Journal of Universal Computer Science, vol. 18, no. 2 (2012), 218-263
submitted: 18/12/10, accepted: 25/1/12, appeared: 28/1/12 © *J.UCS*

Performance Evaluation of Recent Windows Operating Systems

Goran Martinovic
(Faculty of Electrical Engineering, Josip Juraj Strossmayer University of Osijek, Croatia
goran.martinovic@etfos.hr)

Josip Balen
(Faculty of Electrical Engineering, Josip Juraj Strossmayer University of Osijek, Croatia
josip.balen@etfos.hr)

Bojan Cukic
(Lane Department of Computer Science & Electrical Engineering, West Virginia University
Morgantown, WV, USA
cukic@csee.wvu.edu)

Abstract: The primary goal of most OSs (Operating Systems) is the efficient use of computer systems software and hardware resources. Since Windows OSs are most widely used OS for personal computers, they need to satisfy needs of all different kind of computer systems users. In comparison with Windows XP, new versions of the Windows OS; namely Windows Vista and Windows 7, introduce a number of new features and enhancements. Furthermore, performance improvement was imposed as one of the key design goals for both Windows Vista and Windows 7. This paper presents a performance evaluation of three latest versions of the Microsoft OS for personal computers; namely Windows XP, Windows Vista and Windows 7. OS performance measurement is done by means of a set of benchmark applications in the controlled environment. To ensure accurate, reliable and repeatable performance measurement results, we have created a performance measurement process and a performance evaluation model. Special emphasis is placed on evaluation areas with the greatest impact on the performance: CPU scheduling, memory management, graphic subsystem management, hard disk drive management and network performance. To determine the Windows OSs performance in different environments, performance measurement is done in three experiments. Experimental results indicate that Windows Vista and Windows 7 have several performance improvements on the stand-alone high-end computer system, but Windows XP outperforms Windows Vista and Windows 7 on the stand-alone low-end computer system. Furthermore, on network computer system Windows Vista and Windows 7 show network performance improvements mostly for the traffic with medium-sized packets.

Keywords: Benchmark, Operating System, Performance Evaluation, Performance Measurement, Windows 7, Windows Vista, Windows XP
Categories: D.4, D.4.8, K.6.2

1 Introduction

An OS (Operating System) extends the machine and gives programmers a simpler way to work with the hardware. Since it manages resources in time and space, programmers and users do not need to allocate system resources to their applications.

An OS that provides a more efficient use of software and hardware resources of a computer system achieves better performance. For each new version of an OS it is expected to have many new features and better performance than the last one. When Microsoft® released two last versions of the Windows® OS for personal computers, i.e. Windows Vista™ in 2007 and afterwards Windows 7™ in 2009, a significant amount of attention was given to their new features, as described in [Bott et al. 09, Hassel et al. 07]. However, performance evaluation and comparison to their predecessors were not conducted in the literature. Although one of the most important design goals for Windows Vista, besides new features and security, was the performance improvement, after it was released most complaints addressed its performance. One of the main reasons was that average computer systems were not powerful enough to take advantages of all Windows Vista features. In contrast to the aforementioned, upon its release, the performance improvement was the most frequently mentioned and praised characteristic of Windows 7. Meanwhile, eternal question appeared among users: “Which Windows OS for personal computers has the best performance?” Furthermore, various comparisons of different versions of the Windows OS for personal computers have emerged on the Internet [see Schmid 07, Smith 09a, Smith 09b, Williams 09]. However, from these performance measurements and comparisons a representative conclusion cannot be drawn since they do not have clear and unified performance measurement process and evaluation methodology. Furthermore, in the literature there are only a few scientific studies on the Windows OS performance, its evaluation and comparison with different versions of the Windows OS for personal computers. Consequently, there is no unified or standard method, process or approach for the Windows OS performance measurement and evaluation. Therefore, the main motivation in this paper is to develop a model for the Windows OSs performance measurement, evaluation and comparison that will be unified and that can be reused. Furthermore, the goal is to use this model for the performance evaluation of the last three versions of the Windows OSs for personal computers.

This paper presents performance evaluation of the three latest 32-bit versions of the Windows OS; namely Windows XP® Professional SP3 (Service Pack 3), Windows Vista Business SP2 and Windows 7 Professional. The main goal is to determine which of these last three versions of the Windows OS handles system resources more efficient, and consequently has better performance. The major problem of Windows OSs performance measurement and evaluation is that there are no benchmark applications whose purpose is to directly measure the Windows OS performance. There are many benchmark applications that measure the computer system performance, but they are mostly used for measuring the computer system hardware performance. However, in a controlled environment, these benchmark applications can be used for measuring the performance of different Windows OSs. We have chosen seven different benchmark applications available in [Niemela 05, Renquist et al. 06, SPEC 09, ScienceMark 06, Lavalys 09, Xtreme 08, Botta et al. 07]. First six benchmark applications are used for measuring the performance of Windows OSs in working with hardware components that have most impact on the performance of the stand-alone computer system: the memory, the CPU (Central Processing Unit), the graphics subsystem and the HDD (Hard Disk Drive). The seventh benchmark application is used for measuring the network performance of Windows OSs. To ensure reliable and consistent performance measurement results we have created a performance measurement process. An OS performance evaluation can be performed with techniques that include designing an experiment, analytical modeling or simulation [see Fortier et al. 03]. We have created a performance evaluation model that has a large number of performance indicators comparable between different Windows OSs. To determine the Windows OSs performance in different environments, the performance measurement is done in three experiments. In Experiment 1, a stand-alone high-end computer system with more recent hardware is used. Windows Vista and Windows 7 are expected to take advantage of numerous new features, enhancements and optimization mechanisms and to have better performance measurement results since this computer system satisfies all their needs for hardware resources. In Experiment 2, an older stand-alone low-end computer system is used, and due to its poor hardware resources, Windows Vista and Windows 7 are expected to have lower performance than Windows XP. In Experiment 3 two identical network computer systems were used. The goal is to determine if newer version of Windows OSs bring performance improvements in area of networking.

This study will discover if Windows Vista and Windows 7 deliver performance improvements compared to Windows XP, and which Windows OS is preferable for running on stand-alone computer system (on the low-end and on the high-end, respectively) and which on the network computer systems. Although Windows XP is becoming obsolete these days, there are still a huge number of computer systems that are using Windows XP. The main advantages of the Windows XP are that it is old and therefore reliable and stable, and it requires less hardware resources than Windows Vista or Windows 7. According to the Net Application statistics [Net Applications 11] shown in [Fig. 1], in October 2011 Windows XP was the leading OS on desktop computers with 45.23% of the total share. However, falling trend is obvious. Since Windows Vista also shows falling trend and Windows 7 shows strong rising trend, it can be concluded that a vast majority of Windows XP users are migrating to Windows 7.

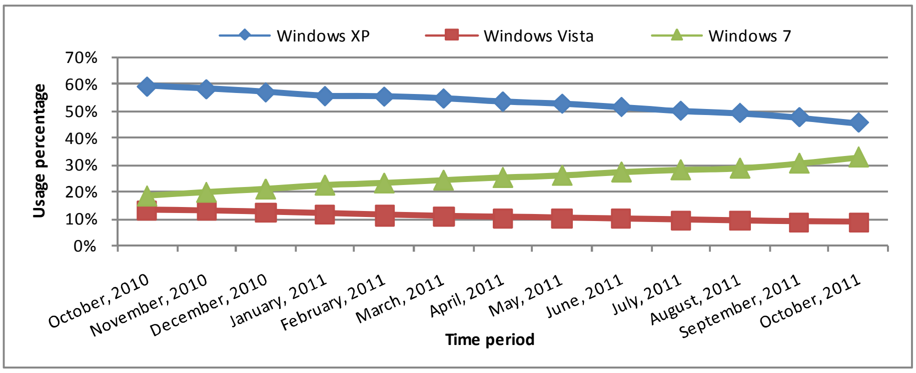

Figure 1: Windows OSs share trend from the October 2010 to the October 2011

The remainder of the paper is organized as follows. In Section 2, a survey on related work is presented. Section 3 presents an overview of Windows XP, Windows

Vista and Windows 7 OSs. In Section 4, benchmark applications are presented. The performance measurement methodology is described in Section 5. Section 6 depicts performance measurement results, which are discussed in Section 7. Section 8 leads to the conclusion.

2 Related work

Performance evaluation of different Windows OSs is complex to perform because many internal, constructional or external factors have influence on the Windows OS performance. However, in the literature several segments of different Windows OSs were benchmarked. In [Martinovic et al. 10], we studied how different host OSs influence virtual machine performance. Windows XP, Windows Vista and Windows 7 were used as host OSs, while Windows Vista was used as a virtual OS. Virtual OS performance evaluation was done in the same controlled conditions for all three host OSs using five different benchmark applications and by performing two resources demanding operations: video encoding and data compression. Performance measurement results show that the virtual OS (Windows Vista) has the best performance when Windows 7 is used as the host OS. In [Pfeiffer 07], Windows Vista and Windows XP were included in a UIF (User Interface Friction) benchmark and results show that Windows Vista fared less well than Windows XP. In [Kalakech et al. 04], authors benchmarked dependability of Windows NT4, Windows 2000 and Windows XP. They showed that these three versions of the Windows OS are equivalent from the robustness point of view and that Windows XP has the shortest reaction and restart time.

There exist a handful research articles dealing with the network performance of different OSs. Performance evaluation of Windows 2003, Windows XP and Windows Vista in wireless LAN IEEE 802.11g environment is performed in [Kolahi et al. 08]. The authors showed that Windows Vista has lower performance in terms of the bandwidth and the round trip time than the other two Windows OSs. In [Salah et al. 09] packet-forwarding performance on the kernel and user level of Linux, Windows Server and Windows XP was compared. Performance measurement results indicate that Linux has the best packet-forwarding performance on the kernel level in terms of throughput, packet loss and delay. However, Windows Server had smallest delays and highest throughput when forwarding packets on the user level. In [Salah et al. 10a] authors continued their work and evaluated the impact of the running CPU-bound application on IP forwarding (kernel level forwarding) in Linux and Windows XP. They concluded that in Linux IP forwarding is not affected by running CPU-bound application, whereas Windows XP network performance is degraded in terms of throughput, packet loss and delay. Network performance evaluation comparison of Snort NIDS (Network Intrusion Detection System) under Linux and Windows Server is performed in [Salah et al. 10b]. Results show that Linux obtains better performance gain for Snort under malicious traffic and Windows Server shows better performance for Snort under normal traffic. In [Narayan et al. 09], IPv4 and IPv6 performance for TCP (Transmission Control Protocol) and UDP (User Datagram Protocol) traffic has been compared on Windows Vista and Linux Ubuntu. Compared to Linux Ubuntu, Windows Vista shows lower throughput and delay, and higher CPU usage. Furthermore, in [Narayan et al. 10] previous work was continued and network performance evaluation of five versions of Windows OS was conducted with the similar experimental setup as in [Narayan et al. 09]. Authors concluded that there is no clear winner, however in most measurements Windows Server 2003 shows worse performance when compared to other. For TCP traffic the average difference between versions range from 2% to 5%, whereas for UDP traffic is around 3%.

3 Windows OSs

As described in [Tanenbaum 08], an OS has two major roles. First, it provides abstraction of raw hardware devices for application programs. Second, it manages system resources like the memory space, the CPU time, graphics resources, the file-storage space and I/O (Input/Output) devices. The primary goal of most OSs is the efficient execution of user applications. An OS that manages computer system resources better will consequently have better performance. OS structure and source code have a great impact on the performance, but an OS behavior depends on actual workloads and external requirements [see Fortier et al. 03, Joukov et al. 06].

General purpose OSs, such as Windows XP, Windows Vista and Windows 7, work with business and desktop applications and they are designed for a wide range of users. Besides personal computers, they find their place in embedded computer systems.

3.1 Windows XP Performance Techniques

Compared to previous versions of the Windows OS, Windows XP design goals included improved security and reliability, Windows and POSIX (Portable Operating System for UNIX) application capability, high performance, extensibility, portability and international support [see Silberschatz et al. 05]. One of the most important design goals of Windows XP was high performance. Windows XP improved the performance by reducing the code-path length in critical functions, using better algorithms and per-processor data structures, using memory coloring for NUMA (Non-Uniform Memory Access) machines, implementing more scalable locking protocols, optimizing threads priority and enabling symmetrical multiprocessing. As described in [Martinovic et al. 07], Windows XP is very good as a soft real-time OS that can tolerate some lateness.

3.2 Windows Vista Enhancements and Comparison with Windows XP

Compared to Windows XP, Windows Vista design goals included advanced security, enhanced performance and reliability, improved graphic-user interface, enhanced power state transitions, and introduction of new features like adjusting to user behavior and solving problems automatically. One of the most important design goals was performance improvement without increasing hardware requirements. The idea was to improve the responsiveness of the OS and to optimize memory management. Windows Vista added numerous enhancements to processes, threads and sections. Each process has one or more threads, which make the basic executable unit and are dispatched by the kernel. Both Windows XP and Windows Vista use a priority algorithm with 32 priority levels where the highest priority thread is always running [see Tanenbaum 08]. However, [Fig. 2] shows the CPU time accounting in Windows XP, which does not provide a fair thread scheduling [see Russinovich 07a].

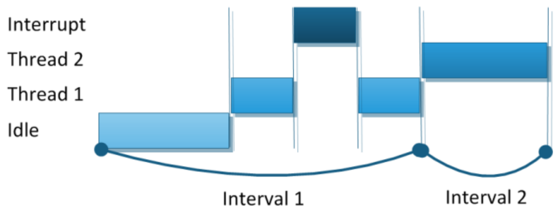

Figure 2: Windows XP thread scheduling

In Windows Vista, the scheduler uses a cycle time counter for fair CPU scheduling, as shown in [Fig. 3]. The thread gets at least one turn without counting the interrupt time [Russinovich 07a]. The CPU is one of the most important resources of the computer system and the OS's CPU scheduling policy has a huge influence on the OS overall performance. Therefore, in measurements that require significant CPU time, Windows Vista should show better results.

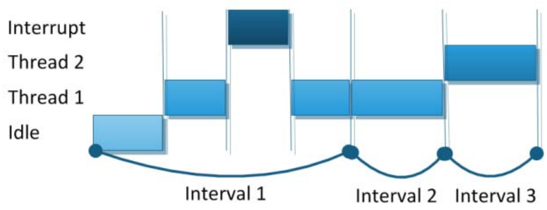

Figure 3: Windows Vista thread scheduling

According to [Peng et al. 07], memory management affects the OS performance more significantly (in terms of the throughput, response time, workload, etc.) than does CPU management. Memory management in Windows Vista is completely changed in comparison to Windows XP, and it has numerous improvements, like the dynamic kernel address space, memory priorities, enhanced NUMA and large page support, paging video memory, improved I/O bandwidth, section access, robustness and self-diagnostics [Russinovich 07b]. These improvements should result in faster memory read/write/copy operations and shorter memory latency.

In addition to the aforementioned improvements, Windows Vista presented completely new features that use a new technology and should enhance the performance of the memory management: SuperFetch™, ReadyBoot, ReadyBoost™ and ReadyDrive™.

SuperFetch [see Russinovich 07b] improves management of the physical memory. Windows XP introduced prefetching support by preloading memory with data based on previous boots and application launches. However, Windows Vista goes a step further and takes into account the frequency of page usage and usage of page in the context of other pages in the memory. Therefore, memory manager logs user activities and preload pages based on the actual usage. These changes should result in a faster application launch, faster resume from hibernate and suspend as well as better performance. To see this improvement in performance measurement results SuperFetch should be "trained" in this way to reboot the computer system several times and after every reboot benchmark application must be started. ReadyBoot analyzes the boot process after every boot and calculates a boot-time caching plan for the next boot. On computer systems with less than 700MB of physical memory Windows Vista uses the same boot prefetch as in Windows XP. According to [Russinovich 07b], Microsoft performance tests show that ReadyBoot provides performance improvements of about 20% in comparison with the Windows XP prefetcher. Apart from ReadyBoot, startup has other improvements such as a new boot mechanism that is platform independent and it is called BCD (Boot Configuration Database), a new flow and organization of system startup processes, a new logon architecture, and support for delayed-autostart services.

ReadyBoost [see Russinovich 07b] improves the performance by using a USB (Universal Serial Bus) flash disk or a SSD (Solid State Drive) exposed by ITM (Intel® Turbo Memory) as a logical extension of the system DRAM (Dynamic Random Access Memory) for system caching, as described in [Matthews et al. 08]. The maximum size that can be used for caching is 4 GB [Hargreaves et al. 08]. The reason for using ReadyBoost is that a USB flash disk or the SSD have several times faster random accesses than a typical HDD. ReadyBoostis expected to improve the memory and HDD performance in Experiments 1 and 2.

ReadyDrive [see Russinovich 07b] enables usage of new hybrid HDDs with extra flash memory onboard. In addition to hybrid HDDs, ITM enables the computer system to support ReadyDrive for disk caching on an internal PCI-Express (Peripheral Component Interconnect) device. ReadyDrive speeds up the computer system and saves the power. Windows Vista also introduces support for HDDs with a larger physical sector size. Windows XP supports only HDDs with a physical sector size of 512 bytes, whereas Windows Vista supports HDDs with the sector size of 1 KB, 2 KB or 4 KB. This Windows Vista feature should provide better performance, capacity and reliability for HDDs. However, components such as the hybrid HDD, ITM and a large sector HDD are not included in the hardware configuration in experiments.

One of the most important functions of desktop computer systems for average users is the multimedia reproduction. As described in [Russinovich 07a], new MCSS (Multimedia Class Scheduler Service) boosts thread priorities of multimedia applications and therefore, they can deliver glitch-free audio and video streaming in Windows Vista. Furthermore, Windows Vista included a new display driver architecture called WDDM (Windows Display Driver Model) that offers users better performance, stability and security. Most of WDDM have been moved from the kernel space to the user space. In Windows XP, display drivers execute entirely in the kernel mode and if a single problem occurs, the OS may crash. Moreover, DirectX® 10 was launched only for Windows Vista. Compared to DirectX 9 used in Windows XP, it has more resources and it heavily enhances 3-D graphics-rendering capabilities by lowering command cycle counts per frame and allowing the graphics processing unit to render more complex scenes without assistance from the CPU, as described in [Blythe 06]. Therefore, it is expected that Windows Vista will show better graphic performance in Experiments 1 and 2.

As described in [Russinovich 07a], I/O completion processing in Windows Vista does not have to be done with the thread that issued the I/O request, as in Windows XP. It can be done with a different thread. This new behavior of an asynchronous I/O operation should enhance the overall application and system performance.

3.3 Windows 7 Enhancements

Windows 7 is built on the Windows Vista code base and therefore, it maintains applications compatibility. Furthermore, all features from Windows Vista are retained in Windows 7 and most of them are enhanced in order to get better system performance, power efficiency, scalability and responsiveness. As described in [Microsoft 09], the underlying design goal for Windows 7 was the performance improvement in key user scenarios with focus on the user responsiveness. Most important enhancements that have influence on various aspects of the performance are described in [Microsoft 10]:

- ReadyBoost changes include support for the caching pagefile-backed pages, concurrent use of multiple flash devices and support for 32 GB cache.
- ReadyBoot is improved by using the compression and reducing memory footprint. The increased parallelism of driver initialization and improved prefetching logic and mechanisms should result in a faster boot process.
- Memory manager improvements include enhanced page replacing policy, which uses the 3-bits value (in Windows Vista this is a 2-bits value) to decide which page to leave in the process working set. System cache, paged pool, and pageable system code now each have their own working set (in Windows Vista they are in the same working set). Also, registry operations are enhanced by removing memory mapping.
- Improved DWM (Desktop Window Manager) has up to 50% smaller memory consumption per window.
- SMT (Simultaneous MultiThreading) is enhanced by using SMT Parking as a further guide for avoiding use of logical pairs. Since it enables better performance on hyper-threaded, multi-core Intel processors, it should improve the Windows 7 performance in Experiment 1.
- Kernel dispatcher lock is removed for some operations and for others it is replaced by a set of synchronization techniques and a new “pre-wait” thread state. The main benefit is that Windows 7 can scale up to 256 processors.
- UMS (User Mode Scheduling) improves the performance by separating a user-mode thread and a kernel-mode thread. User threads can be scheduled in the user-mode without kernel transition.
- Scalability for applications that manage large amounts of memory is improved by removing the memory manager PFN (Physical Frame Number) global lock.
- UBPM (Unified Background Process Manager) is a trigger-based component that starts or stops services on a certain event. It improves the performance by minimizing the number of background running services.

- DirectX 11 improves the scalability and the performance by introducing new features such as the tessellation, multithreading, dynamic shader linkage, improved texture compression and compute shader.
- Core Parking improves power efficiency by dynamically selecting a set of processors (sockets) that should stay idle, based on their recent utilization and by keeping load on as less processors as possible.
- Timer management API (Application Programming Interface) has two improvements that reduce power consumption. The first is ITTD (Intelligent Timer Tick Distribution), which does not permit timer interrupts on the processor that are idle, and the second is Timer coalescing, which coalesces different applications timer notifications together.
- MinWin is an independent bottom part of Windows 7 that consists of a kernel, HAL (Hardware Abstraction Layer), TCP/IP (Transmission Control Protocol/Internet Protocol), file systems, drivers and core system services. The main benefit is that it can be built, booted and tested separately from the rest of the system.

4 Benchmark Applications

Since there are no benchmark applications that can directly measure, describe and evaluate the performance of different Windows OSs, we have scrutinized a large number of benchmark applications for measuring the computer systems performance. Afterwards, we have selected seven benchmark applications that are most popular, most interesting and most extensive when compared with others: PCMark05, 3Dmark06, SPECviewperf, ScienceMark, Everest, SuperPI and D-ITG. Three benchmark applications, i.e.PCMark05, 3Dmark06, and SPECviewperf, create challenging real world workloads (video animations) while measuring the computer system performance. PCMark05 is one of the most widely used benchmark applications among computer users. Since it measures performance of the CPU, the memory, the graphics subsystem and the HDD, it provides the most complete picture of the computer system performance. 3Dmark06 and SPECviewperf represent a reference in graphics benchmarking and provide the comprehensive graphics subsystem performance measurement. ScienceMark and SuperPI perform mathematical calculations that require a huge amount of CPU resources. Besides, ScienceMark will show which Windows OS has best performance when performing scientific calculations. Everest is used for measuring the raw hardware performance of the cache memory, main memory and HDD. D-ITG is a network performance measurement tool that is capable to produce different types of traffic and to measure network performance related metrics.

Since Windows OSs are very complex, many various factors have influence on the performance. However, selected benchmark applications provide a large number of measured values that represent a real world performance of Windows OSs and show how efficiently they assign software and hardware resources to the benchmark applications. These values are classified and used in performance evaluation modeling. In Experiment 1 and Experiment 2 special emphasis is placed on evaluation areas with the greatest impact on the performance of standalone computer systems: CPU scheduling, memory management, graphic subsystem management and HDD management. To eliminate potential errors, each area is covered with at least two different benchmark applications. Selected benchmark applications measure particular and overall performance of computer systems and therefore, given results will be useful to different types of computer system users. Furthermore, every benchmark application has different performance measurement logic and this also contributes to the completeness of measurement results. In Experiment 3 special emphasis is placed on the network performance.

4.1 Futuremark PCMark05 v1.2

PCMark05 [see Niemela 05] is composed of five different testing suites and four of them are used in experiments: memory, CPU, graphics, and HDD test suite. The memory test suite measures the performance of the main memory, the CPU internal (L1 cache) and the external cache (L2 cache) while reading, writing and copying data blocks and their latency. Furthermore, during the execution of the memory test suite, following parameters were also measured: PageFaults/s, CacheFaults/s and PageReads/s. The CPU test suite isolates performance of the CPU. Furthermore, two of the test scenarios are run multithreaded: the first includes two simultaneous tests, and the second runs four tests simultaneously. The remaining six tests are run single threaded. The Rijndael/AES Algorithm is used for the file encryption and decryption task. The graphics test suite measures 2D and 3D graphics performance. The HDD test suite measures the data throughput for five different purposes. The workload is designed to stress the personal computer in the same manner as typical home usage does. PCMark05 is used with default settings.

4.2 Futuremark 3DMark06 v1.1

3DMark06 [see Renquist et al. 06] is a popular application for 3D graphics benchmarking. It focuses on the performance of the graphics subsystem and the CPU with a set of four graphics tests and two CPU tests in advanced real-time 3D rendering of a 3D scenario. The first two graphics tests require the graphics subsystem support for Shader 2.0 and the other two use complex shaders and HDR (High Dynamic Range) rendering. Since CPU tests are optimized for multi-core processors, Experiment 1 will show which OS has better CPU management on multi-core processors. Feature tests that isolate performance of some key 3D features are also included in the performance evaluation. Due to insufficient graphics resources of the computer system in Experiment 2, 3DMark06 performance measurements cannot be conducted. Display resolution in all measurements is 1280x1024 pixels and 3DMark06 is used with default settings.

4.3 SPECviewperf 10

SPECviewperf [see SPEC 09] is a synthetic benchmark designed to be a predictor of application performance and a measure of graphics subsystem performance. It measures 3D graphics performance of systems running under the OpenGL application programming interface. The benchmark test files, called viewsets, are developed by tracing the graphics content from actual applications. Current viewsets represent graphics functionality in 3ds max, CATIA, EnSight, Maya, Pro/ENGINEER, SolidWorks, UGS Teamcenter Visualization Mockup and UGS NX. Display resolution in all measurements is 1280x1024 pixels. Results for each viewset are calculated using the weighted geometric mean W_G as a single composite metric, as shown in (2).

4.4 ScienceMark 2.0

ScienceMark 2.0 [see ScienceMark 06] is based on scientific calculations used in theoretical scientific and engineering computing. It is comprised of seven benchmarks that test various aspects of a CPU and support multi-processor systems. Execution time of the following three benchmarks is measured in experiments: MolDyn, Primordia and Cipher. MolDyn simulates thermodynamic behavior of materials using their forces, velocities and positions. Primordia calculates Quantum Mechanical Hartree-Fock orbitals for each electron in any element of the periodical table. Element Ag (Argentum) was used in experiments. Number crunching involved in the calculation is complex and it stresses the FPU (Floating Point Unit) of a CPU. Cipher tests four different forms of encryption algorithms, AES (Advanced Encryption Standard) 128/256-bit and RSA (named after Rivest, Shamir and Adleman) 512/1024-bit. The calculations involved are mathematically inclined, resulting in stressing the ALU (Arithmetic and Logic Unit) of the CPU. All benchmarks are used with default simulation options.

4.5 Everest Ultimate v5.02

Everest [see Lavalys 09] is a system diagnostics and benchmark application. A memory benchmark module is used for measuring read, write and copy throughput and a latency of the main memory and L1, L2 and L3 cache memory. Furthermore, a HDD benchmark module is used for measuring the HDD read performance. Everest will show how different Windows OSs influence on raw hardware performance.

4.6 Super PI Mod 1.5

Super PI [see Xtreme 08] is a single threaded application that calculates number PI to a specified number of digits after the decimal point, up to a maximum of 32 million. It uses the Gauss-Legendre algorithm for calculation that overloads memory and CPU, and allows testing of the FPU and mathematics performance of the CPU. In experiments, times needed to calculate 4 million and 16 millions digits are measured.

4.7 D-ITG 2.6.1d

D-ITG (Distributed Internet Traffic Generator) is a platform capable to generate traffic at network, transport, and application layer [Botta et al. 07]. It supports both IPv4 and IPv6 traffic generation, as well as various protocols such as TCP, UDP, ICMP (Internet Control Message Protocol), DNS (Domain Name System), Telnet and VoIP (Voice over Internet Protocol). Traffic can be produced with various packet sizes and a variety of probability distributions like constant, uniform, exponential, Pareto, Cauchy, normal, Poisson and gamma. It can measure one-way-delay, round-trip-time, packet loss, jitter and throughput between two separate components called ITG-Send and ITG-Receive.

5 Performance Measurement Methodology

5.1 Hardware Impact on Performance Measurement Results

This section describes how hardware can have an impact on the performance and any special considerations for the performance measurement of Windows OSs, as described in [Microsoft 07, Microsoft 09, Secherest et al. 01].

RAM: The computer system in Experiment 1 has 6 GB of physical memory. However, performance measurements are conducted on 32-bit OSs and only 4 GB can be used. Since Windows 7 requires at least 1GB of memory, a low-end computer system in Experiment 2 has exactly this amount of memory. Performance measurement results will demonstrate whether memory management improvements in Windows Vista and Windows 7 are scalable and how they are affected by different amounts of the physical memory.

CPU: Recommendation from the literature is to include multi-core and 64-bit processors in performance measurement, as we did in Experiment 1. In Experiment 2 performance measurements were conducted on a single core 32-bit processor. This contrast will show how different processor architectures affect different Windows OS performance. Furthermore, when mobile processors are not on the AC power, they lower their performance to save energy, but in our experiments we use only desktop computer systems and therefore this consideration is not taken into account.

Graphics Hardware: Due to the fact that the number of monitors and their resolution can affect the performance, we used only one 19" monitor with the same resolution in all experiments (1280x1024 pixels). Since in Experiment 2 the graphics hardware does not have its own dedicated physical memory and it uses slower system memory, Windows Vista and Windows 7 are expected to have no benefits from their new display architectures. In Experiment 1, the graphics hardware uses its own dedicated physical memory.

HDD: Windows Vista and Widows 7 have better performance when working on hybrid disks with Windows ReadyDrive technology. Furthermore, a bigger physical sector size of the HDD can also improve performance of Windows Vista and Windows 7. However, in our experiments we use HDDs with the physical sector size of 512 bytes since Windows XP does not support larger physical sector sizes. Moreover, the HDD firmware can also affect the performance and therefore it must be updated to the latest version.

Flash memory: ReadyBoost devices like flash memory devices or Secure Digital cards can improve the system responsiveness in Windows Vista and Windows 7. In experiments, we use flash memory device with 4 GB assigned to the ReadyBoost feature.

NIC (Network Interface Controller): Recommendations from the literature [see VMware 07] indicate that all network infrastructure should be appropriately rated. For example, to use appropriately gigabit NIC, other infrastructure must also be gigabit capable e.g. switch, hub and cable. Since most modern NICs can operate in multiple modes (such as 10, 100, or 1000Mbps, half duplex or full duplex), they must be configured to work at their maximum possible bandwidth. To avoid cross-traffic noise over the network while conducting the experiments, direct cables between the computer systems or a private network switch should be used. For the connection, NICs on both computer systems must be dedicated. Furthermore, similar network interface controllers should be used on computer systems under test so that they function well with each other. Using similar or identical NICs also helps to ensure that send and receive have similar performance. Ideally, similar or identical client and server computer systems should be used as well, with similar system bus architectures and configurations. The differences between PCI, PCI-X, and PCIe, for example, can have a significant effect on the networking performance. Therefore, to avoid problems of hardware incompatibility or lowering performance when using different hardware that might not function very well with each other we are using two identical computer systems for measuring network performance.

5.2 Performance Evaluation Model

In the performance evaluation, we use representative metrics that captures best the capabilities of different Windows OSs. Since there is no performance evaluation model for different versions of the Windows OS, we created a model that comprises five major parts of the Windows OS that have most impact on performance: memory management, CPU scheduling, the graphic display, HDD management and network performance.

The performance evaluation requires performance indicators that are measurable, independent and comparable between different versions of the Windows OS. Similarly to [Lilja 00], three different groups of performance indicators that best represent an overall OS performance are defined:

- v - Throughput (rate metric).
- t - Execution time.
- W_G - Weighted geometric mean.

Throughput v represents the quantity of the measured value in the observed time interval. It is calculated by dividing the quantity N by the length time of measurement, as shown in (1):

$$v = \frac{N}{T} \quad (1)$$

In order to standardize and simplify comparison, the length of the observed time interval T is one second. A throughput unit depends on the type of the measured value. A larger number means a higher throughput and better performance. Execution time t refers to the time required to execute a given benchmark application or the length of latency time. It is expressed in nanoseconds (ns), ms (milliseconds), seconds (s) or minutes (min), and the shorter the time, the better the performance. The weighted geometric mean W_G is calculated from various individual tests for each viewset of SPECviewperf 10, as shown in (2), where x_f is the number of frames, t is execution time in seconds, n is the number of individual tests in a viewset, and w

is the weight of each individual test. A weighted geometric mean unit is frames per second, and the higher the number, the better the performance.

$$W_G = \prod_{i=1}^{n} \left(\frac{x_f}{t_i} \right)^{w_i} \quad (2)$$

All performance indicators derived from these three groups are explained in [Tab. 1]. Comparison of performance indicators will determine which OS provides a more efficient use of software and hardware resources and/or it exhibits the following:

- More efficient algorithms for resources management (the memory, the CPU, the HDD and the graphics subsystem).
- More efficient restrictions of using the memory and the HDD space.
- Faster working with registry and other components of the OS.
- Better implementation of the file system.
- More efficient display driver architecture.
- More efficient executions of network bound applications.

Performance indicators	**Explanation**	**Benchmark application**
Graphics display		
v_{Gts}	Graphics test suite	PCMark05
v_{Gs2}	Graphics tests of Shader 2.0 (only in Experiment 1)	3DMark06
v_{Ghdr}	Graphics tests of complex shader and HDR (only in Experiment 1)	3DMark06
v_{G3d}	Graphics tests of 3D features (only in Experiment 1)	3DMark06
W_{G3m}	3ds max viewset	SPECviewperf
W_{Gca}	CATIA viewset (only in Experiment 1)	SPECviewperf
W_{Ges}	EnSightviewset	SPECviewperf
W_{Gmy}	Maya viewset (only in Experiment 1)	SPECviewperf
W_{Gpe}	Pro/ENGINEER viewset (only in Experiment 1)	SPECviewperf
W_{Gsw}	SolidWorksviewset	SPECviewperf
W_{Gut}	UGS Teamcenter Visualization Mockup viewset	SPECviewperf
W_{Gun}	UGS NX viewset (only in Experiment 1)	SPECviewperf
Network performance		
t_{Dnet}	Network delay	D-ITG
t_{Jnet}	Network jitter	D-ITG
v_{Tnet}	Network throughput	D-ITG
p_{CPUnet}	CPU usage (while executing network measurements)	Performance monitor (Win)

Table 1: Performance indicators of Windows OSs

Performance indicators	**Explanation**	**Benchmark application**
Memory management		
v_{Mrd}	Memory read	PCMark05
v_{Mwr}	Memory write	PCMark05
v_{Mcp}	Memory copy	PCMark05
v_{Mla}	Memory latency	PCMark05
v_{MErd}	Memory read	Everest
v_{MEwr}	Memory write	Everest
v_{MEcp}	Memory copy	Everest
v_{L1rd}	L1 cache read	Everest
v_{L1wr}	L1 cache write	Everest
v_{L1cp}	L1 cache copy	Everest
v_{L2rd}	L2 cache read	Everest
v_{L2wr}	L2 cache write	Everest
v_{L2cp}	L2 cache copy	Everest
v_{L3rd}	L3 cache read (only in Experiment 1)	Everest
v_{L3wr}	L3 cache write (only in Experiment 1)	Everest
v_{L3cp}	L3 cache copy (only in Experiment 1)	Everest
t_{Me}	Memory latency in Everest	Everest
t_{L1}	L1 cache latency in Everest	Everest
t_{L2}	L2 cache latency in Everest	Everest
t_{L3}	L3 cache latency in Everest (only in Experiment 1)	Everest
c_f	Memory cache fault	Performance monitor (Win)
p_f	Memory page fault	Performance monitor (Win)
p_r	Memory page read	Performance monitor (Win)
CPU management		
v_{CPUts}	CPU test suite	PCMark05
v_{CPU1}	CPU1 test (only in Experiment 1)	3DMark06
v_{CPU2}	CPU2 test (only in Experiment 1)	3DMark06
t_{CPUmd}	Calculation time of MolDyn benchmark	ScienceMark
t_{CPUpr}	Calculation time of Primordia benchmark	ScienceMark
t_{CPUch}	Calculation time of Cipher benchmark	ScienceMark
t_{CPUs4}	Calculation time of four million digits of number PI	Super PI
t_{CPUs16}	Calculation time of sixteen million digits of number PI	Super PI
HDD management		
v_{Hts}	HDD test suite	PCMark05
v_{Hrd}	HDD read test suite	Everest
t_{Hrd}	HDD average read access	Everest

Table 1 (cont.): Performance indicators of Windows OSs

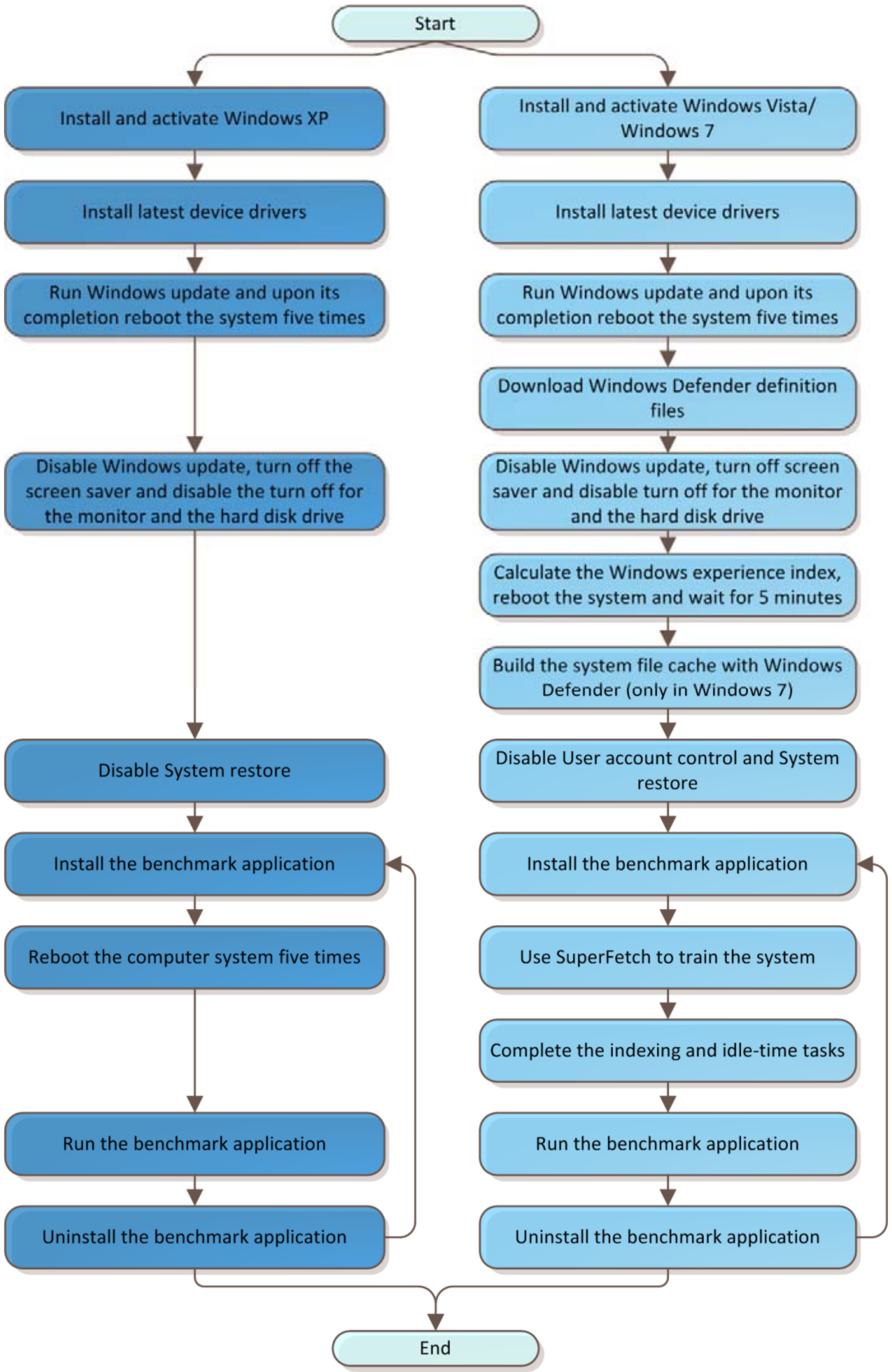

Figure 4: Performance measurement process for Windows OSs

5.3 Performance Measurement Process

[Fig. 4] shows the performance measurement process for Windows XP, Windows Vista and Windows 7 that ensures accurate, repeatable, reliable and consistent performance measurements and results. The proposed process is based on guidelines from [Microsoft 07, Microsoft 09, Secherest et al. 01]. Due to the fact that Windows Vista and Windows 7 attempt to improve own performance over time based on the observed usage patterns, their performance measurement process is more complex and it differs from the Windows XP process. In order to SuperFetch perform the appropriate prefetching and to ensure consistent workload measurements, Windows Vista and Windows 7 systems must be “trained”. Prior to performance measurement, every benchmark application must be started for five times and after every running, system must be rebooted. Furthermore, since post-boot activities in Windows Vista can impact performance measurement results, benchmark application must not be started ten minutes after boot.

6 Experimental Results

6.1 Experimental Setup

Performance measurement is performed on two desktop computer systems on similar editions of Windows XP, Windows Vista and Windows 7. There are five main editions of Windows XP: Home, Professional, Media Center, Tablet PC and Professional x64 edition [see Russinovich et al. 04]. Since performance measurement is performed on the desktop computer system, a 32-bit edition of Windows XP Professional with SP3 was selected. It is targeted at power users, businesses and enterprise clients, and it has extra features and better support than other editions. Furthermore, there are four main editions of Windows Vista: Home Basic, Home Premium, Business and Ultimate edition. Upon exploration of all editions of Windows Vista, a 32-bit Business edition with SP2 was selected. It is also intended for business clients and it includes all features the Windows XP Professional edition has. Windows 7 is available in four different editions: Starter, Home Premium, Professional and Ultimate edition. Since a 32-bit edition of Windows 7 Professional represents the best match to the Professional edition of Windows XP and the Business edition of Windows Vista, it was selected for the performance measurement. As Windows 7 has been recently released, it does not have any SP. For Windows XP and Windows Vista the latest SPs are used since they include updates, hotfixes and enhancements that improve performance, security, and stability of Windows OSs.

