

Content-aware resource allocation model for IPTV delivery networks

Suliman M. Fati¹, Putra Sumari², Wou Onn Choo³

¹ College of Computers and Information Sciences, Prince Sultan University, Saudi Arabia

² School of Computer Sciences, Universiti Sains Malaysia, Malaysia

³ Faculty of Information Technology and Science, INTI International University, Malaysia

Article Info

Article history:

Received Apr 9, 2018

Revised Jul 23, 2018

Accepted Aug 11, 2018

Keywords:

Content status

Content-aware

Hybrid genetic algorithm

IPTV delivery networks

Resource allocation

ABSTRACT

Nowadays, with the evolution of digital video broadcasting, as well as, the advent of high speed broadband networks, a new era of TV services has emerged known as IPTV. IPTV is a system that employs the high speed broadband networks to deliver TV services to the subscribers. From the service provider viewpoint, the challenge in IPTV systems is how to build delivery networks that exploits the resources efficiently and reduces the service cost, as well. However, designing such delivery networks affected by many factors including choosing the suitable network architecture, load balancing, resources waste, and cost reduction. Furthermore, IPTV contents characteristics, particularly; size, popularity, and interactivity play an important role in balancing the load and avoiding the resources waste for delivery networks. In this paper, we investigate the problem of resource allocation for IPTV delivery networks over the recent architecture, peer-service area architecture. The Genetic Algorithm as an optimization tool has been used to find the optimal provisioning parameters including storage, bandwidth, and CPU consumption. The experiments have been conducted on two data sets with different popularity distributions. The experiments have been conducted on two popularity distributions. The experimental results showed the impact of content status on the resource allocation process.

*Copyright © 2019 Institute of Advanced Engineering and Science.
All rights reserved.*

Corresponding Author:

Suliman M. Fati,

College of Computers and Information Sciences,

Prince Sultan University,

Riyadh 12435, Saudi Arabia.

Email: smfati@yahoo.com

1. INTRODUCTION

IPTV, as a promising technology, is growing rapidly in terms of subscribers and revenue to become in the near future the standard means to deliver home and business entertainment contents [1]. Thus, Telecommunication companies have entered a hectic competition to increase their customer base and profit by delivering the IPTV services [1], [2]. On the other hand, the key concern of service providers in this hectic competition is how to provide high quality service with a minimum cost.

Consequently, the cost reduction is one of the service providers' worries. Such cost reduction can be achieved by controlling the resources allocation process [3]. Resource Allocation is an important issue in designing the shared delivery network. Resource allocation determines the locations and the amount of resources to minimize the cost under certain constraints. In the shared delivery networks, many service providers may share resources and compete to deliver good quality services. For that, these resources should be shared in a more effective manner [4]. However, IPTV systems suffer from the sudden peak workloads. Thus, allocating a large amount of resources to cope with the sudden workload leads to low resources utilization at non-peak hours. On the other hand, non-considering the peak workload leads to user

dissatisfaction during the peak hours. For that, pay-as-you-go scheme is suggested by some delivery network providers. Amazon CloudFront is a global example of this scheme. In this scheme, the service provider has to pay for the used resources only without any upfront, commit, or service contract.

Particularly, the fluctuation in the workload (i.e. the number of incoming requests targeting the hosted contents) is a serious issue threatens IPTV systems. Such workload fluctuation issue figures out the content status in IPTV system. Gaber and others [5], [6], [7] suggested a contents' workload estimation model that estimates the contents' status according to their characteristics. Of these characteristics are huge size (i.e. in Mega Bytes or Giga Bytes), interactivity, and rapidly changing effective period [8]. Such distinct characteristics make the design of IPTV delivery networks challengeable. On the other hand, ignoring the status of diverse contents leads to load imbalance and waste of resources [9]. Therefore, estimating the content's status according to these characteristics is a crucial factor in IPTV systems. Content status modeling refers as estimating the portion of concurrent requests that target the content according to its characteristics. Such modeling helps a lot in handling both the load balancing and resource allocation based on the anticipated load of contents [5].

By means of such contents' status estimation model, the service provider can control the resources allocation according to the content status and the users' demand to reduce the hosting cost [10] without affecting the users' satisfaction. This integration allows the service provider to adjust the required resources according to the necessitated replication scheme. For that, Laoutaris et al. [11] argue that replica placement problem and resource allocation problem should be solved jointly due to the dependency between them. Moreover, considering the contents' status during the integration gives the service providers the possibility to control the resources and content distribution according to the demand fluctuation, as well as, enables the delivery network to absorb the sudden load fluctuation.

Therefore, this paper concerns about how to build content-aware resource allocation model that considers the content status to reduce the waste of resources. In this paper, Content-Aware Resource Allocation Model (CARAM) is proposed. CARAM model considers the contents' status including contents' load and contents' popularity distribution to decide the required resources. CARAM model allocates the resources according to the replication strategy. Thus, it can save the resources in terms of storage, bandwidth, and CPU. For this purpose, the resource allocation problem and the replica placement problem are integrated to allocate the resources depending on the necessitated replication scheme. The content-aware resource allocation problem is studied and formulated mathematically as a constrained binary optimization problem. Next, this problem is solved using Hybrid Genetic Algorithm aiming to evaluate the effect of content status on the content-aware resource allocation. For this purpose, two popularity distributions are considered through this paper. These popularity distributions are Zipf's distribution and the exponential distributions, which follow the normal density function [12].

The rest of this paper is organized as follows. Section 2 is dedicated for background and related works. Section 3 presents the system model and problem formulation. Section 4 introduces the proposed hybrid Genetic algorithm to solve this problem. Section 5 is dedicated for experimental results and discussion. Finally, Section 6 presents the conclusion and future work.

2. BACKGROUND AND RELATED WORKS

Although the high benefits of IPTV delivery networks to both service provider and Content consumer, designing efficient delivery networks is not a trivial task and has potential problems, which must be solved. IPTV delivery networks composition, Content distribution and management, resource allocation, and request redirection are the main classes of issues that must be considered during building the delivery networks. IPTV delivery networks composition includes delivery network placement, determining the servers' specifications, and interaction among servers. Content distribution and management includes content selection and distribution based on demand and cache/replica management. Resource allocation chooses the type and size of resources that are required for each surrogate server. Request redirection includes the techniques to distribute the requests among servers in a manner to maintain the load balanced and avoid the traffic congestion [4], [13].

Content placement is an important point in the delivery networks for multimedia streaming. In large multimedia systems (e.g. IPTV), the effective content placement is highly required to cope with load imbalance, insufficient bandwidth, large delay, and packet loss [14]-[16]. In the context of content placement, a considerable number of works are proposed to cope with the growing demand on these contents. On the other hand, Replica placement Problem, as a global strategy, is widely investigated in the literature review to improve the overall performance and/or minimize the allocation cost [17]. In the literature, the Replica Placement problem is formally expressed as a problem definition (cost function), which moves

toward an overall goal (performance or cost improvement). Such problems typically require heuristic algorithms to find approximate solutions within a feasible time, due to that they are NP-complete [18],[19].

In the context of problem definition, the researchers investigate the content replication problem from two perspectives. The first perspective is to improve the quality of service [20]-[23]. The second perspective is to minimize the delivery cost [11], [12], and [24]-[28]. These works can also be classified according to the solution proposed. These solutions can be classified into heuristic solutions and evolutionary solutions. The heuristic solutions include the approximation algorithms and the greedy algorithms. In contrast, the evolutionary algorithms include Genetic, ant colony, and particle swarm algorithms.

Because of the Replica Placement Problem is NP-Complete [29]-[31], they typically require meta-heuristic algorithms to find approximate solutions within a feasible time [18]. Employing such meta-heuristic algorithms help in gaining the maximum benefit of replication that is crucial for the success of the IPTV systems. However, the aforementioned works, which exploit heuristic algorithms, fail to produce an optimal effective placement strategy [30], [31]. Thus, due to the fruitful application of evolutionary algorithms to solve a variety of optimization problems, many attempts are recorded in the direction of using these algorithms in solving replica placement problem. Khan and Ahmad [32] compared the heuristic algorithms with evolutionary algorithms (Genetic Algorithm as an example). They introduced a unified cost model that captures the minimization of the total object transfer cost in the system, which in turn, leads to effective utilization of storage space, replica consistency, and fault-tolerance. The study shows that Genetic Algorithm outperforms the heuristic algorithms (e.g. greedy algorithms) in terms of solution quality. Although, Evolutionary algorithms produce better solutions as close as possible to the optimal solutions, the heuristic algorithms are faster than evolutionary algorithms in the running time. These findings also confirmed by the authors in [31], [33].

