OPTIMIZATION OF DATA CENTER ENERGY CONSUMPTION USING GENETIC ALGORITHMS

By

Faisal A. Alrwilan

Supervisor

Mohammad Al-Jarrah

Program: Master of Science in Computer Engineering / Embedded Systems Engineering

Nov. 2016

OPTIMIZATION OF DATA CENTER ENERGY CONSUMPTION USING GENETIC ALGORITHMS

By

Faisal A. Alrwilan

B.Sc. Computer Science, Al-Juof University, 2009

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Department of Computer Engineering, Yarmouk University, Irbid, Jordan.

Signature of Author ------------------------------

Committee Member Signature and Date

Dr. Mohammad Al-Jarrah (Chairman) ------------------------------

Dr. Ali Eyadeh (Member) ------------------------------

Dr. Mwaffaq N. Otoom (Member) ------------------------------

DEDICATION

***To* my daughter, Raghad**

ACKNOWLEDGMENTS

First and foremost all praises and thanks to Allah for giving me the ability to write this thesis.

I would like to thank my supervisor for the guidance and support. Without your helpful suggestions, advice and encouragement, this work would not have been possible, and I would like to thank the instructors and the staff in my department.

I would also like to thank my friends and colleagues for all the discussions and advices that we have shared throughout the years.

I am very grateful to my family and my warmest thanks to my dear mother for her support, patience, and love.

DECLARATION

I am Faisal A. Alrwilan, I hereby declare that this thesis entitled "OPTIMIZATION OF DATA CENTER ENERGY CONSUMPTION USING GENETIC ALGORITHMS", submitted to the "Department of Computer Engineering at Hijjawi Faculty for Engineering Technology" is my original work and it has not been presented for any degree in any other university, and that any additional sources of information have been properly cited.

Name: Faisal A. Alrwilan.

ID Number: 2009971020

Date: November, 2016

Signature:

Table of Contents **Page**

Table of Contents **Page**

DIDICATION	**iii**
ACKNOWLEDGMENTS	**iv**
TABLE OF CONTENTS	**v**
LIST OF TABLES	**vi**
LIST OF Figures	**viii**
ABSTRACT	**ix**
Introduction	**1**

List of Figures **Page**

ABSTRACT

Faisal A. Alrwilan. OPTIMIZATION OF DATA CENTER ENERGY CONSUMPTION USING GENETIC ALGORITHMS. Master of Science Thesis, Department of Computer Engineering, Yarmouk University, 2014 (Supervisor: Dr. Mohammad Al-Jarrah).

A data center is a facility that is used to accommodate high computing and storage capacity systems such as computers, allied components, telecommunications devices and storage systems. It generally requires backup power supplies, surplus data communications links, air conditioning, environmental systems, fire suppression and security equipment. With growing energy expenses and increasing power usage due to the emergent demand for computing power (servers, storage, networks), energy bills have been developed into a major cost for today's data centers. This thesis presents an optimization methodology on the data center energy using genetic algorithm. The optimization is done through forecasting the computational power of the data center, and thus, supplying only the demanded power. The core problem is that the extra power that has not been used is wasted. The presented approach is considered as one of the early research to find the optimum solution for saving energy consumed in data centers using the genetic algorithm. The results were measured and recorded using MATLAB and discussed carefully in this thesis. This thesis achieved power optimization up to 78%.

CHAPTER ONE

INTRODUCTION

1.1 Introduction

A data center is a facility that used to accommodate high computing and storage capacity systems such as computers, allied components, telecommunications devices and storage systems. It generally requires backup power supplies, surplus data communications links, air conditioning, environmental systems, fire suppression and security equipment. With growing energy expenses and increasing power use due to the emergent demand for computing power (servers, storage, networks), energy bills have been developed into a major cost for today's data centers. Moreover, the usage of such amount of energy will have environmental impacts too.

Data Center is known as a host which contains multiple types of equipment to perform specific functions like storing, processing, exchanging, and managing digital data in order to support the IT requirement of large networks structure, and ensure services for different processing types of data [1].

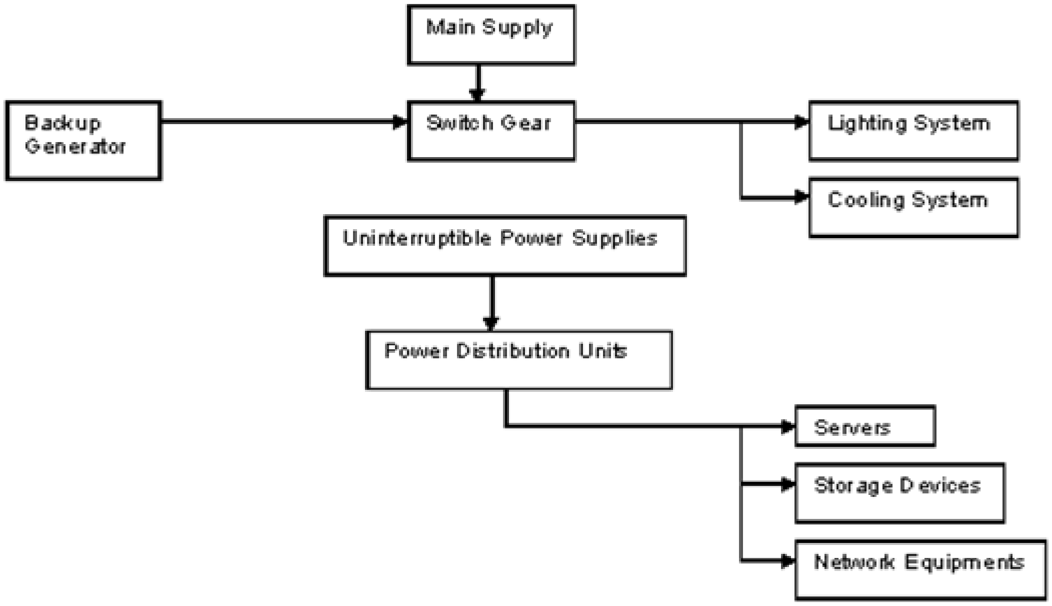


Fig.1.1: Data Center Equipment

There are many types of data centers are nowadays used in a large spectrum of industries, and different applications facilities. It is also used in research institutes, teaching organizations and others. There are various types of equipment that operate in the data center, such as different types of computers, storage, and network components, in addition to load balance equipment, wires, air conditioning systems, and related equipment for security. Fig. 1.1 displays the equipment in the data center. The power flow in the data center is shown in Table 1.1. Total power consumed by IT equipment and the uninterrupted power supplies is around 50% in a typical data center [2].

The data center could be implemented by the organization or the company itself or got sourced by a software systems vendor as in the well-known enterprise CRM and ERP tools. The data centers could be considered to handle only the operations architecture or it could handle other services. Normally, the applications are being composed of multiple hosts; everyone is operating one component. Common components that are well-known are file server, middleware, databases, application servers.

Table.1.1: Power Consumed in the Data [3]

Load	Power Consumption
Chiller	33%
Humidifier	3%
Computer room air conditioner	9%
IT Equipment	30%
Uninterrupted power supplies.	18%
Lighting.	1%
Power distribution units	5%
Switchgear /Generator	1%

One main application of data centers is the offsite backups. Organizations are subscribing to data center's backup services. This is normally used in synchronization with NAS backup tapes, NAS storage or SAN storage. Backup is possible to be performed without tapes by the local servers. However, tapes should be saved in site that is susceptible to fire and flooding and pose a security threat. Big companies could send backup out of the site for more safety and security. This is possible to be done by getting back all data to the data center.

For fast deployment or recovery of disaster, different large hardware vendors have developed mobile solutions in order to install and operating it in a very short interval. Companies such as IBM, Bull, Sun Microsystems, Cisco Systems, Huawei, HP and Google have implemented systems that can be used for such goals.

Infrastructure management is performed using special software, and hardware – including sensors – enables easy, real-time management and monitoring for all sub-systems though infrastructures.

Infrastructure management of the data center providers is increasingly linked to the providers of computational fluid dynamics in order to calculate the system complexity of airflow. The computational fluid dynamics elements are important to improve the future ahead impact on HVAC and efficiency.

Existing data centers have had three main operational and fiscal constraints – space, cooling, and power. As common data centers are needed to support increasingly dense configurations, cooling and power requirements could outstrip the data center capabilities infrastructures. Actually, the space issue becomes argue issue, because of that the existing data centers are subjected to run out of power not out of space.

Many vendors understand that individual challenges of data center like processor and rack cooling and level power cannot be monitored and handled as disconnected issues. Some vendors developed solutions for the basic issues of power and cooling at the server, processor, facility infrastructure, and rack levels, and implemented proven tools for management in order to provide a unified approach to handle the cooling and power in data centers.

1.2 Problem and Motivation

Usually all the attention during the design process and the implementation of data center is concentrated on the computational performance. Minimum attention or no attention at all is being paid to the data center energy requirements and efficiency. There is a need for the concern about the computational output per watt of energy consumed and targeting to increase the efficiency of the data center, since the devices in the data center are operating 24 hours / 7 days a week [3].

The continuous increase in the data center energy consumption and the cost of energy are the two main dimensions highlight the importance of reducing the energy consumption. This will play a major role in the strategy of reducing running costs of the data centers, managing the capacity, and promoting the environmental responsibility. Especially, the consumed power is not fit to the computational performance over time, but it almost much larger than the effective percentage.

Due to the large number and complexity of the hardware used in the data center, there is an urgent need to use optimization methods to reach the optimum design of data centers and the hardware populating them [4]. For example, the simulation of airflow in the data center is computationally intensive. This situation may be more complex when this simulation integrated with an optimization method. Also, the thermal design of data center is a hardware problem considered as multi-objective optimization problem. The efficient technique for solving this kind of problem is to use multi-object genetic algorithm optimization (MOGA) [5].

In fact, the main concern of this thesis is the wasted energy in contrast with the computational performance. Fig. 1.2 shows how the computational payload changes over time while the power capacity is constant. The figure shows that, the demand on the resources is not constant and it varies over time; in most cases, it varies periodically. In contrast, the power capacity resources are constant and much higher than the required. So, the wasted capacity is large and it also varies over time.

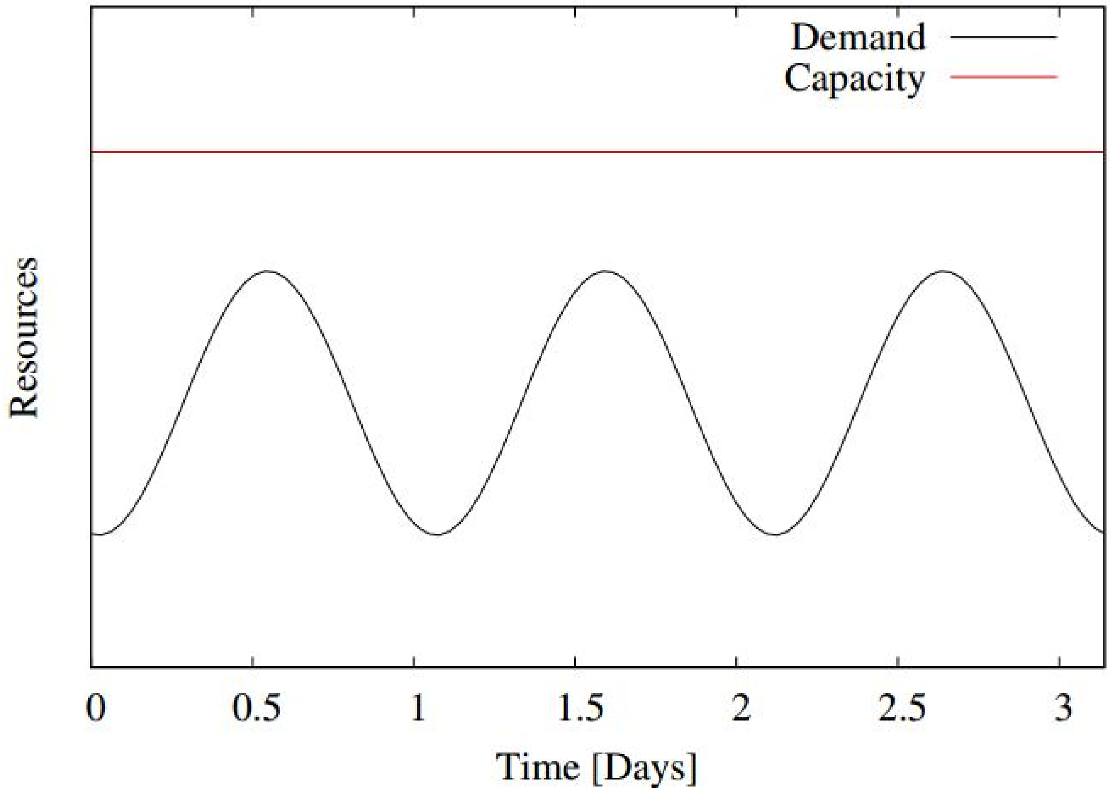


Fig.1.2: Data center resources, demand versus capacity

This thesis aims to optimize the data center energy consumption by minimizes the wasted energy. While in such two approaches are achieved together; computational payload forecasting, and minimal power capacity optimization.

