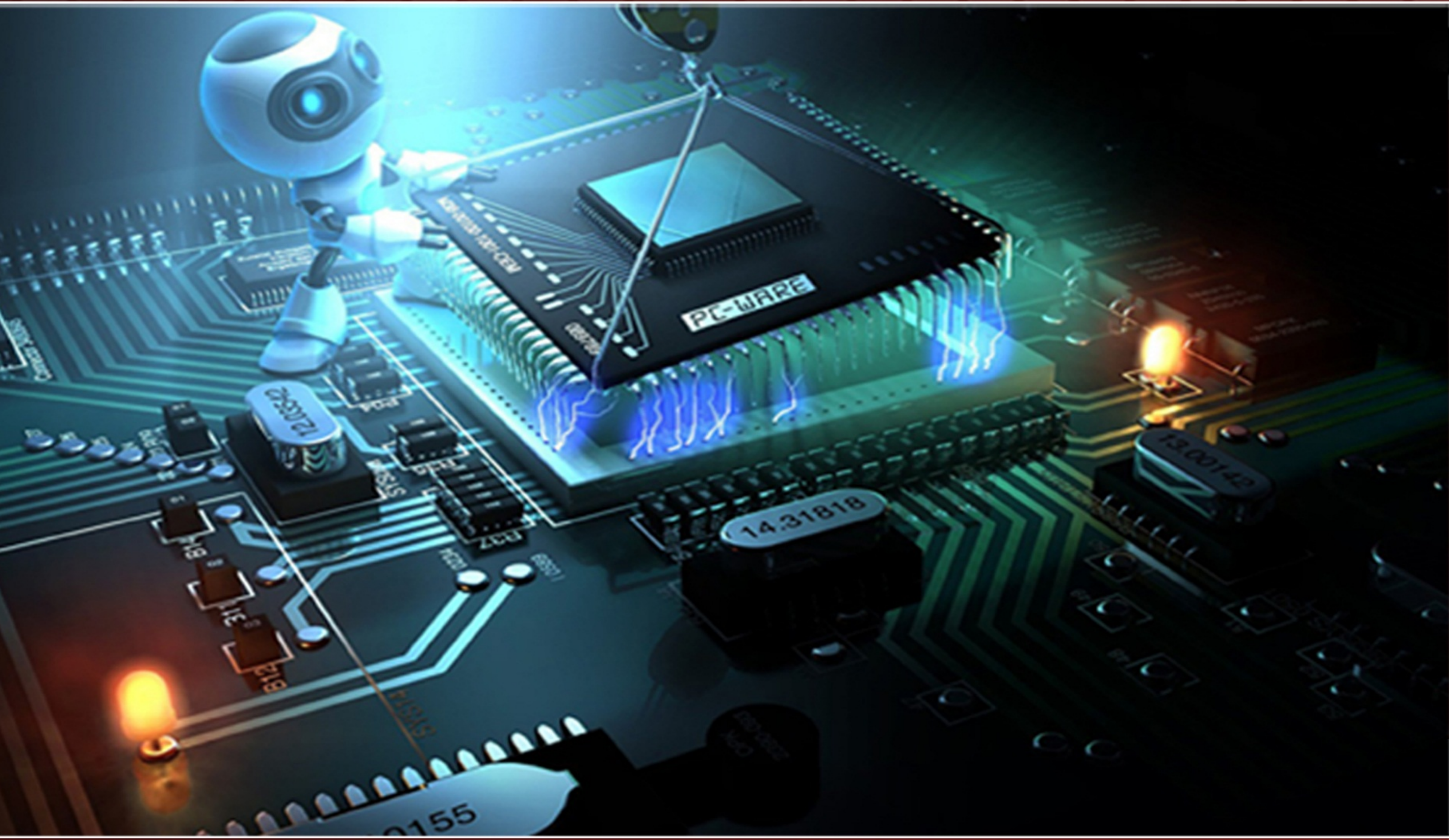
ISSN: 2054-7390
Volume 6 No 5, October 2018

Transactions on Machine Learning and Artificial Intelligence



Society for Science and Education
UNITED KINGDOM

TABLE OF CONTENTS

EDITORIAL ADVISORY BOARD

DISCLAIMER

Society for Science and Education
United Kingdom

TMLAI Transactions on Machine Learning and Artificial Intelligence

Volume 6 No. 5
ISSN 2054-7390

An Algorithm for Generating Sets of Maximally Different Alternatives Using Population-Based Metaheuristic Procedures

[1] **Julian Scott Yeomans**
[1] *OMIS Area, Schulich School of Business, York University, Toronto, ON, M3J 1P3 Canada;*
syeomans@schulich.yorku.ca

ABSTRACT

"Real world" problems typically possess complex performance conditions peppered with inconsistent performance requirements. This situation occurs because multifaceted problems are often riddled with incompatible performance objectives and contradictory design requirements which can be difficult – if not impossible – to specify when the requisite decision models are formulated. Thus, it is often desirable to generate a set of disparate alternatives that provide diverse approaches to the problem. These dissimilar options should be close-to-optimal with respect to any specified objective(s), but remain maximally different from all other solutions in the decision space. The approach for creating maximally different sets of solutions is referred to as modelling-to-generate-alternatives (MGA). This paper outlines an MGA algorithmic approach that can simultaneously generate a set of maximally different alternatives using any population-based metaheuristic.

Keywords: Modelling-to-generate-alternatives, Metaheuristics, Population-based algorithms.

1 Introduction

"Real world" decision-making environments involve complex problems containing design specifications that are frequently difficult to incorporate into underlying mathematical programming formulations and are often overwhelmed with numerous unquantifiable components [1]-[5]. Whereas "optimal" solutions can be calculated for the modelled representations, whether these are truly the best solutions to the real problems can be questionable, as there are always unmodeled components when mathematical models are constructed [1][2][6]. Generally, it is more desirable to create a discrete number of dissimilar alternatives that afford contrasting perspectives to the particular problem [3][7]. All of these options should be close-to-optimal with respect to any specified objective(s), but should be maximally different from each other within the decision space. Numerous procedures referred to as *modelling-to-generate-alternatives* (MGA) have been created to address this multi-solution approach [6]-[8].

The primary impetus behind MGA methods is to produce a set of alternatives that can be considered good when measured by the specified objective(s), but which are inherently distinct from one another within the decision domain. The resulting solution set should deliver alternative perspectives that perform similarly with respect to all modelled objectives, yet very differently with respect to any unmodelled aspects [5]. Decision-makers must conduct a subsequent assessment of the set of alternatives to determine which alternative(s) would most nearly achieve their specific requirements. Consequently,

DOI: 10.14738/tmlai.65.5184
Publication Date: 23rd September, 2018
URL: http://dx.doi.org/10.14738/tmlai.65.5184

MGA methods are classified as decision support procedures rather than as solution determination processes as assumed for explicit optimization.

Earlier MGA methods have employed direct, iterative approaches for alternative generation by incrementally re-running their algorithms whenever new solutions need be constructed [6]-[10]. These iterative methods imitate the seminal MGA approach of Brill *et al*. [8] where, once the initial mathematical formulation has been optimized, all supplementary alternatives are produced one-at-a-time. Consequently, these incremental approaches all require n+1 iterations of their respective algorithms – initially to optimize the original problem, then to produce each of the subsequent n alternatives [7][11]-[13].

In this study, it is demonstrated how a set of maximally different solution alternatives can be *simultaneously* generated using any population-based metaheuristic algorithm by extending several earlier MGA approaches [12]-[18]. All of the earlier MGA procedures employed the Firefly Algorithm (FA) for their solution procedure. The FA is a very specific instance of one population-based metaheuristic procedure that can be used for solving optimization problems. In this paper, a new algorithm is provided that has been updated and generalized so that now a simultaneous MGA solution process can be achieved using *any* population-based mechanism. This new MGA algorithmic approach advances the earlier concurrent procedures of Imanirad *et al*. [13][15]-[18] by permitting the simultaneous generation of the overall best solution together with n distinct alternatives in a single computational run. Stated explicitly, to generate the n maximally different solution alternatives, the new MGA algorithm would run exactly the same number of times that a procedure would need to be run for function optimization alone (i.e. once) irrespective of the value of n [19]-[23]. Furthermore, a new data structure is created that permits simultaneous alternatives to be constructed in a very novel, highly computationally efficient way. It is the implementation of this data structure which facilitates the above-mentioned generalization to solution by all population-based methods. Consequently, this simultaneous MGA algorithmic approach proves to be extremely computationally efficient.

2 Modelling to Generate Alternatives

Mathematical programming methods appearing in the optimization literature have focused almost exclusively upon generating single optimal solutions to single-objective formulations or, equivalently, producing a set of noninferior solutions for multi-objective problems [2][5][8]. While such methods may establish solutions to the derived complex mathematical models, whether their outputs actually generate "best" solutions to the real, underlying problems is somewhat less certain [1][2][6][8]. Within most "real world" decision circumstances, there are countless system requirements and objectives that will never be explicitly apparent or included in the problem formulation stage [1][5]. Furthermore, it may not be possible to explicitly convey all of the subjective requirements as there are frequently numerous incompatible, design components and adversarial stakeholders involved. Therefore, most subjective aspects remain unavoidably unmodelled and unquantified in the constructed decision models. This commonly occurs where final decisions are constructed based not only upon modelled objectives, but also upon more subjective stakeholder preferences and socio-political-economic goals [7]. Numerous "real world" illustrations of such incongruent modelling dualities are discussed in [6][8]-[10].

When unmodelled issues and unquantified objectives exist, non-conventional methods are needed to not only search the decision region for noninferior sets of solutions, but to also explore the decision region for alternatives that are obviously *sub-optimal* for the problem modelled. Specifically, any search for alternatives to problems suspected or known to contain unmodelled components needs to focus not only on a non-inferior set of solutions, but necessarily also on an unambiguous exploration of the problem's inferior solution space.

To demonstrate the consequences of an unmodelled objective in a decision search, assume that the quantifiably optimal solution for a single-objective, maximization problem is $\boldsymbol{X}^*$ with a corresponding objective value $Z1^*$. Now suppose that a second, unmodelled, maximization objective $Z2$ exists that subjectively incorporates some unquantifiable "politically acceptable" component. Now assume that some solution, $\boldsymbol{X}^a$, belonging to the 2-objective noninferior set, exists that represents a potentially best compromise solution for the decision-maker if both objectives had somehow been simultaneously evaluated. While $\boldsymbol{X}^a$ could reasonably be considered as the best compromise solution for the real problem, in the quantified mathematical model it would appear inferior to solution $\boldsymbol{X}^*$, since it must be the case that $Z1^a \leq Z1^*$. Therefore, when unmodelled components are incorporated into a decision-making process, mathematically inferior options to the modelled problem could actually be optimal for the real underlying problem. Consequently, when unmodelled issues and unquantified objectives potentially exist, alternative solution procedures are needed to not only explore the decision region for noninferior sets of solutions, but also to concurrently search the decision region for inferior solutions to the problem modelled. Population-based algorithms permit concurrent searches throughout a decision space and prove to be particularly proficient solution methods.

The principal drive for MGA is to create a manageably small set of alternatives that are as different from each other as possible within the solution space, yet are quantifiably good with respect to all modelled objectives. By achieving this, the resultant set of solution alternatives is able to supply truly different choices that perform similarly with respect to the known modelled objective(s) yet very differently with respect to any unmodelled issues. By generating such good-but-different solutions, the decision-makers are able to examine potentially desirable qualities within the alternatives that might satisfactorily be able to address the unmodelled objectives to varying degrees of stakeholder acceptability.

In order to motivate the MGA search process, it becomes necessary to apply a more formal mathematical definition to the goals of MGA [6], [7]. Assume that the optimal solution to an original mathematical model is $\boldsymbol{X^*}$ with corresponding objective value $\boldsymbol{Z^*} = F(\boldsymbol{X^*})$. The ensuing difference model can then be solved to produce an alternative solution, $\boldsymbol{X}$, that is maximally different from $\boldsymbol{X}^*$:

$$\text{Maximize} \quad \Delta(\boldsymbol{X}, \boldsymbol{X^*}) = \sum_i (X_i - X_i^*)^2 \tag{1}$$

$$\text{Subject to:} \quad \boldsymbol{X} \in D \tag{2}$$

$$| F(\boldsymbol{X}) - \boldsymbol{Z^*} | \leq T \tag{3}$$

where Δ represents an appropriate difference function (for clarity, shown as a quadratic difference in this instance) and T is a specified targeted tolerance limit relative to the original optimal objective value $\boldsymbol{Z^*}$. T is user-supplied and quantifies what proportion of the inferior region must be explored in the solution search for acceptable alternatives. This difference function concept can be extended into a difference measure between any *set of alternatives* by replacing $\boldsymbol{X^*}$ in the objective of the maximal difference model and calculating the overall sum (or some other function) of the differences of the pairwise comparisons between each pair of alternatives – subject to the condition that each alternative is feasible and falls within the specified tolerance constraint.

The population-based MGA procedure to be introduced is designed to generate a pre-determined small number of close-to-optimal, but maximally different alternatives, by adjusting the value of T and solving the corresponding maximal difference problem instance by exploiting the population structure of the metaheuristic. The survival of solutions depends upon how well the solutions perform with respect to the problem's originally modelled objective(s) and simultaneously by how far away they are from all of the other alternatives generated in the decision space.

3 Population-based Simultaneous MGA Computational Algorithm

In this section, a novel data structure is introduced that permits alternatives to be simultaneously constructed in a computationally efficient way that also enables an algorithmic generalization to solution by any population-based procedure. Suppose that it is desired to be able to produce P alternatives that each possess n decision variables and that the population algorithm is to possess K solutions in total. That is, each solution is to contain one possible set of P maximally different alternatives. In this representation, let $\boldsymbol{Y}_k$, $k = 1,\ldots, K$, represent the k^{th} solution which is made up of one complete set of P different alternatives. Namely, if $\boldsymbol{X}_{kp}$ is the p^{th} alternative, $p = 1,\ldots, P$, of solution k, $k = 1,\ldots, K$, then $\boldsymbol{Y}_k$ can be represented as

$$\boldsymbol{Y}_k = [\boldsymbol{X}_{k1}, \boldsymbol{X}_{k2},\ldots, \boldsymbol{X}_{kP}] \ . \tag{4}$$

If X_{kjq}, $q = 1,\ldots, n$, is the q^{th} variable in the j^{th} alternative of solution k, then

$$\boldsymbol{X}_{kj} = (X_{kj1}, X_{kj2},\ldots, X_{kjn}) \ . \tag{5}$$

Consequently, an entire population, $\boldsymbol{Y}$, consisting of K different sets of P alternatives can be written in vectorized form as,

$$\boldsymbol{Y'} = [\boldsymbol{Y}_1, \boldsymbol{Y}_2,\ldots, \boldsymbol{Y}_K] \ . \tag{6}$$

The following population-based MGA method produces a pre-determined number of close-to-optimal, but maximally different alternatives, by modifying the value of the bound T in the maximal difference model and using any population-based metaheuristic to solve the corresponding, maximal difference problem. Each solution within the population contains one potential set of p different alternatives. By exploiting the co-evolutionary solution structure within the metaheuristic, the algorithm collectively evolves each solution toward sets of different local optima within the solution space. In this process, each desired solution alternative undergoes the common search procedure of the metaheuristic. However, the survival of solutions depends both upon how well the solutions perform with respect to the modelled objective(s) and by how far away they are from all of the other alternatives generated in the decision space.

A straightforward process for generating alternatives would be to iteratively solve the maximum difference model by incrementally updating the target T whenever a new alternative needs to be produced and then re-running the algorithm. This iterative approach would parallel the original Hop, Skip, and Jump (HSJ) MGA algorithm of Brill *et al*. [8] in which, once an initial problem formulation has been optimized, supplementary alternatives are systematically created one-by-one through an incremental adjustment of the target constraint to force the sequential generation of the suboptimal solutions. While this approach is straightforward, it requires a repeated execution of the optimization algorithm [7][12][13].

To improve upon the stepwise alternative approach of the HSJ algorithm, a concurrent MGA technique was subsequently designed based upon the concept of co-evolution Imanirad *et al*. [13][15][17]. In a co-evolutionary approach, pre-specified stratified subpopulation ranges within an algorithm's overall population were established that collectively evolved the search toward the creation of the specified number of maximally different alternatives. Each desired solution alternative is represented by each respective subpopulation and each subpopulation undergoes the common processing operations of the procedure. The survival of solutions in each subpopulation depends simultaneously upon how well the solutions perform with respect to the modelled objective(s) and by how far away they are from all of the other alternatives. Consequently, the evolution of solutions in each subpopulation toward local optima is directly influenced by those solutions contained in all of the other subpopulations, which forces the concurrent co-evolution of each subpopulation towards good but maximally distant regions within the decision space according to the maximal difference model [7].

By employing this co-evolutionary concept, it becomes possible to implement an MGA procedure that concurrently produces alternatives which possess objective function bounds that are somewhat analogous to those created by the sequential, iterative HSJ-styled solution generation approach. While each alternative produced by an HSJ procedure is maximally different only from the overall optimal solution (together with its bound on the objective value which is at least x% different from the best objective (i.e. x = 1%, 2%, etc.)), a concurrent procedure is able to generate alternatives that are no more than x% different from the overall optimal solution but with each one of these solutions being as maximally different as possible from every other generated alternative that was produced. Co-evolution is also much more efficient than a sequential HSJ-style approach in that it exploits the inherent population-based searches to concurrently generate the entire set of maximally different solutions using only a single population [12][17].

While a concurrent approach can exploit population-based solution approaches, the co-evolution process can only occur within each of the stratified subpopulations. Consequently, the maximal differences between solutions in different subpopulations can only be based upon aggregate subpopulation measures. Conversely, in the following simultaneous MGA algorithm, each solution in the population contains exactly one entire set of alternatives and the maximal difference is calculated only for that particular solution (i.e. the specific alternative set contained within that solution in the population). Hence, by the evolutionary nature of the population-based search procedure, in the subsequent approach, the maximal difference is simultaneously calculated for the specific set of alternatives considered within each specific solution – and the need for concurrent subpopulation aggregation measures is circumvented.

Using the terminology introduced above, the steps in the co-evolutionary population-based MGA procedure are as follows ([14][19]-[23]:

Preliminary Step. In this initialization step, solve the original optimization problem to determine the optimal solution, ***X****. As with prior solution approaches Imanirad *et al*. [13][15]-[18]) and without loss of generality, it is entirely possible to forego this step and construct the algorithm to find ***X**** as part of its solution processing. However, such a requirement increases the number of computational iterations of the overall procedure and the initial stages of the processing focus upon finding $\boldsymbol{X^*}$ while the other elements of each population solution remain essentially "computational overhead". Based upon the objective value $F(\boldsymbol{X^*})$, establish P target values. P represents the desired number of maximally different alternatives to be generated within prescribed target deviations from the $\boldsymbol{X^*}$. Note: The value for P has to have been set *a priori* by the decision-maker.

Step 1. Create the initial population of size K in which each solution is divided into P equally-sized partitions. The size of each partition corresponds to the number of variables for the original optimization problem. $\boldsymbol{X_{kp}}$ represents the p^{th} alternative, $p = 1,\ldots,P$, in solution $\boldsymbol{Y_k}$, $k = 1,\ldots,K$.

Step 2. In each of the K solutions, evaluate each $\boldsymbol{X_{kp}}$, $p = 1,\ldots,P$, with respect to the modelled objective. Alternatives meeting their target constraint and all other problem constraints are designated as *feasible*, while all other alternatives are designated as *infeasible*. A solution can only be designated as feasible if all of the alternatives contained within it are feasible.

Step 3. Apply an appropriate elitism operator to each solution to rank order the best individuals in the population. The best solution is the feasible solution containing the most distant set of alternatives in the decision space (the distance measure is defined in Step 5). Note: Because the best solution to date is always retained in the population throughout each iteration, at least one solution will always be feasible. A feasible solution for the first step can always consists of P repetitions of $\boldsymbol{X^*}$.

Step 4. Stop the algorithm if the termination criteria (such as maximum number of iterations or some measure of solution convergence) are met. Otherwise, proceed to Step 5.

Step 5. For each solution $\boldsymbol{Y_k}$, $k = 1,\ldots, K$, calculate D_k, a distance measure between all of the alternatives contained within the solution.

As an illustrative example for determining a distance measure, calculate

$$D_k = \Delta(\boldsymbol{X_{ka}}, \boldsymbol{X_{kb}}) = \sum_{a=1 to P} \sum_{b=1 to P} \sum_{q=1\ldots n} (X_{kaq} - X_{kbq})^2. \quad (7)$$

This represents the total quadratic distance between all of the alternatives contained within solution k. Alternatively, the distance measure could be calculated by some other appropriately defined function

Step 6. Rank the solutions according to the distance measure D_k objective – appropriately adjusted to incorporate any constraint violation penalties for infeasible solutions. The goal of maximal difference is to force alternatives to be as far apart as possible in the decision space from the alternatives of each of the partitions within each solution. This step orders the specific solutions by those solutions which contain the set of alternatives which are most distant from each other.

Step 7. Apply appropriate metaheuristic "change operations" to the each of the solutions within the population and return to Step 2.

It should be apparent that the stratification approach outlined in this algorithm can be easily modified to accommodate any population-based solution procedure.

4 Conclusion

"Real world" decision-making situations inherently involve complicated performance components that are further confounded by incongruent requirements and unquantifiable performance objectives. These decision environments frequently contain incompatible design specifications that are problematic – if not impossible – to incorporate when ancillary decision support models are constructed. Invariably, there are unmodelled elements, not apparent during model formulation, that can significantly affect the adequacy of its solutions. These confounding features require the decision-makers to integrate numerous uncertainties into their solution process before an ultimate solution can be determined. Faced with such incongruencies, it is unlikely that any single solution can simultaneously satisfy all ambiguous system requirements without significant compromises. Therefore, any decision support approach must somehow address these complicating facets in some way, while simultaneously being flexible enough to condense the potential effects within the intrinsic planning incongruities.

This study has provided an updated computational procedure, a new data structure, and a significant solution-approach generalization to what has appeared previously in the literature. This new computationally efficient MGA procedure demonstrates how population-based metaheuristics can simultaneously construct entire sets of close-to-optimal, maximally different alternatives by exploiting the evolutionary characteristics of any population-based solution approach. In this MGA role, the simultaneous algorithm efficiently generates the requisite set of disparate alternatives, with each solution generated outlining a completely different perspective to the problem. Since population-based methods can be applied to a diverse spectrum of problem types, the efficacy of this new simultaneous MGA algorithm can be extended to wide range of "real world" applications. These extensions will become the topic of future studies.

REFERENCES

[1] Brugnach, M., A. Tagg, F. Keil, and W.J. De Lange, *Uncertainty matters: computer models at the science-policy interface*. Water Resources Management, 2007. 21: p. 1075-1090.

[2] Janssen, J.A.E.B., M.S. Krol, R.M.J. Schielen, and A.Y Hoekstra, *The effect of modelling quantified expert knowledge and uncertainty information on model based decision making*. Environmental Science and Policy, 2010. 13(3): p. 229-238.

[3] Matthies, M., C. Giupponi, and B. Ostendorf, *Environmental decision support systems: Current issues, methods and tools*. Environmental Modelling and Software, 2007. *22(2)*: p. 123-127.

[4] Mowrer, H.T., *Uncertainty in natural resource decision support systems: Sources, interpretation, and importance*. Computers and Electronics in Agriculture, 2000. *27(1-3)*: p. 139-154.

[5] Walker, W.E., P. Harremoes, J. Rotmans, J.P. Van der Sluis, M.B.A. Van Asselt, P. Janssen, and M.P. Krayer von Krauss, *Defining uncertainty – a conceptual basis for uncertainty management in model-based decision support*. Integrated Assessment, 2003. 4(1): p. 5-17.

[6] Loughlin, D.H., S.R. Ranjithan, E.D. Brill, and J.W. Baugh, *Genetic algorithm approaches for addressing unmodelled objectives in optimization problems*. Engineering Optimization, 2001. *33(5)*: p. 549-569.

[7] Yeomans, J.S., and Y Gunalay, *Simulation-Optimization Techniques for Modelling to Generate Alternatives in Waste Management Planning*. Journal of Applied Operational Research, 2011. 3(1): p. 23-35.

[8] Brill, E.D., S.Y. Chang, and L.D Hopkins, *Modelling to generate alternatives: the HSJ approach and an illustration using a problem in land use planning*. Management Science. 1982. 28(3): p. 221-235.

[9] Baugh, J.W., S.C. Caldwell, and E.D Brill, *A Mathematical Programming Approach for Generating Alternatives in Discrete Structural Optimization*. Engineering Optimization. 1997, 28(1): p. 1-31.