Resource allocation is an important process in designing IPTV delivery networks. This process consists of determining the location, the type, and the amount of networking resources to deploy. The aim of the resource allocation problem is to minimize the cost while certain constraints are respected [19]. Thus, the optimization strategies are a good option to solve such problems [34]. In the literature review, the resource allocation problem has been addressed using heuristic solutions [35]. Wauters et al. [21] address the resource allocation problem of determining the equipment required for building transport network for VoD services. They focused on determining the number of ports of each server, as well as, the number of multiplexers and switch ports at each node. In their study, the decentralized network is divided into regional sub networks with a ring topology. Next, a single server is installed on each regional sub network. Similarly, Kitjongthawonkul and Ko [36] examined the resource allocation problems in a VOD network to find the optimal locations of the video servers. They considered the users' demand to select the location of those servers. Moreover, the authors addressed the allocation problem using the dynamic programming technique. The objective of Kitjongthawonkul and Ko is to minimize the total operating cost, including transmission cost, storage cost, and server installation cost.

Allocating the resources to build an overlay network over multiple substrate networks is proposed by Houidi et al. [37] and Pandey et al. [38]. Houidi proposed a dynamic overlay network creation model according to the load and the available resources. The proposed model composed of request splitting algorithm and embedding algorithm. The request splitting algorithm aims at helping the overlay network provider to distribute the incoming requests among the available substrate network providers. The authors proposed two requests splitting algorithms; max-flow min-cut algorithm and linear programming algorithm. Based on the splitting algorithm, embedding algorithm is proposed to assign virtual nodes and links into each substrate network provider simultaneously. In this work, the authors formulated the overlay network embedding process to be solved by means of mixed integer program. The aim of the proposed embedding algorithm is to increase the request acceptance rate and decrease the leasing cost of infrastructure.

Pandey et al. [38] focused on determining the optimum network deployment strategy for VoD services in a heterogeneous networking environment. The deployment decisions can be made considering the following parameters: the required number of VoD servers, the minimum physical distance of server from the community and the bandwidth capacity requirement. The authors' aim is to determine the minimum capacity requirements that maintain the minimum requirement of Quality of Experience (QoE) to be met. For that, they estimated three QoE metrics: the server waiting time, the minimum one-way delay, and the access network bandwidth consumption. Such estimation is done based on the analysis of user requirements and network configurations. Afterwards, they formulated the server location problem and path selection problem jointly as an Integer Linear Programming (ILP) problem. The authors solved this problem by means of ILP techniques, as well as, heuristics.

One of the flaws in most proposed works to solve the resource allocation and/or replica placement problem is that they consider the problems of resource allocation or replica placement problem independently. Solving these two problems independently produces suboptimal solutions; this is due to the

direct effect of replicas locations on the amount of required resources and vice versa [11], [39]. The main advantage of combining replica placement and resource allocation in one joint problem is that the resources will be utilized efficiently and as necessary. Once these resources become no longer required, they will be released to be available for other users who require it or able to pay for it. For that, integrating the replica placement problem with the resource allocation is suggested firstly by [11]. Laoutaris et al. [11] argue that the replica placement problem and the resource allocation problem should be solved jointly to avoid a suboptimal solution. They proposed a realistic framework to build optimal resource/ content allocation model for a hierarchical architecture that minimizes the total cost. The cost here is represented by the distance between the storage location and the end user. The framework of Lautaris took into consideration the load balance and request peering among servers as constraints. In this framework, Genetic Algorithm is exploited to solve the resource allocation problem. Nguyen et al. [40] addressed the problem of overly network provisioning for multimedia contents to build a cost-effective virtual delivery network. Their framework aims to build an overly delivery network over a substrate network. They formulated the provisioning problem as a combination between content replication and resource allocation. Next, they solved this provisioning problem using a Lagrangian relaxation heuristic algorithm.

Similarly, Nakaniwa et al. [41] introduced a mathematical model to join the content replication with server placement for the hierarchical architecture. Their aim was to maximize the system reliability of delivery network. They introduced an integer programming technique to maximize the reliability of the system. This technique finds the optimal scheme for both content and resource allocation problems that maximize the reliability of all the paths from each user to each file location. They applied their model on a real delivery network for internet service provider in Japan called "B Bit-Japan" with subject to delay and storage and/or transmission cost constraints.

On the other side, Aldana Diaz and Huh [42] argue that allocating the IPTV resources including storage and bandwidth requires creating a cost function based on contents management and distribution. They proposed three cost-based policies, particularly, local hosting, external hosting, and partial hosting policies for 3rd party content service on a distributed architecture. In their detailed cost function, they considered the storage and bandwidth requirements that maintain the least level of Service Level Agreement (SLA). The results of them showed that partial hosting of contents is the most suitable technique wherein some of the contents are hosted locally while the others can be requested upon the demand. Such partial hosting may maintain the trade-off between the storage cost and bandwidth cost. However, this work lacks for real constraints on which contents should be hosted, and, which contents should be requested upon demand. Li and Wu [12] proposed a heuristic algorithm that finds the optimal number of popular contents and the optimal number of servers to store these contents. Their model ignores the contents' replication and the load balancing.

Replicating the contents according to their status helps the service provider to decide on the required resources to allocate. Thus, considering the content status is an important point to achieve optimal resource allocation scheme depending on the replica placement. Therefore, in this paper, the content-aware resource allocation model is proposed aiming at investigating the impact of content status on resource allocation.

3. CARAM FRAMEWOK

In traditional delivery networks, as CDN, the service provider has to allocate the contents according to long-term service contract with a certain commitment. This pricing scheme hinders the service provider to cope with the sudden workload at the peak busy times. On the other side, this scheme leads to waste the resources due to the low utilization at the non- peak busy times. For that, pay-as-you-go scheme is suggested by the delivery network providers. Amazon CloudFront is a global example for pay-as-you-go scheme; in which, the service provider has to pay for the used resources only without any upfront, commit, or service contract. Particularly, the service provider has to control the resources allocation according to the demand to minimize the hosting cost [10] without affecting the user satisfaction.

Therefore, the aim of Content-Aware Resource Allocation Model (CARAM) is to decide on the resources amount and the content replication scheme according to the demand so as to achieve a minimum cost subject to given constraints. The cost, here, is expressed in terms of the amount of payment to the network provider to host the contents. Thus, allocating the storage and bandwidth resources can be performed in conjunction with replica placement. The main advantage of combining replica placement and resource allocation in one joint problem is that the resources will be utilized efficiently and as necessary [10]. Once these resources become no longer required, they will be released to be available for other users who require it or able to pay for it.

Content-Aware Resource Allocation problem focuses on determining the server locations in each service area, along with, the processing power, storage space, and bandwidth capacity at each location.

To achieve an optimal solution for this problem, the service provider has to obtain the size and popularity distributions of contents in each service area. Consequently, the replication degree, the popularity distribution, and the expected load of contents along with their requirement in terms of processing power, storage, and bandwidth must be estimated clearly. Replicating the contents according to their requirement helps the service provider to decide the required resources to allocate. Thus, this interdependency emphasizes the need to address resources allocation and replica placement jointly.

Content-aware Resource Allocation as an optimization process should take into account the available resources and their cost, as well as, the contents and their characteristics as follows.

- a. Available resources: In the context of shared hosting environment, each resource provider advertises “for hire” resources with information of resource type, location, and cost. The processing power, storage space, and bandwidth capacity of servers are also provided. In such environment, the service provider can choose different resources, which belong to different resources providers.
- b. Contents requirement: each video in IPTV systems requires a storage space, processing power, and bandwidth capacity proportional to its popularity. Thus, replicating the contents according to their popularity distribution gives the optimal replication scheme, which in turn optimizes the resource allocation process.