In this thesis, a substantial and efficient genetic algorithm will be formulated through an analysis for data center space. First, three factors are defined (watts consumed, computational output, energy cost) which is used to form the objective and constraint functions of an optimization model. Then, an optimizing model will be generated through some steps which are: initialization of the population, evaluating the population, selecting the parents, performing a cross over, performing mutation, evaluating off spring, and repeating the previous steps until the stop criteria achieved.

1.3 Thesis Objectives

The main objective of this research is to find a way for data center in order to reduce the energy consumption. This is will be satisfied by using a genetic algorithm and taking into account the several conflicting constrains especially economic and environmental performance. The two concerns are the forecasting of computational payload and the power capacity optimization according to the forecasted payload. Such, in this thesis, genetic algorithm will be used to enhance the saving of energy consumed in data center.

1.4 Contribution

This thesis aims to produce friendly and cost effective designing system for modern data centers using advanced technology and techniques, through effective and efficient utilization of resources, to play a vital part in developing community and protecting the environment by saving the resources, and reducing the loses.

The proposed research adds new knowledge through using the genetic algorithm for designing the data center. This design will focus on reducing consumed energy, which is not mentioned in the previous researches even though each subject is studied well separately. The original work in this research will be the optimum and effective design for the data center for saving appreciated reasonable amount of energy by using the evolutionary genetic algorithm.

The main contribution of this research is that, using hybrid forecasting and optimization model to optimize the energy of the data center. The forecasting of computational payload to predict the demand on the data center, and such, optimize the power capacity to achieve minimal wasted energy.

The second contribution point is to build a model for forecasting and optimization, both using genetic algorithm. The genetic algorithm is adopting advanced intelligent systems and intelligent computations to perform non-analytically constrained model. Thus, it is expected to achieve high power optimization and energy saving in comparison with the related researches.

The genetic algorithm has many strength points such as: it operates with any sort of model; it allows very specific parameterization of conditions; it is flexible to fitness criteria and to allow optimization over fitness measures.

1.5 Literature Review

Genetic algorithm was first appeared in 1975 by John Holland in the paper under the topic of, "Adaptation Natural and Artificial System" [6]. The algorithm seed a number of iterations or generations; then each of them will be populated with a predefined number of the binary strings. These binary strings are decoded into a format to represents parameters either for output or some controller.

Computer program can solve different types of problems by using genetic programming paradigm. The first step of genetic programming is the randomly creation of computer programs and population them. Darwinian reproduction operation and genetic crossover is applied iteratively in order to have better programs [7]. The resulted computer program from genetic programming paradigm represents a consequence of fitness which leads to the produce the needed program structure [7].

In designing of internet data centers, the energy efficiency and energy consumption must be considered as critical considerations. The design should increase the energy efficiency and reduce energy consumption while keeping the performance of data centers. In [8] the proposed approach consists of different scenarios to eliminate the power dispatch and saving power of multi-tier data centers with satisfying the specified user service level agreements.

Genetic algorithm was developed to schedule the application in [9]. It is implemented by special cost function, modified genetic operators and selection method of data transition weight.

In [10] the energy efficiency of a virtualized, heterogeneous server infrastructure increased by combining dynamic consolidation and workload forecasting. This method reduces the energy consumption of data centers. It also presented a cost model feeding into the Genetic Algorithms (GA) fitness function.

In [11] three different types of enhanced genetic algorithms procedures were presented. These genetic algorithms are interval genetic algorithm as well as hybrid interval genetic algorithm. The combining of these genetic algorithms optimized the design by increasing the energy efficiency and reducing the energy consumption [12].

Goldberg [13] shows that the simulated evolution of a solution through genetic algorithms in some cases is efficient more than random search or calculus based techniques.

BarnyCapehart [14] works on energy management as an efficient and effective use of energy maximizing profits and minimizing costs as well as to enhance competitive positions. The main focus in the energy management in buildings, manufacturing, IT facilities and data centers [15].

1.6 Thesis Organization

This thesis is divided into five chapters, in addition to the abstract and references. The first chapter demonstrates the problem and scope of the research in addition to demonstrating the motivation and illustrates the problem sides. In addition, it shows literature survey to show the scope of modern researches about the data center, genetic algorithms, and energy optimization. The contributions and objectives are illustrated carefully in this chapter.

The second chapter presents theoretical review that concerns about data centers and genetic algorithms. The data center structure, energy models, concerns, and constraints are illustrated in chapter two. In addition, the theory of genetic algorithm is demonstrated and illustrated in this chapter.

The methodology is presented in chapter three. The research method, problems, concerns, and design model are illustrated.

The results are shown in chapter four. The study sample and performance evaluation are defined. The measurement and recorded results are shown. In addition to that it presents a comparison with related works to demonstrate the importance and efficiency of this thesis research. Chapter five demonstrates the conclusions of this thesis research in addition to discussion about future scope of the research.

CHAPTER TWO

THEORETICAL REVIEW

2.1 Data Centers

The data centers are computer sever rooms that are the grown in the early years of the technology. Currently computer technology becomes difficult to be operated and maintained. It needs special environment to operate safely. One mainframe needs a great power handling and special cooling. On the other hand, security is an important issue. Basic guidelines for design and access CCR room were suggested in the form of data center.

As the rise of the microcomputer industry, the 1980s age comprises that computers started to be deployed anywhere, with small or no care about environment and operating requirements. With the existence of free source Linux compatible computer's OS in the age of 1990s, and in synchronization of that the MS-DOS finally developed multi-tasking capable Windows OS, personal computers started. The old computers were named servers as timesharing OS such as UNIX depend heavily on the client-server model to enable sharing of computer resources among multiple users. The cheap networking tools availability, with new working standards of network structured cabling, enables the hierarchical design to be possible and putting the servers in a special room inside the organization. The "data center" term was used and applied to specially designed computer areas or rooms, and it becomes gaining common definitions and recognition over time. Figure 2.1 shows a sample of data center with tray cable installations.

Fig.2.1:Example data center with cable tray installation

The data centers rise came with synchronization of internet and World Wide Web services. The organizations required fast Internet line connection working to adapt systems and to establish an internet existence. Installation of such tools was not possible for the small-scale organizations. Many organizations started to building large facilities, and named it Internet data centers (IDCs), it provides solutions range for operation and deployment of systems. Most recent practices and technologies had been designed to process the scale and the operations specifications and needs of the large companies. These functions migrated to private DC's, and were adopted because of its studied practical results. Cloud data centers (CDCs) are a special type of data centers that is specialized for cloud computing's. Today, all the previous types of data centers become integrated into the "data center" term.

As up-taking of cloud computing is increasing, government and business profit companies are scrutinizing DC's to a high level in availability, and security and standards. Accredited professional groups' documents like the Telecommunications Industry Association (TIA), determines the requirements for the design of data center. Measurements those are well-known for the availability of data center could be used to estimate business effects. Implementations are still developed in practice. Actually, DC's are very expensive to be implemented.

Functions of the IT are a critical of most company's practices. Business continuity is a main concern while in such, organizations depends on their information systems to keep its operation. When the system becomes unavailable, the operation of the company is possible to be completely inhibited. It is so important to ensure reliable IT running infrastructure for, to minimize any disruption possibility. Security of information is a concern too, so, the data center must ensure save working environment. DC's should keep standards for functionality and integrity assuring of its hosts and servers environment. In most data centers, that requirement is being achieved by redundancy of both power and fiber optic cables.

The standard DC needs tools, devices, and area gives information for DC design and0 communications networks required for construction and installing according to that area spaces. The requirements should be applied to DC's spaces that are holding high data processing or information tools and equipment's. While the tools and devices may be used to one or more objectives of the following:

- Providing a combination of the applications of data center
- Management and operation of telecommunication network carrier's
- Providing applications hosting for third party
- Providing applications that are based on data center directly to customers

Efficient data center functionality needs operation in the housing and the facility. Initial point is to design an initial facility. Modularity and standardization can result efficiencies in telecommunications DC's and design.

Standardization means that integrated rooms or even complex with engineering equipment and tools. On the other hand, modularity has many advantages including scalability, maintainability, and easier to be grown, even when planning prediction are not the optimal. The centralized systems usage needs accurate anticipation to avoid the so called over and under construction not meet ahead actual requirements.

The darkened or lights-out DC's is a special DC's that minimizes personnel direct access. The need for workers for entrance of the room of data center minimizes the need for lighting, so, it becomes dark room. Figure 2.2 shows an example of lights-out data center room. All components are controlled by automated tools and programs that are implemented to complete any unattended tasks. Also, it saves energy and minimizes the required staff, thus, minimizes costs. The lights-out data center minimizes the malicious attacks threat on infrastructure.

Fig.2.2: Lights-out data center

Studies of International Data Corporation (IDC) determines the age mean of a data centers to be approximately nine years old. While other researches considered that, the data centers those are older than seven years should be obsolete.

A report by DC organization of research Uptime Institute (UI), said that, more than 30 percent of the large organizations that is studied expect to exhaust its capacity of IT during the next 1.5 year.

Transformation of DC takes a step by one procedure through projects achieved in long time. That takes different time and steps from old DC's migration that got serial and soled depending on what kind of data center fixture is. The ideal projects through a data center transformation initiative comprise virtualization consolidation, security and automation.

Consolidation or standardization objective is to minimize the number of data centers that is required for large company. That process minimizes the amount of software, hardware, tools, platforms, and processes in data center. Companies replace old age data center equipment and tools with new replaces that ensures increased performance and capacity.

Automation of data center involves tasks such as compliance, patching, release management, configuration, and provisioning. Automating enables data centers to be operated more efficiently with less number of staff workers. Security on virtual systems in advanced data centers is integrated with available physical infrastructure security. The advanced data center security should consider data and user security, physical security, and network security.

2.2 Considerations in Data Centers

Recommendations of the design generally undergo modeling phase criteria. The optimal infrastructure of the technology is formulation and criteria planning are implemented, like capacities of critical power, overall power requirements of data center agreed on power utilization efficiency (PUE), kilowatts per cabinet, capacities of mechanical cooling, resiliency level for the facility, and raised floor space.

Mechanical design infrastructure addresses mechanical system related in keeping the interior environment and surrounding of the data center, like heating, cooling, ventilation, pressurization, dehumidification equipment and such. This design process stage should be planned to save space and costs, while keeping reliability and business are met with achieving power utilization efficiency. Current advanced architecture includes scaling and modularizing IT workloads, in addition to ensure capital spending on the construction of building optimization.

Electrical design infrastructure is concentrated on electrical configurations for different requirements of reliability and sizes of DC. It could include distribution, service planning, uninterruptable power supplies (UPS), and systems switching and bypass power supplies.

The described design considerations should deal with energy standards and better practices while in the same time achieving business goals. Commonly used electrical system is almost capable for scaling and is available for suitable voltage specs.

The data center physical environment should be controlled in rigid way. Cooling and heating is used to control the humidity and temperature [1]. The heat is naturally increase in data center because the electrical power that is used by the processing elements. So, the ambient temperature will increase, causing electronic equipment to be malfunctioned. Excessive humidity will also be followed by water existence that could be condensed on internal components of the electronic systems. When the atmosphere is dry, ancillary humidification systems is possible – if the humidity is too low – to add water vapor, that result in electrostatic discharge problems that could damage the electronic elements.

Modern DC's are using cooling economizers, where they handle outside air to make the data center at desired temperature. Data center needs to cool servers by the use external air in the winter. They are not using chillers, which consumes much energy.

Power backup consists of number of uninterruptible power supplies (UPS), bank of battery, and turbine generators.

