



Project Number	IST-2006-033789
Project Title	PLANETS
Title of Deliverable	Second version of Dioscuri
Deliverable Number	D6
Contributing Sub-project and Work-package	PA/5
Deliverable	Internal
Dissemination Level	Public
Deliverable Nature	Report
Contractual Delivery Date	M30
Actual Delivery Date	M31
Author(s)	KB-NL

Abstract:

This deliverable describes the second version of Dioscuri, one of the tools for environments in PA/5. This follow-up version incorporates some important improvements, primarily better support for GRATE integration and backwards compatibility with earlier version of the Java Runtime Environment (JRE 1.5). This document is intended for both project participants and technical specialists. It describes the history, design, functionality, technical environment and future work of the latest release of Dioscuri within the Planets project. It does not contain a user manual for the tool, because user manual and installation notes can be downloaded from the appropriate project website, together with the released software.

Document Control

Contributors

Person	Role	Partner	Contribution
Jeffrey van der Hoeven	PA/5	KB-NL	Document owner and author
Remco Verdegem	PA/5	NANETH	Review and remarks

1 Executive Summary

One of the main goals of PA/5 is to develop and test emulation tools, and tools related to emulation. Within PA/5 three tools are being developed: Dioscuri (a modular emulator for digital preservation), UVC (the universal virtual computer) and GRATE (remote emulation access tool).

All tools were already under development by Planets partners before the Planets project started. Within the Planets context these tools are further extended. Deliverable PA/5-D2 showed the first results of the evolved tools.

This document, as part of PA/5-D6, is focusing on the second release of Dioscuri developed within the Planets project. In this document the history, design, functionality, technical environment and future work of the latest version of Dioscuri is described. Further user documentation can be downloaded from the project website, together with the released software.

1.1 Dioscuri (second version)

Development of Dioscuri started in 2006 by KB-NL, NANETH and TSS leading the technical development team. This project officially ended on 1 July 2007, but development of Dioscuri continued under the umbrella of Planets by KB-NL and TSS.

Dioscuri is designed following a modular architecture in which each computer hardware component is represented as a module in software. All modules together form the actual emulation process. This new version of Dioscuri is capable of emulating 16 bit and experimental 32 bit processing which offers support for Intel 8086 up until Intel 80386 based computer systems. It runs technical environments such as MS-DOS, FreeDOS, Windows 3.0 and a simple Linux kernel. It can execute old DOS applications such as WordPerfect 5.1, Norton Commander and many old games.

Future work will consist of improving the virtual processor to fully cover 32-bit processing, improving performance and extending the set of peripheral devices with mouse, sound and network (optional).

Dioscuri version 0.4.0 can be downloaded from: <http://dioscuri.sourceforge.net>

Table of Contents

1	<i>Executive Summary</i>	3
1.1	Dioscuri (second version)	3
2	<i>Introduction</i>	5
3	<i>Dioscuri (second Planets version)</i>	6
3.1	Overview	6
3.2	Description	6
3.3	Screenshots	7
3.4	History	7
3.5	Design	8
3.5.1	Modular emulator	9
3.5.2	Universal Virtual Machine (UVM).....	9
3.5.3	Controller, module library and ESD	9
3.6	Functionality	10
3.7	Technical environment	10
3.8	Development notes	10
3.9	Testing	11
3.9.1	User software testing	11
3.9.2	Remote emulation.....	11
3.10	Future work	12
3.10.1	KEEP	13
3.11	References	13

2 Introduction

One of the main goals of PA/5 is to develop and test emulation tools, and tools related to emulation. Within PA/5 three tools are being developed: Dioscuri (a modular emulator for digital preservation), UVC (the universal virtual computer) and GRATE (remote emulation access tool).

All tools were already under development by Planets partners before the Planets project started. Within the Planets context these tools are further extended. Deliverable PA/5-D2 showed the first results of the evolved tools.

Development of Dioscuri started in 2006 by KB-NL, NANETH and TSS leading the technical development team. This project officially ended on 1 July 2007, but development of Dioscuri continued under the umbrella of Planets since then by KB-NL and TSS.

