Applying Hyperthreading Technology for evaluating the performance of HTTP server for stored audio/video retrieval

A. Madhura V Phatak, B. Rakhi Dongaonkar,Jr.

Abstract -
This paper discusses the proposed system that analyses the performance of the prototypes of threaded HTTP servers on Hyper-Threading as well as non-Hyper-Threading platforms. Once the performance benefits for the prototype HTTP servers are recorded, these would be helpful in building various client-server interfaces over the network.
This paper aims at writing HTTP Web-Servers for the 4 basic Threading models.
These models are analysed for stored audio and video streaming on Hyper-Threading enabled and Hyper-Threading disabled platform.
The differences in performance analysis are recorded and studied.The benefits of Hyper-Threading platform are emphasized .
The system consists of the following modules:-

HTTP server modules for the 4 basic Threading modules:-

- *Non-threaded server and simple client.*
- *Single thread of the server and multiple simple clients.*
- *Single thread per simple client request.*
- *Threaded serverr module and multiple simple clients.*

All the above threading modules are used for the purpose of stored audio and video streaming.
The server stays resident on the Hyper-Threading enabled machine,the client machine could have any basic configuration or the client and the server could be on the same machine. The audio and video retrieval is based on HTTP streaming protocol,HTTP itself is capable of the streaming effect.The server response is measured under the audio and video workload stress.Hyper-Threading benefits are analysed for the applications.

Forming suitable test cases for analysing the client-server interface is an important aspect of the analysis.The test cases would be further analysed to prove and explain the benefit levels of Hyper-Threading Technology.

KEY WORDS
HTTP server,Hyper-Threading and Threading models,HT(hyperthreading),UT(Uniprocessor Time).

I. INTRODUCTION

To date computational power has typically increased over time because of the evolution from simple pipelined designs to complex speculation and out of order execution of many of todays deeply pipelined,super-scalar designs.While the processors are now much faster than they used to be,the rapidly growing complexity of designs also makes achieving significant additional gains more difficult.
Consequently,processors/systems that can run multiple software threads have received increasing attention as a means of boosting overall performance.
In this system the workloads of simple HTTP requests and replies are characteristized,alongwith the workloads of stored audio and video,ie the multimedia loads.

The workloads are first charaterized on the current super-scalor architectures and then on the Hyper-Threading platform.Thus the major goal is to provide a better understanding of performance improvements in multimedia applications on processors with Hyper-Threading technology.

Intel introduced Hyper-Threading Technology for its Intel Pentium 4 family of processors. Hyper-Threading involves the execution of two streams of instructions (threads) that could be interleaved to share key portions of the processor's facilities. Hyper-Threading introduced processor-level threading.

Let's consider a single processor. It will have a single execution engine, it's own Architecture-state and it's own Advanced Programmable Interrupt Controller (APIC). Now let's consider a Hyper-Threading enabled processor. This processor has it's own execution engine, but two architecture states and two Advanced Programmable Interrupt Controllers(APIC).This means that it's Architecture States(AS) and Advanced Programmable Interrupt Controller(APIC) are indeed duplicated.

Let us view the following diagram:-

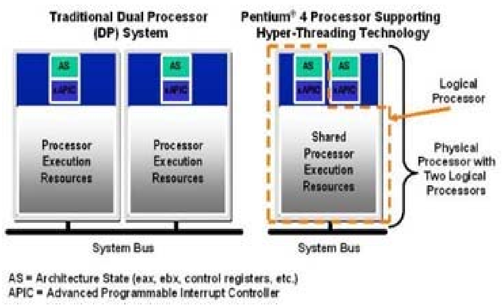

Traditional Dual Processor (DP) System
Pentium® 4 Processor Supporting Hyper-Threading Technology
AS
APIC
Processor Execution Resources
Shared Processor Execution Resources
Logical Processor
Physical Processor with Two Logical Processors
System Bus
AS = Architecture State (eax, ebx, control registers, etc.)
APIC = Advanced Programmable Interrupt Controller

Figure 1: block diagram of Hyper-Threading.

What we are viewing above is the diagrammatic comparison of a dual processor system with that of a Hyper-Threading enabled system. A dual processor system is a collection of 2 individual processors, while a Hyper-Threading processor is a collection of 2 Architectural states and 2 Advanced Programmable Interrupt Controllers, with the same execution engine.

Hyper-Threading Technology is mainly a server side technology, though it is applied to desktop applications.