2.3 Optimization in Data Centers

Energy strategic sustainability performance plan (SSPP) states that the organizations should implement aims for performance targets according to FDCCI (Federal Data Center Consolidation Initiative) objectives while working with IT technologies and services. These support of DC technologies include:

- Modern hotels get efficiently encouraging workers to enter into support an increasing workforce and edge devices equipment's (i.e. end user devices)
- Active power, air conditioning DC's measurement and assessments ensures the evaluation of DC's power performance and to identify consolidation projects

The goals that should be defined in data center planning for optimization focuses about minimizing DC's number in the same organization, enhance the cooling infrastructure in still part of DC's, and to enhance efficiency of the IT system in these DC's, with development of information architecture and technologies that support cloud services, mobile computing, and other transformational and innovative services that match user needs while minimizing energy costs. Such qualitative goals include:

- Optimization of the infrastructure
- Decrease the system cost for information and software
- Minimize the total cost of DC and its operation
- Optimization in the number of DC's
- Development of measurements in running of DC's

As the discussion above, the most important criterion that should be planned and managed in the data centers is the power and energy. The power has many aspects, which represented in terms of heating and ventilation air conditioning, processing capabilities and power, space, and such.

2.4 Green Buildings and Energy Management of Data Centers

Processor power and the power of processor cooling in addition to the cooling requirements depend on the performance per watt of a single processor. The ability to control this performance in relation to needs at both the circuit and chip levels is crucial to avoiding power dispatch and cooling issues.

The latest server's processor from both Intel and AMD has power state registers those are accessible to the programmer. With the suitable ROM firmware or OS interface, such hardware registers can be used to switch the processor between different performance states. This is useful in controlling dynamic scaling of the processor, so, dynamic voltage frequency scaling.

Changing the performance state – the processor frequency and voltage scaling – enables the processor to operate at different power scales. Tables 2.1 list the states exposed by the Intel Xeon processor which is 3.8 GHz and the Quad Core 2.66-GHz processors.

Table 2.1: example of processor state and operating voltage scaling [28]

Processor	Processor state	Core frequency	Core voltage
Xeon 3.8 GHz	Maximum performance	3.8 GHz	1.4 V
	Minimum power	3.2 GHz	1.2 V
Quad Core 2.66 GHz	Maximum performance	2.66 GHz	1.2 V
	Minimum power	2.0 GHz	1.0 V

The administrators can manage processor states by one of two basic methodologies [29]; via the operating system drivers, or via firmware in the computer BIOS ROM, as illustrated in Fig. 2.3.

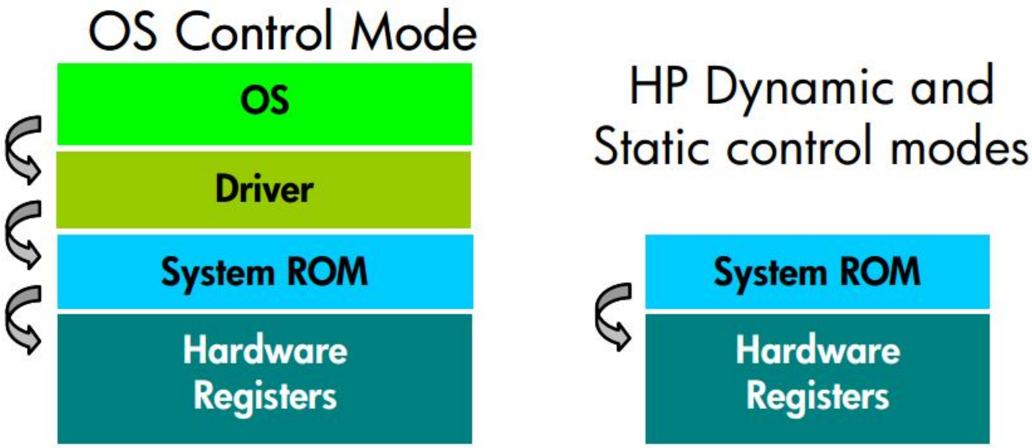


Fig.2.3: Methods of controlling the processor state [25]

Computer equipment manufacturers ideally provide power and heat dissipation load information in their product datasheets. They provide a rack / site installation and preparation utility to help customers in approximating the power and heat dissipation load per rack for facilities management. The site installation and preparation utility uses the power calculators for single platforms so that users can calculate the complete environmental effect of racks with changing configurations and loads [26].

An additional technique that is used in power management and performance control is the power regulation that is an OS independent power control feature of servers. Power regulation technology ensures dynamic or static changes in the CPU performance and thus, power states. It effectively performs automated policy based power control at the single server level. It could be enabled on individual system from the system ROM, or command line. A more robust control capability is possible by using it, while it allows administrators to access some processor power (i.e. voltage and frequency) features of multiple servers in a data center environment [27].

The power regulation and control of the processor monitors the CPU operation of single servers at high and low performance scales and report the percentage of time at each scale over 24 hours. This gives an indication of CPU power efficiency, and the result is accessible the system programmer. When the administrators use this tool the historical information is possible for multiple servers along with more extensive report options.

Previously, when data centers essentially housed large mainframe computers, energy and cooling design criteria were designated in the average watt per unit area that is W/ft2 or W/m2, or BTU/hr. Such design criteria were based on the hypothesis that the power and cooling needs were uniform over the entire data center. Nowadays, IT managers are populating data centers filled with heterogeneous variety of high density hardware devices as they try to extend the life of their used space, making it important to study power density distributions along the facility.

So, as the administrators and data center staff are well aware, facility power needs involve much more than server power needs. The percentage of total power dispatch that is used by cooling alone the today average data center exceeds in many cases 70% percent of the total power.

The front to rear airflow in the data center devices allows racks to be arranged in rows front to front and also back to back in order to form alternating hot and cold aisles. The devices draw in cold source air from the front and exhausts hot air out the rear of the rack into hot aisles. Many of the data centers use a downdraft airflow design in which air currents are heated and cooled in a continuous convection cycle flow. The downdraft airflow design needs a raised floor design that forms an air source plenum beneath the raised floor. The computer room air cooling unit draws in hot air from the top, cools it, and discharges the air into the source plenum beneath the floor. The cycle is illustrated in Fig. 2.4.

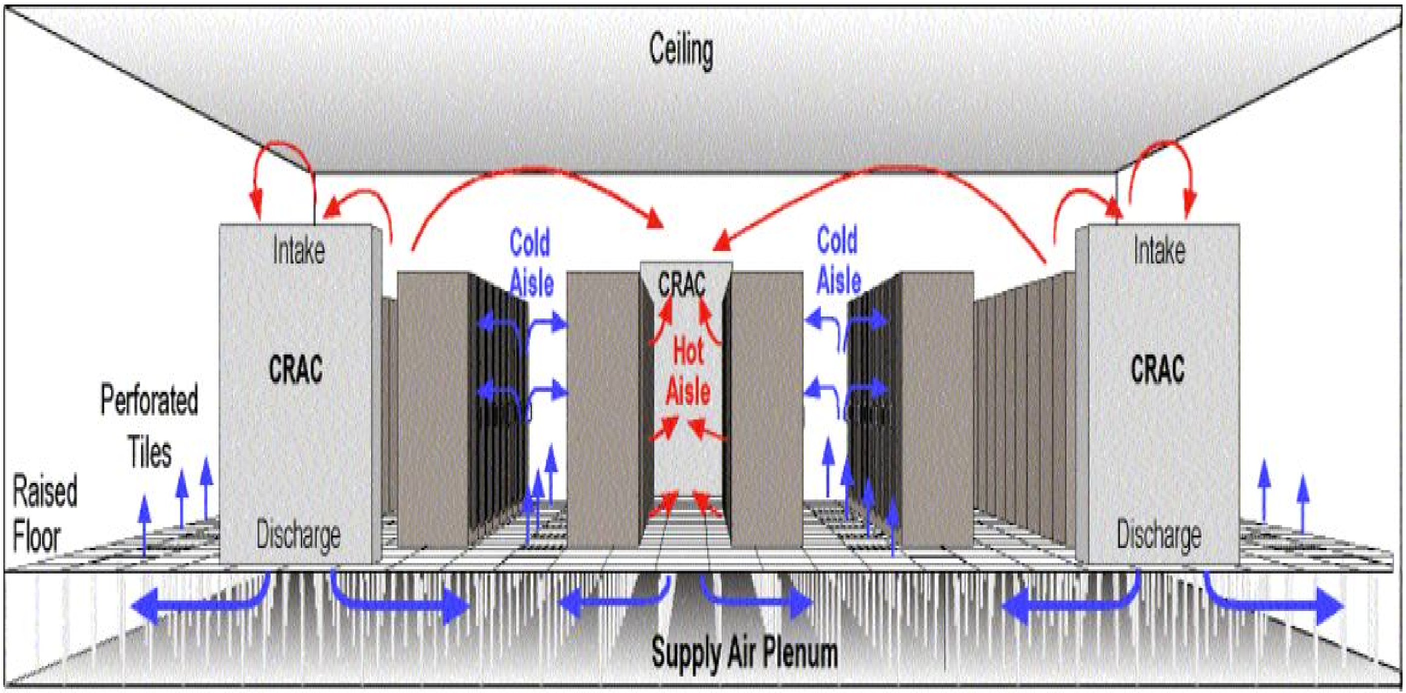


Fig.2.4: Air flow of cooling cycle in a typical data center [30]

To perform an optimal downdraft airflow design, warm exhaust air should be returned to the computer room air conditioning unit with minimal obstacle or redirection. Typically, the hot exhaust air will get up to the ceiling and get back to the computer room air condition (CRAC) unit intake. Actually, only the warm air near to the intake could be captured; the rest may mix with the source air. Mixing occurs if out air goes into the cold aisles, if cold air goes to the warm aisles, or if there is low ceiling height to allow for separation of the cold and hot air zones.

If the warm exhaust air mixed with source air, then, two possible cases could be done, those are:

- The temperature of the exhaust air goes down, thus, lowering the useable CRAC unit capacity.

- The temperature of the source increase, thus, causing warmer air to be re-circulated via the computer devices.

Thus, administrators require minimizing the mixing of warm and cold air by using advanced planning practices.

Heat loads is variable across a data center due to the heterogeneous mix of different hardware types and models, the added or removed of racks over time, and changing compute loads. The variation in heat load could be too complex to be estimated intuitively or to determine by adding cooling capacity. Different methods to control these heterogeneous mixes of hardware and densities tend to be changeable and proprietary.

Dynamic smart cooling manages the heat distribution throughout a data center using what so called computational fluid dynamics (CFD). Thermal estimation services use CFD modeling to help planners in designing the physical layer of the data center for optimal distribution of cooling resources and heat loads. These modeling services also could anticipate the changes in heat dissipation of each CRAC unit when the rack topology and devices heat load are varied.

Dynamic smart cooling presents higher level of automated facility control. It provides intelligent data centers that dynamically provision air conditioning supplies to meet the changing heat dissipation of computing, storage equipment, and networking. It also redistributes computing workloads depending on the most efficient use of cooling supplies within a data center or the global network of the data centers.

So, as discussed in the previous four sections of this chapter, the most critical issue that needs management and control in the data center is the energy or power consumption. This energy is mostly consumed by the air conditioning units which work to cool the computing devices that heats up depending on their workload.

This thesis offers a model for energy optimization for such data centers that is described shortly in the first chapter, and which will be described carefully in the next chapter (third chapter). This model is based on forecasting and optimization that is structured using genetic algorithm artificially controlled computations. The genetic algorithms theory overview will be described in the next section.

2.5 Genetic Algorithm

The (GA) genetic algorithm is a computational method in computer science which considered being a part of the artificial intelligence techniques. GA is a search heuristic computation simulates the natural selection process. It is used to get solutions for optimization issues. GA belongs to the big class of the well-known algorithms that produces solutions for optimization issues the meaning of simulating the natural evolution, like selection, inheritance, crossover, and mutation.

Genetic algorithms application was extended to bioinformatics, phylogenetic, engineering, in addition to many another fields. An ideal genetic algorithm needs representation of genetic solution domain, and fitness function that assess the solution domain. The next stage is generating second generation solutions population from those selected via a combination of genetic operators; mutation and crossover.

For each new generated solution to be produced, a pair of parent solutions is being selected for breeding from the group that is selected previously. A child solution producing using the previous methods of crossover and mutation, new solution is generated which ideally shares many of the characteristics of its parents. New parents are used for each new generated child, and the process keep continuing until a new population of solutions of suitable size is generated. Although reproduction methodologies that are based on the use of two parents are biology inspired. A lot of researches recommend the use of more than two parents to generate higher quality chromosomes.

Such processes finally result in the next generation population of chromosomes that have difference from the initial generation. In fact, the average fitness will be increased by the procedure of the population, since only the better organisms of the first generation are used for breeding, along with a small proportion of minimal fit solutions. These minimal fit solutions provide genetic diversity within the genetic group of the parents and therefore improve the genetic diversity of the subsequent generation of children.