Since the location of system files and the number of volumes on the HDD can affect performance measurement results, only one volume on the HDD is created in all experiments. All OSs are installed with default settings. After installation, OSs are updated through Windows update with the latest available updates. The listing of Windows updates for all OSs in all three experiments can be found in the Appendix. Windows XP does not contain device drivers for some hardware components but Windows Vista and Windows 7 have all necessary device drivers built-in installation. However, the latest device drivers for each hardware component in each OS are installed and the listings of installed device drivers can be also found in the Appendix

[Tab. 18, Tab. 20]. Device drivers can impact the overall Windows OSs performance but due to a large number of different hardware manufacturers and different versions of device drivers, there is no study or model which defines the exact impact of device drivers on the performance. The recommendation is to use the newest device drivers as they should provide the best performance. Prior to the Windows Vista and Windows 7 performance measurement in Experiments 1 and 2, a USB flash drive was included in the computer system configuration and dedicated to the ReadyBoost feature. Due to limitations of the FAT32 file system, the upper limit of flash memory that can be used for ReadyBoost is 4 GB. This flash memory drive serves as an additional memory cache and improves Windows Vista and Windows 7 performance. All three OSs were prepared for the measurement procedure according to the performance measurement process described in [Section 4.3]. This process enables the best operating conditions in which all tested OSs should show their best performance. In addition to the OS and benchmark application, there was no external or third party application installed on the computer system, except in two cases: (i) Prior to the Windows XP performance measurements with PCMark05 in Experiments 1 and 2, we installed a 32-bit version of Windows Media Encoder 9 and Windows Media Player 11 since they are required for PCMark05 in order to run properly; (ii) Since D-ITG GUI (Graphical User Interface) used in Experiment 3 is written in Java, it is required to install Java for running. During performance measurement there was no user activity on the system and on stand-alone computer systems in Experiments 1 and 2 the network was disconnected.

To ensure results stability, each measurement was repeated for five times in the same conditions. In every repetition, the last five steps of the performance measurement process for Windows Vista and Windows 7 and the last four steps for Windows XP were executed, as shown in [Fig. 4]. The final result of each performance measurement is expressed as an arithmetic mean of five repetitions, as shown in (3). Deviation from the arithmetic mean for vast majority of repetition results is in the range of 1%. This very small deviation indicates that our performance measurement process is effective and can be reused. Furthermore, it also indicates that results are consistent and reliable and that there is no need for more than five repetitions of each measurement. Owning to the mentioned deviation of repetition results, in experiments we report only significant digits of performance measurement results. The fourth digit represents the difference in the results less than 1%, and consequently it enters the area of the measurement error. Therefore, performance measurement results are reported with three significant digits.

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \rightarrow N = 5 \quad (3)$$

$$x_{i\,\min} \leq \bar{x} \leq x_{i\,\max}$$

In tables that are reporting performance measurement results of all three Windows OSs, the percentage error formula is used for calculation of the percentage difference between the measured values, as shown in (4). Windows XP values are used as referent values, and Windows Vista and Windows 7 values are compared with respect to Windows XP values.

$$Diff_{\%} = \frac{WinVista\_or\_Win7\_value - WinXP\_value}{WinXP\_value} \times 100\% \quad (4)$$

Performance measurement is done in two experiments. Considering the fact that there exist a huge number of computer systems with various hardware combinations, two completely different hardware are used in experiments: newer (high-end) in Experiment 1 and older (a low-end computer system) in Experiment 2. In both experiments we completed the following steps:

1. Format HDD.
2. Apply the performance measurement process from [Fig. 4] on Windows XP.
3. Format HDD.
4. Apply the performance measurement process from [Fig. 4] on Windows Vista.
5. Format HDD.
6. Apply the performance measurement process from [Fig. 4] on Windows 7.

6.2 Experiment 1

In Experiment 1, performance measurement is performed on the high-end computer system with the hardware and software configuration shown in [Tab. 2]. A more detailed configuration can be found in the Appendix [Tab. 17]. Furthermore, listings of all installed drivers and updates in all three Windows OSs used in Experiment 1 are also part of the Appendix.

Component	**Specifications**
Hardware	
CPU	QuadCore Intel Core i7 940, 2.93 GHz
RAM	3 x 2 GB DDR3
Graphics hardware	NVIDIA GeForce GTX 295, 2 x 896 MB
HDD	SATA 500 GB
Motherboard	Asus P6T WS Pro
USB Flash drive	16 GB (4 GB dedicated to ReadyBoost)
Operating systems	
Windows XP	Microsoft Windows XP Professional SP3 32-bit (version 5.1.2600)
Windows Vista	Microsoft Windows Vista Business SP2 32-bit (version 6.0.6002)
Windows 7	Microsoft Windows 7 Professional 32-bit (version 6.1.7600)
DirectX	
DirectX XP	DirectX 9.0c
DirectX Vista	DirectX 10.1
DirectX Win7	DirectX 11.0

Table 2: Hardware and software configuration in Experiment 1

Performance measurement results of the Windows OS memory management are shown in [Tab. 3, Tab. 4].

Performance indicators	Windows XP	Windows Vista	Windows 7	Unit	Difference Vista to XP	Difference Win7 to XP
v_{Mrd} - 16 MB	10.9	10.7	10.7	GB/s	-1.83%	-1.83%
v_{Mrd} - 8 MB	22.0	25.3	25.4	GB/s	15.00%	15.45%
v_{Mrd} - 192 KB	22.6	22.2	22.5	GB/s	-1.77%	-0.44%
v_{Mrd} - 4 KB	46.0	45.5	45.8	GB/s	-1.09%	-0.43%
v_{Mwr} - 16 MB	9.22	9.17	9.18	GB/s	-0.54%	-0.43%
v_{Mwr} - 8 MB	9.22	9.17	9.19	GB/s	-0.54%	-0.33%
v_{Mwr} - 192 KB	28.0	27.5	27.4	GB/s	-1.79%	-2.14%
v_{Mwr} - 4 KB	46.9	46.3	46.7	GB/s	-1.28%	-0.43%
v_{Mcp} - 16 MB	9.79	9.70	9.70	GB/s	-0.92%	-0.92%
v_{Mcp} - 8 MB	12.9	13.2	13.3	GB/s	2.33%	3.10%
v_{Mcp} - 192 KB	22.1	21.6	22.0	GB/s	-2.26%	-0.45%
v_{Mcp} - 4 KB	46.2	45.7	46.0	GB/s	-1.08%	-0.43%
v_{Mla} - Random 16 MB	13.1	12.4	19.3	MAccesses /s	-5.34%	47.33%
v_{Mla} - Random 8 MB	33.1	41.7	56.0	MAccesses /s	25.98%	69.18%
v_{Mla} - Random 192 KB	305	244	307	MAccesses /s	-20.00%	0.66%
v_{Mla} - Random 4 KB	768	767	769	MAccesses /s	-0.13%	0.13%
v_{MErd}	11.7	11.7	11.8	GB/s	0.00%	0.85%
v_{MEwr}	9.68	9.68	9.68	GB/s	0.00%	0.00%
v_{MEcp}	13.0	12.9	14.1	GB/s	-0.77%	8.46%
v_{L1rd}	49.2	49.2	49.2	GB/s	0.00%	0.00%
v_{L1wr}	49.1	49.1	49.1	GB/s	0.00%	0.00%
v_{L1cp}	98.3	98.3	98.3	GB/s	0.00%	0.00%
v_{L2rd}	32.8	32.8	32.8	GB/s	0.00%	0.00%
v_{L2wr}	29.7	29.8	29.8	GB/s	0.34%	0.34%
v_{L2cp}	37.7	38.0	37.9	GB/s	0.80%	0.53%
v_{L3rd}	23.2	23.3	23.4	GB/s	0.43%	0.86%
v_{L3wr}	15.2	15.2	15.3	GB/s	0.00%	0.66%
v_{L3cp}	22.6	23.1	23.0	GB/s	2.21%	1.77%
c_f (lower is better)	155	284	165	CacheFaults/s	83.62%	6.62%
p_f (lower is better)	482	1039	780	PageFaults/s	115.29%	61.72%
p_r (lower is better)	0.31	1.27	0.08	PageReads/s	301.62%	-74.60%

Table 3: Memory management performance measurement results in Experiment 1 (more is better)

Two performance indicators show significantly better results in Windows Vista and Windows 7: v_{Mrd} shows 15.00% (Vista) and 15.45% (Win7) higher throughput when reading 8 MB data blocks from the memory; v_{Mla} shows 25.98% (Vista) and even 69.18% (Win7) better memory latency performance, again for 8 MB blocks. The memory copy performance on 8 MB data blocks size is also lower in Windows XP. It can be concluded that Windows Vista and Windows 7 have significant improvements in memory management with 8 MB data blocks. In addition, Windows Vista has lower performance in memory management with 192 KB data blocks and v_{Mla} (random 192 KB) result is 20.00% lower than in Windows XP and even 25.82% lower than in Windows 7.

Another interesting fact is that although computer system has 6 GB of RAM memory, Windows XP showed that the total amount of physical memory is 3.25 GB (which is limit for 32-bit system), whereas Windows Vista and Windows 7 showed only 2.5 GB. Furthermore, the initial memory load in Windows XP was only 259 MB, whereas in Windows Vista it was much higher (1059 MB) and in Windows 7 it was 516 MB. However, compared to Windows XP, Windows Vista shows a much higher rate of cache faults c_f (86.62% higher), page faults p_f (115.29% higher) and page reads p_r (301.62% higher). Since p_r represents a hard page faults that occur when the page is not located in physical memory or a memory-mapped file created by the process and therefore they are highly time consuming, it can be concluded that new memory manager in Windows Vista is not more efficient than in Windows XP. Furthermore, although Windows 7 shows higher rates of c_f (6.62%) and p_f (61.72%), memory manager reduces the number of hard page faults for 74.60% compared to Windows XP.

In all other measurements, results are nearly the same for all three Windows OSs except for two performance indicators: v_{Mla} (random 16 MB) and v_{MEcp} show 47.33% and 8.46% better results in Windows 7 than in Windows XP, and 55.65% and 9.30% better than in Windows Vista, respectively.

As shown in [Tab. 4], memory latency performance indicators have almost identical results for all tested OSs. As they are measured with the Everest benchmark application, it can be concluded that these measurements represent raw memory performance.

Performance indicators	**Windows XP**	**Windows Vista**	**Windows 7**	**Unit**	**Difference Vista to XP**	**Difference Win7 to XP**
t_{ML}	67.7	69.0	68.1	ns	1.92%	0.59%
t_{L1}	1.30	1.30	1.30	ns	0.00%	0.00%
t_{L2}	3.30	3.30	3.30	ns	0.00%	0.00%
t_{L3}	4.30	4.30	4.32	ns	0.00%	0.47%

Table 4: Memory latency performance measurement results in Experiment 1 (shorter is better)

CPU management results are given in [Tab. 5, Tab. 6]. Performance indicator v_{CPUts} measurement results are approximately equal in all measurements, except in audio and file compression, where Windows Vista shows the lowest performance. Performance indicators v_{CPU1} and v_{CPU2} show lower performance of Windows Vista and Windows 7. This is surprising as these CPU tests are optimized for multi-core processors and we expected that newer versions of the Windows OS will have better CPU scheduling on multi-core processors. Calculation times in [Tab. 6] are shorter for Windows Vista and Windows 7 in MolDyn and Primordia benchmarks and from this we can conclude that they have faster execution of operations on floating point numbers (FPU part of CPU) than Windows XP.

Performance indicators	**Windows XP**	**Windows Vista**	**Windows 7**	**Unit**	**Difference Vista to XP**	**Difference Win7 to XP**
v_{CPUts}						
File Compression	13.9	13.9	13.9	MB/s	0.00%	0.00%
File Decompression	199	199	199	MB/s	0.00%	0.00%
File Encryption	84.9	84.5	84.8	MB/s	-0.47%	-0.12%
File Decompression	84.4	84.1	84.2	MB/s	-0.36%	-0.24%
Image Decompression	42.7	42.4	42.8	MPixels/s	-0.70%	0.23%
Audio Compression	4.46	4.27	4.45	MB/s	-4.26%	-0.22%
File Compression	13.9	12.8	13.9	MB/s	-7.91%	0.00%
File Encryption	84.9	84.7	84.7	MB/s	-0.24%	-0.24%
File Decompression	199	198	198	MB/s	-0.50%	-0.50%
File Decryption	84.4	84.1	83.7	MB/s	-0.36%	-0.83%
Audio Decompression	3.70	3.66	3.69	MB/s	-1.08%	-0.27%
Image Decompression	42.6	42.5	42.8	MPixels/s	-0.23%	0.47%
v_{CPU1}	1.99	1.84	1.81	Frames/s	-7.54%	-9.05%
v_{CPU2}	2.77	2.57	2.49	Frames/s	-7.22%	-10.11%

Table 5: CPU management performance measurement results in Experiment 1 (more is better)

Performance indicators	**Windows XP**	**Windows Vista**	**Windows 7**	**Unit**	**Difference Vista to XP**	**Difference Win7 to XP**
t_{CPUmd}	35.5	34.1	34.0	s	-3.94%	-4.23%
t_{CPUpr}	192	186	184	s	-3.13%	-4.17%
t_{CPUch}	9.28	9.39	9.32	s	1.19%	0.43%
t_{CPUs4}	67.0	67.9	67.7	s	1.34%	1.04%
t_{CPUs16}	337	335	337	s	-0.59%	0.00%

Table 6: Benchmarks calculation times in Experiment 1 (shorter is better)

Graphics display results are shown in [Tab. 7, Fig. 4]. Performance indicators in [Tab. 7] show a big diversity of performance measurement results. Generally, Windows XP shows better results in comparison to Windows Vista and Windows 7. However, throughput of drawing transparent windows in one second is even 826.90% higher in Windows Vista and 199.49% in Windows 7. This huge difference in results can be assigned to new display driver architecture in Windows Vista and Windows 7.

Performance indicators	**Windows XP**	**Windows Vista**	**Windows 7**	**Unit**	**Difference Vista to XP**	**Difference Win7 to XP**
v_{Gts}						
Transparent Windows	0.985	9.13	2.95	MWindows /s	826.90%	199.49%
Graphics Memory 64 lines	3.52	2.98	2.92	MFrames/s	-15.34%	-17.05%
Graphics Memory 128 lines	1.84	2.08	2.08	MFrames/s	13.04%	13.04%
WMV Video Playback	59.0	44.5	44.9	Frames/s	-24.58%	-23.90%
3D - Fill Rate Multi-Texturing	37.1	36.3	36.1	GTexels/s	-2.16%	-2.70%
3D - Polygon Throughput Multiple Lights	186	142	145	MTriangles /s	-23.66%	-22.04%
3D - Pixel Shader	1.30	1.18	1.15	MFrames/s	-9.23%	-11.54%
3D - Vertex Shader	143	142	142	MVertices/s	-0.70%	-0.70%
v_{Gs2}						
GT1 - Return To Proxycon	62.9	61.8	61.5	Frames/s	-1.75%	-2.23%
GT2 - Firefly Forest	62.2	60.1	59.3	Frames/s	-3.38%	-4.66%
v_{Ghdr}						
HDR1 - Canyon Flight	114	119	119	Frames/s	4.39%	4.39%
HDR2 - Deep Freeze	80.9	79.6	79.1	Frames/s	-1.61%	-2.22%
v_{G3d}						
Fill Rate Single-Texturing	23.1	22.5	22.8	GTexels/s	-2.60%	-1.30%
Fill Rate Multi-Texturing	86.7	77.7	81.3	GTexels/s	-10.38%	-6.23%
Pixel Shader	1.91	1.63	1.71	MFrames/s	-14.66%	-10.47%
Vertex Shader - Simple	463	472	473	MVertices/s	1.94%	2.16%
Vertex Shader - Complex	290	296	296	MVertices/s	2.07%	2.07%
Shader Particles (SM3.0)	220	242	250	Frames/s	10.00%	13.64%
Perlin Noise (SM3.0)	605	567	579	Frames/s	-6.28%	-4.30%

Table 7: Graphics display performance measurement results in Experiment 1 (more is better)

Graphics display performance measurement results from [Fig. 5] show that Windows XP outperforms Windows Vista and Windows 7. Only W_{Gsw} performance indicator gives better results for Windows Vista and Windows 7. Another interesting fact we noticed in the graphics display results, is that almost all measurement results are nearly the same for Windows Vista and Windows 7. This indicates that Windows Vista display architecture is retained in Windows 7 and that Windows 7 enhancements do not result in better performance.