The information about the available resources, as well as, the requirements is gathered from the resource providers and service providers respectively. The service provider is responsible for doing the allocation process based on the requirements. Upon completion the allocation process, the service provider sends the process results to the resource provider to create the topology. In case of the existence of resource broker, the resource broker can complete the whole process according to the given information from both parties. The flow of Content-Aware Resource Allocation Model (CARAM) is depicted in Figure 1.

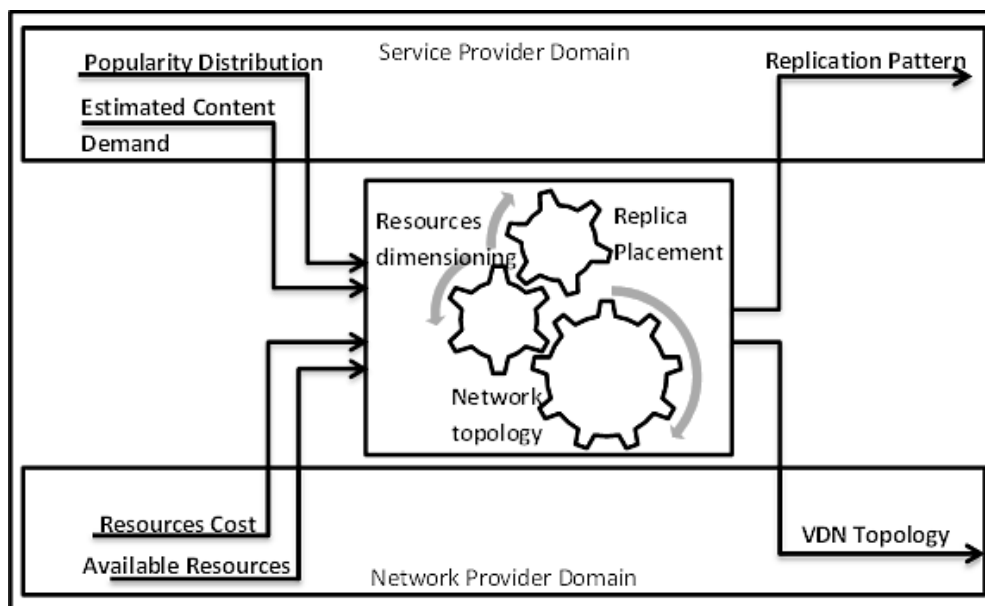


Figure 1. CARAM framework

As depicted in Figure 1, in CARAM model, the information on both the content popularity distribution and the demand of each video are gathered from the service provider. This information on the resources from each physical network is provided by the resources providers. Then, the desired optimization tool is employed to compute the optimal topology including the replication pattern. Lastly, the resultant topology is passed to the resources providers or the resource broker to create the delivery network.

4. CONTENT-AWARE RESOURCE ALLOCATION PROBLEM FORMULATION

Content-aware Resource Allocation problem over the peer-service area architecture can be modeled as a set of inter-connected service areas. Each service area contains a set of servers, which can be physical servers or virtual servers. Each server is expressed by its processing power, storage space, and bandwidth capacity.

This paper considers a delivery network comprising of A interconnected service areas with PL_a potential locations in each service area. Each potential location is with known storage capacity m_j^{max} , processing power u_j^{max} , bandwidth b_j^{max} , and Load capacity L_j^{max} . Moreover, each potential location advertises the cost of storage unit S_j , bandwidth unit β_j , and CPU consumption unit δ_j . Let there be a set of Video contents N . The size of each object is denoted by m_i and measured in Megabytes (MB). Each object i is associated with its processing power u_i , bandwidth b_i , expected load L_i , and popularity value $p_i \in [0,1]$. Each object has a possibility to be replicated among the service area or requested from another service area based on its popularity and the trade-off between the hosting cost and redirection cost. In the cost estimation stage of Content-Aware Resource Allocation Model (CARAM), it is necessary to distinguish between two terms server initiation cost (C_L) and replication pattern cost (C_R).

4.1. Server initiation cost (C_L)

In CARAM, the service provider has the ability to lease the resources from different physical network providers. Each physical network provider has own pricing plan for leasing the resources. The pricing plan includes the cost of configuring the server at the first time, deploying software, and maintaining the servers including the backup and security [40]. For instance, in the virtual networks, the leasing cost CL for the virtual network is the cost of initiating and deploying the candidate locations, which are selected as a virtual server in all service areas, as in (1). The cost C_j represents the leasing cost of a server j in the service area a .

$$C_L = \sum_{a=1}^A \sum_{j=1}^{PL_a} y_{aj} * C_j \quad (1)$$

Furthermore, y_{aj} is a binary indicator to indicate whether the potential location is selected $y_{ja} = 1$ or not $y_{ja} = 0$. To facilitate the communication among the service areas, the service areas are connected via inter-service areas links. These links enable the dispatcher in any service area to redirect the requests targeting the unavailable contents to other service areas. The cost of these links can be obtained from the redirection process in the replication pattern as will be explained in the next sub-section.

4.2. Replication pattern cost (C_R)

The contents replication represents a challenge in CARAM. In this step, the content replication cost must be envisaged. The replica placement problem should be extended to include the cost of other resources. The resources cost includes the cost of storage, processing power, and bandwidth for hosting the popular contents. Besides, the cost of contents redirection is included. Redirection cost determines the cost of requesting unpopular content from other service areas in case of not replicating it due to the resources limitations. The following equation explains the replication pattern cost based on the above illustration.

$$C_R = \sum_{a \in A} \sum_{j \in PL_a} \sum_{i \in N} x_{ija} (m_i S_j + u_i \delta_j + b_i \beta_j) + \sum_{a \in A} \sum_{j \in PL_a} \sum_{i \in N} p_i (1 - x_{ija}) \frac{m_i}{CBR_i} \beta_a \quad (2)$$

The first part of (2) depicts the hosting cost of contents inside the service area including the storage cost $m_i S_j$, the processing power consumption cost $u_i \delta_j$, and the bandwidth cost $b_i \beta_j$. On the other hand, the second part shows the transmission cost of the streams that have to be redirected for non-existent contents inside the service area. To estimate the transmission cost, the number of streams is calculated firstly by dividing the size of redirected object m_i by the bit rate CBR . Next, the bandwidth unit cost of the service area β_a is provided and multiplied by the number of streams for redirected objects to calculate the redirection cost. The inclusion of popularity value p_i of redirected object in the second part of this equation is important to control which contents should be redirected. Only the low popular contents can be redirected due to their low redirection cost.

Upon completion of CARAM process, the optimal topology can be identified through the following components:

1. Servers and Resources– A server is defined through its location, the allocated computation power, storage space and bandwidth. The server's bandwidth is allocated for delivering the contents to the end-users. Thus, it is dimensioned based on the aggregated amount of requests targeting the contents stored in this server.
2. Replication Pattern- The contents are replicated according to their popularity and expected load such as the popular contents can be replicated in the service area while, the low popular contents can be redirected to the other service area. For that, the replication pattern determines the replicated contents, the replication degree, and the redirected contents, as well.

4.3. Decision variables

In fact, there are two binary decision variables are introduced to formulate the content-aware resource allocation problem. The first binary variable Y_j is employed to decide which potential location must be belonged within the optimal topology while, the second one X_{ija} to govern the replication pattern.

$$y_{aj} = \begin{cases} 1 & \text{if site } j \text{ in service area } a \text{ is selected as a server} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

$$x_{ija} = \begin{cases} 1 & \text{if content } i \text{ is replicated into a server } j \text{ in service area } a \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

In fact, the two binary decision variables are related to each other and affected by each other. For instance, these two variables determine the potential locations that should be selected according to the minimum cost of the replication scheme, at the same time; they decide whether the contents can be replicated in a potential location or not. Such relation supports the argument that the network topology, resources dimensioning, and content replication should be conducted mutually due to the mutual interaction between them.

4.4. The proposed optimization model

The resources allocation problem as a constrained binary optimization model can be formulated mathematically as explained in (5-11).

