Software Engineering

Lecture 1
Introduction, principles & architecture

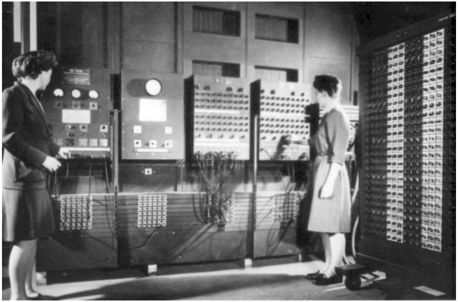
David Baelde
baelde@lsv.ens-cachan.fr

MPRI

September 16, 2016

Introduction

Prehistory



Main control panel of ENIAC (1946)

First Turing-complete computers

- Huge and expensive (30 tons, $167m^2$, 150kW, 6M$)
- One-off, built for specific purposes (military computations)
- Focus on making hardware reliable

Industrialization

IBM System/360 (1964)

Mainframe computers

- Wide range of applications, scientific to commercial
- Separation of architecture and implementation
- Software complexity rises

Birth of software engineering

1960' software crisis

- Spectacular failures: bugs, cost, overtime, cancellation

Birth of software engineering

1960' software crisis

- Spectacular failures: bugs, cost, overtime, cancellation
- Frederick P. Brooks about OS/360:

  > *The flaws in design and execution pervade especialy the control program. [...] The product was late, it took more memory than planned, the costs were several times the estimate, and it did not perform very well until several releases after the first.*

Birth of software engineering

1960' software crisis

- Spectacular failures: bugs, cost, overtime, cancellation
- Frederick P. Brooks about OS/360:

  > *The flaws in design and execution pervade especialy the control program. [...] The product was late, it took more memory than planned, the costs were several times the estimate, and it did not perform very well until several releases after the first.*

1968 NATO conference on Software Engineering

> *Need for software manufacturers to be based on the types of theoretical foundations and practical disciplines that are traditional in the established branches of engineering.*

Software Engineering

A lot of parameters:

- Product: is it critical? clearly defined? meant to evolve?
- Economical aspects: cost of machines and workers, competition and time pressure.

Software Engineering

A lot of parameters:

- Product: is it critical? clearly defined? meant to evolve?
- Economical aspects: cost of machines and workers, competition and time pressure.
- Technology: languages, RCS, communication means.
- Science: PL theory, verification, static analysis, etc.

Software Engineering

A lot of parameters:

- Product: is it critical? clearly defined? meant to evolve?
- Economical aspects: cost of machines and workers, competition and time pressure.
- Technology: languages, RCS, communication means.
- Science: PL theory, verification, static analysis, etc.
- Humans: client, users, developpers, managers, etc.
- Ideology: more or less hierarchy, secret, etc.

Software Engineering

A lot of parameters:

- Product: is it critical? clearly defined? meant to evolve?
- Economical aspects: cost of machines and workers, competition and time pressure.
- Technology: languages, RCS, communication means.
- Science: PL theory, verification, static analysis, etc.
- Humans: client, users, developpers, managers, etc.
- Ideology: more or less hierarchy, secret, etc.

Relies on common sense, computer science but also social sciences: psycho, ethno, experiments, etc.
No best approach.

Software Engineering

Problems	Goal
Complexity	Correctness
Change	Evolutivity

Software Engineering

Problems	Goal
Complexity	Correctness
Change	Evolutivity

Approach

- Principles behing good software products and processes.
- Methodologies that apply and promote those principles.
- Tools to implement and help follow methodologies.

Software Engineering

Problems	Goal
Complexity	Correctness
Change	Evolutivity

Approach

- Principles behing good software products and processes.
- Methodologies that apply and promote those principles.
- Tools to implement and help follow methodologies.

Scope

Activities	implem.
Products	code

Software Engineering

Problems	Goal
Complexity	Correctness
Change	Evolutivity

Approach

- Principles behing good software products and processes.
- Methodologies that apply and promote those principles.
- Tools to implement and help follow methodologies.

Scope

Activities	spec.	design	implem.	validation	evolution
Products	doc	doc	code	tests	history

Rigor

Rigor

How to ensure correctness?

- Ideally, formal methods!
- In practice, mostly through rigorous methodologies.

Rigor

How to ensure correctness?

- Ideally, formal methods!
- In practice, mostly through rigorous methodologies.

Correctness is meaningless without a spec!

- Always specify precisely what you need, and no more
- Informal specs (*i.e.*, doc) are *much* better than nothing
- Make sure the spec is visible to the implementer

Rigor

How to ensure correctness?

