A Systematic Mapping Study on High-level Language Virtual Machines

Vinicius H. S. Durelli, Katia R. Felizardo, and Marcio E. Delamaro

Computer Systems Department
University of São Paulo (ICMC-USP)
13560-970 – São Carlos – SP – Brazil
{durelli,katiarf,delamaro}@icmc.usp.br

October 17, 2010

Agenda

Research on High-level Language Virtual Machines

- A great deal of the contemporary high-level languages have their execution environment based upon high-level language virtual machines (HLL VMs).
- There is a **large body of literature** on research in virtual machine for high-level languages.
  - A mature research area means a **sharp increase** in the number of results made available, thus it becomes **essential** to **summarize** and provide an **overview** of such area.
- To the best of our knowledge **there are no comprehensive studies** focusing on an overview of this research area and its most investigated subjects.

Motivation: First Step Towards Filling in Such *Gap*

- In order to fill in such a gap it is needed to ascertain the **nature**, **extent**, and **quantity** of published research papers.

Contributions:

1. Areas that have been most subjected to investigation.

   Side effect: Areas that require further research.
2. The relevant publication forums.
3. HLL VM implementations that are the most widely used within the academic community.

Evidence-based Paradigm

Definition → Systematic Mapping

Methodology that involves searching the literature in order to aggregate and categorize primary studies, thereby yielding a synthesized view of the research area under consideration [Petersen et al., 2008].

Advantages:

- The approach used for searching and inclusion and exclusion criteria are defined in a research protocol and reported as an outcome.

  Side effects: Transparent;
  Replicable;
  Updatable.

Systematic Mapping Process: Overview

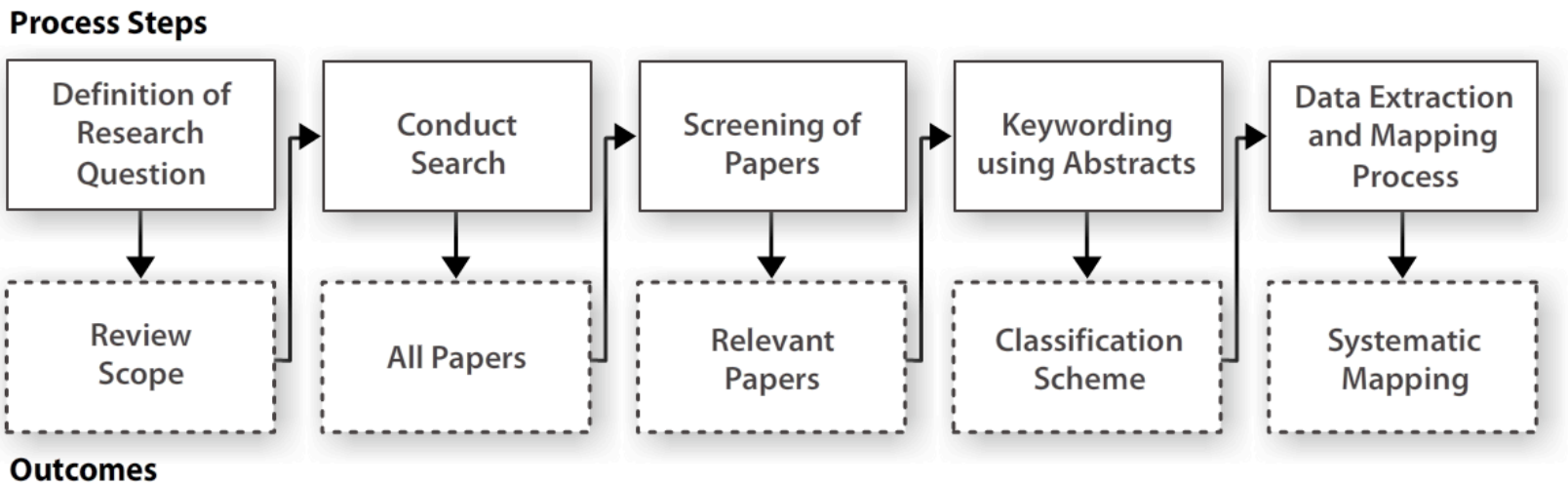


Figure: The systematic mapping process [Petersen et al., 2008].

Research Questions

Research questions must **embody the mapping study purpose**.

RQ_1: which functionalities/features/characteristics of HLL VMs have been most investigated?