Minimize:

$$C_T = C_L + C_R \quad (5)$$

Subject to

$$\sum_{i \in N} m_i x_{ija} \leq m_j^{max} \quad \forall j \in PL_a, \forall a \in A \quad (6)$$

$$\sum_{i \in N} u_i x_{ija} \leq u_j^{max} \quad \forall j \in PL_a, \forall a \in A \quad (7)$$

$$\sum_{i \in N} b_i x_{ija} \leq b_j^{max} \quad \forall j \in PL_a, \forall a \in A \quad (8)$$

$$\sum_{i \in N} L_i x_{ija} \leq L_j^{max} \quad \forall j \in PL_a, \forall a \in A \quad (9)$$

$$\sum_{j \in PL_a} x_{ija} \leq P \quad \forall i \in N, \forall a \in A \quad (10)$$

$$\sum_{j \in PL_a} x_{ija} = p_i * P \quad \forall i \in N, \forall a \in A \quad (11)$$

$$x_{ija} \in [0,1] \quad \forall i \in N, \forall j \in PL_a, \forall a \in A \quad (12)$$

$$y_{ja} \in [0,1] \quad \forall j \in PL_a, \forall a \in A \quad (13)$$

$$x_{ija} = 0 \quad \forall i \in N \text{ not allocated in } j \text{ inside } a \quad (14)$$

$$x_{ija} \leq y_{ja} \text{ for all } i, j \in a \quad (15)$$

$$\sum_{j=1}^{PL_a} y_{ja} = P \quad (16)$$

In this model, the cost function is illustrated in (5) as the sum of server initiation cost and replication cost. Constraints (6–9) enforce the capacity of servers in terms of storage, processing power, bandwidth, and workload, respectively. These constraints (6–9) state that the total storage space, processing power, bandwidth, and workload of all contents replicated at server j should not exceed this server capacity. The number of copies for content i inside a service area has to be less than or equal number of selected potential servers P in that service area as depicted in equation (10). In addition, the number of copies for content i should be bounded by its popularity value P_i as depicted in equation (11). The integrality and non-negativity constraints are presented in equations (12-14). The contents should not be replicated on a potential server unless this potential location is selected as a server according to equation (15). In other words, the contents will not replicate on a site j unless this site is selected as a server. Finally, the number of servers that

should be selected in any service area should be bounded by a predefined number P as in (16) where P is a predefined number of locations in the service area.

4.5. The complexity of the proposed optimization model

The proposed optimization model for Content-Aware Resource Allocation problem is a combinatorial optimization problem and belongs to NP-complete class. We can prove the completeness of this problem by considering the whole network consists of a single service area with a single server, which allocates all the objects. Moreover, we assume that the resources cost, except the storage cost, equals zero. According to these assumptions, the proposed optimization problem will be written mathematically as.

Minimize:

$$C_T = \sum_{i \in N} m_i S_j \quad (17)$$

Subject to

$$\sum_{i \in N} m_i x_{ija} \leq m_j \quad \forall j \in PL_a, \forall a \in A \quad (18)$$

The simplified form of this problem is similar to the well-known knapsack problem, which is an NP-complete problem [40]. For clarity, the knapsack problem tries to minimize/maximize a certain quantity with respect to some constraints. For instance, the knapsack problem aims at maximizing the obtained profit or minimizing the cost without exceeding the knapsack capacity. Similarly, CARA problem aims to minimize the allocation cost with respect to some enforced constraints.

5. HYBRID GENETIC ALGORITHM FOR SOLVING CARA PROBLEM (HGA_CAR)

There is no algorithm with a polynomial time complexity can solve the NP-complete combinatorial optimization problems. On the other side, applying the evolutionary algorithms is a good choice to solve such problems. Thus, we integrate the Genetic Algorithm with two heuristic repair algorithms to solve CARA problem. Such integration reduces the search space and improves the ability of the proposed model to produce optimal solutions efficiently.

5.1. Heuristic repair algorithms

The problem of producing infeasible chromosome can be solved by incorporating problem-specific knowledge. For CARA problem, we have proposed two heuristic repair methods. The first one, called Restricted Search Operator (RSO), is to choose a certain number of servers in each service area based on the constraint 16. The second one, called Heuristic Replica Repair (HRR) algorithm is proposed to control the replica placement constraints.

5.2. Restricted search operator (RSO)

In many combinatorial optimization problems, there is a need to deal with chromosomes that hold a specified number of ones; for example the problem of selecting a subset of candidate locations to find the optimal solution, as in CARA problem. In this case, the individuals, which have more or less than the specified number of ones, will be rejected. Thus, Restricted Search Operator (RSO) is proposed to deal with this constraint. According to [43], RSO was firstly proposed by Salcedo-Sanz in 2004; next it is considered an extra operator added to the conventional GA. The idea of RSO is that the individual generated from crossover and mutation operators x may hold p ones. These ones may be different from the desired number M , thus, if $p < M$ the RSO adds $M - p$ ones in randomly chosen positions. On the other side, if $p > M$ the RSO selects $p - M$ ones randomly to be removed from the binary vector. The following Pseudo-code describes the Restricted Search Operator.

```

Let M be the specified number of 1s in an individual.
For every generation of GA
  For every individual in the population
    Check the number of 1s (p)
    If (p < M) add_ones(M-p)
    Else Remove_ones(p-M)
  end(if)
end(individual)
end(generation)

```

Figure 2. Restricted search operator [42]

In CARA problem, the restricted search algorithm is adapted to choose a certain number of servers where each server is represented as a string of binary values instead of a set of genes. In other words, the restricted search operator in the resource allocation case will work on a set of bulks instead of a set of genes; each bulk in this case consists of a set of sequential genes. As will be explained through this paper, each potential location j in the service area a consists of a set of sequential genes starting from the gene number $((a - 1) * PL_a + (j - 1)) * N + 1$ and ending with the gene number $((a - 1) * PL_a + (j - 1)) * N + N$. The restricted search operator represents the constraint 16 as follows: depending on the predefined number of locations P , the restricted search operator will choose randomly P bulks in each service area to be unchangeable. After that, the rest bulks in the service area will be change to zero. Thus, if the value of all cells for any location equal zero, this location will not be selected to be server. After applying the restricted search operator (RSO), which determines the number of selected servers, HRR algorithm checks each chromosome by counting the number of replicas for each video in the selected servers that are produced from RSO process.

5.3. Hybrid genetic algorithm for CARA problem (HGA_CARA)

To solve CARA problem, the genetic algorithm is hybridized with two repair algorithms. In this hybrid model, the objective function is to minimize the hosting cost (i.e. equation 5) with respect to the listed constraints.

In CARA problem, there are a set of candidate locations to construct servers in each service area. The aim is to select a subset of these locations to replicate the contents over them based on predefined criteria. For this problem, the binary encoding for the chromosomes is exploited. In such binary encoding, the chromosome is represented as a string of ones and zeros. This binary string represents the candidate solutions, in which each cell indicates for the replica position in the service area. The total length of the chromosome is $A * PL_a * N$ where A represents the number of service areas, PL_a indicates to the number of candidate locations in the service area, and N denotes the number of contents.

For instance, if the delivery network comprising of 2 service areas with two potential locations for each and the number of contents equals to 5, then the total length of chromosome equals to 20 genes, as depicted in Figure 3. Each cell in the chromosome holds a binary value $x_{ija} \in \{0,1\}$ that indicates whether or not an object i is replicated at the server j inside the service area a . The position of any gene can be identified according to the following (20).

$$gene(X_{ija}) = ((a - 1) * PL_a + (j - 1)) * N + i \quad (20)$$

	PL ₁₁					PL ₁₂					PL ₂₁					PL ₂₂				
y_{ja}	1					0					1					0				
x_{ija}	1	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

Figure 3. Chromosome representation for CARA problem

For example, the gene 13 holds a replica for the 3rd content into the first server in the second service area a according to Figure 3. We can find this gene using equation (20). $gene(x_{3,1,2}) = ((2 - 1) * 2 + (1 - 1)) * 5 + 3 = 10 + 3 = 13$. The potential location j in the service area a holds the genes from $((a - 1) * PL_a + (j - 1)) * N + 1$ to $((a - 1) * PL_a + (j - 1)) * N + N$. For instance, if there are 20 contents to be replicated over a delivery network with 5 service areas, each area contains 4 servers. The total length of the CARA problem chromosome will be $4 * 5 * 20 = 400$ genes. The third server in the second service area will be represented by the genes from the position 121 to the position 140.