[10] Zechman, E.M., and S.R. Ranjithan, *An Evolutionary Algorithm to Generate Alternatives (EAGA) for Engineering Optimization Problems*. Engineering Optimization. 2004, 36(5): p. 539-553.

[11] Gunalay, Y., J.S. Yeomans, and G.H. Huang, *Modelling to generate alternative policies in highly uncertain environments: An application to municipal solid waste management planning*. Journal of Environmental Informatics, 2012. *19(2)*: p. 58-69.

[12] Imanirad, R., and J.S. Yeomans, *Modelling to Generate Alternatives Using Biologically Inspired Algorithms*. in *Swarm Intelligence and Bio-Inspired Computation: Theory and Applications*, X.S. Yang, Editor 2013. Amsterdam: Elsevier (Netherlands). p. 313-333.

[13] Imanirad, R., X.S. Yang, and J.S. Yeomans, *A Computationally Efficient, Biologically-Inspired Modelling-to-Generate-Alternatives Method*. Journal on Computing. 2012, 2(2): p. 43-47.

[14] Yeomans, J.S., *An Efficient Computational Procedure for Simultaneously Generating Alternatives to an Optimal Solution Using the Firefly Algorithm*, in *Nature-Inspired Algorithms and Applied Optimization*, Yang, X.S. Editor 2018. Heidelberg (Springer), Germany. p. 261-273.

[15] Imanirad, R., X.S. Yang, and J.S. Yeomans, *A Co-evolutionary, Nature-Inspired Algorithm for the Concurrent Generation of Alternatives*. Journal on Computing. 2012, 2(3): p. 101-106.

[16] Imanirad, R., X.S. Yang, and J.S. Yeomans, *Modelling-to-Generate-Alternatives Via the Firefly Algorithm*. Journal of Applied Operational Research. 2013. 5(1): p. 14-21.

[17] Imanirad, R., X.S. Yang, and J.S. Yeomans, *A Concurrent Modelling to Generate Alternatives Approach Using the Firefly Algorithm*. International Journal of Decision Support System Technology. 2013, 5(2): p. 33-45.

[18] Imanirad, R., X.S. Yang, and J.S. Yeomans, *A Biologically-Inspired Metaheuristic Procedure for Modelling-to-Generate-Alternatives*. International Journal of Engineering Research and Applications. 2013, 3(2): p. 1677-1686.

[19] Yeomans, J.S., *Simultaneous Computing of Sets of Maximally Different Alternatives to Optimal Solutions*. International Journal of Engineering Research and Applications, 2017. *7(9)*: p. 21-28.

[20] Yeomans, J.S., *An Optimization Algorithm that Simultaneously Calculates Maximally Different Alternatives*. International Journal of Computational Engineering Research, 2017. *7(10)*: p. 45-50.

[21] Yeomans, J.S., *Computationally Testing the Efficacy of a Modelling-to-Generate-Alternatives Procedure for Simultaneously Creating Solutions*. Journal of Computer Engineering, 2018. *20(1)*: p. 38-45.

[22] Yeomans, J.S., *A Computational Algorithm for Creating Alternatives to Optimal Solutions*. Transactions on Machine Learning and Artificial Intelligence, 2017. 5(5): p. 58-68

[23] Yeomans, J.S., *A Simultaneous Modelling-to-Generate-Alternatives Procedure Employing the Firefly Algorithm*, in *Technological Innovations in Knowledge Management and Decision Support*, Dey, N. Editor, 2019. Hershey, Pennsylvania (IGI Global), USA. p. 19-33

Society for Science and Education
United Kingdom

TMLAI Transactions on Machine Learning and Artificial Intelligence

Volume 6 No. 5
ISSN 2054-7390

The Impact of Technology on Contemporary Accounting: An ABCD Perspective

Lawrence A. Gordon
EY Alumni Professor of Managerial Accounting and Information Assurance at the Robert H. Smith School of Business, University of Maryland, USA
lgordon@rhsmith.umd.edu

1 Introduction

We live in an interconnected digital world of rapidly changing technology. This technology has created dynamic changes in nearly every field of study. The field of accounting is no exception in this regard. Four aspects of modern technology are having profound effect on the education, practice and research of contemporary accounting. These aspects of technology are what I call the **ABCD** of technology, where **A** stands for **A**rtificial Intelligence/Machine Learning, **B** stands for **B**lockchain, **C** stands for **C**ybersecurity, and **D** stands for **D**ata Analytics.

There are two major objectives of this paper. The first objective is to briefly discuss the impact of the **ABCD** aspects of technology on contemporary accounting. The approach to pursuing this objective will be to initially provide a definition of each of these aspects of technology, followed by a discussion on how each aspect of technology is transforming contemporary accounting. This objective is addressed in the next section of the paper. The second objective of this paper, which is addressed in the third section of this paper, is to discuss how the field of accounting needs to change in order to adapt to the ABCD aspects of technology. The fourth section of the paper provides some concluding comments.

2 The Impact of The ABCD of Technology on Contemporary Accounting

Artificial Intelligence/Machine Learning: Artificial Intelligence (AI) is usually considered a sub-field within computer science and is concerned with intelligent behavior by computers. The underlying assumption of AI is that computers (or more broadly, machines) can learn and subsequently perform tasks that previously were considered to require human intelligence. Hence, machine learning is essentially a sub-field of AI.

AI, as a separate field of study, began in the mid-1950s. However, the major advances in AI were not possible until the memory capacity and processing speed of computers reached an adequate level to simulate human reasoning and problem solving tasks. Early advances in AI were often referred to as expert systems because they were simulating structured decisions performed by an expert. Today, AI is applied to all forms of structured and unstructured decisions, including decisions where inductive learning and deductive reasoning are an essential part of the decision process.

Accounting focuses on measuring, recording, summarizing, analyzing, processing, and communicating financial, as well as nonfinancial, information about economic activities of organizations and individuals. Furthermore, accounting information is used in making all types of structured, as well as unstructured, decisions. Thus, an accounting information system is a perfect candidate for embracing machine

DOI: 10.14738/tmlai.65.5196
Publication Date: 30th September, 2018
URL: http://dx.doi.org/10.14738/tmlai.32.5196

processing of information. Indeed, various forms of adding machines and calculators have been used to assist in carrying out accounting activities for centuries. In addition, computers have been used to carry out many of the routine accounting functions for decades.

Recent advances in AI, especially machine learning aspects of AI, have opened up new vistas for the field of accounting. For example, machine leaning is used to develop models for forecasting revenues and costs. The machine (i.e., computer) "learns" how to revise the forecasting models based on a variety of past experiences and real time relevant events, simulating inductive and deductive reasoning. More specifically, based on past deviations between actual and forecasted revenues and costs, combined with additional information such as market demand forecasts, the computer can revise the forecasting model in a manner that resembles inductive and deductive reasoning by human beings. Of key importance is the fact that the revised forecasting model does not have to be based on predetermined programs or algorithms. Other examples of where machine learning is having a significant impact on accounting are in cost allocation models, profit planning models, fraud detection, and tax planning.

Blockchain: A blockchain is a digital decentralized ledger that consists of blocks of transactions between parties. The information in the blockchain is available, and transparent, to the members of the blockchain on a real time basis. Every block in the blockchain includes a link, based on cryptography, which comes from the previous block. Technically speaking, the link is a cryptographic hash function. In order to change the information in a block, a majority of all parties to transactions in subsequent blocks within the chain need to agree to the change. Thus, a large blockchain is extremely secure, even close to immutable (i.e., unchangeable), in terms of protecting the integrity of the information in the blocks. In addition, the availability of the information in a blockchain is stronger than a traditional centralized ledger system because the blockchain's decentralized ledger stores the information in all the blockchain nodes. A blockchain that is open to the public is referred to as a public blockchain. In contrast, a blockchain among a select set of users (e.g., supply chain partners) is referred to as a private blockchain. Hybrid blockchains (i.e., somewhere between public and private blockchains) also exist. Although, blockchains typically use a form of cryptocurrency, a specific country-based currency (e.g., dollars) could also be used.

Besides providing information to the members of a blockchain on a real time basis, transactions conducted via the blockchain can eliminate intermediaries to the transactions (e.g., financial institutions). Accordingly, blockchains facilitate the efficiency of operations and decisions within and between organizations. Hence, accounting systems and controls designed around a blockchain are more cost efficient, as well as more secure, than those based on traditional accounting ledgers. Auditing of the information in a blockchain is also efficiently and quickly carried out. In addition, blockchains permit the preparation of financial reports and the filing of tax information in a speedy and secure manner. Private blockchains among supply chain partners have particular appeal for accounting related transactions and services.

Cybersecurity: Security of information contained in computers became a serious concern ever since personal computers became common back in the late 1970s and early 1980s. However, the commercialization of the Internet in the mid-1990s made computer security a prominent topic of concern. In today's interconnected digital environment, it is common to refer to computer security as cybersecurity. Cybersecurity is concerned with the protection of information accessed and transmitted over the Internet and other computer networks. The primary objectives of cybersecurity are to protect the *confidentiality*, *integrity*, and *availability* (often referred to as CIA) of information. Protecting the confidentiality of information essentially means protecting private information from those that are not authorized to see or use the information. Protecting the integrity of information essentially means protecting the accuracy, reliability and validity of information. Protecting the availability of information essentially means making sure that information is available to authorized users on a timely basis. Two other objectives of cybersecurity that are often considered a subset of availability are *authentication* and *nonrepudiation*. Authentication refers to the notion that information has been made available to an authorized user whose identity has been authenticated. Nonrepudiation refers to the notion that an authorized user's identity cannot be repudiated.

Given the role that computers, the Internet, and smart devices play in today's interconnected digital world, cybersecurity breaches are a fundamental concern to the security of nations, businesses and individuals. Unfortunately, cybersecurity breaches have become common within the last two decades. Although most cybersecurity breaches do not threaten the survival of firms, many breaches do result in large costs associated with detecting and correcting the cause of the breach. In addition, some breaches result in significant costs in terms of their negative effects on a firm's brand image and legal liability. The potential negative effects resulting from cybersecurity breaches have resulted in the creation of a vibrant cybersecurity insurance marketplace as a means of transferring the risks associated with potential cybersecurity breaches.

Contemporary accounting systems are dependent on computers and the use of the Internet. Consequently, cybersecurity has become a fundamental concern for accounting. In fact, cybersecurity is a necessary, although not sufficient, condition of reliable and valid financial reporting and management accounting systems. To see how accounting systems are dependent on cybersecurity, consider the requirement for publicly traded firms to report material weaknesses under Section 404 of the Sarbanes-Oxley Act (SOX) of 2002. According to SOX, firms are required to report to the Securities and Exchange Commission (SEC) any material weaknesses in the reliability of their financial reports. However, firms operating in a computer-based environment (which, of course, means all firms) cannot have reliable financial reports without the presence of cybersecurity. Furthermore, if the reporting systems are not reliable, they lack validity and cannot be verified for accuracy. Accordingly, cybersecurity is a fundamental concern to any contemporary computer-based financial reporting or managerial accounting system. In addition, reliable auditing and tax preparation services are also dependent on secure computer-based information systems (i.e., cybersecurity). In this latter regard, the security of tax filings has become a serious concern to tax preparers, as well as the Internal Revenue Service, within the last two decades.

Cybersecurity not only has an impact on accounting, accounting also impacts cybersecurity. In fact, during the past two decades, accountants have examined several issues of importance to cybersecurity. These issues include: the cost of cybersecurity breaches [1], the amount to invest in cybersecurity activities to prevent cybersecurity breaches [2], and the impact on a firm's value derived from disclosing cybersecurity risks and incidents in 10-K reports filed with the SEC [3]. All three of these issues have been, and continue to be, of particular concern to accounting researchers, as well as practitioners.

Data Analytics: Data analytics involves analyzing and interpreting patterns and trends based on large sets of data. The focus of data analytics is on gleaning insights from large sets of data that are not obvious to the naked eye. Analyzing large sets of data is not a new phenomenon. In fact, analyzing and interpreting large sets of data have been a part of the field of statistics for centuries. During the 1990s, analyzing large sets of data via statistical techniques to determine patterns not easily identified by the naked eye became popular under the heading of data mining. However, contemporary computer processing capabilities and visualization software, combined with sophisticated statistical analyses, have opened up new doors for those interested in data analytics. Furthermore, the amount of data available for such analyses and interpretation prior to the commercialization of the Internet paled in comparison to what is currently available to organizations. Indeed, the large amount of data that is currently available and can be stored in computers, statistically analyzed, and presented in sophisticated graphs, tables, figures and reports within seconds, far exceeds what most imagined possible as recently as the turn of the 21st century.

The explosion of data availability resulting from the Internet created a large demand for user friendly software for analyzing, interpreting, and visually presenting vast amounts of data. The field of data analytics was the response to this demand. Data analytics software allows data users to analyze, interpret, and visually present, in terms of easy to understand tables, figures, and reports, large amounts of data via a simple click of a key on the computer. Although the underpinnings of data analytics software are based on sophisticated statistics and complex computer programs, users of such software only need a basic understanding of these underpinnings. In other words, the data analytics software has become user friendly.

Data analytics is transforming the way contemporary accounting is analyzing, interpreting and reporting information. In fact, the technology associated with data analytics is shifting the focus of contemporary accounting from performing functions like gathering and summarizing data to that of analyzing and interpreting large data sets. In managerial accounting, for example, it is possible to create real-time analyses, interpretations and visualizations of budgets and forecast variances. It is also possible to use a variety of data available on the Internet (e.g., macroeconomic forecasts of Gross Domestic Product, Investments, etc.) to enhance forecasting and profit planning models. In addition, it is now possible to audit an entire population, rather than a sample, while at the same time producing a comprehensive visualization of the audit's findings. Reporting of company financial reports are also being drastically changed by the field of data analytics. In fact, one can easily imagine the day when publicly traded corporations will provide a link to the company's financial information on a real time basis, supported by tables, reports, figures and graphs in the same way stock market information are currently available.

Interactions in the ABCD Aspects of Technology and Accounting: The above discussion on the impact of the ABCD aspects of modern technology on contemporary accounting has considered each aspect as being independent of each other. In reality, there are important interrelationships among them. For example, the developments related to data analytics are facilitated by the developments related to machine learning and vice versa. The developments related to data analytics are also facilitating the developments related to blockchain. In addition, cybersecurity affects all three of the other aspects of modern technology discussed in this paper. More to the point, since computer networks and the Internet are at the heart of modern technology, cybersecurity is a necessary, although not sufficient, requirement for artificial intelligence/machine learning, blockchain, and data analytics.

Figure 1 illustrates the interrelationships among the ABCD aspects of technology and their direct impact on contemporary accounting. For illustrative purposes, Figure 1 subdivides accounting into the following categories: financial reporting, managerial accounting, auditing, and taxes. Accounting information systems is shown in Figure 1 as the thread that links all the subcategories of accounting.

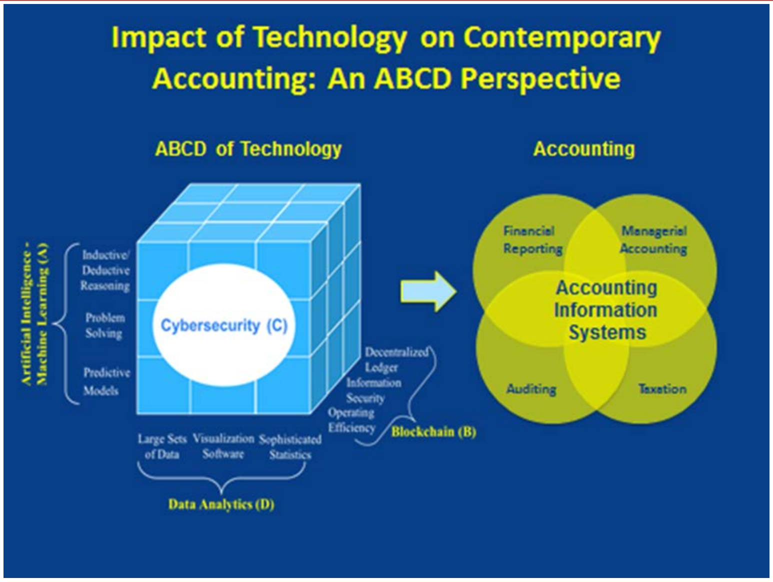


Figure 1

3 Required Changes

Accounting educators, practitioners and researchers have all recognized the importance of the ABCD aspects of technology on accounting. However, the implementation of the requisite changes has been slow when compared to the way several other fields are changing in response to the ABCD aspects of technology. As a result, the field of accounting appears to be losing ground compared to other disciplines in terms of attracting the best and brightest individuals. For example, there was a time when majoring in accounting was commonly viewed as the most intellectually challenging and most marketable major for Business School undergraduates. During that time, starting salaries upon graduation were also at the top, or at least near the top, for accounting majors compared to other majors graduating with a bachelor's degree. Much of that has changed in today's technological-based environment. Fields such as business analytics and information systems, which have fully embraced the advances in modern technology, are becoming (and in many cases have become) more prominent majors in many Business Schools. Furthermore, new graduates majoring in these other fields often receive higher starting salaries than new graduates majoring in accounting. A similar phenomenon is occurring in the hiring of rookie faculty members for business schools. Whereas a rookie Ph.D. in accounting once enjoyed the greatest marketability for a new Ph.D. out of a business school, a clear shift has been occurring. New business school faculty in fields that are more fully grounded in technology than accounting are now occupying the limelight.

The above developments notwithstanding, accounting information is still the bloodline of nearly all organizations. However, technological developments have changed, and continue to change, the role and skill requirements of accountants within organizations. Thus, the accounting industry needs to quickly and fully embrace the changes discussed above related to the ABCD aspects of technology. In this regard, accounting educators need to embrace these changes by substantially revising the curriculum requirements for both undergraduate and graduate degrees in accounting. Of course, substantial changes in the educational requirements of accounting programs means that existing, as well as new, accounting faculty need to be prepared to teach courses that are in line with these changes.

The competitive market for accounting related services provides a powerful economic incentive for accounting practitioners to embrace the ABCD aspects of technology. Thus, it is not surprising that accounting practitioners are currently leading the movement to embrace the integration of the ABCD aspects of technology. However, accounting practitioners need to move more quickly so as to keep up with other fields. Fields like data science and business analytics are more fully, and more quickly, embracing the ABCD aspects of technology in a manner that often incorporates what has previously been thought of as part of the accounting domain (e.g., analyzing and interpreting large sets of financial data derived from financial statements filed with the SEC). If this trend continues, the marketplace may ultimately end up with data science consulting firms performing a large part of the services that were traditionally thought of as accounting related services. To stem this trend, the accounting industry needs to hire individuals who have both the requisite accounting skills and the modern technological skills as part of their formal college education. Many of the master of science degrees in accounting (i.e., MSA degrees) have been designed to produce such graduates, but a large percentage of the top candidates for these programs are entering masters level programs in other fields (e.g., business analytics) due to higher starting salaries for graduate degrees in these other fields. The above is not suggesting that the accounting industry should not be hiring experts from other fields, but rather that the accounting industry needs to provide accounting students who possess the right technological skills, in addition to their accounting knowledge, the economic incentives that are required to attract them to the field of accounting.

Accounting researchers also need to focus on where the field of accounting is heading rather than focusing on the field's past and current activities (i.e., what hockey players refer to as keeping your eye on where the puck is going, not where it is). Starting in the late 1960s, and gaining substantial momentum through the 1980s, a major paradigm shift took place among accounting researchers. This shift was from examining the impact of accounting related transactions and standards on a firm's accounting income and accounting asset values, to focusing on the impact of accounting related transactions and standards on a firm's market value. This new research paradigm was, and still is, usually referred to as "capital markets research" among accounting scholars. Capital markets research is still the dominant focus in most major accounting Ph.D. programs in the U.S. Given the impact of the ABCD aspects of technology on contemporary accounting, accounting researchers need to recognize that the time has come for a new, or at least modified, dominant accounting research paradigm. For example, the capital markets research paradigm needs to take into consideration the fact that technology is changing the way capital markets operate, the infrastructure underlying capital markets, and how participants in the capital markets interact with each other. Technology is also beginning to change the currency used in capital markets (i.e., recent developments with cryptocurrencies). In essence, technology is changing the very nature of what we call capital markets. Accordingly, a research paradigm that focuses on the interactions among technology and capital markets is needed. This shift in paradigm also needs to incorporate risk management concepts that explicitly consider the transformation that is occurring in society due to rapidly changing technologies. If accounting researchers do not take up this challenge for a new, or at least modified, research paradigm, researchers in related fields will do it for them by incorporating accounting information into their research agenda (a trend that is already occurring).

4 Concluding Comments

The interconnected digital world has changed the way nations, organizations and individuals function. These changes are largely the result of technological advances related to computers and the Internet.

Indeed, the routine activities handled by people in the past, such as collecting and processing information, are efficiently handled by computers today. The same is true for highly structured activities and decisions that utilize internal and external financial information. In contrast, there is a greater need today for individuals who have sophisticated analytical and computer-based skills, especially associated with the ABCD aspects of technology discussed in this paper. Indeed, as a result of the ABCD aspects of technology, all fields are experiencing a fundamental transformation. As with all major technological advances, some fields will thrive in this new environment, while other fields will struggle to survive and even go the way of the horse-and-buggy.

Accounting is at a crossroad in the technological transformation currently taking place. The required changes have been recognized by the accounting industry, but the implementation of these changes is occurring at a slow pace relative to many other business-related fields. Thus, survival of the accounting industry depends on speeding up this technological transformation. The objective of this article has been to argue that artificial intelligence/machine learning, blockchain, cybersecurity, and data analytics are four fundamental aspects of technology that the accounting industry needs to fully embrace, and to do so in a rapid fashion. Of course, there are other aspects of technology that are also important for accounting, many of which the accounting industry has already fully embraced (e.g., social media, mobile devices, and cloud computing).

REFERENCES

[1] Gordon, L.A., M. P. Loeb, and L. Zhou, "The Impact of Information Security Breaches: Has There Been a Downward Shift in Costs?" *Journal of Computer Security*, Vol. 19, No. 1, 2011, pp. 33-56.

[2] Gordon, L.A. and M.P. Loeb, "The Economics of Information Security Investment," *ACM Transactions on Information and System Security*, (November 2002), pp. 438-457.

[3] Gordon, L.A., M.P. Loeb and T. Sohail, “Market Value of Voluntary Disclosures Concerning Information Security,” *MIS Quarterly*, Vol.34, No.3, 2010, pp.567-594.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM

TMLAI TRANSACTIONS ON
MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

Volume 6 No. 5
ISSN 2054-7390

A Proposed Control System for Checkmating an Aeronautics Agency Riddled with Fraud

Onwudebelu Ugochukwu[1] and Olusanjo Olugbemi Fasola[2]
[1]*University Paris 13, Informatics Laboratory of Paris-Nord (LIPN), UMR CNRS 7030*
[2]Department of Computer Science,University of Ibadan, Ibadan, Nigeria

anelectugocy@yahoo.com; sanjo.fasola@mbrcomputers.net

ABSTRACT

In recent year the nature of airports has changed vastly. Aviation problem is a chain and some time it trends is complex. It is a sector that changes with time and shared various challenges. Corruption is now the centre of gravity in most nations and it sucks every value, every moral and every drop of human compassion into its vortex. Fraud (be it procurement or maintenance frauds)in the aviation ministry in any country can only continue to make the aviation sector unsafe and thus exacerbate the increasing level of violation of the citizens' human right to life and to security and dignity of the person. Consequently, the immediate future for airlines will be shaped by how well they cope with the current economic crisis. Imagine the following information scenario: the aviation industry is generating a lot of money in many countries of the world but what if the reverse is true in some countries. This will definitely cause such aviation industry to be weak and stagnant. In this paper, we look at the causes and then design a system for detecting and controlling the causes of these abnormalities. This becomes especially important since airports are unique in nature and what works at one may not be successful at another as the operating environment may not be comparable. Non-aviation revenue will be looked at as a key factor in generating revenue for this all important sector. In our research, we aim at checkmating and eliminating criminal activity against the aviation industry, people, customers, suppliers or other stakeholders (with a 'zero tolerance' strategy in relation to crime and corruption) as well as revenue generation. This system, if implemented will be able to manage and control crime, corruption, fraud as well as generate fund for the industry.