RQ_2: which are the mainstream HLL VM implementations within the academic community?

Search for Primary Studies

Search String → combination of these keywords and acronyms

virtual machine, VM, high-level language virtual machine, and HLL VM.

We used the search string on the following **electronic databases**:

- ACM Digital Library,
- EngineeringVillage,
- IEEE Xplore,
- Springer Lecture Notes in Computer Science (LNCS), and
- ScienceDirect.

- **No limits** were placed on **date of publication**.

Screening: Inclusion Criteria

The inclusion criteria **devised** and **applied** are:

- if several papers reported similar studies, only the most recent was selected;
- papers describing more than one study had each study individually evaluated;
- it has to describe at least a prototypical implementation of the proposed improvement, thereby mentioning the HLL VM implementation that was modified.

Screening: Exclusion Criteria *(i)*

and the following exclusion criteria:

- papers that do not present studies pertaining to HLL VMs, e.g., papers describing research on system VMs;
- studies describing the introduction of improvements that consist in solely modifying the intermediate language of the HLL VM under consideration;
- studies whose proposed enhancements do not imply in making changes to the underlying HLL VM, e.g., papers describing features implemented atop HLL VMs;

Screening: Exclusion Criteria *(ii)*

- studies whose target HLL VM is either a co-designed (e.g., composed of both software and hardware portions) or an entirely implemented in hardware HLL VM;
- technical reports, documents that are available in the form of either abstracts or presentations (i.e., elements of *"grey"* literature), and secondary literature reviews (i.e., mapping studies).

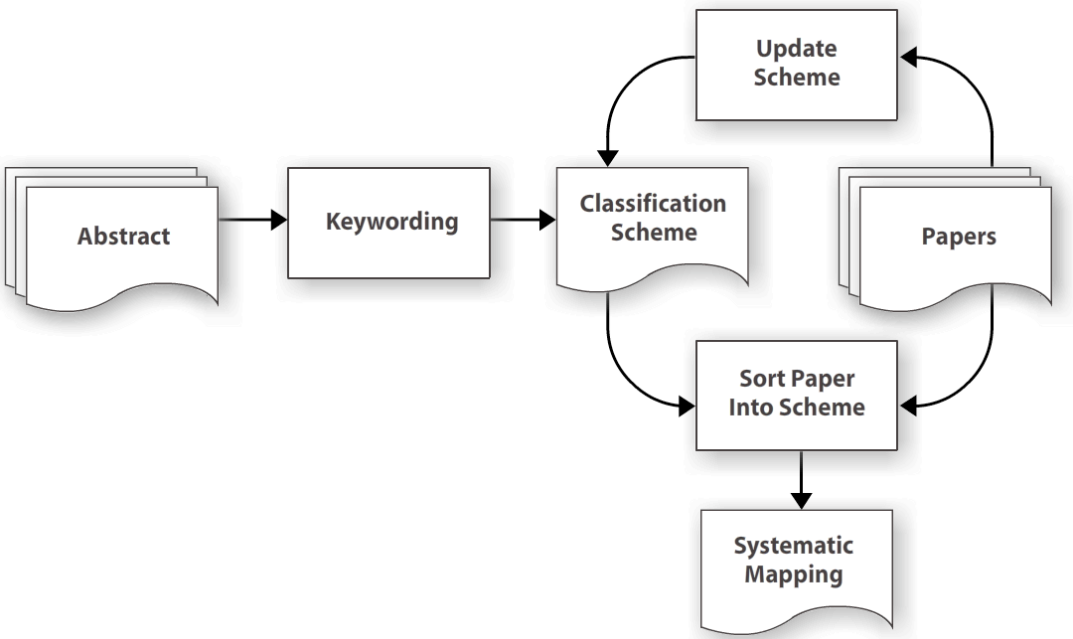
Final Set of Selected Primary Studies

Electronic Database	**Number**
ACM Digital Library	1554
EngineeringVillage	1395
IEEE Xplore	309
Springer LNCS	640
ScienceDirect	1123
Total	5021
Candidates	142
Final set	**128**

Table: Papers retrieved from each electronic database, total of candidate studies, and the final set.

Keywording

- The aim of this step is to devise our own **classification scheme** and **categories** for the selected primary studies.
- **Certain sections** are read for the purpose of finding **keywords and concepts** that reflect their contribution.



Resulting Categories

Categories
Optimization
Garbage Collection (GC)
Debugging
Memory Leak Tolerance (MLT)
New Language Construct (NLC)
Profiling
Aspect-Oriented Programming (AOP)
Embedded System (ES)
Security
Real-Time
Distributed Computing (DC)
Fault Tolerance (FT)
Resource Sharing among HLL VMs (RSVM)
Testing

Resulting Category Frequencies

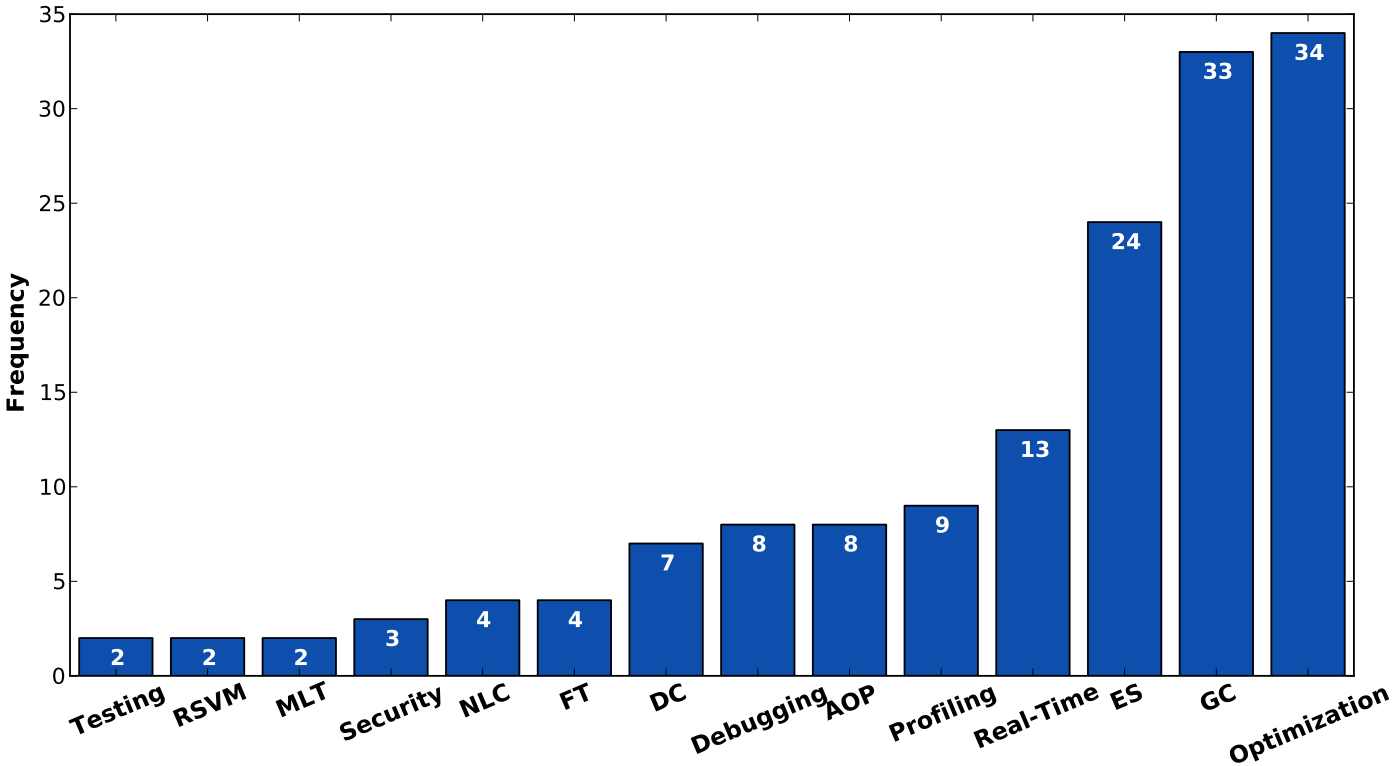


Figure: Frequency of studies in each category*.