The initial population of this problem is generated randomly by selecting a random binary number for each gene in the chromosome. During the evolution process, the roulette wheel selection technique is implemented to select two chromosomes at each time. The uniform crossover operator with a standard crossover probability is applied on the selected pair. After that, the reproduced offspring are mutated, with an extremely small mutation probability, to ensure the diversity in the search space. During the reproduction process, the new offspring may lie in the region of unfeasible solutions; therefore, two heuristic repair algorithms are used to repair this offspring to be obeyed by the proposed constraints as depicted in Figure 4.

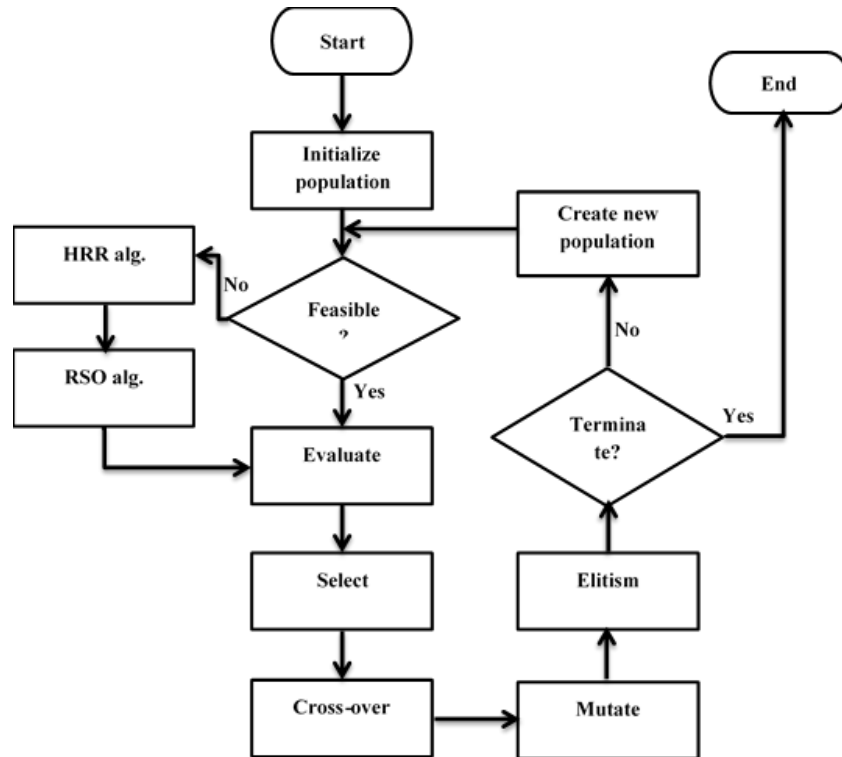


Figure 4. Hybrid GA for CARA problem (HGA_CARAs)

6. THE EXPERIMENTAL RESULTS

To evaluate the performance of the proposed content-aware resource allocation model (CARAM), an experiment using the proposed evolutionary approach HGA_CARAs is conducted. Here, the numerical results are reported. The experiments have been conducted on a system model of 40 servers divided among 10 service areas and 100 video contents. The predefined number of servers P equals 2. This means that the model has to choose two potential locations among four locations to become servers. These potential locations are associated with the cost of storage, processing power, and bandwidth resources. Moreover, the video contents have to be replicated on the selected servers in each service area during the experiment. In this section, the effect of content status on the resource allocation process is investigated. The effect of content status is studied using two popularity distributions.

The aim of using two popularity distributions in this experiment is to study the impact of the change in the popularity distributions on resource allocation process. This impact is measured in terms of storage, CPU consumption, bandwidth, and the number of popular contents. The popularity distributions have been sampled using two different equations. The first popularity distribution follows the normal density function [10]. This popularity distribution has been sampled for 100 contents and depicted in equation (21). We refer to this distribution by the symbol (DF).

$$p_i = \frac{e^{-((i-1)\Delta)^2/2}}{\sum_{i=1}^n e^{-((i-1)\Delta)^2/2}} \quad (21)$$

The other popularity distribution has been sampled for the same content size according to Zipf's law distribution. Like the symbol k in Zipf distribution, the symbols Δ represents for the degree of skewness in the popularity distribution (DF). The expected load of contents is estimated using the proposed IPTV content status model for the two popularity distributions.

Figure 5 depicts the two popularity distributions with different skewness values 0.01, 0.04, and 0.08. According to this figure, the curve of small skewness values (i.e. 0.01 and 0.04) indicates that the contents popularity are distributed in an uniform manner, which means that all contents have the same popularity value. However, the curve tends to be more skewed when the skewness value increased. Notably, DF distribution tends to be more skewed than Zipf's distribution in the figure when the skewness value increased (equals 0.08 as in the figure). Consequently, the number of popular contents in DF distribution is less than

the number of popular contents in Zipf's distribution. For instance, Figure 4 shows that 36% and 66% of contents are popular for DF and Zipf's distributions, respectively. The same thing can be observed in Figures 6 and 7 where Figure 6 shows that only 6% of the contents are popular at the skewness value equals 1 for DF distribution, whereas the other contents having the popularity value zero.

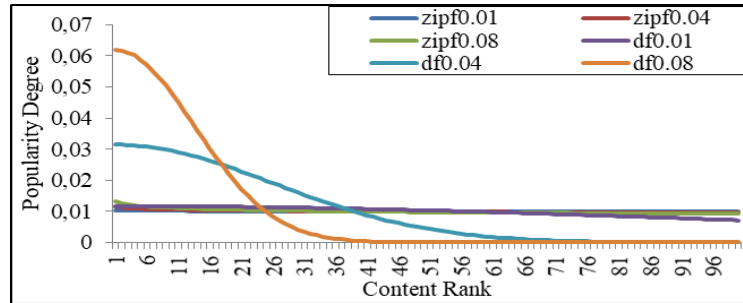


Figure 5. The popularity distributions with different skewness values

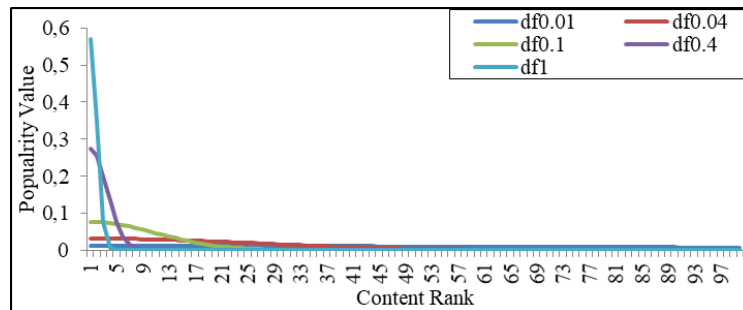


Figure 1. DF popularity distribution with different skewness values

On the other side, at the same skewness value, Figure 7. shows that more than 16% of contents are popular with a long tail of the curve reaches 70% of contents for Zipf's law distribution. The popularity value of these contents is low, but it does not equal zero. That means that Zipf's distribution replicates more contents than DF distribution.

Therefore, it can be concluded that at the higher skewness values, a small percentage of contents are popular. However, in DF distribution, this percentage is less compared to Zipf's distribution. This behavior can be interpreted as follows. In the DF distribution, the popularity degree divided among the contents in a smooth manner wherein the contents popularity values are close to each other at the low skewness degree. However, when the skewness value tends to increase, the disparity in popularity degree of contents becomes clearer. Unlike DF, the popularity curve starts with popular contents and then cascaded down gradually. On the other side, the DF distribution looks similar to Zipf's distribution at the lower skewness degree. However unlike the Zipf's distribution, the DF popularity curve lacks the smooth gradient at higher skewness degree. In general, the popularity degree of popular contents is higher in DF distribution than of those in Zipf's distribution at the same skewness values.