Fig. 2.5 shows the general flow diagram of the genetic algorithm process. It starts by creating initial design population, and goes through objects evaluation, selection and pre-producing, and replacing the old design with new one. And it repeats replacing each generation by the next generation until a satisfactory solution is gotten.

The flow chart that is illustrated in Fig. 2.5 is general and describes the basic operations of the genetic algorithm optimization processes. The detailed design and parameter controls are the task of the system and program design. Its task is to select the best parameters and tuning variable to perform the optimal design that could generate the optimal solution of the targeted systematic problem.

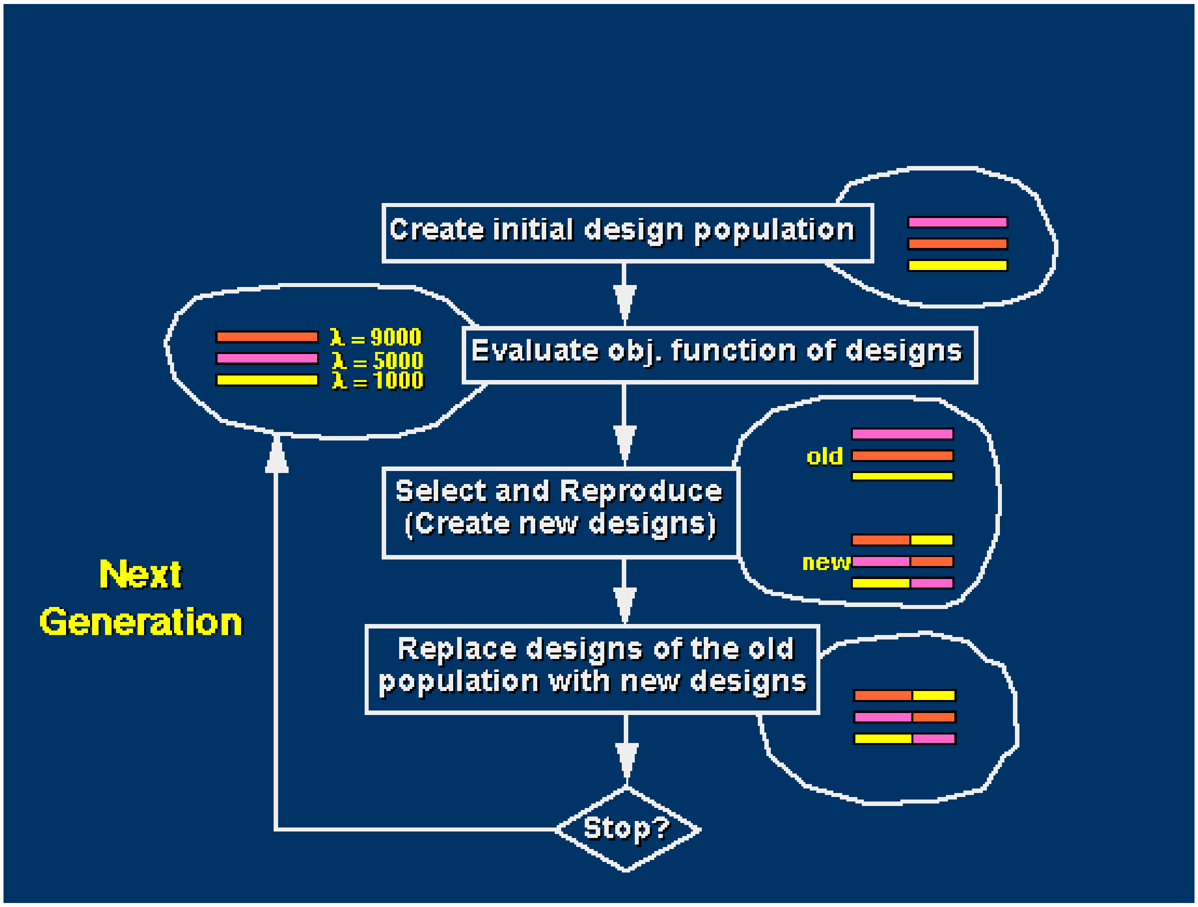


Fig.2.5: Genetic algorithm flow chart [7]

The thought is divided over the importance of mutation versus crossover. There are several references that justify the importance of mutation based search. So, it will be more used in this thesis. In addition, the crossover and mutation are well known as the main genetic operators. It is available to use other operators like colonization-extinction, regrouping, or migration in genetic algorithms.

The generation process should be repeated to reach to the termination condition. There are different termination conditions, the most common terminating conditions and most useful in different applications are the following:

- A solution is achieved that gets minimum criteria
- A predefined number of generations reached
- Allocated budget is being reached, that is could be a computational time or even money
- A high-ranking solution fitness function is being reached or has reached a plateau like that successive iterations no more produce better results
- Manual inspection of the genetic algorithm overall process

In fact, combinations of mentioned above termination conditions is normally selected for a specific genetic algorithm design, order to achieve better and more rigid design customization and performance.

Genetic algorithms are in fact simple to be implemented, but the problem is that their behavior is difficult to understand. Specifically, it is not easy to understand why these algorithms frequently succeed in generating solutions of high fitness if it applied to specific problem.

The two main hypothesis of the genetic algorithm are the following:

- A description of a heuristic that achieve adaptation by formulating and recombining building blocks, that is low order, and low defining length schemes with above average fitness.

- A genetic algorithm achieves adaptation by implicitly and effectively implementing the heuristic.

Actually, the heuristics is defined as short, low order, and highly fit scheme that are sampled, recombined using crossed over, and resampled to represent strings of higher fitness. By working with those particular schemes which represent the building blocks of the genetic algorithm, it is possible to reduce the complexity of the problem. Instead of building high-performance strings by experiencing every conceivable combination, construction of better strings from the better partial solutions of past samplings is being performed. Fig. 2.6 shows a random sample of chromosomes in genetic algorithm iterations.

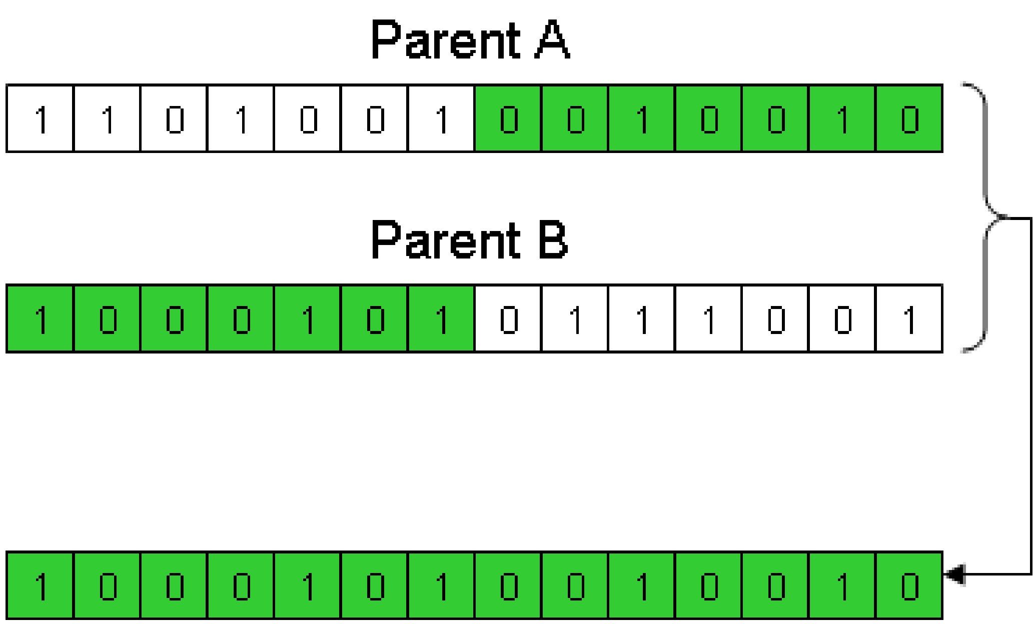


Fig.2.6: Chromosomes example in genetic algorithms [5]

Fig. 2.7 shows the gene distribution over the selected chromosome of the genetic algorithm.

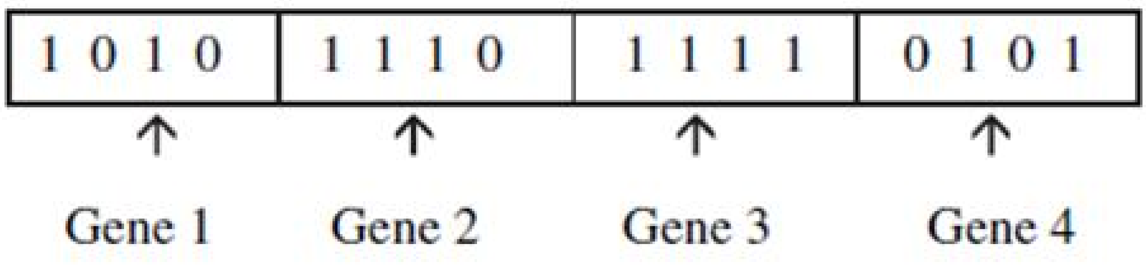


Fig.2.7: Chromosome with gene illustration [5]

The purpose of the selection operator is to ensure better chances to replicate for the better chromosomes of a population. The selection is done by taking into account the fitness of the chromosomes. The most commonly used selection methodologies are Monte Carlo and Tournament.

Selecting the number and position of crossover points for the crossover operator is based on the system topology. If such points are chosen in an inadequate mode bas chromosomes will be obtained. That is un-connected systems with isolated points or connected systems that contain loops. To minimize the number of such cases, the number of cut points should be equal to the value of CN − 1. Where CN is the cyclamate number which is the count of essential loops, according to the attached graph, and it is being calculated as equation 2.2.

$$CN = m - n + p \tag{2.2}$$

Where, m: number of branches

n: the number of nodes

p: the number of connected components

Genetic algorithm has wide application scope in computer era and artificial intelligence. The main use of genetic algorithms is for optimization problems. Also, it could be used in forecasting and intelligent problem solving techniques and applications.

The genetic algorithms in addition to the fuzzy logic and neural networks is intuitively named intelligent controllers because of that, they all are iterative and non-mathematical modeling computation techniques. Also, they don't matter about the number of inputs or outputs of the system; it means that, they can easily handle single input, single output, multi input, or multi output systems. In addition to that, they don't face big deal about the non-linearity of the system, because of that there are no analytical mathematical modeling that make the problem representation much complex than the real.

Adoption of genetic algorithms to solve multiple objectives optimization and complex problems requires dealing with the twin issues of searching big and complex solution domains and dealing with potentially conflicting multiple objectives. Chosen of a solution from a set of available solution on the basis of many criteria is considered as non-easy problem. According to this complexity, most of the researchers minimize the problem to a mono-criterion one. Mathematical programming methods and the popular weighted-sum approach have been implemented as method for such problems solving. On the other hand, the meta-heuristic was recognized as possible method for exploiting evolutionary algorithm to handle multiple objectives problems.

Conventional genetic algorithms use fitness based selection, and so, it requires scalar fitness information. Thus, the objectives are usually artificially combined into a scalar function. This thesis uses multi objective technique in genetic algorithm for forecasting and optimization of energy in data centers. This technique will be illustrated and discussed carefully in the next chapter (i.e. chapter 3).

CHAPTER THREE

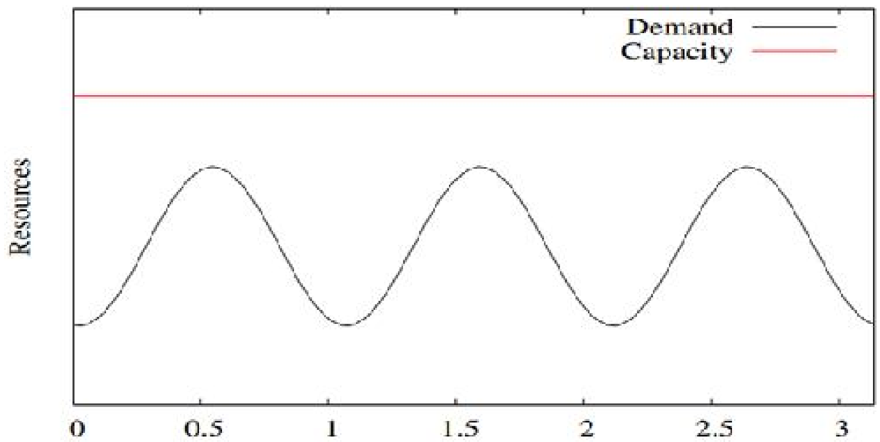
METHODOLOGY

3.1 Research Methodology

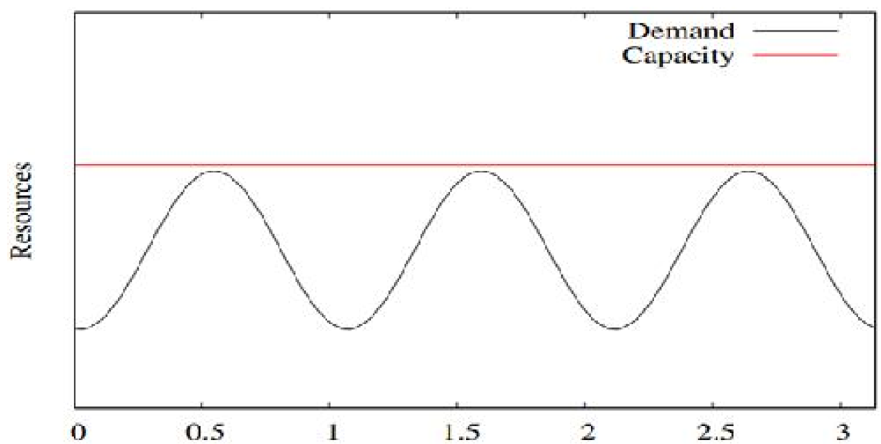
This research was aimed to design a genetic algorithm based data center energy optimization. The problem that is illustrated in figure 1.2 is the major concern of this research. The data center energy is consumed by four categories those are:

- Processing power
- Air conditioning
- Lighting
- Access controls and extras

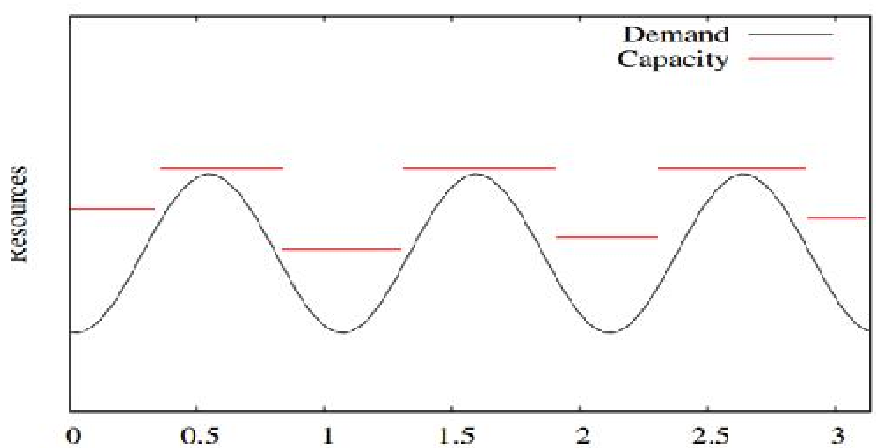
Most of the energy is being consumed by air conditioning. The losses come from the fact that, the operators of the data center didn't know the actual load of the system and how much it need to operate optimally, thus, the power capacity that offered to the data center in most cases becomes much higher than the actual load. So, rest of the capacity (i.e. the un-used part) will be lost.



(a)



(b)



(c)

Fig.3.1: Data center power capacity control schemes

Fig. 3.1 shows different schemes of energy capacity control of the data center. Part (a) of the figure shows that, a very high capacity was dedicated for the data center where it will never consume it in any case. Most of the energy is lost and this is the basic scheme without any optimizations. Part (b) shows a very few optimization, where the capacity of the overall data center is considered to be the maximum energy that could be consumed over time. Even though, the losses in this scheme is less than the previous one, but it also has a large amount of losses because of that, when the payload is less than the max, then, the remaining power will be lost.

The third scheme (c) saves more energy by stepping the power capacity. It's very similar to the previous one in (b) but the difference is that, it divides the time to small periods and calculates the maximum power capacity for each period individually. So, the max will change in stepping form depending on the time series payload behavior. But also, in this case, there is no optimization and a lot of energy is being lost.

This thesis is presenting an optimization model that changes the available power capacity continuously depending on the payload. The power that is could be consumed is being forecasted and thus, the capacity will be changed depending on the forecasted power. The power scheme that is suggested by this thesis is shown in figure 3.2

The methodology that followed in this research was based on five steps, those are the following:

- Problem formulation and system requirement
- System modeling and design issues
- Implementation

- Testing and performance measurements

- Modifications

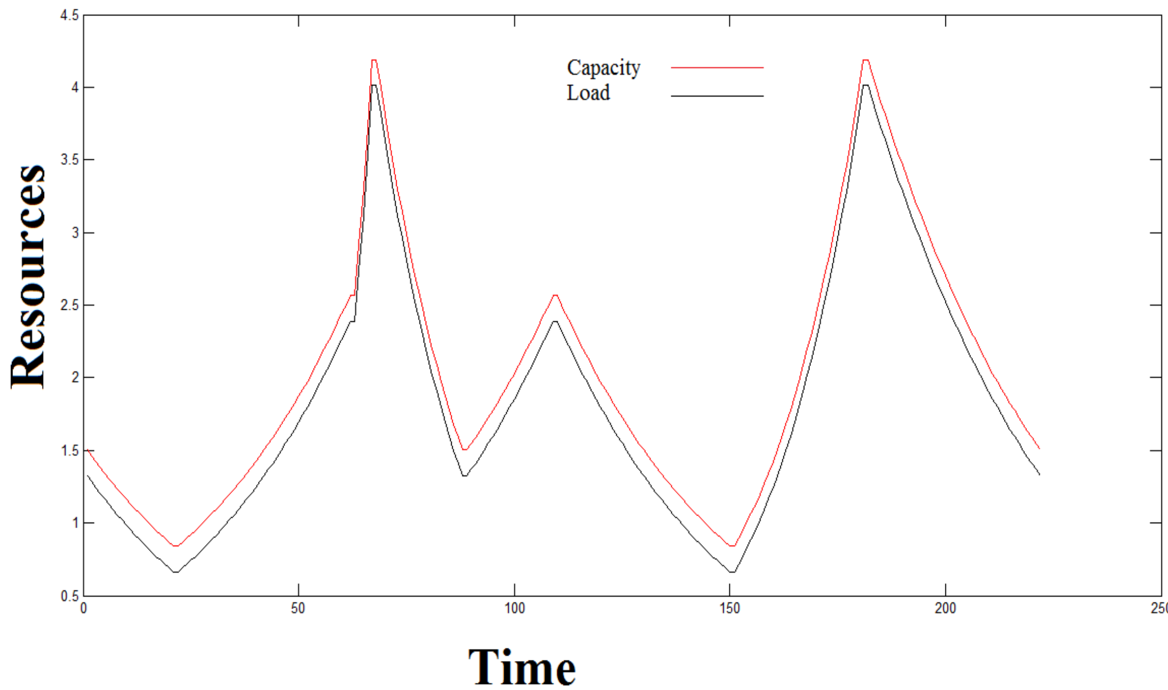


Fig.3.2: Suggested optimization scheme of the data center power capacity (where time in hours and resources as a payload of the data center)

The system requirement was that, designing a system to formulate the power consumption of data center cooling in terms of processing payload of the computing and storage equipment. Then, designing a forecasting algorithm to predict the computational payload power and thus, optimizing the power capacity that is expected to be consumed by the forecasted computing payload of the data center. Those two functions – forecasting and optimization – should be based on genetic algorithm intelligent computing.

The implementation stage was done using MATLAB the language of engineering and technical computations. It was because of that, MATLAB is a perfect tool to develop algorithms and especially for engineers and scientists. In addition to that, it has special library to deal with such technologies like genetic algorithm. For example, the global optimization toolbox contains a library functions for different optimization methodologies, including the genetic algorithms.

The results and performance evaluation was gotten in terms of power consumption and the capacity of energy that is assigned to the data center at specific time period. The genetic algorithm characteristics also studied and considered in evaluation.

Finally, the modifications were done many times, where time the results are not satisfactory. This makes continuous improvement of the presented system until getting powerful and meaningful results.

3.2 Data Center Energy Optimization

The processors' generated heat is a function of the processor's power, where the processor power itself is a function of the working voltage and frequency. In the technology of dynamic voltage dynamic frequency scaling (DVDFS), the frequency of the processor changes depending on the processing payload, and it directly proportional to it. The voltage and frequency ration should be constant, where in such; the relationship is between the frequency and the square of the voltage. The voltage and frequency are the main factors those are affecting the power dissipation and consumption of the processors and storage elements. Another factor is the capacitance of the system. Equation 3.1 shows illustrates how the power of the system could be calculated from the payload parameters.

$$P = C.\ V^2.\ F \qquad (3.1)$$

Where, P: is the total processing power

C: is the total system capacitance

V: is the working voltage

F: is the operating frequency

The generated heat is a scaling function of the processing power in Joule [J]. It could be calculated by multiplying the processing power "P" by a factor "K" as shown in equation 3.2.

$$Pheat = Px\ K \qquad (3.2)$$

In fact, to make more realistic mathematical model for the heat, a historical data that was measured for the electrical power and the heat power was considered and plot between them was drawn and shown in figure 3.3. The figure shows that, the relationship is linear.

By regression or curve fitting of the relationship that is described in the figure 3.3, a mathematical equation could gotten as shown in equation 3.3 .

$$Pheat = -0.02*Pelec^3 + 3.6*Pelec^2 - 283.9*Pelec + 8510 \qquad (3.3)$$

Where the Pheat is the thermal power and the Pelec is the electrical power those are generated at specific processing payload. The result of curve fitting is shown in figure 3.4. This linear model is the final model that will be used by substituting the Pelec by its value as in equation 3.1 to be used by the genetic algorithm optimization step.

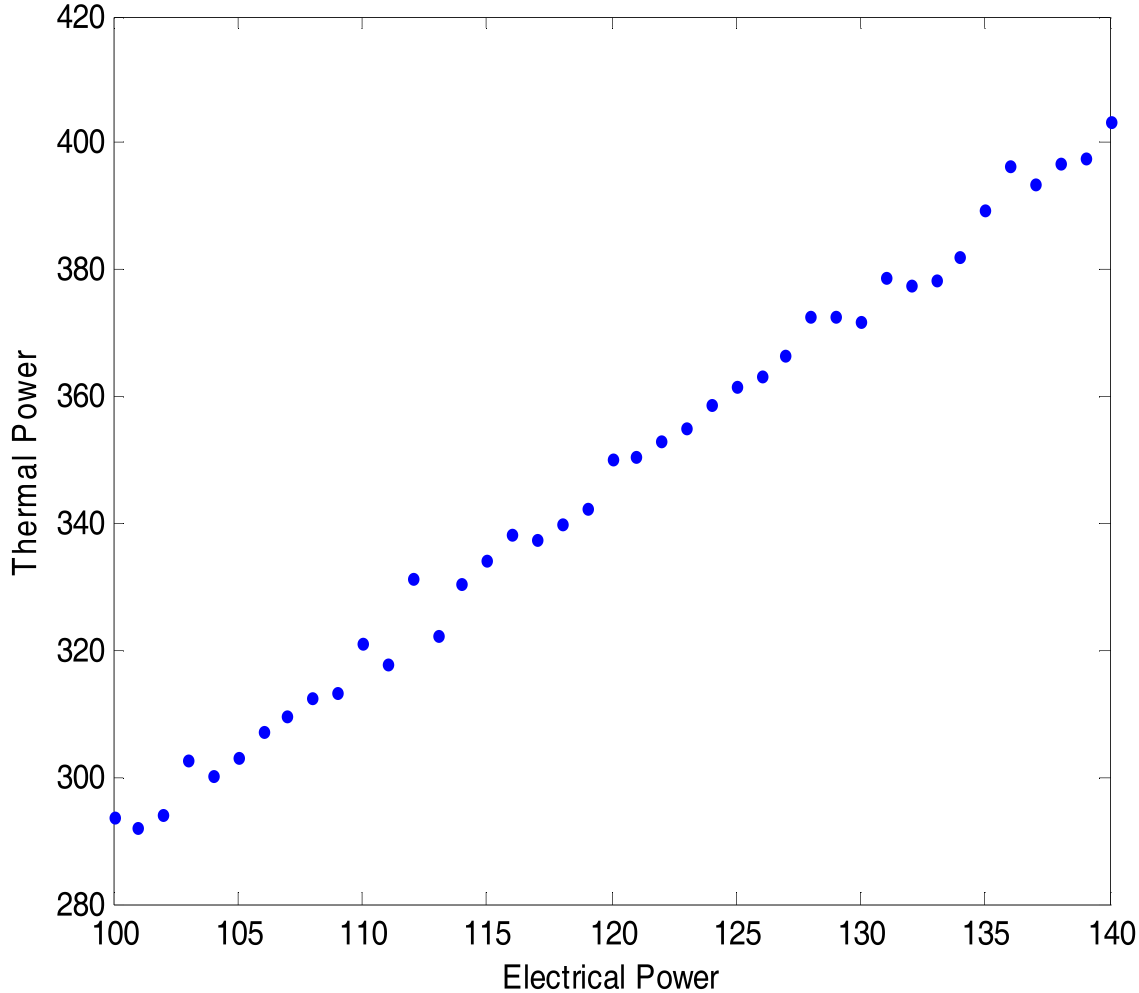


Fig.3.3: Measurement relationship between electrical processing power and heat power of a sample data center, the units are in mega watt