*Certain studies were grouped in **more than one category**

Most Researched Subjects Evolution

According to our results, these are the *"trendy"* subjects:

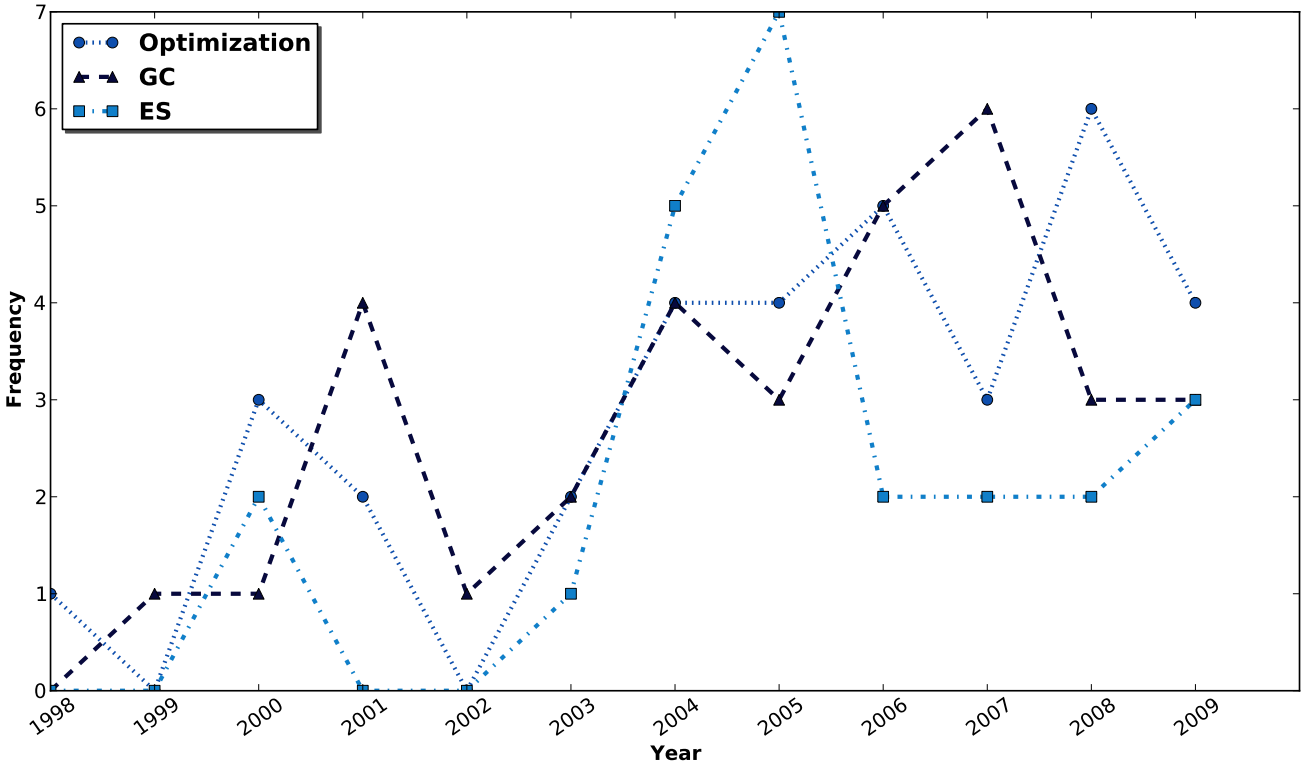
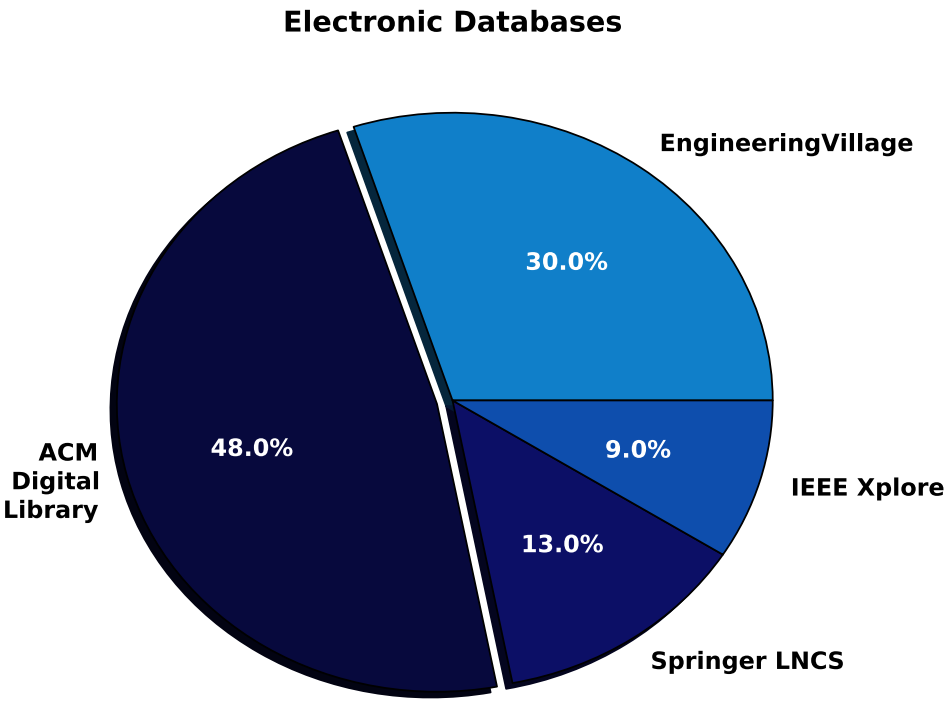