Key Words: non-aeronautical revenue, airports, fraud, control system,

1 Introduction

Aeronautics is the science or art involved with the study, design, and manufacturing of air flight capable machines, and the techniques of operating aircraft and rockets within the atmosphere [4]. Ghana is approximately one-sixth the population size of Nigeria, however, a visit to the Kotoka International Airport, Accra, Ghana will surprise any Nigerian, as he will notice to his shock that the airport is a frail facility which is smaller than the Murtala Muhammed International Airport in Lagos [2], nevertheless, it is busiest airport in Ghana. In it, the immigration officers, aviation security and others efficiently carry out their duties resulting in a touch of patriotism wafting from those airport personnel. That is what is

DOI: 10.14738/tmlai.65.4416
Publication Date: 28th October, 2018
URL: http://dx.doi.org/10.14738/tmlai.65.4416

conspicuously lacking in Nigeria's airport. Consequently, it is not surprisingly that all the airlines that operate in Nigeria equally operate into the airport; these include British Airways, Lufthansa, KLM, Air France and others. There are also many other international carriers that also operate from the same airport, which do not operate from any airport in Nigeria.

Some Nigerian airport officials are self-centered not minding the effect of their decision on the entire nation or the sector. For this reason, it is difficult for the Nigeria economy to grow, more especially, such attitude is affecting the airport economic performance and with this attitude nobody will be willing to do business with Nigeria and this will affect job creation which can benefit even the families of this airport officials. A travel expert said in [2] that it is basically easier to travel through Accra airport than to travel through Lagos. Reasons given includes that Lagos airport is stressful; check in and transfer in Accra is easy; the town is also less stressful; fares for first class and business class cabins are lower. As a result a passenger flies to London from Accra at a cheaper fare.

Since 1999, President Obasanjo [3] stopped politicians from abusing the aviation industry but the other abuses have not stopped. Many see the aviation industry in the corruption continuum. Corruption in the aviation ministry in any country can only continue to make the aviation sector unsafe and thus exacerbate the increasing level of violation of the citizens' human right to life and to security and dignity of the person. They only cry when their corrupt practices result in death of loved ones but after few crocodile tears, these people settle down to life as usual, taking bribes in order to allow cutting corners in maintenance of private air fleets or turning deaf ears to airline operators buying condemned unmaintained aircrafts from foreign lands.

Another issue is that of security. Lasting development cannot be realized without sustainable security. The Levels of safety and security in civil aviation of a nation are determinant factors of economic growth and indices for investment prospects [14]. It is very important therefore for managers and operators in civil aviation to acquaint themselves with what safety and security are about because they form the bedrock for operational growth and continuity in any developed or developing country. Safety and security issues as well as plan to audit the airlines have become critical to the health of the industry and the nation. The plan to audit the airlines will be one in a series of airlines' audit which had trailed the aviation industry since aircraft started falling off the skies since 2005. After the crash of Bellview Airlines in October 2005; Sosoliso in December 2005 and ADC in October 2008, each accident had come with safety audit of airlines, yet planes keep dropping off the skies.

Corruption has eaten deep in our society that human compassion in areas such as aviation industry that is supposed to be corruption free is not spared [3]. Corruption in aviation industry is also affecting many different countries. For instance, in China, the southern regional director of the Civil Aviation Administration of China (CAAC), Liu Yajun, was struck and killed by a train. CAAC authorities responded to public concerns by posting statements in an official newspaper. They said "chose to commit suicide because of depression," and thus denied any connection between his death and possible misconducts [5]. A corrupt aviation ministry will clearly lack the moral authority to effectively carry out important oversight functions or to ensure due diligence regarding the registration and operation of aircraft with safety concerns. The international community has finally appreciated and come to terms with the reality of the "cancer of corruption" and its ubiquitous nature [10]; hence, we are all in agreement that being a persistent feature of human existence worldwide, its solution lies in the collective action of key global institutions with organized international joint efforts against corruption. These efforts have produced a lot of anti-corruption measures including bi-lateral and multilateral agreements, enactment of national anti-corruption laws such as the Nigeria's Economic and Financial Crimes Commission Act, 2004, the designing of international framework and strategies for the prevention of corruption and the making of an all important United Nations Convention Against Corruption (UNCAC) which has now become the reference point for anti-corruption fight all over the world is very essential in this struggle.

The causes of corruption are multiple and have been discussed by scholars under numerous headings but we will briefly list some of the major causes of corruption in a political economy that we have identified as investigators [10, 12, 19]. They include: weak institutions, poverty, poor remuneration, poor incentive system etc. Without addressing corruption in the aviation industry the country will continue to struggle to meet global standards in this sector rather than operate maximally and profitably to the growth of the economic. The former Military Head of State, Gen. Buhari, noted that "two former Ministers of Aviation are undergoing prosecution for corruption [1]. Thus, corruptions have ruined this important sector of the nation that industry's best practices are often circumvented at will by the airline perpetrators. The result of these infractions is what was witness with Air disasters such as the ill-fated Dana airliner which crashed in June killing 153 people onboard, the committee setup to investigate the crash said it was an old plane that was banned from use in the United States before it was brought into Nigeria in 2009 [2] as well as the most recent crash of Associated Airlines' plane in Lagos which claimed the lives of at least 15 people [11].

The activities of the aviation sector include: the Nigeria Civil Aviation Authority (NCAA); Nigerian Airspace Management Agency (NAMA); Federal Airports Authority of Nigeria (FAAN) and other key agencies under the aviation sector. The ministry of aviation is becoming bigger in the scheme of things and this must be redressed. In most countries the regulatory body runs the industry [17] – Federal Aviation Administration (FAA) in USA, Civil Aviation Authority (CAA) in the UK, Ghana Civil Aviation Authority in Ghana, etc. unfortunately the ministry appears to be taking over the powers of the NCAA in Nigeria. This must be discontinued otherwise the industry will revert to the old era of constant incidence. Policy makers must be separated from the regulators.

Another important issue that needs mentioning is the issue of lack of any fundamental strategy to develop a revenue generation plan for the non-aviation sector, that is, non-aviation revenue. Zenglein and Müller define non-aviation revenue as [20]: Any revenue which is not directly or indirectly associated with the handling of aircraft as well as getting anything (passengers, fuel, cargo) to and from the aircraft including any indirectly received revenue originating from providing necessary infrastructure and services for any kind of flight operation. Most notably all commercial activity within the perimeters of the airport as well as services to other external companies, as long as they are not part of the operational activity at the relevant airport are to be considered as non-aviation.

2 SWOT Analysis

One of the simplest tools for mapping the signals relevant to change in an agency or industry is the SWOT analysis [18]. SWOT stands for Strengths, Weaknesses, Threats and Opportunities. It is a simple, structured way of exploring the key challenges facing an agency. The SWOT analysis in Table 1 reveals several opportunities but limited strengths. It is essential for nations in Africa and Nigeria in particular to make the most of the strengths and opportunities when directing efforts and crafting policies with regards to the aviation sector. A number of the identified threats can be limited through appropriate legislation.

Table 1: SWOT Analysis for Aeronautics Agency in Nigeria

Strength	*Weakness*
a. *Advantage of later adopter.* b. *Borrow solutions from others such as USA, UK, Canada, India etc.*	*a. Weak internal controls or compliance programs.* *b. Public corruption.* *c. Lack of effective enhanced compliance policies.* *d. Lack of non-aviation revenue figure.* *e. Lack of Whistleblower Policy.*
Opportunities	***Threats***
a. *Using personal liability as a deterrent to corruption.* *b. Effective exploitation of the non-aviation revenue domain.* *c. Establishing a strong local operational presence of airlines.* *d. Developing modern airports with non-aviation in their initial design.*	*a. Domestic and international tax fraud.* *b. Corruption and money laundering.* *c. Illicit payments to third parties agent.* *d. Political appointment to key position in the industry which is not based on experience but on closeness to the powers that be.* *e. Lack of patriotism.* *f. Recruiting the wrong people into the right jobs.*

3 Design and Methodology

The fact that huge sum of money was invested in the aviation industry in Nigeria but only debt was incurred, demonstrated the fact that there appears to be a fundamental problem in the system that has not been addressed and that has to do with policies enunciated for the industry by government or any other factor. Nigerians need to develop a new strategy and let a system function to the benefit of all in the aviation industry. The aviation sector needs to look again at the following: aircraft operation, take off and landing and maintenance schedule of the engines which are measured in operation hours. The areas of expenditures include but not limited to the following key air carrier operation processes:

a) Operational management
b) Air transportation
c) Aircraft maintenance
d) Personnel training
e) Operational resources provision

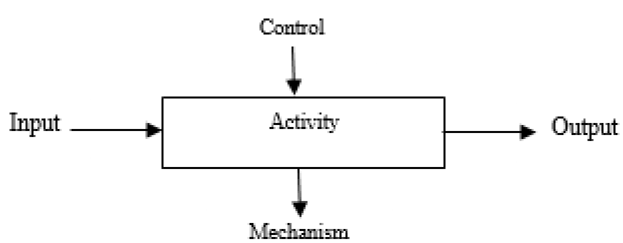


Figure 1: System Design Approach

Design is the structured creation of artifacts. The system design approach use here is called model-driven design. It specifies how a system does what it does. A system generally consists of input, process and output which can be represented by the acronym ICOM which stands for Input, Control, Output, and Mechanism (see Figure 1). The control and mechanism represent (the process) box from different sides of the box: controls connect at the top, inputs connect at the left, outputs connect at the right, and mechanisms connect at the bottom as shown in figure 1.

The diagram in Figure 2, shows our proposed control system for checkmating an aeronautics agency riddled with fraud. We also discussed ways for realizing revenue in the aviation industry which have not been exploited fully by the sector in section IV.

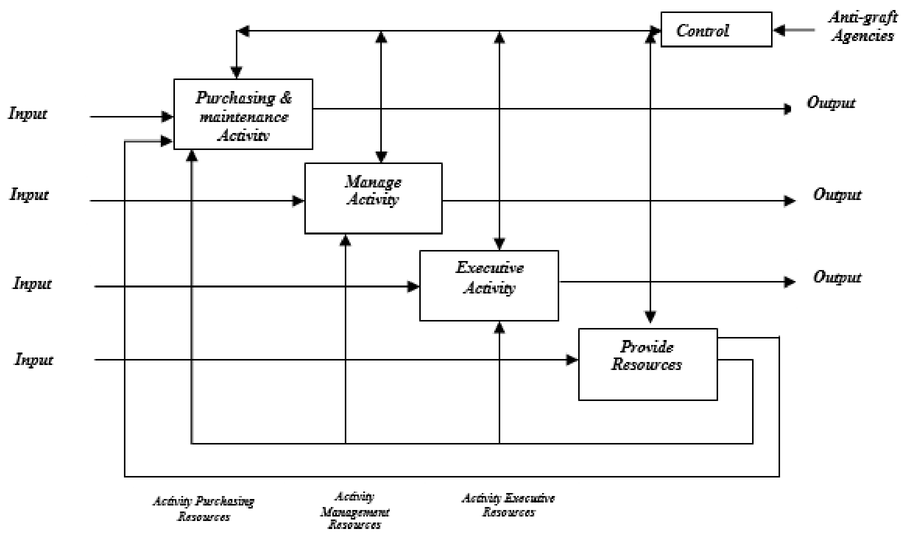


Figure 2: Control System in Aeronautics Agency

The aviation agency should carefully consider whether their existing anti-bribery compliance programs and controls adequately address the risks of doing business, particularly with previously untested third parties, in countries perceived to be at higher risk for corruption. The issue of corruption in the public service especially government functionaries who act with impunity in the spending of public funds, without attending to the reasons for which they were employed or appointed as a result of no control and feedback measures from an independent anti-graft agency will be brought to minimum by the new system (see Figure 2).

Maintenance Section: There are numerous tasks waiting to be pushed such as sufficient oversight of regulatory authority, a more transparent operating mechanism, the purchase of new parts to replace worn-out once, etc. In all this areas, there is need for transparency and an effective checks and balances as well as official accountability especially in the maintenance sector of the industry. Any reasonable designer of machines especially airplanes knows that the use of such machines airplanes involve risks which must be minimized through active maintenance culture and human factor considerations. Therefore, there is need for requisite engineers to do independent aircraft maintenance checks on the planes. In addition, the aerospace agency has to become a member of an International Contract Corruption Task Force (ICCTK), to help identify potential violations of the law with particular focus on contract corruption, procurement fraud as well as maintenance fraud.

In addition to fines, penalties and prison sentences, companies and individuals may be barred from doing business with the Federal government if found to have violated the law in the sector as compliance with the law means observing the letter and spirit of the law. Company executives face the daunting challenge of continually reassessing compliance risk with the law, monitoring the adequacy of compliance programs, and taking proactive measures where necessary to establish or augment processes in order to further mitigate the risks of noncompliance. This can be achieved as follows:

i. Periodic risk assessments that evaluate country risk, business partner risk, internal control weaknesses and other factors that may indicate a vulnerability to law violations by either employees or third parties;
ii. A regular performance of internal audit using an Internal Audit programs to monitor and enforce compliance into a recurring internal audit plan of the aviation agency;

Specialized computer applications can help identify transactions in the accounting system with characteristics of potential interest such as [9, 20]: round naira /dollar payments; one time payments to vendors; duplicate payment amounts to the same vendor; and duplicate invoice numbers. Other leading potential red flags that the law enforcement agency needs to be aware of include the following:

a) Transaction files lacking supporting documentation such as diligence records, evidence of services rendered, etc.;
b) Significant cash advances or reimbursements for travel with insufficient detail;
c) Payments for consulting services with vague descriptions of services rendered.

4 Discussion

An important part of managing risk is ensuring that the critical controls in which an agency relies on are in place and remain effective. It is equally important that the sector proactively check and monitor those controls so that they would be able to take action to avoid failure. In the purchasing & maintenance section of the diagram in Figure 2, a control by the anti-graft agencies such as the Economic and Financial Crimes Commission (EFCC) or Independent Corrupt Practices and Other Related Offences Commission (ICPC) will have to monitor and investigate any corrupt practices in the aviation sector in that domain to checkmate misappropriation or mismanagement of fund provided by the "Provide Resource" in the diagram. Others area that will be monitored and be investigated include the management activities as well as the executive activities (minister) as concerned resources provided. The outputs provided by various activities must however correspond to the control activities output provided by the anti-graft agencies (their findings). This will help in curbing corrupt tendencies, mismanagement of aviation funds and incompetence, in addition to the abuse of power by top managers (for they use their powers to create avenues for unjust enrichment or use the discretionary powers at their disposal to manipulate the system). Consequently, this will help to expose the depth and breadth of corruption and mismanagement of public funds in the aviation sector.

It is important that the government through the anti-graft agencies stepped in to ensure that the level of corruption in the sector is eliminated. As a matter of fact, the anti-graft agency must be independent from the Presidency no matter the political pressure. Consequently, they must also be allowed to act on their findings. The anti-graft agencies should be empowered to investigate all the aviation parastatals. From the diagram in Figure 2, it is clear that the anti-graft agencies will investigate all, not only the minister. This will make everyone from the top management to the middle down to the lower management to be responsible for his/her actions. The chief executive must take proactive measures where necessary to establish processes in order to further mitigate the risks of noncompliance to any compliance program in the industry. Any culprit must be made to face the full wrath of the law without nepotism. Therefore, the anti-graft agencies mentioned above need to eradicate the following in the aviation ministry [17]:

i. Its weak regulatory regime,
ii. Declining and low professional competence in the NCAA, and
iii. Non-adherence to laid down procedures.

Any breach of applicable laws, prevailing business ethics or other requirements set out in any policy document should result in disciplinary action. Such disciplinary action may include, depending on the severity of the breach, counselling, formal warning, demotion or termination of employment. Employees must learn to conduct themselves and the businesses in the aviation sector with the highest level of ethics, integrity and in the best interest of the sector. This obligation also applies particularly to dealings with shareholders, customers, suppliers, competitors, governments, regulators, and all other stakeholders. Furthermore, people must: understand and comply with relevant policies, supporting materials and processes; immediately report suspected, attempted or actual non-compliance (however minor) to the applicable law.

4.1 The Need to Encourage Whistleblower Policy

Aviation companies, as well as their officers, directors, and agents who engage in illegal activities for their financial benefit, should be held accountable for their actions. Consequently, any person or persons found culpable must be made to face the wrath of the law in order to serve as deterrent to such lapses in future. The Federal government must be commended for the action it took concerning the former minister of aviation Stella Oduoh (sacking and replacing her) who spent $1.6 million on two BMW armored cars [8]. Another case of interest is an astonishing $9.3m cash found stashed in the luggage of two Nigerians and an Israel abroad a private jet reportedly owned by a popular Nigeria cleric pastor Ayo Oritsejafor [16] which is still under investigation. The Nigeria government ought to be steadfast in their determination to combat domestic and International tax fraud, corruption and money laundering via aircraft.

To understand their obligations, people working in the aviation industry must read: code of conduct and ethics which references crime and corruption control policy. Employees must fully cooperate with internal or external auditors and must not make a false or misleading statement or conceal any relevant information from these auditors. If there is a breach of any legal or regulatory requirement or of any policy, it is employees' responsibility to report the matter. It is preferred that such concerns or potential breaches are reported directly to executive managers or as a last resort to an anti-graft agency. In instances where people are fearful of any possible unfavourable repercussions as a result of raising a concern, the whistleblower policy can assist with protecting their identity. Recently, the Chinese authorities [6] said that whistle-blowers will be protected amid an ongoing crackdown on corruption; the Nigerian government should follow in their footsteps.

4.2 Placing and Recruiting the Right People in Mission-Critical Position

Nigeria's aviation potential is arguably the best in the continent, but we need the right people with the right policies and character to make this happen. The aviation sector needs to be reposition so as to function maximally and profitably as their counterpart in other parts of the world. The aviation ministry in Nigeria must learn to follow directions and mimic policies of those whose aviation industries have survived for decades (see Table 1). To win the hearts and minds of the governed, government must deliver goods and services especially in the mission critical areas such as aviation, roads and hospital. Consequently, there is a need to place an aviator (a technocrat) in such mission-critical position that will listen to other genuine aviation professionals, taking proactive measures and diplomatically respond accordingly and knowledgeably to anything that has to do with the aviation industry [15]. He/she must not lack honesty and affirmation of morality. To the best of our knowledge and through our reach, aviation ministers have hardly be aviators themselves. For example, Professor Babalola A. Borishade, a onetime aviation minister was a brilliant electronic and electrical professional, though, he could learn quickly, but would not it had been better to put an aviator in such mission-critical position? Furthermore, Femi Fani-Kayode who replaced him was the former Minister of Culture and Tourism, who himself was not an aviator. In our research we discovered that most of the past aviation minister were not professionals when it comes to the aviation sector (See Table 2). The current President, Mr. Muhammadu Buhari, must be applauded for appointing a former pilot as the aviation minister, the first in the history of the aviation sector.

Table 2: Federal Minister of Aviation and What They Studied

Year in Office	Federal Minister of Aviation	Study
May 1999 - 2001	Olusegun Agugu	Geology
2001 – May 2003	Kema Chikwe	French
May 2003 – June 2005	Mallam Isa Yuguda	Economics
July 2005 – Nov. 2006	Prof. Babalola Borishada A.	Electrical Engineering
Nov. 2006 – May 2007	Femi Fani-Kayode	Law
June 2007 – Oct. 2008	Felix Hyatt Hassan	Unknown
Dec. 2008 – March 2010	Babatunde Omotoba	Civil engineering
April 2010 – May 2011	Fidelia Njeze	Pharmacy
July 2011 – July 2014	Princess Stella Oduah	Accounting
July 2014 – May 2015	Osita Chidoka	Management
November 2015- Till Date	Hadi Sirika	Former pilot

The aviation sector must provide a fair and transparent recruitment process based on: sourcing, selecting and appointing the best possible candidates for available positions based on merit rather than appointing people with little or no background concerning the industry. Transparency must also be seen in developing people wherever possible; positively managing candidates and the selection process as well as ensuring that new employees are onboard as quickly and as smoothly as possible.

4.3 Diversification in Revenue Generation

Low revenue generation has amplified pressure on airports to identify other revenue sources ensuring continued revenue growth and increasing profitability. In light of changing airport economics, non-aviation has therefore become an important component of revenue generation for modern airports. The Nigerian government needs to exploit this area also. Airports regardless of size have attempted to increase their share from non-aviation, as profit margins from this sector are typically higher. For example in some countries revenue streams also come from utility service or supply of services like de-icing, waste removal, aviation fuel or engineering services, thus boosting their income from this sector. Furthermore, other revenues can be generated from parking, property, security services, Information Communication and Technology (ICT) services, staff cafeteria etc. Clearly, some can be considered as non-aviation activities, whereas for others (e.g. security services) the allocation is questionable and the exact contribution of each will remain in the dark.

Airports can be very attractive locations for a variety of companies and its infrastructures have a tremendous influence in its environ and the people in the area. Being near the airport and benefiting directly from its operation can attract further corporate customers which can benefit by locating in the vicinity of the airport. For example, such a business location will generally benefit from a well-developed infrastructure leading to the airport and can thereby minimize the commute of its customers e.g. conference centres, hotels, business parks, commercial development, event management, supply services, rents, leases, transportation companies (including taxis), consulting services, real estate development etc. Modern airports which have taken non-aviation into consideration during their initial design will have better opportunities to benefit from non-aviation revenue. In addition, good transportation infrastructure with highway accessibility, rail and bus access and the distance to the city may further influence non-aviation potential.

Non-aviation activities have become increasingly important for airports. In recent years the sector has become more complex and diverse. Consequently, measurements of non-aviation activities are an important indicator of an airport's performance. But lack of a common definition and variations in data availability as well as limited transparency make interpreting such performance problematic [20]. Annual reports are often the only source of data and an analysis of this reveals that non-aviation remains a very ambiguous topic complicating data collection. Aviation charges still are an important revenue source but it is no longer the nucleus. In addition, non-aviation activity leaves airport operators with more creativity to generate additional revenue. At this point it should be stressed that the passenger is not the only potential customer at an airport. The passenger volume airports generally have five possibilities to generate revenue: retail, services, food and beverages, parking, and passenger access. To a much smaller amount some revenue streams are not based directly on the passenger's presence but on the airport's infrastructure and flight operation competencies. It allows the airport to generate a more diversified revenue stream and to maximize its non-aviation revenue.

5 Conclusion

No development can take hold without hard work, clear vision, and security in the air, land and sea. Government needs to start addressing issues raised here rather than see aviation as one sector to compensate loyalists who lacks the necessary professional skills needed in such a mission critical sector of the country. It must generator revenue in this sector to make up for shortfall in oil revenue. Without addressing corruption in the aviation industry the country will continue to struggle to meet global standards in the sector. In this respect, the President should ask anticorruption agencies to monitor the spending made by the ministry of Aviation on everything from airport safety to fleet maintenance and pilot training and certification. Every aviation industry must place high performing individuals in mission-critical positions for maximum risk assessment. Hence, an aviation industry even in a politically charged environment like in Nigeria must have professionals with strong work ethics, problem-solving ability, interpersonal skills and personal integrity who must have passed through a genuine selection process. Simply put: lack of transparency, massive corruption in the Nigeria has reduced the amount of money needed for development just as it does in any other political economy. Thus, there is a need to enhance transparency, bring corruption under control and thoroughly review the lessons of these recent events in this sector. In the end, lack of economic growth and corruption can be wiped out if comprehensive reform is promoted and rule of law improved. Revenue generation can grow exponentially if more individuals adhere to clear, explicit and written rules. Then, the lack of financially growth that has clouded the industry will finally clear out, bringing blue skies to civil aviation in Nigeria. Finally, there must be a behavioural changes on the part of the Nigerian government aimed at eliminating inappropriate behaviours that are designed to exploit and strangle the growth of the industry.

6 Recommendations

Urgent action must be taken to correct the many ills of the sector and set the industry on the path to recovery and prosperity. The following are our recommendations:

a) A technocrat should always be appointed as aviation minister so as to reposition the ministry.

b) All aircrafts must be thoroughly inspected to ensure their airworthiness without cutting corner.

c) The ministry must stop aviation officials and their collaborators who charged for new helicopter parts but installed used parts instead.

d) The establishment of a national hangar by the Federal Government as no one airline could shoulder the cost of establishing it, this will help in revenue generation.

e) NCAA should increase the number of safety inspectors (currently only 10 tested safety inspectors for the whole country).

f) There is a need for a *Special Anti-corruption Court* and strong policy enforcement.

g) The Presidency should show leadership by ensuring full accountability in the sector and by also making sure that the management of the ministry is fully compliant with anti-corruption legislation and treaties that Nigeria has ratified.

h) Anti-graft agencies need to sell the properties obtained from corruption in the industry as perpetrators would know that they will not be allowed to profit from their crimes.

REFERENCES

[1] Abiodun Badejo, (2012) Corruption in aviation sector caused the plane crash – Buhari, CPC http://dailystar.com.ng/

[2] Chinedu Eze, (2012) Aviation: Industry Riddled with Corruption http://www.thisdaylive.com/news/ 14 September 2012

[3] Christopher Odetunde, (2006) Partial Defense of the Aviation Minister But Not of the Industry.