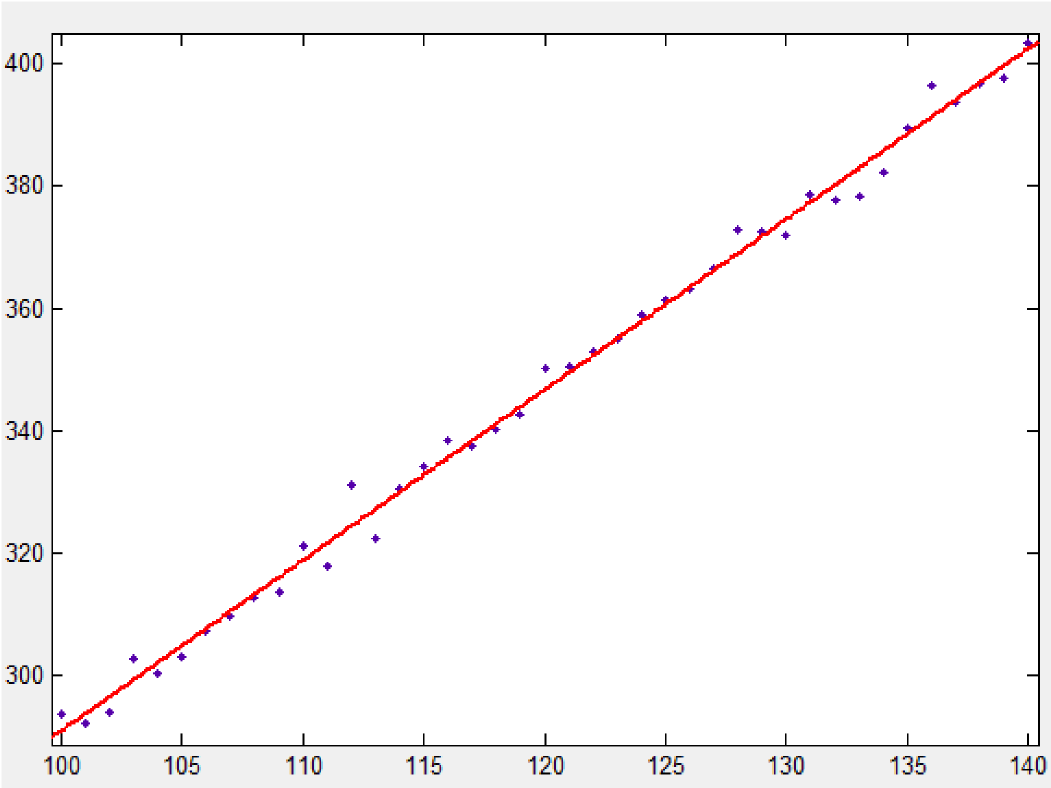


Fig.3.4: Curve fitting result of the data center power (where x axis is the time in hours unit and y axis is the power in watt unit)

3.3 Presented Genetic Algorithm

As stated in chapter 2; the genetic algorithm is a computational methodology to solve constrained or unconstrained problems of optimization. The genetic algorithms are based on natural selection that is derived from biological evolution theories.

This thesis uses genetic algorithm to optimize the energy consumption in computers data center by forecasting the required payload and minimize the energy capacity that is required to operate the data center, especially the air cooling power.

The genetic algorithm is developed by the use of MATLAB Global Optimization Toolbox that contains library function to develop the genetic algorithms and other optimization techniques. The model of the data center heating power versus the computational payload – that is described in the previous section – was used as the fitness function. A payload variant model was created to test and verify the system design. Two variables were structuring the fitness function model; those are the frequency and the voltage of the processing unit.

A crossover scattered function was selected to perform the crossover. The genetic was designed without parallel processing of the chromosomes, and with forward migration direction. The main stopping criteria is considered to be the average change in the fitness function and it should be greater than the tolerance function to keep the iteration going on. When it goes less than the tolerance function, then, the genetic algorithm iterations will stop and the process is being terminated.

The population size was selected to be twenty samples, with approximately 900 generations of possible processing. The options those are selected for the genetic algorithm controls the accuracy, precision, and speed of optimization process.

The result of optimization was great and in will be discussed and demonstrated in the next chapter (Results). As normal engineering research, many problems and issues were faced the research process. The constraints, problems, and issues that this research was handled will be presented and discussed in the next section.

3.4 Problems and Issues

Many problems and constraints were faced during this research. Some of them were solved easily, some of them solved hardly, and some of them remains unsolved and oriented to be solved in the future works.

The mathematical modeling of the data center environment is the first issues that was faced and solved in this thesis. The forecasting of the next load represents a big issue that is solved hardly in this research. In fact, variation of load with respect to un-systematic and emergency conditions is un-modeled situations and behavior of the data center that this research is not intended to deal with it. So, this problem was oriented for future further enhancement on this research.

First of all, the most energy that is consumed by the data center (i.e. that is could be 70% or may be grater, as mentioned in chapter 2) is the HVAC (Heating and Ventilation Air Conditioning) power consumption. The cooling that is required for cooling down the computing and computer storage equipment's is a function of the processing power. Thus, it was not easy to deal with such problem in order to make a model of electrical power that depends on the computational payload of the processing element.

The solution of this problem was to get an approximation model that is used in some researches and recommended case studies. The used model as shown previously in this chapter is approximation and not so exact. Thus, even though this thesis research optimizes the data center and approximates the minimal losses in energy, but these minimal losses itself is an approximation that gets behind the optimal value [16].

The relationship between the computational power and the thermal power is not strait forward, and it has uncertainty. This comes from the fact that, the data center contains different types of processors, different types of storage elements, different types of built-in computer fans for heat exchange and thus different types of heat sinking fixtures, and also different methodologies in components alignments [17].

The variety in components and designs those are used in the data center [18], makes anticipating an accurate mathematical model for the heat as a function of processing power is not easy. This thesis overcame this problem by using conventional processing power equations those are being well-known in the researches those are intended to work in dynamic voltage dynamic frequency scaling (DVDFS) of the computer processors [19]. And so, the electrical power that is needed for cooling was estimated as a function of electrical power scaling that is used for processing [20].

It is big deal in forecasting the processing power or processing payload that the system is going to process it in the next time sample step. In fact, the genetic algorithm is not the suitable computational artificial intelligence technique for forecasting. In this thesis, a hybrid model for forecasting and optimization was built using the same genetic algorithm model. But in fact, the forecasting itself needs separate powerful forecaster to perform it high accuracy [21].

The reason that makes the forecasting in this thesis to be performed accurately is the simplicity of the payload model, where the dynamic voltage scaling is always linear, and also the frequency dynamic scaling is linear. But when dealing with processing payload with high uncertainty, it’s a big deal to forecast such time series using genetic algorithm [22].

As mentioned, this problem was avoided by dealing with simple linear system [23], where the complexity is leaved to future a head modifications and improvements. In fact, the better forecaster for the data center behavior is the fuzzy logic forecasters [24]. Where in such the system could be defined adaptively to anticipate many criteria and parameters that indicates the system change. In addition to the self-adoption of the fuzzy forecaster it by itself to avoid any accumulated errors in the entire system.

CHAPTER FOUR

RESULTS

4.1 Performance Evaluation

Hence, this thesis aims to optimize data center energy consumption, it focuses on two points. The first point is to forecast the required power of the data center, and the second point is to determine the optimized power capacity that should be supplied to the data center at specific time. The computational load of the data center is varied and supposed to be changed linearly over time, rising up and falling down.

Figure 4.1 illustrated the curve of computational power variations in the study sample. Different slopes of variations were assumed where the peak also changes over time. This model is a comprehensive model that includes different behaviors with respect to time. Even though, the linearity makes the forecasting possible (i.e. non-linearity makes the forecasting to be very hard issue in every system and specially when dealing with such proposed system) but also contains different variations those make the system study interesting and considering different styles of power consumption and operational dispatch.

The system's power dispatch that is shown in Fig. 4.1 requires an electrical power capacity that fits to it solving the problems those are shown previously in Fig. 3.1. Two criteria should be achieved; good estimation of a head power and the optimization of minimal power capacity that could be fit for each time sample.

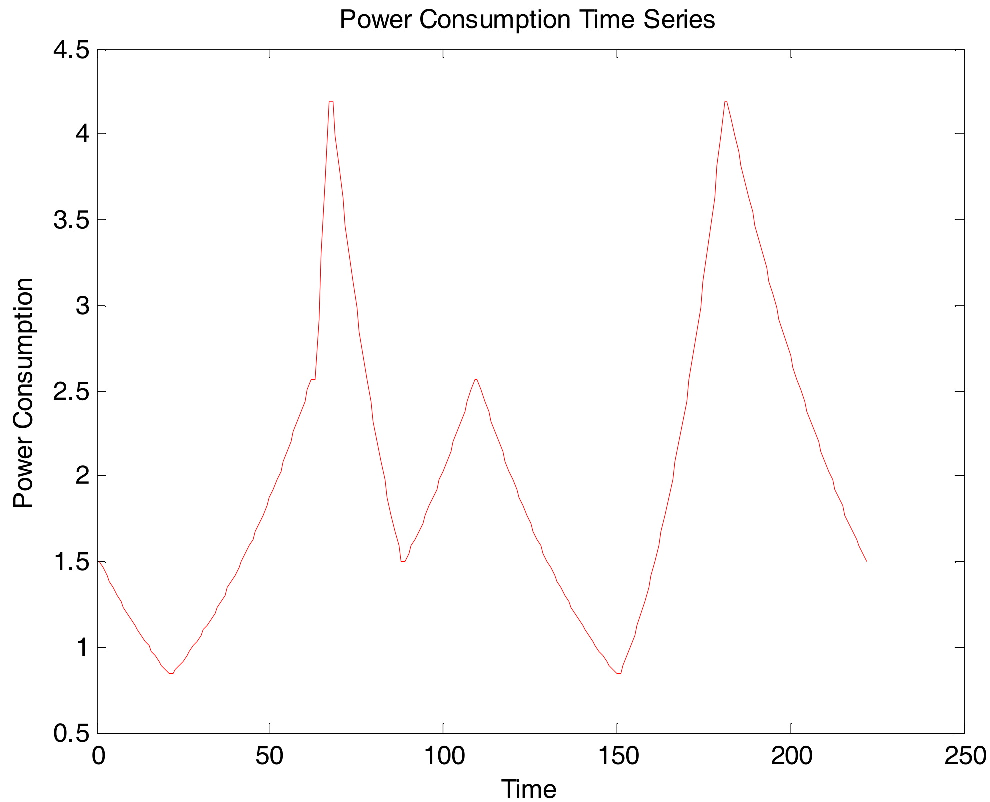


Fig.4.1: Computational power variations in the assumed data center of the study sample in this thesis; power is in watt, and time is in hours

In order to evaluate the performance of the genetic algorithm and the contributed systems, the evaluation points are the following:

- Forecasting of the power consumption with total error that is less than 10% as average, that should be calculated as illustrated in equation 4.1.
- Forecasting error of the power consumption with error that is less than 5% for each forecasting sample step.

- The optimization of minimal power capacity is determined by the genetic algorithm with its intelligent computational process, but the measurement of that accuracy is measured by another statistical factors. This thesis considered the quantitative peak signal to noise ratio (PSNR) to determine the efficiency of optimization. If the PSNR is less than 40, then, the optimization is not so optimal. In order to make it optimal, the PSNR should be greater than 40. The PSNR is being calculated as shown in equation 4.2

- The fourth performance evaluation is that, the energy capacity curve should be very similar to the power consumption curve that is shown in figure 4.1. Where this ensures that no capacity offered to the system that will be lost. In fact, the peak signal to noise ration here in this thesis is not used to measure the signal distortion, but it's used to measure the similarity of two matrices, where both matrices contain data similar to each other with slight change.

