A Greedy Heuristic Algorithm for Location-based Group Formation

Anon Sukstrienwong

School of Information Technology and Innovation, Bangkok University, Thailand

***Abstract* –Presently, location-based Group formation has become more important in many areas. Fundamentally, it is a way to assign all members to suitable groups in order to gain the maximum benefits and accomplish their objectives. However, the process of group formation based on member's location may cause a problem as it can be time-consuming if the number of members is large and more factors are considered in forming the groups. In order to cope with this problem, we proposed two variants of heuristic algorithm based on greedy algorithm in order to generate the optimized groups of nearby individual members, who are located in different locations. In this paper, we present our experimental results and illustrate that it produces optimal results for small problem sizes. Additionally, the experimental results demonstrated that the greedy algorithms produce the solutions with the intent of finding a global optimum.**

***Keywords* – Greedy Algorithm, Group Composition, Group Formation, Location-based Formation, Optimization.**

1. Introduction

Group formation can be found in both industrial and academic institutions. In the company, it is typically responsibility of the manager, which does not always yield optimal results.

DOI: 10.18421/TEM83-03
https://dx.doi.org/10.18421/TEM83-03

Corresponding author: Anon Sukstrienwong,
School of Information Technology and Innovation, Bangkok University, Thailand
Email: *anon.su@bu.ac.th*

Received: 24 January 2019.
Revised: 10 May 2019.
Accepted: 07 June 2019.
Published: 28 August 2019.

Sometimes, group formation causes several problems when the fairness is not fulfilled. Some members might complain that the team formation is biased. Some may have placed in the improper group making them having a hard time to cooperate with others. There are several methods and criteria for building groups. Nevertheless, an interesting issue that we have concerned is how to build all groups equally in terms of mean distance among individuals in a group. Every location on Earth can be specified by a set of numbers, letters or symbols in a geographic coordinate system. In order to deal with this, we proposed a greedy heuristic algorithm to generate the optimized groups of nearby individual members when the user's location information is already available.

Some previous papers in the literature present the quality of the greedy algorithm, and variants mainly formalize the greedy algorithm and then focus on the tradeoff between quality and speed in a general and abstract setting [1]. A few researches work on an optimal and efficient algorithm that ascertains equality and fairness among generated groups. In addition, it is quite useful when the time available to solve a problem is limited [2]. The greedy algorithm always gives the best solution of the current stage while finding an answer. Also, its time complexity is good. Hence, in this paper we consider the greedy algorithm for composing groups based on geographic location of members.

There are five sections including this introductory section. In the following sections, we briefly review some works done in the area of group formation. Then, we detail our proposed heuristic algorithms based the greedy algorithm for group composition, where members spread among multiple locations. Two variant of our algorithms based on the greedy algorithm are introduced. The experimental results show that the proposed algorithm produces the solutions within a reasonable amount of time. The conclusions and future work are presented in the last section.

2. Literature Review

In this era of globalization, people have more opportunity to work with others in all kinds of activities, rather than on their own. It is because a single person has limited knowledge and capability to achieve the common goal. Therefore, a significant amount of work in literature addressed that team formation is a critical activity. Group composition has become an essential part for most organizations. This type of problem has been studied extensively in the past. By definition, it is usually considered in terms of how individual members are assembled together in order to achieve optimal outcome. Several researchers have confirmed that various characteristics of individual are important considerations, when forming groups are both in science and in other contexts [3]. Additionally, Crawford et al. [4] presented that team formation has been found in various fields, such as robotics, engineering projects and business projects. Some authors considered the quality of groups, where individuals are not treated differently because the success of the groups depends on assembling the right combination [5].

Omar et al. [6] modeled the software for team formation by balancing the Belbin team role using fuzzy technique. They have claimed that a good balance of mixed characteristics in the team may have more potential to exhibit a good team work. Also, Kurade [7] implemented a greedy algorithm for selecting players to build optimized sports team. Burney [8] illustrated the method for composing the cricket team based on the previous outcomes by using generic model. The author claimed that formation of the cricket team can be awkward and unclear due to manual process and personal judgments; which may lead to a disappointing result. To cope with the problem, a genetic-like algorithm is designed to optimize the selection of multiplayer sports. Abnar et al. [9] proposed an algorithm for forming mixed groups of learners in web based collaborative learning environments. Then, Mallet et al. [10] presented a method for forming of user device groups using the location information for sharing of the content, photographs, audio content, video content, and textual content. Li et al. [11] proposed the mechanism of forming the optimized group of online buyers in an e-marketplace, where buyers can form coalitions to take advantage of volume based discounts. Furthermore, Gong et al. [12] presented a mechanism for social group utility maximization in mobile networks. Lübke et al. [13] developed a software for mobile phones, named Mobilis, for a service environment supporting the collaboration functionality. Rantapuska et al. [14] implemented a computer-implemented method for creating the group formation using mobile computing devices.

