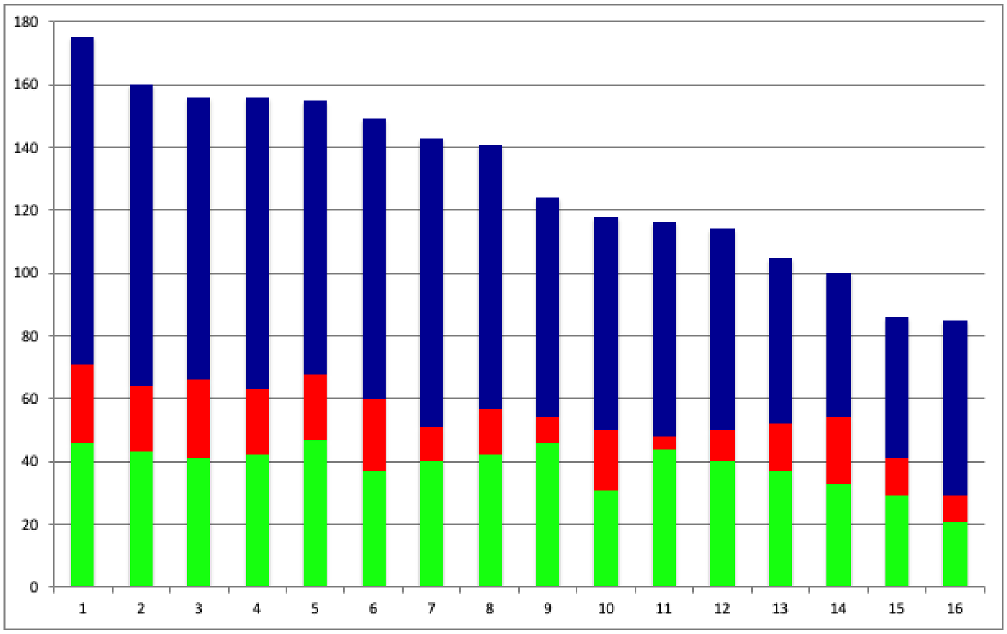
COS 109 Final Exam, Fall 2024

I graded this myself. Last year's median was 121 and the quartiles were 139 and 104. This year's class was much smaller so it's not clear that a comparison is really meaningful, but the medians are about the same, as is the lower quartile; this year's upper quartile is noticeably higher. The colored bars are for parts 1, 2 and 3, reading up from the bottom.



1. (50 points, 2 each) Short Answers. **Circle the right answer or write it in the space provided.**

(a) Which of these got only very belated recognition for their pioneering work on public-key cryptography?

[✓ Clifford Cocks (GCHQ)] **Whitfield Diffie & Martin Hellman (Stanford)**

Ron Rivest, Adi Shamir & Len Adleman (MIT) **Bruce Schneier (Harvard)**

Alan Turing (Bletchley Park) **John von Neumann (IAS)**

Cocks independently and earlier invented a public-key crypto system, but he couldn't publish, one of the problems of working for a spy agency.

(b) The word "**qwerty**" is usually high on the list of most common bad passwords, but variations like **QwertY12** also occur. How many different passwords like the latter can be constructed by different choices of upper and lower case letters in **qwerty** followed by exactly two decimal digits.

6400. 2^6 x 100

(c) Which of these people has/have won both a Turing Award and a Nobel Prize? Circle all that have.

Tim Berners-Lee **[✓ Geoff Hinton]** **Tony Hoare** **John Hopfield** **Yann LeCun** **Alan Turing**

Not Hopfield, who shared the Nobel but is a real physicist, not a computer scientist.

(d) The hexadecimal value **DEADBEEF** could represent many things. Which of the following might it be? Circle all that are possible.

an Ethernet address [✓ **an IPv4 address**] **an IPv6 address**

one pixel of a 24-bit RGB color [✓ **a pair of 16-bit Unicode characters**]

The others are all longer or shorter than 32 bits.

(e) Alice wants to digitally sign a document to convince Bob that she wrote it, and is wondering whether to use AES, RSA or something else. How would you advise her?

AES is clearly better **RSA is clearly better** **both work well** **neither would be good**

RSA. If Alice's private key is not compromised, she's the only one that can encrypt a message with it, and others can decrypt with her public key..

(f) Famous Stanford computer scientist Don Knuth said, "I have just celebrated my 10000th birthday (in base 3)." How old is Don (in decimal)?

81. I think everyone got this one.

(g) To compile Fortran, C and C++ programs for a specific kind of computer, say a Mac running macOS, I would need three different compilers. How many different assemblers would I need?

0 **1** **2** **3** **4** **6**

1. The same assembler would work for any compiler.

(h) Suppose I create an ordinary ASCII text message with my initials "bwk" at the end and save it in a file. Then I change "bwk" to "BWK" and save it in a different file. How many bits are different between the two files? How many bytes are different?

number of different bits ______________________ **number of different bytes** ____________________

3 different bits, in 3 different bytes

(i) Suppose that I use SHA-256 to compute a cryptographic hash of the original message in the previous question. Then I change "bwk" to "BWK" and re-compute the hash. What is the relationship between the first cryptographic hash value and the second?

they are the same **about 128 bits are different** **about 128 bytes are different**

every bit is different **no way to predict**

About half (128) of the bits are different. It can't be every bit; otherwise, just flip each one and you have the original back, which isn't much of a hash.

(j) Each of the following files is exactly 100 MB long and each contains typical information of the type indicated by its filename extension. Which one of these files would likely be smallest after Lempel-Ziv compression is applied to it?

