÖÖ, Bilgen S (2011) A framework for reviewing domain specific conceptual models. Comput Stand Interfaces 33(5):448–464
4. Knuth DE (1964) Backus normal form vs. backus naur form. Commun ACM 7(12):735–736

5. van Deursen A, Klint P, Visser J (2000) Domain-specific languages: an annotated bibliography. SIGPLAN Not 35(6):26–36
6. Fowler M (2010) Domain-specific languages. Pearson Education, New Jersey
7. Mernik M, Heering J, Sloane AM (2005) When and how to develop domain-specific languages. ACM Comput Surv (CSUR) 37(4):316–344
8. Hudak P (1997) Domain-specific languages. Handbook of programming languages, vol 3. pp 39–60
9. Ringert JO et al (2015) Code generator composition for model-driven engineering of robotics component & connector systems. arXiv preprint:1505.00904
10. Edwards G, Brun Y, Medvidovic N (2012) Automated analysis and code generation for domain-specific models. In: Software architecture (WICSA) and European conference on software architecture (ECSA), IEEE/IFIP conference on joint working. IEEE
11. Prout A et al (2012) Code generation for a family of executable modelling notations. Softw Syst Model 11(2):251–272
12. Koch N et al (2008) UML-based web engineering. In: Web engineering: modelling and implementing web applications. Springer, London, pp 157–191
13. Ceri S, Fraternali P, Bongio A (2000) Web modeling language (WebML): a modeling language for designing Web sites. Comput Netw 33(1):137–157
14. Groenewegen DM et al (2008) WebDSL: a domain-specific language for dynamic web applications. In: Companion to the 23rd ACM SIGPLAN conference on object-oriented programming systems languages and applications. ACM
15. Ceri S, Fraternali P, Matera M (2002) Conceptual modeling of data-intensive Web applications. IEEE Internet Comput 6(4):20–30
16. Moreno N et al (2008) Addressing new concerns in model-driven web engineering approaches. In: International conference on web information systems engineering. Springer
17. Linaje M, Preciado JC, Sanchez-Figueroa F (2007) Engineering rich internet application user interfaces over legacy web models. IEEE Internet Comput 11(6):53–59
18. Distante D et al (2007) Model-driven development of web applications with UWA, MVC and JavaServer faces. In: International conference on web engineering. Springer
19. White SA (2004) Introduction to BPMN, vol 2. IBM Cooperation
20. Dumas M, ter Hofstede AH (2001) UML activity diagrams as a workflow specification language. In: ł UML 2001 the unified modeling language. Modeling languages, concepts, and tools. Springer, Berlin, pp 76–90
21. Davis R (2001) Business process modelling with ARIS: a practical guide. Springer Science & Business Media, London
22. van der Aalst WMP, ter Hofstede AHM (2005) YAWL: yet another workflow language. Inf Syst 30(4):245–275
23. IBM (2010) WebSphere MQ workflow flowMareket definition language (FDL)
24. Zeng L et al (2004) Qos-aware middleware for web services composition. IEEE Trans Softw Eng 30(5):311–327
25. JBoss (2008) jBPM process definition language (jPDL)
26. Maker R et al (1992) IDEF3 – process description capture method report. Information integration for concurrent engineering (IICE), Armstrong Laboratory, Wright-Patterson AFB
27. Mili H et al (2010) Business process modeling languages: sorting through the alphabet soup. ACM Comput Surv (CSUR) 43(1):4
28. Recker J et al (2009) Business process modeling-a comparative analysis. J Assoc Inf Syst 10(4):1
29. Diouf M, Maabout S, Musumbu K (2007) Merging model driven architecture and semantic web for business rules generation. In: International conference on web reasoning and rule systems. Springer
30. Musumbu K, Diouf M, Maabout S (2010) Business rules generation methods by merging model driven architecture and web semantics. In: 2010 IEEE international conference on software engineering and service sciences. IEEE

31. Pahl C, Mani N, Wang M-X (2013) A domain-specific model for data quality constraints in service process adaptations. In: Advances in service-oriented and cloud computing. Springer, Heidelberg, pp 303–317
32. Bézivin J (2005) On the unification power of models. Softw Syst Model 4(2):171–188
33. Visser E (1997) Syntax definition for language prototyping
34. Group OM (2003) Unified modeling language specification version 1.5. formal
35. Visser E (2001) A survey of rewriting strategies in program transformation systems. Electron Notes Theor Comput Sci 57:109–143

Chapter 5
Hybridizing Bat Algorithm with Modified Pitch Adjustment Operator for Numerical Optimization Problems

Waheed Ali H.M. Ghanem and Aman Jantan

5.1 Introduction

Innumerous problems in the real life involve a set of possible solutions, from which the one with the best quality is termed as the optimal solution, and the process of searching for such a solution is known as (mathematical) optimization. The quality of solutions is represented by the ability to maximize or minimize a certain function, called the objective function, while the pool of possible solutions that can satisfy the required objective is called the search space. One can traverse all possible solutions, examine the result of the objective function in each case, and select the best solution. However, many real problems are intractable using this exhaustive search strategy. In these problems, the search space expands exponentially with the input size, and exact optimization algorithms are impractical. The historical alternative in such situations is to resort to heuristics, similar to simple rules of thumb that humans would utilize in a search process. Heuristic algorithms implement such heuristics to explore the otherwise prohibitively large search space, but they do not guarantee finding the actual optimal solution, since not all areas of the space are examined. However, a close solution to the optimal is returned, which is "good enough" for the problem at hand.

W.A.H.M. Ghanem (✉)
School of Computer Science, Universiti Sains Malaysia, Gelugor, Penang, Malaysia

Faculty of Education-Saber, University of Aden, Aden, Yemen

Faculty of Engineering, University of Aden, Aden, Yemen
e-mail: waheed.ghanem@gmail.com

A. Jantan
School of Computer Science, Universiti Sains Malaysia, Gelugor, Penang, Malaysia
e-mail: aman@cs.usm.my

P. Vasant et al. (eds.), *Modeling, Simulation, and Optimization*,
EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-319-70542-2_5

The next step would be to generalize those heuristics in higher-level algorithmic frameworks that are problem independent and that provide strategies to develop heuristic optimization algorithms. The latter are known as metaheuristics [1]. Early metaheuristics were based on the concept of evolution, where the best solutions among a set of candidate solutions are selected in successive iterations, and new solution is generated by applying genetic operators such as crossover and mutation to the parent solutions.

Similar to and including evolutionary algorithms, many metaheuristics were based on a metaphor, inspired by some physical or biological processes. Many recent metaheuristics mimic the biological swarms in performing their activities, in particular the important tasks of foraging, preying, and migration. Popular examples of developed metaheuristic algorithms in this category include particle swarm optimization (PSO) [2], which is inspired by the movement of swarms of birds or fishes; ant colony optimization (ACO) [3, 4], which is inspired by the foraging behavior of ants, where ants looking for food sources in parallel employ the concept of pheromone to indicate the quality of the found solutions; and artificial bee colony (ABC) algorithm, inspired by the intelligent foraging behavior of honey bees [5, 6].

The idea of deriving metaheuristics from natural-based metaphors proved so appealing that much more of such algorithms have been and continue to be developed. A few more examples include cuckoo search (CS) [7, 8], biogeography-based optimization (BBO) [9], animal migration optimization (AMO) [10], chicken swarm optimization (CSO) [11], grey wolf optimization (GWO) [12], krill herd (KH) [13], and monarch butterfly optimization (MBO) [14]. The bat algorithm (BA) [15] also belongs to the metaheuristics that are based on animal behavior, inspired by the echolocation behavior of bats in nature. On the other hand, several metaphor-based metaheuristics are derived from physical phenomena such as simulated annealing (SA) [16] which is inspired by the annealing process of a crystalline solid. The harmony search (HS) algorithm [17] belongs to this category and is inspired by the process of improvising musical harmonies by musicians in an orchestra.

The aforementioned metaheuristics are classified as stochastic optimization techniques. To avoid searching the whole solution space, they include a randomization component to explore new solution areas. Though these random operators are essential, they can introduce two types of problems. First, if the randomization is too strong, the metaheuristic algorithm might keep moving between candidate solutions, loosely examining each localized region and failing to exploit promising solutions and find the best solution. Second, if the search process is too localized, exploiting the first-found good solutions very well but failing to explore more regions, the algorithm might indeed miss the real optimal solution (called the global optimum) and trap into some local optima.

The perfect balance between *exploitation* and *exploration* is essential to all metaheuristics. In fact, it is whether and how this balance is achieved that distinguishes most metaheuristics from each other and forms a source of new attempts to improve existing algorithms, possibly by hybridizing ideas from more than one metaheuristic strategy. The work in this paper follows this path, focusing on two of the known metaheuristics: the BA and the HS algorithms. The most similar attempt in the literature is the hybrid metaheuristic method of harmony search/bat algorithm (HS/BA) [18]. HS/BA tries to improve the tendency of BA to trap into local optima by adding a pitch adjustment operation in HS serving as a mutation operator during the process of the bat exploration, in an attempt to increase its diversity [18].

Motivated by the mutation operator and by the HS/BA algorithms, we introduce in this work a new hybrid algorithm that improves the diversity of BA by adding a mutation operator. Unlike HS/BA we do not employ the same pitch adjustment as the original HS algorithm. Rather, we employ a custom operator that we consistently found superior to other mutations in our research, during numerous experiments with the algorithms. We name the resulting algorithm the Hybrid Bat Harmony (HBH) algorithm and evaluate its performance compared to the original BA, HS, and HS/BA.

The rest of this article is organized as follows. Section 5.2 introduces the proposed HBH method, while Sect. 5.3 explains the setup of experimental evaluation. Section 5.4 presents and discusses the obtained results, and finally Sect. 5.5 concludes the paper.

5.2 The Hybrid Bat Harmony Algorithm

This section introduces the Hybrid Bat Harmony (HBH) algorithm, which is based on the standard BA [15] and HS [17] algorithms. The main idea behind the new algorithm is to augment the BA with a very effective operator from the HS algorithm. In particular, the principle of pitch adjustment in HS is further modified and fine-tuned to increase the diversity of BA and allow for more mutations in the BA search, in order to jump out of potential local-optima traps.

In the standard BA, all bats use echolocation to sense distance, and they also "know" the difference between prey and obstacles in some way. Bats fly randomly with velocity v_i at position x_i with a frequency f_i, varying the wavelength λ and loudness A0 to search for prey. They can automatically adjust the wavelength (or frequency) of their emitted pulses and adjust the rate of pulse emission $r \in [0, 1]$, depending on the proximity of their target. It is also assumed that the loudness varies from a large (positive) A0 to a minimum constant value Amin. The critical aspect of metaheuristic search mechanisms in solving optimization problems is the correct balance between exploitation and exploration as discussed in the introduction.

The standard BA controls the capability of the exploration process by Eqs. (5.1), (5.2), and (5.3) and the capability of the exploitation process by Eq. (5.4):

$$f_i = f_{\mathrm{min}} + (f_{\mathrm{max}} - f_{\mathrm{min}})\,\beta \tag{5.1}$$

where $\beta \in [0, 1]$ is a random vector drawn from a uniform distribution.

$$v_i^t = v_i^{t-1} + \left(x_i^t - x_*\right) f_i \tag{5.2}$$

where x_* is the current global best location (solution), which is located after comparing all the solutions among all the bats.

$$x_i^t = x_i^{t-1} + x_i^t \tag{5.3}$$

Initially, each bat is randomly given a frequency which is drawn uniformly from $[f_{\min}, f_{\max}]$. For the local search part, once a solution is selected among the current best solutions, a new solution for each bat is generated locally using random walk where $\varepsilon \in [-1, 1]$ is a scaling factor which is a random number, while $A^t = < A_i^t >$ is the average loudness of all the bats at time step t.

$$x_{\text{new}} = x_{\text{old}} + \varepsilon A^t \tag{5.4}$$

Furthermore, the loudness A_i and the rate r_i of pulse emission are updated as follows:

$$A_i^{t+1} = \alpha A_i^t, \qquad r_i^{t+1} = r_i^0 \left[1 - \exp\left(-\gamma t\right)\right] \tag{5.5}$$

where α and γ are constants.

However, BA at times falls in the trap of local optima after a quick convergence into a promising area. The main improvement by adding an operator from HS algorithm is to introduce more mutation into the elements of the current solution, either drawing from features of a previous good solution or from a random distribution according to the value of a random parameter called the harmony memory consideration rate (HMCR), which is inspired from the HS algorithm. Further, depending on another parameter called the pitch adjustment rate (PAR), the algorithm might pull the search back to better position with respect to the best and worst solutions recorded so far, which proves very useful in case the BA traps in a local optimum. The modification of the pitch adjustment operator is listed in Algorithm 1 and in Eqs. (5.6) and (5.7):

$$x_{\text{new}}(d) = x_{\text{old}}(d) + \text{bw} \times (x_{\text{worst}}(d) - x_{\text{old}}(d)) \times (2 \times \text{rand} - 1) \tag{5.6}$$

$$x_{\text{new}}(d) = x_{\text{old}}(d) + \text{bw} \times (x_{\text{old}}(d) - x_{\text{best}}(d)) \times (2 \times \text{rand} - 1) \tag{5.7}$$

In the equations above, x_{new} is a new bat solution, and $x_{\text{old}}$is the current solution, while variables x_{worst} and x_{best} represent the worst and best solutions ever found, respectively. d refers to a single dimension of the solution (an element of the solution vector). The bw is the control parameter called bandwidth, which is an arbitrary distance bandwidth for each generation, and rand is a random uniform real number between [0, 1].

```
Algorithm 1
Modified pitch adjustment operation (mutation operator)
For d = 1: D
If (rand < HMCR) then                              //memory consideration
If (rand < PAR) then                                  //pitch adjustment
x_new(d)=x_old(d)+2bw(x_worst(d)-x_old(d))*(rand-1)        // Eq(6)
Else
x_new(d)=x_old(d)+2bw(x_old(d)-x_best(d))*(rand-1)         // Eq(7)
End if
Else
        x_new(d)=x_min(d)+rand×(x_max(d)−x_min(d))     //random selection
End if
End for d
```

The introduced mutation maintains the attractive features of the original bat algorithm, especially in terms of fast convergence, while allowing the algorithm to make use of more mutation toward a better diversity. Based on the aforementioned analyses, the pseudocode of the HBH algorithm is shown in Algorithm 2.

```
Algorithm 2
Begin
Stage 1: Initialization stage
Initialize the population of NP bats;
Set the generation counter t = 1; define loudness A_i, frequency
f_iposition x_iand the initial velocities v_i;set pulse rate
r_i(i = 1, 2... NP);
Set the parameters and initialize the HM, HMCR, PAR and bw;
Evaluate the quality f for each bat determined by the objective
function (x);
Stage 2: Update stage
While (t < Maximum Generation)
fori = 1: NP (all bats) do
Generate a new solution by adjusting frequency, and updating
velocity and position by (3), (4), and (5);
I f (rand >r_i)
Select a solution among the best solutions;
Generate a local solution around the selected best solution;
End if
Algorithm 1                                  // mutation operator
Generate a new solution by flying randomly
If (rand <A_i& x_i<f (x_*))
Accept the new solution;
Increase r_i&reduce A_i;
End if
Rank the bats and find the current best x_*;
End for
    t = t  + 1;
End while
Process the results and visualize them
End.
```

5.3 Experimental Evaluation

In this section, we lay out the experimental setup through which we have evaluated the proposed algorithm, HBH.

5.3.1 General Setup

5.3.1.1 Hardware and Software Implementation

All the experiments were conducted on a laptop with an Intel Core i5 processor running at 2.4 GHz and 8 GB of RAM. The software implementation of the proposed HBH algorithm was based on the implementation of BA and HS in [15, 17] and the description of HS/BA in [18]. All software is compiled using MATLAB R2009b (V7.9.0.529) running under Windows 7.

5.3.1.2 Compared Algorithms

To put the performance of HBH in perspective and illustrate its merits among similar metaheuristic methods, we compare its performance with the three closest techniques: the original BA and HS algorithms, which are the basic components of the proposed method, and HS/BA, which is the most related previous work on hybridizing BA with HS. This set of comparisons is benchmarked using a group of 14 global optimization functions.

5.3.1.3 Parameters

Table 5.1 lists the set of parameters used in all experiments. The listed parameters include those for the compared methods for the sake of reproducibility.

In all cases, the population size NP was set to 50; function dimension was set to 20, 50, and 100 in three sets of experiments; and the maximum number of generations was 50. To mitigate the impact of randomness in individual runs, we report the results over a 100 implementation runs for each algorithm on each benchmark function (Tables 5.2).

Table 5.1 Set of used parameters in all experiments

Metaheuristic	Parameter	Symbol/abbr.	Value
HBH	Loudness	A	0.95
BA	Pulse rate	r	0.6
HS	Harmony memory consideration rate	HMCR	0.95
HS/BA	Pitch adjustment rate	PAR	0.1
	Bandwidth	bw	0.9

Table 5.2 Benchmark global numerical functions used for evaluating optimization methods

No.	Equation	Low	Up
F01	$f(x) = \sum_{i=1}^{n} \lvert x_i \rvert$	−100	100
F02	$f(x) = \sum_{i=1}^{n} i x_i^4 + \text{rand}[0, 1)$	−1.28	1.28
F03	$f(x) = \sum_{i=1}^{n} \left[x_i^2 - 10\cos 2\pi x_i + 10\right]$	−5.12	5.12
F04	$f(x) = \frac{1}{4000}\sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n} \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$	−600	600
F05	$f(X) = \frac{\pi}{n} \times \left\{10\sin^2(\pi y_1) + \sum_{i=1}^{n-1} (y_i - 1)^2\left[1 + 10\sin^2(\pi y_{i+1})\right] + (y_n - 1)^2\right\} + \sum_{i=1}^{n} u(x_i, 10, 100, 4)$	−50	50
F06	$f(x) = \sum_{i=1}^{n} (x_i - 1)^2 - \sum_{i=2}^{n} x_i x_i - 1$	−100	100
F07	$f(x) = -\sum_{i=1}^{n} \sin(x_i) \sin^{2m}\left(i x_i^2/\pi\right)$	0	3.1416
F08	$f(x) = \sum_{i=1}^{n} x_i^2 + \left(\sum_{i=1}^{n} 0.5 i x_i\right)^2 + \left(\sum_{i=1}^{n} 0.5 i x_i\right)^4$	−5	10
F09	$f(x) = \sum_{i=1}^{n} \lvert x_i \rvert^{i+1}$	−1	1
F10	$f(x) = \left[1 + (x_1 + x_2 + 1)^2\left(19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2\right)\right] \times$ $\left[30 + (2x_1 - 3x_2)^2\left(18 - 32x_1 + 12x_1^2 - 48x_2 + 36x_1x_2 + 27x_2^2\right)\right]$	−2	2
F11	$f(x) = -\cos(x_1)\cos(x_2)\exp(-(x_1 - \pi)^2 - -(x_2 - \pi)^2)$	−100	100
F12	$f(x) = 0.5 + \sin^2\left(x_1^2 - x_2^2\right) - 0.5/\left[1 + 0.001\left(x_1^2 + x_2^2\right)\right]^2$	−100	100
F13	$f(x) = -\sum_{i=1}^{m}\left(\sum_{j=1}^{4}\left(x_j + c_{ji}\right)^2 + \beta_i\right)^{-1}$	0	10
F14	$f(x) = -1/1.94\left[2.58 + \sum_{i=1}^{4} \alpha_i \exp\left(-\sum_{j=1}^{6} A_{ij}\left(x_j - P_{ij}\right)^2\right)\right]$	0	1

5.3.2 Benchmark Function

This paper uses a set of 14 test functions for global numerical optimization. These functions are listed in Table 5.2 alongside their respective equations and properties.

5.4 Results

Table 5.3 lists the optimization results when applying the 14 optimization test functions to HS, BA, HS/BA, and our HBH methods. The listed values are the optimal value of the objective function achieved by each algorithm after iterating 50 generations. The mean values in the table are averaged over 100 runs (each run constitutes 50 iterations) and listed along the standard deviation. The min values, however, are the best results achieved by each algorithm at all. By the "best result," we mean the closest result to the actual optimal value of the function, as per Table 5.2.

It is evident from Table 5.3 that the HBH algorithm can reach a better optimum on average, at least with respect to the set of benchmark functions used in the experiments (HBH has better average results in the case of 12 out of 14 test functions). For ease of recognition, the best average result is marked with bold font and bordered cell. The min values are shaded in gray to identify the absolute best minimum achieved for each function. Note that this value is meaningful because it happened that the minimum achieved values by the algorithms for the selected benchmark functions are closest to the real optimum. With respect to the set of test functions used in our evaluations, HBH could achieve the best result in 12 out of 14 cases.

On another perspective, we also graphed the optimization process of each algorithm (for each benchmark function) as the value of the so-far best solution versus the current iteration. That is, to show the search path in terms of selected best solution per iteration. The curve of this kind is expected to decline overall at a slope that reflects the convergence speed of the algorithm (there is no degradation during the process of any included metaheuristic algorithm, as the best solution is either improved or kept unchanged at all iterations). Therefore, these graphs can be called the convergence plots of the algorithms. Because of the large number of plots, we include hereby a representative sample of the convergence plots in Fig. 5.1, which compares the convergence of HBH with the three most related metaheuristic techniques: BA, HS, and HS/BA.

Figure 5.1 (a–d) shows that the HBH algorithm enjoys not only a superior overall performance in terms of the quality of the found optimal solution but also a faster convergence especially in the earlier stages. Although the starting points of the algorithms are close to each other inthe plots of the four testing functions in the figure, the proposed HBH method does not trap into a quick local optimum, unlike the original BA and HS algorithms, for example.

Table 5.3 The min, mean, and standard deviation of test function values found by HS, BA, HS/BA, and the proposed HBH algorithms, averaged over 100 experimental run. The best mean for each function is marked in bold font and bordered. The min value is the best optimization result found by each algorithm (closest value to the global optimum over all runs) and is shaded in gray. Functions are set with 20, 50, and 100 dimensions

		HS			BAT			HS/BA			HBH		
Function	Dim	Min	Mean	Std. dev	Min	Mean	Std. dev	Min	Mean	Std. dev	Min	Mean	Std. dev
F1	20	9.9E+02	2.1E+03	5.7E+02	1.3E+04	3.8E+04	6.1E+03	4.1E−16	7.1E−01	7.0E+00	7.5E−23	2.5E+01	7.6E+01
	50	1.5E+04	2.1E+04	2.6E+03	3.3E+04	1.2E+05	1.3E+04	4.1E−16	7.1E−01	7.0E+00	0.0E+00	2.0E−01	9.3E−01
	100	6.6E+04	8.4E+04	6.8E+03	5.2E+04	2.7E+05	2.5E+04	1.9E−09	3.7E+01	2.9E+02	0.0E+00	6.0E−01	3.6E+00
F2	20	6.5E+00	8.3E+00	6.4E−01	1.8E+01	4.5E+01	9.2E+00	4.4E+00	5.9E+00	4.8E−01	4.7E+00	5.9E+00	4.3E−01
	50	3.6E+01	5.0E+01	5.4E+00	5.9E+01	3.9E+02	6.2E+01	4.4E+00	5.9E+00	4.8E−01	1.6E+01	1.8E+01	6.2E−01
	100	2.4E+02	3.3E+02	3.9E+01	2.0E+02	1.8E+03	2.3E+02	3.8E+01	4.0E+01	9.4E−01	3.6E+01	4.0E+01	1.0E+00
F3	20	3.0E+01	4.6E+01	8.0E+00	1.2E+02	1.7E+02	2.1E+01	6.3E−13	1.1E+01	1.2E+01	0.0E+00	7.4E+00	1.2E+01
	50	1.7E+02	2.6E+02	2.4E+01	4.4E+02	5.9E+02	3.6E+01	6.3E−13	1.1E+01	1.2E+01	0.0E+00	9.8E−01	6.3E+00
	100	1.7E+02	2.6E+02	2.4E+01	4.4E+02	5.9E+02	3.6E+01	6.3E−13	1.1E+01	1.2E+01	0.0E+00	9.8E−01	6.3E+00
F4	20	1.1E+01	2.0E+01	4.7E+00	1.1E+02	3.5E+02	5.1E+01	1.0E−14	3.9E+00	8.5E+00	0.0E+00	6.0E−01	8.1E−01
	50	1.5E+02	1.9E+02	2.1E+01	2.6E+02	1.1E+03	1.1E+02	1.0E−14	3.9E+00	8.5E+00	0.0E+00	1.3E−01	6.2E−01
	100	5.8E+02	7.5E+02	6.2E+01	6.2E+02	2.4E+03	2.2E+02	1.7E−10	2.9E+00	6.2E+00	0.0E+00	6.1E−02	3.1E−01
F5	20	4.0E+03	2.3E+05	2.3E+05	1.3E+07	2.6E+08	8.6E+07	1.5E−15	1.7E+00	3.3E+00	3.0E−23	6.8E−01	2.1E+00
	50	1.1E+07	3.4E+07	1.0E+07	9.4E+06	1.0E+09	2.1E+08	1.5E−15	1.7E+00	3.3E+00	9.4E−33	2.4E−03	2.0E−02
	100	1.7E+08	3.0E+08	5.8E+07	1.6E+08	2.7E+09	3.7E+08	3.1E−11	3.1E−01	9.4E−01	4.7E−33	7.3E−04	3.6E−03
F6	20	6.2E+02	2.8E+03	9.7E+02	8.8E+03	3.0E+04	5.7E+03	−3.9E+02	−3.8E+02	3.5E+01	−4.5E+02	−3.9E+02	1.9E+01
	50	1.8E+04	2.8E+04	4.3E+03	3.2E+04	1.0E+05	1.3E+04	−3.9E+02	−3.8E+02	3.5E+01	−2.6E+03	−2.5E+03	2.5E+01
	100	8.1E+04	1.0E+05	9.6E+03	6.5E+04	2.4E+05	2.4E+04	−9.9E+03	−9.9E+03	2.3E+01	−1.0E+04	−9.9E+03	1.5E+02

(continued)

Table 5.3 (continued)

Function	Dim	HS			BAT			HS/BA			HBH		
		Min	Mean	Std. dev	Min	Mean	Std. dev	Min	Mean	Std. dev	Min	Mean	Std. dev
F7	20	−1.8E+01	−1.6E+01	5.4E−01	−1.2E+01	−8.6E+00	1.2E+00	−1.1E+01	−8.3E+00	7.4E−01	−1.4E+01	−1.0E+01	1.6E+00
	50	−3.5E+01	−3.1E+01	1.5E+00	−2.3E+01	−1.8E+01	2.2E+00	−1.1E+01	−8.3E+00	7.4E−01	−2.0E+01	−1.7E+01	2.2E+00
	100	−5.4E+01	−4.8E+01	2.1E+00	−3.8E+01	−3.0E+01	2.7E+00	−4.0E+01	−2.9E+01	7.4E+00	−4.0E+01	−2.8E+01	4.3E+00
F8	20	8.9E+01	1.9E+02	4.3E+01	1.5E+02	4.8E+06	2.2E+07	5.0E−12	4.9E+00	3.5E+01	1.6E−19	2.9E+00	2.4E+01
	50	5.1E+02	7.2E+02	1.1E+02	9.8E+02	1.3E+11	1.5E+11	5.0E−12	4.9E+00	3.5E+01	4.4E−38	4.5E−02	3.2E−01
	100	1.3E+03	2.5E+03	5.2E+03	1.7E+04	1.4E+14	8.0E+13	9.3E−03	2.2E+05	1.1E+06	4.0E−32	2.8E+02	2.7E+03
F9	20	2.2E−05	1.3E−03	1.7E−03	6.9E−06	1.8E−02	1.7E−02	1.2E−19	1.6E−11	3.5E−11	1.6E−24	7.1E−12	3.7E−11
	50	5.4E−04	5.1E−03	3.9E−03	7.3E−06	6.6E−02	5.8E−02	1.2E−19	1.6E−11	3.5E−11	3.3E−48	9.7E−07	8.3E−06
	100	1.8E−03	8.6E−03	5.5E−03	2.2E−05	1.1E−01	8.6E−02	6.1E−10	1.8E−05	3.2E−05	7.4E−46	1.8E−07	1.8E−06
F10	2	3.0E+00	1.1E+01	1.3E+01	3.0E+00	1.2E+01	1.8E+01	3.0E+00	9.0E+00	1.1E+01	3.0E+00	8.7E+00	1.5E+01
F11	2	−9.9E−01	−9.5E−02	2.3E−01	−1.4E−05	−1.5E−07	1.4E−06	−1.0E+00	−3.0E−01	4.4E−01	−1.0E+00	−3.7E−01	4.7E−01
F12	2	1.7E−04	8.2E−02	6.4E−02	1.5E−04	2.1E−01	1.1E−01	6.4E−15	7.1E−02	9.4E−02	4.4E−15	4.5E−02	6.8E−02
F13	4	−9.3E+00	−3.7E+00	1.4E+00	−7.7E+00	−1.8E+00	1.1E+00	−1.1E+01	−6.9E+00	3.0E+00	−1.1E+01	−7.6E+00	3.4E+00
F14	6	−3.0E+00	−3.0E+00	2.3E−02	−3.0E+00	−3.0E+00	4.5E−02	−3.0E+00	−3.0E+00	5.9E−02	−3.0E+00	−3.0E+00	3.9E−02

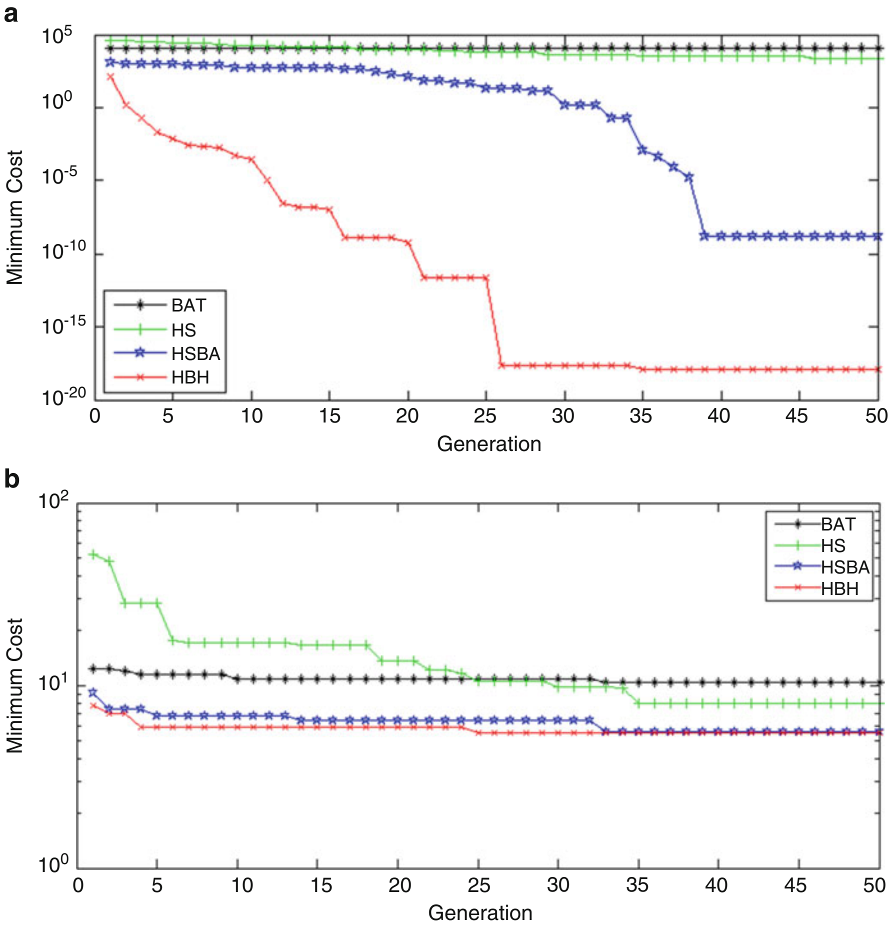

Fig. 5.1 Performance of BA, HS, HS/BA, and HBH algorithms for (**a**) F1 Step, (**b**) F2 Quartic, (**c**) F3 Rastrigin, and (**d**) F4 Griewank benchmark functions

5.5 Conclusion

This paper presented a new metaheuristic, combining the original bat algorithm with a mutation operator to increase its diversity. The introduced operator resembles the pitch adjustment operator from the harmony search metaheuristic but is modified so as to allow for a larger mutation rate while preserving the strength of BA in swift and efficient exploitation. Experimental evaluations against a set of 14 benchmark numerical optimization functions showed that the proposed HBH algorithm converges faster than other metaheuristics and achieves better or at least competitive performance in most cases. We believe that these results are very promising and hope to further explore the specific applications that will benefit from the merits of HBH the most, which is left for future experimentation.

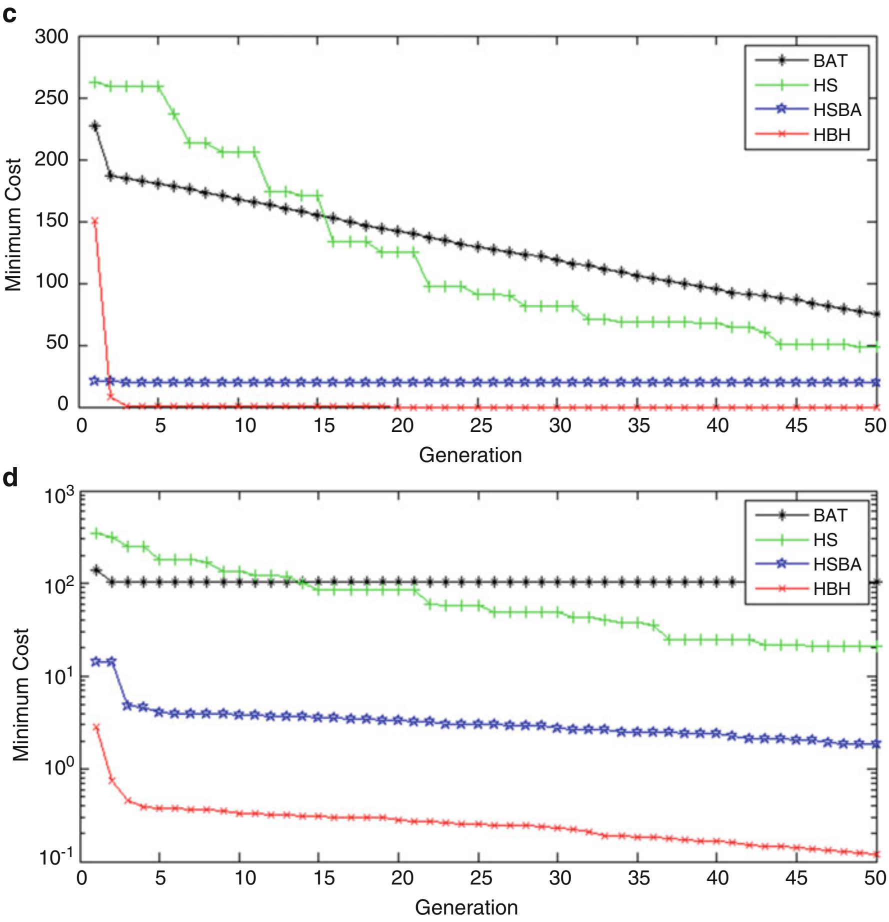

Fig. 5.1 (continued)

Acknowledgments This research was funded by Universiti Sains Malaysia, APEX (308/AIPS/415401) and was also supported by the Fundamental Research Grant Scheme (FRGS) for "Content Based Analysis Framework for Better Email Forensic and Cyber Investigation" [203/PKOMP/6711426].

References

1. Sörensen K, Glover FW (2013) Metaheuristics. In: Encyclopedia of operations research and management science. Springer, Berlin, pp 960–970
2. Eberhart RC, Kennedy J (1995). A new optimizer using particle swarm theory. In: Proceedings of the sixth international symposium on micro machine and human science, New York

3. Dorigo M, Birattari M, Stützle T (2006) Ant colony optimization. IEEE Comput Intell Mag 1(4):28–39
4. Dorigo M, Maniezzo V, Colorni A (1996) Ant system: optimization by a colony of cooperating agents. IEEE Trans Syst Man Cybern Part B Cybern 26(1):29–41
5. Karaboga D (2005) An idea based on honey bee swarm for numerical optimization. Technical report-tr06, Erciyes university, engineering faculty, computer engineering department
6. Karaboga D, Basturk B (2007) A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. JGlob Optim 39(3):459–471
7. Yang XS (2010) Nature-inspired metaheuristic algorithms. Luniver press, Frome
8. Yang XS, Deb S (2009) Cuckoo search via Lévy flights. In: Nature & Biologically Inspired Computing, 2009. NaBIC 2009. World Congress on. IEEE
9. Simon D (2008) Biogeography-based optimization. IEEE Trans Evol Comput 12(6):702–713
10. Li X, Zhang J, Yin M (2014) Animal migration optimization: an optimization algorithm inspired by animal migration behavior. Neural Comput Applic 24(7–8):1867–1877
11. Meng X et al (2014) A new bio-inspired algorithm: chicken swarm optimization. In: Advances in swarm intelligence. Springer, Berlin, pp 86–94
12. Mirjalili S, Mirjalili SM, Lewis A (2014) Grey wolf optimizer. Adv Eng Softw 69:46–61
13. Gandomi AH, Alavi AH (2012) Krill herd: a new bio-inspired optimization algorithm. Commun Nonlinear Sci Numer Simul 17(12):4831–4845
14. Wang G-G, Deb S, Cui Z (2015) Monarch butterfly optimization. Neural Comput Applic:1–20
15. Yang X-S (2010) A new metaheuristic bat-inspired algorithm. In: Nature inspired cooperative strategies for optimization (NICSO 2010). Springer, Berlin, pp 65–74
16. Kirkpatrick S, Vecchi MP (1983) Optimization by simmulated annealing. Science 220(4598):671–680
17. Geem ZW, Kim JH, Loganathan G (2001) A new heuristic optimization algorithm: harmony search. Simulation 76(2):60–68
18. Wang G, Guo L (2013) A novel hybrid bat algorithm with harmony search for global numerical optimization. J Appl Math 2013:1–21

Chapter 6
Consolidation of Host-Based Mobility Management Protocols with Wireless Mesh Network

Wei Siang Hoh, Bi-Lynn Ong, R. Badlishah Ahmad, and Hasnah Ahmad

6.1 Introduction

The Internet consolidated itself as a very powerful platform that has forever changed the way human communicates and their behavior. The mobile communications technology had made it possible for much greater reach of the Internet and increase in the number of Internet users through the mobile devices in wireless environment. The dependency of human toward the Internet has increased the usage of the Internet which causes congestion and intermittent connection issues that rise rapidly. Present telecommunication infrastructures have some coverage issues as it cannot cover specific areas in its coverage area such as alpine area, underground facilities, and forest region. Furthermore, people from rural areas have been receiving limited coverage from the entire major telecommunication service providers. Cost of upgradation of this phase can cause a fortune and it is less profitable based on business and market point of view.

Finding the ways to solve the congestion and intermittent connection problems, Internet Engineering Task Force (IETF) introduces the mobility management protocols. The mobility management is the essential part for mobile devices that automatically connects to the Internet while simultaneously can roam freely without disturbing the communication. Mobility management provides routing support and permits Internet Protocol (IP) nodes using either IPv4 or IPv6 to seamlessly roam among IP subnetworks and media types. Host-based mobility management protocols are in IPv6, and it is also one of the main parts of mobility management protocols. It includes Mobile Internet Protocol version 6 (MIPv6) [1]

W.S. Hoh (✉) • B.-L. Ong • R. Badlishah Ahmad • H. Ahmad
Universiti Malaysia Perlis (UniMAP), School of Computer and Communication Engineering, Arau, Perlis, Malaysia
e-mail: weisiangkelvin1990@gmail.com; drlynn@unimap.edu.my; badli@unimap.edu.my

P. Vasant et al. (eds.), *Modeling, Simulation, and Optimization*,
EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-319-70542-2_6

and its enhancements such as Fast Handover Mobile Internet Protocol version 6 (FMIPv6) [2], Hierarchical Mobile Internet Protocol version 6 (HMIPv6) [3], and Fast Handover for Hierarchical Mobile Internet Protocol version 6 (FHMIPv6) [4].

The wireless mesh network (WMN) is selected as wireless environment that implemented all the host-based mobility management protocols. The WMN can be connected to wireless networks such as Worldwide Interoperability for Microwave Access (WiMAX), generic wireless fidelity (Wi-Fi), and cellular and sensor networks. Third- (3G) and fourth-generation (4G) networks include all Internet Protocol (IP) which are wired, and wireless networks interwork together as heterogeneous networks [5]. However, the challenge is to connect to host-based mobility management. Host-based mobility management protocols rely on the good performance of an infrastructure-based network [6]. However, a typical WMN topology tends to be an unplanned graph and its route dynamically changes [7]. Mobility management provides an undisrupted support of real-time and non-real-time services to mobile network users. Additionally, mobility management also facilitates the maintenance of connections for users on the move when they change their points of attachment from one access point (AP) to another. Host-based mobility allows a mobile node (MN) to change its point of attachment to the network, without interrupting IP packet delivery to or from the node [8]. The current location of all the MNs in the network is maintained by Access Network Procedures [9]. On the other hand, the WMN primary advantages lie in its inherent fault tolerance against network failure, broadband capability, and simplicity of setting up network. The WMN is reliable and offers redundancy [10]. When one node of WMN is failure, the rest of the nodes can still able to communicate with each other, directly or through one or more intermediate nodes.

In this research paper, all the host-based mobility management protocols are investigated firmly on wireless mesh network topology environment. MIPv6, HMIPv6, FMIPv6, and FHMIPv6 are developed and analyzed in wireless mesh network (WMN) environment which are considering the performance parameters: packet delivery ratio (PDR), delay/latency, and throughput. The wireless mesh network (WMN) topology is developed by using network simulation software, and the results obtained are analyzed to agree with the best mobility management protocols to handle the inter-domain mobility with wireless mesh network (WMN) topology.

6.2 Related Works

Jong-Hyouk Lee et al. [11] had investigated a simulation research on analytical comparison of IPv6 mobility management protocols' handover scheme. The researchers compared the host-based mobility management protocols and network-based mobility management protocols to identify the optimized routing protocol for mobile network. The host-based mobility management protocols include Mobile IPv6 and its extensions such as Fast Mobile IPv6 and Hierarchical Mobile IPv6, while network-based mobility management protocols include Proxy Mobile IPv6 (PMIPv6) and Fast Proxy Mobile IPv6 (FPMIPv6). These mobility management protocols have been standardized. The existing IPv6 mobility management protocols are developed by the IETF and have been analyzed and compared in terms of handover latency, handover blocking probability, and packet loss. The conducted analysis results can be used to identify each mobility management protocol's characteristic and performance indicators. The results obtained are used to facilitate decision-making in developing a new mobility management protocol.

K. Vasu et al. [12] had investigated a survey and comparative analysis for MIPv6 protocols. The researchers had performed various mobility management protocols in terms of handover latency, and the number of hops is needed to evaluate these protocols. The IPv6 mobility management protocols such as MIPv6, FMIPv6 (reactive), FMIPv6 (predictive), HMIPv6, PMIPv6, FPMIPv6 (reactive), and FPMIPv6 (predictive) are analyzed and compared in terms of average hop delay, wireless link delay, wired part delay, binding update, and registration delay. PMIPv6 and FPMIPv6 have been compared with the host-based mobility management protocols to make a decision that suits the future networks. The conclusion that the authors made among these protocols is the following: reactive mode protocols perform better in terms of delay compared to predictive-based protocols. The performance is measured in terms of delay during AP to MAG/AR and binding update/registration components. Predictive-based protocols perform better performance in terms of wireless link delay for faster radio access technologies and perform rather slower performance for slower radio access technologies.

J. H. Sun et al. [13] had investigated the mobility management techniques for next-generation wireless networks. The researcher had performed macro and micro mobility protocols in terms of handover performance. The macro and micro mobility protocols such as Mobile Internet Protocol version 6 (MIPv6), Fast Handover Mobile Internet Protocol version 6 (FMIPv6), Hierarchical Mobile Internet Protocol version 6 (HMIPv6), Fast Handover for Hierarchical Internet Protocol version 6 (FHMIPv6), and Proxy Mobile Internet Protocol version 6 (PMIPv6) are analyzed and compared in terms of handover latency. The conclusion that the authors made is that the best handover latency is achieved by FHMIPv6. The result is signaling load reduction, improvement in latency, and less packet losses apart from aiding the handover process.

M. Sko et al. [14] had investigated a simulation research based on analytical comparison of Mobile IPv6 handover schemes. The authors have done a comparison for four most common handover schemes in terms of the cost of packet delivery of Mobile IPv6, that is, MIPv6, FMIPv6, HMIPv6, and FHMIPv6. The access network for using in this research is based on IEEE 802.11b, and the transport core network is Ethernet – IEEE 802.3 100BaseT. The researchers used analytical methods for the comparison. The researchers used network simulation software to run the simulation. The researchers have taken into account these two performance matrices during the comparison, that is, the handover cost and handover latency. The authors concluded that the hierarchical Internet Protocol which consists of HMIPv6 and FHMIPv6 performed well compared to other Internet Protocols in terms of packet delivery cost. The problem of the intra-network issue is not addressed in detail in this research. As a conclusion, the WMN can enhance the Mobile Internet Protocols to perform better in handover latency either in intra-condition or inter-condition compared to the basic wireless network like 802.11b.

C. Makaya et al. [15] had investigated an analytical framework for performance evaluation of IPv6-based mobility management protocols. The researchers have developed an analytical framework for the performance analysis of IPv6 mobility management protocols. MIPv6, FMIPv6, HMIPv6, and a combination of FMIPv6 and HMIPv6 have been compared and evaluated in terms of signaling cost, binding refresh cost, packet delivery cost, required buffer space, and handover latency. The researchers presented the effect of subnet residence time, packet arrival rate, and wireless link delay to the different IPv6 mobility management protocols. The packet delivery ratio, throughput, and the delay are considered in this research.

Shaojian Fu et al. [16] had done an investigation on handover latency in SIGMA, FMIPv6, HMIPv6, and FHMIPv6 protocols. The researchers had designed a new scheme called Seamless IP diversity-based Generalized Mobility Architecture (SIGMA) which can provide low latency and low packet loss mobility. The researchers compared the handover latency of SIGMA with FMIPv6, HMIPv6, and FHMIPv6. The researches have taken into account various parameters such as layer 2 handover/setup latency, layer 2 beacon period, mobile host moving speed, and IP address resolution latency. The software Network Simulator version 2 (NS-2) is used to run the simulation. The researchers concluded handover latency of SIGMA is lower than that of MIPv6 enhancements under various simulated scenarios. SIGMA could also seamlessly handle relatively high-speed movement. Having studied the previous research, this research is not considering the handover latency of SIGMA, but it is considering the latency for network-based and host-based mobility managements. The reason is because cellular network is not one of the considerations of this research.

Yan Zhang et al. [17] had done research on hierarchical Mobile IPv6 with fast handover. In the study, the authors have compared four types of mobile routing protocol to identify the best routing protocol for mobile network, which are MIPv6, FMIPv6, HMIPv6, and FHMIPv6. NS-2 has been used to conduct the simulations. Performance metrics that have been taken into account are handover delay and jitter. This research review has been useful to our research as the authors mentioned about performance metrics jitter and delay which is similar to our initial research plan to consider those performance metrics. Delay is crucial to any performance investigation as it is the most demanding performance metric in the field of networking as speed is everything in networking. At the end of research, it’s been concluded that FHMIPv6 performed extremely better compared to other MIP. The reason of low performance of the other MIP protocols like MIPv6, FMIPv6, and HMIPv6 is not presented in the research. Thus, in this research, investigation is carried out to find the reason of low performance.

M. K. Murtadha et al. [18] have proposed a fully distributed mobility management scheme for future heterogeneous wireless network. The researchers develop distributed mobility management (DMM) scheme based on the PMIPv6 and compare it with the centralized mobility management (CMM) scheme PMIPv6. The proposed approach removes any central anchor node and disables the signaling between MN and the access networks. The performance metrics include the signaling cost, handover latency, and packet loss. The DMM scheme reduced the signaling cost, handover latency, and packet loss compared to the CMM scheme PMIPv6. The main advantages of the DMM scheme are the effectively reduced processing requirement and the power consumption of the MN.

6.3 Terminology

6.3.1 Host-Based Mobility Management Protocols

Host-based mobility management protocols include Mobile Internet Protocol version 6 (MIPv6) and its enhancements such as Fast Handover Mobile Internet Protocol version 6 (FMIPv6), Hierarchical Mobile Internet Protocol version 6 (HMIPv6), and Fast Handover for Hierarchical Mobile Internet Protocol version 6 (FHMIPv6). Host-based mobility management protocols are deployable in wireless mobile communication infrastructures, communication service providers, and standard development organizations [19]. These mobility management protocols have identified that such conventional solutions for mobility service are not suitable, in particular, for telecommunication service. The reason is because the mobile node (MN) is required to perform mobility functionalities at its network protocol stack inside, and thus, modifications or upgrades of the MN are needed. It obviously increases the operation expenses and complexity for the MN. Hence, the extension of Mobile IPv6 (MIPv6) had been introduced to overcome the handover latency problem. The extension of MIPv6 includes HMIPv6, FMIPv6, and FHMIPv6.

6.3.1.1 Mobile Internet Protocol Version 6 (MIPv6)

Internet Engineering Task Force (IETF) brought into use the Mobile Internet Protocol version 6 (MIPv6) to allow mobile nodes (MN) to be reachable and maintain ongoing connection while changing location within topology without changing the allocated IP address [20]. The operation of MIPv6 is illustrated in Fig. 6.1.

The operation begins as MN detects movement to a foreign agent (FA) and autoconfigures itself with a new care-of address (NCoA) using either stateful or stateless method [21]. MN sends binding update (BU) to its home agent (HA) to notify the new address, and HA returns back binding acknowledgment (BAck). Then, all packets are tunneled to MN's NCoA with the help of HA as HA encapsulates packets and sends to MN's NCoA and MN decapsulates the packets received from HA. However, the tunneling also causes drawback due to the long path between mobile node (MN) and correspondent node (CN). This leads the air interface

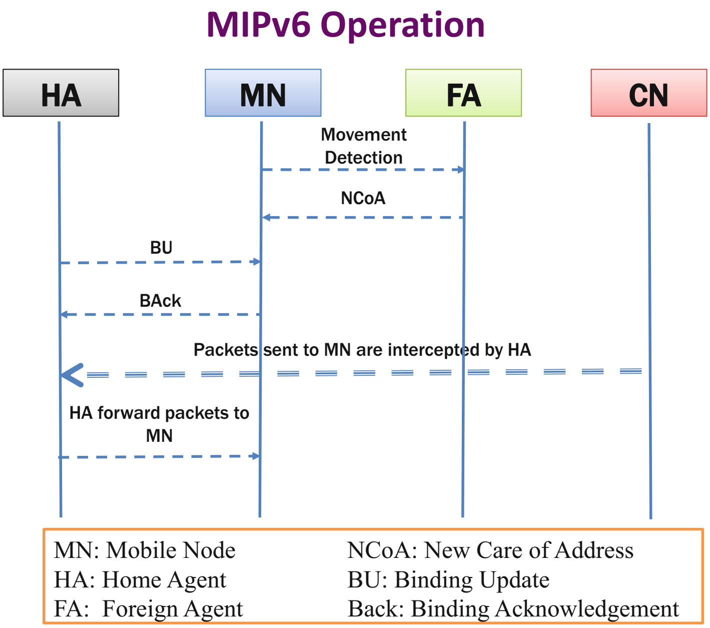

Fig. 6.1 MIPv6 flow diagram

traffic overhead high and it also causes high tunneling overhead at MN. Hence, an additional mode for MIPv6 is route optimization (RO). It allows the datagrams to be delivered using the shortest path. This process requires MN to register its current binding to CN. This allows CN to triangulate datagrams to be delivered to MN without concerning HA. This measure decreases the signaling overhead and handover latency between MN and CN, and it also reduces congestion at MN's HA and home link.

6.3.1.2 Hierarchical Mobile Internet Protocol Version 6 (HMIPv6)

The Internet Engineering Task Force (IETF) has introduced the HMIPv6 based on its predecessor MIPv6 and has implemented new technologies to it to ensure the increment in the performance of mobile networking. One of the latest features in HMIPv6 is mobility anchor point (MAP). The introduction of MAP in HMIPv6 has improved the handover latency and reduced the amount of signaling between the mobile node (MN), its correspondent nodes (CNs), and its home agent (HA) [22]. Figure 6.2 illustrates the process flow of HMIPv6.

Operation of HMIPv6 involves three phases, namely, MAP discovery, MAP registration, and packet forwarding. The first step is the MAP discovery procedure which is to obtain a successful connection. Normally, MAP is a router which is located in a network that is visited by the MN. When the visit point router starts to

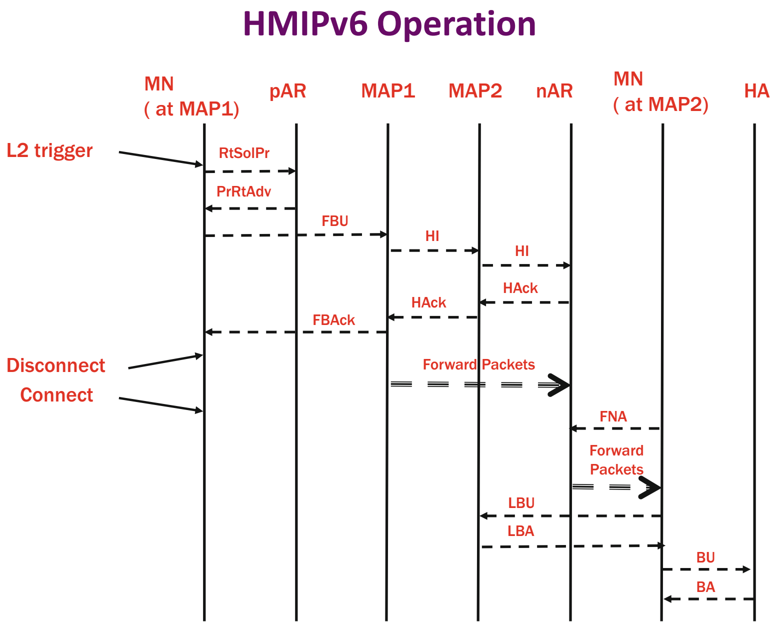

Fig. 6.2 HMIPv6 flow diagram

advertise, the discovery begins. There are two discovery options in HMIPv6: static configuration and dynamic MAP discovery.

The HMIPv6 operation starts when the MN obtains its MAP IP address through the discovery, and then it calculates the distance of MAP from the current AR. This is to verify the connection strength between AR and MAP. The second step is MAP registration that is to register the MN to MAP. Visit point router is assigned as regional care-of address (RCoA) which is obtained by the MN from the visited network. RCoA is formed using the prefix advertised by visit point router. MN assigns as on link care-of address (LCoA). LCoA is configured on a MN's interface based on the prefix advertised by its default router (AR). After that, MN creates a binding between RCoA and LCoA at MAP. Then, MN sends Local Binding Update (LBU) to newly discovered MAP. Next, HA performs duplicate address detection (DAD) and updates the binding cache. Then, MAP sends Binding Acknowledgement (BAck) to MN. After this process, MN sends binding update to its HA and active CNs with RCoA as its source address and HA. CN's address is set as destination address. The final step is to forward packets which are performed after the MAP discovery and MAP registration. A bi-directional tunnel between MAP and MN is established to allow the packet forwarding. All packets sent by the MN are tunneled through MAP, and also all packets destined to the MN's RCoA are intercepted by MAP and tunneled to MN's LCoA.

As simplified to figure out the HMIPv6 mechanism, when an MN moves in MAP domain, MAP acts as HA of the MN locally, and the MN registers only to its location

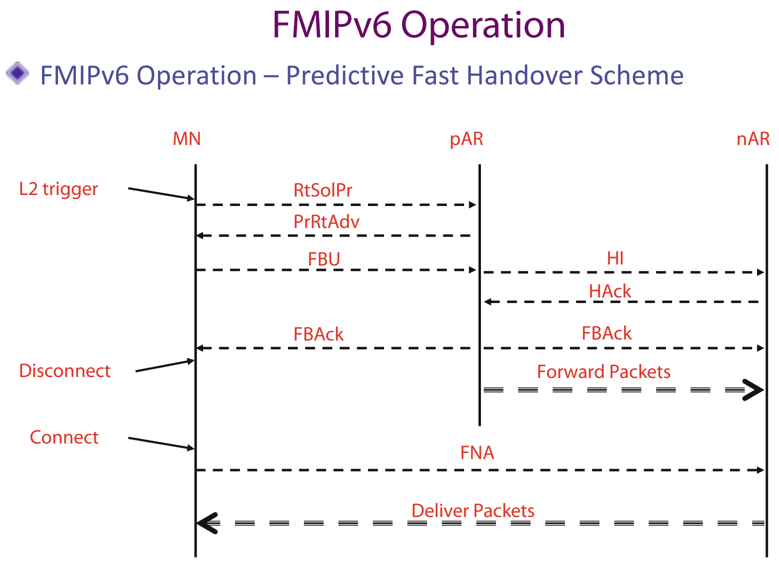

Fig. 6.3 FMIPv6 predictive fast handover scheme flow diagram

information to MAP. Therefore, HMIPv6 has a smaller location update cost than MIPv6 which updates an MN's location information to the HA and CNs. Thus, HMIPv6 is more efficient compared to the previous MIPv6 in terms of the handover and broadcasting. The technology also improves the inter-network connection and smoothen the users' intermittent connections.

6.3.1.3 Fast Handover Mobile Internet Protocol Version 6 (FMIPv6)

FMIPv6 is another initiative by the Internet Engineering Task Force (IETF) to improve the mobile network for the mobile users. The FMIPv6 is also designed based on the previous version of MIPv6. The FMIPv6 handover schemes introduce three signaling messages which involved in the anticipation phase, namely, Router Solicitation for Proxy Advertisement (RtSolPr), Proxy Router Advertisement (PrRtAdv), and Fast Binding Update (FBU) [23]. Figure 6.3 shows the FMIPv6 predictive fast handover flow diagram. Figure 6.4 shows the FMIPv6 reactive fast handover flow diagram.

For FMIPv6 predictive fast handover scheme, the MN obtains the new CoA before the actual movement to new subnet through newly defined messages: Router Solicitation for Proxy Advertisement (RtSolPr) and Proxy Router Advertisement (PrRtAdv) [24]. Next, when MN sends Fast Binding Update (FBU) to the pAR and requests pAR to send a Handover Initiate (HI) messages to the nAR, this is to obtain the new care-of address (NCoA). So all the packets arriving to the previous care-of address (PCoA) can be tunneled to the NCoA. The nAR performs duplicate address

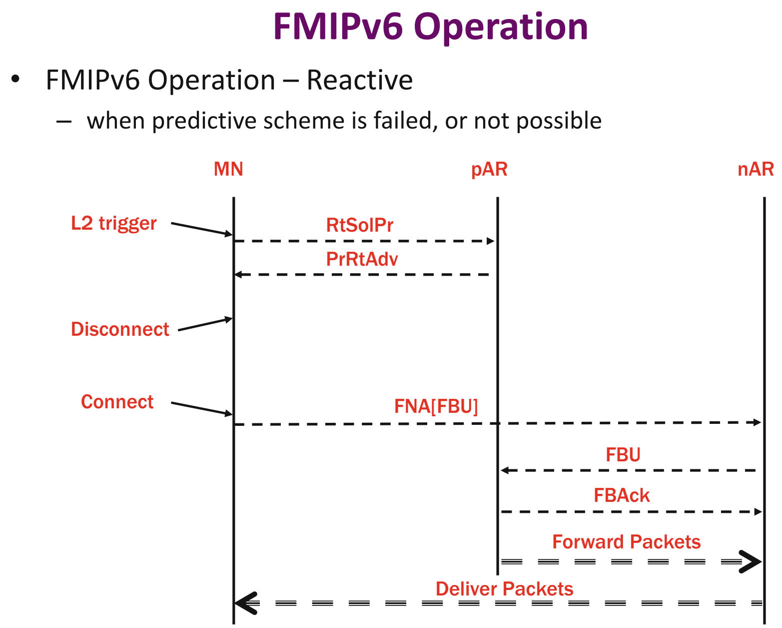

Fig. 6.4 FMIPv6 reactive fast handover scheme flow diagram

detection (DAD) and returns Handover Acknowledgement (HAck) to the pAR with the tunnel establishment. Next, pAR sends Fast Binding Acknowledgement (FBAck) to MN and nAR. In the last step, the MN sends the Fast Neighbor Advertisement (FNA) to the nAR. This informs that the MN is in the nAR subnet and the nAR returns the FNA Acknowledgement (FNA-Ack) to the MN.

