Unified Approach for Operating System Comparisons with Windows OS Case Study

Raja Zahilah, Farhaj Tahir, Anazida Zainal, Abdul Hanan Abdullah, Abdul Samad Ismail
Department of Computer Science
Faculty of Computing, Universiti Teknologi Malaysia
Johor Bahru, Johor, Malaysia

Abstract — **The advancement in technology has changed how people work and what software and hardware people use. From conventional personal computer to GPU, hardware technology and capability have dramatically improved so does the operating systems that come along. Unfortunately, current industry practice to compare OS is performed with single perspective. It is either benchmark the hardware level performance or performs penetration testing to check the security features of an OS. This rigid method of benchmarking does not really reflect the true performance of an OS as the performance analysis is not comprehensive and conclusive. To illustrate this deficiency, the study performed hardware level and operational level benchmarking on Windows XP, Windows 7 and Windows 8 and the results indicate that there are instances where Windows XP excels over its newer counterparts. Overall, the research shows Windows 8 is a superior OS in comparison to its predecessors running on the same hardware. Furthermore, the findings also show that the automated benchmarking tools are proved less efficient benchmark systems that run on Windows XP and older OS as they do not support DirectX 11 and other advanced features that the hardware supports. There lies the need to have a unified benchmarking approach to compare other aspects of OS such as user oriented tasks and security parameters to provide a complete comparison. Therefore, this paper is proposing a unified approach for Operating System (OS) comparisons with the help of a Windows OS case study. This unified approach includes comparison of OS from three aspects which are; hardware level, operational level performance and security tests.**

Keywords— *Benchmarking; Windows OS; Operating System*

I. INTRODUCTION

In today's era of technology, the world has seen development of multitude of Operating System (OS) but only a few has managed to grab the attention and of those three popular environments namely, Windows, Macintosh (MAC) and Linux distributions. According to a survey (Choney, 2013) Microsoft dominates the OS market share with over 90% of personal desktop computers running different variants of available Microsoft Windows OS while MAC OS X and others are following at a desperate and disappointing rate of approximately 5% market share combined together.

Microsoft enjoyed a monopoly of being the software giant for over 20 years but as technology is advancing researchers are claiming 2012 to be a post personal computer (PC) era (Hughes, 2012) in which PC are being replaced by tablets as primary consumption devices. To keep up with technology, Microsoft has jumped into hardware business and manufactures tablets and another variant of OS called Windows 8 that runs on desktops, laptops, and tablets. With the introduction of Windows 8 and surface tablets, Microsoft anticipates of being successful and upgrades its position from top desktop OS to top device OS. Furthermore, Microsoft has started to cut down its hardware partners and is selling the tablet itself to pull a profit margin similar to the PC revenues where PC is bundled with Windows and Office.

The purpose of this study is to propose a unified approach for Windows Operating System (OS) comparisons with the help of Windows OS case study. As based on current industry practice where most of the researchers; when comparing OS, either choose benchmarking to benchmark the hardware level performance or do penetration testing for checking security features of the OS. It is unknown whether the latest OS is in fact having better performance than its predecessor. A thorough comparison of OS should include hardware level and operational level performance. Thus, this research is aiming to integrate different aspects of the operating systems into comparison process which were neglected before to achieve a complete comparison result. The tests were performed on Windows XP, Windows 7 and Windows 8.

The organization of the paper is includes Introduction, related works on benchmarking, automated hardware level benchmarking methodology, performance analysis using different benchmarking software, proposed unified approach user oriented tasks benchmarking and finally conclusion and future work.

II. RELATED WORKS ON BENCHMARKING

Many researchers' and organizations focuses on the hardware-level performance for the components benchmarking

This work is supported by research university grant 12J19.

978-1-5386-0725-1/17/$31.00 ©2017 IEEE

that can be done through the automated and popular benchmarking tools available worldwide. Dr. Bob Steigerwald and Abhishek Agrawal (2011) from Intel software and services group mentioned Cinebench as their favourite benchmarking tool whereas Chanik Park (2006) from Samsung Electronics considers PCMark and SiSoft Sandra as Industry standard benchmarking software. Similarly, 3DMark11 is ranked as most popular tool for graphics benchmarking among the industry (Ungureanu M, 2011). Dain K Schroeder (2008) also performed MAC and PC performance analysis with the help of such software and similarly Goran Martinovic et al. (2012) who compared performances of Windows XP, Windows Vista and Windows 7 through the software and concluded Windows 7 and Windows Vista do not provide a better overall performance on the high-end computer system compared to Windows XP.