- Ideally, formal methods!
- In practice, mostly through rigorous methodologies.

Correctness is meaningless without a spec!

- Always specify precisely what you need, and no more
- Informal specs (*i.e.*, doc) are *much* better than nothing
- Make sure the spec is visible to the implementer

There *will* be bugs!

- Be paranoid, seek to detect anomalies early on
- Design precise tests, run them after each change

Change

Anticipation of change

Code *will* evolve

- Bugs will have to be fixed
- Requirements and the environment may change
- Components could be re-used in a (slightly different) context

Anticipation of change

Code *will* evolve

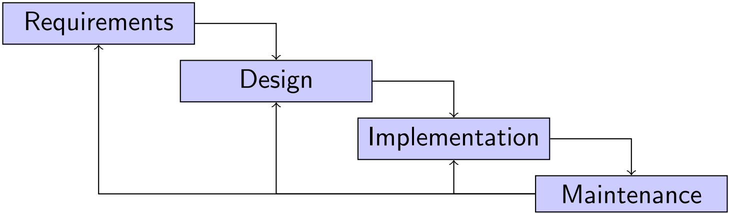
- Bugs will have to be fixed
- Requirements and the environment may change
- Components could be re-used in a (slightly different) context

Be ready

- Actively work to identify potential changes
- Design code so that change and re-use is facilitated
- Use tools that help to keep track of change
- Organize work around upcoming changes

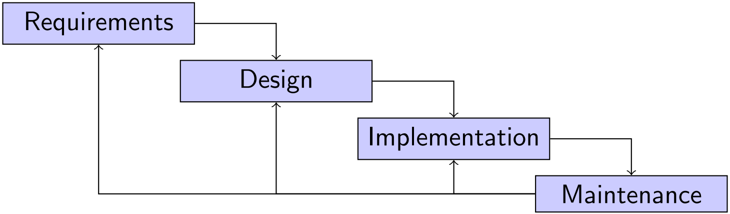
Software development processes

Waterfall model



Software development processes

Waterfall model



- Prevalent at least until 70'
- Probably inspired from other engineering fields
- DoD guidelines for military software: mandatory until 88
  remains reference after that (until recently?)

Incrementality

Proceed step by step to get early feedback and adjust.

- Start by implementing a subset of features.
- Start with functional correctness only.

Incremental development model

Incrementality

Proceed step by step to get early feedback and adjust.

- Start by implementing a subset of features.
- Start with functional correctness only.

Incremental development model

Pros/cons

- ⊕ Early feedback. Opportunity to fix requirements and design. May be necessary if requirements are not initially clear.
- ⊕ Good for the morale of developers and clients!
- ⊖ Requires refactoring to maintain good structure.
- ⊖ Hard to keep track of change in large projects.

The Linux kernel

The main invention in Linux is . . .

The Linux kernel

The main invention in Linux is its development model.

- Wide distribution and invitation to contribute,
  thanks to personal computers and the internet.
- Active integration of patches and frequent releases,
  initially by hand, then with dedicated tools.
- Pre-requisites in the code:
  - precise documentation
  - extensibility through modules for drivers, file system, etc.

E. S. Raymond, *The Cathedral and the Bazaar*, O'Reilly, 1999.

More development models

Collaborative software development

Incremental with collaboration and involvement of the public

Main model for open-source software:

- More testers $\rightarrow$ earlier bug reports
- Massive peer review (?)

More development models

Collaborative software development

Incremental with collaboration and involvement of the public

Main model for open-source software:

- More testers $\rightarrow$ earlier bug reports
- Massive peer review (?)

Agile software development

Incremental process + focus on collaboration & self-organization

`http://agilemanifesto.org/principles.html`

Various methodologies (XP, SCRUM...)

Modularity

Modularity

Segment project in modules with clearly defined interfaces.

Modularity

Segment project in modules with clearly defined interfaces.

A slogan: High Cohesion, Low Coupling

- Maximize modularity:
  parallelizability of the software process, chances of re-use
- Minimize interactions:
  separately test, modify... understand, then integrate

Example (types of cohesion)

- Coincidental: no (good) reason
- Temporal: executed around the same time, *e.g.*, init
- Functional: realize a task, *e.g.*, convert file
- ...

Modularity

Segment project in modules with clearly defined interfaces.

A slogan: High Cohesion, Low Coupling

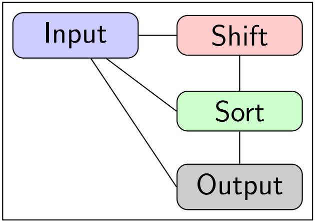
- Maximize modularity:
  parallelizability of the software process, chances of re-use
- Minimize interactions:
  separately test, modify... understand, then integrate

Example (types of cohesion)

- Coincidental: no (good) reason
- Temporal: executed around the same time, *e.g.*, init
- Functional: realize a task, *e.g.*, convert file
- ...

What is a good modularization?

Modularity



Modules export arrays

No text in shifted/sorted arrays

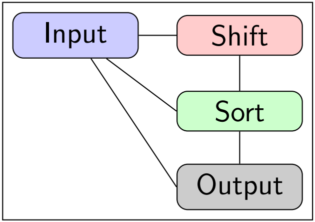
A possible modularization for a *KWIC index generator*:

Input: lines of text

Output:
all permutations of those lines, sorted alphabetically

David Parnas, *On the Criteria To Be Used in Decomposing Systems into Modules*, Communications of the ACM, 1972.

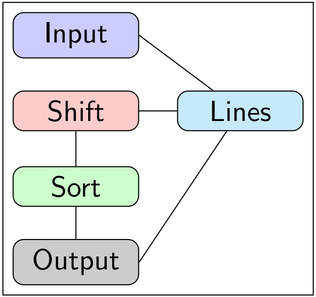
Modularity



Modules export arrays

No text in shifted/sorted arrays

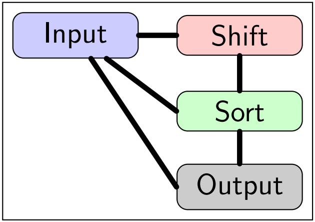
vs.



Modules export `get/set()`

David Parnas, *On the Criteria To Be Used in Decomposing Systems into Modules*, Communications of the ACM, 1972.

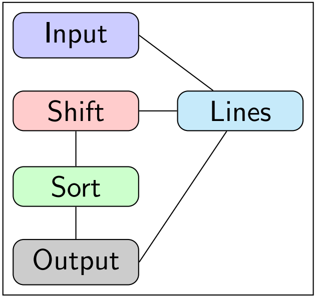
Modularity



Modules export arrays

No text in shifted/sorted arrays

vs.



Modules export `get/set()`

David Parnas, *On the Criteria To Be Used in Decomposing Systems into Modules*, Communications of the ACM, 1972.

Modularity

Segment project in modules with clearly defined interfaces.

A slogan: High Cohesion, Low Coupling

- Maximize modularity:
  parallelizability of the software process, chances of re-use
- Minimize interactions:
  separately test, modify... understand, then integrate

Example (types of cohesion)

- Coincidental: no (good) reason
- Temporal: executed around the same time, *e.g.*, init
- Functional: realize a task, *e.g.*, convert file
- Informational: independent operations on same data, *e.g.*, list

Modularization goes hand in hand with information hiding, aka ...

Abstraction

Ignoring details

Design

- Do not specify implementation details.
- Details are things that can easily change:
  maximum waiting time, password length, etc.

Code

- Code in a high-level language, far from the machine.
- Code for correctness first, then optimize if needed.
  "*Premature* optimization is the root of all evil." – Knuth
- Don't hardcode:
  no magic numbers, any constant should be justified.

Modularity + Abstraction

Segment project in modules with clearly defined interfaces.

Maximize information hiding in interfaces:

- Minimize coupling.
- Plan for evolution.

Language support

More or less constraining/helpful

- Modules and abstract types in ML-like languages
- Classes in object oriented programming
- Separate compilation units in other languages
- Procedures in *structured programming languages*!

Proof assistants

Concerns of computer-aided theorem proving

- Soundness: the whole point is to have trustworthy proofs!
- Usability: undo, notations, automation, efficiency, user extensions, etc.

Proof assistants

Concerns of computer-aided theorem proving

- Soundness: the whole point is to have trustworthy proofs!
- Usability: undo, notations, automation, efficiency, user extensions, etc.

Edinburgh LCF (70's)

- Proof objects cannot be maintained for performance reasons
- Small trusted kernel provides sound manipulations of abstract datatype `theorem`
- Tactics and tacticals built on top of this sound kernel
- By-product: ML language and module system!

M. J. C. Gordon, *From LCF to HOL: a short history*, 2000.

Proof assistants

Concerns of computer-aided theorem proving

- Soundness: the whole point is to have trustworthy proofs!
- Usability: undo, notations, automation, efficiency, user extensions, etc.

Coq v7 (2000)

- Proof objects are maintained: relevant, non-local checks
- Isolated kernel: breaking dependency on undo-able objects
- (OCa)ML modules still used: abstraction ensures safety
- Kernel is purely functional, 1/3 of the code
- 2013, `v8.4pl2`: same design, impure kernel, 1/10 of the code

J-C. Filliâtre, *Design of a proof assistant: Coq version 7*, 2000.

Exercises

Intermediaries

Discuss the following Java function:

```
public void
showDeadline(User u, Conference c) {
  TimeZone tz = u.getLocation().getTimeZone();
  Date d = c.getPaperDeadline();
  ... // something involving only tz and d
}
```

Pairs

(Based on 2013 MPRI project "Geriatric Terrorist Anarchy")

Two C++ classes use pairs:

- The UI performs drawing using SFML's `Vector2f` class.
- The simulator moves characters around the world, also using 2D floating point coordinates.

Alternatives to discuss:

- Use `Vector2f` for the simulator code.
- Create a new class for pairs of floats.
- (Use `std::pair`.)

strtok

```
char *strtok(char *str, const char *delim);

The strtok() function parses a string into a sequence
of tokens. On the first call to strtok() the string
to be parsed should be specified in str.  In each
subsequent call that should parse the same string,
str should be NULL.
```

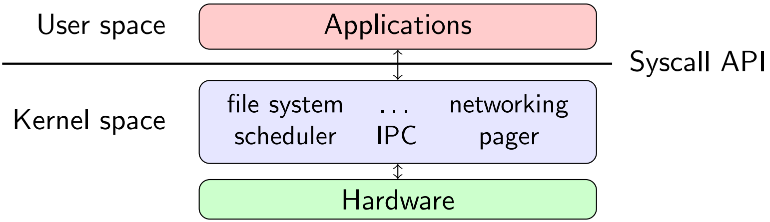
Discuss

- What's wrong with this spec?
- Give examples of when the function is unusable.
- Propose other designs, not necessarily in C.

Software architecture examples

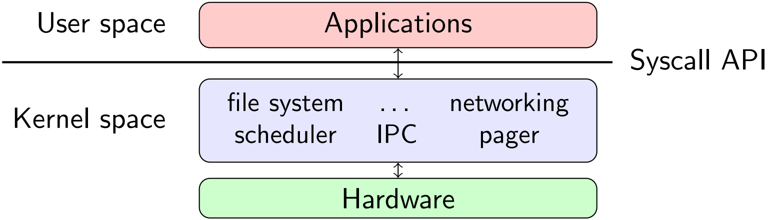
Layers

Monolithic kernel architecture



Layers

Monolithic kernel architecture



Unix

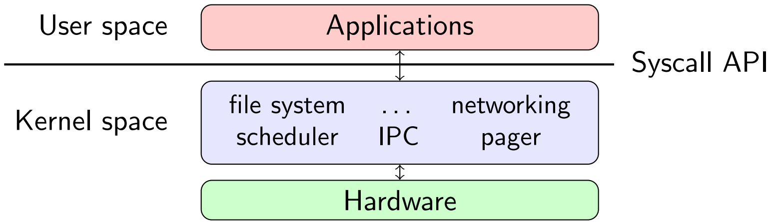
Powerful abstractions such as processes and file descriptors.

> *The success of Unix lies not so much in new inventions but rather in the full exploitation of a carefully selected set of fertile ideas.*

N. Gordon, *Ghosts of the UNIX past: a historical search for design patterns*, LWN, 2010.

Layers

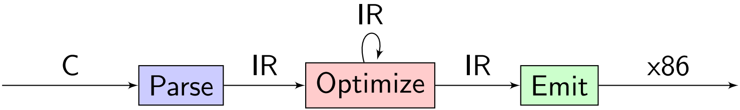
Monolithic kernel architecture



Exercise

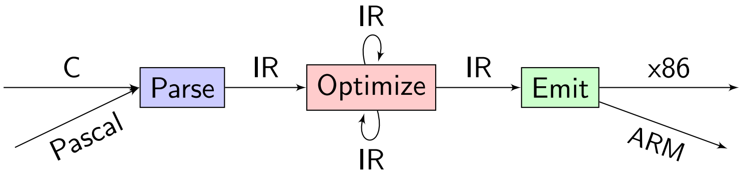
Does Unix follow a strict layered architecture?
Reflect the abstraction level of memory or file descriptors.