When the predictive fast handover scheme is failed or not possible, FMIPv6 can operate in reactive fast handover scheme. The reactive fast handover scheme is almost the same as the predictive fast handover scheme. The main difference is that in predictive fast handover scheme, it allows the MN to send FBU even before it is attached to the nAR. But in reactive fast handover, it only allows MN to send FBU after it is attached to the nAR.

As summary, when MN moves to the new subnet and connect with the new link, it can receive the forwarded packets from pAR. The buffers exist in pAR and new access router (nAR) for protecting packet loss. Therefore, it reduces the service disruption duration and handover latency. The MN must need to update the HA and CNs. As a conclusion, the introduction of the FMIPv6 has minimized the packet loss and latency due to the handover process; thus these have improved the network connection and smoothen the users' intermittent connections.

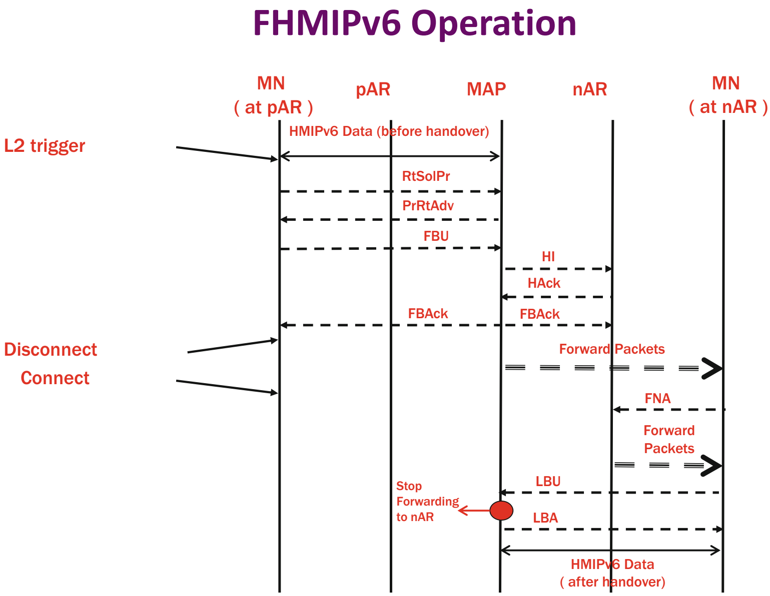

Fig. 6.5 FHMIPv6 flow diagram

6.3.1.4 Fast Handover for Hierarchical Mobile Internet Protocol Version 6 (FHMIPv6)

FHMIPv6 is the combination of two mechanisms that are Fast Handover Mobile Internet Protocol version 6 (FMIPv6) and Hierarchical Mobile Internet Protocol version 6 (HMIPv6). Fast Handover for Hierarchical Mobile IPv6 (FHMIPv6) reduces the signaling overhead and binding update (BU) delay during handover by using HMIPv6 procedures [25]. Furthermore, movement detection latency and new CoA configuration delay during handover are reduced by utilizing FMIPv6 processes.

The FHMPv6 contains the Router Solicitation for Proxy Advertisement (RtSolPr), Proxy Router Advertisement (PrRtAdv), and Fast Binding Update (FBU) technology from the FMIPv6 mechanism and also the mobility anchor point (MAP) technology from the HMIPv6 mechanism which is combined into one single technology, namely, FHMIPv6. When the MN associates with a new MAP domain, HMIPv6 procedures are performed with the HA and the mobility anchor point (MAP). If the MN moves from a previous access router (pAR) to a new access router (nAR) within the domain, it follows the local BU process of HMIPv6. Packets are sent to the MN by the CN during handover which are tunneled by the MAP en route for the nAR. Figure 6.5 shows the FHMIPv6 flow diagram.

Based on the FHMIPv6 operation flow figure, the MN sends RtSolPr message containing the information of nAR to MAP. Continuingly, the MAP sends out

PrRtAdv message to the MN, which contains information of new link care-of address (NLCoA) for MN to be used in nAR region. Next, the MN sends out the Fast Binding Update (FBU) to the MAP, which encloses previous link care-of address (PLCoA) and IP address of the nAR. Once the MAP receives FBU from the MN, MAP sends out Handover Initiate (HI) to nAR. In response to the HI message, nAR sets up a host route for the MN's PLCoA and responds with a Handover Acknowledgement (HAck) message. It means that a bi-directional tunnel between MAP and nAR is established. After that, MAP sends out Fast Binding Acknowledgement (FBAck) toward MN over pAR and nAR. Then, MAP begins to forward data packets destined to MN to the nAR by using the established tunnel. Once the MN is in nAR region, it sends out Fast Neighbor Advertisement (FNA) to the nAR and nAR returns the FNA-ACK to the MN. Then, MN sends Local Binding Update (LBU) to MAP. Next, the HA performs duplicate address detection (DAD) and updates the binding cache. Then, MAP sends a Binding Acknowledgement (BAck) to MN. After this process, MN sends binding update to its HA and active CNs with NLCoA as its source address or HA or CN's address is set as destination address.

6.3.2 Wireless Mesh Network

Wireless mesh network (WMN) is one of the multi-hop infrastructure-based networks. A WMN is a communication network which is made up of radio nodes organized in a mesh topology. The WMN consists of mesh client, mesh router, and gateways. The mesh clients are often laptops, cell phones, and other wireless devices, while the mesh routers forward traffic to and from the gateway, but need not connect to the Internet [26].

The WMN primary advantages lie in its inherent fault tolerance against network failure, broadband capability, and simplicity of setting up a network. Compared to cellular networks, unavailability of communication services over a large geographical area occurred when a single base station (BS) failed [27]. The WMN is reliable and offers redundancy. When one node failed, the rest of the nodes can still communicate with each other, directly or through one or more intermediate nodes. Besides that, the administration and maintenance costs for WMN are lower. In addition, a wireless mesh overcomes the line-of-sight issues that may occur when a space is crowded with buildings or industrial equipments.

A WMN can be seen as a special type of WANET. A WMN often has a more planned configuration and may be deployed to provide dynamic and cost-effective connectivity over a certain geographic area [28]. An ad hoc network, on the other hand, is formed ad hoc when wireless devices come within the communication range of each other. The mesh routers may be mobile and may be moved according to specific demands that are arising in the network. Often, the mesh routers are not limited in terms of resources compared to other nodes in the network. Thus, mesh routers can be exploited to perform more resource-intensive functions. In this way, the WMN differs from an ad hoc network, since these nodes are often constrained by resources.

6.4 Simulation Design

6.4.1 Simulation Setup

To perform a comparison between MIPv6, HMIPv6, FMIPv6, and FHMIPv6 mobility protocols, some configurations and parameters need to be fixed to obtain the optimum results for each mobility management protocol. The environment for all host-based mobility management protocols are set up in wireless mesh network topology environment, and the data rate is fixed in 100 Mbps. Table 6.1 shows the type of parameters and values that are needed to be fixed for the whole simulation process.

The network topology consists of MIPv6, HMIPv6, FMIPv6, FHMIPv6, and WMN. Inter-network section comprises of eight routers with five wired routers and three wireless routers that are accomplished as base stations. Intra-network portion includes nine wireless mesh routers which have been set up in a grid formation to maximize the coverage area. Figure 6.6 shows the simulation design of inter- and intra-network environment for host-based mobility management protocol. Table 6.2 shows configuration details of inter- and intra-network topology of each link and node.

All the host-based mobility management protocols are implemented in the inter-network packet transmission. The inter-network consists of corresponding node (CN), home agent (HA), mobility anchor point (MAP), previous access router (pAR), new access router (nAR), and three routers, namely, N1, N2, and N3. The connection between CN–N1, HA–N1, and N1–MAP is set to 100 Mbps. These represent the local area network with high connectivity. MAP–N1 and MAP–N2 are set to 10 Mbps which represent 10 Mbps Ethernet network. N2–pAR and N3–nAR are set to 1 Mbps which represent Wi-Fi networks.

The network environment needs to be constructed before implementing the mobility management protocols. The intra-network is set up by wireless mesh network (WMN). The WMN topology is in grid formation to fully cover the coverage area which is either scattered or concentrated in a certain geographical terrain. The grid formation WMN topology built up by nine mesh routers and the

Table 6.1 Type of parameters and value

Wireless mesh network data rate	10 Mbps
Window size (byte)	32
Transport protocol	TCP
Link delay	2 ms

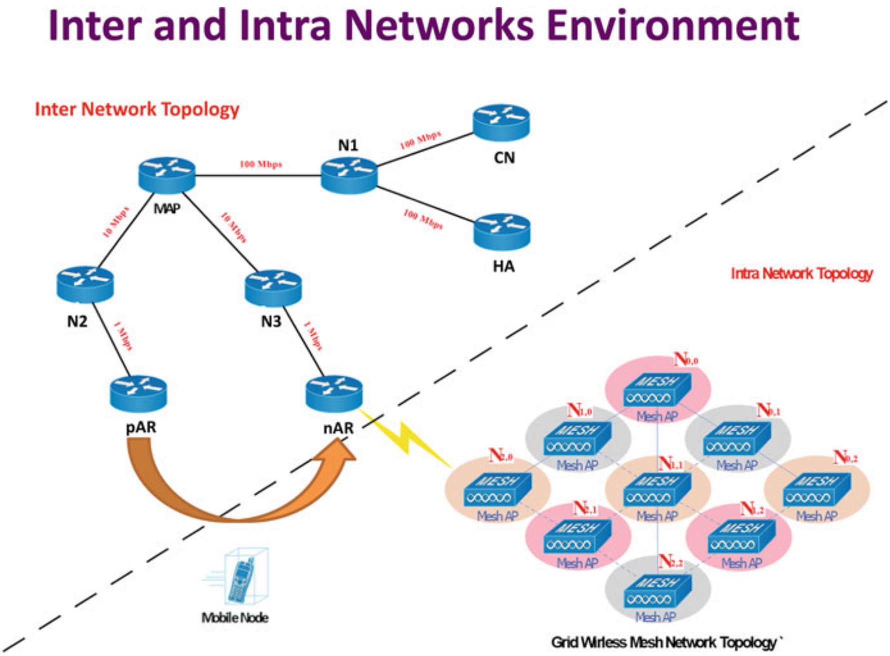

Fig. 6.6 Inter- and intra-network environment

Table 6.2 Configuration details for inter- and intra-network topology

Link connection	Link speed	Queue type
CN – N1	100 Mbps	RED
HA – N1	100 Mbps	RED
N1 – MAP	100 Mbps	RED
MAP – N2	10 Mbps	RED
MAP – N3	10 Mbps	RED
N2 – pAR	1 Mbps	DropTail
N3 – nAR	1 Mbps	DropTail

mobility management protocols act as the mesh client. The bandwidth for WMN is setting up at 10 Mbps, and the data transfer rate among the mesh routers is up to 100 Mbps. The grid WMN topology allows the data transmitted from different devices simultaneously and can withstand high load traffic. The mesh router can exchange and transfer messages directly with neighbors which are located in a vertical or horizontal position. The two mesh routers in the diagonal of a rectangle cannot generally reach each other directly, as obstacles like a tall building or mountain that highly blocks the wireless transmission. For grid formation of WMN network topology, one mesh router act as the main controller to whole topology. The main mesh router, N0, connects with the WLAN controller which works as main controller.

For inter-network, MN is set to move into the connection area of pAR and establish connection with pAR. The pAR belongs to the corresponding home networks. The MN assigned static address from home network, which means that the MN will never change its IP address once it gets that IP address. The HA represents its home default network, and it's also used to tunnel traffic when the mobile device is moving into nAR network when route optimization is not used. The CN keeps communicating with MN through the pAR.

When MN starts moving from pAR toward nAR connection area, MN sends it's binding messages while detecting motion and keeps looking for routing advertisements and solicitation messages from the neighbor router. The MN associates with the nAR. If the association is successful, then the handover process starts. If the association is not successful, the MN repeats the current steps. For MIPv6 and FMIP6, the MAP router is functioned as normal access router, while for HMIPv6 and FHMIPv6, MAP router is functioned as access router that provides new technology, mobility anchor point (MAP) service.

6.4.2 Performance Metrics

In order to get a full understanding about the behavior of all host-based mobility management protocols performed in grid formation of WMN network topology, the following performance metrics were selected for this simulation.

(i) Throughput – represents the average rate of successful packet delivery per unit time over a communication channel.
(ii) Packet delivery ratio (PDR) – represents the ratio between the number of packets received by the receiver and the number of packets sent by the source.
(iii) Latency – represents the delay from the packets sent by the host (computer user) to the server (Internet).

6.5 Results and Discussion

Simulation results are presented in this section. In Table 6.3, the result of each performance metric for all host-based mobility management protocols in wireless mesh network (WMN) environment is shown. The packet sizes used for simulation are started from 256 bytes and increase to 512 bytes, 1024 bytes, and 2048 bytes. The host-based mobility management protocols include MIPv6, HMIPv6, FMIPv6, and FHMIPv6.

Based on the results obtained from the simulation experiments, MIPv6 with WMN performs well in terms of latency mean. The reason that MIPv6 with WMN has low latency is because the packets have been dropped that contribute

Table 6.3 Performance of various types of host-based mobility management protocols in WMN environment

Packet size (bytes)	Latency mean (ms)	Packet delivery ratio (%)	Throughput (bps)
Mobile Internet Protocols version 6 (MIPv6) with WMN			
256	175	74.68	69617.23
512	179	75.22	70860.98
1024	178	74.66	70215.13
2048	183	82.43	91023.06
Hierarchical Mobile Internet Protocol version 6 (HMIPv6) with WMN			
256	201	81.76	77824.00
512	204	82.22	78807.04
1024	202	84.51	80858.21
2048	237	92.65	104133.43
Fast Handover Mobile Internet Protocol version 6 (FMIPv6) with WMN			
256	195	79.87	75243.53
512	200	81.05	77672.90
1024	198	81.51	77987.84
2048	215	89.65	100761.60
Fast Handover for Hierarchical Mobile Internet Protocol version 6 (FHMIPv6) with WMN			
256	200	81.76	77824.00
512	202	82.22	78807.04
1024	201	84.51	80858.21
2048	234	92.65	104133.43

to a low packet delivery ratio and throughput. The highest packet delivery ratio that can be reached is only 82.43%, and in the same time the throughput only reaches 91023.06 bps. HMIPv6 and FHMIPv6 with WMN perform well in terms of throughput and packet delivery ratio (PDR). This is because the mobility anchor point (MAP) is implemented, and the hierarchical movement pattern performs better. The FHMIPv6 with WMN performs slightly better than HMIPv6 in terms of latency mean. FMIPv6 with WMN has a fair performance outcome as it is based on the anticipated handover with a slightly lower packet delivery ratio and throughput compared to HMIPv6 and FHMIPv6.

Having studied Fig. 6.7, it is observed that as the packet size increases, latency mean increases. The reason is because as the packet size increases, the network needs more time to send packets over the Mobile Internet and through Wireless Mesh Network. The handoff latency is also included. However, in Wireless Mesh Network, it has minimized the time for handover process. Additionally, it is observed that MIPv6 performs better compared to HMIPv6, FMIPv6, and FHMIPv6. It is followed by FMIPv6 that performs better than HMIPv6 and FHMIPv6. Although HMIPv6 and FHMIPv6 have been implemented with MAP, FHMIPv6 has slightly lower latency mean compared with HMIPv6. However, latency mean for host-based mobility management protocols does not have many

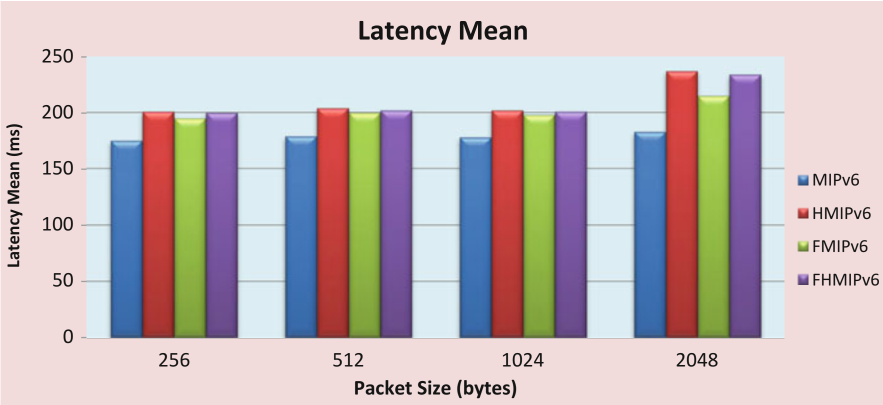

Fig. 6.7 Latency mean chart

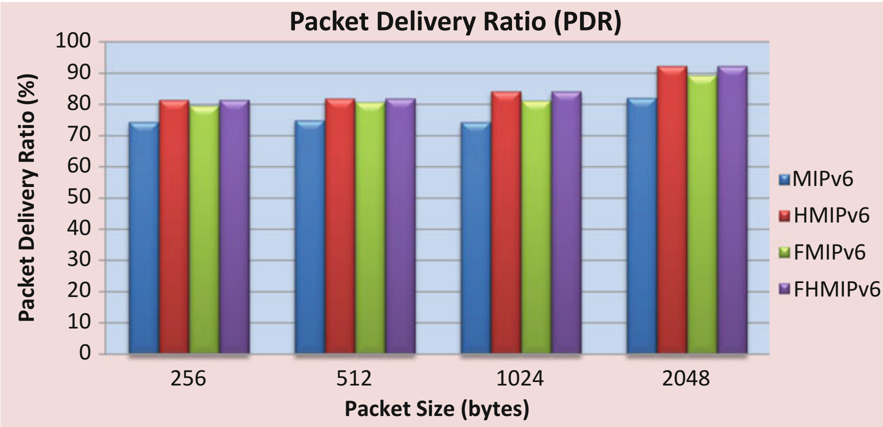

Fig. 6.8 Packet delivery ratio (PDR) chart

differences. Thus, it can be concluded that all host-based mobility management protocols have not much improvement between the mechanisms with the consolidation with WMN.

As shown in Fig. 6.8, the packet delivery ratio of MIPv6, HMIPv6, FMIPv6, and FHMIpv6 for various packet sizes is presented. It can be observed that HMIPv6 and FHMIPv6 perform better compared to MIPv6 and FMIPv6 in terms of packet delivery ratio (PDR). For HMIPv6 and FHMIPv6, both have the same amount for the packet delivery ratio for various packet sizes. MIPv6 performs the least among these four mobility management protocols. The reason is because HMIPv6 and FHMIPv6 perform the MAP mechanism where the same hierarchical network does not need to be sent over to the higher hierarchical network, whereas in FMIPv6 and FHMIPv6, the fast handover mechanism informs the new network about

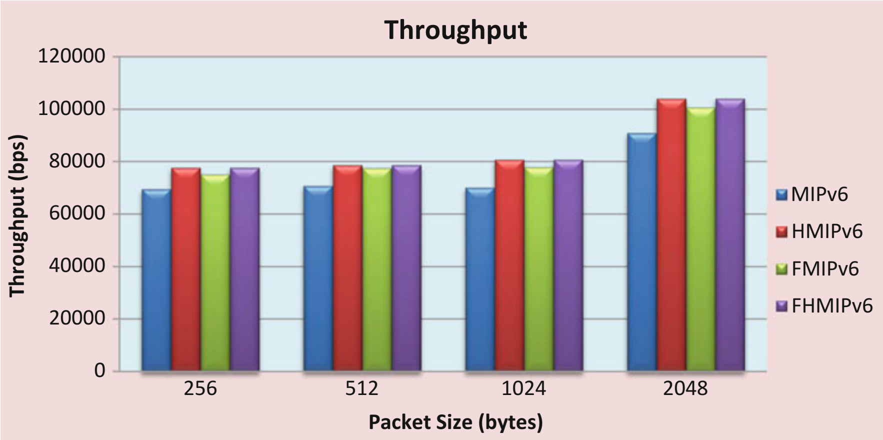

Fig. 6.9 Throughput chart

the handover process before it performs the handover processes. Thus, HMIPv6, FMIPv6, and FHMIPv6 perform better than the original MIPv6.