In this research, it came to our attention that the above-mentioned researchers do not consider user oriented tasks; hence the results seemed incomplete when comparing an OS. Furthermore, we have noticed that available industry standard benchmarking software merely focus on hardware-level system performance of OS. Thus, this research is to find the issues with the automated hardware level benchmarking method. The same performance measurement was used to train the system throughout the test.

III. AUTOMATED HARDWARE LEVEL BENCHMARKING METHODOLOGY

There is industry standard benchmarking software to benchmark the hardware level performance for Windows OS. Industry standard software gives us benchmarking results in terms of performance of the OS. Some of these tests are able to measure the following:

- Central processing unit tests such as mathematical operations, compression, encryption, and streaming SIMD extensions.
- 2D graphics test such as drawing lines, bitmaps, fonts, text, and GUI.
- 3D graphics test such as simple to complex DirectX 3D graphics and animations.
- Disk tests such as reading, writing and seeking within disk files.
- Memory tests such as allocating and accessing memory speed and efficiency.
- DVD burning test such as test the speed of drive for burning a DVD using windows disk burning tools.

The automated-hardware level benchmarking software used in this research are listed as in Table I with their capabilities.

TABLE I. AUTOMATED HARDWARE LEVEL BENCHMARKING SOFTWARE

Software	Capabilities
Maxon Cinebench 11.5	Graphics and CPU performance testing.
FutureMark PCMark 7	Accurate Series of tests including Video playback, DirectX graphics, Image manipulation and Web browsing.
PassMark performance test	Perform series of tests such as floating point match, image filtering, processing complex 3D vectors, read cached and un cached data, sequential read and write on disk which measures CPU, GPU, Memory and hard drive performance and give scores.
Sisoft Sandra Lite	User can test entire machine or specific components.
NovaBench	All in one benchmark suite testing CPU, GPU and hard drive performance with floating point operations, Integer operations, capability of processing 3d frames per second and hard drive writing speed with fast results
3DMARK 11	Series of tests measuring graphics and CPU performance. Tests include physics test, real time graphics rendering with detail and complexity.

IV. PERFORMANCE ANALYSIS OF AUTOMATED HARDWARE LEVEL BENCHMARKING SOFTWARE

The automated-hardware level benchmarking testing using different software was completed and the results are given below. The results include comparisons with Windows XP, Windows 7 and Windows 8.

A. PassMark Performance Test 8.0

PassMark Performance Test 8.0 evaluation version mainly provides benchmarking for CPU, hard disk, memory modules and graphic card. The evaluation version of PassMark Performance Test 8.0 has following benchmarking capabilities.

- CPU Mark: It is used to measure the CPU processing which includes benchmarks on integer math, extended instructions, compression, floating point math, CPU physics, prime numbers, encryption, sorting and single threaded performance of processors.
- 2D Graphics Mark: It is used to measure 2D graphics processing which includes benchmarks on Simple vectors, complex vectors, fonts and text, windows interface, image filters, image rendering and Direct 2D performance of graphic cards.
- 3D Graphics Mark: To measure 3D graphics processing which includes benchmarks on DirectX 9 Simple vectors, complex vectors, DirectX 10, DirectX 11 and Direct Compute performance of graphic cards.
- Memory Mark: To measure the memory processing which includes benchmarks on database operations, read cached and uncached data, data write, latency, available RAM and threaded performance of memory module.

- Disk Mark: It is used to measure the performance of hard disk and other drive such as DVD drive processing which includes benchmarks on sequential read and write on the disk and random seek and rewrite capabilities.

PassMark Performance Test 8.0 shows the performance rating of the computer in the software itself and gives an option to check results online where you can compare your results with similar systems to see if your hardware is performing, as it should be.

From the PassMark results shown in Table II, we have gathered that for overall performance, Windows 8 performs better. The results also show that Windows XP performs better in CPU tests whereas Windows 8 excels in graphics and memory, and Windows 7 in hard disk tests.

B. Maxon Cinebench 11.5

Maxon Cinebench 11.5 performs two tests each for CPU and Graphics card. OpenGL benchmark test stresses the Graphics card with a complex 3D car chase scene that uses the complete processing power of the GPU in OpenGL mode to benchmark the various factors. According to Cinebench (2013), the graphic card would need to display a huge amount of geometry textures, effects, bump maps, transparency and more to evaluate the capabilities of the graphic card. The result of this benchmark is provided in frames per second (fps), the greater the value indicates the better performance of the card.