Dioscuri is designed following a modular architecture in which each computer hardware component is represented as a module in software. All modules together form the actual emulation process. This new version of Dioscuri is capable of emulating 16 bit and experimental 32 bit processing which offers support for Intel 8086 up until Intel 80386 based computer systems. It runs technical environments such as MS-DOS, FreeDOS, Windows 3.0 and a simple Linux kernel. It can execute old DOS applications such as WordPerfect 5.1, Norton Commander and many old games.

This document, as part of PA/5-D6, is focusing on the second official version of Dioscuri developed within the Planets project. In this document the history, design, functionality, technical environment and future work of the latest version of Dioscuri is described. Further user documentation can be downloaded from the project website, together with the released software.

This deliverable consists of two parts:

- Latest release of Dioscuri (0.4.0)
- This document

This document is intended for both project participants and technical specialists. It describes the history, design, functionality, technical environment and future work of the latest release of both tools within the Planets project. Some parts may be a bit technical, but are not necessary to understand the functionality of the tools.

Furthermore, this document does not contain a user manual for the tool, because user manual and installation notes can be downloaded from the appropriate project websites, together with the released software.

3 Dioscuri (second Planets version)

3.1 Overview

Name: Dioscuri – modular emulator for digital preservation

Current version: 0.4.0

Release date: 11 December 2008

Availability: public, open source

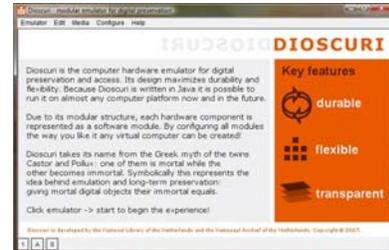
License: GNU GPL version 2

Programming language: Java (version 1.5 or higher)

Organisation: KB-NL

Reference: <http://dioscuri.sourceforge.net>

User manual: see project website



3.2 Description

Dioscuri is an x86 computer hardware emulator written in Java. It is designed by the digital preservation community to ensure documents and programs from the past can still be accessed in the future.

The Dioscuri emulator has two key features: it is durable and flexible. Because it is implemented in Java, it can be ported to any computer platform which supports the Java Virtual Machine (JVM), without any extra effort. This reduces the risk that emulation will fail to work on a single architecture in the future, as it will continue to work on another architecture.

Dioscuri is flexible because it is completely component-based. Each hardware component is emulated by a software surrogate called a module. Combining several modules allows the user to configure any computer system, as long as these modules are compatible. New or upgraded modules can be added to the software library, giving the emulator the capability to run these.

Dioscuri takes its name from the Greek myth of the twins Castor and Pollux: one of them is mortal while the other becomes immortal. Symbolically this represents the idea behind emulation and long-term preservation: giving mortal digital objects their immortal equals.

