CMPSCI 630 Systems **Fall 2014**

Lecture 10

Lecturer: Emery Berger *Scribes: Shaylyn Adams, Theodore Sudol*

The collection of papers discuss in this lecture all deal with system experiences and software engineering and architecture.

10.1 MIT Vs. New Jersey

The Rise of 'Worse-is-better' by Richard Gabriel is a widely cited and used paper. It describes two opposing design philosophies that are known as the **MIT** and the **New Jersey** (worse-is-better) approaches.

The "Right Thing" or MIT approach:

- Strive for simplicity in the interface, even at the cost of complicating implementation
- The design should be as "perfect" as possible, i.e. both correct and complete.
- Make the design as consistent as possible.

The "Worse-is-Better" or New Jersey approach:

- Strive for simple implementation (even at the risk of complicating interface)
- Perfection is not achieved first
- Users can easily port and adapt the system
- "first to market wins" (although classically not true, i.e. VHS vs. BetaMax, Linux vs. Windows)

The New Jersey approach was so named because it is the home of UNIX, C, and C++ which are examples of this design view.

10.1.1 The Canonical Case: Multics versus UNIX

Multics was from MIT and General Electric and is the example of the MIT design approach. It was a groundbreaking OS with multi users, lots of API sets, lots of built in functionalities etc. Multics was formed by design-by-committee which was a main cause of added complexity. The system was supposed to be released in 1963 but it took them until 1969/70 to finish.

UNIX conversely was created by just 2 men who wanted their system to be up and running as soon as possible so they could use it. Hence, UNIX was made by the end user, design-by-user. In the process they designed the C language in order to achieve portable gaming and based it on BCPL which is simple and has one type, the byte. UNIX in the style of NJ, hid complexity via the abstraction where "everything is a file". It avoided the monolithic approach and was built out of many parts. All the basic utilities were easy to reach and right there for the user.

Obviously when looking at systems today, UNIX is the winner in this case and out survived Multics because it was simple and easier to adopt and incorporate into systems.

10.2 "Worse-is-Better" in Programming Language Design

10.2.1 Cult-of-personality Languages

Many programming languages have been designed in the New Jersey style with the main goal of solving a real world problem. **Cult-of-personality languages** refers to languages designed by a *guru*, without any real regard to standard design practices.

The "terrible" language of Perl is example of this. It was created to solve the performance and portability problems with using the command line to do things like:

```
\%ls | grep -v pdf | wc -l
```

Perl (*P*ractical *E*xtraction *R*eport *L*anguage) combines the ideas of grep and awk (a scripting language for the command line) and has many ways to do everything. The main take away being that Perl provides tools to *solve a specific problem*. Examples of cult-of-personality languages are:

- Perl *(Larry Wall)*
- Python *(Guido van Rossum)*
- Ruby *(Yukihiro "Matz" Matsumoto)*
- PHP *(Rasmus Lerdorf, later became design-by-committee)*
- JavaScript *(Brendan Eich)*

They are often known as "curly braces style" languages, meaning they draw inspiration from C-style syntax. They all can do useful things quickly and often interoperate with C easily. These languages can also be used to write programs in a functional style.

The "cult of personality" languages tend to be practical, easy to use, and "nicer". The environment for them has been established. Developed ecosystems and answered certain concerns people had in reality.

10.2.2 Functional Languages

Functional languages are the other kind of programming language ideology which fits under the MIT approach. These languages have mathematical foundations and often have a significant learning curve. Examples include:

- LISP
- Scheme *(made by MIT)*
- Haskell
- ML

Purely functional means programs cannot see state, thus manipulation of state is not allowed.

```
a = f(x);
b = f(x);
a and b are guaranteed to have the same result.
```

This is also known as *referential transparency.*

These functional languages are often build using the MIT approach. They are designed to replace existing systems or to realize an academic goal. In contrast, the "cult-of-personality" languages were designed to solve practical problems.

Functional languages are not as popular since they were made to improve upon existing programming languages instead of solving a real world problem. They reside in the realm of academia. Haskell research is still done today, but in the context of how Haskell applies to the real world, which is something obvious to languages like PHP, C and Python. *Note: Java is the middle of these two classifications*

Scala is a quasi-functional, multi-paradigm language. It runs on a Java virtual machine and can use Java libraries natively. Used by Twitter as a replacement for their original Ruby backend, which could not scale to the demands on their system.