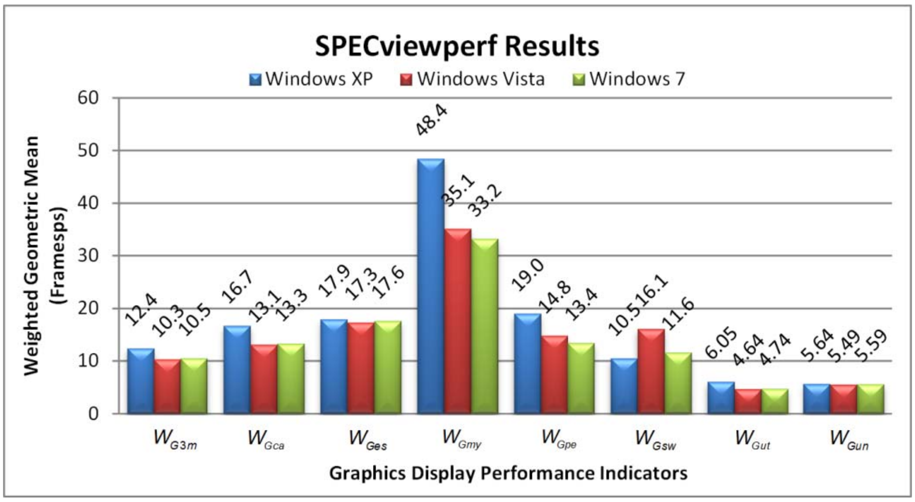

Figure 5: Graphics display performance indicators in Experiment 1 (more is better)

Performance indicators	**Windows XP**	**Windows Vista**	**Windows 7**	**Unit**	**Difference Vista to XP**	**Difference Win7 to XP**
v_{Hts}						
HDD - XP Startup	10.3	10.2	9.92	MB/s	-0.97%	-3.69%
HDD - Application Load	8.60	8.39	8.15	MB/s	-2.44%	-5.23%
HDD - General Usage	7.24	7.18	7.03	MB/s	-0.83%	-2.90%
HDD - Virus Scan	163	151	156	MB/s	-7.36%	-4.29%
HDD - File Write	93.8	79.1	87.5	MB/s	-15.67%	-6.72%
v_{Hrd}						
Linear Read (Begin)	128	127	127	MB/s	-0.78%	-0.78%
Linear Read (Middle)	110	105	111	MB/s	-4.55%	0.91%
Linear Read (End)	67.9	64.3	68.1	MB/s	-5.30%	0.29%
Random Read	115	93.9	109	MB/s	-18.35%	-5.22%
Buffered Read	245	238	234	MB/s	-2.86%	-4.49%
t_{Hrd} (shorter is better)	13.8	13.9	14.0	ms	0.72%	1.45%

Table 8: HDD management performance measurement results in Experiment 1 (more is better)

As shown in [Tab. 8], it is obvious that Windows XP has better HDD management than Windows Vista and Windows 7. Windows 7 and especially

Windows Vista achieve worse results in all measurements. When measuring performance of the HDD file write and random read, Windows Vista has 15.67% respectively 18.35% worse results than Windows XP and 10.62% respectively 16.08% worse results than Windows 7.

6.3 Experiment 2

In Experiment 2, Windows OSs performance measurement is performed on the computer system that satisfies minimum hardware requirements of all three Windows OSs, as shown in [Tab. 9]. The goal of this experiment is to compare Windows OSs performance on the low-end computer system. Although high-end computer systems are today in vast majority, there are still users with low-end computer systems that consider upgrading to a new version of Windows OS. This experiment will discover whether they will have to move to the new hardware also. A more detailed configuration and listings of installed drivers and updates in all three Windows OSs used in Experiment 2 can be found in the Appendix [Tab. 19, Tab. 20].

Component	**Specifications**
Hardware	
CPU	Intel Celeron D 330, 2.66 GHz
RAM	2 x 512 MB DDR
Graphics hardware	Integrated: Intel 82865G Graphics Controller
HDD	SATA 40 GB
Motherboard	FUJITSU SIEMENS D1561
USB Flash drive	16 GB (4 GB dedicated to ReadyBoost)
Operating systems	
Windows XP	Microsoft Windows XP Professional SP3 32-bit (version 5.1.2600)
Windows Vista	Microsoft Windows Vista Business SP2 32-bit (version 6.0.6002)
Windows 7	Microsoft Windows 7 Professional 32-bit (version 6.1.7600)
DirectX	
DirectX XP	DirectX 9.0c
DirectX Vista	DirectX 10.1
DirectX Win7	DirectX 11.0

Table 9: Hardware and software configuration in Experiment 2

Performance indicators in [Tab. 10] show that Windows XP and Windows 7 have a much more efficient memory manager than Windows Vista. In almost all measurements Windows Vista has much worse results than other two OSs, especially when it comes to memory read/write/copy operations and memory latency of data blocks size 192 KB. Memory latency performance indicator v_{Mla} shows that Windows 7 has the fastest memory access. Performance measurement results of memory management indicate that on low-end computer systems with a limited amount of memory, Windows Vista has slower execution of read/write/copy operations on the main memory, and L1 and L2 cache memory.

Similarly to Experiment 1, the initial memory load in Windows XP was the lowest (only 235 MB), whereas in Windows Vista it was again much higher (756 MB) and in Windows 7 it was 348 MB. Furthermore, compared to Windows XP, Windows Vista shows higher rates of c_f (47.71% higher) and p_f (48.76% higher) and much higher number of hard page faults per second (5100.97% higher).

Compared to Windows XP, Windows 7 shows a similar rate of c_f and a higher rate of p_f (62.19%). The number of the hard page faults per second is reduced for 73.54%.

Performance indicators	**Windows XP**	**Windows Vista**	**Windows 7**	**Unit**	**Difference Vista to XP**	**Difference Win7 to XP**
v_{Mrd} - 16 MB	2.05	1.87	2.04	GB/s	-8.78%	-0.49%
v_{Mrd} - 8 MB	2.05	1.89	2.04	GB/s	-7.80%	-0.49%
v_{Mrd} - 192 KB	20.0	10.4	19.2	GB/s	-48.00%	-4.00%
v_{Mrd} - 4 KB	33.4	31.3	32.0	GB/s	-6.29%	-4.19%
v_{Mwr} - 16 MB	1.97	1.83	1.99	GB/s	-7.11%	1.02%
v_{Mwr} - 8 MB	1.97	1.83	1.99	GB/s	-7.11%	1.02%
v_{Mwr} - 192 KB	8.93	4.92	8.52	GB/s	-44.90%	-4.59%
v_{Mwr} - 4 KB	9.00	8.35	8.55	GB/s	-7.22%	-5.00%
v_{Mcp} - 16 MB	1.96	1.81	1.94	GB/s	-7.65%	-1.02%
v_{Mcp} - 8 MB	1.95	1.80	1.93	GB/s	-7.69%	-1.03%
v_{Mcp} - 192 KB	7.71	4.26	7.43	GB/s	-44.75%	-3.63%
v_{Mcp} - 4 KB	9.01	8.33	8.57	GB/s	-7.55%	-4.88%
v_{Mla} - Random 16 MB	5.90	5.46	6.92	MAccesses /s	-7.46%	17.29%
v_{Mla} - Random 8 MB	6.38	6.33	6.85	MAccesses /s	-0.78%	7.37%
v_{Mla} - Random 192 KB	97.6	59.4	111	MAccesses /s	-39.14%	13.73%
v_{Mla} - Random 4 KB	664	664	665	MAccesses /s	0.00%	0.15%
v_{MErd}	2.20	2.16	2.24	GB/s	-1.82%	1.82%
v_{MEwr}	2.12	2.08	2.19	GB/s	-1.89%	3.30%
v_{MEcp}	2.13	2.08	2.18	GB/s	-2.35%	2.35%
v_{L1rd}	42.4	42.4	42.4	GB/s	0.00%	0.00%
v_{L1wr}	9.46	9.46	9.46	GB/s	0.00%	0.00%
v_{L1cp}	18.6	18.9	18.8	GB/s	1.61%	1.08%
v_{L2rd}	19.4	19.1	19.3	GB/s	-1.55%	-0.52%
v_{L2wr}	8.95	8.70	8.76	GB/s	-2.79%	-2.12%
v_{L2cp}	14.9	14.5	14.6	GB/s	-2.68%	-2.01%
c_f (lower is better)	153	226	154	CacheFaults/s	47.71%	0.76%
p_f (lower is better)	476	707	771	PageFaults/s	48.76%	62.19%
p_r (lower is better)	0.22	11.6	0.06	PageReads/s	5100.97%	-73.54%

Table 10: Memory management performance measurement results in Experiment 2 (more is better)

With one exception, memory latency performance indicators shown in [Tab. 11] are almost equal for all tested OSs. This exception refers to a much longer L2 cache latency in Windows Vista (45.76% compared to Windows XP and 44.94% when compared to Windows 7).

Performance indicators	**Windows XP**	**Windows Vista**	**Windows 7**	**Unit**	**Difference Vista to XP**	**Difference Win7 to XP**
t_{ML}	144	146	141	ns	1.39%	-2.08%
t_{L1}	1.50	1.50	1.50	ns	0.00%	0.00%
t_{L2}	8.85	12.9	8.90	ns	45.76%	0.56%

Table 11: Memory latency performance measurement results in Experiment 2 (shorter is better)

With the few exceptions, Windows XP has better CPU management performance measurement results than Windows Vista and Windows 7, as shown in [Tab. 12]. Windows Vista has the lowest throughput during audio, image and file decompression and Windows 7 during file compression. In [Tab. 13] performance indicators t_{CPUmd} and t_{CPUch} show much longer calculation times in Windows Vista and Windows 7. Performance indicators t_{CPUs4} and t_{CPUs16} show that Windows Vista has the fastest calculation of 4×10^6 and 16×10^6 digits of the number PI. The results indicate that Windows XP best allocate CPU resources to the benchmark applications.

Performance indicators	**Windows XP**	**Windows Vista**	**Windows 7**	**Unit**	**Difference Vista to XP**	**Difference Win7 to XP**
v_{CPUts}						
File Compression	6.11	6.20	5.12	MB/s	1.47%	-16.20%
File Decompression	102	96.7	99.2	MB/s	-5.20%	-2.75%
File Encryption	55.1	57.1	59.0	MB/s	3.63%	7.08%
File Decompression	53.4	53.5	54.8	MB/s	0.19%	2.62%
Image Decompression	21.1	18.9	20.5	MPixels/s	-10.43%	-2.84%
Audio Compression	1.74	1.64	1.68	MB/s	-5.75%	-3.45%
File Compression	3.09	3.12	2.46	MB/s	0.97%	-20.39%
File Encryption	27.4	28.1	29.1	MB/s	2.55%	6.20%
File Decompression	25.8	22.3	24.0	MB/s	-13.57%	-6.98%
File Decryption	13.2	12.4	13.1	MB/s	-6.06%	-0.76%
Audio Decompression	435	348	402	KB/s	-20.00%	-7.59%
Image Decompression	5.40	4.62	4.96	MPixels/s	-14.44%	-8.15%

Table 12: CPU management performance measurement results in Experiment 2 (more is better)

As shown in [Tab. 14], the throughput of drawing transparent windows in one second is more than two times higher in Windows Vista and Windows 7 than in

Windows XP. However, other performance measurement results of a graphics subsystem are much higher in Windows XP. 3D graphics tests failed in Windows 7 mostly because of low graphics hardware resources of the tested computer system.

Performance indicators	**Windows XP**	**Windows Vista**	**Windows 7**	**Unit**	**Difference Vista to XP**	**Difference Win7 to XP**
t_{CPUmd}	109	257	285	s	135.78%	161.47%
t_{CPUpr}	737	697	701	s	-5.43%	-4.88%
t_{CPUch}	18.8	23.9	23.3	s	27.13%	23.94%
t_{CPUs4}	422	342	400	s	-18.96%	-5.21%
t_{CPUs16}	30.9	26.4	30.2	min	-14.56%	-2.27%

Table 13: Benchmarks calculation times in Experiment 2 (shorter is better)

Performance indicators	**Windows XP**	**Windows Vista**	**Windows 7**	**Unit**	**Difference Vista to XP**	**Difference Win7 to XP**
v_{Gts}						
Transparent Windows	53.8	108	109	Windows /s	100.74%	102.60%
Graphics Memory 64 lines	198	181	79.3	Frames/s	-8.59%	-59.95%
Graphics Memory - 128 lines	164	157	72.1	Frames/s	-4.27%	-56.04%
WMV Video Playback	25.2	24.2	12.4	Frames/s	-3.97%	-50.79%
3D - Fill Rate Multi-Texturing	960	931	Failed	MTexels/s	-3.02%	-
3D - Polygon Throughput Multiple Lights	3.15	1.30	Failed	MTriangles /s	-58.73%	-
3D - Pixel Shader	1.00	1.00	Failed	Frames/s	0.00%	-
3D - Vertex Shader	1.00	1.00	Failed	MVertices/s	0.00%	-

Table 14: Graphics display performance measurement results in Experiment 2 (more is better)

As shown in [Fig. 6], Windows Vista and Windows 7 show approximately the same performance in all graphics display performance measurements, similarly to Experiment 1, but their performance are significantly lower than Windows XP. The results indicate that Windows Vista and especially Windows 7 need much more graphics resources than Windows XP. Since they are newer OSs than Windows XP, they contain much more graphics elements like animations, shapes, textures and colors.