3.3 Screenshots

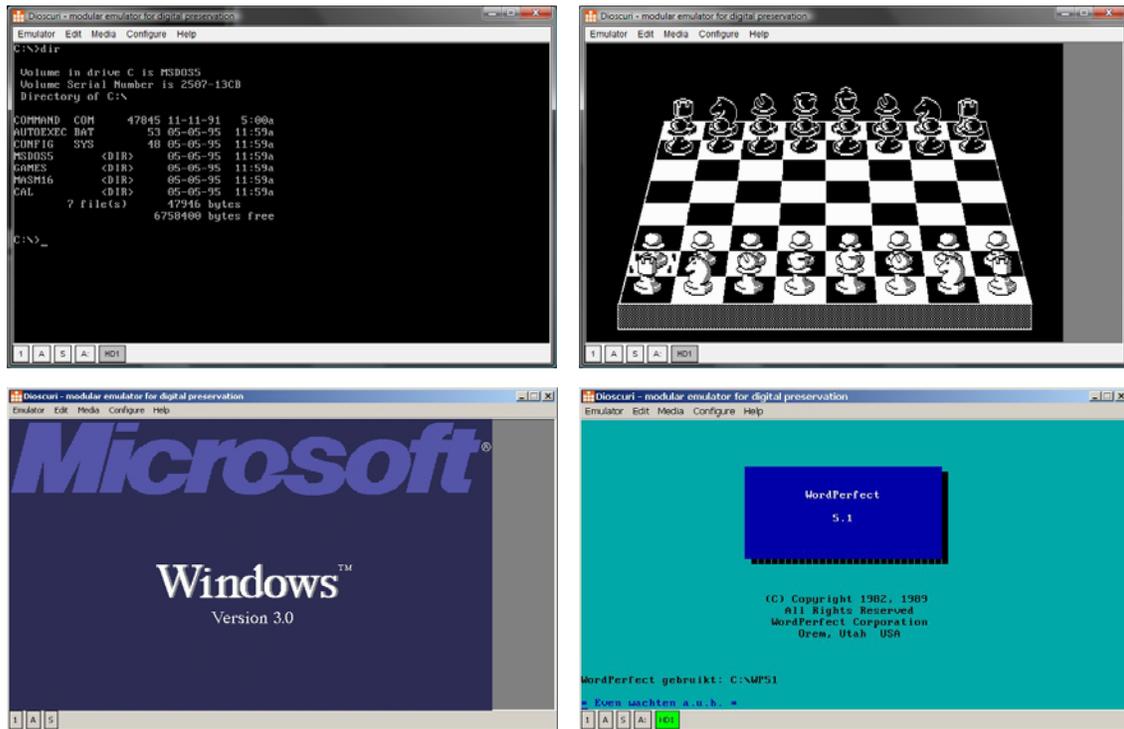


Figure 1: screenshots of Dioscuri

3.4 History

To prevent a digital black hole, the Koninklijke Bibliotheek (KB) - National Library of the Netherlands - and the Nationaal Archief of the Netherlands started a joint project to research and develop a solution. Both institutions have a large amount of traditional documents and are very familiar with preservation over the long term. However, the amount of digital material (publications, archival records, etc.) is increasing with a rapid pace. To manage them is already a challenge. But as cultural heritage organisations, more has to be done to keep those documents safe for hundreds of years at least.

The National Library and the National Archives of the Netherlands have been active in the field of digital preservation for many years. Various researches pointed out that basically two strategies exist to retain access to digital material: migration and emulation.

Where migration alters the original document into a new version, emulation changes the current computer environment to mimic the environment in which the document was originally created and rendered. For example, with migration an old WordPerfect document could be migrated to MS Word. This gives the user the benefit of using the document in his current environment. However, migration has the risk of changing the document's content, layout or structure significantly which may be undesirable.

With emulation, the old MS-DOS environment with WordPerfect as text processing software can be run. This allows the document to be viewed in its native environment and guarantees the user that the document's representation has not changed. This can be quite significant, especially for applications which cannot be migrated.

Although many people recognise the importance of having a digital preservation strategy based on emulation, it has never been taken into practice. Of course, many emulators already exist and show the usefulness and advantages they can offer. But none of them have been designed to be digital preservation proof. For this reason the National Library and Nationaal Archives of the Netherlands started a joint project on emulation.

The aim of the emulation project was to develop a new preservation strategy based on emulation. Our goal is to build an emulator that is durable and able to replace the Reference Workstation (RWS, a modern computer environment using MS Windows 2000 that is used as a reference for what the emulator should be capable of.). As a result, all kinds of digital material, like multimedia programs, databases and PDF documents will remain accessible in the long term. The emulation project started in April 2005 and ended in July 2007. From then on, development has taken place within the Planets project.

3.5 Design

To develop an emulator for digital preservation, two main requirements have to be met: the emulator must be durable and flexible. Durability is essential for keeping the emulator sustainable for the long term. Flexibility is desired to ensure several different computer environments can be emulated with a minimum amount of effort. Based on a unique design that is defined by the National Library of the Netherlands, in cooperation with emulation-expert Jeff Rothenberg, a Modular Emulator has been developed. The illustration below shows the following basic elements of this design:

- Modular emulator
- Universal Virtual Machine (UVM)
- Controller
- Module library
- Emulator specification document (ESD)

