Decompilation of Binary Programs

Christina Cifuentes & K. John Gough
School of Computing Science
Queensland University of Technology
Presented by Conny Chan

Overview

- Usage of decompiler
- Phases of the decompiler
  - Front-end, UDM, Back-end
- The decompiling system
  - Signature generator
- Conclusion
- Discussion

Reverse Compiler

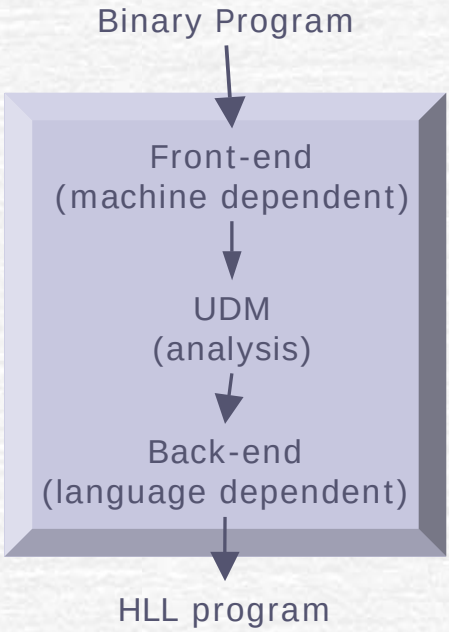
- Perform inverse process of a compiler

- Usage:
  - Maintenance of Code
    - Lost code recovery
    - Migration of application to new HW platform.
    - Translation of the obsolete code into new code.
  - Software Security
    - Malicious code detection

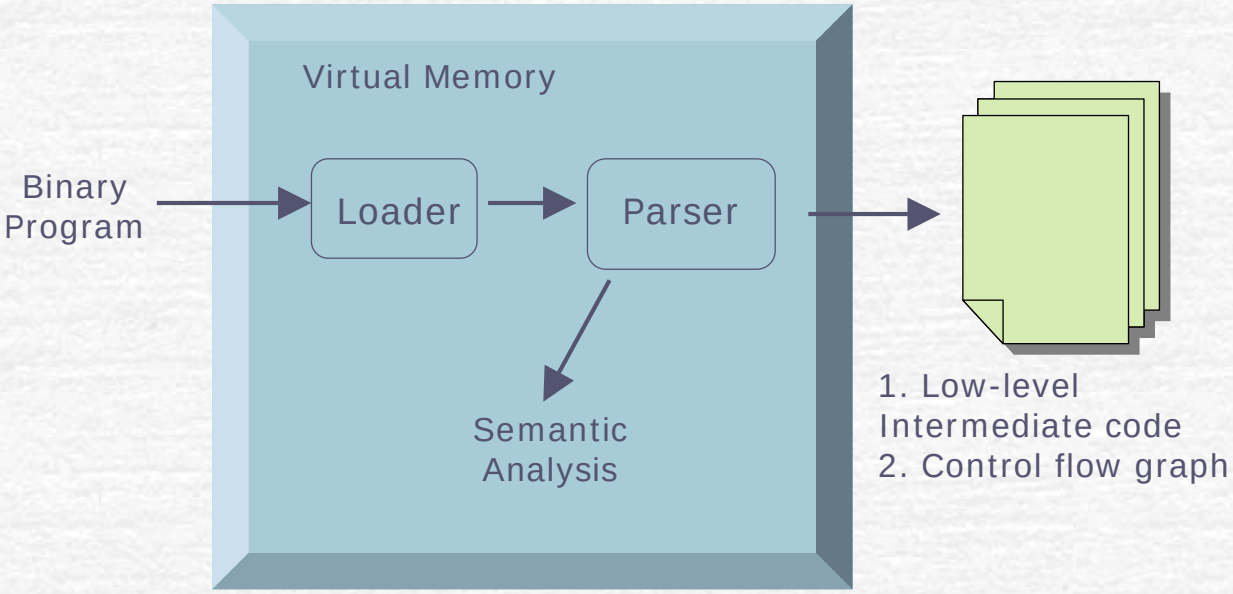
Note in advance:

- Decompiler in this article:
  - Experimental decompiler for the DOS OS
  - Intel i80286 architecture
  - Read .com and .exe files
  - Produce C programs as output.