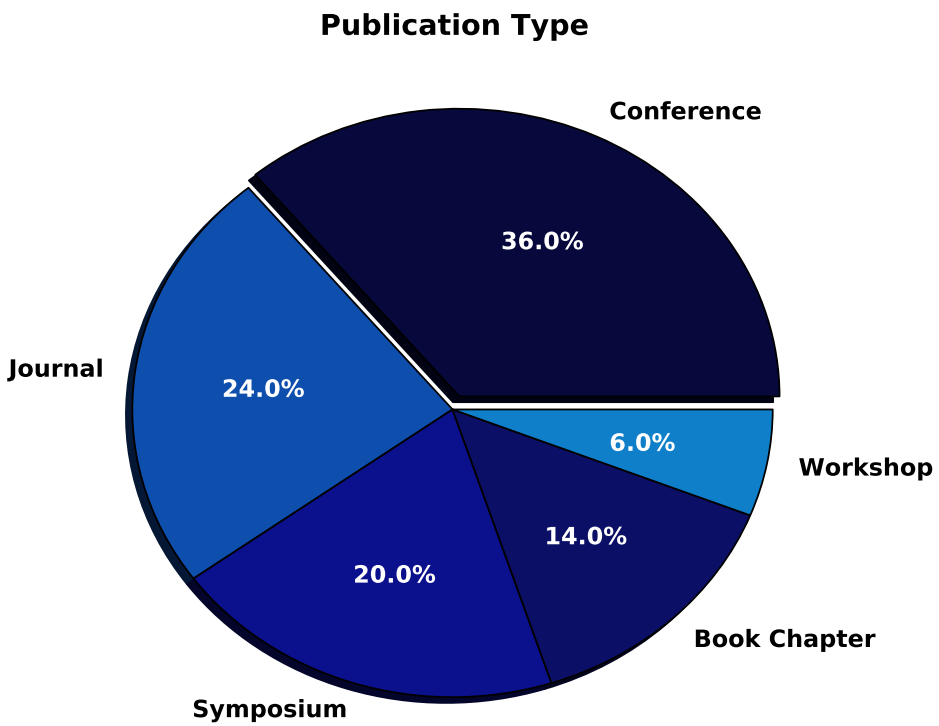
Figure: Year-wise distribution of publications on the most investigated categories.