Numerous studies demonstrated the method in establishing mixed groups of members to indicate equality team prior academic achievements among team members, such as [5]. Venkatamuni and Rao [15] introduced the approach for multi-functional teams in product design and development for lean manufacturing. Moreover, in the real world, people who work on the remote site usually face various problems when they are required to work in the team. Based on [16], Weisband argued that geographic distance between team members raises the challenge of managing interdependencies. He had confirmed that if group members are located in the same area and they are able to come in contact with peers in the same place, it generally makes better outcomes for the groups in many aspects.

3. Problem Formalization and Algorithm

We have proposed an algorithm based on greedy algorithm to generate the optimized groups of nearby members, where members spread among multiple locations. The problem formalization is described as follows.

Let $M = \{m_1, m_2, m_3, \ldots, m_n\}$ denote the set of n members who are requested to form several groups based on their location. Each member named m_i, where $1 \leq i \leq n$, has been located in the specific position. Let us divide M into p groups, denoted $g_1, g_2, g_3, \ldots, g_p$, where no member belongs to two or more groups. Then, $M = \bigcup_{i=1}^{p} g_i$ and $\bigcap_{i=1}^{p} g_i = \emptyset$. As mentioned previously, our method form of the groups depends on the geographic distance among members. For geographic coordinate system, each location can be presented in the units of Latitude and Longitude. Therefore, every member's location can be specified by the latitude and longitude coordinates. We assume $(latitude_i, longitude_i)$ to be the position of m_i. That is the user's location information already available. In this work, we employ the greedy algorithm in order to generate all groups, where all members are the ones who are located in different locations.

Basically, greedy algorithm is acknowledged as an algorithmic paradigm that follows the problem solving heuristic by making the locally optimal choice at each stage. Two greedy algorithms for composing p groups of n members based on members' location are shown in Fig. 1. and Fig. 2.

The heuristic 1 starts from the first stage by randomly selecting a member for the group g_j, where $i = 1$ to p. With regard to the greedy behavior, the algorithm searches for the closest neighbor to be allocated in the group. It then picks the other neighboring one, who has the shortest

Algorithm 1: Greedy heuristic

Let $M = \{m_1, m_2, m_3, \ldots, m_n\}$ and q_j as the size of each group, where $1 \le j \le p$, $g_1 = \emptyset$, $g_2 = \emptyset, \ldots, g_p = \emptyset$ and $W = \emptyset$.
For j = 1 to p
Begin
  $i \leftarrow 1$
  Randomly select m_i where $m_i \in M$ and $m_i \notin W$
  $W \leftarrow W \cup \{m_i\}$
  $g_j \leftarrow g_j \cup \{m_i\}$
  For k = 1 to $q_j - 1$
  Begin
    Find $m_k \in M$ and $m_k \notin W$, where the distance between m_k and other members currently in g_j is the shortest.
    $W \leftarrow W \cup \{m_k\}$
    $g_j \leftarrow g_j \cup \{m_k\}$
  End.
End.

Figure 1: a greedy algorithm (heuristic 1) for group formation based on member's location.

distance from all other members currently in the group. Once all members of group g_j are selected, the following stage will be started to form the group g_{j+1} in the same way.

It is noted that we aim to build the all groups with the same size (p). Hence, the final result becomes $\{g_1, g_2, \ldots, g_p\}$. In order to find the complexity of the algorithm, we count the total number of comparisons in searching of the shortest distance of nearby members that occurred in forming all groups. It can be illustrated as below.