Moreover, Cinebench Release 11.5 stresses the CPU to render a photorealistic 3D scene that utilises various algorithms to benchmark all available processor cores. According to Cinebench (2013), this test contains approximately 2000 objects, which in turn contain more than 300,000 polygons in total. This test measure the CPU by testing sharp and blurred reflections, shadows, procedural shaders, ant aliasing and much more. The result of this particular test is given in points (pts), the bigger the number shows the better performance.

From the Maxon Cinebench 11.5 results shown in Table III, we have gathered that Windows 8 still performs better. Windows 8 perform better in CPU tests whereas; Windows XP excels in graphics stress tests.

C. FutureMark PCMark 7

The basic version of FutureMark PCMark 7 provides benchmarking for the CPU, GPU and more. Several tests conducted are video playback, DirectX 9 graphics, image manipulation and web browsing.

Video playback test is a workload used to measure the performance of playing back video file. The intention of this workload is to test the system for its capability of playing a video content with the desired frame rate and without any glitches. The workload uses high definition H.264 content (1080p, 24fps, 12Mbps) and plays it with windows built-in video codec so that there will be no issues of invalid codec installations. DirextX 9 graphic test is a workload used to measure the performance of DirectX 9 graphics. The test was done with the help of firefly scene borrowed from 3DMark06. Image manipulation is a workload used to measure the performance of playing back images onto computer display. The intention of this workload is to use the Windows Imaging Component (WIC) to simulate the everyday image manipulations such as color correction, stretch, rotate and flip. Finally, web-browsing test is a workload makes use of embedded internet explorer to open pages in the tabs. Three tabs were used and the time was measured in terms of reload and rendering of all three tabs concurrently.

From the FutureMark PCMark 7 results, we have gathered that Windows 8 performs better overall but the breakdown of the scores can be seen in Table IV. As per PC Mark 7 results, Windows 8 performs better in CPU tests whereas Windows 7 excels in graphics stress tests as for Windows XP it does not support these workloads. The software stops responding abruptly under Windows XP as the workloads require features that Windows XP does not support.

TABLE II. PASSMARK PERFORMANCE TEST 8.0 RESULTS ANALYSIS

PassMark	**Windows XP**	**Windows 7**	**Windows 8**
CPU Mark	2607	2157	2329
Disk Mark	451	472	389
2D Graphics Mark	321	328	471
3D Graphics Mark	545	670	667
Memory Mark	1076	863	1074
PassMark Rating	1103	1076	1133

TABLE III. MAXON CINEBENCH 11.5 RESULTS ANALYSIS

Cinebench	**Windows XP**	**Windows 7**	**Windows 8**
OpenGL	28.50 fps	28.29 fps	28.36 fps
CPU	1.73 pts	1.94 ts	1.97 ts

TABLE IV. FUTUREMARK PCMARK 7 RESULTS ANALYSIS

FutureMark PC Mark 7	**Windows XP**	**Windows 7**	**Windows 8**
Video Playback	Not supported	24.0 fps	23.9 fps
Video Transcoding - Downscaling	Not supported	1855.4 kB/s	1740.1 kB/s
System Storage – Gaming	Not supported	3.63 MB/s	3.64 MB/s
Graphics – DirectX 9	Not supported	23.9 fps	24.1 fps
Image Manipulation	Not supported	10.8 Mpx/s	15.1 Mpx/s
System Storage – Importing Pictures	Not supported	5.36 MB/s	4.75 MB/s
Web Browsing	Not supported	9.13 pages/s	9.46 pages/s
Data Decrypting	Not supported	36.9 MB/s	64.2 MB/s
System Storage – Windows Defender	Not supported	1.1 MB/s	1.19 MB/s
FutureMark PC Mark 7 Score	**Not supported**	**1859**	**2036**

D. Sisoft Sandra Lite

The lite version of Sisoft Sandra software has the following benchmark capabilities:

- Processor Arithmetic: It is used to measure the arithmetic and floating point performance of processors.
- Processor Multi-Media: It is used to measure the media processing and multimedia performance of processors.
- Cryptography: It is used to measure the cryptographic performance of processors which includes checking for encryption/decryption, hash and sign capabilities.
- .NET Arithmetic and Multi-Media: It is used to measure the .NET Common language runtime arithmetic, floating point and media processing performance of processors.
- Memory Bandwidth: It is used to measure the performance of memory systems streaming capabilities.
- Cache & Memory Latency: It is used to measure the performance related to access time of the cache and memory systems.
- File System Bandwidth: It is used to measure the bandwidth performance of mounted file systems which are connected to storage adapters and hosts.
- File System I/O: It is used to measure the performance of I/O devices such as removable disks which are connected to ports and hubs.
- GP (GPU/CPU/APU) benchmarks: It is used to measure the computational, memory performance and cryptographic processing performance of general purpose (GP) GPUs, CPUs or other accelerators.

Sisoft Sandra Lite results show the overall score of the computer in the software itself and it can compare the results with similar systems for hardware performance differences. Table V shows the result of Windows OS benchmarks taken from the Sisoft Sandra Lite software the greater the value shows greater performance.

E. 3DMark 11

The basic version of 3DMARK 11 provides benchmarking for the graphics card. The basic version of 3DMARK 11 has the following benchmark capabilities:

- Ice Storm: It is used to measure the GPU and CPU performance by including two graphic tests on GPU and a physics test for CPU.
- Cloud Gate: It consists of the same tests as Ice Storm but Cloud gate uses a DirectX 11 engine limited to Direct3D feature level 10 in comparison to Ice Storm that uses feature level 9 thus cloud gate is suitable of testing DirectX 10 compatible hardware.
- Fire Strike: Fire strike is the most technical and ambitious DirectX 11 benchmark used to measure the real-time graphics rendered with detail and complexity far beyond any benchmarking suite. Fire strike is designed for high performance gaming PCs.

From the 3DMARK11 results, it shows that Windows 8 performs better overall but the breakdown of the scores can be seen in the Table VI. As per 3DMARK11 results, Windows 8 is clearly the winner in performing graphic intensive tasks whereas Windows XP lacks supports for DirectX11 which definitely is not good for PC games enthusiasts that still uses Windows XP as their hardware supports DirectX 11 but OS does not. In Table VI, the higher the scores mean the better performance. Note that the software does not install under Windows XP as the workloads require DirectX 11 features which Windows XP do not support so as far as this tool is considered. The comparison is based on Windows 7 and 8 as the previous version which supports Windows XP. These results are incomparable as the latest version does not support previous OS.

TABLE V. SSISOFT SANDRA LITE RESULTS ANALYSIS

Sandra Lite Benchmark	**Windows XP**	**Windows 7**	**Windows 8**
Processor Arithmetic	31.74 GPOS	30.71 GPOS	33.98 GPOS
Processor Multimedia	45.43 MPix/s	51.88 MPix/s	45.68 MPix/s
Cryptography	0.20 GB/s	0.22 GB/s	0.19 GB/s
.NET Artihmetic	12.04 GPOS	9.12 GPOS	8.91 GPOS
.NET Multimedia	10.26 MPix/s	9.28 MPix/s	8.06 MPix/s
Memory Bandwidth	8.63 GB/s	8.51 GB/s	8.81 GB/s
Cache and Memory Latency	48.00 ns	48.70 ns	48.30 ns
File System Bandwidth	49.91 MB/s	59.801 MB/s	43.755 MB/s
File System I/O	689.67 IOPS	597.60 IOPS	418.40 IOPS
GP (GPU/CPU/APU) Processing	108.92 MPix/s	109.80 MPix/s	110.52 MPix/s
GP (GPU/CPU/APU) Cryptography	1.011 GB/s	1.010 GB/s	1.013 GB/s
GP (GPU/CPU/APU) Bandwidth	4.30 GB/s	4.32 GB/s	4.35 GB/s
Overall Score	**2.09 kPT**	**2.12 kPT**	**1.95 PT**

TABLE VI. 3DMARK 11 RESULTS ANALYSIS

3DMARK	**Windows XP**	**Windows 7**	**Windows 8**
Ice Storm Graphics Score	Not supported	28582	55099
Ice Storm Physics Score	Not supported	29280	12541
Ice Storm Total	**Not supported**	**28734**	**31411**
Cloud Gate Graphics Score	Not supported	2682	5388
Cloud Gate Physics Score	Not supported	2374	1546
Cloud Gate Total	**Not supported**	**2606**	**3471**