The forecasting error should be calculated using equation 4.1.

$$\boldsymbol{Ferror} = \sum \frac{\boldsymbol{pEst(i)} - \boldsymbol{pAct(i)}}{\boldsymbol{pAct(i)}} \qquad (4.1)$$

Where, pEst (i): is the forecasted power for the sample point index at specific time sample.

pAct: is the actual power

i: is the sample point index at specific time sample

Ferror: is the total forecasting error in % percentage

The peak signal to noise ratio is being calculated using equation 4.2, and since it requires the mean square error (mse) value, then the mean square error is being calculated using equation 4.3 as illustrated bellow.

$$PSNR = 10\log_{10}\left(\frac{R^2}{MSE}\right) \quad (4.2)$$

Where, R: is the data sample point

MSE: is the mean square error that is being calculated as equation 4.3

$$MSE = \frac{\sum_{M,N} [I_1(m,n) - I_2(m,n)]^2}{M * N} \quad (4.3)$$

Where, I1, I2: are the two array those require to calculated the error between them

m, n: are the indexes of the I1, and I2 array

M, N: the size of the two arrays I1, and I2

In addition, to support the performance evaluation with real virtualization results, graph will be illustrated for the detailed results bellow.

4.2 Measurement and Records

This section illustrated the practical results and records in addition to analysis and discussion on it. The performance evaluation that is described in the previous section is being applied here in this section.

The genetic algorithm was designed with two variables, those are, electrical power and cooling power. The population size was selected to be 20 elements. Thus, Figure 4.2 shows the iterations of the fitness function with respect to the fitness value. It's clear from the figure that, the generations (iterations) were reached the stopping goal after 51 generation chromosome.

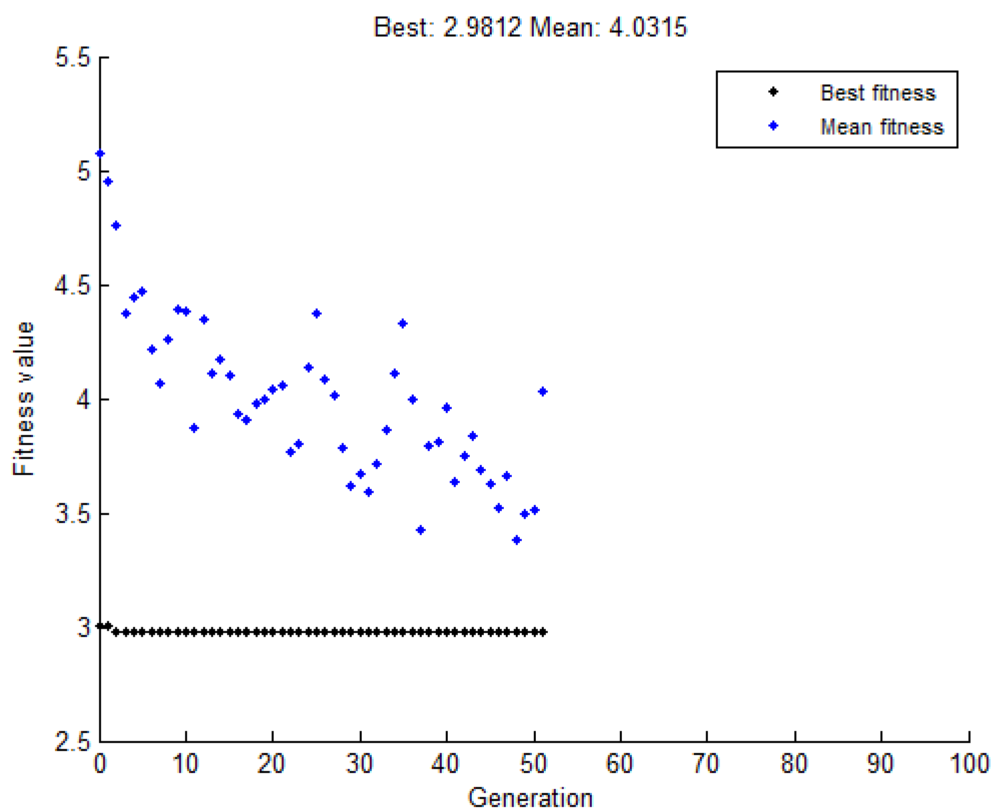


Fig.4.2: Genetic algorithm fitness function behavior with generations (iterations)

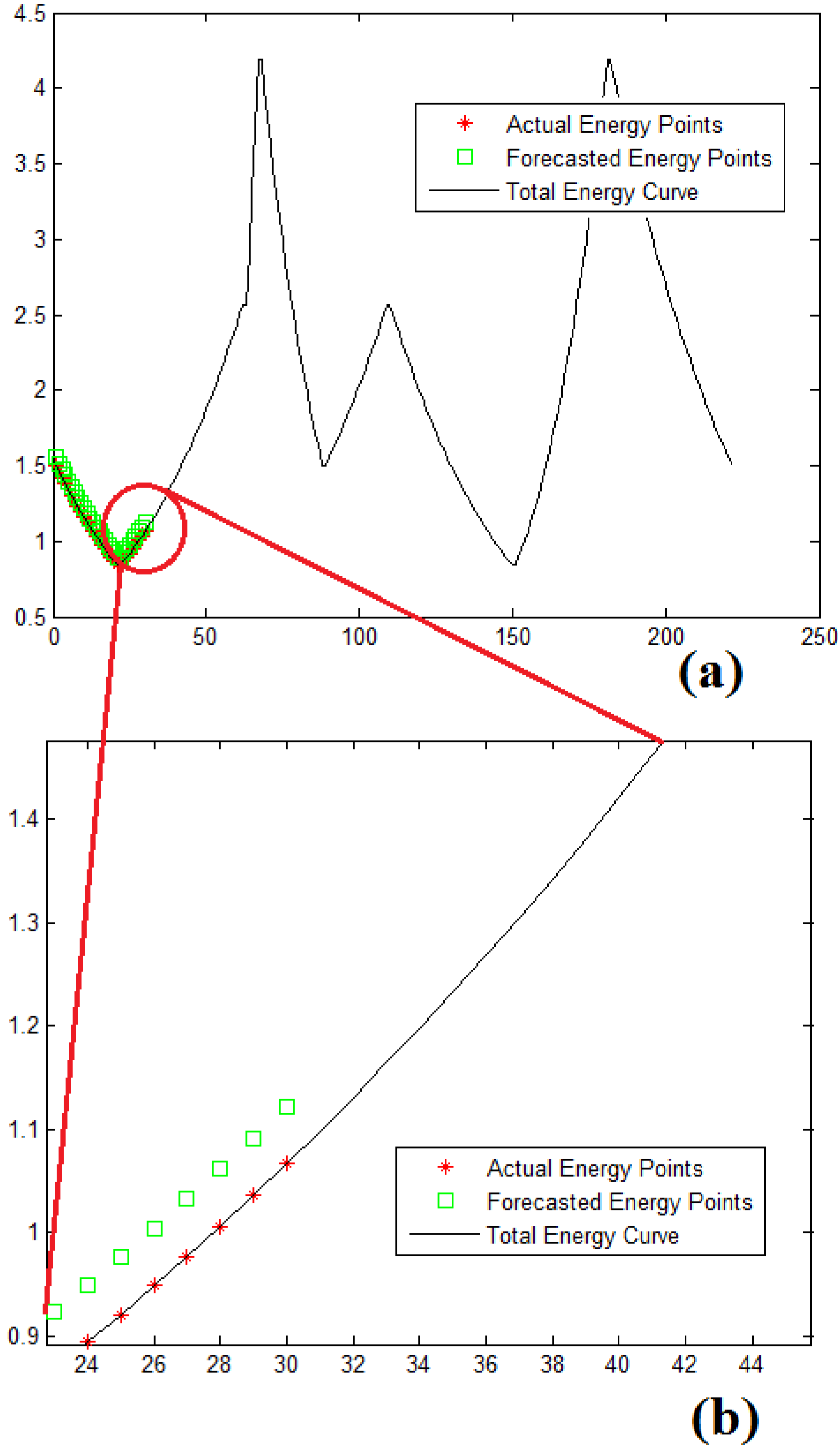


Fig.4.3: Forecasting Results

The behavior of the genetic algorithm shows that, how it fast getting the best goal value. In fact, this indicates also that, the use of genetic algorithm can be very suitable for such systems ad such optimization applications.

Fig. 4.3 illustrates the forecasting result. The Fig. 4.3.a shows the total curve with forecasting. The black curve represents the total power behavior curve, which is virtual curve, where the red stars represents the actual power capacity value, and the green squares represents the forecasted power capacity value. The power capacity here is measured in megawatts.

It is illustrated from the Fig. 4.3 above that the forecasting is accurate with some error value. The error of the forecasting is shown in Table 4.1 as numerical value. The reason that makes the forecasting accurate as that the power change in the data center is linear, thus, it is behavior is could be simple predictable and can be fitted to almost time stamp of the running time that could be considered to be over very long period of time.

Also, Fig. 4.3 shows a drift between the actual record of the power and the forecasted one. This is the forecasted error. But in fact, this error is very small and in most cases his limit considered to be zero. The main criterion that judge on the estimation result is the error value as numerical value, which is illustrated in Table 4.1 bellow.

Fig. 4.4 shows the total power optimization scheme in the study sample virtual data center over long period of time. It shows that, the capacity offered by the resources is slightly higher than the actual demand over long period of time. This means that, the amount of lost energy is very low and could be neglected.

Table 4.1 Performance evaluation

Criterion	**Total forecasting error**	**PSNR**
Power forecasting	0.11%	
Resources optimization		77.7710

This optimization scheme is unique and improves the power utilization in the data center. If we talk about air cooling of the data center, it's possible to operate the conditioning units in the required capacity for cooling only and there is no need to operate the air conditioning and heat compressors and ventilation fans without any efficiency.

Table 4.1 shows the quantitative value of performance evaluation of the presented thesis. As mentioned in the previous section, the forecasting accuracy should be measured as the total forecasting error that should be calculated as equation 4.1, where the resources capacity optimization is being calculated in terms of peak signal to noise ratio (PSNR).

The peak signal to noise ratio represents the difference between the demands and the offered resources, and as illustrated in Table 4.1 the PSNR is very high value. The target value in assumptions was to be greater than 40, but what gotten is greater than 70 which are very good. In fact, this value means that, the optimization is very accurate and the resources capacity is just higher than the required capacity for real application.

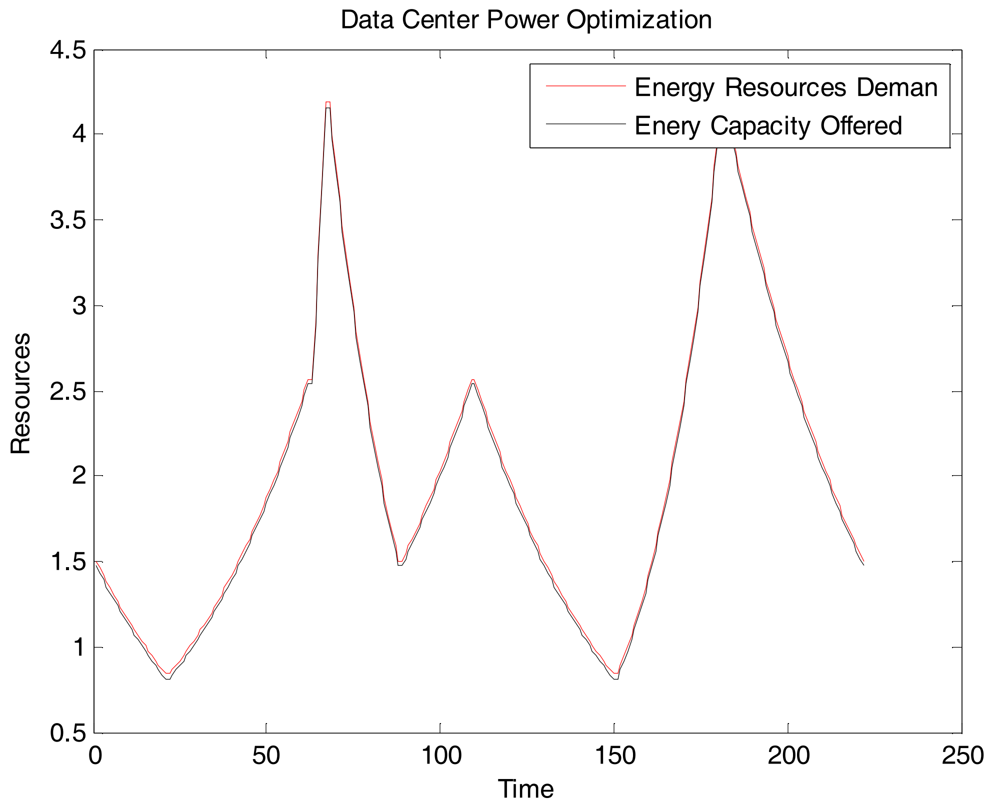


Fig.4.4: Optimize resources versus resources demand (where time in hours and resources as a payload of the data center)

On the other hand, Table 4.1 indicates that, the forecasting accuracy is near to 0% (i.e. 0.11% is most likely zero). This value is great value and thus, the forecasting is very accurate. This numerical result supports the results shown in Fig. 4.3 previously in this section. The graphical result in Fig. 4.3 shows that, the forecasting is very accurate and very near to the actual data value of the system, while the numerical value indicates that, the quantitative error in the processing is almost neglected

The target value of the power forecasting error was stated to be less than 5% of the total forecasting error. But the actual one that has been gotten is very less than that, achieving a high rate of performance and optimization.

