Plagiarism in Computer Science

• Copying and editing someone else’s code (with or without attribution) is a standard practice in industry: it greatly reduces cost.

• GitHub.com is a site that specializes in storing, indexing, and sharing code.

• Some CS instructors require their students to use GitHub.

• A student in my Compilers class in a previous semester used GitHub but failed to set file protection; his code was plagiarized by six students (at least) in Spring 2017.
Compilation and Code Obfuscation

A Compiler translates code in a more human-readable source code form into unreadable machine code that actually runs on the computer.

Code Obfuscation tries to change the source code into a form that looks different or is hard to read, but compiles to the same machine code.

Text editors make changes of variable names easy:

Four score and seven years ago, our four fathers made a fourplex using four-by-four lumber.

Using an editor command:

M-X replace-string four purple

yields:

Purple score and seven years ago, our purple fathers made a purpleplex using purple-by-purple lumber.

The compiler doesn’t care what the names are. Students who plagiarize code often try to disguise it using this technique. Those trying to prove plagiarism could use the same technique to restore the code. Edit distance: how many editing steps change one form to the other.
Comments

Comments in code are notes left by humans, for humans. The compiler ignores comments and discards them.

// You can say any #!@& thing you want in a coment.

Students who plagiarize code may try to defend their use of similar code by saying that this code is clearly the correct way to do it. That could even be a legitimate defense.

Typos in comments are likely to be overlooked by the student, but they are a clear red flag: the same typos in the same places are nearly always a sign of plagiarism.
Moss

The Moss program is fairly effective at identifying plagiarism. Moss works by identifying common substrings.

I use Moss to identify the most likely suspects, then look at the code myself. I would not be comfortable bringing a misconduct case based on Moss alone.

Prosecuting a plagiarism case is very time-consuming and unpleasant. I only do it when a TA brings it to my attention, and most TAs do not take the time to do it.

OMG! No, Not the Dean!

I tell students that they have a choice: they can sign the faculty disposition form, or go talk to the Dean of Students.

Most students are terrified of the Dean, and will sign anything to avoid it.