INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

AN OPTIMIZED APPROACH FOR EFFICIENT TRANSMISSION OF AN ARTIFICIAL NEURAL NETWORK PROCESSED IMAGES OVER THE INTERNET

BY

OMER MAHMOUD MOHAMED

A thesis submitted in fulfilment of the requirement of the degree of Doctor of Philosophy in Engineering

Kulliyyah of Engineering
International Islamic University Malaysia

June 2008

ABSTRACT

Internet growth in recent years has encouraged many new applications to be provided for the user such as video conferencing and video telephony. Most of these new applications require data involving images characterized by a large number of bits. Therefore, it is a difficult task to implement these types of applications since the data needs to be transferred from one place to another over the network. Moreover, these applications are intended to be real time applications implying that they are very sensitive to delay and jitter which add to the challenges of providing such applications. Consequently, two strategies are needed in order to meet the challenges of transferring and storing very high volume of data for the application while maintaining strict requirement for delay and jitter. The first strategy involves qualitative and quantitative optimization of the application's data. The qualitative optimization ensures that the data of the application meets at least a minimum quality requirement (e.g. image resolution). Quantitative optimization is accomplished by using compression techniques to reduce the number of bits that are required to be transmitted over the network. The second strategy involves optimizing the utilization and performance of the network. This research, which consists of two phases, examines an optimized approach for transferring and storing images. Phase one deals with the processing and compressing of the images. Processing the images is done by applying an adaptive filter in order to enhance the visual quality while compression is used to reduce the amount of data needed to be stored or transmitted. The study developed a Multilayer Feed Forward Artificial Neural Network (MFFANN) for image compression. Based on Gradient Descent, Conjugate Gradient, Quasi-Newton techniques, three different error back propagation algorithms were developed for use in training the MFFANN. The essence of this study is to investigate the most efficient and effective training methods for use in image compression and its subsequent applications. The results show that the Quasi-Newton based algorithm has a better performance as compared to the other two algorithms. In Phase two, the study proposes a new admission control mechanism that aims to enable Multi Protocol Label Switching (MPLS) tunnel differentiations. It also presents a simulation based evaluation for the proposed mechanism. The results demonstrate the effectiveness of the proposal, as it is able to maintain a robust and stable end-to-end quality of service for selected flows, which leads to a better performance in terms of throughput, delay and jitter. The implementation of the optimization approach presented in this research could be highly beneficial in providing online real time services such as medical services, where consultation or even remote medical operations could take place using the Internet.