Pipes and filters



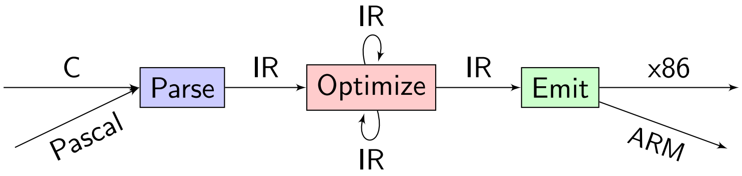
- Parser generators are engineering pearls in themselves

Pipes and filters



- Parser generators are engineering pearls in themselves
- Reason separately about individual “filters” (cf. CompCert)
- Easy extension with new front-ends, back-ends or optimizers?

Pipes and filters



- Parser generators are engineering pearls in themselves
- Reason separately about individual “filters” (cf. CompCert)
- Easy extension with new front-ends, back-ends or optimizers?
- LLVM took this architecture seriously: truly decoupled phases, documented interfaces, ships as library, provides dynamic configuration tools
  $\rightsquigarrow$ maximum re-use, huge community, lots of features

Chris Lattner, *The Architecture of Open Source Applications, volume I*, chapter 11: *LLVM*, 2012.

Plugins

Plugins are dynamically loaded software components:

- Core software defines interfaces
  for acquiring data, converting file formats, building UIs, etc.
- Core software loads plugins dynamically
- Plugins register new implementations of interfaces
- User may explicitly trigger new feature
- In case of implicit use, a selection mechanism is needed

Plugins

Plugins are dynamically loaded software components:

- Core software defines interfaces
  for acquiring data, converting file formats, building UIs, etc.
- Core software loads plugins dynamically
- Plugins register new implementations of interfaces
- User may explicitly trigger new feature
- In case of implicit use, a selection mechanism is needed

Remarks

- Full exploitation of modularity and abstraction
- Eases extension and configuration of software
- Enables external contributions
- Simple "static plugins" already useful for configurable builds

Events

When things are so decoupled, interactions seem quite limited...

Events

When things are so decoupled, interactions seem quite limited. . .

"Please call me whenever event E occurs."

- Sender and receiver don't need to know each other
- Central event manager or peer-to-peer system
- Set of events may or may not be fixed

Events

When things are so decoupled, interactions seem quite limited...

"Please call me whenever event E occurs."

- Sender and receiver don't need to know each other
- Central event manager or peer-to-peer system
- Set of events may or may not be fixed

Limitations

- Some form of "runtime coupling"
- Isolated tests of limited use
- Abuse may lead to messy control flow

Scripting

Video game (Battle for Wesnoth)

- Separate game logic (campaigns, characters. . .) from engine, graphics, etc.
- Simpler script language for high-level stuff
  ⇝ more contributors

R. Shimooka and D. White, *The Architecture of Open Source Applications, volume I*, chapter 25: *Battle for Wesnoth*, 2012.

Scripting

Video game (Battle for Wesnoth)

- Separate game logic (campaigns, characters. . .) from engine, graphics, etc.
- Simpler script language for high-level stuff $\rightsquigarrow$ more contributors

Configuring characters

How to handle combinations such as elvish warriors, drunken ogres, invisible men, and so on?

R. Shimooka and D. White, *The Architecture of Open Source Applications, volume I*, chapter 25: *Battle for Wesnoth*, 2012.

Scripting

Video game (Battle for Wesnoth)

- Separate game logic (campaigns, characters. . .) from engine, graphics, etc.
- Simpler script language for high-level stuff $\rightsquigarrow$ more contributors

Configuring characters

How to handle combinations such as elvish warriors, drunken ogres, invisible men, and so on?

- Theoretician's approach: new language to describe these traits
- WML approach: keep it simple, fixed set of possible features

R. Shimooka and D. White, *The Architecture of Open Source Applications, volume I*, chapter 25: *Battle for Wesnoth*, 2012.

References

- Frederick P. Brooks, *The Mythical Man-Month (20th anniversary edition)*, Addison-Wesley, Prentice Hall, 1995.
- Ian Sommerville, *Software Engineering (9th edition)*, Addison-Wesley, 2011.
- C. Ghezzi, M. Jazayeri, D. Mandrioli, *Fundamentals of Software Engineering*, Prentice Hall, 1991.
- A. Hunt, D. Thomas, *The Pragmatic Programmer*, Addison-Wesley, 2000.

. . . and many others cited in the slides.