The optimization of data center is not so simple problem. But this research achieved high accuracy results. This comes from the fact that, the genetic algorithms analysis is very accurate and can achieve optimization result with the accuracy those shown and discussed previously in this chapter and this section.

CHAPTERFIVE

CONCLUSION AND FUTURE REMARKS

5.1 Conclusion

Data centers consumes large amount of energy, where most of these energy is being consumed for cooling purposes. The heat is almost generated by the electronic components heat dissipation, in addition to the surrounding temperature itself.

The energy that is consumed by the data center for different resources, and especially for cooling, has a trend between resources available and demands. The capacity of energy that the resources offer to the data centers is either consumed by the data center equipment's or either lot. Thus, any offered power resources capacity that isn't used it will be lot. This loss is very high value and thus very costly.

This thesis aims to optimize the energy of the data center in order to save as much as possible of the lost energy part of the offered resources power capacity in each data center. This task is possible by changing the way of planning of power resources and by intelligent idea that can change the way of resources availability in the data center.

A genetic algorithm was designed to optimize the data center energy consumption in order to save the possible part of the lost energy over time. The optimization was designed in the idea that limits the energy resources to the value that is near to the needed value. This is known by estimating how much the system will draws power in the next time and thus, offering the required resources capacity.

Once this research is designed, implemented, tested and completed, in addition to that the results were measured, recorded and tested. The conclusion points those are gotten over the total period of the research are summarized as the following:

- The most losses of energy in the data center is being done by wrong management of the resources
- The power resources those are available for specific task are either used or lost, thus, excessive resources will cause excessive loose of electrical energy in addition to minimize the life time of the preparatory equipment
- Limiting the resources to the value that is required is the best way to save the unneeded amount of electrical power, which increases the power utilization in addition to stop unneeded maintenance of the equipment's
- In order to make energy capacity equals or near to the actual required power, anticipation of the next time stamp is required. This anticipation – or as most commonly known in the field of energy "forecasting" – should be done with high reliability in order to make the optimization really available. The optimization is done when know how much the data center requires energy.
- Genetic algorithm is adaptive intelligent way that performs the required optimization in high efficiency and reliability

- Even the genetic algorithm can afford high efficiency, accuracy, and precision. It needs effort in implementation and development in order to design the suitable efficient model for the system in addition to get the best solution, as it depends on iterative process

- The data center environment is complex with respect to resources those are consuming the electrical power. Thus, complex modeling is needed to present the problem formulation in mathematical model. The required mathematical model should be accurate and expressive in the way it represents all component of the electromechanical system including power dispatch and thermal energy requirements of the engineering system.

- The exact modeling leads to exact, efficient, and reliable design, while errors or inability to express the engineering model in suitable model will decently lead to misguidance energy model and may lead to more loses in energy.

5.2 Future Work

Any research – even though it has an important contributions and good results – presents a scope for future work in the same field. This thesis research started optimization of the data center in the way that use smart idea about optimization using genetic algorithm and mathematical modeling of different system component by simple way.

So, a large scope to future development in the same field and starting from the problems and ideas those are faced in this research process are summarized in the points below:

- It's important to build researches that model the energy of the data center completely in terms of the computational power that could be expressed in different computer science engineering scales. In fact, exact model for such system requires long time of research and extensive work in different areas of interest, including computer science and operating systems, electrical engineering, mechanical engineering, and non-linear engineering system modeling.

- To build a hybrid model for optimization of the data center that combines both; fuzzy logic and genetic algorithm. The genetic algorithms are excellent solver for the complex optimization models and can get unique solutions for the systems like the one concern of interest. But the genetic algorithm is not the best forecaster for such applications. On the other hand, the fuzzy logic is excellent for such forecasting purposes and can handle different parameter of the system in non-mathematical way that lead to better expression efficiency and minimal approximation and estimation errors.

- Trying to minimize the energy of the data center by hardware researches convert heat into another shape of energy is in my mind a good task for future. In fact, most of energy that is consumed in the shape of heat is lost, not only that, but we pay another amount of energy to cool it. Thus, trying to use that energy in efficient and useful way is good future scope.

REFERENCES

[1] W. Tschudi, T. Xu, D. Sartor, and J. Stein," High-Performance Data Center", Public Interest Energy Research Industrial Program Report, California Energy Commission, 2002.

[2] S. Garg, "Benchmarking and Maturity Models for Green Practices", R.K.Puram, 110066, New Delhi, 2010.

[3] Minoli,"Designing Green Networks and Network Operations — Saving Run-the-Engine Costs with Green Networks and Data Centers', CRC Press, Taylor & Francis, NY, 2009

[4] Y. Fan, Q. M. Liang, Y.M. Wei, N. Okada, "A model for China's energy requirements and CO2 emission analysis", Environmental Modeling and Software 378-393, 2007

[5] M. Mitchell, "An introduction to genetic algorithm", A Bradford Book the MIT Press Cambridge, Massachusetts, 1999.

[6] J.H, Adaptation in Natural and Artificial Systems. Sci. Am. 267 (1), 66– 72 University of Michigan Press. Ann Arbor, MA.1992.

[7] J. R. Koza, "Genetic Programming", MIT Press Edition, 1992.

[8] S. Lim, B. Sharma, B. Tak, and C. R. Das, "A Dynamic Energy Management Scheme for Multi-tier Data Centers", Transaction on energy management IEEE, 2010.

[9] G. Wen, S. Feng, Y. Wan, P. Jiang, S. Zhang, "Energy-aware Application Scheduling based on Genetic Algorithm", Seventh International Conference on Natural Computation, 2011.

[10] H. Hlavacs, T. Treutner, "Genetic Algorithms for Energy Efficient Virtualized Data Centers", 6th International DMTF workshop on Systems and Virtualization Management, 2012.

[11] P. Guo, X. Wang, Y. Han, "The Enhanced Genetic Algorithms for the Optimization Design", International Conference on Biomedical Engineering and Informatics, 2010.

[12] R. Raghavendra, P. Ranganathan, V. Talwar, Z. Wang, X. Zhu, "No Power Struggles: Coordinated Multi-level Power Management for the Data Center". IJCIT, Volum 4, issue 8, 2012

[13] D. E. Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning", Addison – Wesley, 1999.

[14] B. Capehart, "Wayne Turner: Guide to energy management", The Fairmont press.5th edition, 2007.

[15] A. Woods," What can be done to make cooling systems in data centers more energy efficient", Article development led by ACM Queue, vol. 53, no. 4, 2010.

[16] C. Tomory, and P. Corp, "Lowering the Carbon Emissions Footprint of Enterprise Data Centers through Energy Efficiency Gains", University of Oregon, 2010.

[17] L. A. Barroso and U. Holzle. "The Case for Energy-Proportional Computing", IEEE Computer, 40, 2007.

[18] G. Chen. "Energy-Aware Server Provisioning and Load Dispatching for Connection-Intensive Internet Services". In Proceedings of NSDI, April 2008.

[19] K. Le, R. Bianchini, and T. D. Nguyen. "A Cost-Effective Distributed File Service with QoS Guarantees". In Proceedings of Middleware, November 2007.

[20] C. Guo, G. Lu, D. Li, H. Wu, X. Zhang, Y. Shi, C. Tian, Y. Zhang, and S. Lu. Cube. "A High Performance, Server-centric Network Architecture for Modular Data Centers". In ACM SIGCOMM, August 2009.

[21] B. Heller, S. Seetharaman, P. Mahadevan. "ElasticTree: Saving Energy in Data Center Networks". In NSDI'10, Apr 2010.

[22] J. Chabarek, J. Sommers, P. Barford. "Power Awareness in Network Design and Routing". INFOCOM'08, Apr 2008.

[23] S. Srikantaiah, A. Kansal and F. Zhao. "Energy Aware Consolidation for Cloud Computing". HotPower'08, Dec 2008.

[24] Z. Abbasi, G. Varsamopoulos, and S. Gupta. "Thermal Aware Server Provisioning and Work Load Distribution for Internet Data Centers". In Proceedings of ACM International Symposium on High Performance Distributed Computing (HPDC10), USA, 2010.

[25] Al-Qawasmeh, "Heterogeneous Computing Environment Characterization and Thermal-Aware Scheduling Strategiesto Optimize Data Center Power Consumption". Ph.D. Thesis, Colorado State University, USA, 2012

[26] D. Bouley. "Estimating a Data Centers Electrical Carbon Footprint". White Paper66; APC by Schneider Electric, USA, 2011.

[27] G. Callou, E. Sousa. And F. Magnani. "A Formal Approach to the Quantification of Sustainability and Dependability Metrics on Data Center Infrastructures". In Proceedings of DEVS—Discrete Event System Specification, USA, 2011.

[28] X. Zhang, X. Zhao, Y. Li, and L. Zeng. "Key Technologies for Green Data Center". In Proceedings of 2010 3rd International Symposium on Information Processing (ISIP), China, 2010.

[29] G. Callou, P. Maciel, D. Tutsch, J. Ferreira, J. Arajo, and R. Souza. "Estimating Sustainability Impact of High Dependable Data Centers: A Comparative Study between Brazilian and U.S". Energy Mixes. Computing 2013.

[30] K. Baker, "The Case for Blade System Servers", hp, 2006

ABSTRACT (عربي)

مركز البيانات هو عبارة عن بنية يتم انشاؤها لتقوم بالعمليات المحوسبة ذات القدرات العالية والتي تتضمن خادم او خوادم الحواسيب, ادارة الشبكة, حماية الحواسيب والشبكة, الانترنت, قواعد البيانات, مقسم الاتصالات واجهزة تخزين البيانات وما شابه ذلك. مركز البيانات بشكل عام يتحاج لمصدر احطياطي لتزويد الطاقة الكهربائية, وصلات اتصال ذات موثوقية عالية, انظمة تكييف للمحافظة على درجات الحرارة ضمن الحد المطلوب وانظمة مراقبة المحيط والتي تتمثل في انظمة انذار ومكافحة الحريق والحماية ضد الدخول والسرقة. ومع اتساع وزيادة متطلبات استهلاك الطاقة الكهربائية المطرد مع ازياد المطلبات والقدرات الانتاجية لعمليات الحوسبة والاتصالات في مركز البيانات اصبحت تكاليف استهلاك الطاقة الكهربائية لا يمكن تجاهلها وفي الغالب تشكل عبء مالي بحاجة الى الترشيد. وهذه الرسالة تقدم طريقة مثلى لاستخدام للطاقة في مراكز البيانات باستخدام الخوارزميات الجينية. الطريقة المقدمة تتمثل في التنبؤ بالحمل المتوقع على مركز البيانات وبالتالي تشغيل مصادر الطاقة الكهربائية حسب القدرة المتوقة المطلوبة. المشكلة الاساسية تتكون من ان الانظمة الحالية تقوم بتشغيل مصادر الطاقة الكهربائية اعلى من الحمل الاقصى للاستهلاك خلال اليوم, وبالتالي تتشكل نسبة كبيرة من الضياعات بدون استخدام وبدون فائدة. الحل المقدم في هذه الرسالة تقدم حلا لهذه الضياعات من خلال التنبؤ الامثل في الاستهلاك المتوقع خلال فترة زمنية مستقبلية بالتخدام الخوارزميات الجينية, وبالتالي التشغيل الامثل لمصادر الطاقة الكهربائية وتقليل الضياعات بنسبة عالية. في هذه الرسالة يتم عرض النتائج من خلال تطبيق الحل المقدم ومناقشته باستخدام برمجية MATLAB.