A. *Need for the system*

Hyper –Threading Technology was mainly developed as a server side technology to improve the server response time.
Hyper-Threading Technology allows multithreaded server software applications to execute threads in parallel within each each processor of a server platform.
Thus by analyzing the response time of the multi-threaded server would go a long way
in judging the Hyper-Threading technology.The workloads would also determine a lot about the response time of the server running on the Hyper-Threading platform.
The IIS(Internet Information services) and the apache server are web servers that are available for installation on the windows platform and both the web-servers support the concept of multithreading ,but I chose to write the 4 prototype web-servers for analysing the performance of the application on the Hyper-Threading platform.

B. Current System

- HTTP servers that run on architectures other than Hyper-Threading platform.
- The clients expect the response time to be less and replies to be quick.
- For large and complex servers the response time is very critical.
- On non-HT platforms the server response time is low, especially when critical applications are concerned.

Comparison with the existing system

When HT platform is compared with single processor non-HT platform. The former is definitely efficient. In HT Technology the idle states in a processor are few as compared to the non-HT platform. This benefits the threaded applications. Thus, this would benefit the threaded servers in lowering their response time towards their clients. In HTTP servers, response time is the key performance parameter. In this case also, the HT platform can prove its benefits over the non-HT platform.

C. Proposed system

Server platforms based on the Intel Xeon processor family have implemented the necessary changes in the platform BIOS in order to recognize the logical processors so that the operating system and software can utilize Hyper-Threading technology.
Todays server software is multithreaded and can take advantage of Hyper-Threading Technology
without changes.
The proposed system consists of writing 4 HTTP server threading modules for the purpose of audio and video streaming.The performance of the server side response times is measured and analysed once on Hyper-threading enabled platform and then on Hyper-Threading disabled platform.

Applications that exhibit good threading methods and scale well on Multi-processor based servers today are likely to take advantage of Hyper-Threading Technology.The performance increase seen is highly dependent on the nature of the application,the threading model it uses as well as system dependencies.It has been proved that some server applications can experience up to 30 percent additional performance due to Hyper-Threading Technology in the Intel Xeon processor implementation.
The system comprises of writing 4 HTTP server modules for audio/video retrieval .
The users can connect to the servers through the web-client and request the web server for web
Audio video files .The response time,that is the time at which the first request is made to the time at which the first response is received is critical.
The response time depends upon factors like:

- The kind of request
- The number of clients
- The time the request spends in the wait queue.
- The time taken to sojourn through the work process.

D. Module details

There are 4 basic threading modules:-

- Non-threaded server and simple client
- Single thread per simple client request
- Single thread of the server and multiple simple clients
- Worker thread module and multiple simple clients

Description of the execution of the 4 modules
Non-threaded server and simple client.

The Input specified is HTTP audio/ video file request and the Output is HTTP audio/ video file reply.

The HTTP server services the Client request
The response of the audio video reply is analyzed on Hyper-Threading enabled and Hyper-Threading disabled platform.The results are documented.

Single thread per simple client request

The Input specified is HTTP audio/ video file request and the Output is HTTP audio/ video file reply.

The HTTP server services the Client request.

The response of the audio video reply is analyzed on Hyper-Threading enabled and Hyper-Threading disabled platform.The results are documented.

Single thread of the server and multiple simple clients.
The Input specified is HTTP audio/ video file request and the Output is HTTP audio/ video file reply.

The HTTP server services the Client request and the response of the audio video reply is analyzed on Hyper-Threading enabled and Hyper-Threading disabled platform.The results are documented.

Threaded Server and multiple simple clients
The Input specified is HTTP audio/ video file request and the Output is HTTP audio/ video file reply.

The HTTP server services the Client request
The response of the audio video reply is analyzed on Hyper-Threading enabled and Hyper-Threading disabled platform.The results are documented.

II. RESULTS

Proper testing procedures were followed,the and response of audio and video retrieval was checked from less load to about 50 clients,by using appropriate testing tool,such as webserver stress tool.The results are obtained in a controlled environment.

Results for Non-threaded server and simple client,such as IE/any browser.