[4] http://en.eikipedia.org/wiki/aeronautics

[5] http://english.caixin.com/editorial/ Clear Corruption from Aviation's Skies, 6 July 2010

[6] http://www.bbc.com/news/world-asia-china-29797985, China 'to protect whistle-blowers' amid corruption fight, 28 October 2014.

[7] http://www.envisionnigeria.com/ Nigeria's Aviation sector weak, stagnant — Report, 2013

[8] http://www.indepthnigeria.tumbir.com/ E-X-P-O-S-E-D: How Aviation Minister Stella Oduah Forced Cash-Strapped Aviation Agency to Spend $1.6 Million On Two BMW Armored Cars For Her, 15 October 2013

[9] http://www.qantas.com, Qantas Airways Limited, Working Toward Our Vision: An Overview of Qantas Group Business Practices

[10] Kayode Oladele, (2013) Causes and Consequences of Corruption: The Nigerian Experience, http://chatafrik.tumblr.com/

[11] Kenneth Ehigiator, (2013), Vanguard: Nigeria: Frightening Revelations - Why Planes May Keep Dropping Off the Skies, http://allafrica.com/transport/?aa_source=aans-tags-top

[12] McFarlane K., Cross M., Gross N., Mosher D., (2010) Mitigating corruption risk Perspectives for Aerospace & Defence industry, Deloitte Development LLC.

[13] Mu'Sodiq Adekunle, (2013), Aviation Car Scandal: Jonathan Should Sell Armored Vehicles – SERAP http://dailytimes.com.ng/

[14] Nkem Osuagwu, (2013) The Unending Crises in Aviation Industry, http://www.leadership.ng/news/071113/katsina-reps-exchange-blows-over-nass-seat

[15] Nkem Osuagwu (2014), Why Professional is required as Aviation Minister, Leadership, Tuesday, February 18, 2014, page 20

[16] Onwukwe, D., 2014, The Controversial Arms Deal, Dailysun Tuesday, September 23, 2014, page 21, http://www.sunewsoline.com

[17] Simon Tumba (2013), Strange shindig in Nigeria’s aviation sector

[18] http://www.businessdayonline.com/category/personal-finance/

[19] Tidd, J., Bessant, J., & Pavitt, K., (2005) Managing Innovation Tools: SWOT Analysis. http://www.wileyeurope.com/college/tidd

[20] Udeme Ekwere , (2013), FG urged to stem corruption in aviation sector, http://www.punchng.com/am-business/

[21] Zenglein M. J., and Müller J., (2007) Non-Aviation Revenue in the Airport Business – Evaluating Performance Measurement for a Changing Value Proposition, Berlin School of Economics, WWW.GAP-PROJEKT.DE

Society for Science and Education
United Kingdom

TMLAI Transactions on Machine Learning and Artificial Intelligence

Volume 6 No. 5
ISSN 2054-7390

Optimization of a Joint Economic Lot Size Model for a First-tier Supplier with Sequential Processes Using a Genetic Algorithm

[1]DongJin Jin, [1]MoonHyeung Lee, [1]ByungDo Chung
[1]*Department of industrial engineering, Yonsei University, Seoul, Republic of KOREA;*

dj.jin@yonsei.ac.kr; mh_lee@yonsei.ac.kr; bd.chung@yonsei.ac.kr

ABSTRACT

As corporate competition intensifies in the 21st century, optimal in the integrated supply chain is more important than optimization of individual company. Our research examines a series of integrated supply chain systems comprising a single raw material supplier, a first-tier supplier with multiple processes, and a single original equipment manufacturer in the Korean automotive industry. Unlike other papers, we have studied the situation in which the first-tier supplier has an assembly process. We have also analyzed the situation in which the demand for semi-finished products occurs in the first-tier supplier process or in which semi-finished products are purchased from subcontractors and put into production. The objective function is to minimize the sum of production costs, inventory holding costs, ordering costs, and setup costs in the integrated supply chain. To solve this problem, we formulated nonlinear programming, and developed a genetic algorithm. The results showed that using a dynamic lot size is cheaper than using a fixed lot size. In addition, the lower the setup cost, the smaller the lot size, and when a certain level is reached, the lot size will be the same even if the cost changes.

Keywords: Integrated supply chain, First-tier supplier, Lot size, Nonlinear programming, Genetic algorith m

1 Introduction

The supply chain in the automotive industry comprises raw material suppliers, first-tier suppliers, original equipment manufacturers (OEMs), dealers, and distributors. Supply chain management in the domestic automotive industry is underway to optimize the supply chain structure from manufacturing procurement to delivery. OEMs use just-in-time (JIT) systems to supply the required quantity of parts in a timely manner. However, some first-tier suppliers do not use a JIT system. Consequently, first-tier suppliers are inefficient in terms of production or inventory to avoid problems of OEMs production.

In this study, we analyzed the first-tier supplier of three processes of the blank line, press line, and assembly line of companies manufacturing automobile body parts. Determining the appropriate lot size for each line is very important as it affects efficiency in the manufacturing line and directly affects inventory and production costs. If the production lot size increases, the number of setups decreases. However, a large number of works in process (WIPs) occur, resulting in inefficiency in the supply chain. On the other hand, reducing the size of the production lot reduces inventory. However, the number of

DOI: 10.14738/tmlai.65.5351
Publication Date: 28th October, 2018
URL: http://dx.doi.org/10.14738/tmlai.65.5351

setups increases, resulting in process unavailability [12]. The manufacturing industry in Korea, especially the automobile industry, thinks that the setup is the cause of productivity deterioration. Among them, replacement work of coils and molds increases the intensity of work. Owing to the nature of these industries, the production lot size cannot be minimized. This study addresses the joint economic lot size problem (JELP), which reflects the process characteristics of first-tier suppliers. Previous research has considered a two-tiered supply chain system that includes only suppliers and distributors. Recently, research has been underway to determine the lot size in a supply chain system with three or more levels. In addition, most previous studies have used two approaches to the JELP methodology. In the first, the heuristic method is used after mathematically proving the optimality condition or convexity to simplify the problem. The second method involves formulating the problem with a mixed integer non-linear programming (MINLP) model and using heuristic methods. We formulate the problem with a non-linear programming (NLP) model in which all decision variables are integers, as in the second method, and we solve the problem using a genetic algorithm. The reason for using integers is that the lot size produced by the first-tier supplier and the number of production and delivery are actually represented as integer values. In addition, the lot used once for production must be fully used.

This study considers the raw material suppliers, first-tier suppliers, and OEMs in the supply chain. The purpose of this study is to determine the lot size that minimizes the cost of the total supply chain, including production, setup, ordering, and inventory holding costs. We also propose a genetic algorithm to solve the NLP problem. This is because the computation time exponentially increases as the number of products or processes increases in the NLP model. In addition, the actual automobile parts manufacturers have the characteristic of making the production plan by using a fixed lot size. Therefore, we compare the total cost of the supply chain based on the actual production lot size of the actual automobile body parts manufacturer and the production lot size obtained by the proposed algorithm.

The remainder of the paper is organized as follows. Section 2 describes the literature review, Section 3 presents the problem definition and assumptions, Section 4 explains the modeling, Section 5 describes the algorithm, and Section 6 presents the experimental results. Finally, Section 7 offers a conclusion.

2 Literature Review

The JELP began in 1977 when Goyal [1] began to study a two-stage supply chain that involved only a vendor and a retailer. He proposed a solution to solve this problem, assuming that the producer could produce indefinitely without any restrictions on the production volume. Next, Banerjee [2] investigated problems closer to actual situations by mitigating the assumption that they could produce indefinitely from models such as Goyal. Hill [3] conducted research that restricted the production volume and placed no restrictions on the shipment policy. Alizade et al. [4] covered a supply chain comprising two or more suppliers and retailers in a two-tier supply chain. This study considered demand to be probabilistic or that lead time can be changed. In addition, they considered a situation in which the number of items in the supply chain is more than one.

The following research comprises three echelons of the supply chain. Lee [5] studied a JELP problem by minimizing the sum of the six costs resulting from focusing on the three-level supply chain. He also considered the manufacturer’s production batches and the number of orders. Munson and Rosenblatt [6] studied the JELP problem by combining quantity discounts in a three-tier supply chain comprising a single supplier, a single manufacturer, and a single retailer. Wang and Sarker [7] conducted a studied the JIT system operated by kanban and they also analyzed the number of deliveries, the lot size, and the number of appropriate kanban.

The following is a study comprising a three-echelon supply chain with several participants in each stage. Khouja [8] published a paper that expands the thesis of a three-phase supply chain in which one company participates in each stage of Mundson and Rosenblatt. Khouja dealt with the non-serial supply chain and studied a model with several participants. Since then, Jaber and Goyal [9] have discussed order quantity adjustments in a three-echelon supply chain of suppliers, manufacturers, and buyers. Bendaya and AlNassar [10] extended the model to one with several participants, reflecting Khouja's assumption that the same size lot is delivered to the next step.

Adeinat and Ventura [11] studied pricing and supplier selection as well as lot size in an integrated supply chain comprising three or more stages. The study showed that the JELP problem is applicable to various problems.

The preceding literature has confirmed that lot size problems in the integrated supply chain have been investigated. Among them, we analyze a three stage supply chain that considers raw material suppliers, first-tier suppliers, and OEMs. The first-tier suppliers comprise blank, press, and assembly processes. We also analyze lot sizes for situations in which there are two lines per process. To the best of our knowledge, there were no studies have considered the assembly line in a supply chain comprising three or more stages and situations where there is a demand or outsourcing item between first-tier suppliers process. Since some automobile parts manufacturers have assembling processes to assemble several parts to form one part, our consideration of assembly lines is an important factor in reflecting the reality.

3 Problem Definition

3.1 Problem Description

We analyze the lot size in a three stage supply chain comprising raw material suppliers, first-tier suppliers, and OEMs. The integrated supply chain that we consider is shown in Figure 1. A raw material supplier is a company that supplies raw materials for a first-tier supplier. A first-tier supplier is an automobile parts manufacturer. And, we analyze the situation of three lines (blanks, presses and assembly lines) which have two machines exist for each line. The automobile parts manufacturer's blank line is the line of processing raw materials for use in the press line, and the press line is the pre-process of the assembly line that forms the parts. The assembly line is then involved in the process of combining parts to make the finished product. There is also a storage area for all parts between all processes. OEMs have demand for semi-finished products and outsourcing materials. Outsourcing materials coming into the manufacturer's storage area can be sent to the customer or the manufacturer's internal processes. Finally, the OEMs represents companies that sell the finished automobile to customers, and first-tier suppliers deliver the quantity to meet the OEMs demand.

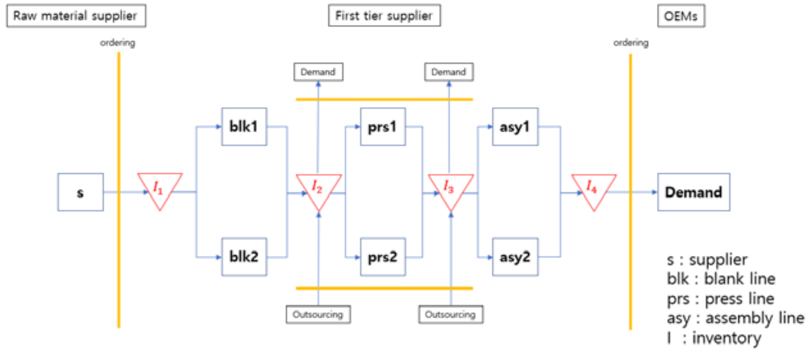


Figure 1. Integrated supply chain

3.2 Assumption

The assumptions of this study are as follows:

- The rates of production and demand are constant
- Raw materials (coils) are not scarce
- They produce 500 blank sheets using one raw material (coil)
- In the assembly line, they produce one product per line (considering BOM)
- There is no delivery lead time
- They do not allow backlogging
- The cost of setup, ordering, holding, and production are the same for each stage
- Ordering costs occur between the raw material supplier and first-tier supplier and between the first-tier supplier and OEM. For the first-tier supplier, this cost occurs when they receive the product from a subcontractor or when they ship the product to OEM
- If they use a one lot, they should use it completely

4 Modeling

4.1 Model Description

The following is a description of the mathematical model. The Bill of Material is shown in Figure 2. Each stage item represents the index of the item stored in the stepwise storage area. In Figure 2, the raw material supplier has two items, the blank line has three items, the press line has four items, and the assembly line has two items. Blank item 3 and press item 4 is the outsourcing item from the subcontractor. For example, blank item 1 is created through the blk process using the coil item 1 from the raw material supplier. In Figure 2, blank item 3 and press item 4 refers to the item produced by the subcontractor. It will be stocked in a blank or press storage area. Finally, the assembly line uses the press items 1 and press item 2 to create assembly item 1. Similarly, press item 3 and press item 4 are used to create assembly item 2. We want to know the optimal lot size in an integrated supply chain with multiple items.

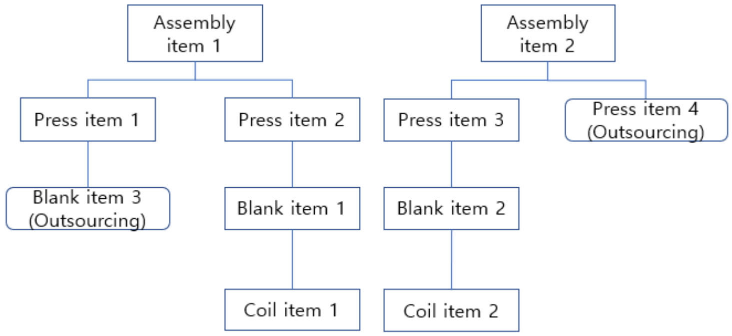


Figure 2. BOM (Bill of Materials) structure

4.2 Mathematical Model

The indices, parameters, and decision variables used in the mathematical model are summarized in Tables 1, 2, and 3. These notations are used to formulate a mathematical model. The decision variables include lot size, the number of production and delivery. These decision variables are also set to integer values because they actually have integer values.

Table 1

Indices
i : set of stage i $= 1,\ldots, I$
j : set of customers in the coil process index. j = 1, … , J
b : set of customers in the blanking process index. b =1 , … , B
c : set of customers in the press process index. c =1 , … , C
d : set of customers in the assembly process index. d = 1 , … , D
α : set of the supplier item index. α = 1 , … , e
β : set of the blank item index. β = 1, … , f
γ : set of the press item index. γ = 1 , … , r
δ : set of the assembly item index. δ = 1, … , g
ε : set of the blank outsourcing item index. ε = 1, … , v
ζ : set of the press outsourcing item index. ζ = 1, … , w

Table 2

Parameters
s_i : setup cost of i stage
h_i : inventory holding cost of i stage
o_i : ordering cost of i stage
ci_1^{α} : initial inventory of a αth item ($ci_1^{\alpha}, ci_2^{\beta}, ci_3^{\gamma}, ci_4^{\delta}$)
pc_i : production cost of i stage

D_{2b}^{β} : demand of the βth item to b-customer in the blank warehouse

D_{3c}^{γ} : demand of the γth item to c-customer in the press warehouse

D_{4d}^{δ} : demand of the δth item to d-customer in the assembly warehouse

C_i : warehouse area in i stage

p_{ij} : production rate of i-stage j-machine

T : planning horizon

SS_i : safety stock level in i-stage

L : maximum number of production and delivery

$$A_{\alpha\beta j} = \begin{cases} 1 \text{ , if the αth item is used in the j machine to create the βth item} \\ 0 \text{ , otherwise} \end{cases};$$

$$A_{\varepsilon\beta} = \begin{cases} 1 \text{ , if the } \varepsilon \text{th item is used to create the βth item;} \\ 0 \text{ , otherwise} \end{cases}$$

$$A_{\beta\gamma b} = \begin{cases} 1 \text{ , if use the } \beta \text{th item is used in the b machine to create the } \gamma \text{th item;} \\ 0 \text{ , otherwise} \end{cases}$$

$$A_{\zeta\gamma} = \begin{cases} 1 \text{ , if use the } \zeta \text{th item is used to create the } \gamma \text{th item;} \\ 0 \text{ , otherwise} \end{cases}$$

$$A_{\gamma\delta c} = \begin{cases} 1 \text{ , if use the } \gamma \text{th item is used in the c machine to create the } \delta \text{th item;} \\ 0 \text{ , otherwise} \end{cases}$$

Table 3

Decision Variables

Q_i: lot size in i-stage

k_1^{α} : the number of production for the αth item

$k_2^{\alpha\beta j}$: the number of production for the αth item produced as the βth item for the blank j-customer

$k_2^{\varepsilon\beta}$: the number of production for the εth item produced as the βth item

$k_3^{\beta\gamma b}$: the number of productions for the βth item produced as the γth item for the press b-customer

$k_3^{\zeta\gamma}$: the number of production for the ζth item produced as the γth item

$k_4^{\gamma\delta c}$: the number of production for the γth item produced as the δth item for the c-customer

m_{1j}^{α} : the number of deliveries for the αth item for blank j-machine from the supplier

m_{2b}^{β} : the number of deliveries for the βth item for the press b-customer from the blank process

m_{3c}^{γ} : the number of deliveries for the γth item for the assembly c-customer from the press process

m_{4d}^{δ} : the number of deliveries for the δth item for the d-customer for the assembly process

u^c : maximum number of items that come into the same machine to account for BOM.

Next, the objective function is to minimize the total cost, which is the sum of the production cost, setup cost, ordering cost, and inventory holding cost. The costs are expressed as follows.

$$\text{Total Cost = Production Cost + Setup Cost + Ordering Cost + Holding Cost} \quad (1)$$

Production Cost = $\sum_\alpha pc_1 k_1^\alpha * Q_1 + \sum_\alpha \sum_\beta \sum_j pc_2 * Q_2 * (k_2^{\alpha\beta j} + \sum_\varepsilon k_2^{\varepsilon\beta}) + \sum_\beta \sum_\gamma \sum_b pc_3 * Q_3 * (k_3^{\beta\gamma b} + \sum_\epsilon k_3^{\epsilon\gamma}) + \sum_\gamma \sum_\delta \sum_c pc_4 * Q_4 * k_4^{\gamma\delta c}$ (2)

Setup Cost = $\sum_\alpha s_1 k_1^\alpha + \sum_\alpha \sum_\beta \sum_j s_2 * (k_2^{\alpha\beta j} + \sum_\varepsilon k_2^{\varepsilon\beta}) + \sum_\beta \sum_\gamma \sum_b s_3 * (k_3^{\beta\gamma b} + \sum_\epsilon k_3^{\epsilon\gamma}) + \sum_\gamma \sum_\delta \sum_c s_4 * k_4^{\gamma\delta c}$ (3)

Ordering Cost = $\sum_\alpha o_1 * k_1^\alpha + \{\sum_b \sum_\beta o_2 * m_{2b}^\beta + \sum_\varepsilon \sum_\beta o_2 * k_2^{\varepsilon\beta}\} + \{\sum_c \sum_\gamma o_3 * m_{3c}^\gamma + \sum_\zeta \sum_\gamma o_3 * k_3^{\zeta\gamma}\} + \sum_d \sum_\delta o_4 * m_{4d}^\delta$ (4)

Holding Cost = $\sum_\alpha \{2ci_1^\alpha + Q_1(k_1^\alpha - \sum_j m_{1j}^\alpha)\} \frac{h_1}{2} + \sum_\beta \{2ci_2^\beta + Q_2\left(\sum_\alpha \sum_j k_2^{\alpha\beta j} + \sum_\varepsilon k_2^{\varepsilon\beta} - \sum_b m_{2b}^\beta\right)\} \frac{h_2}{2} + \sum_\gamma \{2ci_3^\gamma + Q_3\left(\sum_\beta \sum_b k_3^{\beta\gamma b} + \sum_\zeta k_3^{\zeta\gamma} - \sum_c m_{3c}^\gamma\right)\} \frac{h_3}{2} + \sum_\delta \{2ci_4^\delta + Q_4\left(\sum_\gamma \sum_c k_4^{\gamma\delta c} - \sum_d m_{4d}^\delta\right)\} \frac{h_4}{2}$ (5)

In Equation (1), the objective function is expressed as the sum of the production cost (2), setup cost (3), ordering cost (4), and inventory holding cost (5) included in the integrated supply chain. The production and inventory holding costs are calculated by the number of production, and the remaining costs are by the number of production and delivery. The production cost is calculated by multiplying production volume and production costs per unit. The setup cost is calculated by multiplying the number of production per process by the setup cost per lot. The ordering cost occurs between raw material suppliers and first-tier suppliers, as well as between first-tier suppliers and OEMs. The order cost is calculated as the number of production and shipments multiplied by the order cost per lot in this relationship. The inventory holding cost is calculated by multiplying the average inventory quantity by the inventory holding cost per unit.

Next, the constraints are as follows:

$$\begin{aligned}
&\sum_\alpha \{ci_1^\alpha + Q_1(k_1^\alpha - \sum_j m_{1j}^\alpha)\} \le C_1 \\
&\sum_\beta \{ci_2^\beta + Q_2\left(\sum_\alpha \sum_j k_2^{\alpha\beta j} + \sum_\varepsilon k_2^{\varepsilon\beta} - \sum_b m_{2b}^\beta\right)\} \le C_2 \\
&\sum_\gamma \{ci_3^\gamma + Q_3\left(\sum_\beta \sum_b k_3^{\beta\gamma b} + \sum_\zeta k_3^{\zeta\gamma} - \sum_c m_{3c}^\gamma\right)\} \le C_3 \\
&\sum_\delta \{ci_4^\delta + Q_4\left(\sum_\gamma \sum_c k_4^{\gamma\delta c} - \sum_d m_{4d}^\delta\right)\} \le C_4
\end{aligned} \tag{6}$$

$$Q_1 = \frac{Q_2}{500}, \quad Q_i \ge Q_{i+1} \qquad i = 2,3 \tag{7}$$

$$\begin{aligned}
&ci_1^\alpha + Q_1(k_1^\alpha - \sum_j m_{1j}^\alpha) \ge SS_1 && \forall \alpha \\
&ci_2^\beta + Q_2\left(\sum_\alpha \sum_j k_2^{\alpha\beta j} + \sum_\varepsilon k_2^{\varepsilon\beta} - \sum_b m_{2b}^\beta\right) \ge SS_2 && \forall \beta \\
&ci_3^\gamma + Q_3\left(\sum_\beta \sum_b k_3^{\beta\gamma b} + \sum_\zeta k_3^{\zeta\gamma} - \sum_c m_{3c}^\gamma\right) \ge SS_3 && \forall \gamma \\
&ci_4^\delta + Q_4\left(\sum_\gamma \sum_c k_4^{\gamma\delta c} - \sum_d m_{4d}^\delta\right) \ge SS_4 && \forall \delta
\end{aligned} \tag{8}$$

$m_{1j}^{\alpha} * Q_1 = k_2^{\alpha\beta j} * Q_2$ $\forall\, \alpha, \beta, j$

$m_{2b}^{\beta} * Q_2 = k_3^{\beta\gamma b} * Q_3$ $\forall\, \beta, \gamma, b = 1,2$ (9)

$m_{3c}^{\gamma} * Q_3 = k_4^{\gamma\delta c} * Q_4$ $\forall\, \gamma, \delta, c = 1,2$

$k_4^{\gamma\delta c} \geq u^c$, $u^c \geq 0$ $\forall\, \gamma, \delta, c$ (10)

$\sum_{\alpha} k_1^{\alpha} * Q_1 \leq P_1 T,$

$\sum_{\alpha} \sum_{\beta} k_2^{\alpha\beta j} * Q_2 \leq P_{2j} T$ $\forall\, j$

$\sum_{\beta} \sum_{\gamma} k_3^{\beta\gamma b} * Q_3 \leq P_{3b} T$ b = 1,2 (11)

$\sum_{\gamma} \sum_{\delta} k_4^{\gamma\delta c} * Q_4 \leq P_{4c} T$ c=1,2

$m_{2b}^{\beta} * Q_2 \geq D_{2b}^{\beta}$ $\forall \beta, b = 3$

$m_{3c}^{\gamma} * Q_3 \geq D_{3c}^{\gamma}$ $\forall \gamma, c = 3$ (12)

$m_{4d}^{\delta} * Q_4 \geq D_{4d}^{\delta}$ $\forall \delta, d = 1$

$k_2^{\alpha\beta j} = 0$, $m_{1j}^{\alpha} = 0$ $A_{\alpha\beta j} = 0$

$k_2^{\varepsilon\beta} = 0$, $A_{\varepsilon\beta} = 0$

$k_3^{\beta\gamma b} = 0$, $m_{2b}^{\beta} = 0$ $A_{\beta\gamma b} = 0$ (13)

$k_3^{\zeta\gamma} = 0$ $A_{\zeta\gamma} = 0$

$k_4^{\gamma\delta c} = 0$, $m_{3c}^{\gamma} = 0$ $A_{\gamma\delta c} = 0$

$k_1^{\alpha}, k_2^{\alpha\beta j}, k_2^{\varepsilon\beta}, k_3^{\beta\gamma b}, k_3^{\zeta\gamma}, k_4^{\gamma\delta c}, m_{1j}^{\alpha}, m_{2b}^{\beta}, m_{3c}^{\gamma}, m_{4d}^{\delta} \leq$

L integer, $\forall\, \alpha, \beta, \gamma, \delta, j, b, c, d$

$k_1^{\alpha}, k_2^{\alpha\beta j}, k_2^{\varepsilon\beta}, k_3^{\beta\gamma b}, k_3^{\zeta\gamma}, k_4^{\gamma\delta c}, m_{1j}^{\alpha}, m_{2b}^{\beta}, m_{3c}^{\gamma}, m_{4d}^{\delta} \geq$ (14)

0 integer, $\forall\, \alpha, \beta, \gamma, \delta, i, j, b, c, d$

$Q_i \geq 1$ integer, $\forall\, i$

Equation (6) is related to the storage space per process. Equation (7) shows that the lot size of the preceding process is greater than or equal to the lot size of the next process. Equation (8) demonstrates the inventory balance constraint that the sum of the initial inventory and production must be greater than the sum of the safety inventory level and shipment. Equation (9) shows that the quantity of products entering the production process must be equal to the number of quantity at the time of production completion. Equation (10) relates to the BOM, and Equations (11) and (12) show that the maximum production per line cannot be exceed. The constraint in Equation (13) indicates that the raw material or semi-finished product is produced as a semi-finished and finished product in a specific blank, press, or assembly line. Finally, Equation (14) limits the number of production and delivery times, and the decision variable has a non-negative integer value. In addition, lot sizes have an integer value greater than 1.