Decompiler Structure



The Front-end



The Semantic Analysis

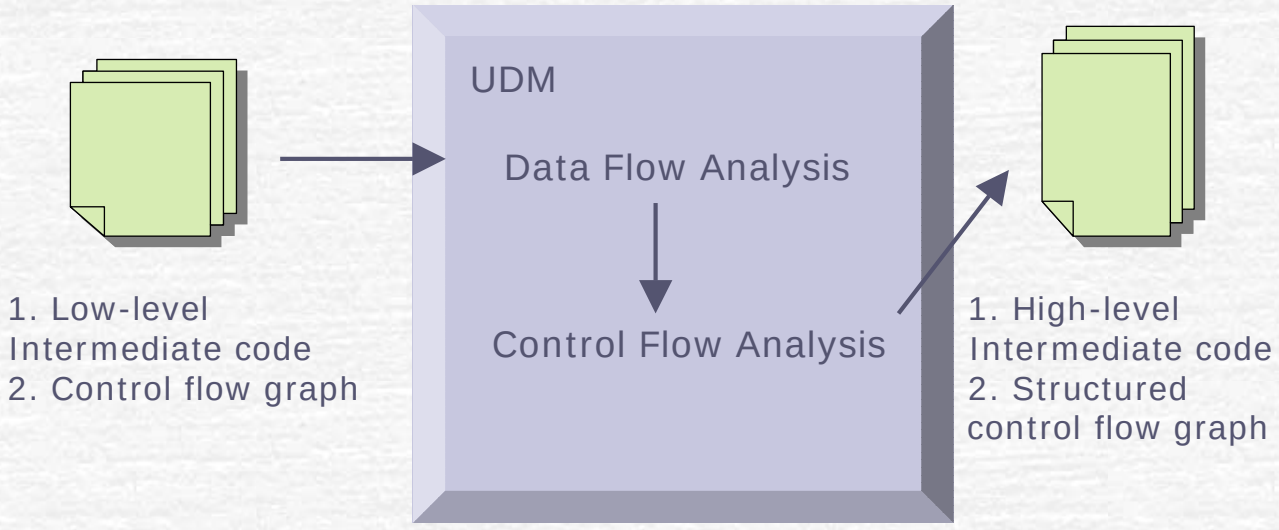
- Performs idiom analysis e.g.

```
neg dx
neg ax          ──────►   neg dx:ax
sbb dx, 0
```

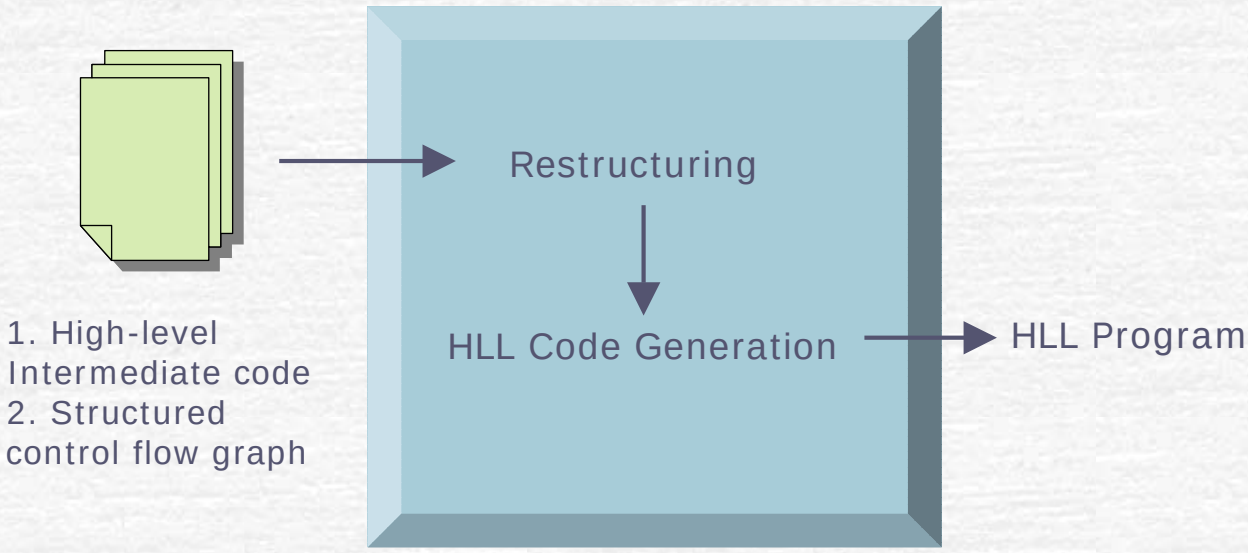
- type propagation.
  - E.g. long variable found at –1 to –4.

```
[bp-2]
  ...          merged      [bp-2]:[bp-4]
[bp-4]
```