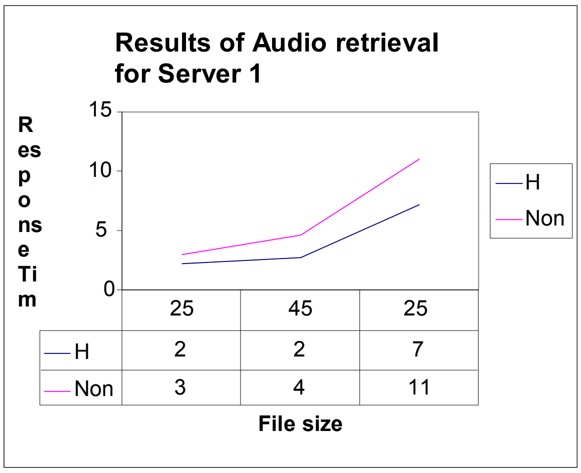


Figure 2: Graph of results of audio retrieval for server 1

HT scaling for above is UT/HT,ie 62.33/41.3 = 1.5

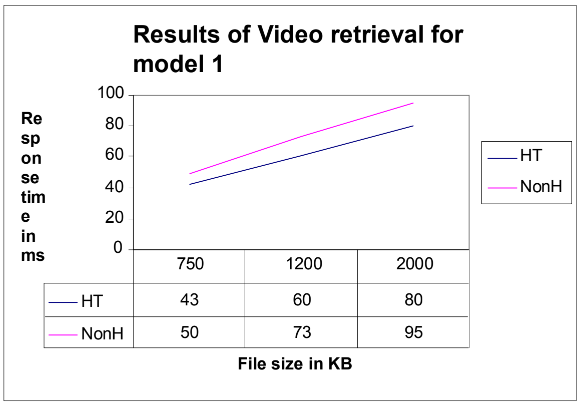


Figure 3: Graph of results of video retrieval for server 1

HT scaling for above is UT/HT,ie 72.66/61=1.19

Results of Single thread of server per simple client (such as IE/any browser).

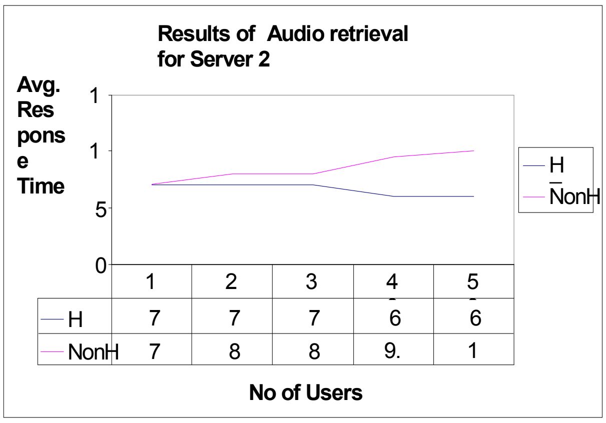


Figure 4: Graph of results of audio retrieval for server 2

HT scaling for above is UT/HT,ie 8.5/6.6=1.28

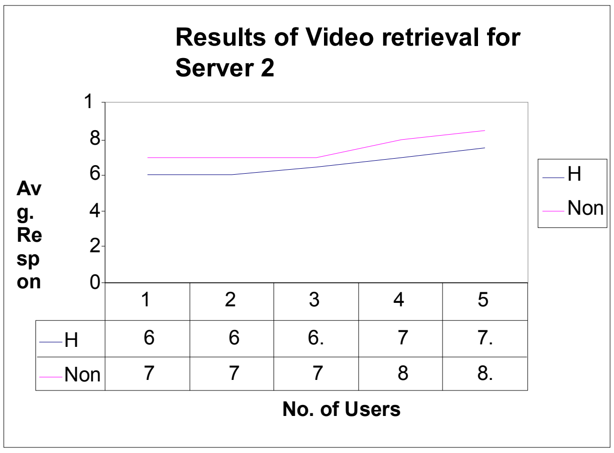
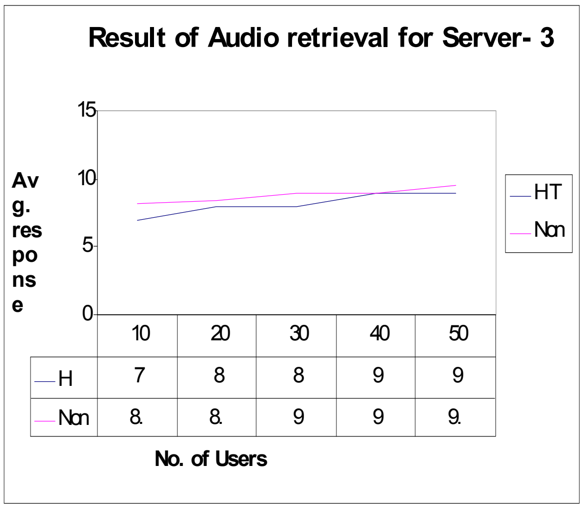