ملخص البحث

النمو الكبير فى شبكة الانترنت فى السنوات الاخيرة شجع على تقديم الكثير من التطبيقات الجديدة للمستخدمين مثل عقد المؤتمرات باستخدام الفيديو واتصالات الفيديو الهاتفية عبر الإنترنت. ومعظم هذه التطبيقات الجديدة تحتوى على الصور ولذا تتسم بحجم هائل جدا من البيانات مما يشكل جزأ من المصاعب في تنفيذ هذا النوع من التطبيقات حيث ان هذه البيانات يجب نقلها من مكان الى آخر عبر الشبكه . وعلاوة على ذلك فان هذه االتطبيقات يهدف لها ان تكون من التطبيقات الوقت الحقيقي مما يعني الها حساسه جدا للتأخير Delay والجتر Jitter وهذا يضاف الى تحديات فى توفير هذه االتطبيقات. وبالتالي هناك حاجة ماسه الى استراتيجه من شقين من أجل مواجهة التحديات المتمثلة فى نقل وتخزين كم هائل من البيانات مع الحفاظ على شرط صارم لتأخير و الجتر. الشق لأول من لاستراتيجية ينطوي على تحسين نوعي وكمي لبيانات التطبيق. يضمن التحسين النوعي توفر الحدا الادنى على الاقل من الجوده المطلوبه للبيانات. اما التحسين الكمي فيتم عن طريق استخدام تقنيات ضغط البيانات التي يتعين بثها على الشبكه للحد من حجمها. والشق الثانى من الاستراتيجيه يعنى بتكيف الشبكه لرفع ادائها و لاستفادة الى اقصى حد ممكن من مواردها. هذا البحث يقدم دراسه لايجاد نهج امثل لنقل وتخزين الصورويتكون من جزئين. الجزء الأول يحتوى على معالجة الصور ثم ضغطها. تتم معالجة الصور من خلال تطبيق مرشح قابل للتكيف من اجل تحسين جودة الصورة في حين ان ضغط يستخدم للحد من كمية البيانات التي تحتاج الى ان تخزن او تنقل عبر الشبكه. استخدمت الدراسه الشبكة العصبية Multilayer Feed Forward Artificial Neural Network الاصطناعيه متعدد الطبقات أماميه التغذية لضغط الصور. تم تدريب الشبكه العصبيه الاصطناعيه باستخدام ثلاث تقنيات تدريب مختلفة وهى تقنية الميل النسبي Gradient Descent وتقنية الميل المتقارن Gradient Conjugate وتقنية كواسو نيوتن Quasi-Newton. جوهر هذه الدراسه هو ابراز الخوارزميات الاكثر كفاءه وفعالية في تدريب الشبكه العصبيه الاصطناعيه متعدد الطبقات اماميه التغذيه لاستخدامها في ضغط الصور . وتبين النتائج ان خوارزميه كواسو نيوتن تقوم باداء افضل بالمقارنة مع الخوارزميات الاخرى. وفي الجزء الثاني ، تقترح الدراسه آلية جديدة للتحكم فى قبول التدفقات والتي تمكن الشبكه من تقديم افضليه لبعض التدفقات . كما تقدم الدراسه تقييما للاليه المقترحة على اساس برنامج المحاكاه . النتائج المتحصل عليها اثبتت فعالية هذا الاقتراح حيث انه يودى الى اداء متماسك وثابت من حيث جودة الخدمه المقدمه الى المجموعه المختاره من التدفقات. ان الاستراتجيه المقدمه فى هذا البحث يمكن ان تكون ذات فائده عظيمه لتقديم خدمات تطبيقات الوقت الحقيقى على الاترنت مثل الخدمات الطبيه حيث يمكن تقديم الاستشاره الطبيه او حتى المساعده فى اجراء عمليات جراحيه من على البعد باستخدام الانترنت.

APPROVAL PAGE

The thesis of Omer Mahmoud Mohamed has been approved by the following:

Farhat Anwar
Supervisor

Momoh J. E. Salami
Supervisor

Othman O. Khalifa
Internal Examiner

Mohammed Umar Siddiqi
Internal Examiner

Ali Chekima
External Examiner

Muhammad Nasir Din
Chairman

DECLARATION

I hereby declare that this dissertation is the result of my own investigations, except where otherwise stated. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at IIUM or other institutions.

Omer Mahmoud Mohamed

Signature …………………………………… Date ……………………..

INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

DECLARATION OF COPYRIGHT AND AFFIRMATION OF FAIR USE OF UNPUBLISHED RESEARCH

AN OPTIMIZED APPROACH FOR EFFICIENT TRANSMISSION OF AN ARTIFICIAL NEURAL NETWORK PROCESSED IMAGES OVER THE INTERNET

No part of this unpublished research may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission of the copyright holder except as provided below.

1. Any material contained in or derived from this unpublished research may only be used by others in their writing with due acknowledgement.

2. IIUM or its library will have the right to make and transmit copies (print or electronic) for institutional and academic purposes.

3. The IIUM library will have the right to make, store in a retrieval system and supply copies of this unpublished research if requested by other universities and research libraries.

Affirmed by Omer Mahmoud Mohamed

…………… ……. …………….

Signature Date

ACKNOWLEDGEMENTS

I wish to express deep appreciation and gratitude to my supervisors Prof. Momoh Salami and Dr. Farhat Anwar for their endless help and advice at every stage of my research work.

I also wish to thank examination committee members for their interest in this work.