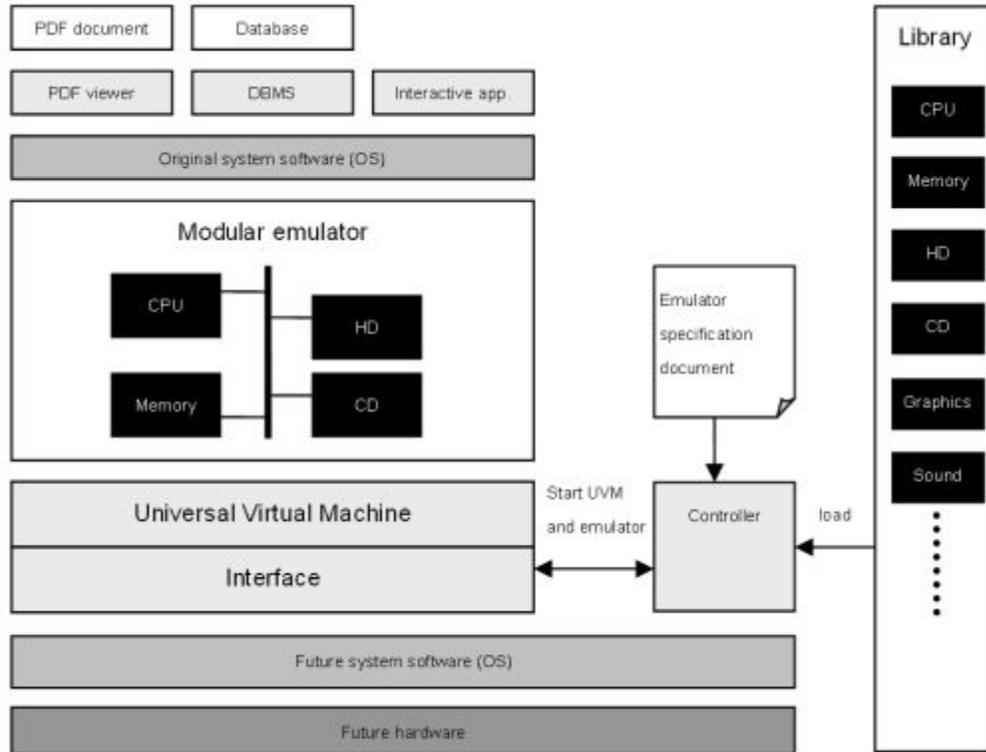


Figure 2: conceptual model of modular emulator

3.5.1 Modular emulator

The emulator is the core element of the design. The task of the emulator is to virtually recreate the hardware of the upper (emulated) platform in such a way that original software can run on it as it used to do on real hardware. Because of the versatility of hardware, system software and applications, the emulator needs to be flexible in its recreation of a particular environment. This can be achieved by designing the emulator in a modular structure. Each module emulates a specific hardware component, like processor, memory, keyboard, etc. The recreation of a platform can then be realized by combining all necessary modules. Changing the emulated platform can simply be done by changing the configuration of the modules. As technology moves forward, new modules can be created based on their analogy with actual hardware components.

3.5.2 Universal Virtual Machine (UVM)

By introducing an intermediate layer (UVM) between the current platform and the emulator, dependency on the underlying system is avoided. This offers more stability and less maintenance of the emulator as changes in the computer platform underneath the UVM do not have any (or have little) effect on the execution of the emulator. In the current implementation, Java Virtual Machine (JVM) has been chosen as the UVM.

3.5.3 Controller, module library and ESD

The controller interconnects all the elements in the model. It configures and controls the emulation process. The Emulator Specification Document (ESD) describes which modules are required to emulate the desired platform. Based on a particular ESD the controller selects the required modules from the module library and configures them. This has now been implemented as part of the GUI and configuration loader, but will be extended in the near future.

3.6 Functionality

Version 0.4.0 of Dioscuri offers a wide range of emulated computer components. The complete list is shown below:

- 16-bit Intel 8086-based CPU
- 32-bit Intel 80286/80386-based CPU
- 1 MB RAM (expandable)
- Storage devices:
 - 2 floppy drives (all formats up to 1.44 MB)
 - 1 Hard Disk Drive (HDD, up to 1 GB)
- Input devices:
 - XT/AT/PS2 compatible keyboard
 - Serial/PS2 mouse (experimental)
- Output devices: virtual screen
- VGA video graphics adapter
- Serial port (UART 16550A with COM-ports 1 to 4)
- DMA-support for fast data transfer
- IRQ-handling based on a Intel 8259 PIC
- Timing mechanism based on an Intel 82C54 PIT and Crystal Clock
- Real-time clock with integrated CMOS for user-defined BIOS settings
- System BIOS using Plex86/Bochs BIOS
- Video BIOS using VGA LGPI'ed BIOS

With this set of components, Dioscuri is able to emulate an IBM compatible Personal Computer (PC) with an Intel x86-based processor up to 80386 processing power.