Based on Fig. 6.9, the throughput for all host-based mobility management protocols has been observed. The HMIPv6 and FHMIPv6 perform better compared to the other two mechanisms. The HMIPv6 and FHMIPv6 share the same throughput for various packet sizes. For example, with the packet size of 2048 bytes, HMIPv6 and FHMIPv6 have the throughput of 104133.43 bps. FMIPv6 has the throughput of 100761.60 bps and MIPv6 has the throughput of 91023.06 bps. The reason why HMIPv6 and FHMIPv6 perform better than FMIPv6 and MIPv6 is explained as before where HMIPv6 and FHMIPv6 do not need to perform higher hierarchical data transmission if the nodes perform lower hierarchical network communication.

By implementing the FHMIPv6 to handle the inter-network scenario, the MAP and fast handover mechanisms are activated. When an MN moves in MAP domain, MAP acts as HA of the MN locally and the MN registers only to its location information to MAP. The MAP minimizes frequent BUs to its HA and CNs, and the BU signaling is localized most of the time. Therefore, FHMIPv6 has a smaller location update cost than MIPv6 which updates an MN’s location information to the HA and CNs. For fast handover mechanism, when the mobile node senses lower signal strength, mobile node advertises to the neighbor network for the need to attach to the new higher signal strength access point. Mobile node informs the new access point of the need to change to the new access point before the process of handover. Thus, movement detection latency and new CoA configuration delay during handover are reduced. With this enhanced mobility management, it is believed that with the FHMIPv6 implementation, this can reduce latency, increase throughput, and decrease distortion. With these criteria, the future aims of wireless communication while moving are made possible which eases the communication between human.

6.6 Conclusion

As a conclusion, due to the various performance metrics results, the FHMIPv6 performs much better compared to other host-based mobility management protocols. By introducing this proposed expansion, it's able to reduce the signaling cost, in order to improve service quality and service range of wireless communication in areas that are affected by coverage problems. In the future, the network-based mobility management protocols such as Proxy Mobile Internet Protocol version 6 (PMIPv6) and Fast Proxy Mobile Internet Protocol version 6 (FPMIPv6) are proposed to be consolidated and make comparisons with host-based mobility management protocols in wireless mesh network (WMN) environment. This simulation research enables the researchers to fully understand the performance of PMIPv6 and FPMIPv6 in inter-network with WMN.

Acknowledgment An acknowledgement is given to MyBrain15 for granting this research. We are deeply grateful to First EAI International Conference on Computer Science and Engineering 2016 (COMPSE 2016) held in Penang Malaysia on November 11–12, 2016. The COMPSE 2016 provides an opportunity to accept and publish our paper. We thank our colleagues from the School of Computer and Communication from Universiti Malaysia Perlis who provided insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations or conclusions of this paper.

References

1. Johnson D, Perkins C, Arkko J (2004) Mobility support in IPv6. Internet Soc., Reston, VA, IETF RFC 3775
2. Koodli R (2005) Fast handovers for Mobile IPv6. Internet Soc., Reston, VA, IETF RFC 4068
3. Soliman H, Castelluccia C, ElMalki K, Bellier L (2005) Hierarchical Mobile IPv6 mobility management (HMIPv6). Internet Soc., Reston, VA, IETF RFC 4140
4. Jung H, Lee JY, Soliman H (2005) Fast handover for hierarchical MIPv6 (F-HMIPv6). IETF, (October), pp 1–14
5. Hui S, Yeung K (2003) Challenges in the migration to 4G mobile systems. IEEE Commun Mag 41(12):54–59
6. Zhang Y, Luo J, Hu H (2007) Wireless mesh networking. Posted at Wireless DevCenter on Jan, (May 2016), pp 507–521
7. Chitedze Z, Tucker W (2012) FHMIPv6-based handover for wireless mesh networks. In: proceedings of the South African telecommunication networks and applications conference, 2012, pp 1–5
8. Shaima Q, Ajaz HM (2014) Mobility management in next generation networks: analysis of handover in micro and macro mobility protocols. Int J Comput Netw Technol 3(3)
9. Ashraf K, Amarsinh V, Satish D (2013) Survey and analysis of mobility management protocols for handover in wireless network. … Conference (IACC), 2013, 2, pp 413–420
10. Li Y, Chen IR (2013) Dynamic agent-based hierarchical multicast for wireless mesh networks. Ad Hoc Netw 11(6):1683–1698
11. Lee J-H, Bonnin J-M, You I, Chung T-M (2013) Comparative handover performance analysis of IPv6 mobility management protocols. IEEE Trans Ind Electron 60(3):1077–1088

12. Vasu K, Mahapatra Sudipta, Kumar CS (2012) MIPv6 protocols: a survey and comparative analysis, Computer Science & Information Technology (CS & IT) 07, pp 73–93. © CS & IT-CSCP 2012
13. Sun JH, Howie D, Sauvola J (2012) Mobility management techniques for next generation wireless networks. Proc SPIE Wirel Mob Commun 4586:155–166
14. Sko M, Klügl R (2011) Analytical comparison of Mobile IPv6 handover schemes. Electrorevue 2(2):22–26
15. Makaya C, Pierre S (2008) An analytical framework for performance evaluation of IPv6-based mobility management protocols. IEEE Trans Wirel Coummn 7(3):972–983
16. Shaojian Fu, Atiquzzaman M (2005) Handover latency comparison of SIGMA, FMIPv6, HMIPv6, FHMIPv6. In: GLOBECOM '05. IEEE Global Telecommunications conference, 2005, St. Louis, MO, pp 3809–3813
17. Zhang Y, Bi H (2012) The simulation of Hierarchical Mobile IPv6 with fast handover using NS2. Procedia Eng 37:214–217
18. Murtadha MK, Noordin NK, Ali BM (2015) Survey and analysis of integrating PMIPv6 and MIH wireless networks. Wirel Pers Commun 82:1351–1376
19. Hoh WS, Muthut S, Ong B-L, Elshaikh M, Warip MNM, Badlishah Ahmad R (2015) A survey of mobility management protocols. ARPN J Eng Appl Sci 10(19):2015
20. Khan RA, Mir AH (2014) Performance analysis of Host Based and Network Based IP mobility management schemes, 6, pp 1798–1803
21. Li J, Zhang P, Sampalli S (2008) Improved security mechanism for mobile IPv6. Int J Netw Secur 6(3):291–300
22. Yadav A, Singh A (2014) Performance analysis and optimization of Hmipv6 and Fmipv6 handoff management protocols. Int J Eng Res 5013(3):305–308
23. Gelogo YE, Park B (2012) Reducing packet loss for Mobile IPv6 fast handover (FMIPv6). Int J Softw Eng Appl 6(1):87–92
24. Muthut Sashikumar, Ong Bi-Lynn, Hoh Wei Siang, Badlishah Ahmad R (2015) Integration of fast handover and Hierarchical Mobile Internet Protocol with wireless mesh network. Aust J Basic Appl Sci 9(25) Special 2015, pp 72–78
25. Ortiz JH, Perea JL (2011) Integration of protocols FHMIPv6/MPLS in hybrid networks. Cyber J Multidiscipli J Sci Technol, J Sel Areas Telecommun (JSAT)
26. Ahmed E, Shiraz M, Gani A (2013) Spectrum-aware distributed channel assignment for cognitive radio wireless mesh networks. Malays J Comput Sci 26(3):232–250
27. Muthut S, Ong B, Adilah N, Zahri H, Ahmad RB (2015) An overview of performance enhancement of FHMIPv6 on wireless mesh network. Int J Futur Comput Commun 4(3):160–164
28. Ghazisaidi N, Kassaei H, Bohlooli MS (2009) Integration of WiFi and WiMAX-mesh networks. In: 2009 second international conference on advances in mesh networks, pp 6–11

Chapter 7
Performance Evaluation of Hot Mix Asphalt Concrete by Using Polymeric Waste Polyethylene

Hakeem Jan, Mohamad Yusri Aman, M. Tawab, K. Ali, and Basit Ali

7.1 Introduction

Asphalt is a mineral aggregate with dark brown to black color, which can be obtained naturally from lakes or artificially from petroleum processing. With increasing road traffic and insufficient maintenance as a result of funding financial constraints, the roads are rapidly deteriorating. This situation has prompted the use of other alternatives, including better-quality materials, cost-effective construction methods, and allocation of adequate funds for road maintenance, improvement, and innovative design. Research has cited various factors that affect flexible course performance like the component properties (binder, aggregate, as well as additive) and their proportion in the mix [1]. It is well established that polymers may be used as one of the additives in asphalt to produce modified asphalt [1, 2].

Moreover, polymers increase the asphalt stiffness as well as ensuring improved temperature susceptibility. According to Awwad and Shbeeb [2] and Coplantz, Yapp, and Finn [3], higher stiffness ensures improved modified asphalt rutting resistance during hot climates besides using relatively asphalt with softer base so as to achieve better performance in low temperature. It is also suggested that polymer-modified binders possess better adhesive together with cohesive properties [3–6]. Polymers may also serve as additives of asphalt concrete mixtures for processing

H. Jan (✉) • M.Y. Aman
Faculty of Civil and Environmental Engineering, Universiti Tun Hussein Onn Malaysia, Batu Pahat, Johor, Malaysia
e-mail: khyberpass08@gmail.com; mdyusri@uthm.edu.my

M. Tawab • K. Ali • B. Ali
Faculty of Civil Engineering SUIT Peshawar, Peshawar, Pakistan
e-mail: tawab4@gmail.com

P. Vasant et al. (eds.), *Modeling, Simulation, and Optimization*, EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-319-70542-2_7

aggregate coating material. Such coatings improve the aggregate surface toughness and thereby ensuring superior engineering properties in production of asphalt mixtures [6, 7].

There is increasing use of asphalt in construction of road pavement to bind aggregates [10]. Nonetheless, the asphalt mixture/coating layer is susceptible to severe temperature like high-temperature rutting to medium-temperature fatigue as well as cracking damages due to low temperature. This explains the use of asphalt mixture modification to ensure improved application. Improved asphalt properties may be achieved through use of suitable starting crude oil and ensuring controlled refinery processes, both of which are very difficult to realize [11]. According to these authors, air blowing hardens asphalt, while fluxing agents/diluent oils may be used to soften it. Moreover, polymer addition ensures significant asphalt quality improvement. This is why asphalt modification with polymers remains the common method for quality improvement, particularly for rheological properties [12].

Apparently, pavements defects like rutting during high temperatures, cracking in regions with low temperature, and others arise due to traffic loads and the inadequate ability of asphalt concrete to resist temperature changes. Moreover, increased traffic factors like heavier loads and higher traffic volume as well as higher tire pressure require road pavements with higher performance, which in turn requires asphalt that is less susceptible to factors such as high-temperature rutting and low-temperature cracking and possesses excellent bonding with stone aggregates. This study evaluated the performance of hot asphalt concrete by use of polymeric waste polyethylene.

With the depletable nature of asphalt binders and the environmental and social hazards of waste plastic, this paper seeks to determine the effectiveness of using polymeric waste polyethylene to modify asphalt mixes and to optimize the penetration values, softening points, fire points, and flash points to make the asphalt mixes more durable and high temperature resistant.

7.2 Materials and Methods

This experiment used polyethylene to modify asphalt mixtures. Black stone was collected and crushed into coarse aggregate for performance evaluation of polyethylene on hot asphalt mixtures. The finer dust (<0.075 mm) was obtained from the aggregate blend of domar sand and used as filler material. During crushing of the stone, a lot of dust was produced, while some stone dust was trapped in crushed stone. To ensure optimum use of stone dust, the dust was mixed with fine sand.

7.2.1 Polyethylene

This experiment used low-density white polyethylene bags from local market as well as domestic wastes. The polyethylene was well cleaned and then shredded into particles of 2–3 mm for recycled polyethylene preparation. To meet the recommended requirements by Rahman, Sobhan, and Ahmed [13], the polythene used had specific gravity of 0.94 and melting temperature of 115 °C.

7.2.2 Bituminous Materials and Preparation

Samples of 80/100 penetration grade binding asphalt were tested in accordance with AASHTO to evaluate various properties of asphalt, namely, flash and fire points, penetration value, as well as softening points. Notably, the asphalt binder properties were kept within the penetrating specification of asphalt grade 80/100.

Asphalt specimens were prepared for evaluating the Marshall stability associated with mixture of asphalt and polyethylene. This was done using test procedure by Bruce Marshall that was advanced by the US engineer corps. Aggregates of about 1200 g were therefore used to prepare their specimen with a diameter of 101.6 mm and a thickness of about 63.5 mm of pure asphalt. In addition, the optimum asphalt content for preparing the mixture was determined by the same method. Notably, the Marshall stability as well as flow test were used to predict the performance measure of stability of the mixture, load carrying capacity, as well as the permanent deformation resistance. Overall, nine test samples were used to determine the content of optimum asphalt binder with varying asphalt proportions.

7.2.3 Aggregates

Black stone was crushed into a coarse aggregate and further broken manually into 25.00 mm downgrade pieces. As stated by the Asphalt Institute [14], all the coarse aggregate had to remain on 2.36 mm sieve. A fine aggregate that passed through 2.36 mm but retained by 0.075 mm sieve was obtained from aggregate blend/coarse sand. To meet the requirements of mineral filler recommended by Rahman, Sobhan, and Ahmed [13], a non-plastic sand sieve with less than 0.075 mm was used, while the apparent filler specific gravity was 2.63.

7.2.4 Gradation of Aggregates

Evaluation of the asphalt concrete mix that is modified with additives in terms of behavior requires a continuously graded bituminous macadam aggregate. However the latter requires uniform aggregate blend gradation so as to achieve a dense mix with air-free controlled optimum content for construction of stable as well as durable flexible pavement [15].

7.3 Results

This study evaluated the performance of hot asphalt concrete based on fire, flash, and softening points as well as the penetration of asphalt. To achieve this, samples of varying grinded plastic (polymer) with asphalt. The results obtained are presented in detail in the following sections.

7.3.1 Penetration Value

The results on comparison between the polymer added and penetration value showed an inverse variation with 0% of polymer added at highest penetration value of 85 and 16% polymer added at penetration value of 0. The penetration value decreased gradually with an increase in the percentage added in asphalt up to a penetration value of 0 with a polymer percentage of 16% (Table 7.1 and Fig. 7.1). The mean penetration value 44.111 ± SD = 27.980 and Std. Error = 9.327.

Table 7.1 Comparison between polymers added (%) and penetration value

Polymer added in asphalt (%)	Penetration value
0	85
2	70
4	64
6	57
8	46
10	38
12	24
14	13
16	0

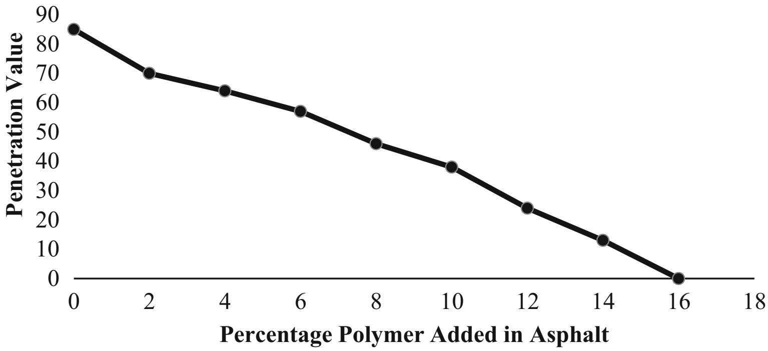

Fig. 7.1 Comparison between polymers added (%) and penetration value

Table 7.2 Comparison between polymers added (%) and softening point

Polymer added in asphalt (%)	Softening point (°C)
0	55
1	57
2	61
3	65
4	68

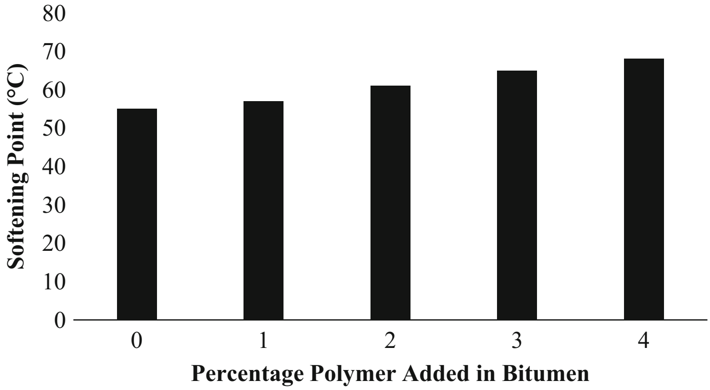

Fig. 7.2 Comparison between polymers added (%) and penetration value

7.3.2 Softening Points

The percentage polymer in asphalt was found to be directly proportional to the softening point of the asphalt. Asphalt with 0% polymer had a softening point of 55 °C, whereas asphalt with 4% polymer had a softening point of 68 °C (Table 7.2 and Fig. 7.2). The mean softening point was ± SD = 27.980 and Std. Error = 9.327.

7.3.3 Percentage of Polymer Added, Fire Point, and Flash Point

When percentage of polymer added, fire point, and flash point are compared, the results showed that the percentage polymer added in asphalt was directly proportional to both the fire point and the flash point. Pure asphalt had flash and fire points of 238 °C and 246 °C, respectively. At 1% polymer added, the flash and fire points were 241 °C and 252 °C, while at 4% polymer added, the flash and fire points were 268 °C and 274 °C, respectively. The mean fire point was 260. ± SD = 11.649 and Std. Error = 5.210. On the other hand, the mean flash point was 250.8 ± SD = 12.194 and Std. Error = 5.453. ± SD = 11.649 and Std. Error = 5.210 (Table 7.3 and Fig. 7.3).

Table 7.3 Percentage of polymer added, fire point, and flash point

Polymer (%) added in asphalt	Flash point(°C)	Fire point (°C)
0	238	246
1	241	252
2	250	263
3	257	269
4	268	274

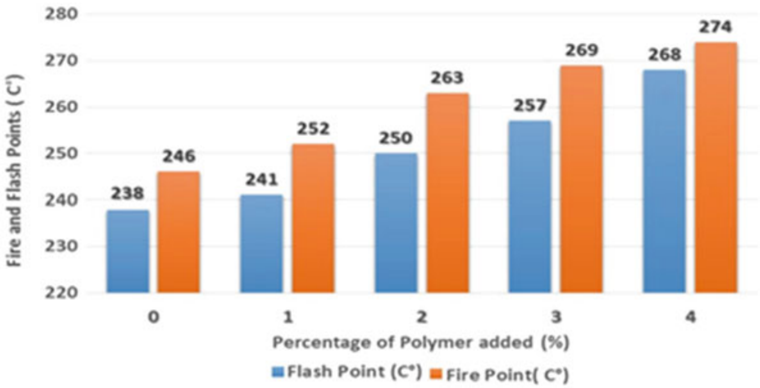

Fig. 7.3 Comparison between polymers added (%) and penetration value

7.4 Discussion

Asphalt is used to construct flexible pavements by binding the aggregates together through coating them. In addition, the material promotes the strength as well as the life of the road pavement. However, due to its poor properties like resistance toward water, it is commonly improved by modifying its rheological properties by mixing with synthetic polymers such as rubber and plastics. Since the use of waste plastic with asphalt is similar to using polymer-modified asphalt, this experiment used waste plastic to evaluate performance of hot asphalt concrete based on fire, flash, and softening points as well as penetration of asphalt.

This study found a decrease in penetration value with increasing polymer content in asphalt, suggesting harder modifying binder. The penetration value of asphalt is a measure of the consistency of the asphalt. The consistency of the asphalt enables the engineer to determine the degree of resistance that the asphalt will have to continuous deformation under stress. From the experiment, the inverse relationship between the percentage polymer added in asphalt and the penetration value indicates that with an increased percentage of polymer in the asphalt, the amount of permissible stresses on the asphalt reduces significantly. These findings support the findings by Justo and Veeraraghavan [16] that the values of penetration and ductility of modified asphalt reduce with increasing plastic content up to a weight of 12%. In another study by Sabina, Khan, Sharma D K, and Sharma B M [17], performance evaluation of bituminous mixes with plastic (8%) and asphalt (15%) was compared to mixture of conventional bituminous concrete (i.e., penetration grade asphalt with 60/70). The findings showed improved properties such as Marshall and retained stability to indirect tensile strength as well as rutting with polymeric modified bituminous concrete. As stated by Imtiyaz [18], the modified mixes are resistant to permanent deformation in elevated temperatures.