I express appreciation to my friends and colleague for their support and encouragement during my research

I reserve my deepest gratitude to my parents, brothers, sister and other family members, for their persevering effort toward my entire academic career

TABLE OF CONTENT

LIST OF FIGURES

Figure No — Page No

LIST OF TABLES

LIST OF ABBREVIATIONS

AF: Assured Forwarding
ANN: Artificial Neural Network
BA: Bandwidth Agent
BB: Bandwidth Broker
BER: Bit Error Rate
CBR: Constant Bite Rate
CD: Conjugate Gradient
COPS: Common Open Policy Service
CRC: Cyclic Redundancy Check
DER: Differentiated Services Edge Router
Diffserv: Differentiated Services
DSCP: Differentiated Services Code Point
EF: Expedited Forwarding
ER: Edge Router
FEC Forward Equivalent Class
FIB: Forwarding Information Base
GD: Gradient Descent
HL: Hidden Layer
IETF: Internet Engineering Task Force
IL: Input Layer
Intserv: Integrated Services
IP: Internet Protocol
JPEG: Joint Picture Expert Group
LSP: Label Switch Path.
LSR: Label Switch Router
MER: MPLS Edge Router
MFFANN: Multilayered Feed Forward Artificial Neural Network
MPEG: Motion Picture Expert Group
MPLS: Multi Protocol Label Switching
NHLFE: Next Hop Label Forwarding Entry
OL: Output Layer
OPSF: Open Short Path First
PHB: Per Hope Behavior
QN: Quasi Newton
QoS: Quality of Service
RED: Random Early Detection
RSVP: Resource reSerVation Protocol
SLA: Service Level Agreement
TCA: Traffic Conditioning Agreement
TE: Traffic Engineering
TOS: Type of Service
TTL: Time to Live
VBR: Variable Bite Rate

CHAPTER ONE

INTRODUCTION

1.1 OVERVIEW

Over the years, the Internet has become one of the most important communication channels. The default service associated with the Internet is a *best-effort* service, where the network treats all traffic in exactly the same way. There is no consistent service outcome from the Internet best-effort service model. When the load level is low, the network delivers a high quality service. The best-effort Internet does not deny entry of traffic to the network so, as the load levels increase, the network congestion levels increase, and service-quality levels decline accordingly. This decline in service is experienced by all traffic passing through a congestion point, and is not limited to the most recently admitted traffic flows. Best effort is acceptable for traditional Internet applications like web, email, file transfer, and the like, however, it is inadequate for new classes of applications such as audio and video applications which are characterized by high volume of raw data and time sensitivity. Therefore, it is essential to reduce the volume of the application data by means of compression and enable end to end quality of service within the network. Providing Quality of Service (QoS) is one of the main challenges in the Internet that has recently gained the interest of many researchers. The Internet Engineering Task Force (IETF) proposes three main architectures in an effort to make IP QoS realty. These architectures are integrated services (Inserv), Differentiated services (Diffserv) and Multiprotocal Label Switching (MPLS). Active research is currently going on in order to combine and to enhance these new architectures [Mahmoud, Anwar and Sallami 2006] [Faucheur, and Lai 2005]

1.2 PROBLEM STATEMENT AND ITS SIGNIFICANCE

In the last few years, Internet has spread widely and prove itself to be one of the most economical and affective ways of communication. This expansion in the internet promotes many new applications to be developed and offered. The Internet was originally designed to provide best effort service; however, due to the nature of these new applications it requires service beyond best efforts which bring up the first challenge of the need to alter the existing Internet protocol in order to meet these demands. The second challenge lies within the characteristics and nature of the data that forms these applications. As these applications contain video and images, therefore, are characterized by high volume of data (bits). Consequently, they require high bandwidth for transmission and for storage. For example, a 1024 X1024 color image with 8 bit/pixel generates 25 Mbits of data, which without compression requires about 7 minutes of transmission time on a 64 kbps line. A Compact Disk with storage capacity of 5 Gbits can only hold about 200 uncompressed images. Therefore, two strategies are needed in order to meet the challenges of transferring and storing applications with very high volumes of data while maintaining strict requirements for delay and delay variation. The first strategy involves the qualitative and quantitative optimization of the application. The qualitative optimization ensures that the data of the application meets - at least -some minimum quality requirement (e.g. image resolution). The quality of the data is highly application dependent. For example, a user of an Internet telephone will be happy if he can get an image that can enable him to recognize the person he is talking to, but a user of an entertainment game will need and expect images with very high resolution. Quantitative optimization is accomplished by using compression techniques to reduce the number of bits that are required to be transmitted over the network.