Fire Strike Graphics Score	Not supported	353	669
Fire Strike Physics Score	Not supported	3449	2297
Fire Strike Total	**Not supported**	**333**	**634**

TABLE VII. NOVABENCH RESULTS ANALYSIS

NovaBench	**Windows XP**	**Windows 7**	**Windows 8**
RAM Score	120	120	120
CPU Tests	378	387	393
Graphics Tests	144	58	142
Hardware Tests	12	15	12
NovaBench Score	**654**	**580**	**667**

TABLE VIII. USER ORIENTED TASK BENCHMARKING SOFTWARE/APPLICATION

User Oriented Tasks Benchmarking		
Operation	**Methodology**	**Software/Technique Required**
Boot Time	Investigate the effectiveness of OS accessibility to user	Bootracer
Shutdown Time, Sleep, Wake-up Time	Investigate the effectiveness of OS in releasing resources	Manual Stopwatch
Encrypting File System	Investigate data recovery possibility in accidental situation	Virtualbox, Advanced EFS Data Recovery
CPU, GPU Processing Unit Temperature	Investigate temperature handling capabilities of the three OS	CPUID Hardware Monitor
Video Encoding	Investigate video and coding capabilities of the three OS	Avidemux
Data Compression	Investigate data archiving and extraction performance of the three OS	Winrar
Large File Transfers	Investigate data transfer speeds of the three OS	Windows Explorer

F. NovaBench

NovaBench is completely free and provides guideline results in less than a minute. This software is popular only because that it provides lightweight benchmark results for CPU, GPU and hard disk that later can be compared online with similar systems.

From the NovaBench results, it shows that Windows 8 performs better overall but the breakdown of the scores can be seen in the Table VII. As per NovaBench results, Windows 8 perform better in CPU tests whereas Windows XP excels in graphics and Windows 7 in Hard disk tests.

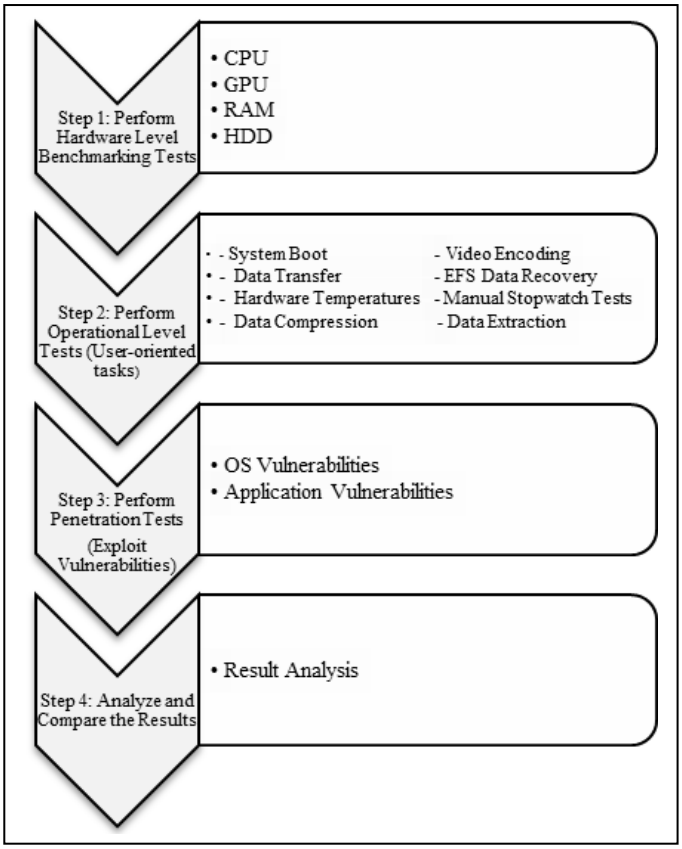

Fig. 1. Proposed Unified Approach Guidelines for Comparison of Operating System.

V. THE PROPOSED BENCHMARKING USING UNIFIED APPROACH

Based on the performance analysis of the available industry standard benchmarking software, it merely focus on hardware-level system performance but none of them measures the OS in terms of performing user-oriented tasks and security parameters altogether. A thorough comparison of Windows OS should not only focus on hardware level performance but must include operational level performance and security tests. Thus, we proposed a unified approach of benchmarking by adding user-oriented tasks performance testing and penetration tests on top of automated hardware-level benchmarking.