All components can be configured in an XML-based configuration file. This file can be altered via an easy-to-use graphical user interface (GUI) or directly in the file itself. The emulation process can be started via the same GUI and takes the XML-file as input.

Since version 0.4.0 Dioscuri can be started without user interaction by supporting a command-line interface. This allows external emulation services to load and run Dioscuri automatically.

3.7 Technical environment

Dioscuri is able to emulate the hardware of a Intel 8086 computer platform with the peripheral support as mentioned in section "Functionality". With this basic set of hardware support, Dioscuri can support the following operating systems:

- MS-DOS 4.0, 5.0 and 6.22 (also supporting the MS Windows 3.0 application)
- FreeDOS 0.9 Beta
- ELKS 16-bit Linux

Within this second development cycle of Dioscuri the focus is on running more operating systems (MS-DOS and Windows) and applications such as DrawPerfect 1.1, Lotus 1-2-3 and an early web browser designed for MS-DOS, called Arachne. Future versions of Dioscuri will be aimed to run more modern versions of the Microsoft Windows family, such as Windows 3.11, 95, 98 and 2000, and better support for Linux.

3.8 Development notes

Until 1 July 2007, development of Dioscuri has been lead by software company TSS. After then, KB-NL (supported by TSS) is performing new improvements to Dioscuri. Although the focus is on building a full 32-bit emulator which is capable of running modern versions of MS Windows, the

intermediate result is focused on 16-bit emulation and experimentally 32-bit as well. Furthermore, support is available for (virtual) floppy disks, hard disks, input devices like keyboard, and a screen as output device. Moreover, the emulator can be configured easily to additional user requirements based on an XML file and can be loaded and started via a command-line interface.

Java has been chosen as programming language. The main reason is the possibility to run Java code on a large variety of computer platforms using the Java Virtual Machine (JVM) without the need to recompile the source code. Recompilation is a risk factor since program code is adapted to a specific underlying computer platform. Although today's compilers are very advanced, it can still result in different functionality of the same emulator compiled for different platforms. By avoiding this step, the emulator will be more stable and portable. Using the JVM instead provides similar results to the use of a UVM as discussed above.

Developing an emulator requires a thorough understanding of the computer platform that is to be emulated. Documentation is crucial. Although much information on computer devices is available, it is widespread over various sources. To ensure this information will not be lost, all documentation is centrally organised and made available via the Dioscuri website.

3.9 Testing

During and after development of Dioscuri, several tests have been performed to check if Dioscuri is mimicking the original hardware correctly. Aside from small tests aimed at specific hardware components, integrated tests have been performed to see if Dioscuri behaves as a "normal computer".

3.9.1 User software testing

Dioscuri has been tested with the following user software:

Table 1: Test results overview Dioscuri

Operating systems	Result
MS-DOS 4.0	OK
MS-DOS 5.0	OK
MS-DOS 6.22	OK
FreeDOS 0.9 Beta	OK
ELKS 16-bit Linux	OK
Applications	Result
WordPerfect 5.1	OK
DrawPerfect 1.0	Unstable
Norton Commander 3.0	OK
Chess	OK
Block-out	OK
Arachne (DOS-based web browser)	OK
Windows 3.0	Unstable

3.9.2 Remote emulation

Furthermore, Dioscuri has been tested in combination with GRATE (Global Remote Access to Emulation-Services) to run emulation processes remotely via a network. This setup complies with *solution 4: web-based remote emulation* proposed in deliverable PA/5-D4 which is depicted in figure 3.