The second strategy which needs to be carried out is to optimize the utilization and performance of the network.

1.3 RESEARCH OBJECTIVES

The main objectives of this research are:

1- Development of an adaptive image enhancing filter to improve image visual quality

2- Development of an image compression technique that reduces the amount of data representing the image while producing decompressed images of good quality.

3- Evaluation of the current Internet protocols and proposal of modifications to enhance the quality of service provided by the networks especially for real time applications.

4- Proposing a new traffic admission control mechanism to realize objective number three.

5- Simulation based evaluation of the suggested mechanism.

1. 4 RESEARCH METHODOLOGY

The research methodology adapted in order to achieve the objectives is presented in this section. It involves the following steps.

1- Conducting a literature review to explore the current state of the art in the research area

2- In order to achieve the first objective an algorithm was developed to design an adaptive image filter

3- This new adaptive filter was implemented and evaluated using MATLAB software

4- In order to achieve the second objective a neural network based image codec was developed

5- Different learning algorithms to train this neural network were developed

6- The neural network based image codec was implemented on MATLAB software.

7- Additional functionalities to the current Internet components were identified to provide a better QoS.

8- A new traffic admission control was proposed to achieve the QoS requirement

9- The performance of this admission control was evaluated by means of simulation

Several experimentation and analysis tools were required to accomplish these steps.

These tools are:

- Linux operating system. This platform is needed because the network simulator works most suitably on Linux
- Network Simulator: this will be the base of the simulation
- Trace graphs will be used to interpret the trace files so that important results can be extracted from them
- Microsoft office will be used for arranging and plotting extracted results for better visual analysis
- MATLAB software will be used to implement and test the neural network based image codec and the adaptive image enhancing filter.

1.5 RESEARCH SCOPE

The thesis presents an optimized approach for transferring and storing images. The approach at hand consists of optimizations on two levels or layers: the Application Layer and the Network Layer. Application layer optimization deals with processing and compressing images (e.g. MRI images). Processing the images aims to enhance visual quality while compression is used to reduce the amount of data needed to be stored or transmitted. Network layer optimization proposes a new admission control mechanism that aims to enable MPLS-TE tunnel differentiations whereby a modified Multiprotocol Label Switching MPLS would be used to provide real time applications with a very high quality of service (QoS). The thesis is divided into two phases. Phase one will consider the optimization in the application layer, while phase two will present optimization on the network layer.

1.5.1 Image Enhancement and MFFANN based Compression

Due to many factors that are related to the sensor or outside environments the acquired image may need to undergo some processing in order to improve its quality.

The most common problems are noise either due to insufficient signal or inherent in the sensor and nonuniform illumination or brightness across the image.

The first issue which will be considered in this phase is to enhance image visual quality generally while giving the edges of the image extra attention. This will be accomplished by using contrast enhancement techniques (e.g. histogram equalization) and using adaptive enhancement strategies by segmenting the image according to transition in grey levels. For examples, the edges are characterized by a sharp transition in grey level and thus using sharpening filters will enhance the edge of the image while in the areas where the grey level changes gradually smoothening filters can be use to reduce the noise within that segment. The block diagram of the enhancement process is shown in Figure 1-1.