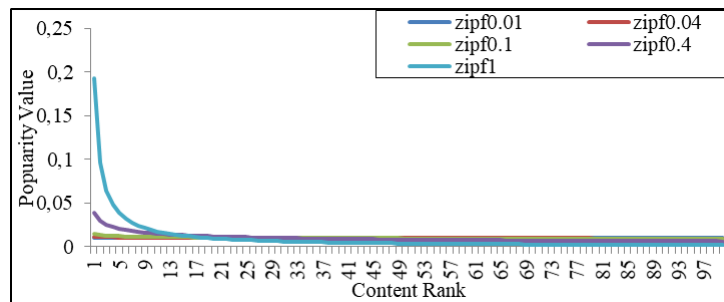


Figure 7. Zipf's popularity distributions with different skewness values

Based on this observation, the load of the popular contents in DF distribution is higher than the load of the same popular contents in Zipf's distribution at the same degree. This behavior of load distribution is depicted in the Figures 8. and 9. where the load of the first content is 800 and 5000 for Zipf and DF distributions, respectively. As mentioned through this thesis, the popularity distribution is not the only factor that affects the anticipated load distribution. However, it is the most influential factor. Thus, the anticipated load distribution of the contents takes the same behavior of the popularity distribution, but with some deviations. Figure 8. depicts the load distribution for DF distribution at low skewness values. This figure depicts that the popular contents take larger workload portion than the other contents. This is due to the highly skew of DF distribution as pointed out in Figure 6.

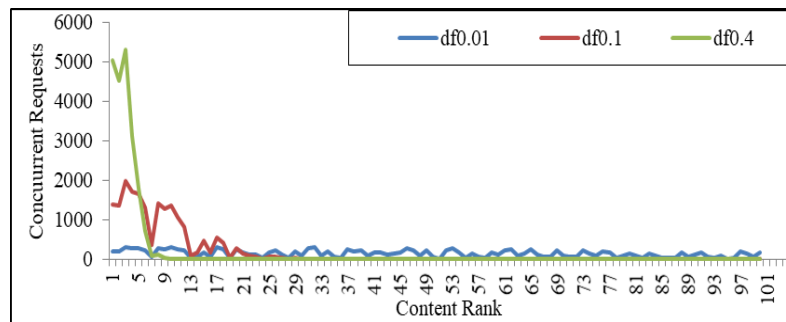


Figure 8. Expected load according to DF distribution

Accordingly, most of the incoming workload will be redirected to a small portion of contents. Unlike Zipf's law, the load distribution following DF distribution will be more skewed than Zipf's distribution. Such distribution follows the popularity skewness. For instance, a small percentage of contents utilize a higher portion of the load where these popular contents take a higher popularity degree. On the other side, the other contents having very low popularity degrees take the rest portion of the load, which is very low or in many cases equals zero. On the other hand, the expected load of contents that follow Zipf's law distribution will be distributed more smoothly than DF distribution as depicted in Figure 9. However, definitely the popular contents will take the lion's share.

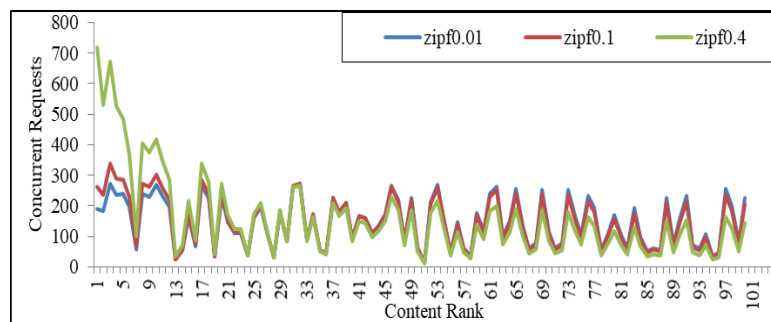


Figure 9. Expected load according to Zipf's distribution

Based on the different skew behaviors of these two popularity distributions, the effect of content status on the aspect of resources allocation process with accordance to these two popularity distribution is studied. The effect of content status on resource allocation process is investigated and reported below in terms storage, bandwidth, and the number of popular contents.

The expected load at different skewness values in the service area is depicted in Figure 10. In this figure, the expected load of DF distribution is higher than the expected load of Zipf distribution. This is due to that the expected load of contents in DF distribution is higher than the contents in Zipf distribution.

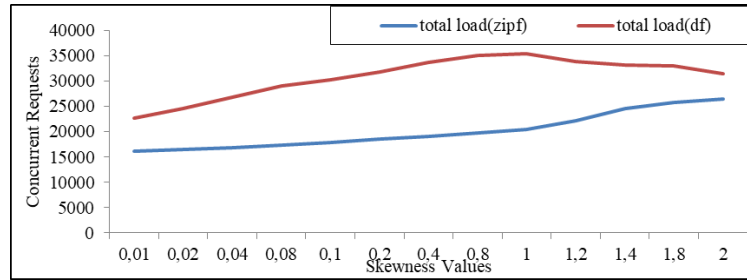


Figure 10. The total load for one service area

What is apparent from the same figure is that in case of the expected load of contents following Zipf distribution, the expected load increases linearly when the skewness value increases. On the other side, the expected load of contents following DF distribution follows the same behavior at the low skewness values. However, at the higher skewness values, the expected load alters its behavior inversely as it decreases when the skewness value increases. The behavior of DF can be interpreted by the sharp disparity in the popularity degree at higher skewness values. Such sharp disparity leads to store very small percentage of contents with high popularity values. Moreover, in case of redirecting any content with lower popularity, a significant portion of the workload can be redirected. Such redirection leads to decrease the load on the servers. On the other side, the contents with low popularity values may be redirected to other service areas. Redirecting these contents leads to reduce the load on servers. Such reduction in load interprets the dwindling in the total load of DF distribution at higher skewness values.

Furthermore, the required bandwidth of any content in the service area is a function of both the active streams at the server and the bit rate of the streamed contents. Thus, the expected bandwidth for the delivery network follows the workload distribution if we assume that the bit rate of all contents is fixed. The expected bandwidth for one service area for bit rate=9.8 Mbps is depicted in Figure 11.

Figure 12 plots the exploited storage versus different skewness degrees for both popularity distributions. The figure shows that the exploited storage space tends to decrease when the skewness degree increases. However, the storage space of contents following DF distribution decreases rapidly more than the storage space of those content following Zipf distribution. As depicted in Figure 12, the storage space of contents following Zipf distribution decreases from 58 GB to 44 GB while the storage space of contents following DF distribution decreases from 82 GB to 22 GB for the skewness values (0.01–2).

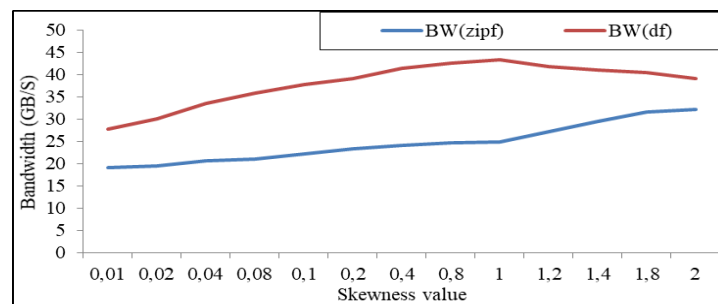


Figure 11. The expected bandwidth for one service area

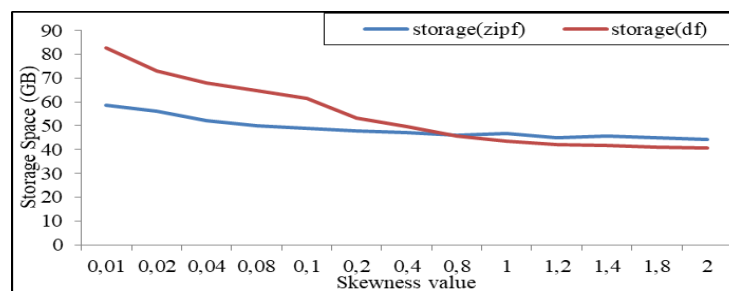


Figure 12. The exploited storage versus different skewness degrees

This behavior can be interpreted as follows. In case of DF distribution, the contents are close to each other in their popularity, which leads to replicate more contents in the service area. Moreover, the number of replica of each video is high because of the popularity values are closer to each other than Zipf's distribution at the same low skewness values, which leads to increase the number of replica for the replicated contents, as depicted in Figure 13. Consequently, the exploited storage space becomes high. From the same Figure, the number of replicas for Zipf distribution-based replicated contents is less than that of DF distribution, which leads to exploit a less storage space for Zipf distribution contents.