5 Algorithm

5.1 Algorithm Description

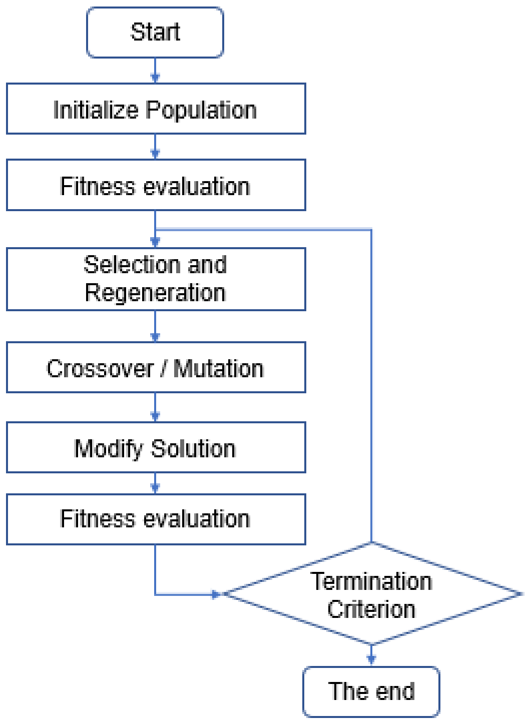


Figure 3. Algorithm flow chart

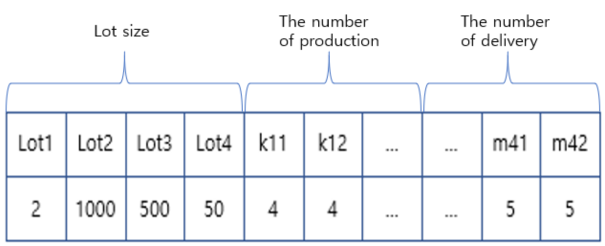


Figure 4. Chromosome expression

We propose a genetic algorithm to solve the NLP problem. A flowchart of the algorithm used in this study is shown in Figure 3. The chromosome comprises the actual values of the decision variables, as shown in Figure 4. The first line shows the location and meaning of the decision variable, and the second line shows the actual value of the decision variable used in the algorithm.

The initial solution is generated as follows: First, the lot size of the assembly process is calculated as a multiple of the pallet size, and the lot size of the press process is then calculated as a multiple of the lot size of the assembly process. When the lot size of the press process is determined, the lot size of the blank process can be calculated in the same manner, and the lot size of the raw material supplier can be obtained. The reason for using a multiple is to make the decision variables have integer values. The reason for using integers is that if one lot is used at the factory, it will not be possible to interrupt the work within the lot in use and proceed with other work. The lot size is determined in the following manner, and the number of production and delivery is determined by the constraints that the modeling reflects. For example, in the assembly process, the quantity delivered per item must be greater than or equal to the OEM demand; thus, the number of deliveries can be determined. In addition, the number of production through assembly can be calculated by the inventory balancing constraint. The number of deliveries of the press can be obtained because of the constraint that the input quantity to the assembly process is equal to the output quantity. Likewise, we can obtain the number of production and delivery for all processes. We calculate the lot size, the number of production, and the number of delivery to complete a single solution.

Next, we evaluate the fitness of each initial solution based on the initial 20 solutions. The fitness uses the reciprocal of the cost. This is because fitness increases as costs decrease. Owing to the fitness assessment, 10 solutions, 50% of previous generations' good solutions, survive in the next generation, and the remaining 50% of solutions are regenerated using the initial solution generation method.

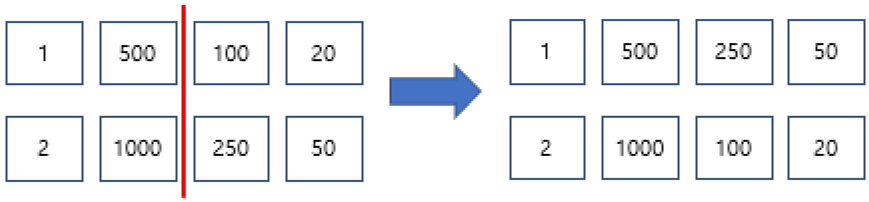


Figure 5. One-point crossover

The crossover operator is performed using the next-generation solution created through selection and playback. The crossover method uses a one-point crossover. In addition, the crossover method uses only 4 lot sizes. The reason is that if the lot size changes, the number of production and delivery will change. The crossover method randomly selects two solutions from the population and determines the location of the lot size for intersection. It then intersects the two solutions at this location. For example, the crossover operator is shown in Figure 5. The red line indicates the point at which the one-point crossover is performed and indicates that the two-parent solutions of [1, 500, 100, 20] and [2, 1,000, 250, 50] are changed to two offspring solutions of [1, 500, 250, 50] and [2, 1,000, 100, 20].

In the mutation operator, one solution is selected from the population in which the crossover has progressed. Next, we select the lot size that one randomly would want to change for the selected solution and change the lot size. The mutation proceeds by changing the lot size of selected process to an arbitrary value that satisfies a divisor relation to the lot size of the previous process and a multiple relation to the lot size of the next process. This is because all decision variables must have integer values. Once the lot size is determined, the number of production and delivery is determined to satisfy the constraint, and all solutions in the population will have a feasible solution.

In the modify solution, a local search is performed with the population in which the crossover and mutation operators have been carried out. We select one solution from the population and choose one lot size from among the four lot sizes. We then change all of the lot sizes of the selected processes to determine the lot size value that satisfies a divisor relation to the lot size of previous processes and a multiple relation to the lot size of the next process. Then, we calculate the number of production and delivery. In this way, we can create multiple solutions with a divisor multiple relationships and replace the originally selected solution with a modified solution with the best fitness value.

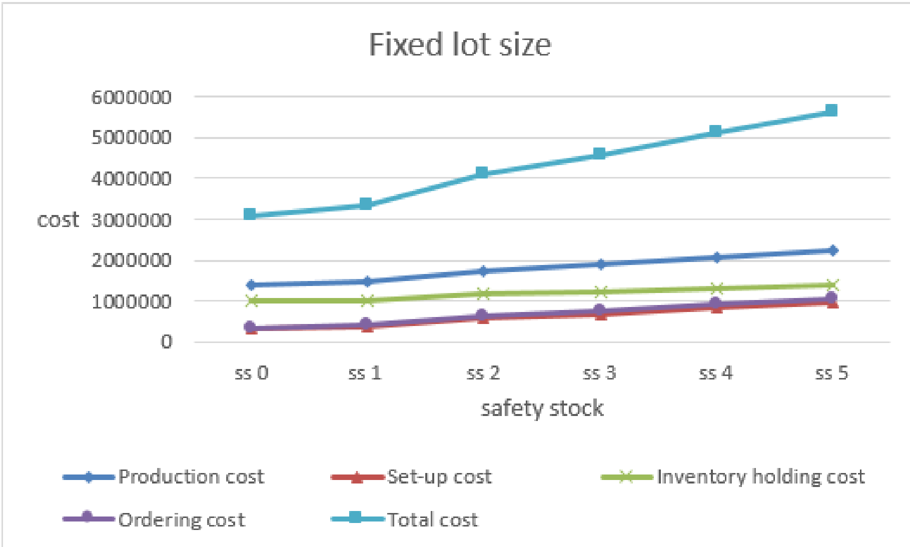


Figure 6. Fixed lot size

After performing the modify solution steps, all 20 solutions are evaluated for fitness. In this step, we assess the fitness of each solution and stored the best solution. Once all the preceding steps are completed, we check the algorithm's termination condition, and it then terminated when the same best solution was repeated 20 times or when the number of repetitions reached 500. Otherwise, we increased the number of iterations by 1 and returned to the selection and regeneration stage and repeat the process.

6 Experiment Results

After formalizing the NLP model, we using a genetic algorithm to solve the NLP model and explain the experimental results. The cost data included in the experiment were used for the following reasons: The production cost per unit was calculated with respect to the product production time. Production time was based on the actual first-tier supplier's production time. However, because the coil, the raw material, is assumed to produce 500 blank products using a one coil, the production cost of the coil was multiplied by 500 to determine the production cost of the blanking product. The setup cost per lot was based on the setup time used by the actual first-tier supplier process. Inventory holding costs were calculated as the volume of individual products. We assume that the same volume per lot was used for each process. The volume is used as the number of products per lot divided by the number of products in the palette. Finally, the ordering cost per lot is the number of individual products included in the lot. The volume, quantity, time, and the actual lot size used here were based on actual first-tier suppliers.

First, some real companies use the fixed lot size. However, if demand or safety stock levels change, using fixed lot sizes may result in inefficiencies. Hence, we compared the case of taking the lot size fixedly and the case of taking the lot size dynamically. We ran an experiment on two cases and checked for cost changes.

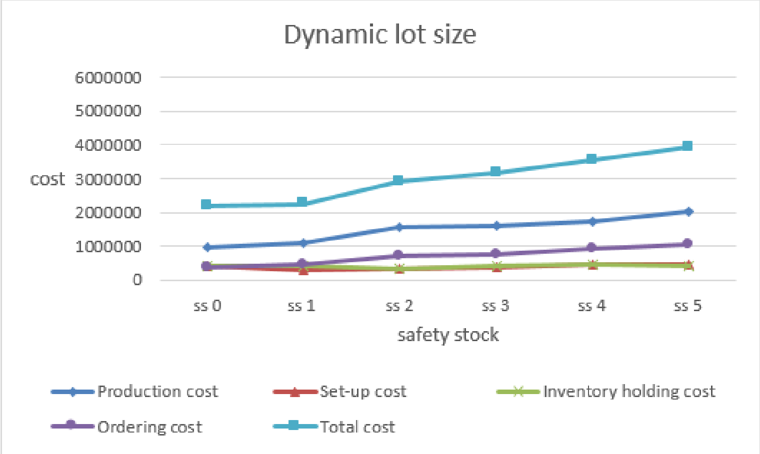


Figure 7. Dynamic lot size

Figures 6 and 7 show a comparison between the fixed lot size and the dynamic lot size. In the Figures 6 and 7, ss 0 represents the result of setting the safety inventory level to 0; ss 1 represents the result of setting the safety inventory level to 1 day, and ss 2, 3, 4, and 5 represent the result of setting the safety inventory level to 2, 3, 4, and 5 days, respectively. The meaning of safety inventory level 1 is set to the production quantity of one day.

The results of the experiment are as follows. We compared the use of fixed lot sizes and the use of dynamic lot sizes. Using a dynamic lot size rather than a fixed lot size reduces the total cost. The reason for this is that it changes the lot size to determine the lot size of the lowest cost. Experiments were performed while increasing the level of safety inventory. Figures 6 and 7 show that the difference between the total cost and the other costs increased slightly when using the dynamic lot size, while the fixed lot size showed a sharp increase in the difference. if one uses a fixed lot size, the higher the safety stock level, the higher the cost; thus, a dynamic lot size is better than fixed a one if a safety stock is needed.

Second, because settings are important in real companies, we experimented to determine how shortening the setup time can affect the cost and lot size. The experiment identified changes in other costs and lot sizes while changing the setup cost.

Table 4. Cost of changing the setup

	Change in setup cost							1 / 10
	10	7	5	3	1	0.9	0.8	Ratio
Production cost	1,623,000	1,539,000	1,285,200	1,077,000	1,077,000	1,077,000	1,077,000	0.66 (-34%)
Setup cost	1,872,000	1,386,000	1,183,500	710,100	299,700	269,730	239,760	0.16 (-84%)
Ordering cost	261,180	274,680	330,480	330,480	415,080	415,080	415,080	1.59 (+59%)
Holding cost	500,464.50	490,009.50	481,564.50	481,564.50	466,077	466,077	466,077	0.93 (-7%)
Total cost	4,256,645	3,689,690	3,280,745	2,807,345	2,257,857	2,227,887	2,197,917	0.53 (-47%)
Sum of other cost	2,384,645	2,303,690	2,097,245	2,097,245	1,958,157	1,958,157	1,958,157	0.82 (-18%)

Table 5. Lot size according to the setup cost change

	Change in setup cost						
	10	7	5	3	1	0.9	0.8
Lot size of raw material process	27	27	27	27	26	26	26
Lot size of blank process	13,500	13,500	13,500	13,500	13,000	13,000	13,000
Lot size of press process	500	450	300	300	200	200	200
Lot size of assembly process	100	90	100	100	100	100	100

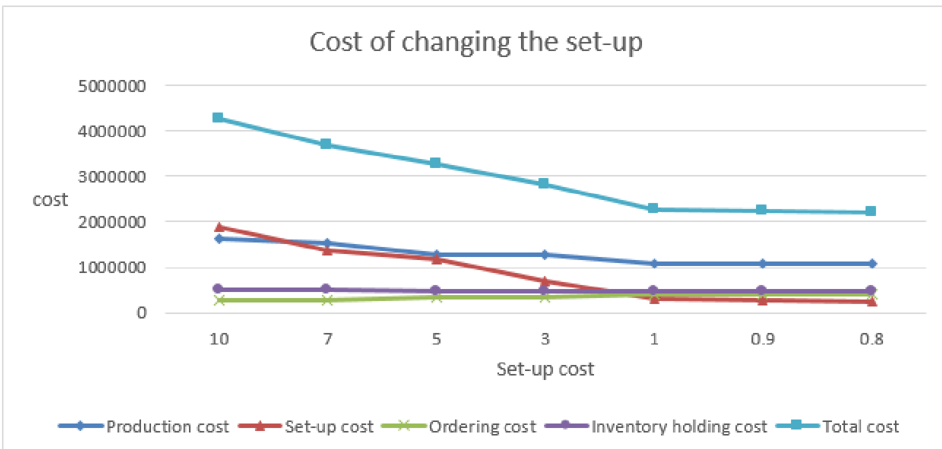


Figure 8. Cost of the changing setup

Figure 8 and Table 4 show changes in costs due to changes in the setup cost, and Table 5 shows changes in lot size due to changes in the setup cost. The change in the setup cost in Tables 4 and 5 indicates that the parameter value of the other cost is fixed and that only the parameter value of the setup cost is changed. For example, 10 is the result of a 10-fold increase in setup costs, while other costs are fixed, and 1 is the initial cost without any changes. The ratio represents the value of 1 divided by 10 in the setup cost change. If the ratio is greater than 1, the cost is increased. If the ratio is less than 1, the cost is decreased. Finally, the sum of the other costs is the total sum of the costs, excluding the setup costs.

The experimental results are as follows. Reducing the setup cost will reduce the production cost by 34%, inventory holding cost by 7%, and total cost by 47%. In addition, the sum of the remaining cost and the setup cost is reduced. Reducing the setup cost from 10 to 1 reduces the sum of the remaining costs by 18%. On the other hand, the ordering cost will increase by 59%. This is because the lot size becomes smaller as the setup cost decreases. For example, when reducing the setup cost from 10 to 1, the production cost decreases from 1,623,000 to 1,077,000, which means that the production cost is reduced by 34%. In addition, if the setup cost is reduced from 10 to 1, then the inventory holding cost is reduced from 500,464.5 to 466,077, which means that the inventory holding cost decreases by 7%. Moreover, if the setup cost is decreased from 10 to 1, then the ordering cost increases from 261,180 to 415,080, which means that the ordering cost increases by 59%. Also, if the setup cost is decreased from 10 to 1, the sum of the other cost (excluding the setup cost) is reduced from 2,384,645 to 1,958,157, which means that the sum of the other cost is reduced by 18%.

In addition, even if the setup cost is reduced, the lot size is the same if it falls below a certain level. For example, if the setup cost change is 10, the lot size is [27, 13,500, 500, 100], and if it is 1, it is reduced to [26, 13,000, 200, 100]. On the other hand, for 0.9 and 0.8, the lot size is the same as 1 for [26, 13,000, 200, 100]. It is therefore important to find and manage specific levels.

7 Conclusion

This study analyzes the lot size of a three stage supply chain comprising raw material suppliers, first-tier suppliers, and OEMs. To the best of our knowledge, there is no paper has considered the assembly line in an integrated supply chain. Some automobile first-tier suppliers assemble parts in the final assembly process to make the finished product. In this study, we built a supply chain comprising several products with assembly lines, unlike most previous research considering a single product. Most JELP research have not considered outsourcing or shipping between production processes; however, we consider outsourcing or shipping products between production processes. Therefore, the contributions of this paper are that it conducts an analysis of a supply chain that includes the assembly process and that it considered outsourcing and semi-finished products between the first-tier supplier processes.

We propose an NLP model and solve the problem by using a genetic algorithm. The results are based on scenarios. First, using a dynamic lot size rather than a fixed lot size reduces the total cost. In addition, as the level of safety inventory increases, the gap between the total cost and other costs slightly increases. Second, if one can shorten setup costs by reducing the setup time, the lot size will be smaller, and the total cost will be lower. However, if the lot size falls below a certain level, the lot size will be the same even if the setup cost is reduced. It is therefore important to find and manage specific levels of costs.

The future research directions of our study are as follows. This study accounts for the multilevel supply chain of a single company; thus, the number of participants analyzed can be increased. It is also possible to increase the number of items or lines in the supply chain. Finally, the present research can be extended to a problem that considers uncertain situations.

ACKNOWLEDGMENTS

This work was supported by the World Class 300 Project (R&D) (Project number S2482274, Development of Multi-Vehicle Flexible Production Platform for Future Smart Body Factory (2/5)) of the MOTIE, MSS (Korea).

REFERENCES

[1] GOYAL, S. K. *An integrated inventory model for a single supplier-single customer problem*. The International Journal of Production Research, 1977, 15.1: 107-111.

[2] BANERJEE, Avijit. *A joint economic-lot-size model for purchaser and vendor*. Decision Sciences, 1986, 17.3: 292-311.

[3] HILL, Roger M. *The optimal production and shipment policy for the single-vendor single buyer integrated production-inventory problem*. International Journal of Production Research, 1999, 37.11: 2463-2475.

[4] TALEIZADEH, Ata Allah; NIAKI, Seyed Taghi Akhavan; BARZINPOUR, Farnaz. *Multiple-buyer multiple-vendor multi-product multi-constraint supply chain problem with stochastic demand and variable lead-time: a harmony search algorithm*. Applied Mathematics and Computation, 2011, 217.22: 9234-9253.

[5] LEE, Wenyih. *A joint economic lot size model for raw material ordering, manufacturing setup, and finished goods delivering*. Omega, 2005, 33.2: 163-174.

[6] MUNSON, Charles L.; ROSENBLATT, Meir J. *Coordinating a three-level supply chain with quantity discounts*. IIE Transactions, 2001, 33.5: 371-384.

[7] WANG, Shaojun; SARKER, Bhaba R. *Optimal models for a multi-stage supply chain system controlled by kanban under just-in-time philosophy*. European Journal of Operational Research, 2006, 172.1: 179-200.

[8] KHOUJA, Moutaz. *Optimizing inventory decisions in a multi-stage multi-customer supply chain*. Transportation Research Part E: Logistics and Transportation Review, 2003, 39.3: 193-208.

[9] JABER, M. Y.; GOYAL, S. K. *Coordinating a three-level supply chain with multiple suppliers, a vendor and multiple buyers*. International Journal of Production Economics, 2008, 116.1: 95-103.

[10] BEN-DAYA, M.; AL-NASSAR, A. *An integrated inventory production system in a three-layer supply chain*. Production Planning and Control, 2008, 19.2: 97-104.

[11] ADEINAT, Hamza; VENTURA, José A. *Integrated pricing and lot-sizing decisions in a serial supply chain*. Applied Mathematical Modelling, 2018, 54: 429-445.

[12] LIM, I. J.; PARK, K. S.; KIM, J. H. *Determining Optimal Lot Size for Batch Processes Using TOPSIS*. Management Education Research, 2015, 30: 163-185.

SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM

TMLAI TRANSACTIONS ON MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

Volume 6 No. 5
ISSN 2054-7390

Enhanced Handover Clustering and Forecasting Models Based on Machine Learning and Big Data

Luong-Vy Le[1], Bao-Shuh Paul Lin[2,3], Li-Ping Tung[3] , Do Sinh[2]
[1]*College of Electrical and Computer Engineering, National Chiao Tung University, Hsinchu, Taiwan*
[2]*Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan*
[3]*Microelectronics & Information Research Center, National Chiao Tung University, Hsinchu, Taiwan*
leluongvy.eed03g@nctu.edu.tw, bplin@mail.nctu.edu.tw, lptung@nctu.edu.tw, dosinhuda.cs04g@nctu.edu.tw

ABSTRACT

In mobile networks, handover (HO) is one of the most important and complex KPIs (Key Performance Indicators), which directly affect to Quality of Service (QoS), Quality of Experience (QoE), and mobility performance. Moreover, its configuration parameters such as handover thresholds and handover neighbor lists are the key factors for implementing network optimization such as load balancing and energy saving. In a study before, the authors proposed clustering and forecasting models using ML algorithms and Time Series models to cluster, forecast, and manage the HO behavior of a huge number of cells. In this study, on the other hand, the authors firstly investigated more network KPIs to analyze their relationship with HO KPIs, and then, proposed new clustering, forecasting, and abnormal detection models that are expected to make them much more comprehensive. Finally, the performances of the proposed models were evaluated by applying them to a real dataset collected from the HO KPIs and other KPIs of more than 6000 cells of a real network during the years, 2016 and 2017. The results showed that the study was successful in identifying the relationship among network KPIs and significantly improving the performance of the HO clustering, forecasting, abnormal detection models. Moreover, the study also introduced the integration of emerging technologies such as machine learning (ML), big data, software-defined network (SDN), and network functions virtualization (NFV) to establish a practical and powerful computing platform for future self-organizing networks (SON).

Keywords: key performance indicators (KPIs); handover, Machine Learning; clustering, forecasting; SDN/NFV, SON; 5G; big data.

1 Introduction

In 5G networks, dense heterogeneous architectures of macrocells and small cells, are expected as a promising solution to overcome the exponential growth of broadband traffic (1000-fold capacity improvement) produced by massive IoT devices (over 25 billion devices are connected by 2020) and new vertical business services (e.g., healthcare, augmented reality, content delivery network (CDN), automotive system, entertainment). Although, the deployment of a huge number of small cells of ultra-dense networks (UDN) brings such benefit for network operators in improving indoor coverages, service quality, system capacity, energy saving, spectral efficiency, and the cost of expenditure, it carries many challenges for network management that may cause of the increase in the operational cost. Therefore,

DOI: 10.14738/tmlai.65.5411
Publication Date: 30th October, 2018
URL: http://dx.doi.org/10.14738/tmlai.32.5411

the SON in 5G must be significantly improved from the current 4G SON and pushed its functions to a next level that be able to provide full intelligence, faster computation, automatic management and optimization to fulfill network QoS and QoE [1][2].