1st group: (n-1)+2(n-2)+3(n-3)+…+(q-1)(n-(q-1))
2nd group: (n-q-1)+2(n-q-2)+…+(q-1)(n-(2q-1))
3rd group: (n-2q-1)+2(n-2q-2)+…+(q-1)(n-(3q-1))
…
pth group:(n-((p-1)q-1))+2(n-((p-1)q-2)))+…+((q-1)(n-(pq-1)))
=(n-(p-1)q-1))+2(n-((p-1)q-2)))+…+((q-1)(2))+(q-1)(1)))

Let T(n) be the summation of all comparisons presented above. We can represent that

T(n) =(n-1)+ 2(n-2)+3(n-3)…+ (q-1)(2))+(q-1)(1))
< q(n-1)+q(n-2)+q(n-3)...+q(2)+q(1)
= q((n-1)+(n-2)+(n-3)…+2+1)
= q(n-1)(n-2)/2

For the worst case, if members form two groups equally, $q = n/2$. Then, T(n) becomes $(n^3 - 3n^2 + 2n)/4$. That is, the algorithm runs in $O(n^3)$ time. If q is small, then T(n) becomes $n^2 + 2n + 2$. For the best case, the algorithm runs in $O(n^2)$

The heuristic 2 is the modification of the first greedy algorithm 1 in which the equality in terms of distance within formed groups is primarily in our consideration. In this algorithm, the mean value of all members' position in a group is specified as the center of the group in order to compare with the others.

Algorithm 2: Greedy heuristic

Let $M = \{m_1, m_2, m_3, \ldots, m_n\}$ and q_j as the size of each group, where $1 \le j \le p$, $g_1 = \emptyset, g_2 = \emptyset, \ldots, g_p = \emptyset$ and $W = \emptyset$.
For j = 1 to q /*First stage*/
Begin
  $i \leftarrow 1$
  For k = 1 to p
  Begin
    Randomly select the first member m_i for g_k, where $m_i \notin W$ far from most others.
    $W \leftarrow W \cup \{m_i\}$
    $g_k \leftarrow g_k \cup \{m_i\}$
  End.
For k = 2 to p /*Other stages*/
Begin
  $i \leftarrow i + 1$
  Find $m_i \in M$ and $m_i \notin W$, where the distance between m_i and other members currently in g_k is the shortest.
  $W \leftarrow W \cup \{m_i\}$
  $g_k \leftarrow g_k \cup \{m_i\}$
End.
End.

Figure 2: a greedy algorithm (heuristic 2) for group formation based on member's location.

Similarly, the algorithm tries to select the best individual for each group while finding an answer toward a globally-optimal solution.

In order to avoid generating groups in the same area; therefore, at the first stage it greedily chooses the first member of each to be furthest from other groups. At each stage, a single member of each group will be selected. Doing so will minimize the opportunity to miss picking up members who are close to many groups. Then, the total distance within groups will be low. All groups will be more likely to contain members who are not scattered in different locations. To determine the worst-case running time for algorithm 2, we found that it is similar to the algorithm 1 and runs in $O(n^3)$ time. For the best case, the algorithm also runs in $O(n^2)$, with n representing the size of members.

Moreover, we employ the Euclidean distance (ED) to calculate the distance between members in a generated group. This value helps to measure the performance of our proposed algorithms. We prefer to use the word "strength" to describe this property of a group. In fact, it is the feature that determines the quality of our groups as well. Two definitions used in this work can be defined in the following mathematical form.

Definition 1: Strength of a group. If $g_k \subset M$ where $1 \leq k \leq p$, then the strength of the group based on the distance between members can be calculated as below:

$$strenth(g_k) = \sum_{i=1}^{s_k-1} \sum_{j=j+1}^{s_k} |m_i - m_j| \quad (1)$$

, where s_k is the group size of g_k.

Definition 2: Strength of the group formation. If all members are divided into p groups, where $\bigcup_{i=1}^{p} g_i = M$ and $\bigcap_{i=1}^{p} g_i = \emptyset$, then the strength of all members is calculated by the average value of all groups given by the following equation:

$$strenth(M) = \sum_i^p strenth(g_i)/p. \quad (2)$$

These definitions help to accurately reflect the characteristics of the generated groups. It is important to remind that the objective of this paper is to form the optimized groups of members based on members' location. Thus, the minimum value of the strength of the group formation defined in (2) is the most preferable in this work. If this strength value of the group is higher, it implies that members, who are located in the same area, are not being together in the same group.

4. Experimental Results and Discussion

In this section, we demonstrate some example results received by both heuristic algorithms. We start the empirical experiment by setting the original position of nine persons ($n = 9$) on a two dimensional map as presented in Fig. 3. Suppose we want to generate three groups ($p = 3$) and each group consists of 3 ($q = 3$) members.