Distribution of Primary Studies by Electronic Database



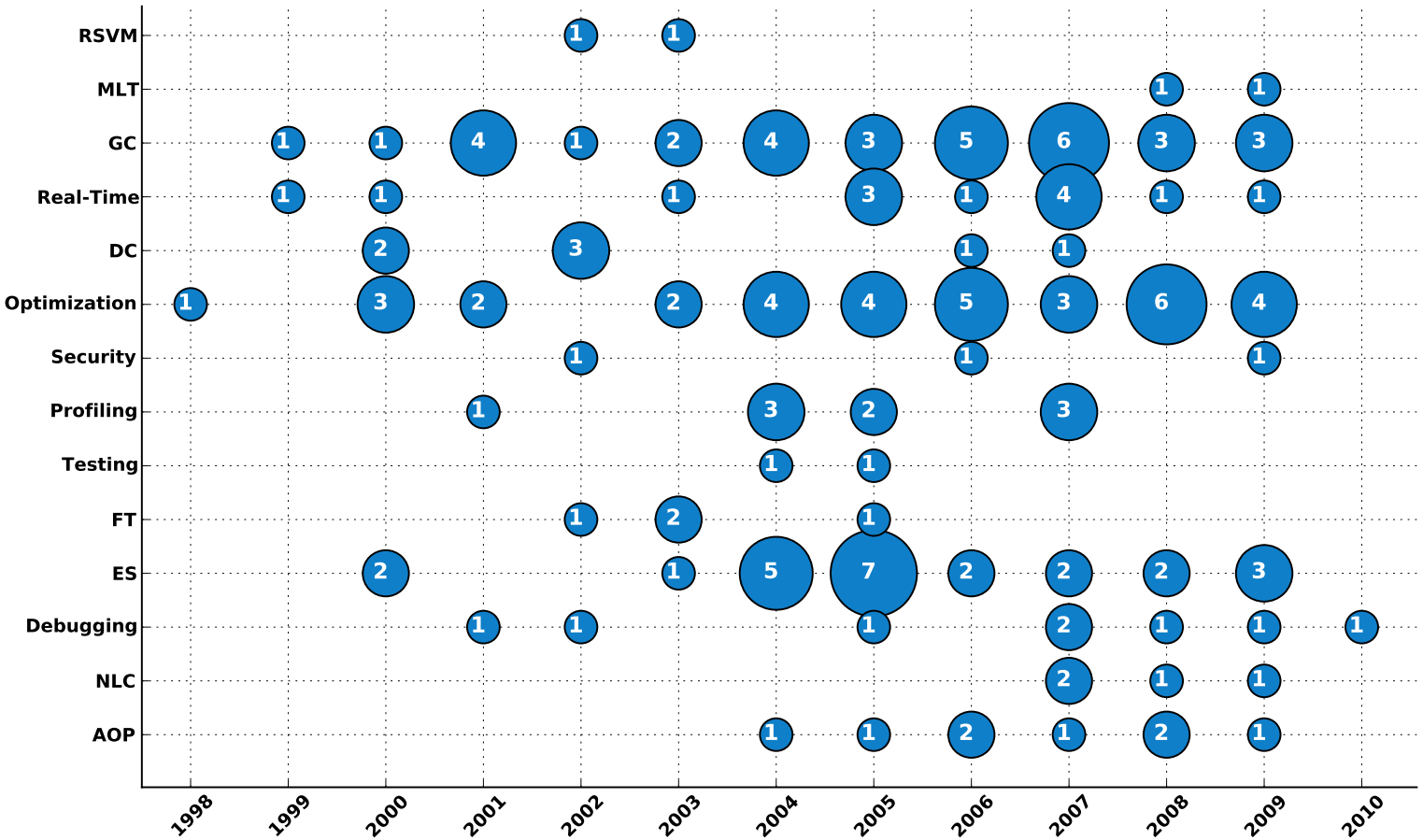
Electronic Database	Number
ACM Digital Library	62
EngineeringVillage	38
Springer LNCS	16
IEEE Xplore	12
ScienceDirect	0