In mobile networks, KPIs are used to judge the network performance, evaluate network operation quality, and statistical network traffic. Generally, they are collected on the hourly period and daily period by Operation and Maintenance Center (OMC) or by counters located at eNodeBs[3][4]. Their values are extremely important for network optimization to ensure that the system is operating normally at the peak of performance through evaluating the success or failure rate of different indicators reflecting the QoS regarding user perspective. For example, Radio Resource Control (RRC) success rate represents for the connection setup success rate, and Handover outgoing success rate (HOSR) represents the proportion of total number handover attempts that result in successfully completed handovers.

Recently, ML, big data, cloud computing, and SDN/NFV have been introduced as emerging technologies for the SON to deal with such requirements by reinforcing the SON talent in processing a massive data of HetNets and UDNs [5][6][7][8][9][10]. Moreover, the current development of many robust platforms like Apache Spark, Kafka, Zookeeper, IBM InfoSphere, and SDN controller (e.g., ONOS, OpenDaylight) carries a great opportunity to empower the SON with intelligence to make it much more comprehensive. This done to fully shift from being reactive to being a proactive SON with minimum human intervention by integrating various capacities: self-configuration, self-optimization, and self-healing. These functions facilitate the manual workload of the network and make the network management more economically. Recently, many researchers focused on developing ML, big data, and SDN/NFV models analyzing network KPIs to empower SON. For example, research [11] introduced a framework to empower the SON (called BSON) based on network KPIs and ML algorithms to solve challenges in the 5G SON; research [3] proposed and analyzed adaptive SON management using KPI measurements; study [4] analyzed the impact of SON function on the KPI behavior in realistic large-scale network scenarios; research [12] proposed and implemented a HO procedure based on a 5G SDN-based network architecture; moreover, research [13][14][15] investigated and developed many SDN-based architectures and algorithms for managing HO in HetNets.

In the previous study [1], we showed that HO management plays a crucial role in improving network quality, management efficiency, and mobility performance in mobile networks[1]. For example, an HO clustering model identifies cells that their HO behaviors are similar. As a result, it can support the self-configuration to provide plug-and-play functions, such as when a new component is deployed in the RAN (radio access network), it automatically configures HO threshold, radio parameters, and neighbor cell list [16][17]. On the other hand, an HO forecasting model that precisely predicts the future HO demand of cells can assist the self-optimization [18][19] to optimize and adjust network parameters (NPs) and handover parameter in response to real-time circumstance to guarantee that the system is working at its peak performance. Meanwhile, an abnormal HO detection model can support the self-healing to monitor the network and trigger rapid fault recovery by diagnosing the failure and then taking appropriate compensation mechanisms such as change the NPs to keep the network operating smoothly.

In this study, the authors propose several applications based on different ML algorithms for managing HO. Firstly, the relationship of HO and other KPIs is analyzed to identify KPIs that primarily affect to HO behaviors. Secondly, an HO clustering model is proposed to extract cells having the same HO patterns by exploring the complicated HO behavior of regularities in a dataset, and then, the HO patterns of the clusters and cells will be extracted. After that, we propose an HO forecasting model to forecast the future HO demand in cells or clusters based on several important KPIs using various ML algorithms. Besides, based on the results of clustering and forecasting models, we propose an effective abnormal HO detection model. Finally, the performance of these models will be evaluated and compared with those in the previous study.

The remainder of the paper is organized as follows: Section 2 introduces the experimental framework and computing platform based on SDN/NFV, big data, and ML; Section 3 reviews the network KPIs, HO characteristics, and ML applications for HO management; Section 4 analyzes the relationships of HO and other KPIs; Section 5 proposes and applies a new HO clustering model; Section 6 proposes and implements new HO forecasting models; Section 7 introduces an abnormal HO detection model based on clustering and forecasting results; and finally, Section 8 concludes the present study.

2 The Integration of Big Data, ML, and SDN/NFV to the 5G experimental platform

SDN/NFV, big data, and ML are considered as key ingredients and solutions for 5G to provide such significant capacities such as softwarization, virtualization, automation, high broadband bandwidth, etc. Specifically, in studies [1][12][13][14][15], they were applied to support new 5G architecture for developing various HO algorithms to perform low-latency handover.

2.1 Applying Big Data, ML, and SDN/NFV to 5G platform at NCTU

Recently, we have applied those technologies to establish a service-oriented architecture for the 4G/LTE&5G testbed located at Broadband Mobile Lab (BML), National Chiao Tung University (NCTU) [2][10][20][21][22][23]. There are essential revolutions in the 5G's main components as described in Fig.1: The C-RANs (Centralized/Cloud-Radio Access Network), Open5GSON, Open5GCore, and applications. The first revolution is in the C-RAN, which exploits small-cells, massive MIMO, and optical fibers to support low latency, high capacity, high-speed connections, cost-effective, and greener communication for a wide-area wireless connectivity of various types access technologies like LTE-E-UTRAN, UMTS-UTRAN, and Wireless Sensor Network (WSN). For example, in this architecture, a BBU (Base-Band Units) can connect to multiple RRUs (Remote Radio Heads) through optical fibers. The second revolution is in the Open5GCore, which is specifically illustrated as in research [23], in which Evolved Packet Core (EPC) is virtualized using hypervisor and container (Docker container) and its components such as P-GW, S-GW, Authorization and Accounting (AAA), home subscriber server (HSS), and mobility management entity

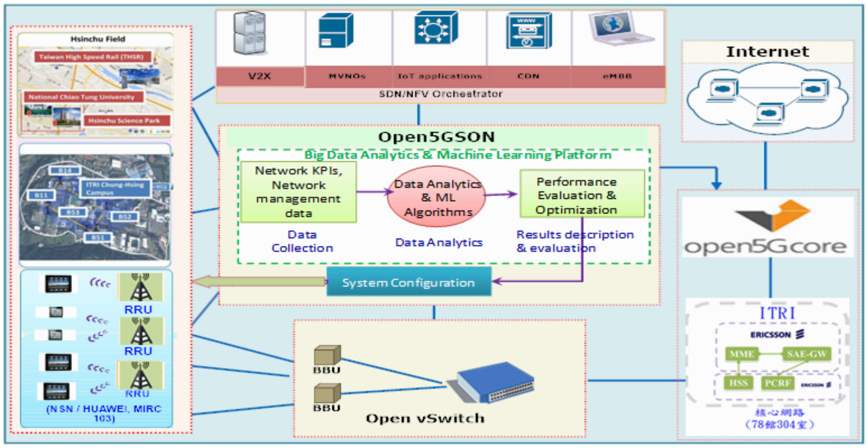


Figure 1. Experimental platform for Big Data analytics & ML at NCTU

(MME) are run on commodity hardware under the management of SDN/NFV orchestrations. As a result, in each component, the control plane is decoupled from the data plane, for example, the P-GW will contain the P-GW-C and P-GW-U. The third significant revolution is in the application layer in which SDN/NFV are deployed to enable a better support for vertical and diversified 5G services under a single network infrastructure [8]. Each application is a virtual entity deployed in a Docker container controlled by SDN/NFV orchestrations. Therefore, the deployment of 5G-based services become more flexible and efficient, such as end-to-end network slicing, network management as a service, IoT services, and massive machine type communications. The final crucial revolution is in the computing platform (e.g., Open5GMEC or Open5GSON), which works as a new component interacting with both the 5G RAN and Core to provide robust and real-time computing capabilities. It can be considered as a promising solution to simplify the core and eliminate a massive amount of data routing and processing at the core. As a result, core components can be simplified. The following section presents the computing platform in detail.

2.2 Applying Big Data and Machine Learning to the Computing Framework

The proposed computing platform described in Fig.2 is driven by 5G service requirements.

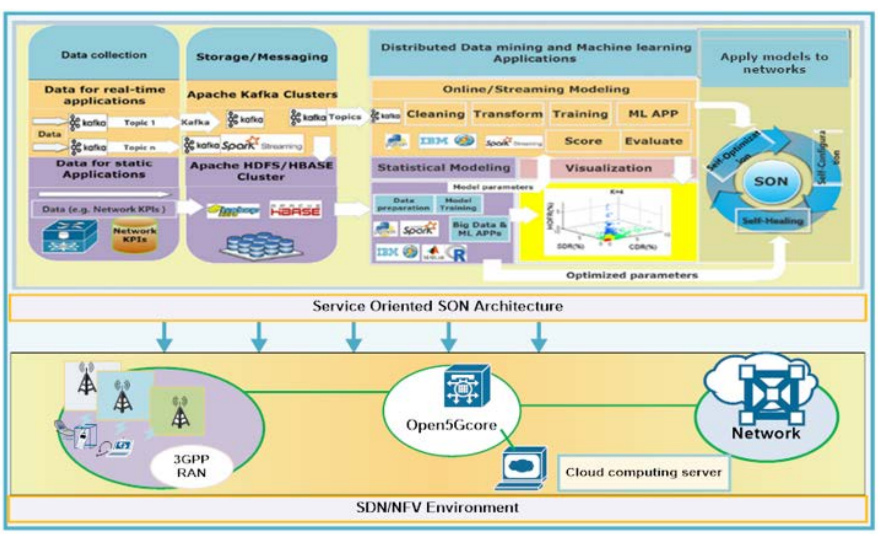


Figure 2. Applying Big Data & ML model to computing framework

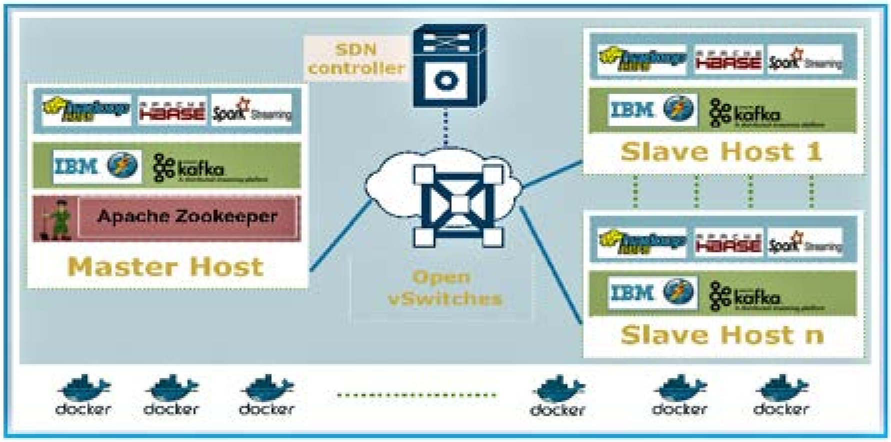


Figure 3. Distributed computing platform

It aims to provide a comprehensive SON with powerful capacities for different types of 5G applications and functions by interacting with the RAN and CORE to modify their configurations. Since it is necessary to collect and process a huge amount of data in a timely and autonomous manner to build SON applications, the computing platform integrates and utilizes state-of-the-art software, computing platforms, and programming languages, for different purposes. For example, Apache Kafka, a robust distributed streaming platform with high fault-tolerant, is used for building real-time streaming applications and data pipelines using publish-subscribe-based messaging; R can support data processing such as filtering, clearing, extraction, aggregation, visualization, and building ML and data mining models; finally, other platforms like HBase, Spark, InfoSphere, ZooKeeper, Matlab are used for data collection, storage, transformation, processing, etc.

The practical deployment of those components is a distributed computing platform as described in Fig.3, in which a host runs as the master and many hosts work as workers or executors. They connect to each other and work in the SDN/NFV environment under the control of an SDN controller. SDN/NFV program and control the network easily because the control plane is centralized with a global view. Moreover, as can be seen, in a physical machine, we can deploy multiple software components concurrently and independently, and each software component (e.g., Kafka) can work separately in its cluster under the monitor of ZooKeeper and SDN/NFV orchestrations. This makes the computing framework more powerful with full of intelligence and automation providing crucial capabilities, such as flexibility, reliability, scalability, and high computing performance. Moreover, since the components in each host run on independent Docker containers running on Linux systems of a sharing the OS' kernel, the deployment of the components becomes instant and efficient without downtime, while consuming small resources. On the other hand, each entity also provides a different approach for fault tolerance such as Apache Kafka can use partition and replication methods to improve its reliability, another example is in the case of a ZooKeeper cluster, the backup master will substitute the role of the primary master when a failure occurs in the primary master.

In the platform, a client can submit its jobs to the master or any worker, and once a computing application is submitted, the master of each relating component (e.g., Spark) distributes the job to its workers, and based on the computing load and the available computing resources, the master automatically adjusts and optimizes number of executors. In summary, this distributed computing for Open5GSON enables parallel data processing and building big data and machine learning models for on-demand services.

3 Handover KPIs and ML-Based Applications

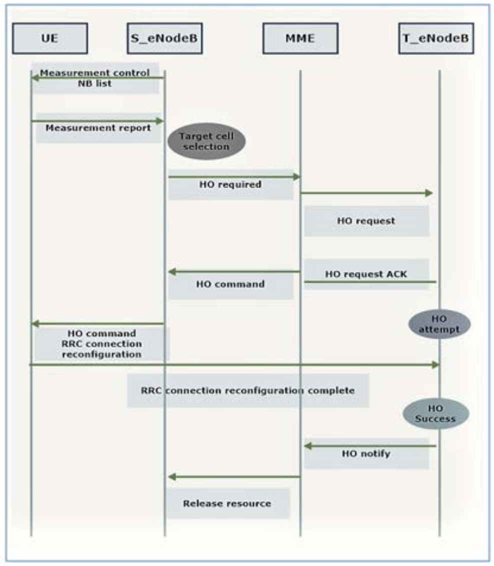


Figure 4. Inter-eNodeBs handover process

In cellular networks, handover KPIs are used to evaluate the mobility performance, which is significant to customer experience or QoE. The HO KPIs are defined for different handover types such as intra-frequency, inter-frequency, intra-eNodeB, inter-eNodeB, and inter-RAT (Radio Access Technology). This section illustrates the HO procedure, how to calculate HO KPIs, describes HO characteristics, and introduces several ML applications to manage HO KPIs.

3.1 Handover Process and KPIs

Fig.4 describes inter-eNodeBs HO process, this is the most popular HO type in cellular networks. Firstly, the UE is requested to send its measurement reports relating to some parameters such as the received signal level known as RSRP (Reference Signal Received Power) and RSRQ (Reference Signal Received Quality) of the serving cell and neighboring cells for the serving cell. If the received power and quality from the serving cell is less than a defined threshold (or much less than other cells a threshold value), the serving eNodeB selects the best target cell from the candidate list and triggers HO procedure by sending HO required message to MME, and the MME sends HO request to the target eNodeB [18] [16]. The latter will determine if there are available resources to provide for the new user. In the favorable case, the response messages are sent to the MME and the serving eNodeB, which then send the HO command to

Table 1. Examples of network KPIs

Date	Time	3G Cell name	No-SHO Attempt	No_SHO_Suc	SHOSR(%)	No_HHO Attempt	No_HHO Suc	HHO SR(%)	CS Traffic (Erl)	PS Traffic (MB)	RAB Success	RAB Attempt	2G Cell name	No SDCCH Attempt	No TCH Ass Attempt	No Incomming HO Attempt	No Outgoing HO Attempt	No SDCCH Attemp	Traffic Volume on SDCCH	TCH Traffic
04.07.2017	0:00	3HSC001	472	472	100	58	58	100	1.3333	287.4	58	58	QN001	602	17	190	280	1506	601	11.67
04.07.2017	1:00	3HSC001	269	269	100	15	15	100	1.575	653.24	139	139	QN001	347	13	110	140	1078	346	11.53
04.07.2017	2:00	3HSC001	131	131	100	12	12	100	1.4167	235.75	149	149	QN001	280	1	20	20	1118	279	1.94
04.07.2017	3:00	3HSC001	115	115	100	16	16	100	0.5583	127.71	109	109	QN001	278	6	10	60	1295	277	2.22
04.07.2017	4:00	3HSC001	245	245	100	40	40	100	1.1917	1014.4	139	139	QN001	308	5	80	70	1565	307	4.17
04.07.2017	5:00	3HSC001	238	238	100	14	14	100	1.1583	496.86	141	141	QN001	484	22	230	320	1203	483	35.28
04.07.2017	6:00	3HSC001	386	386	100	45	44	97.78	0.0167	224.66	75	75	QN001	1231	184	2080	2800	1311	1230	204.03
04.07.2017	7:00	3HSC001	853	853	100	96	96	100	1.0583	38.311	69	69	QN001	2682	265	3310	4030	1467	2681	278.89
04.07.2017	8:00	3HSC001	1376	1376	100	309	308	99.68	0.275	17.809	13	13	QN001	3534	436	5360	6430	1459	3533	430.56
04.07.2017	9:00	3HSC001	1637	1637	100	345	345	100	1.8667	101.02	131	131	QN001	3226	303	3520	4250	1479	3225	282.92
04.07.2017	10:00	3HSC001	2106	2106	100	543	543	100	1.4083	37.67	71	71	QN001	2334	321	4240	5310	1861	2333	407.92
04.07.2017	11:00	3HSC001	2106	2106	100	547	545	99.63	0.3917	26.91	14	14	QN001	1966	258	3530	4290	1434	1965	312.78
04.07.2017	12:00	3HSC001	1910	1908	99.9	380	380	100	1.075	190.73	135	135	QN001	1649	170	3260	3760	1557	1648	329.58

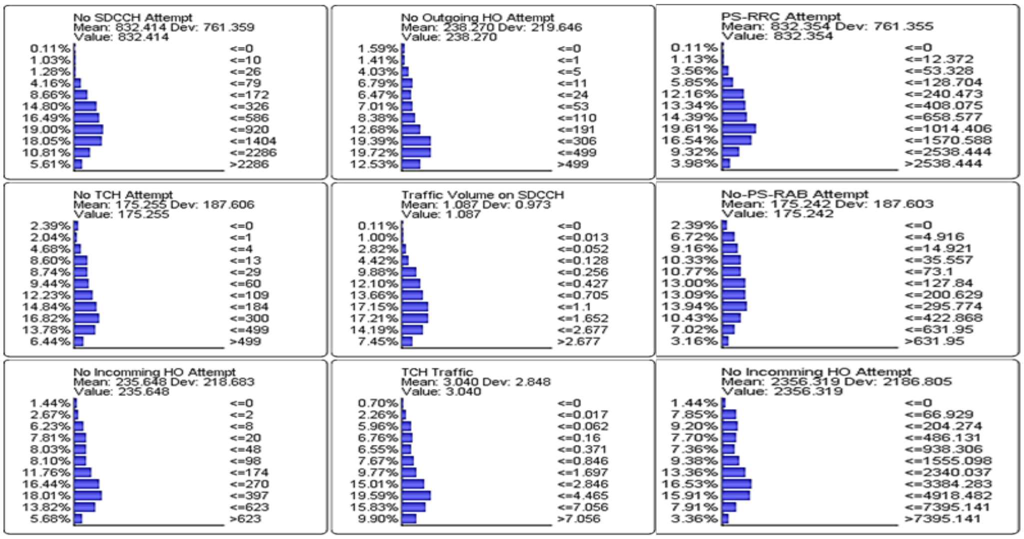


Figure 5. Marginal probability distributions of HO and several KPIs

confirm the target cell and RRC reconfiguration information for the UE. Finally, the connection is switched to the target cell, which then notifies to the MME to release the old resources at the serving eNodeB. During the HO procedure, there are several important HO KPIs are collected at both eNodeBs: the source eNodeB counters collect outgoing HO KPIs, and the target eNodeB counters collect incoming HO KPIs. For example, in Fig.4, the incoming HO attempts is calculated at the point when the target eNodeB sends the Handover Request Acknowledge message to the MME; and the number of successful incoming HO is measured at the point when the target eNodeB receives the RRC Connection Reconfiguration Complete message sent from the UE. Table 1 is an example of several typical KPIs and their hourly values of a 3G cell and a GSM cell, from early morning to noon duration. They represent for number of incoming HO attempts, number of successful incoming HO, incoming HO success rate for soft HO and hard HO; voice traffic (CS traffic), data traffic (PS traffic); number of RAB (Radio Access Bearer) attempts and success, number of TCH (traffic channel) attempts, etc. Moreover, to summarize the distribution values of several important KPIs, their hourly values were clustered in 10 groups by applying K-means algorithm, and the marginal probability distributions are shown in Fig.5, which provides a deeper understanding of the characteristics of these KPIs. This monitor histogram also illustrates a comparison of KPIs statistical values, such as the probability and value of each cluster, the mean values, and the deviation values.

3.2 Applying Machine Learning to HO Management

Recently, Machine learning widely exhibited as the breakthrough solution to reinforce 5G SON and MEC talent for solving various problems automatically [1][2][17]. Fig.6 introduces possible ML-based applications for managing HO in mobile networks, the following describes several HO applications.

Clustering HO models use unsupervised learning to group a set of cells that their HO behaviors are similar. The popular and efficient ML algorithms for clustering models are K-means, relevance determination Automatic (ARD), Mixtures of Gaussians, etc.

HO forecasting aims to accurately forecast the future HO demand of a cell, an eNodeB, even for an area. It has crucial roles in improving QoE and QoS, HO management, load balancing, and congestion avoidance by setting relevant HO parameters. The HO forecasting models usually use real-time HO data collected from various counters, and the prevalent method for HO forecasting is Time Series models, which are based primarily on the historical patterns of HO to forecast the future one, and the suitable ML algorithms are dynamic models like neural networks (NN), Linear Dynamical Systems, and Gaussian Process (GP).

HO diagnosis and decision making models analyze the current condition of a UE or network parameters such as UE tracking to predict HO trend of the UE at an early stage, and then, the SON can provide timely controlling actions to ensure the accuracy of HO decision. The most suitable ML algorithms for decision making are dynamic ML algorithms like HHM, Kalman filter, dynamic Bayesian networks, reinforcement learning, and deep learning.

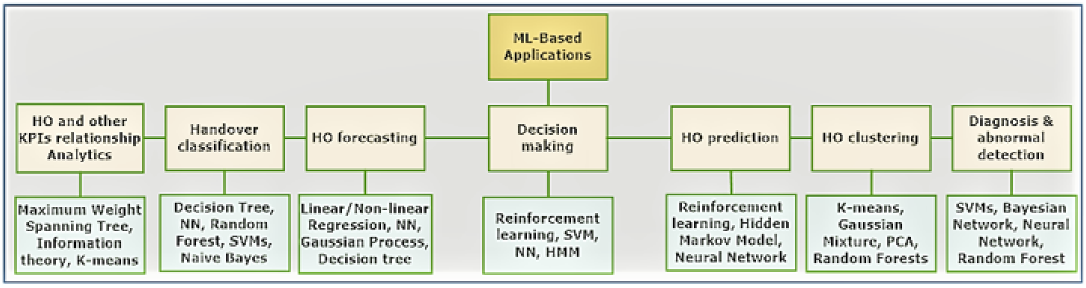


Figure 6. ML-Based applications for HO management

4 Analytics of the relationships of HO and other KPIs using ML

In 5G networks, each eNodeB is expected to have more than 2000 parameters that produce a massive amount of data or KPIs data [4]. Moreover, since network KPIs significantly affect each other, understanding the fundamental relationships between them has an important role in managing and analyzing the network behaviors for optimizing purposes like fault detection. This study proposes Maximum weight spanning tree (MWST), an unsupervised learning algorithm, to tackle this challenge. It analyzes and builds a structural network describing the relationships between KPIs, and then, provides a deeper awareness of KPIs characteristics. Moreover, based on the structure, we can extract the crucial KPIs that relate and impact to the HO KPIs. In general, the model analyzes a set of variables from the pure collected data without having any specific acknowledgment about the relationship between them. The goal is to find the tree that maximizes the data likelihood, in other words, find the tree with the greatest total weight. A weight is a mutual information between 2 nodes.

This study analyzes the hourly values of 9 typical network KPIs from the collected data to find out their relationship, especially, we also analyze the future number of HO attempts as a normal KPI in this model. This is important for HO forecasting applications in the next section. In addition, the Time is also considered as a KPI because it impacts directly to other KPIs such as the traffic of the cell.

MWST is a popular algorithm for structural learning in the Bayesian network. It is a constrained algorithm, which has one parent per node. As the result, the learning time is much faster than those of other algorithms. Fig.7 and Fig.8 show the structural networks representing relationships between the variables after analyzing the dataset for GSM cells and 3G&4G cells, respectively. It is noticeable that two structural networks describe exactly the hypothetical relationships of cellular network KPIs, and they also provide such useful and interesting information. For example, in the case of GSM KPIs, we can learn the procedure of making a call, it starts by requiring an SDCCH (standalone dedicated control channel) for the signaling and the call setup purpose, and then a TCH is assigned to UE (number of TCH attempts) for carrying user information (speech or data), which consumes the traffic of the serving cell.