All stages made by the greedy heuristic 1 can be shown as in Figure 4. The figure shows that only one group is completely formed for each stage. Starting from stage 1, m_6 is randomly selected as the first member of group g_1. Then, other two members, m_8 and m_5, are chosen respectively because they are close to group g_1. After completing the first group, group g_2 randomly selects the first member. Let m_1 be the first member of group g_2. Next, m_2 and m_4 are chosen respectively to be the members of group g_2 because both members are located closest to this group. The last group has been completed in the final stage. We can see that the rest of the members will be altogether in group g_3. Hence, three groups made by this heuristic 1 are listed below.

$g_1 = \{m_5, m_6, m_8\}$,
$g_2 = \{m_1, m_2, m_4\}$ and
$g_3 = \{m_3, m_7, m_9\}$,

Based on [17], every possible way of generating groups of nine members into three equal groups is $\frac{n!}{(p!)^q q!} = \frac{9!}{(3!)^3 3!} = 280$. However, the solution space becomes very large when n is big. It may be unlikely to find a good solution easily. With its greedy behavior, the algorithm 1 will find a solution toward the optimal result. Consequently, for this empirical example, there are only four different ways that can be derived from the algorithm 1 illustrated in Table 1. The strength of the group formation, defined in definition 2, is also shown to see how the algorithm works.

Nevertheless, the heuristic algorithm 2 works different from the algorithm 1 because the other objective of algorithm 2 is to form all groups equally in terms of the mean distance of group members. As the group size is equal to three, the greedy heuristic 2 is completed in three stages, which can be seen in Figure 5. For the first stage, it looks for a member who is remote from the other groups. After that the algorithm creates all groups simultaneously by adding only one neighborhood member in each group. Let m_6 be randomly picked as the first member of group g_1. Then, the group g_2 will choose the first member who is furthest from the group g_1, which is now having only m_6. In this example, m_3 will be selected. After that, the algorithm looks for someone who is furthest from both groups. In our map, m_6 is the best to be chosen as the first member of group g_3. Once the first stage is completely done, we can start the second stage. Remember that for each single stage of the algorithm 2 only one nearby neighboring will be added into each group. The algorithm continues adding members to all groups until all members have been selected. Regarding the greedily search, m_8 is chosen as a member to group g_1. And m_7 is added into group g_2. Lastly, m_1 is added into group g_3. Finally, all groups will be completely formed in the last stage. The final result of the heuristic 2 can be listed as below.

$g_1 = \{m_5, m_6, m_8\}$,
$g_2 = \{m_3, m_7, m_9\}$ and
$g_3 = \{m_1, m_2, m_4\}$,

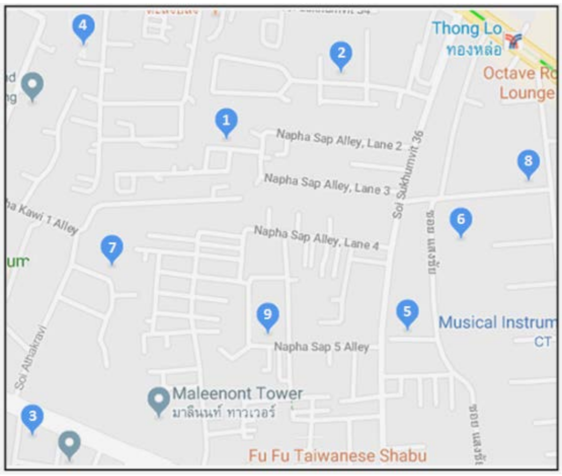

Figure 3: Original position of nine persons on map.

As empirically shown, the algorithm 1 performs well; it yields the group formation with lowest value of the strength of the group formation, while the algorithm 2 does not. This is because it has two objectives for the group formation. First, it does the group formation in order to obtain the higher value of the strength of the group formation. Second, it is designed to ascertain the equality and fairness among created groups in terms of the average distance of group's members. Therefore, it starts forming groups by randomly selecting the first member for the group g_1.