Distribution of Primary Studies by Publication Type

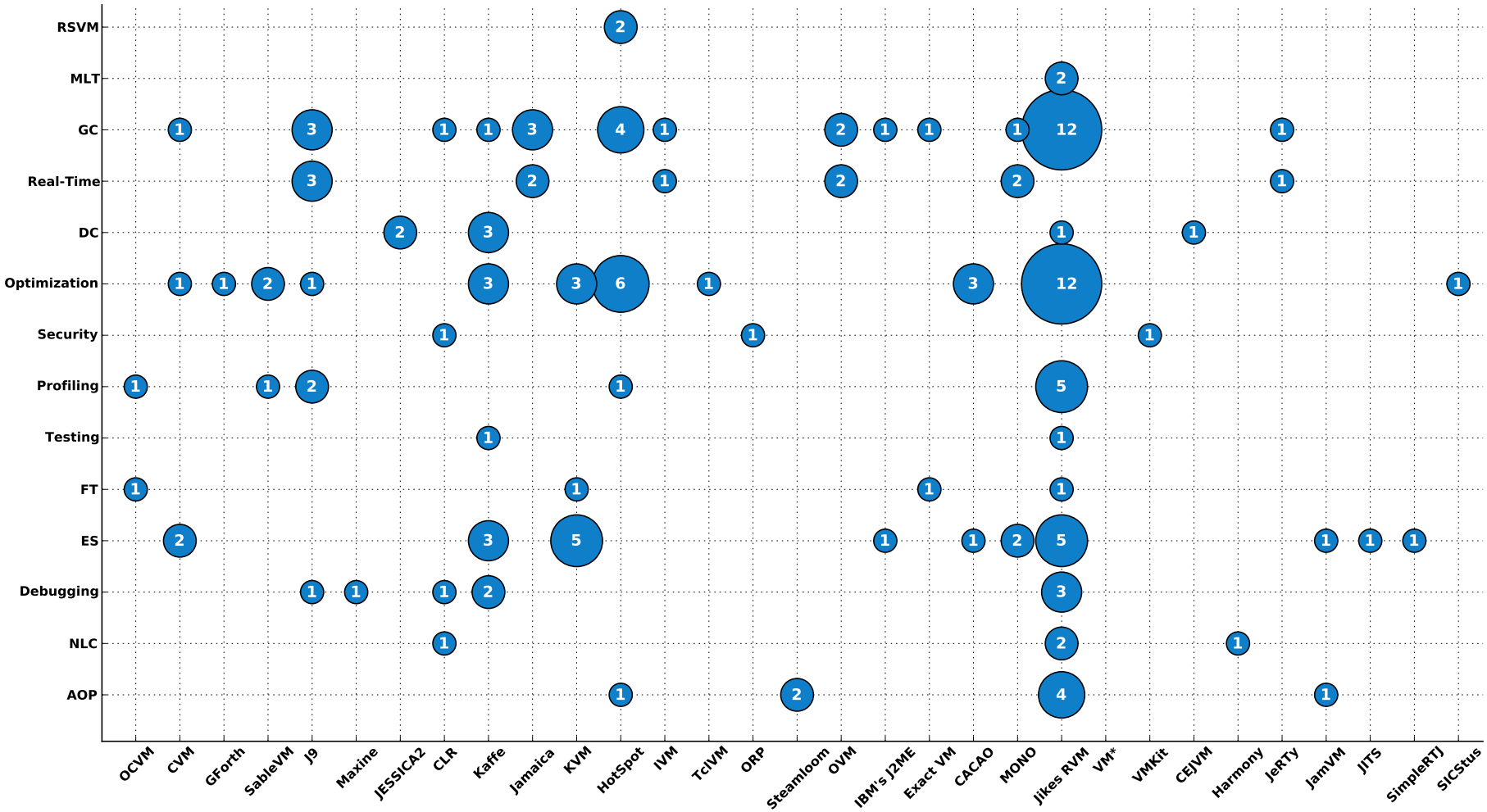


Publication Type	Number
Conference	46
Journal	31
Symposium	25
Book Chapter	18
Workshop	8

Map: Year-Wise Distribution (detailed)



Map: The Most-Widely Used HLL VM Implementations



Threats to Validity

- We cannot rule out threats from a **quality assessment perspective**.
  - (We wanted to be as inclusive as possible) We simply selected **studies without assigning any scores**.
- Another threat consists in whether we have properly identified and selected **all relevant publications**.
- Whether our **resulting classification scheme and categories** are coherent also represents a threat to validity.

Concluding Remarks

The mapping study results, although not entirely surprising (some may argue), can be used to support several claims that are frequently made but not **scientifically backed up**.

- Our mapping study reveals that the majority of research into HLL VMs focuses on optimizing these execution environments, improving their memory management capabilities, and tailoring them to resource-constrained settings.
- As for the publication types, the majority of the studies are **conference publications**.
- Another contribution of this paper is the **map** we have created.

References


K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson.
Systematic Mapping Studies in Software Engineering.
12th International Conference on Evaluation and Assessment in Software Engineering (EASE), pages 71–80, 2008


J. E. Smith and R. Nair
The Architecture of Virtual Machines.
Computer 38(5):32–38, 2005.


J. E. Smith and R. Nair
Virtual Machines: Versatile Platforms for Systems and Processes.
Morgan Kaufmann, 656 pages, 2005.

A Systematic Mapping Study on High-level Language Virtual Machines

Vinicius H. S. Durelli, Katia R. Felizardo, and Marcio E. Delamaro

Computer Systems Department
University of São Paulo (ICMC-USP)
13560-970 – São Carlos – SP – Brazil
{durelli,katiarf,delamaro}@icmc.usp.br

October 17, 2010