The Universal Decompiling Machine (UDM)



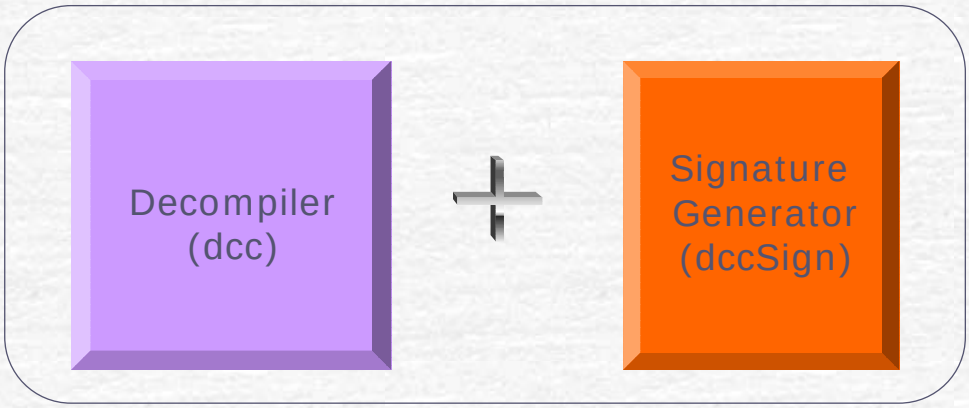
The Back-end



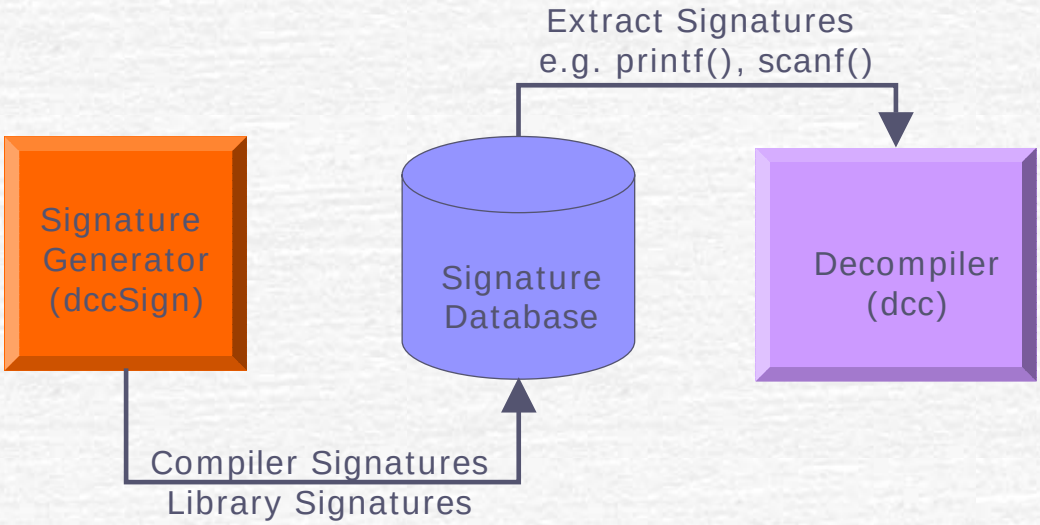
The HLL code generation

- Defines:
  - Global variables
  - Emits code for each function
- In each function:
  - Comments of such procedures
- Variables and procedures named in loc1, proc2, etc.

The Decompiling System



Signature Database

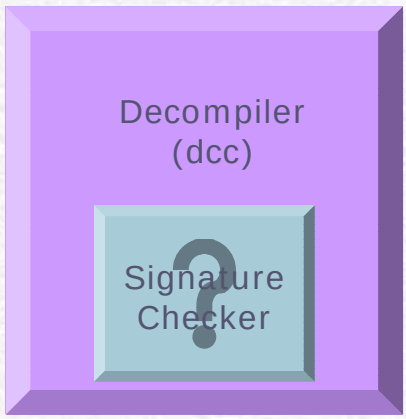


- If a library function matched, replaced by the library name instead of analyzed by dcc.

Signatures

- Library Signature
  - A series of instructions that identifies library function for a compiler.
- Compiler Signature
  - A series of instructions that identifies a particular version of a compiler.

Signature Checker



- Determined if known compiler is used.
- Check first n bytes of instructions with pattern-matching

Conclusion

- Present one way for decompiling binary program.
- Prove the feasibility of writing a decompiler for a contemporary machine architecture.

Discussion

- Is decompilation legal?