10.3 Notes: Worse-Is-Better

NJ style programming languages are used more in the real world than MIT styled ones. For example, Dropbox and Google use Python, while PHP is used by Facebook, who made their own just-in-time compiler for it called HipHop. Meanwhile, Haskell is just beginning to be used in the industry.

Clearly, worse-is-better is not necessarily pretty, but it does achieve useful results. It is important to note that Gabriel presented the worse-is-better case but he was part of the Common Lisp standardization project. A hybrid of the two approaches may be ideal.

MIT	New Jersey
Multics (took far too long)	**UNIX** (uses 1 standard interface for everything)
"MIT" Languages	**"Cult of Personality" Languages**
	Plan 9*

*__Plan 9__ was designed by Ed Wood from Bell Labs and pushed file abstraction to its limits. It introduced the /proc directory, which shows processes as files and captures the state and memory contents. It never took off because it was made to supplant an existing, popular system and didn't solve a problem. Some of its innovations did get absorbed into UNIX.

Take home message: Design a pretty and simple language that explicitly solves a new/real problem. It does not necessarily have to be deep and academically satisfying.

10.4 UNIX

UNIX introduced many innovations and current standards for operating systems:

Symbolic links: Labels that point to any file or directory, target may be nonexistent. Essentially alias for target.

Hard links: Direct link to a file, must be on same volume. Essentially alias for target data.

Inode: Filesystem index, stores attributes and disk location of file

Filesystem: Data structure stored in persistent memory

Everything is a file: Directories, devices, network sockets, processes...

Piping and I/O redirection: Feed output of command into input of another, or redirect output from stdout to a file

10.5 End-to-End Argument

Saltzer and Clark presented the **end-to-end argument** which is about pushing complexity to the 'ends', i.e. the application level. This argument is demonstrated through the development of the Internet. TCP is an example and it provides reliable delivery at end and in sequence. Problems arise with *fault tolerance* and *security*.

10.5.1 Packet vs. Circuit

This distinction arose when deciding how the Internet should provide communication and tolerate some failures.

Circuits were used with telephony and a fixed routing is established once between A and B. There is no overhead because there is no routing during communication. However, if the channel goes down, everything is lost. This is a single point of failure and there is no fault tolerance.

Packets on the other hand, are like the mail system and practice the end-to-end policy. They contain a source, destination, TTL (time to live) and sequence number. If one route goes down or is unavailable, packets can recover by finding a new path. Once an acknowledgment signal, ACK, is received, the sender knows the message was delivered.

Datagrams as seen in UDP are similar to packets, except there is no guarantee of delivery. This is applicable to things like Skype, streaming media, and phone calls. It is up to the applications to decide how to handle a lost packet. For example, Skype could replace a lost packet with just silence.

10.5.2 Fault Tolerance and Security

Fault Tolerance: How well a system handles errors. An example of a fault is a lost datagram or a file that was corrupted during transmission. In the end-to-end argument, the application must be able to handle the faults.

Security: Since the Internet was made for scientists to communicate in the event of a nuclear attack (like how interstate highways were 'created' to protect us against Communists), security was a low priority so there was no built-in security checks.

One way to secure the network is by implementing *checksums*, which should catch modified packets and help to ensure reliable and correct transmission. However, if these checks were implemented for packets on the network level, the network throughput would drop without actually guaranteeing that the packets do not change. Checksums should instead be implemented by the applications that send and receive data on the networks, as this is simpler to implement and allows better methods for resolving failed checksums.

Security is slowly being integrated into the Internet, and it will be the 'next generation' development.

10.5.2.1 Distributed Denial of Service Attacks (DDOS)

DDOS attacks are a good example of how the network may be exploited for malicious ends. "Denial of Service" means that the server for a website cannot accept or handle more connections. This can occur, for example, if a user shops at a store, walks up to the register, but just sits in front of the cashier instead of paying. If only a few computer initiate an attack, they can be blocked by telling the server to ignore their IP addresses. To get around this, networks of compromised computers called botnets, which can flood the server with hundreds of thousands of connections. This is a "distributed" denial of service attack, or DDOS.

One defense against these attacks is *IPSec*, which ensures that packets cannot be modified en route.

Note: Lampson's hints were not thoroughly discussed; it was noted that they do actually work and have important impacts on systems