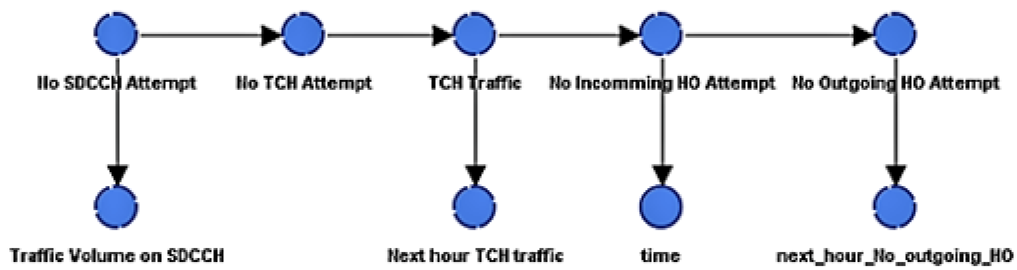


Figure 7. The relationship structure of GSM KPIs

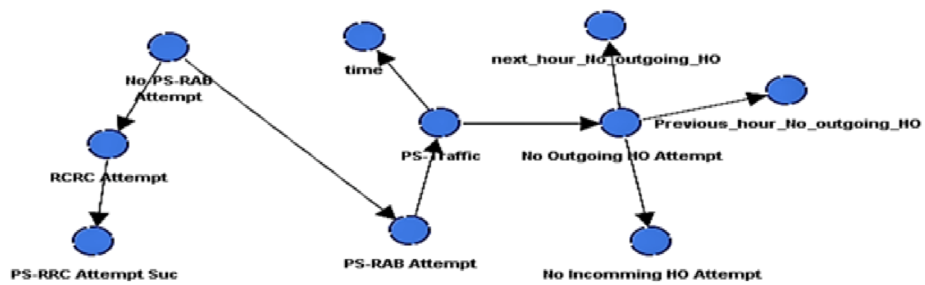


Figure 8. The relationship structure of 3G&4G KPIs

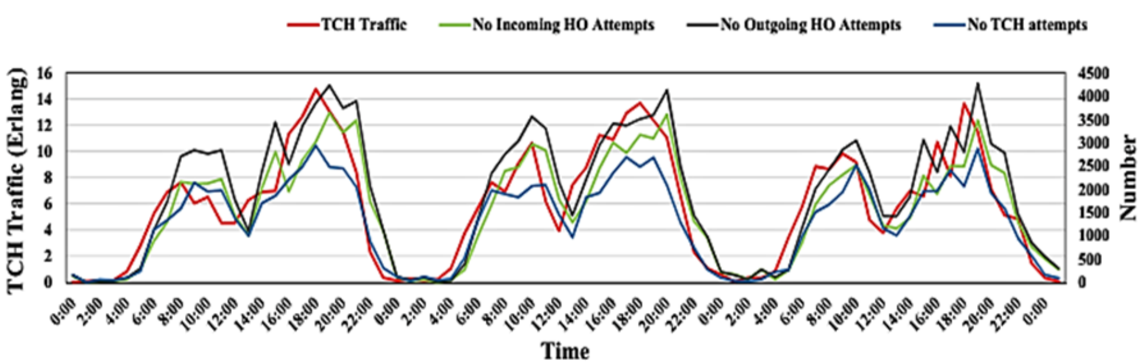


Figure 9. Hourly profiles of several KPIs

Meanwhile, the handovers usually happen during the call. To show the relationships clearer, Fig.9 shows the similar patterns of four KPIs (TCH traffic, number of incoming and out outgoing HO attempts, and number TCH attempts) during 3 days. On the other hand, the structural network also shows that the number of SDCCH attempts, the number of TCH attempts, and the number of incoming handover impact to the SDCCH traffic, the TCH traffic, and the number of outgoing HO, respectively. Finally, it is important to identify several factors that impact to the future number of HO attempts such as the TCH traffic, the time, and the number of incoming HO. The detailed analysis of the relationships will be discussed in the HO forecasting section. In brief, this application is a practical solution for future networks to analyze relationships of a huge number of KPIs and NPs.

5 Improving the Performance of Handover Clustering Model

HO patterns of cells in cellular networks are complicated, depending on time scales (e.g., hour, day), cell configuration, and the geographic location of the cell. Since understanding the HO behavior of cells is important for network management, in the research [1], the authors proposed a clustering model to put cells that have the most similar HO patterns into a cluster. Moreover, the model performance was investigated by comparing the HO behaviors among cells in a cluster and cells between clusters. Besides, we also extracted basic HO characteristics and the distributions of HOs under a day and a week of different clusters. However, the previous clustering model was built based on the HO features only without considering other KPIs, as the result, there were some cells in the same cluster that their HO patterns were similar but the probability of HO that occurs during a call of each cell might be different. To deal with this challenge, this study enhances the previous model by considering other features.

5.1 Features selection

The first main factor that impacts to the HO pattern of a cell is the probability that UEs move into or out of the cell when they are making calls at a specific time of a day. For example, cells that cover highways or high speech rails usually have a higher HO probability than those of cells that cover offices or buildings. Moreover, the HO probability of a cell is represented by the relationship between its traffic and number of HO attempts. As the result, if the traffic of these cells is similar, the number HO attempts of the cells in the first group (cover high-speed rails) must be significantly greater. On the other hand, the number of HO attempts also depend on the time such as during the day, during weekdays and weekends, monthly, even yearly [1]. For example, cells that cover school campuses usually have a higher number of HO in the daytime duration, while cells that cover night market areas usually have a higher HO trend in the night time duration. Hence, to achieve higher performance of clustering models, input features must accurately represent the two main factors affecting the HO behaviors of the cells.

Different from the previous study, the clustering model was built based on the HO features only, this study considers features extracted from different KPIs: the values of traffic KPIs and HO KPIs are used to capture

Table 2. Input features selection for the clustering model

	Time (Hour)	0:00	1:00	2:00	3:00	4:00	5:00	6:00	7:00	8:00	9:00	10:00	11:00
	Avarage HO values	391.13	267.80	194.93	169.53	190.60	269.80	647.80	1248.53	1837.20	2212.93	2074.20	2153.20
	Percentag e HO (%)	1.15	0.79	0.57	0.50	0.56	0.79	1.90	3.66	5.39	6.49	6.09	6.32
	Avarage values	179.57	102.70	3.70	12.02	11.78	51.97	41.31	65.73	111.49	124.45	105.03	155.06
	Percentag e (%)	5.63	3.22	0.12	0.38	0.37	1.63	1.30	2.06	3.50	3.90	3.29	4.86
	Time	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00
	Avarage values	1918.00	1703.40	1808.53	1882.33	2006.40	2131.20	2201.73	2287.13	2546.60	2053.73	1281.20	595.47
	Percentag e (%)	5.63	5.00	5.31	5.52	5.89	6.25	6.46	6.71	7.47	6.03	3.76	1.75
Cell 3QI0015	Avarage Traffic values	180.98	198.90	182.67	103.12	115.89	123.88	119.26	191.86	293.68	254.52	220.99	237.85
	Percentag e Traffic (%)	5.68	6.24	5.73	3.23	3.63	3.89	3.74	6.02	9.21	7.98	6.93	7.46
	Weekday	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Avarage HO values	Average hourly values	Avarage daily values	Avarage weekly values	
	Avarage HO values	33073.50	31112.00	32829.67	36496.00	33527.50	34630.00	37467.00					
	Percentag e HO (%)	13.87	13.04	13.76	15.30	14.06	14.52	15.71		1419.725	34073.4	238514	
	Avarage Traffic values	1675.80	5541.71	2177.91	3516.97	4095.82	2420.91	2725.33	Avarage Traffic values	1419.72	34073.40	238513.80	
	Percentag e Traffic (%)	7.51	9.76	15.76	24.83	18.35	10.85	12.21					

the HO probability of a cell, the percentage values of traffic and HO in hourly, daily and weekly distribution capture the changes of HO behaviors in the time. Table II shows an example of the input features for the clustering model (cell QI0015), the sample consists of 130 features. This section compares the performance of two clustering models. The first model is the clustering model of the previous study, which uses 65 features relating to the number of HO attempts. The second model or the new model uses all 130 features as shown in Table 2.

5.2 Experimental Set Up Performance Analytics

Experimental set up: This experiment uses K-means to cluster 2000 cells into 40 clusters, the number of clusters is high enough to guarantees that the number of clusters is manageable and the HO behaviors of cells in the same group are most similar without many clusters that cover a small number of cells. Since K-mean is a distance-based algorithm, all input attributes are normalized into a comparable range by using z-score normalization.

Performance analytics: Each clustering model divided 2000 cells into 40 groups, whose sizes were summarized as in Table 3. After identifying and assigning cells into their respective cluster, the number HO attempts of each cluster were determined by the average values of all cells belonging to it. Fig.10 and 11 show the hourly HO patterns of all 40 clusters extracted from each model, respectively, for 4 days. As can be seen, the HO pattern of each cluster was identified and distinguishable from others, the difference

Table 3. Handover clustering result summarization

Cluster	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Clustering result (first model)	27	51	19	130	13	33	52	78	29	83	138	7	66	17	6	66	40	17	79	28
Clustering result (second model)	36	57	84	56	53	56	12	20	103	74	72	97	55	37	50	74	46	58	61	62

Cluster	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
Clustering result (first model)	5	3	13	4	72	41	34	19	31	63	89	82	109	78	87	90	16	40	19	126
Clustering result (second model)	43	29	93	43	45	40	48	43	44	60	26	26	43	23	75	33	31	19	10	63

can be average values, patterned shapes, cluster sizes, etc. However, it is noticeable that with the same input dataset, the cluster patterns of the second model separate more smoothly and steadily from others than those of the first model. For example, in Fig.10, there are some clusters that their patterns change dramatically and unpredictably during different days.

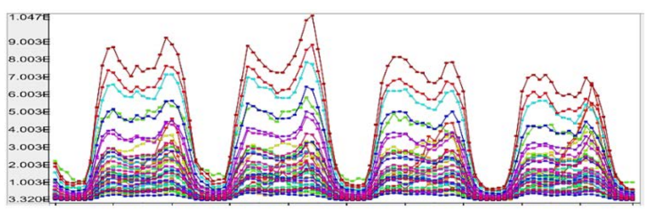


Figure 10. HO profiles of all Clusters extracted from the first model

That means the second model produced the clusters that their behaviors were more straightforward to identify, analyze, and manage.

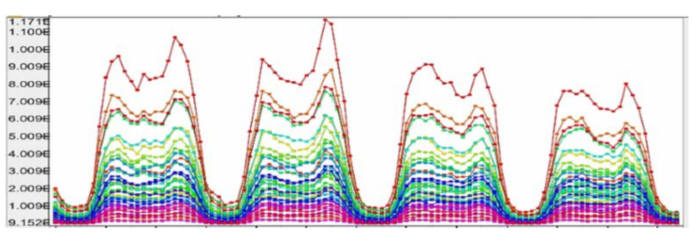


Figure 11. HO profiles of all Clusters extracted from the second model

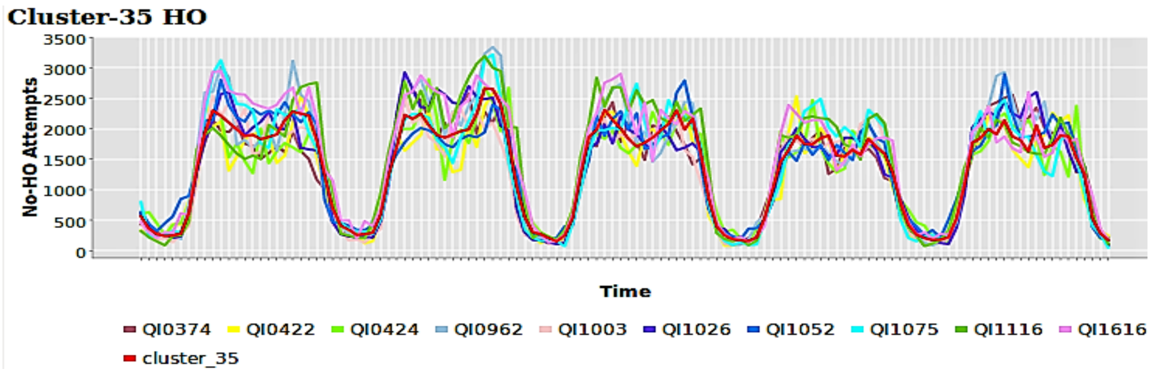


Figure 12. HO profiles of 10 cells in Cluster-35 of the first model

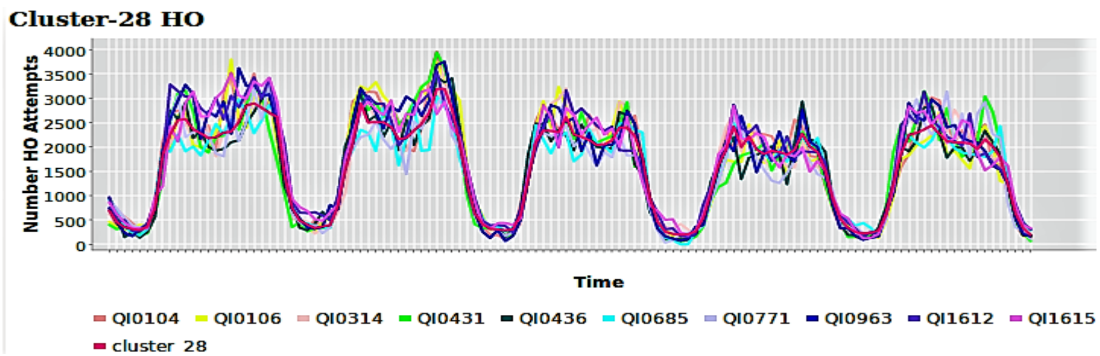


Figure 13. HO profiles of 10 cells in Cluster-28 of the second model

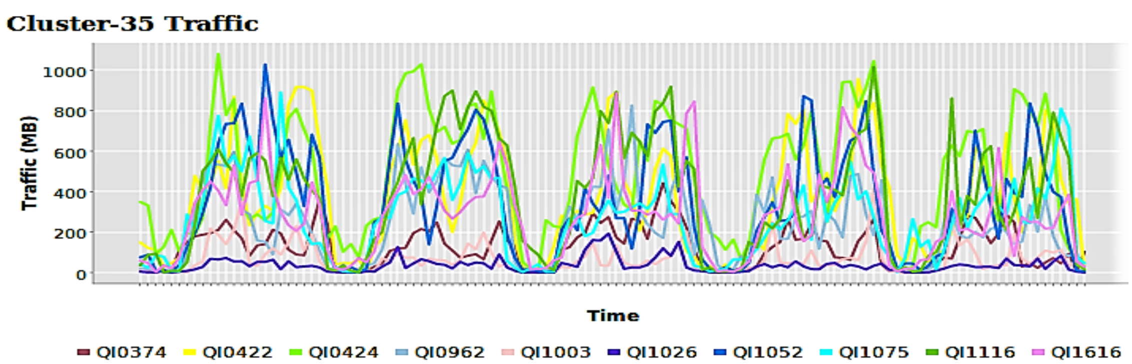


Figure 14. Traffic profiles of 10 cells in Cluster-35 of the first model

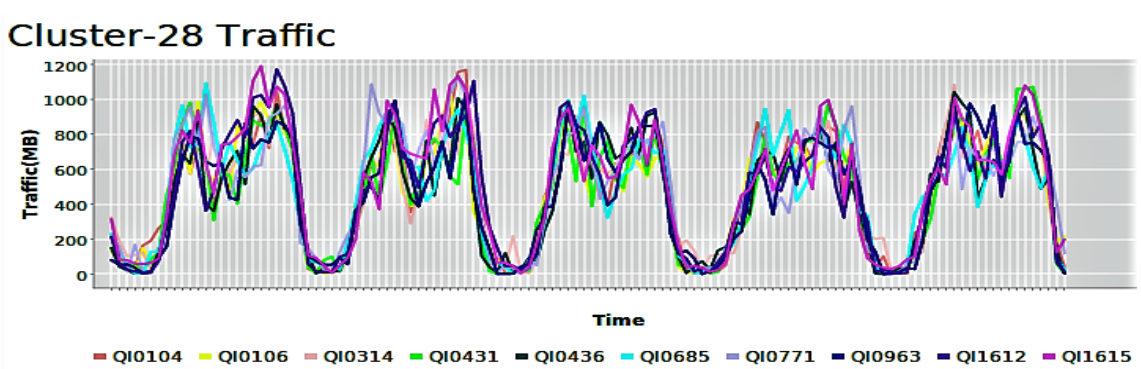
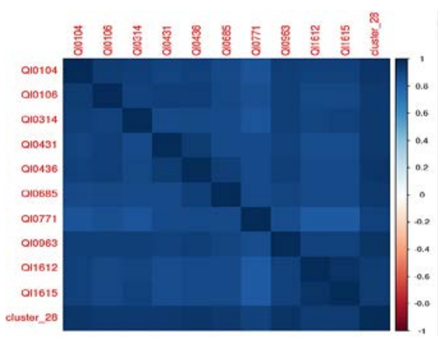


Figure 15. Traffic profiles of 10 cells in Cluster-28 of the second model

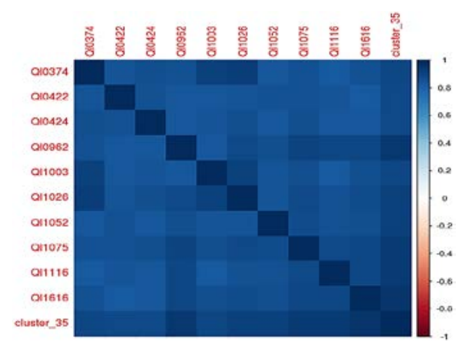
Next, to investigate the HO patterns of a cluster and its cells, each cluster of each model, Cluster-35 of the first model and Cluster-28 of the second model, were selected. Their HO patterns are quite similar and their sizes are large enough (90 and 44, respectively) so that they are able to represent for the common pattern of clusters in each model to make the analysis more generally. Fig.12 and 13 show the HO patterns of these two clusters and their cells for 5 days. Here, for the sake of clarity, only 10 cells of each cluster were randomly selected and displayed. It is visible that all of the HO patterns in the same cluster are quite similar, and they only fluctuate around the cluster patterns. As the result, the HO pattern of a cluster can be used to represent the HO pattern of its cells.

Similarly, the traffic patterns of the cells in Cluster-35 and Cluster-18 are shown in Fig.14 and Fig.15, respectively. It is obvious that traffic behaviors of the cells in Cluster-35 vary significantly and possess different characteristics in time-domain, while those of the cells in Cluster-28 are quite similar during all time. That means the first model clustered cells that have the same HO patterns into a cluster without considering their traffic behaviors. In contrast, the second model clusters cells that the same both HO behaviors and traffic behaviors into a cluster.

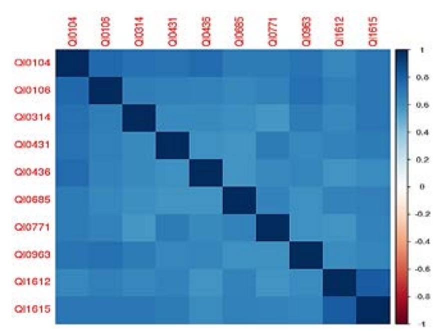
To make the comparison clearer, Figs. 16 (a), (b), (c), and (d) show the correlation matrices of the HO patterns, as well as the traffic patterns of the cells in the two clusters, respectively, represented by colors. As can be seen, the cells in each cluster have high correlations of the HO patterns with each other. In other



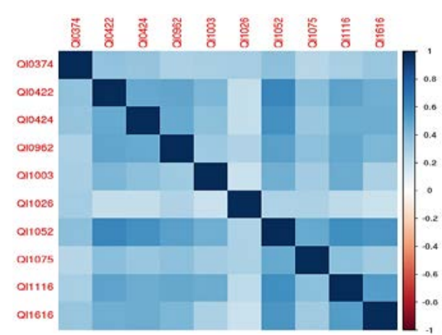
(a)



(b)



(c)



(d)

Figure 16. Correlation matrix of HO and Traffic of cells
(a) HO of 10 cells in Cluster-35; (b) HO of 10 cell in Cluster-28
(c) Traffic of 10 cells in cluster-35; (d) Traffic of 10 cells in cluster-28

words, their HO patterns are similar. However, compared with the cells in Cluster-35, those of Cluster-28 have much higher correlations of traffic patterns with each other.

In summary, the second model with the input features significantly improves the HO clustering model that can group cells into accurate clusters associated with their HO, traffic behaviors, and HO probability. This is crucial for networks operators to manage HO behaviors of a huge number of cells easier and more efficiently. For example, all cells in the same cluster covering areas with a high HO probability (e.g., along high-speech rails) should be assigned with the same suitable parameters (e.g., HO threshold) to ensure the continuity of services. Moreover, in the previous study, based on the clustering result, we extracted and classified clusters with different HO patterns in daily and weekly periods, such as weekday-higher pattern, the weekend-higher pattern, and the equally distributed pattern, which can be utilized for load balancing and energy saving applications [1].

6 Improving Performance of Handover Forecasting Models

Handover forecasting plays a crucial role in improving network quality and influences directly the mobility performance of networks. In the previous study, the authors proposed and implemented different forecasting models to forecast hourly HO demand in cells using various powerful ML algorithms. Especially, the model, which utilized HO clustering results to forecast HO patterns of any cell in a cluster by just using one forecasting model, significantly improved the performances of HO forecasting model. It was considered as a practical solution that can be applied in the industry due to the high accuracy and computing efficiency. Generally, Time-series approach, which bases primarily on the historic patterns of a parameter to predict its future values, is the most popular model for HO forecasting and prediction, while the utilization of other network parameters is limited.

In this study, the authors propose a practical model for HO forecasting based on network KPIs and ML algorithms. its process comprises two main steps: Step 1 analyzes the relationship between the HO attempts in the next hour and other KPIs in the past to determine the KPIs that have a significant impact to the future HO; Step 2 builds forecasting model based on the selected KPIs from the first step using ML algorithms. Furthermore, the performance of the proposed model will be evaluated by applying various popular ML algorithms.

6.1 Analysis of the relationship of future HO demand and network KPIs

The fundamental background for analyzing the relationship of the future HO demand and other KPIs is mainly based on information theory concepts, such as mutual information (MI) and relative mutual information (RMI). Moreover, the Bayesian network (BN) from BayesiaLab is used to analyze the probabilistic relationships among parameters and present their relationships as a directed acyclic graph (DAG) model. In this model, the next-hour HO attempt is the target node, and other KPIs such as current-hour HO attempt, current-hour TCH traffic, etc. are observation nodes. Here, Naïve Bayes, one of the most popular supervised learning algorithm of BN, is used to build the DAG.

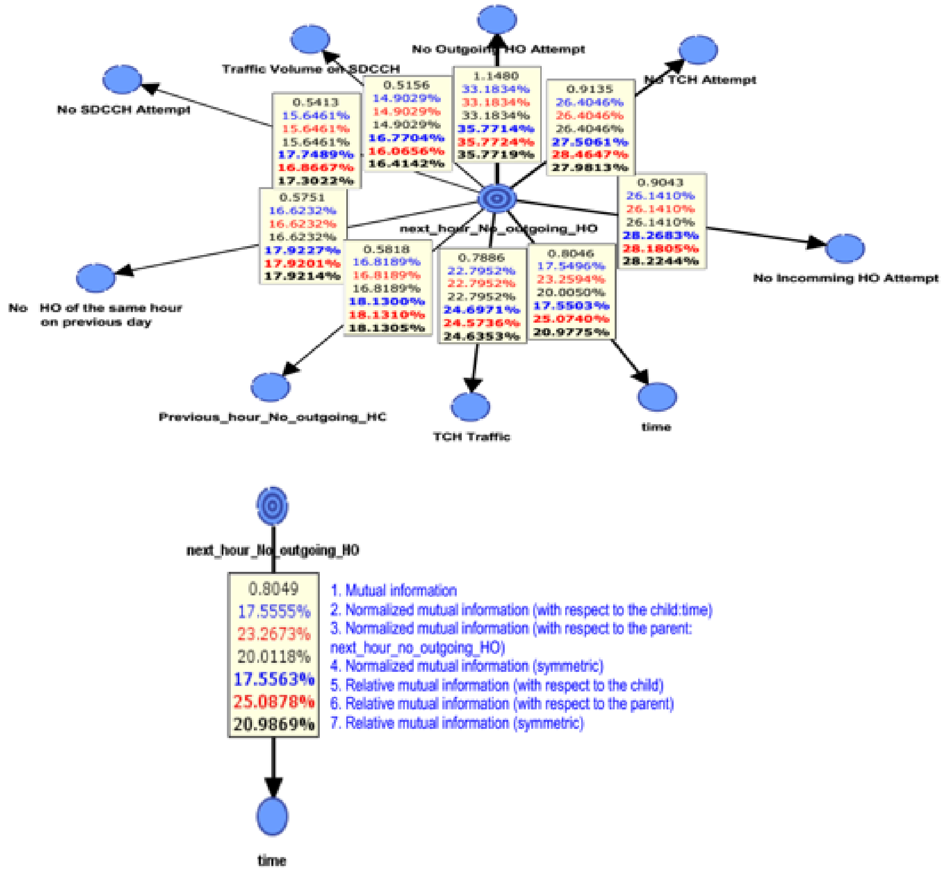


Figure 17. Graphical representation of the relationships between the next-hour HO attempts and the KPIs

Fig.17 presents the analytics results, in which the relationship between an observation node and the target node is represented by 7 information concepts defined in the right side of Fig.17. For example, the top number in the boxes is the Mutual Information, which is a measure of mutual dependence between two variables; it represents the reduction in entropy or the information it would gain, on average, about a variable if the knowledge of the other is given. Therefore, variables that have higher MI values are more significant in forecasting the target variable due to their close relationship. Similarly, the RMI is an important concept, which determines how much information that it would gain in percentage of a variable through observing another. The result shows that the number HO attempt of the current hour is the most important KPI for the HO forecasting model since it provides the highest information gain with 35.77% about the future HO. Other significant KPIs that have a primary impact on the future HO are the number of TCH attempts, the number of incoming HO, the Time, and the TCH traffic. It is noticeable that their MI and RMI values are higher than those of other historic HO KPIs, which are usually used as the main features of time series forecasting models, such as the number HO of the previous hour, the number HO of the same hour on the previous day.