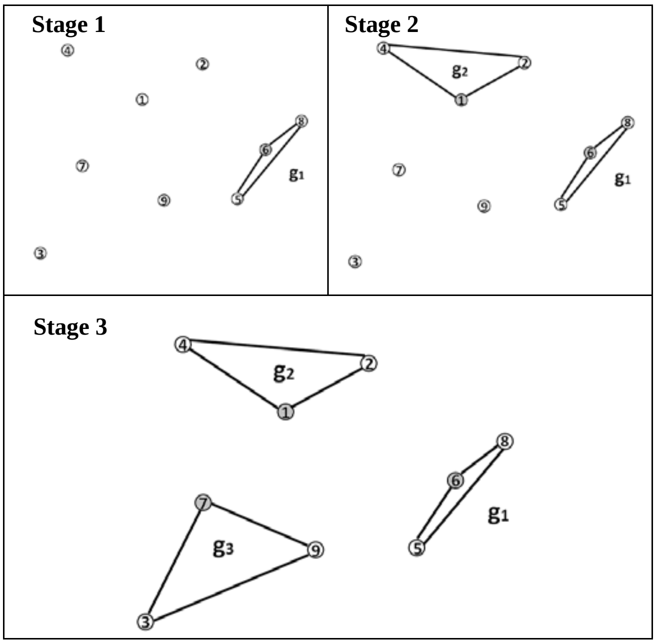

Figure 4: The example results of group composition by heuristic 1.

Then, the start-up process for all groups begins. By doing this, it becomes a limitation in forming groups. For this reason, only two different types of groups can be created, see Table 1. for details. While the algorithm 1 randomly selects the first member of all groups, it is possible to create multiple ways of groups.

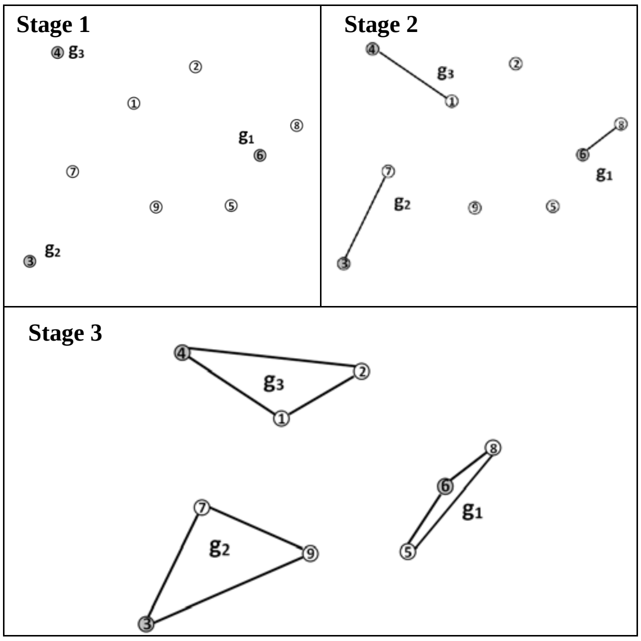

Figure 5: The example results of group composition by heuristic 2.

Table 1: The results of both heuristic algorithms, where $n = 9, p = 3, q = 3$.

No.	Heuristic algorithm 1		Heuristic algorithm 2	
	Strength of the group formation	*Formed groups*	*Strength of the group formation*	*Formed groups*
1	11.0628	$g_1 = \{m_1, m_2, m_4\}$ $g_2 = \{m_5, m_6, m_9\}$ $g_3 = \{m_3, m_7, m_8\}$	9.89794	$g_1 = \{m_5, m_6, m_8\}$ $g_2 = \{m_2, m_3, m_9\}$ $g_3 = \{m_1, m_7, m_4\}$
2	9.89794	$g_1 = \{m_5, m_6, m_8\}$ $g_2 = \{m_2, m_3, m_9\}$ $g_3 = \{m_1, m_7, m_4\}$	11.0628	$g_1 = \{m_5, m_6, m_9\}$ $g_2 = \{m_1, m_2, m_4\}$ $g_3 = \{m_3, m_7, m_8\}$
3	9.3875	$g_1 = \{m_5, m_6, m_8\}$ $g_2 = \{m_1, m_2, m_9\}$ $g_3 = \{m_3, m_4, m_7\}$		
4	9.3725	$g_1 = \{m_5, m_6, m_8\}$ $g_2 = \{m_1, m_2, m_7\}$ $g_3 = \{m_3, m_4, m_9\}$		

5. Conclusion and Future work

In this paper, we proposed a method for location-based group formation. Two variants of greedy algorithms are designed in order to generate the optimized group of nearby individual members, who are scattered in different locations. The property of the group formation is mathematically defined in order to evaluate the generated groups received from both heuristic algorithms. According to the experimental results, both algorithms achieve our goal in forming groups with the intent of finding a global optimum. Furthermore, it has been proven that they have polynomial time complexity of n^2, where n is the number of members; therefore, they are fast enough for our problem and other practical applications.