The results showed an increase in the softening point value with increasing polyethylene content. This is explained by the higher softening temperature associated with polyethylene. Of note, higher softening points are preferred in warm climates since they are associated with lower temperature susceptibility. These findings are in agreement with a previous study by Justo and Veeraraghavan (2002) that the softening point of asphalt mix increases with an increase in the plastic content up to a weight of 8.0%. Waste plastics are thermoplastic polymers and they are non-biodegradable. These thermoplastic polymers have no significant use when disposed at dumping sites rather than being recycled into different plastic products, which are still pollutants to the environment. As stated by Vasudevan and Rajasekan [19], these polymeric materials improve the binding properties of asphalt than conventional asphalt for pavement construction. Moreover, the use of blended asphalt with waste plastic improves the properties of blended asphalt and addresses challenges associated with plastic waste disposal common with many nations and thereby improves the sanitation systems and employment creation for the collectors.

Moreover, the blend increases the softening point but reduces the penetration value with eventual suitable ductility. This view is supported by Punith [20], who reported the likelihood of improving the performance of modified bituminous road pavements. This is attributable to the lack of gas evolution between 130 and 180 °C, yet waste plastics soften on heating at around 130 °C in order to confer the binding property to asphalt. This explains why it is used for binding purposes in road construction. This study found an increase in fire and flash points of the modified asphalt increased with increasing polythene content. This suggests low propensity of hazardous situation since the inflammability of asphalt decreases with increasing polymer content. Put in other words, the blended asphalt has improved resistance to burning, implying that the road with polymer-asphalt blended surfaces is less likely to be affected by fire hazards. Studies have shown that polymer-modified asphalt is highly resistant to temperature and water, explaining why it is used for construction of flexible road pavements [21].

According to Afroz et al. [22], polymer-modified asphalt exhibits better rheological properties essential for construction of highways. For instance, asphalt ensures flexible pavement construction by binding the aggregates together through coating their surface. This coating results in improved strength as well as life of the road pavements. However, asphalt has poor water resistance compared with polymeric modified asphalt which has better temperature and water resistance. Zoorob and Suparma [23] reported that recycled plastics predominantly comprise of low density. Polyethylene (LDPE) plus polypropylene (PP) when used for conventional bituminous concrete mixtures improves durability together with fatigue life. In their research, Apurva and Chavan [24] found that blended polymer-asphalt helps to promote better asphalt binding of plastic coated aggregate as a result of increased bonding and surface contact area between polymeric materials and asphalt. Moreover, the polymer coatings reduce voids and therefore prevent absorption of moisture and asphalt oxidation by entrapped air that collectively cause reduced rutting, raveling, and pothole formation.

In another study by Sreedevi and Salini [25] to explore pavement performance of roads surfaced with bituminous mixtures and coated aggregates, it was concluded that using waste polymeric materials for road construction may prevent environmental pollution, increase the road service life, reduce petroleum product consumption, and generate income to the society through job creation. Evidently, blended asphalt with waste plastics has improved rheological properties. However, the evaluation of the binder performance relies heavily on the proportion of polymeric material blended asphalt. Other benefits associated with the use of polymeric materials in pavement construction are reduction of waste plastics in the environment with eventual significant reduction in environment pollution and improved hygiene and sanitation systems and creation of employment for garbage collectors.

7.5 Conclusion

This study found that waste plastic materials may be effectively used to modify asphalt for construction of flexible pavements. This may be achieved by mixing processed waste plastic of varying proportions by weight of asphalt to ensure substantial improvement in the Marshall stability, strength, and other related properties of the modified bituminous concrete in order to achieve longevity and better performance of pavement with minimal asphalt usage. The process is friendly to the environment. However, the study recommends further studies to ascertain the exact proportions of the mix that achieve the best modified bituminous concrete.

7.6 Acknowledgment

The authors of this study acknowledge the Centre for Graduate Studies (CGS) of UTHM, Malaysia, that contributed to the success of this research work by providing technical and financial support.

References

1. Subagio BS, Kosasih DB, Tenrilangi D (2005) Development of stiffness modulus and plastic deformation characteristics of porous asphalt mixture using tafpack super. Proc East Asia Soc Trans Stud 5:803–812
2. Awwad MT, Shbeeb L (2007) The use of polyethylene in hot asphalt mixtures. Am J Appl Sci 4(6):390–396

3. Coplantz JS, Yapp MT, Finn FN (1993) Review of relationships between modified asphalt properties and pavement performance. SHRP-A-631, Strategic Highway Research Program, National Research Council, Washington, DC
4. Newman K (2004) Polymer-modified asphalt mixtures for heavy-duty pavements: fatigue characteristics as measured by flexural beam testing. FAA Worldwide Airport Technology Transfer Conference, Atlantic City
5. Hansen KR, Robert BM, Brian P, Stonex A (n.d) Current and future use of non-bituminous components of bituminous paving mixtures. Committee on Characteristics of Non bituminous Components of Bituminous Paving Mixtures: A2D02. Onlinepubs.trb.org/onlinepubs/millennium/00079.pdf
6. Robin LS (1994) The use of recycled materials in highway construction. J Public Road 58:2
7. Amit G, Zamre GS, Renge VC, Bharsakalea GR, Tayde S (2012) Utilization of waste plastic in asphalting of roads. Sci Rev Chem Commun 2(2):147–157
8. Gawande A, Zamare G, Renge VC, Tayde S, Bharsakale G (2012) An overview on waste plastic utilization in asphalting of roads. Eng Res Stud J 3(2):1–5
9. Jabarin AS (1996) Polyethylene terephthalate chemistry and preparation. In: The polymeric materials encyclopedia. CRC Press Inc, Boca Raton
10. Akmal N, Usmani AM (1999) Application of asphalt-containing materials. Polym News 24:136–140
11. Yuonne B, Yajaira MPMR (2001) Polymer modified asphalt. Vis Technol 9(1):39–48
12. Uranga OG (2008) Rheological property of asphalt modified with polyethylene and polyethylene based blends
13. Rahman MN, Sobhan MA, Ahmed TU (2012) Effects of waste polyethylene on bituminous mixes. Proceedings of the 1st International Conference on Civil Engineering for Sustainable Development, Khulna
14. The Asphalt Institute (1984) Mix design methods for asphalt concrete and other hot-mix types, 6th edn. Asphalt Institute, Lexington
15. Sobhan MA, Mofiz SA, Rasel HM (2011) Effect of gradation and compactive effort on the properties of bituminous mixes with waste concrete aggregates. Int J Civ Environ Eng 11(4):18–21
16. Justo CEG, Veeraragavan A (2002) Utilization of waste plastic bags in bituminous mix for improved performance of roads. Centre for Transportation Engineering, Bangalore University, Bangalore
17. Sabina, Tabrez AK, Sangita, Sharma DK, Sharma BM (2009) Performance evaluation of waste plastic/polymer modified bituminous concrete mixes. J Sci Ind Res 68:975–979
18. Mohd I (2002) Adhesion characteristics study on waste plastics modified asphalt. Int J Technol Res Eng 2(9):2077–2080
19. Vasudevan R, Rajasekaran S (2007) Utilization of waste plastics in construction of flexible pavements (Reuse of waste plastics – a path – breaking initiative)
20. Punith VS (2001) Study of the effect of plastic modifier on bituminous mix properties; Mohd. Imtiyaz (2002) Adhesion characteristics study on waste plastics modified asphalt
21. Dr. Gupta YP, Shailendra Tiwari, Pandey JK (2010) Utilisation of plastic waste in construction of bituminous roads. NBM & CW, p 92
22. Afroz Sultana SK, Prasad KSB (2012) Utilization of waste plastic as a strength modifier in surface course of flexible and rigid pavements. Int J Eng Res Appl 2(4):1185–1191
23. Zoorob SE, Suparma LB (2000) Laboratory design and investigation of the properties of continuously graded asphalt concrete containing recycled plastics aggregates replacement (Plastiphalt). Cem Concr Compos 22:233–242
24. Apurva CJ (2013) Use of plastic waste in flexible pavements. Int J Appl Innov Eng Manag 2(4):540–552
25. Sreedevi BG, Salini PN (2013) Pavement performance on roads surfaced using bituminous mixtures with coated aggregates. Int J Eng Res Technol 2(9):149–115

Chapter 8
Evaluation of Technical Efficiency of Thermal Power Units in Mexico: Data Envelopment Analysis and Stochastic Frontiers

Román Rodríguez-Aguilar and Jose Antonio Marmolejo-Saucedo

8.1 Introduction

The present work develops an application of different methodologies for the estimation of technical efficiency, for a set of thermoelectric units in Mexico. As part of the recent reform in the Mexican electricity sector, the electricity market has been liberalized, so that the public supplier before the reform must guarantee competitiveness in order to participate in the electricity market. One of the key decisions for the public supplier is to define through robust methodologies the most efficient power plants and to reform their portfolio of generating units to compete in the best possible conditions in the electricity market. In the context of a liberalized market, it is necessary to take into account that it is not only important to determine the most efficient plants in a product input logic, it is also necessary to consider those plants that are more efficient in terms of costs and those will allow to offer energy at competitive prices with the lowest operating cost.

The work developed compares the application of different methodologies: data envelopment analysis (DEA) and stochastic frontiers (SF). Each methodology has pros and cons; for that reason the comparison of the results will allow to select the one that is the most robust possible and that adapts more to the Mexican case. It has been documented that DEA present weaknesses in those cases in which the information of the productive units contain outlier values, and this affects the

R. Rodríguez-Aguilar (✉)
Facultad de Ingeniería. Universidad Anáhuac, Naucalpan, Mexico
e-mail: roman.rodriguez@anahuac.mx

J.A. Marmolejo-Saucedo
Panamerican University, Mexico, Mexico
e-mail: jmarmolejo@up.edu.mx

P. Vasant et al. (eds.), *Modeling, Simulation, and Optimization*,
EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-319-70542-2_8

efficiency levels estimated for all units. In the case of stochastic frontiers, the main restriction is that it is necessary to assume a specific production or cost function form.

Information for the electricity sector of Mexico of the period 2009–2013 was used only for thermoelectric power plants, since they represent the 86% of the electric power generation in Mexico. It is therefore of great importance to identify which plants are more efficient in generation and costs. It was considered a universe of 22 power plants which are still in operation in Mexico. The results of the estimated models show that the thermal generation in Mexico is very costly due to the inefficiency of the plants, for this reason a set of plants that are in a range of acceptable minimum efficiency.

The work is structured as follows: The first section describes the operation of the Mexican electricity sector before the energy reform and subsequently outlines the new framework of operation generated by the liberalization and creation of the wholesale electricity market, as well as the main instruments through which the market will operate. In the second section, the methodological framework is developed briefly analyzing the two methodologies to be compared. The third section presents the results of the analysis performed for the two methodologies. The final section presents the conclusions of the study.

8.2 Power Generation in the Mexican Electricity System

The Mexican electricity sector operated with a parastatal company as sole supplier, the Federal Electricity Commission (FEC). In addition, other generators participated in the sector through self-generation schemes, independent producers, cogeneration, small producers, export units, and continuous own uses (Table 8.1).

As can be seen in Table 8.1, the CFE and Independent Energy Producers (IEP) accounted for 86% of the national gross generation. The objective of the liberalization of the electricity sector is to open the market to private investment by generating lower energy prices, as well as the diversification of the energy matrix allowing the promotion of clean energy.

Table 8.1 Gross generation by type of generator (2014)

Generator	GWh	%
Small producers	212	0.0
Own uses	1002	0.0
Exports	7050	2.0
Cogeneration	15,258	5.0
Self-generation	19,154	6.0
Independent power producers	87,589	29.0
CFE	172,541	57.0
Total gross generation	302,806	100

Source: Ministry of Energy

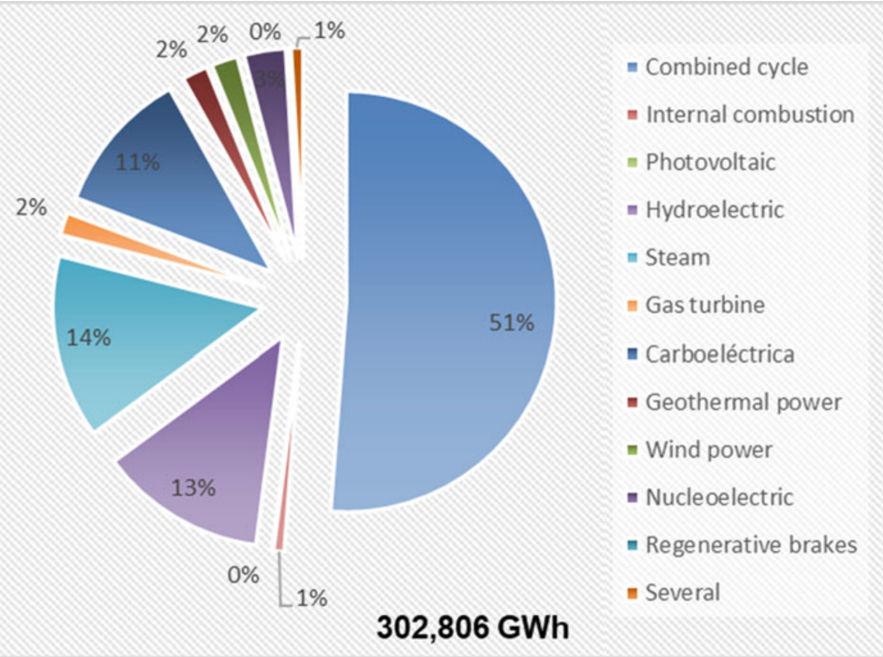

Fig. 8.1 Gross generation by type of technology (2014) (Source: Ministry of Energy)

Table 8.2 Gross generation by primary fuel source (2014)

Primary fuel source	GWh (CFE + PIE)	%
Gas	143,172	55.1
Fuel oil	26,199	10.1
Diesel	1429	0.5
Carbon	33,437	12.9
Hydroelectric	38,145	14.7
Nucleoelectric	9677	3.7
Geothermal power	6000	2.3
Wind	1897	0.7
Photovoltaic	12.7	0.0
Total	259,968	100

By 2014, the total net electricity generation in the Mexican electricity sector was 302, 806 GWh; the energy matrix is concentrated in technologies based on the use of fossil fuels, such as the combined cycle and generation by internal combustion (Fig. 8.1).

Of the total CFE and IEP generation, 65.7% use as primary fuel fossil fuels, 13% coal, and 15% hydro, and the rest is distributed between clean energy sources such as geothermal, nuclear, wind, and photovoltaic (Table 8.2).

The use of fossil fuels for power generation has been severely questioned due to the production of carbon dioxide (CO_2) which contributes to the accumulation of greenhouse gases (GHGs) emitted into the atmosphere.

Therefore, there are two prospective scenarios of capacity expansion for the period 2013–2027. The first is the planned expansion of the public power generation companies with a share of 31.9% with the use of clean technologies in 2027. An 18.4% is of hydro capacity, 4.1% wind power, nuclear 1.8%, and 2.4% remaining with geothermal, solar, and biogas capacity. In the alternative scenario, the expansion program is aligned to the goals set in local laws, seeking to increase its generation with non-fossil sources to a 35% in 2027 [1].

This means that, although fossil fuel generation will be reduced, its use will be located in the worst case in 65% in 2027 [2]. Therefore, the total elimination of fossil fuels for power generation is nearly impossible, because it still represents one of the economic lower cost options in Mexico. Figures 8.1, 8.2, and 8.3 show the behavior of thermal power generation, fuel oil consumption, and useful existence (stock of fuel oil) from 2009 to 2013.

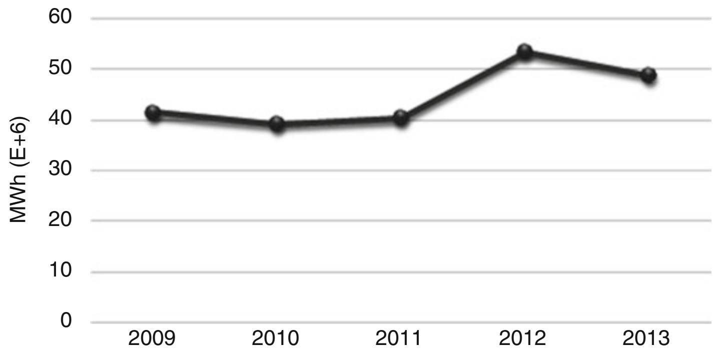

Fig. 8.2 Power generation per year with fuel oil (Source: Salazar, M.S. 2015)

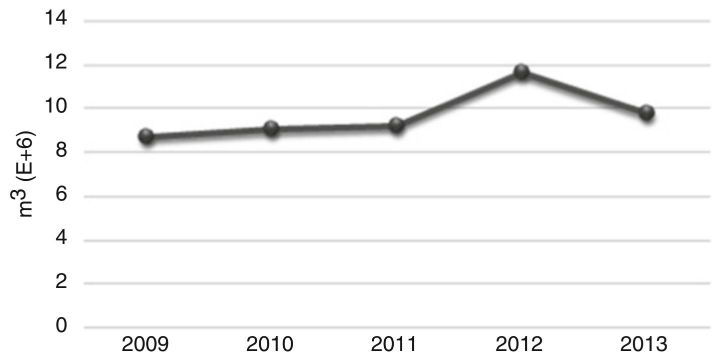

Fig. 8.3 Fuel oil consumption per year (Source: Salazar, M.S. 2015)

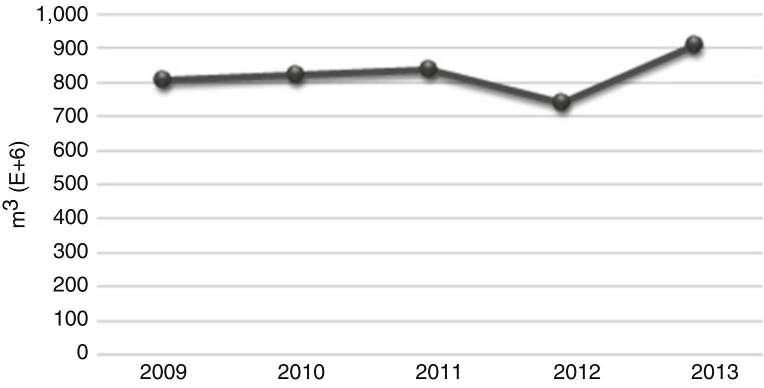

Fig. 8.4 Useful existence (stock) per year (Source: Salazar, M.S. 2015)

The generation of electric energy based on fuel oil represents the most expensive form of electric power generation.

As expected, the consumption of fuel oil has a similar behavior to generation, due to the context of a non-liberalized sector, which was the case of Mexico until 2016. Where the main objective of the parastatal company providing the electric service was to maintain the supply at all times and avoid load cuts at any cost, this is not compatible with a liberalized market; therefore, it is of utmost importance to determine a portfolio of thermal power plants that maximize generation at the lowest cost.

One of the major problems of having a generator focused on ensuring the supply is that it must necessarily maintain levels of fuel inventory large enough to guarantee generation and avoid load cuts. However, this decision involves assuming the cost of maintaining an inventory of these characteristics (Fig. 8.4).

Figure 8.5 shows that the behavior of the fuel oil consumption of power plants over time is directly related to the behavior of generation; i.e., power generation is a function of the consumption. Both behaviors follow the same trend; however they do not match on the same points throughout the timeline; i.e. there is a difference between consumption and generation. Since in some years, the points almost coincide and graphical difference is very low, but in other years, this differential increases and the points are away. It could be concluded that when the points of consumption and generation are far apart, we have a period of technical inefficiency.

In 2013 the production of generation per unit volume (m^3) increased, the lowest difference it can be seen. A total of 4.9 MWh/m^3 was generated in 2013; while in 2010 was of 4.3 MWh/m^3. Ideally, the points must coincide throughout the timeline, i.e., 100% of technical efficiency.

Regarding the level of fuel oil inventory (useful existence), we can mention that it is a policy relating to reducing the risk of not meeting the demand of electricity. The public power generation company must ensure at all times that the fuel oil supply to the thermal plants is guaranteed for any contingency.