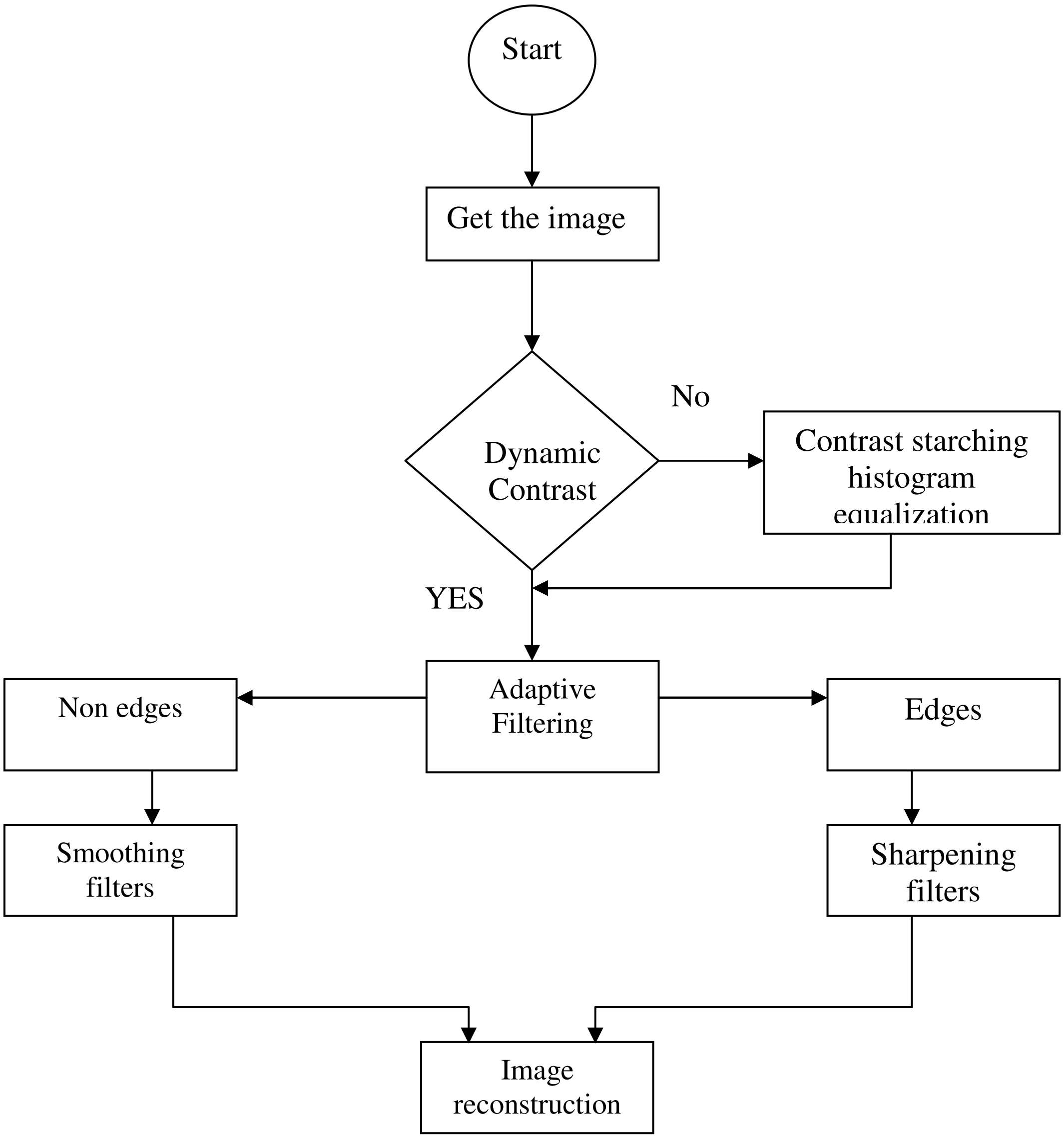

Figure 1-1: Block Diagram of an Image Enhacement Steps

The second issue that will be considered in this phase is the compression of images. By using compression techniques, it is possible to remove some of the redundant information contained in images, requiring less storage space and less time to transmit.

Multilayer Neural Network with Error Back Propagation Algorithms will be used for compressing the images. The process will be as follows (shown in Figures 1-2 and 1-3).

1) Training of the Neural Network:

- Create a neural network system with a selected architecture.
- X Input Neurons
- Y-Z Hidden Neurons (the compression ratio is the ratio of input to hidden neurons)
- X Output Neurons (X>Y>Z)
- Break the training image into blocks.
- Scale each block
- Use the scaled blocks as input to the neural network
- Adjust the weights so as to minimize the difference between the input and output
- Repeat until the error of the training set is sufficiently small

2) Compression:

- Break the image into blocks
- Scale each block
- Use the scaled blocks as input to the neural network
- Take the output of the HIDDEN LAYER
- Quantize this data
- Save the quantized data

3) Decompression:

- Read in the quantized data
- Set the output of the hidden layer equal to the read data
- Find the output of the output layer
- Scale the real output to an integer value
- Put this data into the image
- Repeat for each block of the image
- Reconstruct the image