In contrast, the exploited storage space for both types is decreased at the higher skewness value (e.g. at skewness values equals 1 and 2) due to that only very small percentage of contents will be replicated in the service area, as depicted in Figure 14. However, DF exploits a storage space slightly less than Zipf distribution at higher skewness values. This is due to that DF replicated a percentage of contents slightly smaller than the percentage of contents replicated by Zipf's distribution.

Finally, the number of stored contents in the service area is measured and depicted in Figure 15. From this figure, it can be noticed that the number of stored contents in the service area decreases when the skewness value increases. However, the decreasing for DF distribution is larger than that for Zipf distribution. The interpretation for this phenomenon refers to the fact that the skewness in DF distribution is higher than the skewness of Zipf distribution.

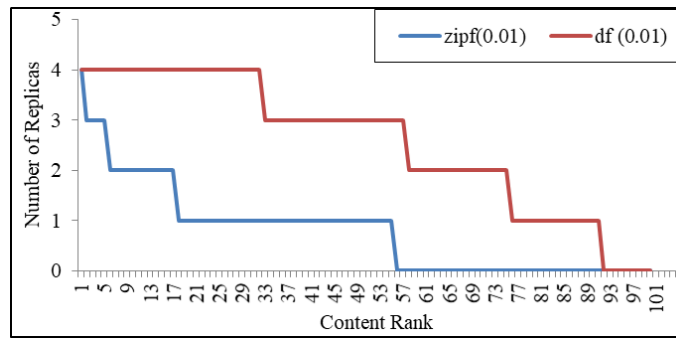


Figure 13. The number of replica at low skewness values (=0.01)

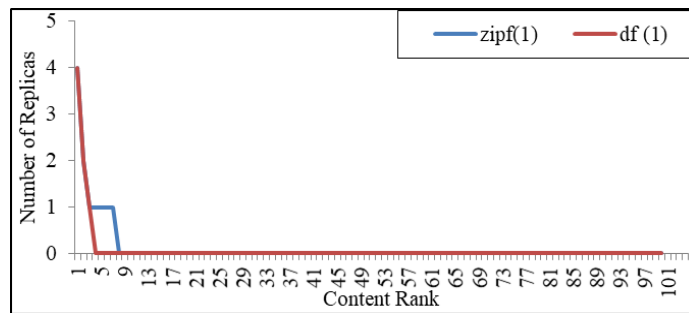


Figure 14. The number of replica at high skewness values

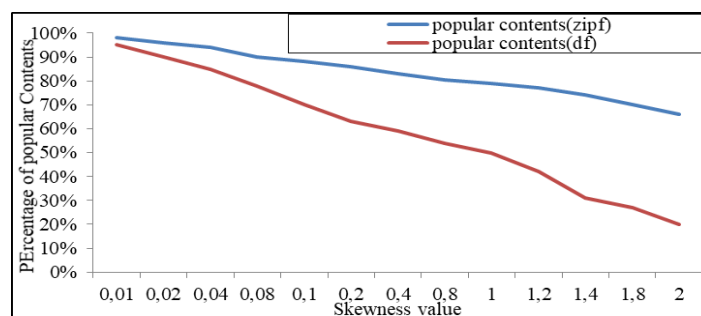


Figure 15. Percentage of popular contents for both distributions

7. CONCLUSION

Integrating the replica placement with resource allocation is a necessity to cope with the fluctuation of demand on IPTV services. In this paper, Content-Aware Resource Allocation Model (CARAM) is proposed. CARAM model considers the contents' status to decide the required resources. For this purpose, the replica placement problem and the resource allocation problem have been integrated aiming to allocate the resources depending on the necessitated replication scheme. The content-aware resource allocation problem is studied and formulated as a constrained binary optimization problem. The proposed model has been solved using Hybrid Genetic Algorithm. The experiments have been conducted on two popularity distributions. The experimental results showed the impact of content status on the resource allocation process. Thus, considering the content's status during the resource allocation process helps in allocating the resources according to the demand without wasting the resources. The future work for this direction will focus on building online optimization method for allocating the resources on the fly. Such online optimization helps in allocation the resources according to the real demand.

REFERENCES

- [1] A. Yarali, and A. Cherry, "Internet Protocol Television (IPTV)," in *Proc. Of IEEE Region TENCON 2005, Melbourne, Australia*, pp. 1-6, 2005.
- [2] S. B. Lee, G. Muntean, and A. F. Smeaton, "Performance-aware Replication of Distributed Pre-recorded IPTV Content," *IEEE Trans. On Broadcasting*, vol. 55, 2, pp. 516-526, 2009.
- [3] T. Ma, Y. Chu, L. Zhao, and O. Ankhbayar, "Resource Allocation and Scheduling in Cloud Computing: Policy and Algorithm," *IETE Technical Review*, vol.3, 1, pp. 4-16, 2014.
- [4] S. Xu, "Replica Placement Algorithms for Efficient Internet Content Delivery," *Doctoral dissertation, The University of Adelaide, Australia*, 2009.
- [5] Gaber, S. M. A., & Sumari, "P. Predictive and Content-aware Load Balancing Algorithm for Peer-service Area based IPTV Networks. *Multimedia Tools and Applications*, pp. 1-24, 2012.
- [6] Fati SM, Sumari P., "Content-aware Replica Placement Strategy for IPTV Services Over Peer-service Area Architecture. *Multimedia Tools and Applications*," vol. 76, no. 7, pp. 10041-65, Apr 2017.
- [7] Fati SM, Sumari P, Yuhaniz SS, Sjarif NN, "Modelling Contents Status for IPTV Delivery Networks, " *Proceedings of the 6th International Conference on Computing and Informatics, ICOCI 2017*, Kuala Lumpur. Universiti Utara Malaysia, pp. 25-27, April 2017.
- [8] I. Joe, J. H. Yi, and K. S. Sohn, "A Content-Based Caching Algorithm for Streaming Media Cache Servers in CDN," *Multimedia, Computer Graphics and Broadcasting. Springer Berlin Heidelberg*, vol. 262, pp. 28-36, 2012.
- [9] C. Vicari, "Distributed Dynamic Replica Placement and Request Redirection in Content Delivery Networks," Ph.D. thesis, Università degli Studi di Roma "La Sapienza", 2008.
- [10] D. Niu, "Demand Forecast, Resource Allocation and Pricing for Multimedia Delivery from the Cloud," Ph.D. thesis. University of Toronto, Canada, 2013.
- [11] N. Laoutaris, V. Zissimopoulos, and I. Stavrakakis, "On the Optimization of Storage Capacity Allocation for Content Distribution," *Computer Networks*, vol. 47,3, pp. 409-428, 2005
- [12] M. Li, and C. H. Wu, "A Cost-Effective Resource Allocation and Management Scheme for Content Networks Supporting IPTV Services," *Computer Communications*, vol. 33, 1, pp. 83-91, 2010.
- [13] M. Pathan, and R. Buyya, "Architecture and Performance Models for QoS-driven Effective Peering of Content Delivery Networks," *Multiagent and Grid Systems*, vol. 5, no. 2, pp. 165-195, 2009.
- [14] C. Huang, J. Li, and K. W. Ross, "Can Internet Video-on-demand be Profitable?," *ACM SIGCOMM Computer Communication Review*, vol. 37, no. 4, pp. 133-144, 2007.
- [15] X. Hei, C. Liang, I. Liang, Y. Liu, and K. W. Ross, "A Measurement Study of a Large-scale P2P IPTV System," *IEEE Trans. on Multimedia*, vol. 9, 8, pp.1672-1687, 2007.
- [16] Kulatunga, C., Kandavanam, G., Rana, A. I., Balasubramaniam, S., & Botvich, D. "HySAC: a Hybrid Delivery System with Adaptive Content Management for IPTV Networks." *In Communications (ICC), 2011 IEEE International Conference on*, pp. 1-5I, 2011.
- [17] M. Karlsson, C. Karamanolis, and M. Mahalingam, "A Framework for Evaluating Replica Placement Algorithms. Internet Systems and Storage Laboratory," *HP Laboratories Palo Alto*. HPL-2002-219 August, 2002.
- [18] Z. Du, J. Hu, Y. Chen, Z. Cheng, and X. Wang, "Optimized QoS-aware Replica Placement Heuristics and Applications in Astronomy Data Grid," *Journal of Systems and Software*, vol. 84, 7, pp. 1224-1232, 2011.
- [19] F. Thouin, and M. Coates, "Video-on-demand Networks: Design Approaches and Future Challenges," *IEEE Network*, vol. 21, no. 2, pp. 42-48, 2007.
- [20] K. S. Tang, K. T. Ko, S. Chan, and E. W. Wong, "Optimal File Placement in VOD System Using Genetic Algorithm," *IEEE Trans. on Industrial Electronics*, vol. 48, 5, pp. 891-897, 2001.
- [21] T. Wauters, J. Coppens, F. De Turck, B. Dhoedt, and P. Demeester, "Replica Placement in Ring based Content Delivery Networks," *Computer Communications*, vol. 29, no. 16, pp. 3313-3326, 2006.
- [22] J. Guo, Y. Wang, K. S. Tang, S. Chan, E. W. Wong, P. Taylor, and M. Zukerman, "Evolutionary Optimization of File Assignment for a Large-scale Video-on-demand System," *IEEE Trans. On Knowledge and Data Engineering*, vol. 20, 6, pp. 836-850, 2008.