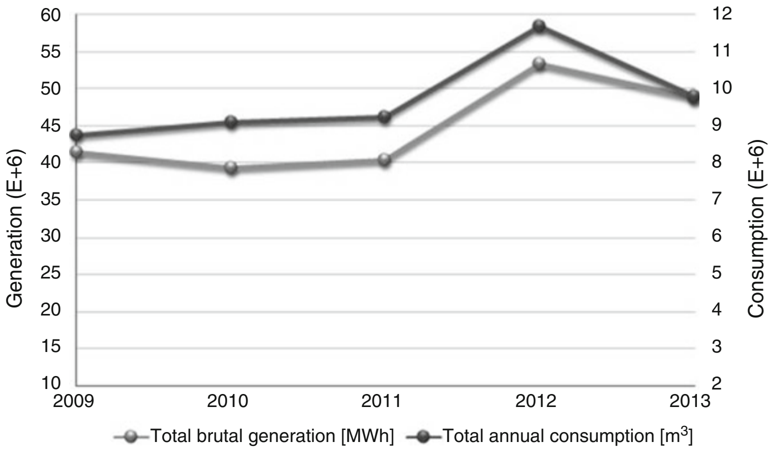

Fig. 8.5 Fuel oil consumption vs. power generation (Source: Salazar, M.S. 2015)

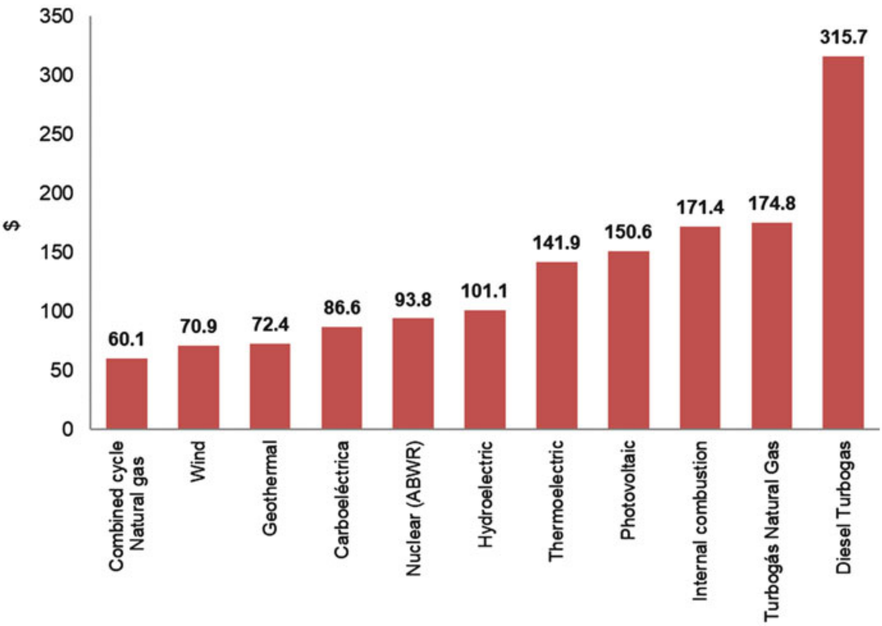

Fig. 8.6 Technology-level cost (dollars per net MW) (Source: Ministry of Energy)

In terms of generation costs, the level cost represents the estimated cost for the entire life of the power unit, which is why it is an efficient mechanism for evaluating generation projects. Figure 8.6 shows the level cost of the different technologies that make up the energy matrix for Mexico, as mentioned the generators with turbo gas technology with diesel and gas, and those of internal combustion are the most expensive plants, with a maximum of up to $ 315.7 per MW in the case of diesel.

8.2.1 The Wholesale Electricity Market

As a result of the energy reform implemented in Mexico, the wholesale electricity market (WEM) was created, a market based on variable costs. The purpose of the WEM is to efficiently cost and ensure the supply of electricity in the country. It is a newly created market whose operation began in February 2016. At present there is a small group of private suppliers that have joined the WEM, but it is expected that as the market matures, market concentration will decrease, allowing the reduction of generation costs.

As a prudent measure to guarantee supply, a legacy contract scheme was created through which generation contracts are assigned to a group of power plants belonging to CFE in such a way as to guarantee a minimum margin of supply according to the expected demand. It is expected that as more suppliers enter to the WEM, the market share of the parastatal company will decrease and only operate with a portfolio of plants that are efficient and economically competitive. Due to the above, it is of great relevance for CFE to know the level of efficiency of its plants to be able to form a portfolio of highly competitive plants and to be able to participate in equal conditions in the WEM.

The WEM is formed by a set of market instruments that facilitate the transition to a liberalized market; among the main instruments that contribute to the development of the MEM, we can mention [3] power market: The power market seeks to guarantee the generation of energy in the long term, by generating incentives for the establishment of generation plants in areas where there is greater demand. Depending on the technology of the plant, the generators can offer power, just as the consumers of electric energy are obliged to acquire power in proportion to their consumption. It is an ex post market, so the plaintiffs and bidders will have total certainty when carrying out their operations. The purpose of this market is to establish price signals in congruence with the conditions of shortage or surplus of generation capacity in the National Electrical System.

Clean energy certificates market: The market for clean energy certificates (wind, solar, geothermal, oceanic, bioenergy, hydroelectric, nucleoelectric, efficient cogeneration, among others) allows to encourage the generation of clean energy, through an economic incentive to make these technologies competitive in their participation in the WEM. The Energy Regulatory Commission will be in charge of determining the number of certificates that corresponds to the generating plants through clean technologies and that they will be able to commercialize in the market. For their part, both consumers and suppliers in the market will have the obligation to acquire such certificates to ensure the reduction of contaminants.

Financial Transmission Rights Market: The Financial Transmission Rights (FTR) grant the owner the right to charge or the obligation to pay the difference of the marginal congestion components of the local marginal prices of the Market of the Day in Advance, between a destination node and a source node. The FTR market seeks to offer coverage to market participants of price volatility generated by the congestion of transmission lines. Since in the case of Mexico, due to the disposition of its territory, there are networks of transmission and distribution hitherto insufficient to bring energy to each area of the country, so that seasonal congestion is generated in the transmission lines, and this generates the price of energy increase.

The mechanism by which these markets will operate will be through auctions carried out by the market operator, the National Energy Control Center (CENACE), which will determine the price and quantities to be marketed in each market, as well as the optimal allocations for the generation in the Mexican electricity sector.

Medium- and long-term auctions: Auctions are a mechanism that gives transparency and certainty to the market. Auctions will be held in two planning horizons, in the medium and long term, allowing basic service providers and other responsible entities to conclude coverage contracts with generators and marketers. In the case of long-term auctions, placement adjustment mechanisms will be considered, giving greater certainty on congestion costs and rewarding or penalizing the generator according to the location, which will generate incentives to locate the new plants where they are most needed.

The WEM started operations in February 2016; in the first year of operation, energy prices recorded in the market show some volatility. The local marginal prices (LMP) that are the prices that reflect the cost of the energy in the delivery node reflect the variable cost of energy generation, plus the cost per congestion and the cost of the losses. The main reasons for the high volatility of LMPs at the onset of WEM may be due to climatic aspects, fuel availability, congestion, and network failures. The evolution of the LMP is shown in Fig. 8.7.

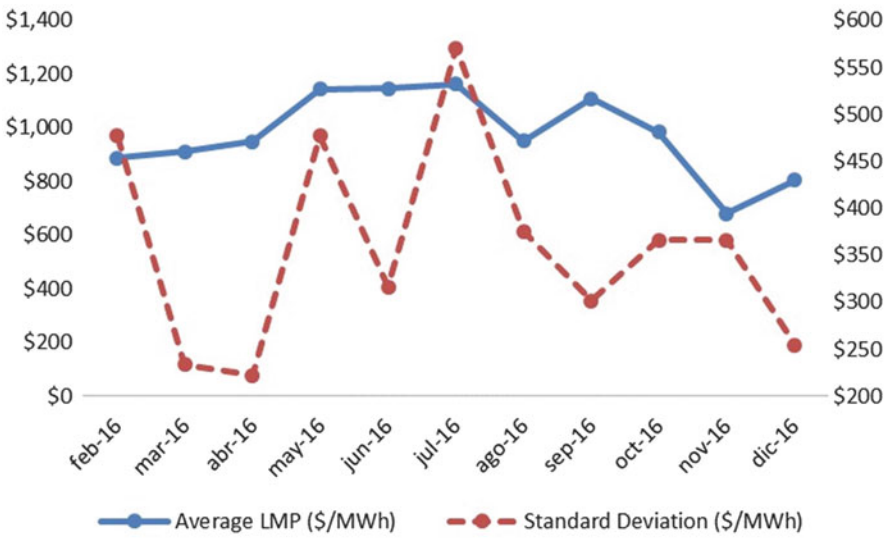

Fig. 8.7 Evolution of LMP in Mexico (MX pesos per MWh) (Source: Ministry of Energy)

8.3 Technical Efficiency

The efficiency analysis is based on microeconomic analysis, through maximizing a given production function using a set of inputs and technology. There are two approaches to the optimization of a production function through the maximization of production and cost minimization given a certain level of production.

The efficiency analysis is based on the concepts of partial and total factor productivity. In the case of partial productivity, the output per input used is calculated. In total productivity, index of products and an input index are constructed to evaluate the total productivity factor. In [4], two conceptual approaches of economic efficiency were proposed:

- Technical efficiency: The ability of a production unit to obtain the highest level of product, a given level of inputs and technology
- Allocative efficiency: The ability of a production unit to use the inputs in optimal proportions, given a price level and technological level

The usefulness of technical efficiency lies in generating information to improve management capacity of the production units, knowing the information of input use and generation of production over to define optimal strategies for improvement. There are several methods of estimating technical efficiency, parametric, nonparametric, deterministic, and stochastic:

- Parametric: Assume a functional form of the production function.
- Nonparametric: Do not assume any functional form of the production function.
- Deterministic: Part of the approach that all distances between the borders of production and production value are observed for a productive unit corresponds to technical inefficiency.
- Stochastic: Production which presents a stochastic component that attaches the technical inefficiency as a part of the error generated by the production unit is away from the production frontier; a component of error is random and one attributable to technical inefficiency.

Techniques for estimating technical efficiency according to the adopted approach can be classified among the primal and dual approaches; in the case of primal approach, technical efficiency is estimated based on maximizing production or minimizing function costs. In the case of dual approach, technical efficiency for both functions is estimated [5–7]. Models can also be categorized in terms of temporality, in models of cross section or panel data.

Generally, the nonparametric models used are estimates based on data envelopment analysis (DEA) and methodologies based on mathematical programming. One advantage of using DEA is that a specific functional structure of the production function is not required. The main disadvantage is that it is a deterministic model that can be affected by the number of inputs used and the presence of outliers [8, 9].

8.3.1 Data Envelopment Analysis

The data envelopment analysis is a well-known technique for conducting technical efficiency studies. The first approach was developed in 1978 by [10]. Mainly, this method follows the basic concepts developed in the work of [4]. Over the years, several modifications and contributions have been made by many other authors. These proposals generated more stylized version of this technique.

The data envelopment analysis is a nonparametric method of measuring efficiency based on obtaining an efficiency frontier from the set of observations that are considered without the estimation of any production function. The objective is to optimize the efficiency measure of each unit analyzed to create an efficient frontier based on the Pareto criterion [10].

First, the frontier of empirical production is constructed and then the efficiency of each production unit. Additionally it is not a parametric method, nor is it statistic, since it does not assume that the unobserved efficiency follows any kind of probabilistic distribution.

Among the advantages of using the DEA are: (a) it applies to sectors that employ n-inputs in their process and generate n-outputs and (b) does not require prior knowledge of the production function. Some disadvantages of the DEA technique are: (a) It requires homogeneity of the productive units. (b) Influence on the results in the presence of outlier data. (c) The difference between the efficient frontier and each productive unit is only attributed to inefficiency.

In recent years, authors such as [11] showed that most representative efficient points can be found using a direct approach and may differ from those obtained by multistage DEA. Assuming the economic production activities, convexity, strong disposability, and constant returns to scale (CRS), we can develop the linear program as a type of piecewise linear frontier. Input-oriented CRS efficiency is defined in Eq. (8.1) by applying the piecewise linear frontier to the input requirement set [12].

$$\max_{v,u} z = uy_j \tag{8.1}$$

Subject to

$$\begin{aligned} &vx_j = 1 \\ &-vX + uY \leq 0 \\ &v \geq 0 \\ &u \geq 0 \end{aligned}$$

u_j being free in sign
DMU$_j$; $j = 1, \ldots, n$ are a set of observed decision-making units (DMU)
x_j is input vector
y_j is output vector
u is the row vector

v are the output and input multipliers

X, Y are the input and output matrices

The goal of the input-oriented DEA model is to minimize the virtual input, relative to a given virtual output, subject to the constraint that no DMU can operate beyond the production possibility set and the constraint to nonnegative weights. The dual form of DEA is expressed according to Eq. (8.2).

$$\min \theta_{\theta,\lambda} \tag{8.2}$$

Subject to

$$\theta x_j - X\lambda \geq 0$$
$$Y\lambda \geq y_j$$

$\lambda \geq 0$, where λ is a semi positive vector in R^k and θ is a real variable. The problem can be expressed as Eqs. (8.3) and (8.4).

$$\min \ \theta_{\theta} \tag{8.3}$$

$$\min_{\lambda,s^+,s^-} \Sigma - s^+ - s^- \tag{8.4}$$

Subject to

$$\theta x_j - X\lambda - s^- = 0$$

$$Y\lambda + s^+ = y_j$$

$$\lambda \geq 0$$

λ, s^+, s^- are semi positive vectors in R^k and θ is a real variable.

The single-stage DEA model solves (3), and two-stage DEA model solves (3) followed by (4), consecutively. In this model, constant returns to scale is assumed.

8.3.2 Stochastic Frontier

In 1977, simultaneously [13, 14] proposed a stochastic frontier. The initial specification was a cross-sectional model with an error term composed of two factors one that measures the random effect and another that measures the technical inefficiency. In [15–18], developments of stochastic frontier models in various applications are presented. In [19], a model for the inefficiency effects of the stochastic frontier production function was proposed. This model is applied in the analysis of data on electricity generation during different time periods.

The most important in terms of the methodology of estimation of stochastic frontiers is that treatment of error term does not assume that all error is attributable to a random factor, but a segment of this is attributable to technical inefficiency. Figure 8.5 shows the comparison between deterministic and stochastic methods to estimate technical inefficiency for a production function.

A number of empirical studies have estimated stochastic frontiers and predicted firm-level efficiencies using these estimated functions and then regressed the predicted efficiencies upon firm-specific variables in an attempt to identify some of the reasons for differences in predicted efficiencies between firms in an industry.

This has long been recognized as a useful exercise; both the two-stage estimation procedures have also been long recognized as one, which is inconsistent in its assumptions regarding the independence of the inefficiency effects in the two estimation stages. The two-stage estimation procedure is unlikely to provide estimates, which are as efficient as those that could be obtained using a single-stage estimation procedure.

This issue was addressed in [19, 20] where stochastic frontier model was proposed in which the inefficiency effects are expressed as an explicit function of a vector of firm-specific variables and a random error. In [21, 22], it was proposed a model, which is equivalent to the specification founded in [20]. In this work the allocative efficiency is imposed, the first-order profit-maximizing conditions are removed, and panel data is permitted. The model specification may be expressed as

$$Y_{it} = x_{it}\beta + (V_{it} - U_{it}), i = 1, \ldots, N \; y \; t = 1, \ldots, T \tag{8.5}$$

where

Y_{it} is a logarithm of the production of the i-th firm in the t-th time period

x_{it} is a kx1 vector of transformation of the input quantities for the i-th firm in the t-th time period

β is a vector of unknown parameters

V_{it} is a random variable which is assumed to be $iid{\sim}N(0, \sigma V^2)$ U_{it} is a nonnegative random variable which is assumed to account for technical inefficiency in production and is assumed to be independently distributed as truncations at zero of the $N(m_{it}, \sigma U^2)$

$$m_{it} = z_{it}\delta$$

z_{it} is a $(px1)$ vector of variables which is influenced by the efficiency of a firm

δ is a $(1xp)$ vector of parameters to be estimated.

Efficiency Predictions

The measures of technical efficiency relative to the production frontier are defined as

$$\text{EFF}_i = E\left(Y_i^* | U_i, Xi\right) / E\left(Y_i^* | U_i = 0, X_i\right) \tag{8.6}$$

where Y_i* is the production of the i-th firm, which will be equal to Y_i

when the dependent variable is in original units and will be equal to exp (Y_i) when the independent variable is in logarithms. In the case of a production frontier, EEF_i will take a value between zero and one, while it will take a value between one and infinity in the cost function case.

8.4 Efficiency of Thermal Power Units in Mexico

The study includes data for a group of 22 thermoelectric units, annual average gross electricity generation, level of consumption, and useful existence of fuel oil are analyzed. A balanced panel data for 22 thermoelectric units was built for the period 2009–2013. The useful existence and consumption of fuel oil are expressed as an annual average in cubic meters and annual average gross generation of electrical energy in MWh.

The objective of measuring technical efficiency in the production of electrical energy is to assess whether levels of consumption and useful existence correspond with the level of production of each thermoelectric, assuming a technology given.

8.4.1 Data Envelopment Analysis

The results generated by the DEA show great variation among the estimated efficiency indexes; the lowest efficiency power plant is Lerdo with a value of 0.29, and the most efficient plants are Tula Vapor, Rio Bravo, and Baja California Sur I, with values of 1 in efficiency.

It is worth mentioning one of the main criticisms of the nonparametric models estimated by DEA, since in this case, a set of three units with very low values of efficiency is observed in contrast to three units with unit values. This is attributed to the values observed in the inputs and outputs for these plants, so the production boundary is influenced by outlier values.

On the other hand, one of the advantages of using DEA is that it allows working with sets of inputs and sets of outputs. The results show very low values of efficiency for most plants since only five plants had efficiency indexes above 0.8 (Fig. 8.8).

An additional analysis that is usually performed using DEA is to identify possible influence variables that represent the causes of variability of technical efficiency. In this case, power plants only have information about whether the total maintenance programmed during the year was carried out at each plant. In order to compare this variable with the estimated technical efficiency level, a dummy variable was constructed. This variable takes the value of 1 if all maintenance programmed during the year was performed and a value of 0 if at least one of them was not performed.

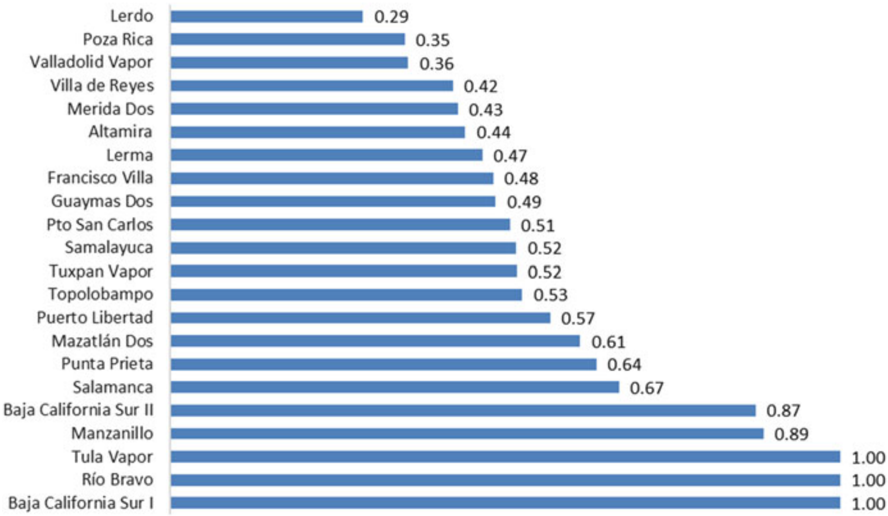

Fig. 8.8 Technical efficiency: DEA (2013) (Source: The authors)

Table 8.3 Tobit regression analysis

Theta	Coef.	Std. Err.	*t*	*P* > *t*
Maintenance	0.043081	0.1062439	0.41	0.689
_cons	0.5871584	0.0712578	8.24	0.000
/sigma	0.2453271	0.0416399		
Log likelihood	−4.043081			
Pseudo R2	0.0197			
Left censored				
Observations	0			
Uncensored observations				
Right censored	10			
Observations at theta ≥1	3			

Source: Data estimated in STATA

The estimated Tobit regression considers as an independent variable the estimated technical efficiency and as an independent variable the dummy variable constructed with the information about maintenance. The results of the Tobit regression show that the dummy variable considered is not statistically significant at 95% confidence (Table 8.3).

These results will be discussed later; one of the main reasons attributed to the low technical efficiency is the lack of maintenance of the power units. But it should be remembered that the results generated by the DEA show that there is very low efficiency and probably the built frontier is influenced by outlier values of the inputs and outputs. This is because although the technology of the comparative plants is similar, the operation and administration of the power units are heterogeneous.

