Distinguished Lecturer Series “Leon the Mathematician” at the Department of Informatics, Aristotle University of Thessaloniki Greece (http://dls.csd.auth.gr)

Department of Informatics
Aristotle University of Thessaloniki

CELEBRATING
1992 - 2012
YEARS

INVITED LECTURE

Professor Diomidis Spinellis (Athens University of Economics and Business, Greece), senior member of the ACM and the IEEE, is going to give a lecture titled

Farewell to Disks: Efficient Processing of Obstinate Data

at Room A, Department of Informatics, Aristotle University of Thessaloniki, **Ethnikis Antistaseos 16 (2nd floor), Thessaloniki 55133** on **Wednesday March 7th, 2012 at 17:00.**

ABSTRACT

Questions whose answer requires sophisticated processing of huge data sets come up increasingly often in our networked and interlinked, and (increasingly) DNA-sequenced world. Attacking such problems with traditional techniques, such as loading data into memory for processing or querying a relational database, is cumbersome and inefficient. Data sizes are growing inexorably, while disk-based data structures and applications relying on them, optimized to handle sequential retrievals and relational joins, often prove inadequate for running complex algorithms. Therefore, sophisticated processing of huge complex data sets requires us to rethink the relationship between disk-based storage and main-memory processing.

Some features of modern systems, namely 64-bit architectures, memory mapped sparse files, virtual memory, and copy on write support, allow us to process our data with readable and efficient RAM-based algorithms, using slow disks and file systems only for their large capacity and to secure the data’s persistence. I demonstrate this approach through a series of C++ programs that run on Wikipedia’s data looking for matching words and links between unrelated entries. Through these programs I will show how we can use STL containers, iterators, and algorithms to access disk-based data without performing any system calls. Although RAM-based processing opens up again many problems that database systems already solve, I will argue that such processing is the right move, because it provides us with a unified programming and performance model for all our data operations irrespective of where the data resides.

About the Speaker:

Professor **Diomidis Spinellis**
Department of Management Science and Technology
Athens University of Economics and Business
Email: dds@aueb.gr
URL: http://www.dmst.aueb.gr/dds/

Diomidis Spinellis is a Professor in the Department of Management Science and Technology at the Athens University of Economics and Business, Greece. From November 2009 to November 2011 he served as the Secretary General for Information Systems at the Greek Ministry of Finance. His research interests include software engineering, computer security, and programming languages. He has written the two award-winning "Open Source Perspective" books: "*Code Reading*" and "*Code Quality*" as well as dozens of scientific papers. He is a member of the *IEEE Software* editorial board, authoring the regular "Tools of the Trade" column. Dr. Spinellis has contributed code to the FreeBSD operating system and is the developer of UMLGraph and other open-source software packages, libraries, and tools. He holds a MEng in Software Engineering and a PhD in Computer Science, both from Imperial College London.

Off campus premises of the Department of Informatics, Aristotle University of Thessaloniki: