Bauhaus
–
A Tool Suite for Program Analysis and Reverse Engineering

Aoun Raza Gunther Vogel Prof. Dr. Erhard Plödereder

Department of Programming Languages
Institute for Software Technology
University of Stuttgart

06/06/06

Content

Introduction

Maintenance and Evolution of Software:

- difficult
- time consuming
- expensive

Maintenance in software life-cycle: $60\% - 80\%$ of costs

Critical Systems:

- High requirements for quality and reliability
- Faults must be prevented under all circumstances
- Impacts of changes must be fully understood
- Understanding of details and overall structure

The Bauhaus Project

Goal:

- Support software engineers in the task of maintenance
- Provide methods, techniques and tools for program understanding on all levels of abstraction
  - source code $\longleftrightarrow$ architecture level
- Tools for practical use in software development
- Improve quality and efficiency of maintenance processes

History

- Foundation in 1996 as a research project of University of Stuttgart and Fraunhofer Institute Kaiserslautern
- Collaboration of University Stuttgart and Bremen
- Commercial distribution by Axivion GmbH
- Over 100 person-years of development
- Methods and tools were validated in industrial practice
- More than 40 scientific publications

Infrastructure

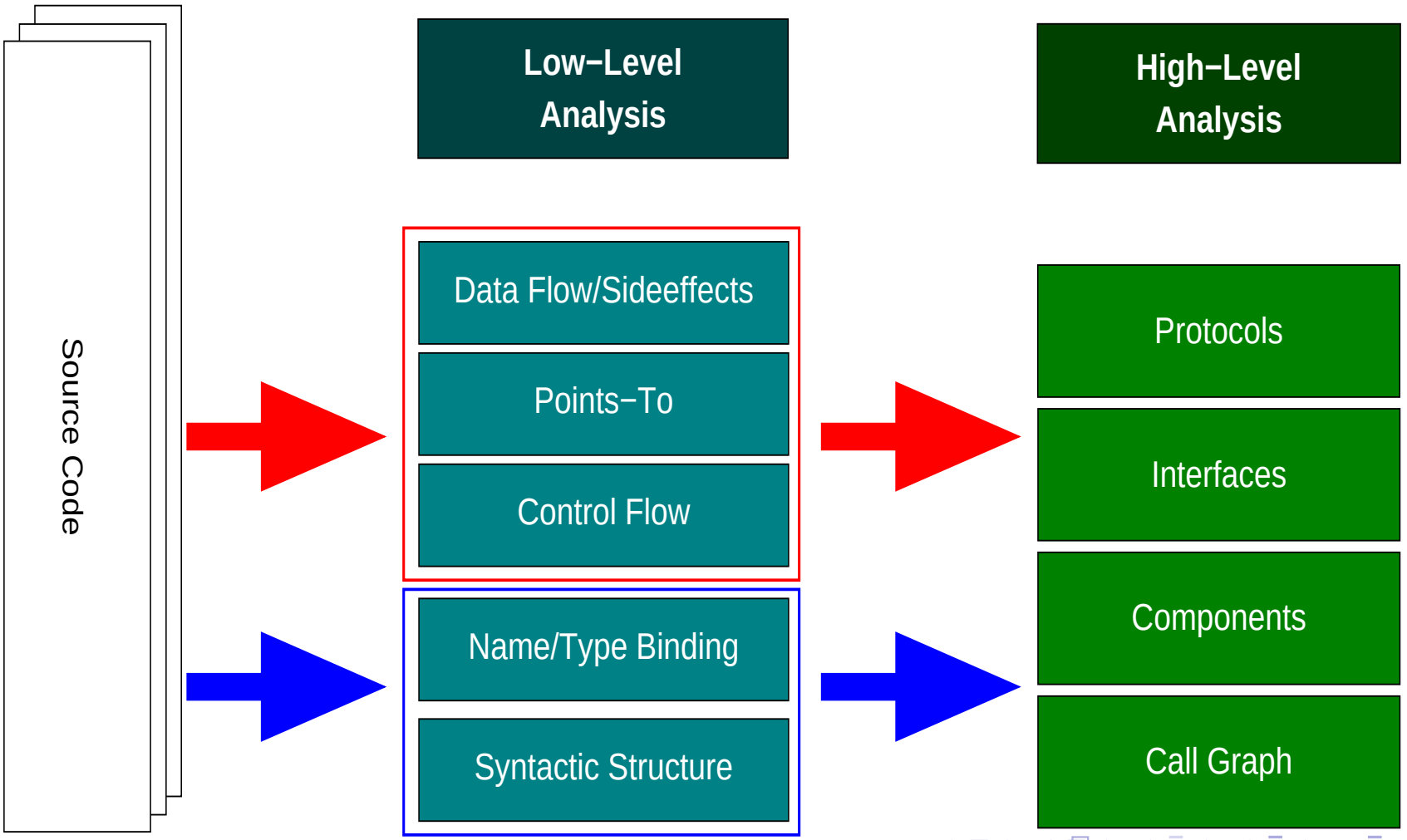
Key idea: Source code as most important source of information

→ Compiler technology

Applications:

- Source code navigation
- Anomaly detection
- Architecture recovery and validation
- Quality assessment

Infrastructure



Infrastructure – Technical Details

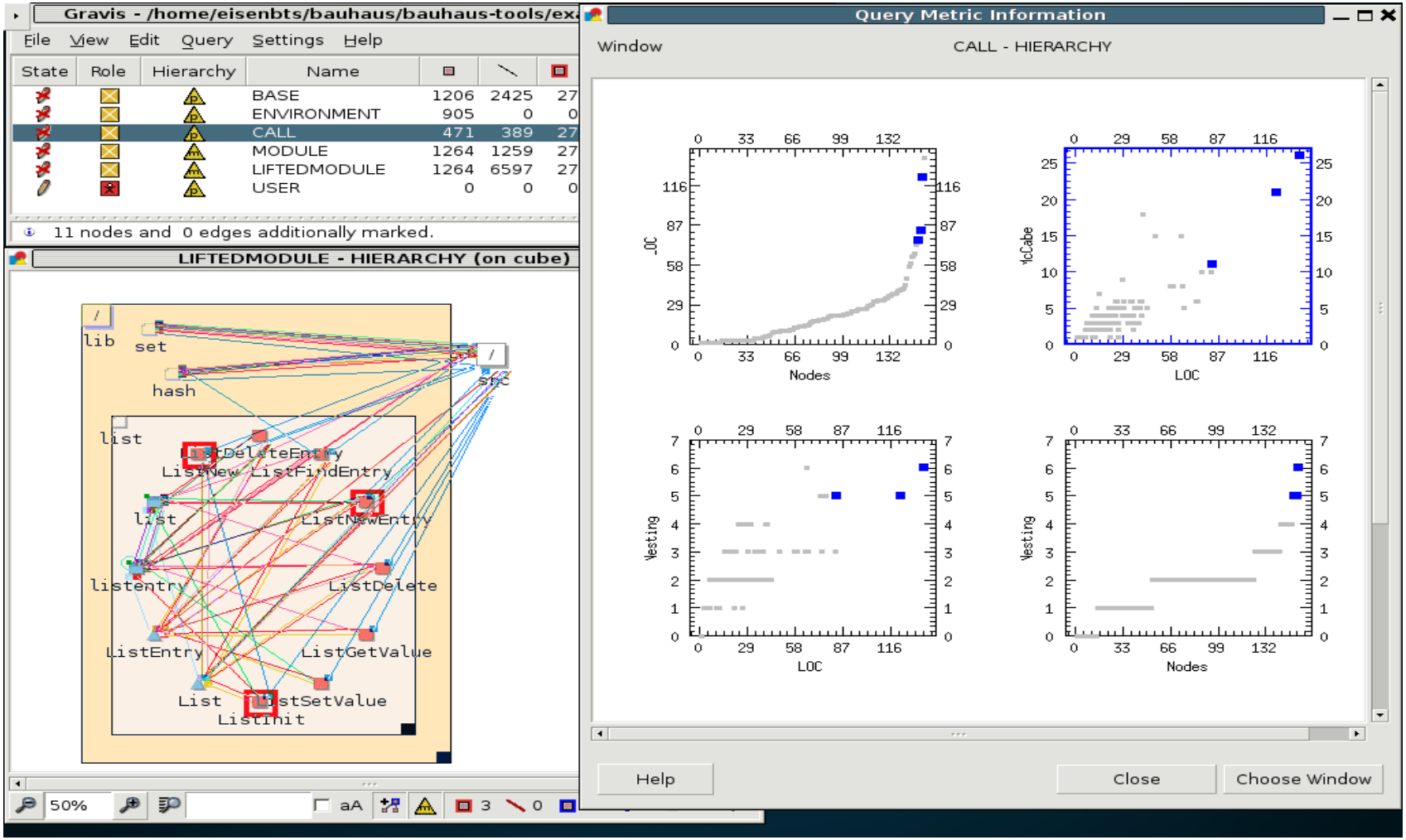
Two program representation for different levels of abstraction:

- IML (InterMediate Language) – run time semantics
- RFG (Resource Flow Graph) – system structure

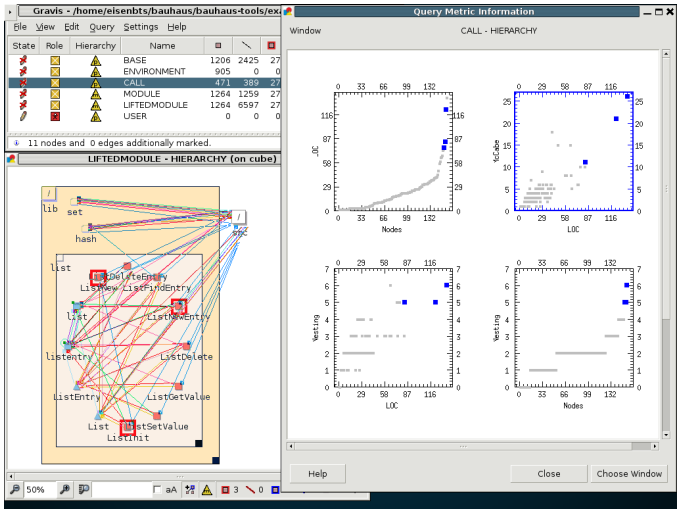
Both representations:

- Language independent (C, C++, Ada, Java)
- Graph-based:
  - Nodes represent constructs of the source program
  - Edges represent relations
- Representation of full programs
- Extensible by analyses

Software Quality – Metrics

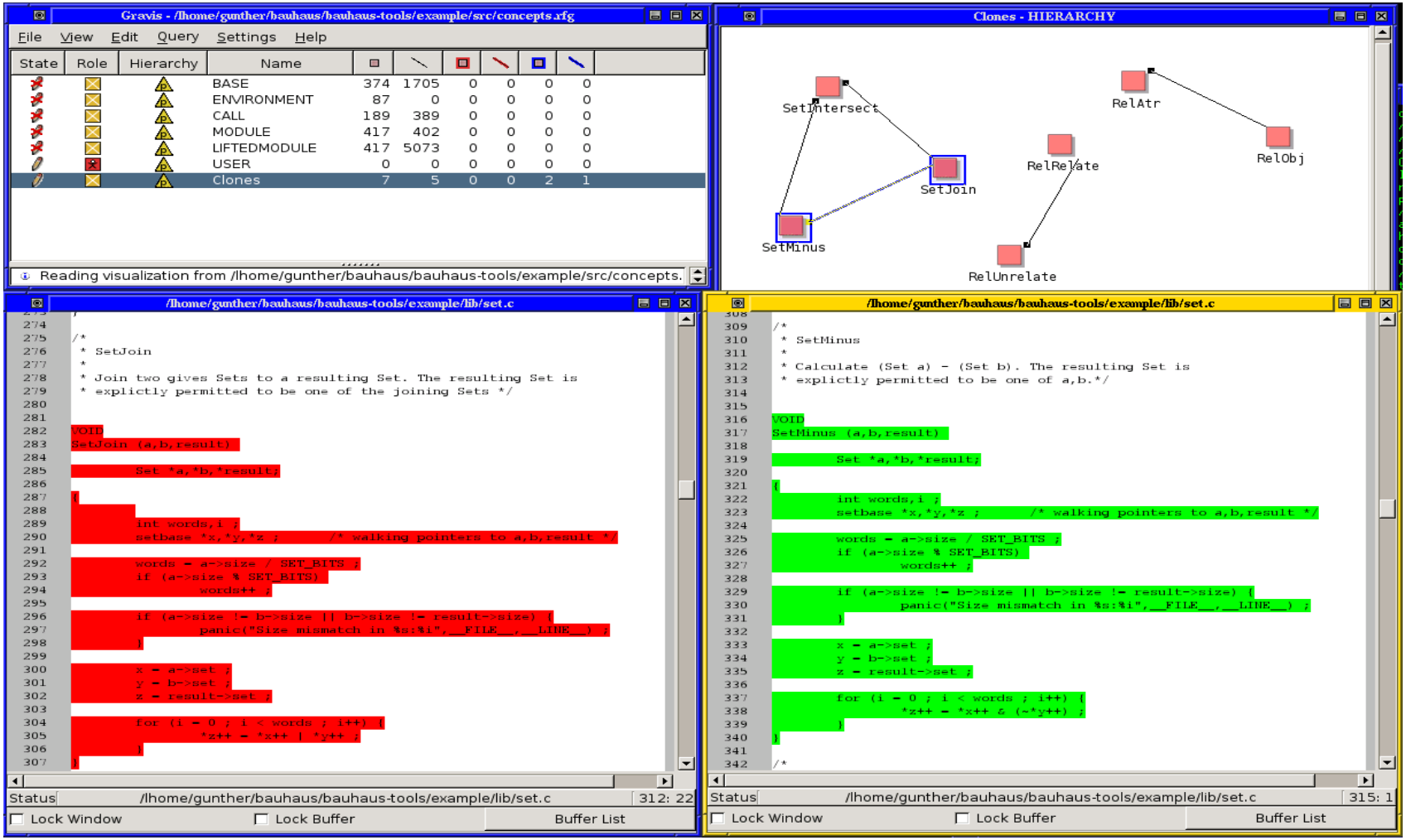


Software Quality – Metrics

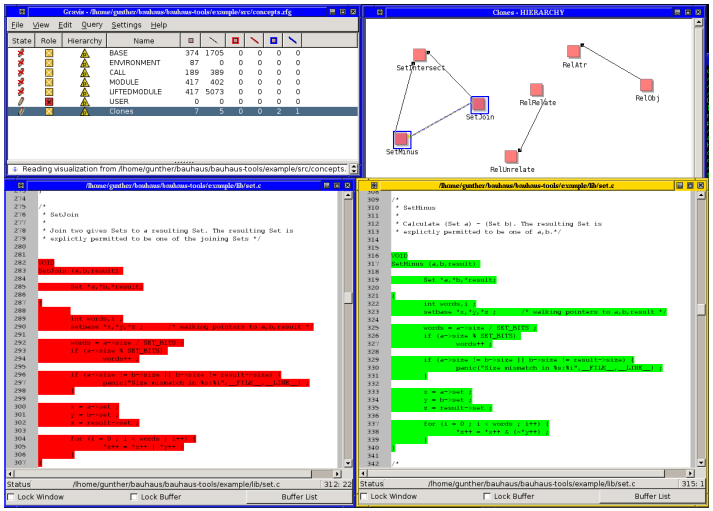


- Code level: lines of code, Halstead, maximum nesting, cyclomatic complexity
- Architecture level: number of methods, classes, and units, coupling, cohesion
- Derived metrics (Python Scripting): average number of methods per class, classes per unit, maintainability index (Coleman, Oman)

Clones


Gravis - /lhome/gunther/bauhaus/bauhaus-tools/example/src/concepts.xfg
File View Edit Query Settings Help
State Role Hierarchy Name
BASE 374 1705 0 0 0 0
ENVIRONMENT 87 0 0 0 0 0
CALL 189 389 0 0 0 0
MODULE 417 402 0 0 0 0
LIFTEDMODULE 417 5073 0 0 0 0
USER 0 0 0 0 0 0
Clones 7 5 0 0 2 1
Reading visualization from /lhome/gunther/bauhaus/bauhaus-tools/example/src/concepts.
Clones - HIERARCHY
SetIntersect
SetJoin
SetMinus
RelAtr
RelRelate
RelObj
RelUnrelate
/lhome/gunther/bauhaus/bauhaus-tools/example/lib/set.c
/*
* SetJoin
*
* Join two gives Sets to a resulting Set. The resulting Set is
* explictly permitted to be one of the joining Sets */
VOID
SetJoin (a,b,result)
Set *a,*b,*result;
{
int words,i ;
setbase *x,*y,*z ; /* walking pointers to a,b,result */
words = a->size / SET_BITS ;
if (a->size % SET_BITS)
words++ ;
if (a->size != b->size || b->size != result->size) {
panic("Size mismatch in %s:%i",__FILE__,__LINE__) ;
}
x = a->set ;
y = b->set ;
z = result->set ;
for (i = 0 ; i < words ; i++) {
*z++ = *x++ | *y++ ;
}
}
Status /lhome/gunther/bauhaus/bauhaus-tools/example/lib/set.c 312: 22
Lock Window Lock Buffer Buffer List
/lhome/gunther/bauhaus/bauhaus-tools/example/lib/set.c
/*
* SetMinus
*
* Calculate (Set a) - (Set b). The resulting Set is
* explictly permitted to be one of a,b.*/
VOID
SetMinus (a,b,result)
Set *a,*b,*result;
{
int words,i ;
setbase *x,*y,*z ; /* walking pointers to a,b,result */
words = a->size / SET_BITS ;
if (a->size % SET_BITS)
words++ ;
if (a->size != b->size || b->size != result->size) {
panic("Size mismatch in %s:%i",__FILE__,__LINE__) ;
}
x = a->set ;
y = b->set ;
z = result->set ;
for (i = 0 ; i < words ; i++) {
*z++ = *x++ & (~*y++) ;
}
}
/*
Status /lhome/gunther/bauhaus/bauhaus-tools/example/lib/set.c 315: 1
Lock Window Lock Buffer Buffer List

Clones

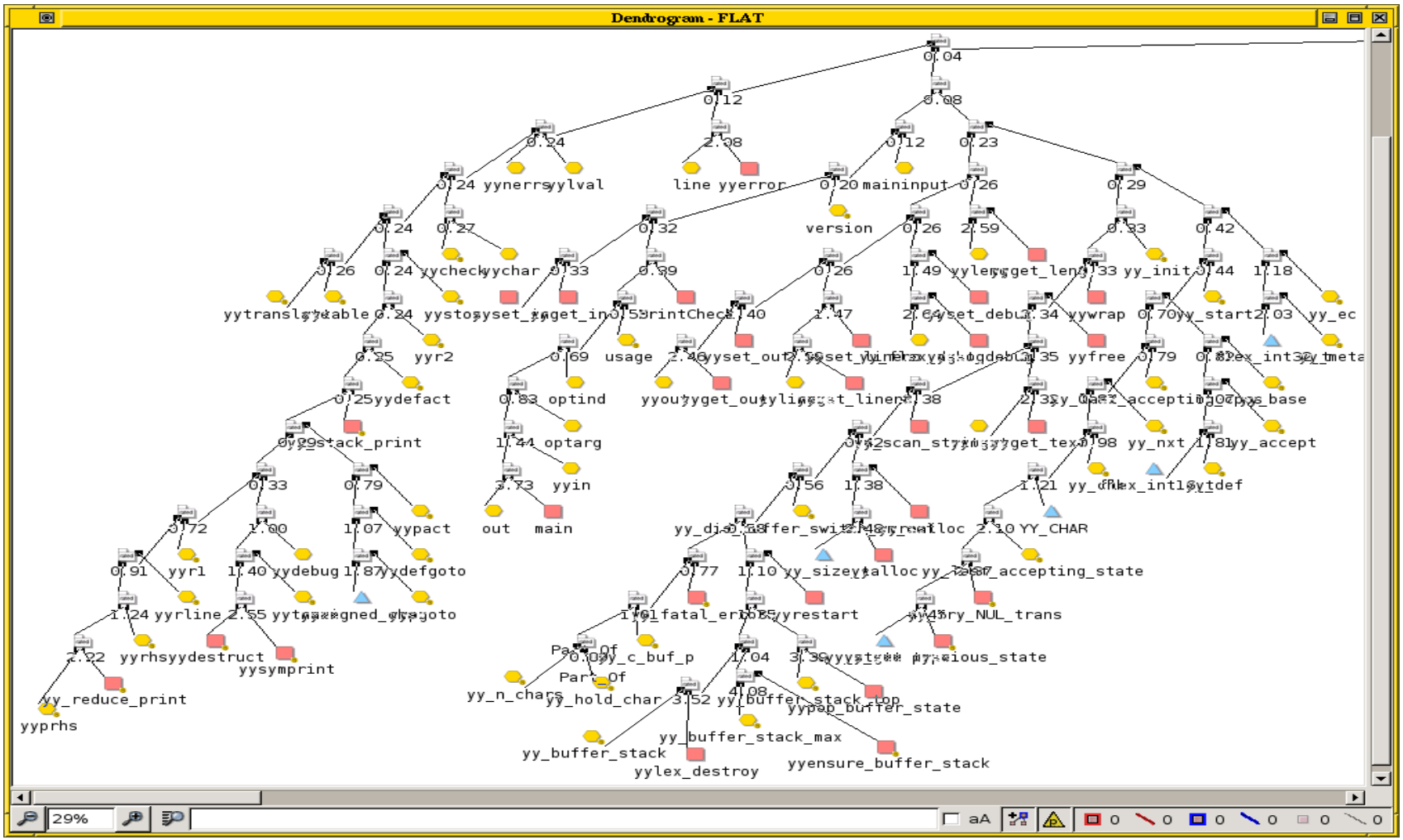


Token-based vs IML-based clone analysis

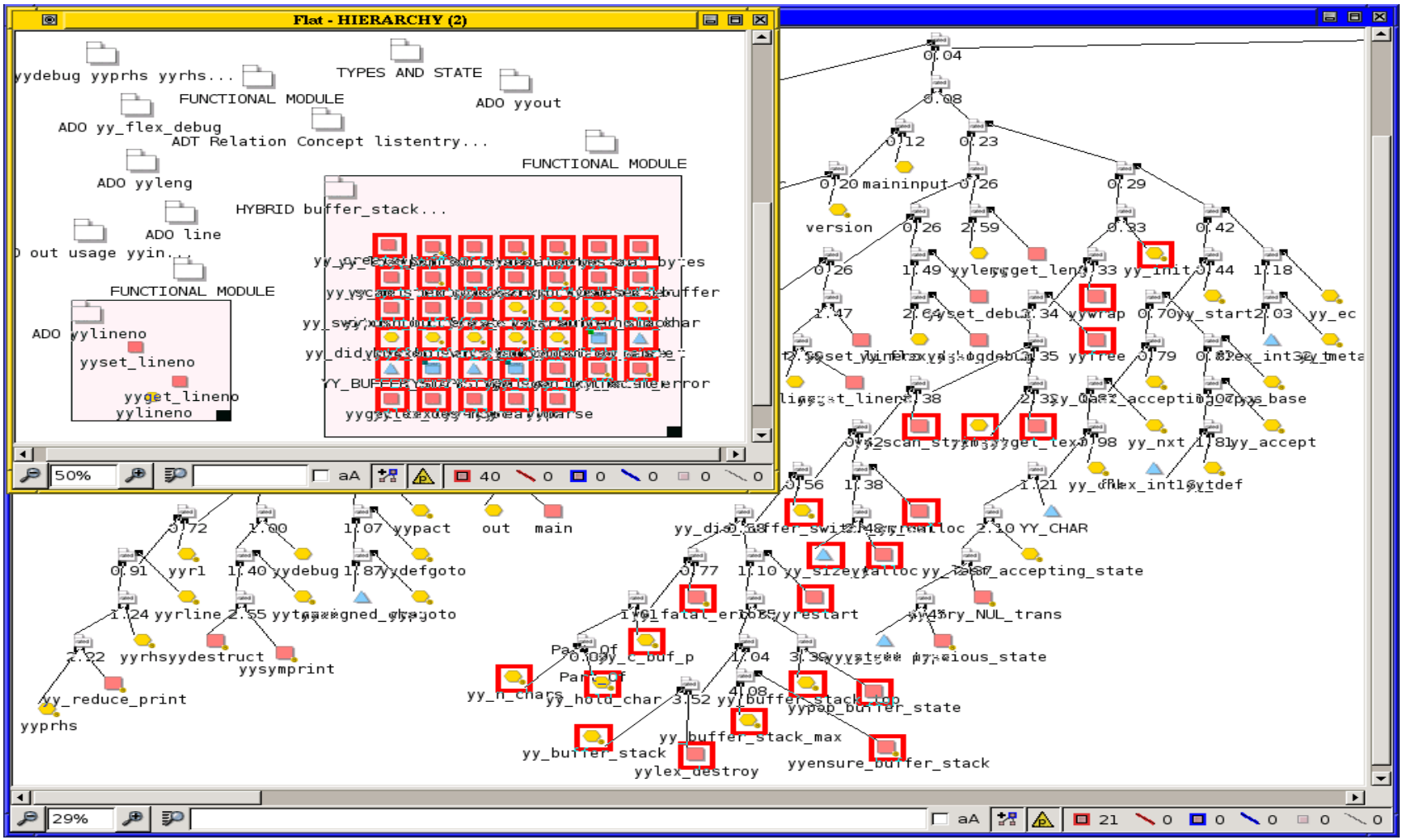
Types of duplication:

- Type 1: exact copy
- Type 2: copy with consistent substitution
- Type 3: additional insertions and deletions

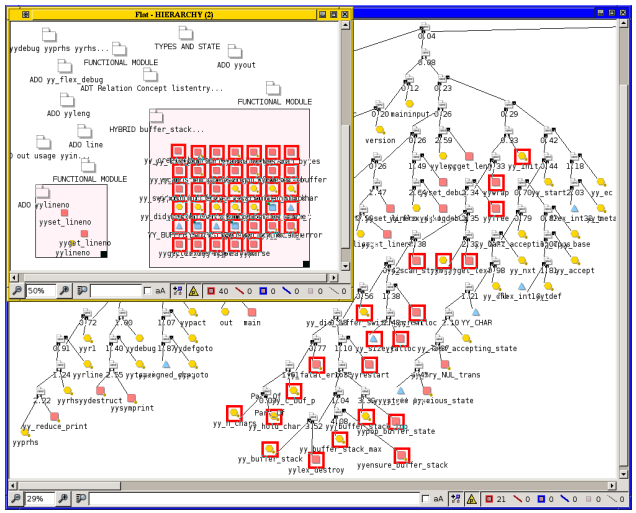
Architecture Recovery



Architecture Recovery



Architecture Recovery



- 14 automatic recovery techniques
- 7 categories of components: ADT, ADO, Function Library, ...
- 1 iterative semiautomatic recovery process

- Validation of hypothetical architectures – Reflection method

Protocol Analysis

Definition

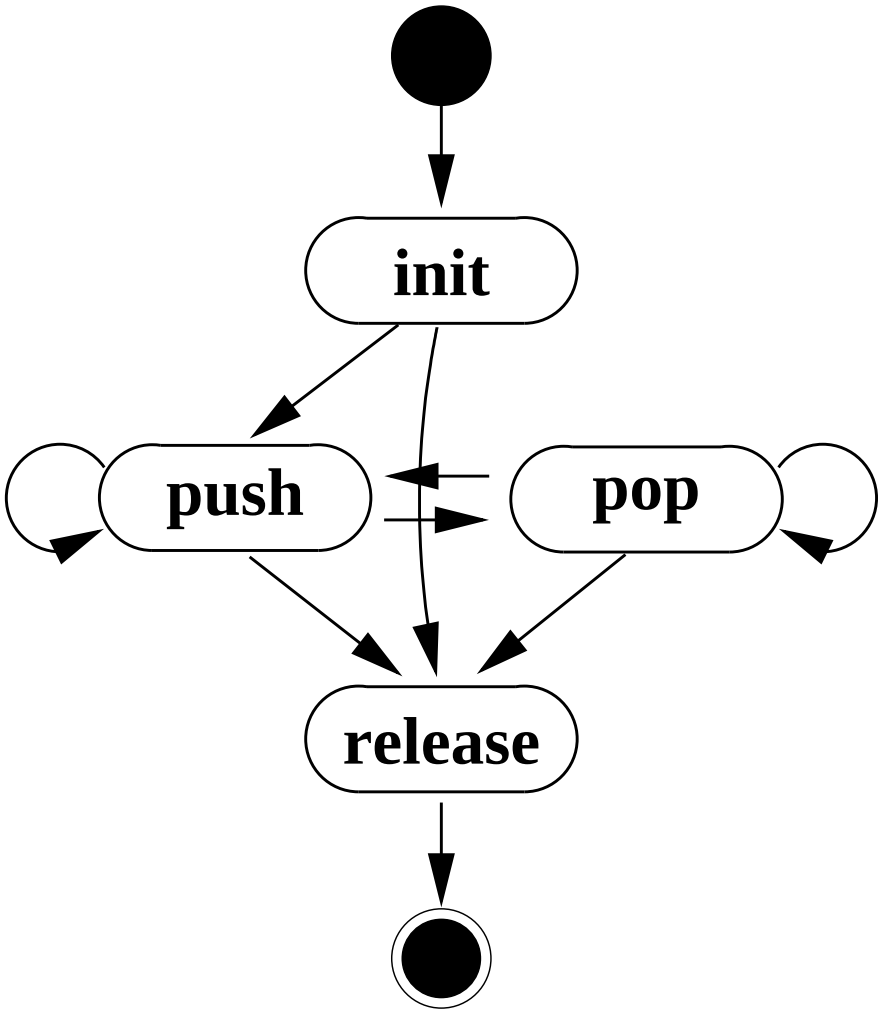
A protocol is a set of rules and conventions for program execution sequences

Applications:

- Program understanding
- Verification

Protocol representation:

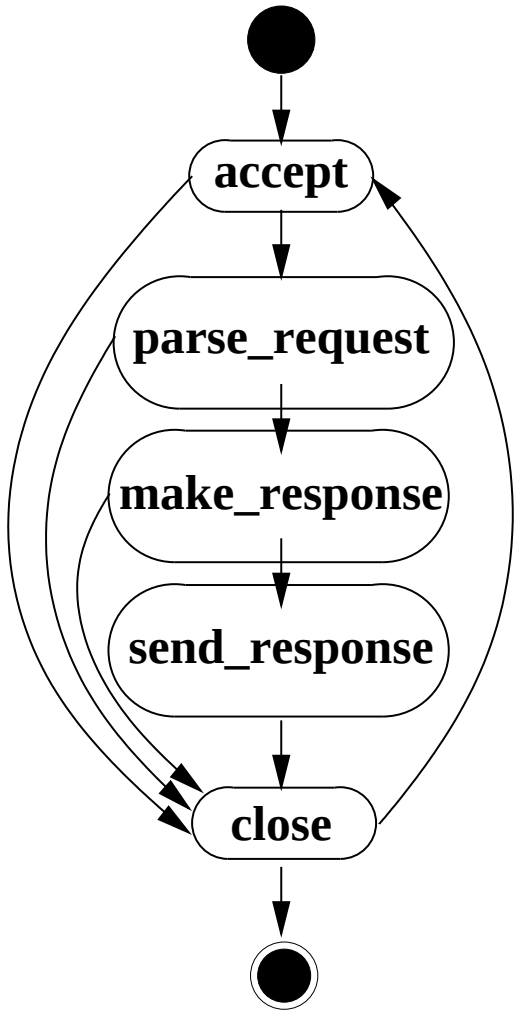
- Control flow graphs
- Finite automatons



Web Server – Network Communication

Implementation of HTTP:

- Sequence of actions is correct
- Connections can be closed at arbitrary times
- But: Only one request per connection



Current Research

- Improved base analyses
- Analyses for parallel programs
- Analyses of programs with GUIs
- Protocol analyses

Experiences

Language	Handwritten	Generated	Total
Ada	589'000	291'000	880'000
C	106'000	0'000	106'000
C++	115'000	177'000	292'000
...	...	...	...
Total	843'000	469'000	1'312'000

- GNAT Coding Style
- Platform-independence: Linux, Windows, Solaris
- Interfacing to other languages:
  - Compiler front-ends (compiler C++, SOOT – Java)
  - Python scripting (C, SWIG)

Summary

Conclusions:

- Bauhaus offers a broad range of tools for reverse-engineering
- Strong base analyses support high level program understanding
- Ada was successful as the main programming language
- High requirements for efficiency and reliability have been met

Wishes for the Future:

- More free tools for Ada software development
- Better library support

More information:

- `http://www.bauhaus-stuttgart.de`
- `http://www.bauhaus-bremen.de`
- `http://www.axivion.de`