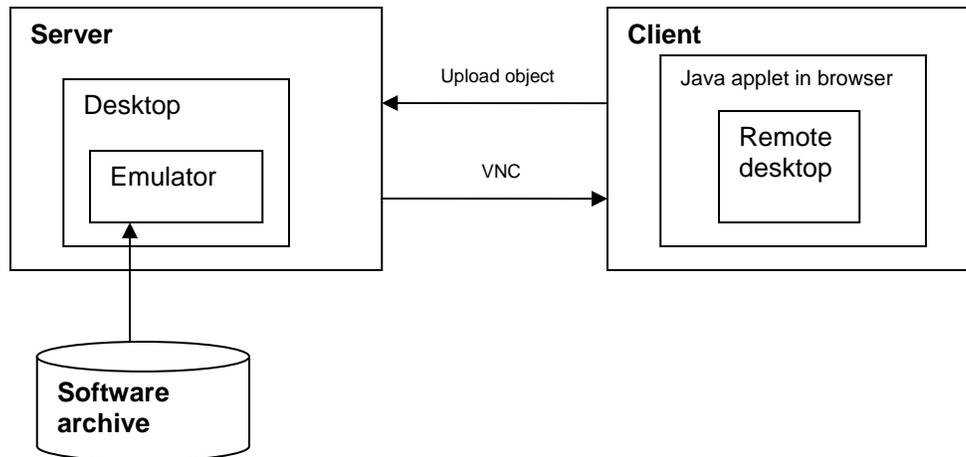


Figure 3: remote emulation solution 4 (PA/5-D4)

This configuration offers the most advantages as the emulation process can be controlled via a client connected to a server where the actual processing takes place. This frees the user from having to install, configure and run the emulation process locally on the computer. With the XML-configurator and the command-line interface, Dioscuri can easily be started in an automated environment such as GRATE. Dioscuri showed to work well remotely, although inserting and ejected floppy disks virtually seem to be problematic still.

Another test concerning remote emulation was performed by running Dioscuri as a Java applet in a web browser. This setup complies best with scenario 1 of PA/5-D4 in which the user sets up an emulation process by first downloading the emulator and then run it locally. The only difference is that the Applet does not have to be installed and configured, because this is done automatically when running the applet in the browser. This test showed that Dioscuri can be ported quite easily to a Java applet. A disadvantage is that an applet is somewhat limited in available resources which hampers switching disk images on the fly. This can be done more easily with GRATE.

3.10 Future work

Although the current version is already very useful for emulating older computer platforms, the next iteration of Dioscuri should include the following functionality:

- Improved support for 32-bit protected mode, for running Windows 3.11 and higher
- Improved performance
- Improved mouse support
- CD-ROM support
- Improved data extraction and insertion (via clipboard and other interfaces)
- Metadata for each module based on metadata schema
- Improved exception handling model

Other future considerations (optional):

- Sound support
- Store/restore state functionality
- Introduce shortcut keys to menu items
- Improve GUI and use a better status update mechanism between GUI and modules
- Module library

- Network support
- Writing a test suite (unit tests, etc.)

3.10.1 KEEP

The abovementioned list is ambitious and requires probably more development time than still available within the Planets project. To ensure that development on Dioscuri continues, the Planets results regarding emulation will be taken into account in a new European project KEEP (Keeping Emulation Environments Portable). The focus of this project is to create a full-fledged emulation access platform offering emulation services that are preservation proof and applicable for a wide range of users (libraries, archives, universities, museums and game industry). Dioscuri will become one of the core technical components of this platform and will be extended where necessary to allow authentic rendering of computer environments.

3.11 References

Dioscuri website and documentation (user manual, requirements, design, implementation, PC architectural documents, etc.):

<http://dioscuri.sourceforge.net>

Project page emulation, Koninklijke Bibliotheek:

http://www.kb.nl/hrd/dd/dd_projecten/projecten_emulatie-en.html

Project page emulation, Nationaal Archief of the Netherlands:

<http://www.digitaleduurzaamheid.nl/index.cfm?paginakeuze=355>

J.R. van der Hoeven, B. Lohman, R. Verdegem, Requirements for applying emulation as a preservation strategy, IS&T Archiving Conference 2008, Bern, Switzerland, 2008. In: Archiving 2008, final program and proceedings, Bern, Switzerland, 2008

R. Verdegem, J.R. van der Hoeven, Emulation: To Be or Not To Be, proceedings IS&T Archiving conference, Ottawa, Canada, 2006.

J.R. van der Hoeven, H.N. van Wijngaarden, Modular emulation as a long-term preservation strategy for digital objects (IWAW, Vienna, Austria, 2005). Available at:

<http://www.iwaw.net/05/papers/iwaw05-hoeven.pdf>