8.4.2 Stochastic Frontier Model

The goal is evaluating the interaction between the inputs to generate electric power and congruence of results. Two specifications of production function, Translogarithmic and Cobb-Douglas, were tested. Cobb-Douglas production function was chosen because it generates the better estimates, so all variables are expressed in logarithmic.

The advantage of having a series of panel data is that it allows evaluating the technical efficiency of thermoelectric units over time to capture the dynamic performance of each production unit. The functional form of the model is presented in Eq. (8.7).

$$\begin{aligned}&\log\left(\mathrm{MWh}_{it}\right) = \beta_0 + \log\left(\mathrm{cons.}_{it}\right)\beta_1 + \log\left(\mathrm{Stock}_{it}\right)\beta_2 + \left(\mathrm{V}_{it} - \mathrm{U}_{it}\right)\\&i = 1, \ldots, 21 \text{ and } t = 2009, \ldots, 2013\end{aligned} \tag{8.7}$$

It was estimated a fixed-effects model by the maximum likelihood method, considering fuel oil consumption and useful existence as independent variables. Power generation is the dependent variable, and it was considered technical efficiency as variable over the time. The results obtained for the proposed model are presented in Table 8.4.

The model results show that general model and the inputs considered are statistically significant. The values of the estimated parameters of inputs considered are consistent with expectations, i.e., both inputs positively involved in the generation of electricity.

The first test (Mu) states that the average truncated normal is zero. It can be seen that the 5% significance level of the null hypothesis is not rejected. The null hypothesis of the second test is that inefficiency is time invariant (Eta). Since the values of the null hypothesis are located in the limit to 95% significance, the hypothesis that inefficiency is time invariant is rejected to 90% significance and shall be deemed inefficiency dynamic over time.

Table 8.4 Results of stochastic frontier model

Variable	Coefficient	Standard error	P-value
Consumption	0.4304761	0.0476022	0.0000
Stock	0.479270	0.0800474	0.0000
Constant	1.414125	0.3824081	0.0000
Mu	0.2714063	0.2098633	0.1960
Eta	−0.1630130	0.0318923	0.0609
Log likelihood	56.822957		
Prob. > Chi2	0.000000		

Source: Data estimated in STATA

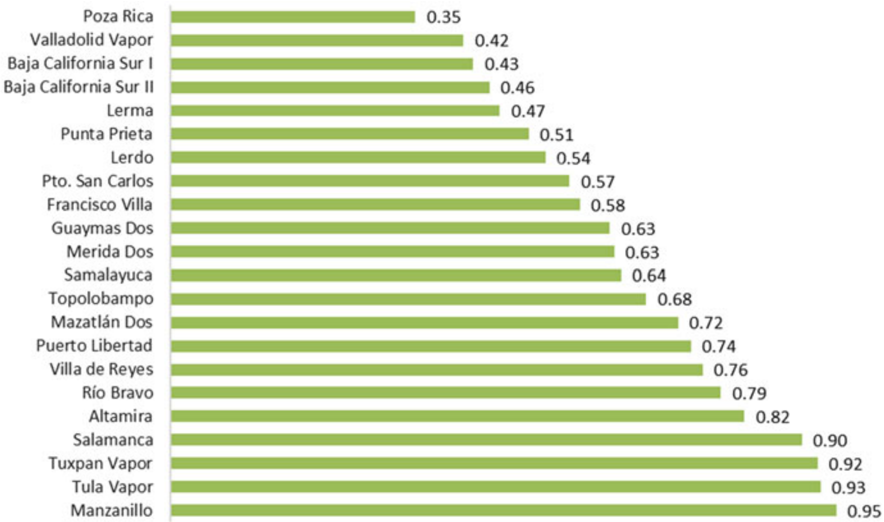

Fig. 8.9 Technical efficiency: Stochastic frontier (2013) (Source: The authors)

Based on the model parameters estimated, the level of efficiency according to Eq. (8.3) was calculated. The level of technical efficiency estimated is between 0 and 1, where 1 means that the production unit is efficient and values below 1 indicate that the unit is inefficient. When the distance from 1 is greater, it means the unit presents greater inefficiency.

Figure 8.9 shows the estimation of efficiency for 2013 of all thermoelectric units considered in the analysis. It is noted that Rio Bravo, Tula Steam, and Manzanillo are the most efficient thermoelectric units. By contrast, the worst performing thermoelectric units are Lerma, Lerdo, and Valladolid Vapor.

Of the 22 thermoelectric units analyzed, only 4 have an efficiency of 90%, 10 have an efficiency of 80%, and 7 show efficiency level below 70%. It is noteworthy that only four thermoelectric show efficiency levels around 90%; this implies that strictly speaking 81% of the thermoelectric units present an inefficient operation.

An efficiency rating was calculated based on estimates of efficiency technical for each thermoelectric in the period of analysis, using 2009 as base year. The index shows the evolution of efficiency for each thermoelectric in time, values greater than 100 indicate that the efficiency has been improved, and values below 100 indicate that the efficiency is worse.

Figure 8.10 shows the evolution of efficiency index of each unit analyzed for 2013. It is noted that in general all units have decreased the level of efficiency from 2009. Especially for the three thermoelectric units with worst performers, their level of inefficiency has fallen on average 3% in the period of analysis.

The results show that 81% of the thermoelectric units are inefficient, considering threshold efficiency values of at least 90% and observing the evolution of efficiency in time; all thermoelectric have decreased levels of efficiency.

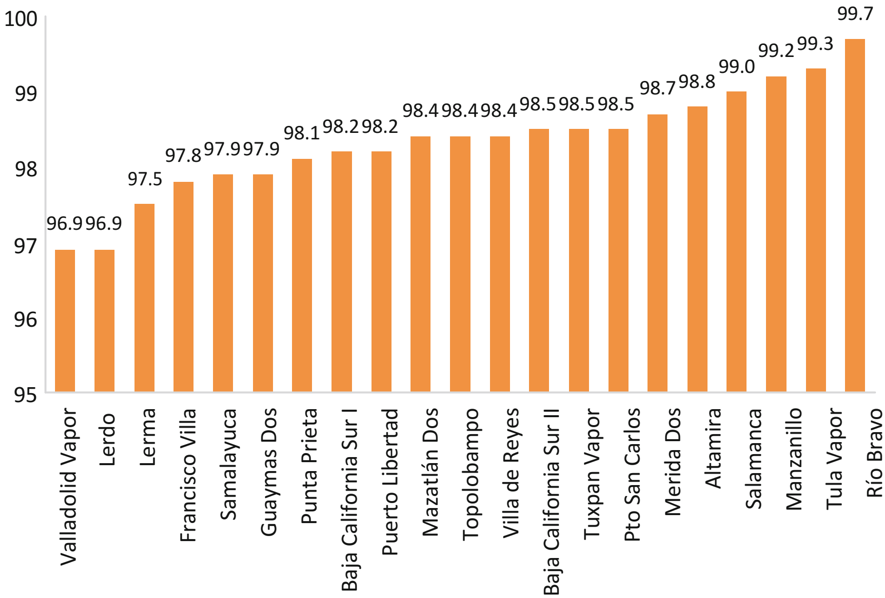

Fig. 8.10 Technical efficiency index (2013) (Source: The authors)

The estimated level of efficiency is an indicator to analyze in more detail the operation of thermoelectric units with lower levels of efficiency, and it is useful to evaluate overall performance of the electric power sector. These results show the need to assess in more detail the operation of 81% of the thermoelectric plants in Mexico.

8.4.3 Comparison of Methodologies: DEA and Stochastic Frontier

The results of both methodologies converge in most of the power units but as expected for the case of the DEA are influenced by outlier values in the inputs and outputs, which generate that one set of plants has very low efficiency and another smaller set with high efficiency. When comparing these results with those generated by the stochastic boundary model, we can observe that in the case of the stochastic boundary model, less dispersed efficiency values were generated. However, it should be emphasized that the efficiency ranking generated by both methodologies is consistent (Fig. 8.11).

The atypical case that stands out is Baja California I, since this DEA plant appears in the first rankings, which is not consistent when observing the results generated by the stochastic frontier where the central is located in the last places.

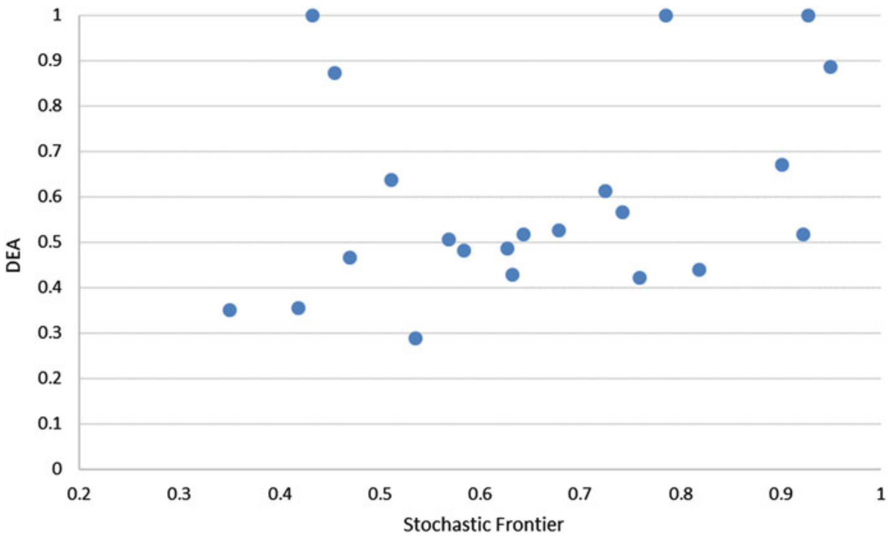

Fig. 8.11 Technical efficiency: DEA vs. stochastic frontier (2013) (Source: The authors)

Due to the above, it is necessary to know well the information that is counted and to evaluate different methodologies to validate the congruence of the results. Due to the nature of the data and the goodness of being able to generate an efficiency index over time, stochastic boundaries seem to be the best option in the present study.

Due to the heterogeneous nature of the data, it does not allow the efficient application of the DEA. For this particular case, the stochastic frontier is the most efficient methodology because it allows to avoid the influence of outlier values. In the same way the heterogeneity of the information allows to classify the power units in better way. Working with data panel allows to follow up the evolution of the efficiency of each plant over the time.

8.5 Cause Analysis of Inefficiency

There are several factors that affect the productivity of the systems, for example, technical efficiency, which can be incorporated in the stochastic frontier.

The reason why technical efficiency showed a downward trend in all thermoelectric plants could be for the wear and tear of the generating units. Since in many cases the right kind of maintenance determined by the number of operation hours (inspection, minor, intermediate, and major) is not given. This situation is mainly due to limited budget or other politic factors, as we can see in Table 8.5.

The diagnosis of the operation of an energy system is to discover and interpret the signs of malfunction of equipment that compose and quantify their effects in

Table 8.5 Type of maintenance that should be given to a steam turbine

Maintenance type	Period (OH)
Inspection	4,000
Lower	8,000
Intermediate	16,000
Major	32,000

Source: Own estimations

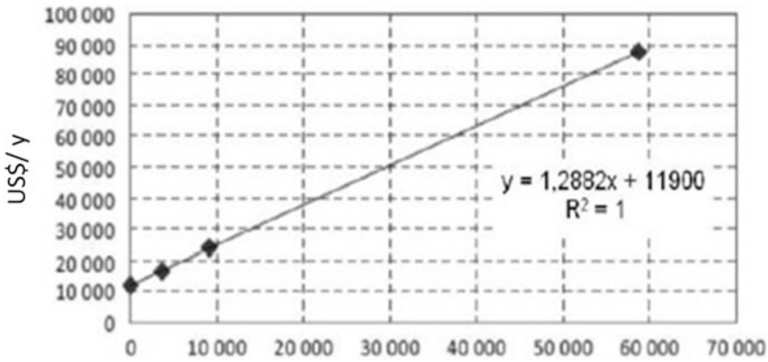

Fig. 8.12 Total cost of maintenance vs. energy production (Source: Salazar, M.S. 2015)

terms of additional consumption of resources, i.e., where, how, and how much the overall consumption of resources can be saved, holding constant the quantity and product specifications of the system.

For thermoelectric plants, a malfunction of certain equipment such as boilers will have a major economic impact, including small deviations in their performance originated for its design. A good diagnosis of the operation must be preceded by a conceptual development that explains the origin of the increase.

In Figs. 8.12 and 8.13, we see an example of the relationship between the total cost of maintaining per year and the relationship between the variable annual maintenance cost (CVM) and the operation time (HO), respectively. Thus, considering the age of most power plants, the maintenance represents a considerable budget, which often cannot be covered in a timely manner, minimizing technical efficiency.

8.6 Conclusions

The Mexican electricity sector went through an important reform in recent years; the liberalization of the electricity market allowed the private participation in the generation of electric energy. However, the transition will not be an immediate process, and that is why CFE should guarantee its participation in the wholesale electricity market through the build of a portfolio of efficient and competitive power units.

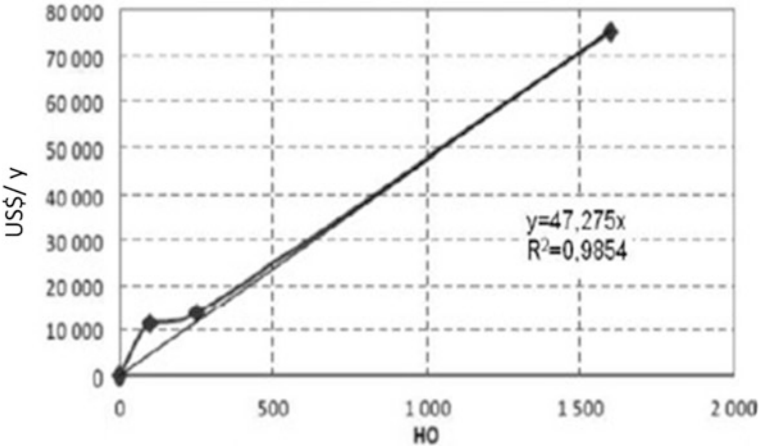

Fig. 8.13 Variable annual maintenance cost vs. hours of operation (Source: Salazar, M.S. 2015)

The study of technical efficiency for a set of thermoelectric power units allows to provide technical evidence for decision-making. In this case, two generally accepted methods, DEA and Stochastic Frontier, were used to test the results. Each has pros and cons, and it is of utmost importance to consider the nature of the information available and the problem to be addressed in order to decide the appropriate methodology to be applied.

The results generated by the DEA were influenced by the presence of outlier values of the inputs and outputs, which generated that a set of power plants presented very low efficiency and a smaller set of high efficiency. Additionally, the estimation of a Tobit model allowed evaluating the effect of the maintenance of the power plants on the efficiency, which apparently was not significant.

In contrast, available information shows that if there is a relationship between maintenance costs and above all the type of maintenance performed at each plant, this should be done with caution until more information is available.

The stochastic frontier of production function that is shown in this work permits an efficiency study of different thermal power units. Additionally, the analysis of panel data allows to evaluate the variation in time of technical inefficiency of electrical production.

In general, 81% of the thermoelectric have technical inefficiency in its operation; one reason for this could be that technical inefficiency effects are significantly related to the maintenance levels.

The evaluation of technical efficiency is a useful indicator for monitoring the operation of thermoelectric units and evaluates their performance, in order to identify which units require particular attention to achieve maximum performance in its operation. Therefore, it is important to conduct a proper diagnosis with (conventional, energy simulation, and thermoeconomic) existing methods for the proper functioning of the energy system.

The results generated in the study show that the efficiency estimates generated by stochastic frontiers are more robust and consistent, besides allowing to follow the evolution of the efficiency over time. Despite the effect of the outlier values on the efficiency estimated by DEA, in general for most plants, their efficiency ranking was congruent with that generated by the stochastic frontier. The results generated are a first indication that all plants should be evaluated more thoroughly to decide if they should be part of the portfolio of plants with which CFE will participate in the wholesale electricity market.

As future research, the effect of considering constant returns to scale and a specific production function should be evaluated in relation to using other nonparametric techniques to mitigate the effect of outlier values of inputs and outputs, such as partial frontiers (order-m and order-alpha).

References

1. Secretaría de Energía. Prospectiva del sector eléctrico (2013) 2013–2027, México, [Online], [date of reference December 19th of 2014]. Available at: http://sener.gob.mx/res/PE_y_DT/pub/2013/Prospectiva_del_Sector_Electrico_2013-2027.pdf
2. González Santaló JM (2009) La generación eléctrica a partir de combustibles fósiles. Boletin IIE. Octubre-diciembre . [Online], [date of reference December 19th of 2014]. Available at: http://www.iie.org.mx/boletin042009/divulga.pdf
3. Ministry of Energy (2015) Bases del Mercado Eléctrico. Ministry of Energy, México
4. Farrell M (1957) The measurement of productive efficiency. J R Stat Soc 120(3):253–290
5. Garzón P and Pellicer E (2010) Organizational efficiency of consulting engineering firms: proposal of a performance indicator. Dyna, 76 (160), pp 17–26, [Online], [date of reference December 19th of 2014]. Available at:http://www.revistas.unal.edu.co/index.php/dyna/article/view/1346 3/14365
6. Calderero-Gutierrez A, Fernandez-Macho J, Kuittinen H et al (2009) Innovacion en las regiones europeas. una alternativa metodologica y actualizada del ris. Dyna 84(6):501–516
7. Diaz J (2009) La eficiencia técnica como un nuevo criterio de optimización para la generación hidroeléctrica a corto plazo. Dyna, 76(157), pp 91–100, [Online], [date of reference December 19th of 2014]. Available at:http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0012-73532009000100009&lng=es&nrm=iso
8. Berndt ER, Wood DO (1975) Technology. Prices and the De- rived Demand for Energy Econ Statist 57:259–268
9. Díaz Serna FJ (2011) Optimización de la operación y evaluación de la eficiencia técnica de una empresa de generación hidroeléctrica en mercados de corto plazo. PhD thesis, Universidad Nacional de Colombia. Medellín, p 186, [Online], [date of reference December 19th of 2014]. Available at:http://www.bdigital.unal.edu.co/3683/#sthash.KcLJiF0R.dpuf
10. Charnes A, Cooper W, Lewin A, Seiford L (1997) Data envelopment analysis theory, methodology and applications. Operational Research Society 48
11. Cherchye L, Puyenbroeck TV (2001) A comment on multi-stage DEA methodology. Oper Res Lett 28:93–98
12. Cooper WW, Seiford LM, Tone K (2006) Introduction to data envelopment analysis and its uses. Springer, New York
13. Meeusen W, Van den Broeck J (1977) Efficiency estimation from Cobb-Douglas production functions with composed error. Int Econ Rev 18:435–444

14. Aigner D, Lovell C, Schmidt P (1977) Formulation and Estimation of stochastic frontier production function models. J Econ 6:21–37
15. Forsund F, Lovell C, Schmidt P (1980) A survey of frontier production functions and of their relationship to efficiency Measurement. J Econ 13:5–25
16. Schmidt P (1986) Frontier production functions. Econometric Reviews 4:289–328
17. Bauer P (1990) Recent developments in the econometric estimation of Frontiers. J Econ 46:36–56
18. Greene W (1993) The econometric approach to efficiency Analysis. In: Fried H, Lovell C, Schmidt S (eds) The measurement of productive efficiency. Oxford University Press, New York
19. Reifschneider D, Stevenson R (1991) Systematic departures from the frontier: a framework for the analysis of firm Inefficiency. Int Econ Rev 32(3):715–723
20. Battese G, Coelli T (1993) A stochastic frontier production function incorporating a model for technical inefficiency effects. Working Papers in econometrics and applied statistics. Department of Econometrics, University of New England, 69, pp 1–22
21. Battese G, Coelli T (1995) A model for technical inefficiency effects in a stochastic frontier production function for panel da- ta. Empir Econ 20:325–332
22. Salazar MS (2015) Una estrategia para mejorar la administración de los inventarios de diesel en las centrales termoeléctricas: un estudio de caso, M.S. Thesis, Engineering Systems Department, National Autonomous University of Mexico, México, p 100

Index

P. Vasant et al. (eds.), *Modeling, Simulation, and Optimization*,
EAI/Springer Innovations in Communication and Computing,
https://doi.org/10.1007/978-3-319-70542-2

E