HDD management performance measurement results are shown in [Tab. 15]. Performance indicator v_{Hts} shows lower performance for Windows Vista and Windows 7 than for Windows XP. Performance indicators v_{Hrd} and t_{Hrd} show equal performance for all three Windows OSs. It can be concluded that the average data throughput is much higher in Windows XP than in Windows Vista and Windows 7.

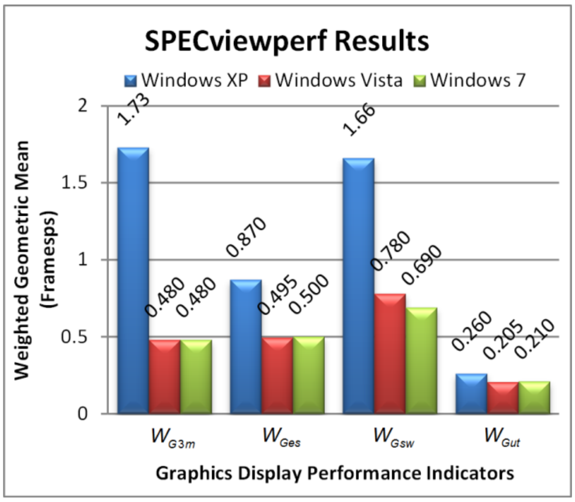

Figure 6: Graphics display performance indicators in Experiment 2 (more is better)

Performance indicators	Windows XP	Windows Vista	Windows 7	Unit	Difference Vista to XP	Difference Win7 to XP
v_{Hts}						
HDD - XP Startup	6.54	5.97	6.17	MB/s	-8.72%	-5.66%
HDD - Application Load.	4.54	4.21	4.37	MB/s	-7.27%	-3.74%
HDD - General Usage	3.51	3.30	3.34	MB/s	-5.98%	-4.84%
HDD - Virus Scan	74.0	52.6	62.2	MB/s	-28.92%	-15.95%
HDD - File Write	52.0	47.5	39.9	MB/s	-8.65%	-23.27%
v_{Hrd}						
Linear Read (Begin)	52.8	52.8	52.7	MB/s	0.00%	-0.19%
Linear Read (Middle)	46.5	46.8	47.0	MB/s	0.65%	1.08%
Linear Read (End)	30.6	30.5	30.7	MB/s	-0.33%	0.33%
Random Read	37.7	36.3	37.2	MB/s	-3.71%	-1.33%
Buffered Read	125	123	123	MB/s	-1.60%	-1.60%
t_{Hrd} (shorter is better)	15.9	16.0	16.1	ms	0.63%	1.26%

Table 15: HDD management performance measurement results in Experiment 2 (more is better)

6.4 Experiment 3

In Experiment 3, performance measurement is performed on the two identical computer systems with the hardware and software configuration shown in [Tab. 16]. A more detailed configuration can be found in the Appendix [Tab. 21]. The goal of this experiment is to compare network performance of the Windows OSs.

Component	Specifications
Hardware	
CPU	Dual-Core E5200 Intel, 2.50 GHz
RAM	2 x 2 GB DDR2
Graphics hardware	Asus EAH3650, 1024 MB
HDD	SATA 250 GB
Motherboard	Asus P5K-VM
Network	Onboard LAN Marvell88E8056 PCI-E Gigabit LAN controllers
Operating systems	
Windows XP	Microsoft Windows XP Professional SP3 32-bit (version 5.1.2600)
Windows Vista	Microsoft Windows Vista Business SP2 32-bit (version 6.0.6002)
Windows 7	Microsoft Windows 7 Professional 32-bit (version 6.1.7600)
DirectX	
DirectX XP	DirectX 9.0c
DirectX Vista	DirectX 10.1
DirectX Win7	DirectX 11.0

Table 16: Hardware and software configuration in Experiment 3

According to the instructions from the literature and the [Section 5.1] we used two identical computer systems connected with 1 Gbps Ethernet crossover cable and the testbed setup is shown in [Fig. 7].

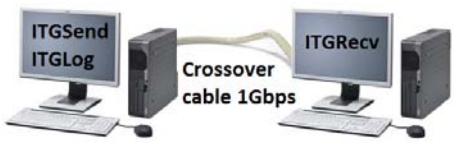

Figure 7: Testbed setup in Experiment 3

Experimental setup of D-ITG is configured similar as in literature [Narayan et al. 09, Narayan et al. 10] and consists of following parameters:

- Duration of each measurement was 60s and each measurement was repeated 5 times in the same conditions.
- In order to gain wide range of data, measurements were conducted for 13 different packet sizes in the range from 64 bytes to 1536 bytes (in regular steps of 128 bytes after second packet size which was also 128 bytes) and packet size was constant during each measurement.
- Inter-departure time option was uniformly distributed and the packet rate of the sender was distributed from 30000 packets/s to 300000 packets/s.
- Performance of TCP and UDP protocols was measured when running over the both IPv4 and IPv6, respectively.

- RTT (Round Trip Time) meter was used.
- ITGSend component was used for sending packets from the first computer system to the ITGRecv component on the second computer system.
- ITGLog on sender computer system was used for generating log file with measurement results.
- Three performance indicators were used: (i) t_{Dnet} for measuring network delay; (ii) t_{Jnet} for measuring network jitter; and (iii) v_{Tnet} for measuring network throughput.
- All other parameters in D-ITG were default.

Beside performance indicators obtained from D-ITG, during the network performance measurement CPU usage was also measured using the Windows performance monitor. This measure indicates processor activity and displays the average percentage of busy time observed during the measurement interval.

Round-trip time delay values for TCP protocol [Fig. 8] indicate that for the small packet sizes up to 384 bytes Windows Vista has smallest delays but for the bigger packet sizes over 384 bytes it has largest delays. Significantly smaller delays for packet sizes over 256 bytes are accomplished in Windows XP for IPv6, in average 146.67% lower than in Windows Vista and 118.47% lower than in Windows 7.

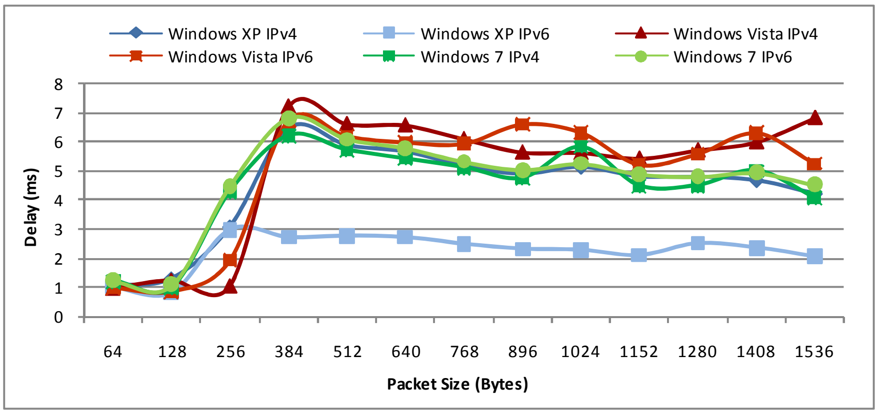

Figure 8: TCP delay (lower is better)

Jitter values for TCP protocol [Fig. 9] indicate that for the small packet sizes up to 256 bytes all Windows OSs show similar performance but for the bigger packet sizes over the 256 bytes Windows Vista and Windows 7 show average 15% and 25% lower jitter values, respectively. Throughput values for TCP protocol [Fig. 10] indicate that Windows XP with IPv6 for packet sizes over 256 bytes has in average 14% lower throughput when compared to other Windows OSs. Windows Vista and Windows 7 show similar throughput performance.

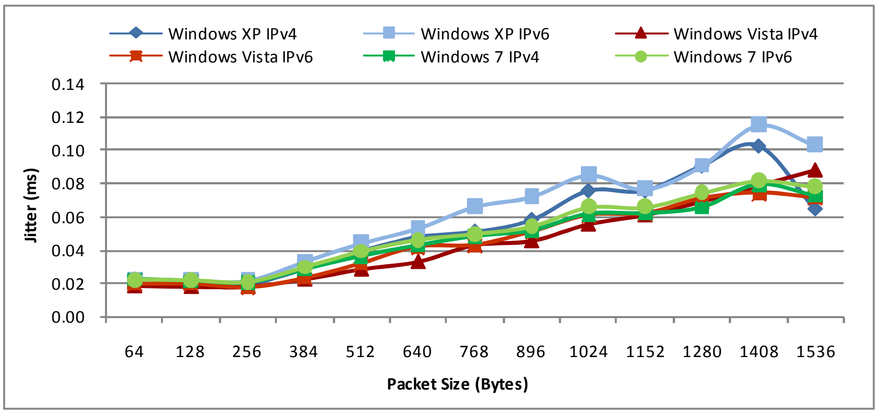

Figure 9: TCP jitter (lower is better)

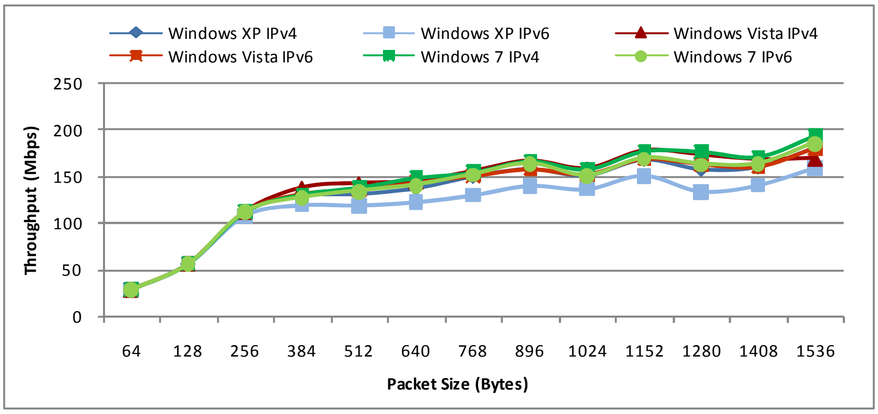

Figure 10: TCP throughput (more is better)

CPU usage percentage values for TCP protocol [Fig. 11] indicate that all Windows OSs require highest processor activity when sending small packets up to 256 bytes. In general, Windows XP consumes smallest amount of CPU processing power. Windows 7 with IPv4 shows very unusual results: for some packet sizes (384, 512, 640, 1408 and 1536 bytes) it consumes lowest amount of CPU processing power (below 20%) and for the other (from 768 to 1280 bytes) it consumes highest amount of CPU processing power when compared to other Windows OSs.

Round-trip time delay values for UDP protocol [Fig. 12] indicate that for the small packet sizes up to 256 bytes and for the bigger packet sizes over 1152 bytes Windows XP has smallest delays. Delays for the other packet sizes are similar in all Windows OSs.

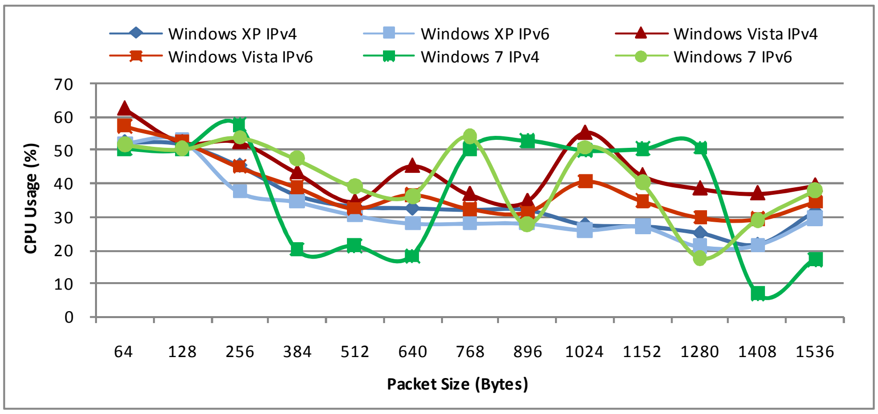

Figure 11: TCP CPU usage (lower is better)

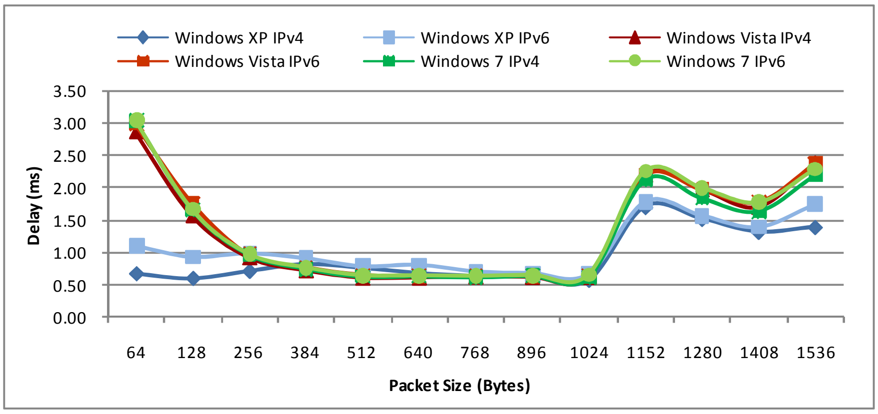

Figure 12: UDP delay (lower is better)

Jitter values for UDP protocol [Fig. 13] indicate that for the packet sizes up to 1024 bytes all Windows OSs show similar performance. However, for the bigger packet sizes over the 1024 bytes Windows Vista and Windows 7 show much higher jitter values when compared to the Windows XP jitter values.

Throughput values for UDP protocol [Fig. 14] indicate that Windows XP has in average 10% higher throughput for small packet sizes up to 256 bytes and 22% in average higher throughput for bigger packet sizes over 1152 bytes when compared to Windows Vista and Windows 7. However, for medium-sized packets (from 256 to 896 bytes) Windows XP shows for 17% in average lower throughput when compared to Windows Vista and Windows 7.

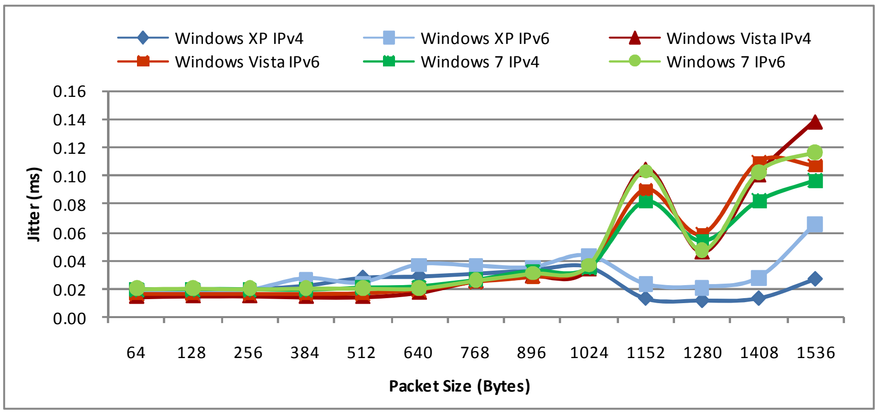

Figure 13: UDP jitter (lower is better)

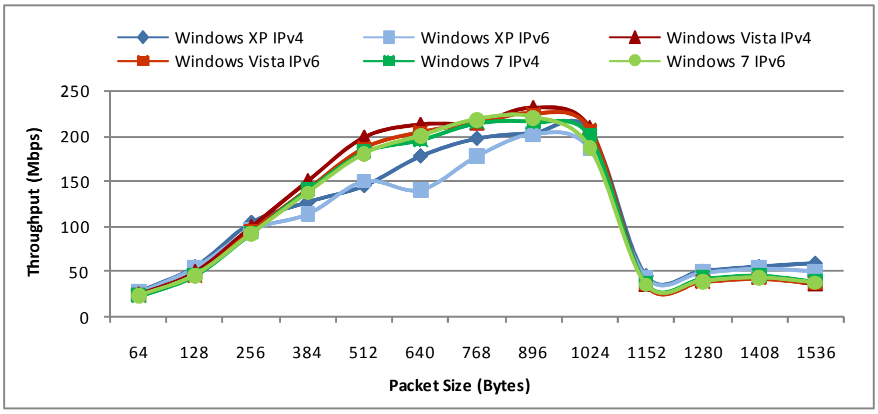

Figure 14: UDP throughput (more is better)

CPU usage percentage values for UDP protocol [Fig. 15] indicate that Windows XP consumes smallest amount of CPU processing power for medium-sized packets (from 256 to 1024 bytes). However, some interesting characteristic can be seen for bigger packets over 1152 bytes. While sending packets Windows 7 CPU activity drops almost to the zero. Furthermore, Windows Vista processor activity has various peaks, like lowest for the packet sizes of 1152, 1408 and 1536 bytes for IPv6 and for the packet sizes of 1152 and 1536 bytes for the IPv4, and highest for the packet size 1280 bytes for IPv6 and for the packet size 1408 bytes for IPv4.

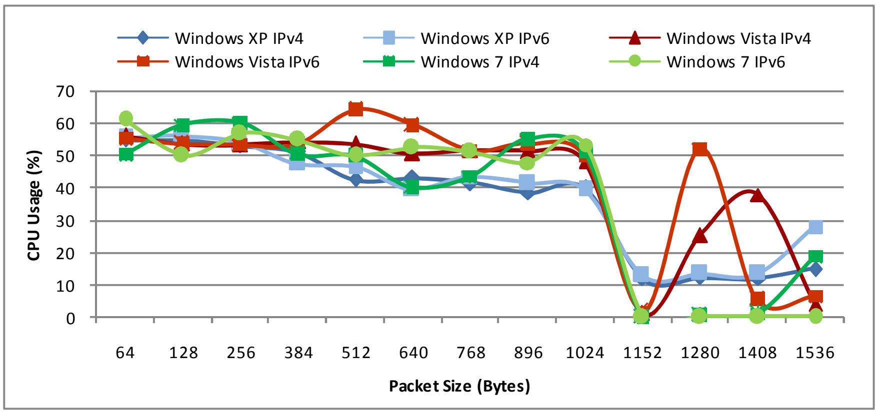

Figure 15: UDP CPU usage (lower is better)

7 Discussion

At first glance, it is not obvious which Windows OS handles system resources more efficient and has best performance. In fact, the same performance indicators in the same OS behaved differently on stand-alone computer systems in the first and the second experiment. This behavior is caused by different hardware architectures in experiments. Our goal is to evaluate various aspects of the Windows OS from a huge amount of performance measurement results. Since the computer system in Experiment 1 offers more hardware resources than in Experiment 2, Window Vista and Windows 7 were expected to obtain better results in Experiment 1. Due to the lowest hardware requirements, Windows XP was expected to have highest performance in Experiment 2.

On stand-alone computer systems, some performance measurement results show a huge difference between different versions of Windows OSs. Owning to the fact that all measurements were repeated five times, this huge difference does not represent the threat to validity of the results. It rather shows in which parts Windows OSs have the largest performance gap.

On the network computer system performance depends on packet size and on the protocol used for sending data. IPv4 and IPv6 have similar performance in Windows Vista and Windows 7 but show deviations when using in Windows XP.

In comparison with Windows XP, Windows Vista and Windows 7 have plenty of new features and enhancements in parts of the OS that mostly impact the performance: memory management, CPU scheduling, the display architecture, HDD management and network optimization. Our performance measurement results show that in all tested environments, in most performance measurements these new features and enhancements do not result in better performance.

7.1 Experiment 1

Compared to Windows XP, Windows Vista and especially Windows 7 memory management improvements result in a higher memory throughput in several measurements. However, other memory management performance measurement results are nearly the same for all three OSs except the number of hard page faults that is four time higher in Windows Vista than in Windows XP. Therefore, it can be concluded that numerous enhancements of memory management in Windows Vista and Windows 7 are not reflected in huge performance improvements. New CPU scheduling policies in Windows Vista and Windows 7 result in lower CPU performance than in Windows XP. This result is in complete contradiction with our expectations as the hardware in Experiment 1 has the multi-core processor and Windows Vista and Windows 7 have numerous enhancements in the CPU scheduling policy (processes, threads and sections) especially for scheduling on multi-core processors. In graphics display performance measurements, various results are obtained. In several measurements, Windows Vista and Windows 7 obtain better results and in other Windows XP achieves better results. As results are not constant, we cannot draw a conclusion which OS has the best display driver architecture. HDD management performance indicators show a lower throughput of HDD read/write operations in Windows Vista and Windows 7 than in Windows XP. Although, Windows Vista and Windows 7 do not have many improvements in HDD management, the results were expected to be at least the same as in Windows XP. Probably, if we had included a new hybrid HDD or a HDD with a larger physical sector size in our measurements, then Windows Vista and Windows 7 would have had better performance than Windows XP. Generally speaking, in Experiment 1, contrary to our expectations, Windows Vista and Windows 7 do not provide better overall performance than Windows XP. Several memory management and graphics display performance indicators show much better performance, but others show equal or even better performance in Windows XP. Except for few measurements, Windows Vista performance measurement results are similar to Windows 7 results. In addition, we can conclude that on the high-end computer systems memory-intensive applications will have best performance under Windows 7. On contrary, CPU-intensive applications will get more CPU processing power under Windows XP. Furthermore, multimedia (graphics-intensive) applications will have similar performance under all three Windows OSs but disk-intensive applications will again benefit under Windows XP.

7.2 Experiment 2

Windows 7 and especially Windows XP have showed much better memory management performance than Windows Vista. It is obvious that the Windows Vista memory management model requires huge amounts of physical memory and on low-end computer systems it reduces the memory performance. In Windows 7 we can see several improvements, especially in managing memory latency and lowering the number of hard page faults. CPU management performance measurement results, with few exceptions, are generally worse for Windows Vista and Windows 7. Times needed to calculate 4×10^6 and 16×10^6 digits of the number PI are shorter in Windows Vista (18.96% and 14.56%) and in Windows 7 (5.21% and 2.27%). This indicates that Windows Vista and Windows 7 have some improvements in managing FPU part of CPU. Graphics display performance measurement results are worst in Windows 7 and 3D graphics performance measurements failed although they finished correctly in Windows XP and Windows Vista. Only the throughput of drawing transparent windows in one second is much higher in Windows Vista and Windows 7, similarly to Experiment 1. This indicates that all new graphics features included in Windows Vista and Windows 7 requires powerful graphics hardware. Similarly to Experiment 1, Windows Vista and Windows 7 show a lower HDD data throughput, compared to Windows XP, when performing common operations. Generally, according to our expectations, in Experiment2, best performance is shown by Windows XP. It outperforms Windows Vista and Windows 7 in almost all measurements. When comparing Windows Vista to Windows 7, performance measurement results are quite different and neither OS shows convincingly better performance. In addition, we can conclude that on the low-end computer systems memory-intensive applications will again have best performance under Windows 7. However, CPU-intensive, multimedia (graphics-intensive) and disk-intensive applications will benefit under Windows XP.

7.3 Experiment 3

In TCP traffic measurements Windows Vista and Windows 7 mostly show similar results for both IPv4 and IPv6 protocols. However, Windows XP shows different performance results for IPv4 and IPv6. Windows XP IPv6 shows smaller delays and lower CPU usage but largest jitter and lower throughput when compared to IPv4. Generally, Windows Vista and Windows 7 have higher throughput and lower jitter values when compared to Windows XP. However, Windows XP has smallest delays, particularly with IPv6 protocol. CPU usage is generally lower in Windows XP but Windows 7 IPv4 has very low peaks for the some packet sizes. Network-intensive applications that require low delays would benefit from Windows XP IPv6. Nevertheless, network intensive applications that require high TCP throughput and lower jitter would benefit under Windows Vista and Windows 7.

In UDP traffic measurements there are again some deviations of Windows XP IPv4 and IPv6 but much smaller than in TCP traffic. For the smaller packet sizes up to 256 bytes and for the bigger packet sizes over the 1024 bytes Windows XP shows better performance of network indicators. However, for the medium-sized packets network performance is better in Windows Vista and Windows 7. For the CPU usage there is no clear pattern but Windows 7 CPU activity drops almost to the zero for bigger packet sizes than 1152 bytes. It can be concluded that the network-intensive applications with small or big packet sizes would benefit under the Windows XP. However, network-intensive applications with medium-sized applications would have better performance under Windows Vista or Windows 7.

7.4 Threats to the Validity of the Results

Since Windows OSs need to satisfy needs of all different kind of computer systems users, newer versions of Windows OSs are becoming increasingly complex and consist of an increasing number of various features and capabilities. Furthermore, in recent years the computer system hardware components are developing increasingly and are also becoming more complex. Consequently, there are numerous different factors that could influence the computer system performance. The hardware impact on the Windows OS performance is described in [Section 5.1]. We took into account all the recommendations and prepared the hardware in the manner to minimize the negative impact of hardware on the performance. By preparing Windows OSs through the performance measurement process described in [Section 5.3], we try to eliminate possible issues affecting performance measurement results. Several Windows Vista and Windows 7 features, as described in [Microsoft 09], are internal threats to the validity of measurement results:

- Background tasks: By executing theProcessIdleTasks function, we ensure that background tasks are run before performance testing. Benchmark applications do not periodically simulate user input and therefore, they do not prevent background tasks from running during idle periods that can occur during the tests.
- Memory management: During testing, the SuperFetch feature was not disabled, rather Windows OSs (Vista and 7) are "trained" to have improved performance. However, if SuperFetch is in the learning mode, it can impair the Windows Vista and Windows 7 performance. Furthermore, in Windows Vista and Windows 7 there is no exact value that shows how much benefit SuperFetch provides. Since benchmark applications run end-user scenarios at a speed that probably differs from the actual end user, these measurements might not benefit from SuperFetch and other performance enhancing features.
- The performance versus power consumption: Windows OSs reduce power consumption by optimizing the CPU performance and by scaling system performance for the current workload. Furthermore, different power plans can be chosen to save energy, maximize system performance, or achieve the balance between the two. In our experiments, the balanced power plan in all tested Windows OSs was chosen. However, we believe that in different versions of the Windows OS the same power plan does not contain exactly the same collection of hardware and system settings.
- Network equipment: Measuring network performance with additional network equipment (switches, hub, etc.) could have negative impact on performance. Furthermore, because of the unpredictable nature of wireless links it is often very hard to get reliable performance results of wireless networks. Therefore, we measured only performance of wired network between two directly connected identical computer systems.

Besides internal threats, several external threats, such as different versions of device drivers, hardware age, etc., can have a pronounced impact upon the Windows OSs performance. Real device driver impact on the OSs performance is not clearly defined in the literature. The recommendation is to use the newest device drivers as they should provide the best quality and performance.

In addition, all performance measurements were conducted on computer systems without any additional software installed in addition to the OS and the benchmark application. Furthermore, there were no additional data files on the HDD or user activities during the tests. Therefore, the obtained results could differ if performance measurements are conducted on loaded computer systems with additional resource demanding applications, such as security or network applications.

8 Conclusion

The main contribution of this paper is determination which of the three recent Windows OSs has better performance in different environments. Other contributions lie in developing the performance measurement process and performance evaluation model for recent Windows OSs. They are used for the performance evaluation of Windows XP, Windows Vista and Windows 7. However, they can be adopted for the future generation of Windows OSs. Performance measurement is done with the set of benchmark applications in three experiments. These experiments determine how OSs performance varies in different environments. A huge amount of performance measurement results allows us to evaluate various aspects of the Windows OSs. Results are evaluated with performance indicators that are measurable, independent and comparable between different versions of the Windows OS.

The obtained experimental results lead to the conclusion that, contrary to our expectations, Windows 7 and especially Windows Vista do not provide a better overall performance on the high-end computer system compared to Windows XP. Some performance improvements can be seen in memory management and graphics display, but other parts of these OSs have equal or lower performance than Windows XP. On the low-end computer system, Windows XP outperforms Windows Vista and Windows 7 in most tested areas. In comparison with Windows Vista in Experiment 1, Windows 7 shows slightly better performance, and in Experiment 2 Windows 7 shows much better performance of memory management, while Windows Vista has much better graphics display performance. In Experiment 3 Windows OS network performance depends on the packet size and used protocol. In general, Windows Vista and Windows 7 show similar results and compared to Windows XP they show better network performance particularly for the medium sized packets.

We also find significant that numerous performance measurement results are approximately equal in Windows Vista and Windows 7. This regularity may be the result of the same basic architecture both OSs share. Our study results could be useful to different types of computer system users, programmers and OS designers as they indicate which parts of Windows OSs have most improvements and where the bottlenecks are. Furthermore, they could be helpful for users who plan to migrate from Windows XP on the newer version of Windows OS.