6.2 HO forecasting model based on important KPIs

This section evaluates the performances of the proposed model called KPI model, which is based on relevant KPIs that have high MI and RMI values with future HO. Here, five KPIs are selected for the model: the number of HO outgoing attempts, the number of TCH traffic attempts, the number of incoming HO attempts, the Time, and the TCH traffic of the current hour. The hidden meanings behind these KPIs are that they present for the observation of the past HO pattern, the number of calls, and the mobility of Users. the performance of the KPI model is compared to the performance of the time series model, which introduced in the previous study [1]. The time series model is based mainly on the historic traces of HO to forecast future patterns. That means it uses x(t-1), x(t-2), …, x(t-N) to predict x(t), where x(t) is the future number of HO, x(t-1) is the current number of HO, and x(t-2),... , are the past number of HO demand. Here, 7 steps behind the past (N=7 (hour) will be used for hourly HO forecasting.

6.3 Machine learning algorithms

In this subsection, several popular forecasting algorithms are used to evaluate the performance of the proposed HO forecasting model. The first type of forecasting models is based on Auto-Regressive (AR) method, which is widely used for practical regression and time series model, both linear regression (LR) and polynomial regression (PR) algorithms are deployed. The second type is Neural Network (NN), which is a state-of-the-art non-linear algorithm providing a remarkable ability to derive meaning from complicated data. It is usually used to handle real data characteristics such as non-linearity to extract and detect patterns, forecast and predict trends that are too complex to be solved by other algorithms. The last ML algorithms investigated for the HO forecasting is Gaussian process (GP) algorithm, which is a kernel-based model that can provide a practical solution with a significant improvement in performance for both classification and regression applications.

6.4 Experimental setup and performance evaluation

Experimental setup:

Practically, building a forecasting model for a whole network can cause its performance degradation because HO behaviors of cells are diversity. However, building a model for each cell, it might be an overwhelming problem and inefficient. Fortunately, the clustering results in the previous section proved that the HO behaviors of all cells in the same cluster were similar to one another and similar to the cluster pattern. Therefore, building a forecasting model for all cells in a cluster can guarantee both accuracy and efficiency. For example, in our case, we only need to build 40 forecasting models for 40 clusters. Hence, the training and testing dataset for each experiment were selected as below: a training dataset of 200000 data samples and a testing dataset of 50000 data samples were selected randomly from the data of different cells in the same cluster, and samples of the testing dataset must be different from the training dataset.

Performance evaluation: The performance of each model and ML algorithm was evaluated through 4 performance metrics: the mean absolute error (E_{ma}), the root mean square error (E_{rms}), the coefficient of determination (R^2), and the correlation coefficient (Correl).

$$E_{ma} = \frac{1}{n} * \sum_{i=1}^{n} |p_i - r_i|$$

$$E_{rms} = sqrt(\frac{1}{n} * \sum_{i=1}^{n} (p_i - r_i)^2)$$

$$R^2 = 1 - \sum_{i=1}^{n} (p_i - r_i)^2 \,/\, \sum_{i=1}^{n} (r_i - \frac{1}{n} * \sum_{i=1}^{n} r_i)^2$$

$$Correl = R$$

Where r_i is the real value, p_i is the forecasted value, and n is the total number of test samples.

6.4.1 Experimental result

We examined the performance of HO forecasting models through HO patterns and the performance metrics. Firstly, Fig.18 shows the forecasted result for a cluster, cluster-25 (of the first clustering model), of the two models and the four ML algorithms for continuous 8 days. For the sake of clarity, in this figure, the forecasted HO traces of each ML algorithm and model appears only for two continuous days. It is obvious that the HO forecasted traces and the actual trace are quite similar and close to each other during all time. That means, all the ML algorithms and 2 models can accurately capture the future HO attempts of the cluster. Similarly, Fig.19 describes the experimental results when applying the same forecasting models for cell QI006 in Cluster-25. Furthermore, the statistical performances metrics of the models and ML algorithms were summarized in Table 4 to provide a detailed comparison of their performances. The first evaluation is the performance of ML algorithms when applying them to both models: LR and PR algorithms gave quite similar performances, PR performance was a little bit higher than LR performance even if the maximum polynomial degree of the PR algorithm was set to a high number; NN and GP gave better performance, and GP gave the best performance. The second evaluation is the performance of the Time Series model and the KPI model. As can be seen, when the same ML algorithm was applied, the KPI models significantly improved the forecasting performance such as higher R^2 , $Correl$, and smaller E_{ma}, E_{rms}.

In summary, the KPI model is a practical solution for handling complex patterns with different characteristics of HO behaviors. It can precisely capture the future HO demand of a cluster and any cell in this cluster with just using one model. This is crucial in industrial networks like HetNets and UDNs, which covers a huge number of cells.

Table 4. Hourly HO forecasting performances

Performance evaluation	Real values			
Model	R^2	Correl	E_{ma}	E_{rms}
LR + Time Series	0.92	0.959	137.00	218.00
LR + KPI	0.98	0.990	58.55	106.46
PR + Time Series	0.93	0.964	135.18	215.14
PR + KPI	0.98	0.990	57.40	102.42
NN + Time Series	0.953	0.976	84.90	148.06
NN + KPI	0.993	0.996	31.62	56.19
GP + Time Series	0.973	0.986	53.63	114.97
GP +KPI	0.995	0.997	25.25	50.94

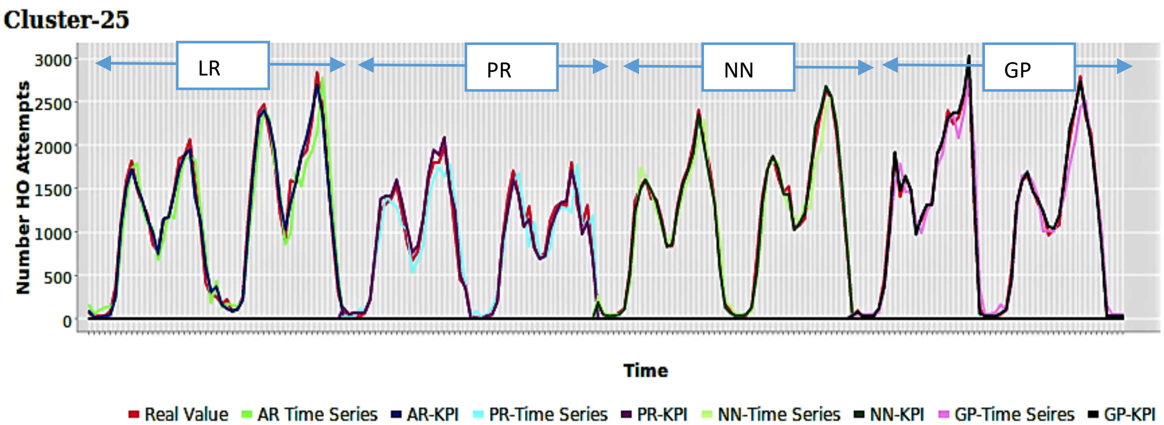


Figure 18. Forecasted hourly HO attempts results for Cluster-25

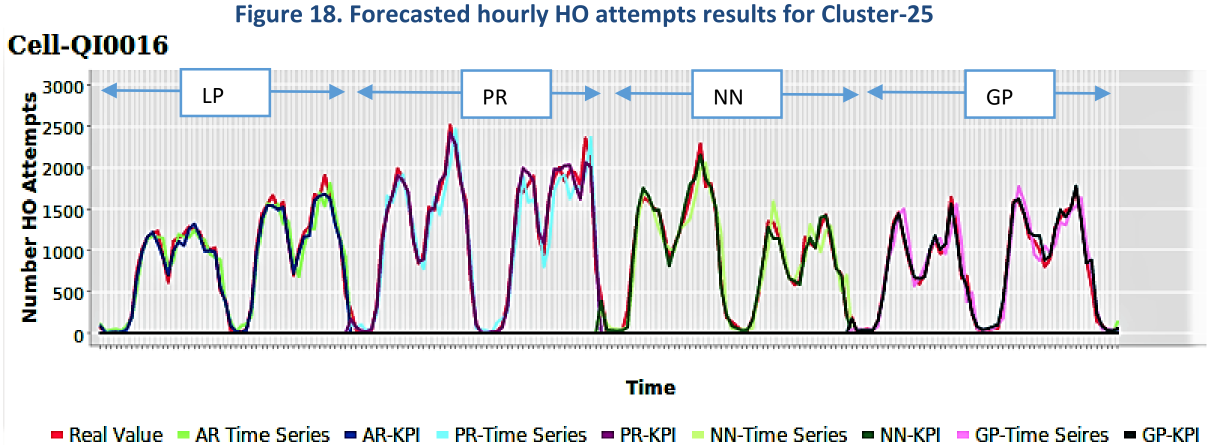


Figure 19. Forecasted hourly HO attempts results for Cell-QI0016 in Cluster-25

7 Abnormal HO detection based on HO Clustering and Forecasting Results

One of the most essential applications based on KPIs analysis, HO clustering, and forecasting is abnormal detection. Abnormal problems are unavoidable challenges of mobile networks while operating, such as equipment failures, frequency interference, and missing configuration of neighbor cells. Since most of the issues result in abnormal KPIs and HO behaviors, in the conventional network, network operators manually analyze a set of factors from a huge amount of collected KPIs dataset or through the driving test to detect abnormal in operating cells. This may take a long time and need a lot of efforts. Hence, it is in need of an urgent approach that is able to automatically and effectively detect abnormal problems.

As discussed before, the second clustering model puts cells with similar probability of HO attempts (the same traffic and HO patterns) together. However, the relationship of HO and traffic patterns is dramatically changed by abnormal issues that often produce abnormal KPIs. For example, Fig.20 shows the driving test result of an abnormal cell, cell QNI006. As can be seen, the test mobile equipment (UE) was consecutively handovered or ping-pong HO among cell QNI006 with its neighbor cells, as the result, there were many HO attempts, even dropped ongoing calls due to low C/I (the carrier to interference ratio) and HO failures. The root cause of the problem was the inter-cell interference of cell QNI006 with its new active neighbor. Thus, their Received Signal Levels (RxLev) were overlapped with each other, and

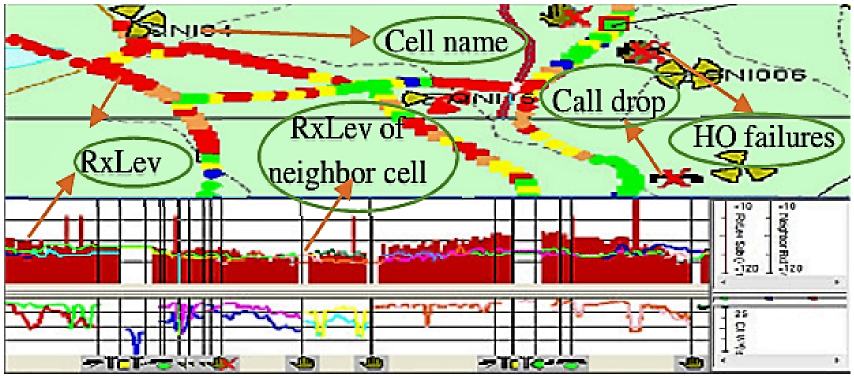


Figure 20. An example of abnormal Handover in a cell

the C/I was much lower than the regular value (C/I >12 dB). Consequently, this cell produced a high number of HO attempts, low traffic, high Call Drop Rate (CDR), and high HO-probability.

In the previous study, we proposed a useful approach called two alarm levels to detect abnormal

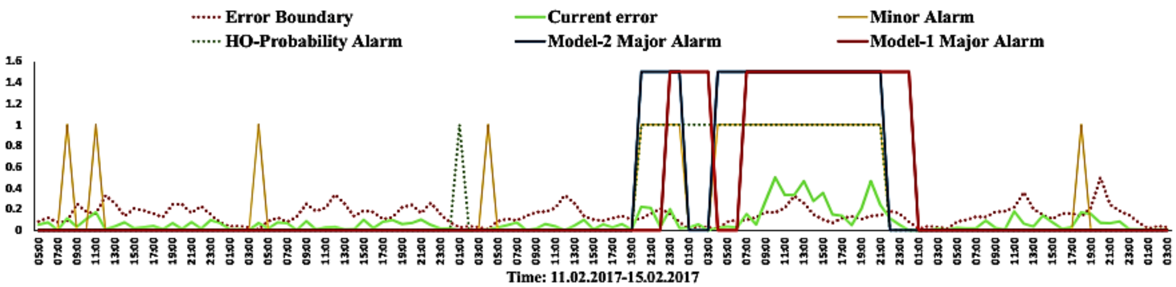


Figure 21. Abnormal detection time series of cell QNI006

situations based on HO forecasting. In this study, we propose a new practical abnormal detection based on HO forecasting and HO relationship analysis. Both models were built based on experimental results after extracting and analyzing the abnormal and failure situations through their KPIs of more than 6000 cells for 2 years. Each model involves three steps, and their first two steps are similar.

The first step: Define a parameter called *error boundary* for each hour based on calculating average errors and standard deviation errors between the forecasted HO and the real one using its moving window of 7 past values at the same hour (or during 7 days):

$$error\ boundary = average\ (error) + 3 * standard\ deviation(error)$$

(the factor 3 here is a practical factor, its value is usually in the range [2, 10]). (In real situations, we also use *error boundary* to define a parameter called HO boundary for HO congestion avoidance applications).

The second step (trigger minor alarm): calculate and compare the current error with the *error boundary*.

$IF\ current\ error\ > Boundary$, then trigger minor alarm (or $minor\ alarm = 1$).

The minor alarm notifies that there was an irregular HO behavior during the last hour. It could due to the movement of group users or an abnormal increase in traffic consumption. Since the minor alarm does not provide much meaning, it is necessary to have a further process to make the abnormal detection more reliable.

The third step of the first model (trigger model-1 Major Alarm): Both HO forecasting models are based on the past HO so that they will update any minor event causing increasing in predicted error. However, if the errors occur continuously, that means the HO behavior has been changed or there was a serious problem in the cell. To detect the major problem, the average 5 past values of minor alarm is calculated. If it exceeds 0.5 then trigger the major alarm (*or Model-1* $Major\ Alarm\ = 1.5$)

The third step of the second model (trigger model-2 Major Alarm): Similar to the *error boundary*, here we define a parameter called *HO-probability boundary* for each hour based on its moving window of 7 past values

HO-probability $= number\ of\ HO\ aptempts/Traffic$

HO-probability boundary $= average\ (value) + 3 * standard\ deviation(value)$

Next, we need to calculate and compare the current HO-probability with its boundary

$IF\ current$ HO-probability $>$HO-probability boundary, then trigger the HO-probability alarm (or HO-probability $alarm = 1$).

This value notifies that the regular relationship between HO and traffic was changed during the last hour. Finally, if both the HO-probability alarm and the minor alarm happen concurrently, then trigger the mode-2 major alarm.

If HO-probability alarm $* minor\ alarm = 1,$ then mode-2 $Major\ Alarm = 1.5.$

Fig. 21 illustrates the time series patterns of error boundary, current error, minor alarm, HO-Probability Alarm, Model-1 Major Alarm, and Model-2 Major Alarm of cell QNI006 since 11.02.2017 to 15.02.2017. In this figure, the values of the current error and error boundary were normalized in the range [0, 0.5] using the min-max approach. As can be seen, there was a critical issue in the cell at around 19 pm on 13.02.2017. The second model immediately triggered its major alarm while the first model required an extra delay. The problem is caused by inter-cell interference of cell QNI006 with its new active neighbor as described in Fig.20.

In summary, the second model based on the HO forecasting and HO probability is a practical solution for abnormal detection that can be applied UDNs due to the fact that it is effective and simple to implement. Moreover, it is possible to improve the model performance by using short-term forecasting models (e.g., minutes, seconds).

8 Conclusion

A comprehensive architecture for handover analytics, accurate HO clustering and forecasting, and effective HO management is the most critical challenge in UDNs and HetNets to improve mobility performances and make networks more coordinated and efficient. Firstly, establishing the relationship between HO KPIs and other KPIs is important to identify the key factors that affect HO patterns in cells, such as time series, the number of TCH attempts, and the traffic pattern. Secondly, the process proposed in this study can capture the dynamic behavior of HO and thus proposes a comprehensive approach for HO clustering, which is important for self-organization to manage HO patterns of a huge number of cells easily and efficiently. Finally, the new HO forecasting model provides an accurate HO forecasting, which is crucial for the SON to develop robust optimization applications, such as handover control, offloading traffic from the Macro-cells, congestion control, and abnormal detection. In this scenario, the SON become more robust, active and efficient in terms of managing, optimizing, and improving the network performance.

ACKNOWLEDGMENT

This paper is particularly supported by "the Center for Open Intelligent Connectivity from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project" by the Ministry of Education (MOE) in Taiwan, and the Ministry of Science and Technology of Taiwan under Grants: MOST 107-2221-E-009-056.

REFERENCES

[1] Le Luong Vy; Li-Ping Tung; Bao-Shuh Paul Lin, "Big data and machine learning driven handover management and forecasting," in *IEEE Standards for Communications and Networking (CSCN), Helsinki, 2017*, 2017, pp. 214–219.

[2] Luong-Vy Le; Bao-Shuh Lin; Do Sinh, "Applying Big Data, Machine Learning, and SDN/NFV for 5G Early-Stage Traffic Classification and Network QoS Control," *Trans. Networks Commun.*, vol. 6, no. 2, pp. 36–50, 2018.

[3] S. Lohmüller, L. C. Schmelz, and S. Hahn, "Adaptive SON management using KPI measurements," *Proc. NOMS 2016 - 2016 IEEE/IFIP Netw. Oper. Manag. Symp.*, no. Noms, pp. 625–631, 2016.

[4] S. Hahn, M. Schweins, and T. Kurner, "Impact of SON function combinations on the KPI behaviour in realistic mobile network scenarios," in *2018 IEEE Wireless Communications and Networking Conference Workshops, WCNCW 2018*, 2018, pp. 1–6.

[5] L. Le, D. Sinh, B. P. Lin, and L. Tung, "Applying Big Data, Machine Learning, and SDN/NFV to 5G Traffic Clustering, Forecasting, and Management," *IEEE NetSoft*, vol. 870, no. NetSoft, pp. 168–176, 2018.

[6] O. G. Aliu, A. Imran, M. A. Imran, and B. Evans, "A Survey of Self Organisation in Wireless Cellular Communication Networks," *IEEE Commun. Surv. Tutorials*, vol. 15, no. 1, pp. 336–361, 2013.

[7] P. V Klaine, M. A. Imran, O. Onireti, and R. D. Souza, "A Survey of Machine Learning Techniques Applied to Self-Organizing Cellular Networks," *IEEE Commun. Surv. Tutorials*, vol. 19, no. 4, pp. 2392–2431, 2017.

[8] D. Sinh, L. Le, B. P. Lin, and L. Tung, "SDN / NFV - A new approach of deploying network infrastructure for IoT," *27th Wirel. Opt. Commun. Conf. (WOCC), Hualien, Taiwan, 2018*, pp. 1–5, 2018.

[9] D. Sinh, L. Le, L. Tung, and B. P. Lin, "The Challenges of Applying SDN / NFV for 5G & IoT," *14th IEEE - VTS Asia Pacific Wirel. Commun. Symp. (APWCS), Incheon, Korea, Sep 2017*.

[10] B. P. Lin, F. J. Lin, and L. Tung, "The Roles of 5G Mobile Broadband in the Development of IoT, Big Data, Cloud and SDN," *Commun. Netw.*, vol. 8, no. 1, pp. 9–21, 2016.

[11] A. Imran and A. Zoha, "Challenges in 5G: How to empower SON with big data for enabling 5G," *IEEE Netw.*, vol. 28, no. 6, pp. 27–33, 2014.

[12] J. Prados-Garzon, O. Adamuz-Hinojosa, P. Ameigeiras, J. J. Ramos-Munoz, P. Andres-Maldonado, and J. M. Lopez-Soler, "Handover implementation in a 5G SDN-based mobile network architecture," *2016 IEEE 27th Annu. Int. Symp. Pers. Indoor, Mob. Radio Commun.*, pp. 1–6, 2016.

[13] T. Nguyen-Duc and E. Kamioka, "An extended SDN controller for handover in heterogeneous wireless network," *2015 21st Asia-Pacific Conf. Commun. APCC 2015*, pp. 332–337, 2016.

[14] E. B. Hamza and S. Kimura, "A Scalable SDN-EPC Architecture Based on OpenFlow-Enabled Switches to Support Inter-domain Handover," *2016 10th Int. Conf. Innov. Mob. Internet Serv. Ubiquitous Comput.*, pp. 272–277, 2016.

[15] C. Chen, Y. T. Lin, L. H. Yen, M. C. Chan, and C. C. Tseng, "Mobility management for low-latency handover in SDN-based enterprise networks," *IEEE Wirel. Commun. Netw. Conf. WCNC*, vol. 2016–Septe, no. Wcnc, 2016.

[16] I. Elgendi, K. S. Munasinghe, and A. Jamalipour, "Mobility management in three-Tier SDN architecture for DenseNets," *IEEE Wirel. Commun. Netw. Conf. WCNC*, vol. 30 October, no. Wcnc, pp. 214–219, 2017.

[17] P. V. Klaine, M. A. Imran, O. Onireti, and R. D. Souza, "A Survey of Machine Learning Techniques Applied to Self Organizing Cellular Networks," *IEEE Commun. Surv. Tutorials*, vol. 19, no. 4, pp. 2392–2431, 2017.

[18] M. Boujelben, S. Ben Rejeb, and S. Tabbane, "SON Handover Algorithm for Green LTE-A/5G HetNets," *Wirel. Pers. Commun.*, vol. 95, no. 4, pp. 4561–4577, 2017.

[19] J. Rizkallah and N. Akkari, "SDN-based vertical handover decision scheme for 5G networks," in *2018 IEEE Middle East and North Africa Communications Conference, MENACOMM 2018*, 2018, pp. 1–6.

[20] L. Le, D. Sinh, L. Tung, and B. P. Lin, "A practical model for traffic forecasting based on big data, machine-learning, and network KPIs," in *2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, 2018, pp. 1–4.

[21] B. P. Lin, L. Tung, F. Tseng, I. Hsieh, Y. Wang, and S. Chou, "Performance Estimation of MAR for Outdoor Navigation Applications based on 5G Mobile Broadband by using Smart Mobile Devices," in *Conference: IEEE VTS APWCS 2015, Singapore*.

[22] B. S. Lin *et al.*, "The design of big data analytics for testing & measurement and traffic flow on an experimental 4G/LTE network," *2015 24th Wirel. Opt. Commun. Conf. WOCC 2015*, pp. 40–44, 2015.

[23] H. C. Chang *et al.*, "Empirical Experience and Experimental Evaluation of Open5GCore over Hypervisor and Container," *Wirel. Commun. Mob. Comput.*, vol. 2018, no. i, 2018.