- [23] S. Zaman, AND D. Grosu, "A Distributed Algorithm for the Replica Placement Problem," *IEEE Transactions on Parallel and Distributed Systems*, vol. 22, 9, pp. 1455-1468, 2011.
- [24] C. C. Bisdikian, , and B. V. Patel, "Cost-based Program Allocation for Distributed Multimedia-on-demand Systems," *IEEE multimedia*, vol. 3, 3, pp. 62-72, 1996.
- [25] I. Cidon, , S. Kutten, , and R. Soffer, "Optimal Allocation of Electronic Content," *Computer Networks*, vol. 40, 2, pp. 205-218, 2002.
- [26] J. M. Almeida, "Streaming Content Distribution Networks with Minimum Delivery Cost," Doctoral dissertation, University of Wisconsin, USA, 2003.
- [27] H. Ebara, A. B. E. Yasutomo, D. Ikeda, T. Tsutsui, K. Sakai, A. Nakaniwa, and H. Okada, "A Cost-effective Dynamic Content Migration Method in CDNs," *IEICE Transactions on Communications*, vol. 88,12, pp. 4598-4604, 2005.
- [28] T. A. Neves, L. Drummond, L. S. Ochi, C. Albuquerque, E. Uchoa, "Solving Replica Placement and Request Distribution in Content Distribution Networks," *Electronic Notes in Discrete Mathematics*, vol. 36, pp. 89-96, 2010.
- [29] M. Karlsson, and C. Karamanolis, "Choosing Replica Placement Heuristics for Wide-area Systems," *In Proc. Of 24th IEEE International Conference on Distributed Computing Systems*, Tokyo, Japan, pp. 350-359, 2004..
- [30] F. Tenzakhti, K. Day, and M. Ould-Khaoua, "Replication Algorithms for the World-wide Web," *Journal of Systems Architecture*, vol.50, 10, pp.591-605, 2004.
- [31] A. Mahmood, "Replicating Web Contents Using a Hybrid Particle Swarm Optimization," *Information processing & management*, vol. 46, 2, pp. 170-179, 2010.
- [32] S. U. Khan, and I. Ahmad, "Comparison and Analysis of Ten Static Heuristics-based Internet Data Replication Techniques," *Journal of Parallel and Distributed Computing*, vol. 68, 2, pp. 113-136, 2008.
- [33] T. Loukopoulos, and I. Ahmad, "Static and Adaptive Distributed Data Replication Using Genetic Algorithms," *Journal of Parallel and Distributed Computing*, vol. 64, 11, pp. 1270-1285, 2004.
- [34] D. Liao, G. Sun, V. Anand, and K. Xiao, "Efficient Provisioning of Hybrid Virtual Network with Stochastic Resource Demands," *IETE Technical Review*, vol. 31, 5, pp. 342-352, 2014.
- [35] A. Belbekkouche, M. Hasan, and A. Karmouch, "Resource Discovery and Allocation in Network Virtualization," *IEEE Communications Surveys & Tutorials*, vol. 14, 4, pp. 1114-1128, 2012.
- [36] S. Kitjongthawonkul, and J. Ko, "Using an Effective Algorithm to Resolve the Video-on-demand Service Network Resource Allocation Problem," *In proc. Of 13th IEEE International Conference on Advanced Communication Technology*, Phoenix Park, Korea, pp. 1031-1036, 2011.
- [37] I. Houidi, W. Louati, W. Ben Ameer, and D. Zeglache, "Virtual Network Provisioning Across Multiple Substrate Networks," *Computer Networks*, vol. 55,4, pp. 1011-1023, 2011.
- [38] S. Pandey, Y. J. Won, J. W. Hong, J. Strassner, "Dimensioning Internet Protocol Television Video on Demand Services," *International Journal of Network Management*, vol. 21,6, pp. 455-468, 2011.
- [39] F. Thouin, "Video-on-demand Equipment Allocation," *Master thesis, McGill University Montreal*, 2007.
- [40] T. Nguyen, F. Safaei, P. Boustead, and C. Tung Chou, "Provisioning Overlay Distribution Networks," *Computer Networks*, vol.49, 1, pp.103-118, 2005.
- [41] A. Nakaniwa, and H. Ebara, "Optimal Allocation of Cache Servers and Content Files in Content Distribution Networks," *In Proc. of European Conference on internet and multimedia systems and applications, Chamonix, France*, pp. 15-22, 2007.
- [42] M. E. Aldana Diaz, and E. N. Huh, "Cost Analysis on IPTV Hosting Service for 3rd Party Providers," *In Proc. Of 5th ACM International Conference on Ubiquitous Information Management and Communication*, Seoul, Korea, p. 114, 2011.
- [43] S. Salcedo-Sanz, "A Survey of Repair Methods Used as Constraint Handling Techniques in Evolutionary Algorithms," *Computer science review*, vol. 3, 3, pp. 175-192, 2009.

BIOGRAPHIES OF AUTHORS



Suliman M. Fati—obtained his BS.c (2002), MS.c. (2009), and Ph.D. (2014) from Ain Shams University-Egypt, Cairo University -Egypt. And Universiti Sains Malaysia (USM)—Malaysia respectively. He is an assistant Professor in College of Computers and Information Sciences, Prince Sultna university, Saudi Arabia. His research interests focus on Internet of Things, Machine Learning, Social Media Mining, Cloud Computing, Cloud Computing Security, and Information Security. He has authored over 20 ISI and Scopus journal/conference papers, books, and book chapters. He is the lead editor for an edited book shortly be published by Wiley entitled "A Comprehensive Guide to IPTV Delivery Networks". Besides, He is a member of different professional bodies as IEEE, IACSIT, IAENG, and Institute of Research Engineers and Doctors, USA. He is a reviewer in many international impact-factor journals, and also technical committee program member in a number of international conferences.



Putra Sumari—obtained his MSc and PhD in 1997 and 2000 from Liverpool University, England. Currently, he is an Associate Professor at the School of Computer Science, Universiti Sains Malaysia. His research areas include video on demand system, multimedia storage server, MPEG standard, image/video compression and retrieval, image Cryptography. He has published more than 70 papers related to these areas. He is the head of the Multimedia Research Group, School of Computer Science, USM.



Choo Wou Onn-specializes in smart school systems, knowledge management systems, ecommerce and learning management systems. He received his Bachelor, Master, and Ph.D in Information Technology, major in Information Science, from National University of Malaysia (Universiti Kebangsaan Malaysia (UKM)). He has been actively involved in conducting research in the areas of multimedia courseware design and development, information systems, and management information systems particularly in education field. He had involved in a few of funded projects such as Seeding Grants, IRPA, FRGS and Techno-Fund by MOSTI. This leads him to be pro-active in both teaching and research. He supervises both undergraduate and postgraduate students (Masters and Ph.D levels). He has more than 40 papers in International and Local Journals, International and National Conference proceedings as well as in Newspaper article. His works received more than 15 International and Local Awards including best invention in ICT. He also serves as member in editorial board and technical reviewer for local and international journals as well as conferences.