In addition, since huge majority of high-end computer systems support 64-bit versions of the Windows OS, it would be beneficial to evaluate performance of 64-bit editions of Windows XP, Windows Vista and Windows 7 in the future work. The main difference between the 32-bit editions and the 64-bit editions of the Windows OS relates to the larger memory space on the 64-bit editions. Consequently, memory management should be improved and processes should be managed more efficiently. Furthermore, performance measurements should be provided with 64-bit versions of benchmark applications.

Acknowledgements

This work was supported by research project grant No. 165-0362980-2002 from the Ministry of Science, Education and Sports of the Republic of Croatia.

References

[Blythe 06] Blythe, D.: “The Direct3D 10 System”; ACM Transactions on Graphics, 25, 3 (2006), 724-734.

[Bott et al. 09] Bott, E., Siechert, C., Stinson, C.: “Windows 7 Inside Out”; Microsoft Press, Washington (2009)

[Botta et al. 07] Botta, A., Dainotti, A., Pescape A.: ”Multi-protocol and multi-platform traffic generation and measurement”; IEEE Conference on Computer Communications, DEMO Session, Anchorage, (2007)

[Fortier et al. 03] Fortier, P. J., Michel, H. E.: “Computer Systems Performance Evaluation and Prediction”; Digital Press, Burlington (2003)

[Hargreaves et al. 08] Hargreaves, C., Chivers, H., Titheridge, D.: “Windows Vista and Digital Investigations”; Digital Investigation Journal, 5, 1-2 (2008), 34-48.

[Hassel et al. 07] Hassell, J., Campbell, T.: “Windows Vista: Beyond the Manual”; Apress, New York (2007)

[Joukov et al. 06] Joukov, N., Traeger, A., Iyer, R., Wright, C. P., Zadok, E.: “Operating System Profiling via Latency Analysis”; In Proceedings of the 7th Symposium on Operating Systems Design and Implementation, Seattle (2006), 89-102.

[Kalakech et al. 04] Kalakech, A., Kanoun, K., Crouzet, Y., Arlat, J.: “Benchmarking the Dependability of Windows NT4, 2000 and XP”; In Proceedings of the International Conference on Dependable Systems and Networks, Florence, (2004), 681-686.

[Kolahi et al. 08] Kolahi, S. S., Narayan, S., Nguyen, D. D. T., Sunarto, Y., Mani, P.: “The Impact of Wireless LAN Security on Performance of Different Windows Operating Systems”; In Proceedings of the IEEE Symposium on Computers and Communications, Marrakech (2008), 260-264.

[Lavalys 09] Lavalys: “Everest”; January 2009. available at http://www.lavalys.com/products/everest-pc-diagnostics

[Lilja 00] Lilja, D. J.: “Measuring Computer Performance: A Practitioner's Guide”; Cambridge University Press, New York (2000)

[Martinovic et al. 07] Martinovic, G., Buljan, Z.: “Measurement of Timing Characteristics of Operating Systems”; In Proceedings of the 11th International Conference on Intelligent Engineering Systems, Budapest (2007), 287-290.

[Martinovic et al. 10] Martinovic, G., Balen, J., Rimac-Drlje, S.: “Impact of the host operating systems on virtual machine performance”; In Proceedings of the 33rd International Convention MIPRO, Opatija (2010), 613-618.

[Matthews et al. 08] Matthews, J., Trika, S., Hensgen, D., Coulson, R., Grimsrud, K.: “Intel Turbo Memory: Nonvolatile Disk Caches in the Storage Hierarchy of Mainstream Computer Systems”; ACM Transactions on Storage, 4, 2 (2008), 1-24.

[Microsoft 07] Microsoft Corporation: "Measuring Performance in Windows Vista"; July 2007. available at http://www.microsoft.com/whdc/system/sysperf/Vista_perf.mspx

[Microsoft 09] Microsoft Corporation: "Performance Testing Guide for Windows"; August 2009. available at http://www.microsoft.com/whdc/system/sysperf/Win7perf.mspx

[Microsoft 10] Microsoft Developer Network, April 2010. available at http://microsoft.com/msdn

[Narayan et al. 09] Narayan, S., Shang, P., Fan, N: "Performance Evaluation of IPv4 and IPv6 on Windows Vista and Linux Ubuntu"; In Proceedings of the International Conference on Networks Security, Wireless Communications and Trusted Computing, Wuhan, (2009), 653-656.

[Narayan et al. 10] Narayan, S., Shi, Y.: "TCP/UDP network performance analysis of windows operating systems with IPv4 and IPv6"; In Proceedings of the 2nd International Conference on Signal Processing Systems, Dalian, (2010), V2-219-V2-222.

[Net Applications 11] Net Applications: "Top Operating System Share Trend"; November 2011. available at http://www.netmarketshare.com/os-market-share.aspx?spider=1&qprid=11#

[Niemela 05] Niemela, S.: "PCMark05 PC Performance Analysis White Paper"; June 2005. available at http://www.futuremark.com/pressroom/companypdfs/PCMark05_Whitepaper_1_0.pdf

[Peng et al. 07] Peng, Y., Li, F., Mili, A.: "Modeling the Evolution of Operating Systems: An empirical Study"; The Journal of Systems and Software, 80, 1 (2007), 1-15.

[Pfeiffer 07] Pfeiffer Consulting: "User Interface Friction Benchmarks: Windows Vista vs. Windows XP and Mac OS X"; 2007. available at http://pfeifferreport.com/trends/Vista_UIF_Rep.pdf

[Renquist et al. 06] Renqvist, N., Kallinen, M.: "3DMark06 Whitepaper v1.0.2"; January 2006. available at http://www.futuremark.com/pressroom/companypdfs/3DMark06_Whitepaper_v1_0_2.pdf

[Russinovich et al. 04] Russinovich, M., Solomon, D.: "Microsoft Windows Internals, 4th Edition: Microsoft Windows Server 2003, Windows XP, and Windows 2000"; Microsoft Press, Washington (2004)

[Russinovich 07a] Russinovich, M.: "Windows Administration: Inside the Windows Vista Kernel: Part 1"; Microsoft Technet, CMP Technology, San Francisco (2007), available at http://technet.microsoft.com/en-us/magazine/2007.02.vistakernel.aspx

[Russinovich 07b] Russinovich, M.: "Windows Administration: Inside the Windows Vista Kernel: Part 2", Microsoft Technet, CMP Technology, San Francisco (2007), available at http://technet.microsoft.com/en-us/magazine/2007.03.vistakernel.aspx

[Salah et al. 09] Salah, K., Hamawi, M.: "Comparative packet-forwarding measurement of three popular operating systems"; Journal of Network and Computer Applications, 32, 5 (2009), 1039-1048.

[Salah et al. 10a] Salah, K., Hamawi, M.: "Impact of CPU-bound Processes on IP Forwarding of Linux and Windows XP"; Journal of Universal Computer Science, 16, 21 (2010), 3299-3313.

[Salah et al. 10b] Salah, K., Kahtani, A.: "Performance evaluation comparison of Snort NIDS under Linux and Windows Server"; Journal of Network and Computer Applications, 33, 1 (2010), 6-15.

[Schmid 07] Schmid, P.: “Windows XP vs. Vista: The Benchmark Rundown”; January 2007. available at http://www.tomshardware.com/reviews/xp-vs-vista,1531.html

[ScienceMark 06] ScienceMark 2.0 build 21mar05, August 2006. available at http://www.sciencemark.org

[Secherest et al. 01] Sechrest, S., Fortin, M.: “Windows XP Performance”; White Paper, Microsoft Corporation, June 2001. available at http://technet.microsoft.com/en-us/library/bb457057.aspx

[Silberschatz et al. 05] Silberschatz, A., Gagne, G., Galvin, P.: “Operating System Concepts, 7th Edition”; Wiley, Danvers, (2005)

[Smith 09a] Smith, A.: “Benchmarks: Windows 7 RTM versus Vista, XP”; August 2009. available at http://www.zdnet.com/blog/btl/benchmarks-windows-7-rtm-versus-vista-xp/22006

[Smith 09b] Smith, W.: “Windows 7 Review: XP vs Vista vs 7 in 80+ Benchmarks”; October 2009. available at http://www.maximumpc.com/article/reviews/windows_7_review

[SPEC 09] SPEC (Standard Performance Evaluation Corporation): “SPECviewperf 10”; September 2009. available at http://www.spec.org/gwpg/gpc.static/vp10info.html

[Tanenbaum 08] Tanenbaum, A. S.: “Modern Operating Systems, 3rd Edition”; Prentice Hall PTR, New Jersey (2008)

[VMware 07] VMware, Inc.: “Performance Tuning and Benchmarking Guidelines for VMware Workstation 6“; May 2007. available at http://www.vmware.com/pdf/WS6_Performance_Tuning_and_Benchmarking.pdf

[Williams 09] Williams, M.: “Windows compared: Windows 7 vs Vista vs XP”; July 2009. available at http://www.techradar.com/news/software/operating-systems/windows-compared-windows-7-vs-vista-vs-xp-615167

[Xtreme 08] Xtreme Systems Inc.: “Super PI”; February 2008. available at http://www.xtremesystems.com/superpi

Appendix

Appendix provides a detailed hardware and software configuration used in Experiment 1, Experiment 2 and Experiment 3.

	Specifications
CPU	
Intel	QuadCore Core i7 940
Instruction Set	64-bit
Internal Clock	2933.0 MHz
External Clock	133.0 MHz
Number of Cores	4
Voltage	1.2 V
Cache Memory	
L1	128 KB
L2	1024 KB
L3	8 MB
RAM	
Corsair Dominator	3 x 2 GB DDR3-1333 SDRAM 1066 MHz
Graphics hardware	
NVIDIA	GeForce GTX 295
Memory	1792 MB GDDR3 (896MB per GPU)
Graphics Clock	576 MHz
Processor Clock	1242 MHz
Memory Clock	999 MHz
Hard disk	
Seagate	Barracuda ST3500418AS
Capacity	500 GB
Interface	SATA 3Gb/s
Cache	16MB
Spindle Speed	7200 rpm (revolutions per minute)
Motherboard	
Asus	P6T WS Pro
Chipset	Intel® X58 / ICH10R
Memory architecture	Triple channel
Bios version	0711
USB Flash drive	
Corsair	VoyagerGT USB 2.0
Capacity	16 GB
File system	FAT32

Table 17: The hardware configuration in Experiment 1

Component	**Driver Provider**	**Driver version**
Windows XP		
CPU	Microsoft	5.1.2600.0
Graphics hardware	NVIDIA Corporation	6.14.11.9062
HDD	Microsoft	5.1.2535.0
Chipset	Intel	9.1.0.1007
Windows Vista		
CPU	Intel	9.0.0.1005
Graphics hardware	NVIDIA Corporation	8.16.11.9062
HDD	Microsoft	6.0.6002.18005
Chipset	Intel	9.1.0.1007
USB Flash drive	Microsoft	6.0.6002.18005
Windows 7		
CPU	Intel	9.0.0.1005
Graphics hardware	NVIDIA Corporation	8.16.11.9062
HDD	Microsoft	6.1.7600.16385
Chipset	Intel	9.1.0.1007
USB Flash drive	Microsoft	6.1.7600.16385

Table 18: Device drivers in Experiment 1

	Specifications
CPU	
Intel	Intel Celeron D 330
Instruction Set	32-bit
Internal Clock	2660.0 MHz
External Clock	533.0 MHz
Number of Cores	1
Voltage	3.3 V
Cache Memory	
L1	16 KB
L2	256 KB
RAM	
Samsung M3	2 x 512 MB PC3200 DDR 333 MHz
Graphics hardware	
Intel	82865G Graphics Controller
Memory	max. 96MB of system memory
Graphics Clock	266 MHz
RAMDAC Clock	350 MHz
Hard disk	
Samsung	SP0411CC
Capacity	40 GB
Interface	SATA 150 Mb/s
Cache	2 MB
Spindle Speed	7200 rpm (revolutions per minute)
Motherboard	
FUJITSU SIEMENS	D1561
Chipset	Intel Springdale-G i865G
Memory architecture	Dual channel
Bios version	5.00 R2.14.1561.02
USB Flash drive	
Corsair	VoyagerGT USB 2.0
Capacity	16 GB
File system	FAT32

Table 19: The hardware configuration in Experiment 2

Component	Driver Provider	Driver version
Windows XP		
CPU	Microsoft	5.1.2600.5512
Graphics hardware	Intel Corporation	6.14.10.3889
HDD	Microsoft	5.1.2535.0
Chipset	Intel	8.3.0.1014
Windows Vista		
CPU	Microsoft	6.0.6000.16386
Graphics hardware	Intel Corporation	6.14.10.4656
HDD	Microsoft	6.0.6002.18005
Chipset	Intel	8.3.0.1014
USB Flash drive	Microsoft	6.0.6002.18005
Windows 7		
CPU	Microsoft	6.1.7600.16385
Graphics hardware	Microsoft	6.1.7600.16385
HDD	Microsoft	6.1.7600.16385
Chipset	Microsoft	6.1.7600.16385
USB Flash drive	Microsoft	6.1.7600.16385

Table 20: Device drivers in Experiment 2

	Specifications
CPU	
Intel	Dual-Core CPU E5200
Instruction Set	64-bit
Internal Clock	2500.0 MHz
External Clock	200.0 MHz
Number of Cores	2
Voltage	1.2 V
Cache Memory	
L1	32 KB per core
L2	32 KB per core
L3	2 MB
RAM	
Corsair Dominator	2 x 2 GB DDR2 800 MHz
Graphics hardware	
Asus	EAH3650
Memory	1024 MB DDR2
Graphics Clock	724 MHz
Processor Clock	724 MHz
Memory Clock	400 MHz
Hard disk	
Western Digital	WD2500AAKS-00VSA0
Capacity	250 GB
Interface	Serial-ATA 3.0 Gbps
Cache	16 MB
Spindle Speed	7200 rpm
Motherboard	
Asus	P5K-VM
Chipset	G33
Memory architecture	Dual channel
Bios version	0702
Network	
LAN (onboard)	Marvell88E8056 PCI-E Gigabit LAN

Table 21: The hardware configuration in Experiment 3

Windows OSs updates in Experiment 1, Experiment 2 and Experiment 3 (without security and driver updates):

- Windows XP updates: KB898461, KB951978, KB967715, KB968389, KB973815; and hotfixes: KB952287 and KB970653-v3.
- Windows Vista updates: KB943729, KB905866, KB968389, KB970653, KB972036, KB948465 (SP), KB955430, KB938371, KB937287 and KB935509.
- Windows 7 update: KB974332.