In the future, we will develop an application for location-based group formation using our greedy algorithms on smartphones to provide users with greater convenience. In addition, we will experiment with real-world data on a large number of users to see how the heuristic algorithm performs. Furthermore, another possible extension of this research is to compare the results to the other heuristic grouping algorithms.

References

[1] Fokkinga, M., A Greedy Algorithm for Team Formation that is Fair over Time, 'The Netherlands version', April, 2010. Retrieved from: https://maartenfokkinga.github.io/utwente/mmf2009l.pdf , [accessed: 10 January 2019].

[2] Viswanathan, V., Sen, A. K., & Chakraborty, S. (2011). Stochastic Greedy Algorithms. *International Journal on Advances in Software*, *4*(1).

[3] National Research Council. (2015). *Enhancing the effectiveness of team science*. National Academies Press.

[4] Crawford, C., Rahaman, Z., & Sen, S. (2016, May). Evaluating the efficiency of robust team formation algorithms. In *International Conference on Autonomous Agents and Multiagent Systems* (pp. 14-29). Springer, Cham.

[5] Syed-Abdullah, S. L., Omar, M., & Idris, M. F. I. M. (2011). Team achievements equality using fuzzy rule-based technique. *World Appl. Sci. J*, *15*(3), 359-363.

[6] Omar, M., Hasan, B., Ahmad, M., Yasin, A., Baharom, F., Mohd, H., & Darus, N. M. (2016, August). Towards a balanced software team formation based on Belbin team role using fuzzy technique. In *AIP Conference Proceedings* (Vol. 1761, No. 1, p. 020082). AIP Publishing.

[7] Kurade, N. P. (2014). An Intelligent method for Selecting and Recommending best Players to help Build Sports Team. *International Journal of Computer Applications*, *105*(7), 9-12.

[8] Burney, S. A., Mahmood, N., Rizwan, K., & Amjad, U. (2012). A Generic Approach for Team Selection in Multi–player Games using Genetic Algorithm. *International Journal of Computer Applications*, *40*(17), 11-17.

[9] Abnar, S., Orooji, F., & Taghiyareh, F. (2012, January). An evolutionary algorithm for forming mixed groups of learners in web based collaborative learning environments. In *2012 IEEE international conference on technology enhanced education (ICTEE)* (pp. 1-6). IEEE.

[10] Mallet, V., Cheng, J., Fiedel, N., Gillum, E. W., Ramanarayanan, G., & Woods, N. J. (2015). *U.S. Patent No. 8,943,137*. Washington, DC: U.S. Patent and Trademark Office.

[11] Li, C., Chawla, S., Rajan, U., & Sycara, K. (2004). Mechanism design for coalition formation and cost sharing in group-buying markets. *Electronic Commerce Research and Applications*, *3*(4), 341-354.

[12] Gong, X., Chen, X., & Zhang, J. (2014, March). Social group utility maximization in mobile networks: From altruistic to malicious behavior. In *Information Sciences and Systems (CISS), 2014 48th Annual Conference on* (pp. 1-6). IEEE.

[13] Lübke, R., Schuster, D., & Schill, A. (2011, March). Mobilisgroups: Location-based group formation in mobile social networks. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on* (pp. 502-507). IEEE.

[14] Rantapuska, O., Kinnunen, J., Lonnfors, M., & Kuismanen, P. (2007). *U.S. Patent Application No. 11/170,723*.

[15] Venkatamuni, T., & Rao, A. R. (2010). Reduction of product development time by team formation method in lean manufacturing. *Indian Journal of Science and Technology*, *3*(5), 578-582.

[16] Weisband, S. (2008). Research challenges for studying leadership at a distance. *Leadership at a distance: Research in technologically-supported work*, 3-11.

[17] Sukstrienwong, A. (2018). A Price-Based Mechanism for Online Buyer Coalition by Genetic Algorithms. *International Journal of Innovative Computing, Information and Control (IJICIC)*, *14*(05), 1653-1679. DOI: 10.24507/ijicic.14.05.1653.