The examples of user-oriented tasks are startup time, shutdown time, wakeup time, video encoding performance, data compression performance and data transfer performance. Table VIII shows a list of related software/applications to test those tasks.

We have performed the testing and in summary, Windows 8 is the best in term of system boot, stopwatch testing, HDD temperatures and data extraction. While Windows XP performs the best in terms of video encoding, CPU and GPU temperatures, data compression and data transfer. Windows 7 and Windows 8 are tied in EFS data recovery.

Fig. 1 illustrates the guideline that can be followed for OS comparisons. There are four steps in the unified approach which are hardware-level testing to benchmark the

performance of CPU, GPU, HDD and RAM in Step 1, followed by user-oriented tasks testing in Step 2. Step 2 is operational level tests, which stress the system with the tasks a normal user would perform every single day. Then, in Step 3, penetration-testing tests the OS and applications vulnerabilities that find anything that would compromise the systems.

VI. CONCLUSIONS AND FUTURE WORK

From all the results gathered, it can be concluded automated benchmarking tools that runs on Windows XP are less efficient as it does not support DirectX 11 and other features, which the hardware could support. Thus, some software could not benchmark the OS comprehensively. Hence, to fulfill this, there is a need of unified approach to compare OS, which this research proposed on additional user-oriented tasks benchmarking and penetration testing to provide a complete comparison data.

The result does agree with the previous research that Windows 7 did not perform better than Windows XP in a high-end system. Thus, we can also conclude from this research that there are instances where Windows XP excels but overall Windows 8 is a superior OS to its predecessors, which performs better and provide more security on the same hardware.

For future work, this research could be enhanced to test vulnerability against rootkit and advanced vulnerabilities. The result also could be used as a base for new benchmarking software to incorporate security and user oriented tasks in their benchmarks to automate the process for a normal user. Furthermore, the guideline can served as benchmarking for new Windows OS.

ACKNOWLEDGMENT

The authors would like to thank colleagues and students for their help and support during the experiments and writing process and to the reviewers for their useful comments.

REFERENCES

[1] 3DMARK11 2013. Available at: http://www.3dmark.com/3dmark11/ [Accessed April 14. 2013].

[2] Adrian Kingsley-Hughes, "Microsoft Surface tablet pulling in PC-level license fees in a post-PC world" Available at: http://www.zdnet.com/microsoft-surface-tablet-pulling-in-pc-level-license-fees-in-a-post-pcworld-7000008909/ [Accessed April 14. 2013].

[3] Brandon LeBlanc, "Windows 8 – 40 Million Licenses Sold." Available at: http://blogs.windows.com/windows/b/bloggingwindos/archive/2012/11/27/windows-8-40-million-licenses-sold.aspx [Accessed April 14. 2013].

[4] Martinovic, G., Balen, J., & Cukic, B., "Performance Evaluation of Recent Windows Operating Systems." *Journal of Universal Computer Science,* vol. 18(2), pp.218–263, 2012.

[5] Maxon Cinebench 2013. Available at: http://www.maxon.net/products/cinebench/overview.html [Accessed April 14. 2013].

[6] NovaBench 2013. Available at: http://novabench.com/about.php [Accessed April 14. 2013].

[7] Pangaria, M., Shrivastava, V., & Soni, P., "Compromising Windows 8 with metasploit's exploit." Technology, 7(2), pp.119–123. 2012.

[8] PCMARK 7 2013 Available at: http://www.futuremark.com/benchmarks/pcmark7 [Accessed April 14. 2013].

[9] Schroeder, D. K. "MAC and PC Performance Analysis" (Doctoral dissertation, Big Bend Community College, Washington). 2008.

[10] Software Sandra lite 2013. Available at: http://www.sisoftware.net/?d=&f=downandbuy&l=en&a [Accessed April 14. 2013].

[11] Steigerwald, B., & Agrawal, Abhishek. "Developing Green Software". *Intel White Paper*. 2011.

[12] Suzanne Choney. "Windows 8 remain behind Vista in desktop OS market" Available at: http://www.nbcnews.com/technology/technolog/windows-8-remains-behind vista-desktop-os-market-1C9157764 [Accessed April 14. 2013].

[13] Ungureanu, M., "Benchmarking–Between Traditional & Modern Business Environment". *CES Working Papers*, (3), pp.440–451, 2011.