Figure 5: Graph of results of video retrieval for server 2

HT scaling for above is UT/HT,ie 7.5/6.6 =1.13

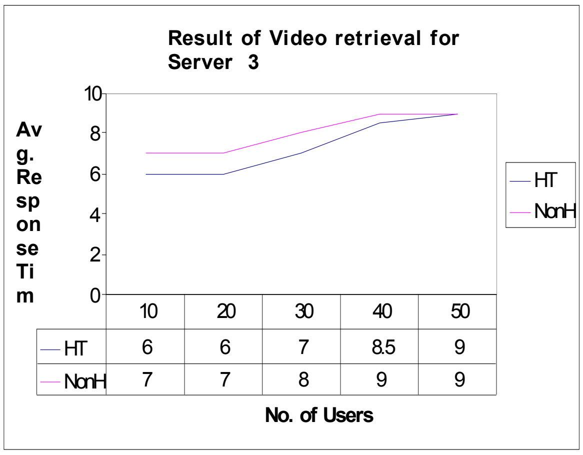
Results for Single thread of the server (non-threaded-server) and multiple simple clients.



	10	20	30	40	50
H	7	8	8	9	9
Non	8.	8.	9	9	9.

Figure 5: Graph of results of audio retrieval for server 3

HT scaling for above is UT/HT,ie 8.86/8.2=1.08

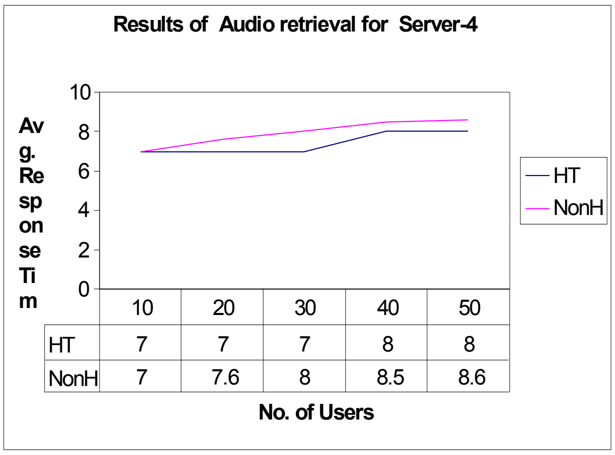


	10	20	30	40	50
HT	6	6	7	8.5	9
NonH	7	7	8	9	9

Figure 6: Graph of results of audio retrieval for server 3

HT scaling for above is UT/HT ie 8/7.3=1.09

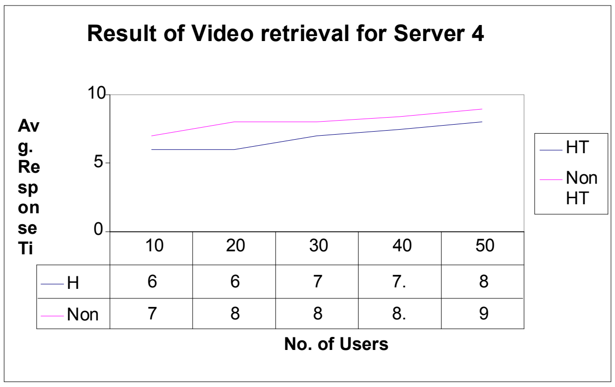
Threaded server and multiple simple clients



	10	20	30	40	50
HT	7	7	7	8	8
NonH	7	7.6	8	8.5	8.6

Figure 7: Graph of results of audio retrieval for server 4

HT scaling for above is UT/HT,ie 7.94 / 7.4 = 1.07



	10	20	30	40	50
H	6	6	7	7.	8
Non	7	8	8	8.	9

Figure 8: Graph of results of video retrieval for server 4

HT scaling for above is UT/HT,ie 8.1/6.9 =1.17

III. CONCLUSION

In conclusion, Hyper-Threading does give performance benefits,but this too is dependent on various factors like the underlying OS,or kind of application written.
Hyper-Threading Technology can be explored for a variety of applications,and benefits for each can be analyzed.
Thus this application does show performance benefits ranging from scaling factor of around 1.

REFERENCES

[1] Richard Gerber and Andrew Binstock , *Programming with Hyper-Threading Technology*, Intel Press 2004.
[2] Charles Petzold, *Programming Windows,*Microsoft Press 1999.
[3] Richter Clark,*Programming server side applications for windows 2000*,Microsoft Press.
[4] Free tutorials
http://www.intel.com
http://www.paessller.com