`F.gif` **`F.jpg`** **`F.mpg`** **`F.mp3`** **`F.png`** **`F.txt`** **`F.zip`** **no way to tell**

F.txt. All of the others are already compressed.

(k) David Auerbach's 2018 book *Bitwise: A Life in Code* includes this sentence: "Store this number here, retrieve this number from there, add or subtract these two numbers, and branch to different bits of code depending on some condition or other." What kind or level of programming language is being described in this excerpt?

Assembly language. “Low-level” is sort of correct but not as specific as I was looking for.

(l) My laptop has several long unidentified identifiers on the bottom, including one or more valid Ethernet addresses. Circle those of the following that are likely to be Ethernet addresses.

[✓ 000D5665BEB7] 88956126-A 006DX1EF7FFE [✓ 08:00:07:1A:3D:56] BD0C63920DZJ

The others are either the wrong length or have non-hex characters (or both).

(m) The running time of a particular algorithm is proportional to the square root of the number of items it processes (i.e., its complexity is $\sqrt{\mathbf{n}}$). Where does $\sqrt{\mathbf{n}}$ belong in the following list? The answer is either one of the expressions given or one of the five positions marked with “---". Circle the correct expression or position.

--- log n [✓ ---] n --- n^2 --- 2^n ---

Between log n and n. It’s n^(1/2) so it can’t be greater than n if n >= 1.

(n) A news story says “Taylor Swift ticket demand crashes Ticketmaster website.” What kind of “attack” is this?

[✓ DDoS] IoT MITM MS-DOS Swift Boat Trojan horse

Almost everyone got this.

(o) If I wanted to buy a WW2 Enigma machine at auction, which is these is the most likely price that I might encounter today, within a factor of say two or three?

\$2,500 \$25,000 \$250,000 \$2,500,000 \$25,000,000

\$250,000 is representative of the most recent sales, as we saw in class one day.

(p) The *NY Times* reported on 12/10/14 that the US government had recovered “144,336 bitcoins found on computer hardware belonging to the creator of Silk Road”, a site largely devoted to the sale of illegal drugs. Which one of these specific hardware components would be the most likely place to “find” bitcoins?

accumulator bus cache CPU disk GPU RAM ROM

disk. Nothing else is storage that is both writeable and permanent. Most people got it.

(q) From James Gleick, *The Information* (2010): “Morse and Vail had realized that they could save strokes by reserving the shorter sequences of dots and dashes for the most common letters. But which letters would be used most often? Little was known about the alphabet's statistics. [To learn more] Vail visited the local newspaper office in Morristown, NJ and looked over the type cases.” What dozen letters would Vail discover are used most often in normal English?

ETAOIN SHRDLU, the famous phrase discussed at length in class when we talked about compression. Things I learned: this Vail is Davis Vail, father of Theodore Vail, who learned telegraphy in Morristown and later founded AT&T.

(r) Suppose we start a Towers of Hanoi game with N disks at the beginning of class. If it takes 30 seconds to move one disk from one pin to another (they're heavy, being made of solid gold), what's the largest value of N for which we could finish a game in an 80-minute class period?

7. 2^7 is 128, so at half a minute each, we could get that done in 64 minutes.

(s) On 10/24/24 the *NY Times* said “Where ________________’s microprocessor chips excelled in rapidly executing calculations one after another, ________________’s chips delivered superior performance in graphics by breaking tasks up and spreading them across hundreds or thousands of processors working in parallel, an approach that would pay off years later in artificial intelligence.” Fill in the blanks with the names of the two chip companies?

Intel, Nvidia. Almost everyone got this one.

(t) Amazon.com and the government of Brazil both want to own the top-level domain `.amazon`. Which one of these organizations is responsible for adjudicating who gets the domain name?

FCC FTC [✓ ICANN] ITU UNESCO USPTO WIPO

(u) Which one of these entities would I have to deal with if I want to license radio frequency spectrum for a new wireless service in the USA? Circle the correct answer(s).

EULA [✓ FCC] FTC ICE IETF NIST RTFM WTF

(v) Zoom uses a point-to-point network architecture where each party in a video conference is connected to all the others through Zoom's servers. How does the number of connections grow in proportion to **N**, the number of participants?

logarithmically linearly N log N quadratically exponentially

linearly. Note "through Zoom's servers". That's a star network.

(w) From *The Dark Hours*, a 2021 detective story by Michael Connelly: "He also uses ____________as a browser. It encrypts his moves and bounces them all over the world. So he's anonymous." Passing over whether this is strictly accurate, what belongs in the blank?

Tor. Everyone got this freebie.

(x) In the factoring challenge sponsored by RSA Labs, RSA-1024 (not factored yet) is 1024 bits long in binary and 309 digits long in decimal. RSA-2048 (also unfactored) is 2048 bits long. ***Approximately*** how many digits would it have if written out in decimal?

618, easiest computed by multiplying 309 by 2. No need to do logs, though I accepted nearby answers, which could be correct..

(y) Crossword quickies from the *NY Times Magazine*:

"Conditional coding word" (9/21/24) 4 letters, ends with E) _____ **ELSE** __________

"In which 1 + 1 = 10" (10/27/24) 6 letters ______ **BINARY** _____________

2. (25 points) Understanding Programs

(a) The following Python code is supposed to simulate flipping a fair coin ***exactly*** 1,000 times. At the end, it should print the number of heads and tails. Sadly, it has several errors and doesn't work. Fix the errors. You do not need to rewrite it if you ***clearly*** indicate the changes you would make. (This is a question about correct logic; don't worry about syntax. The expression for computing random numbers is correct: each call of **`random.random()`** produces a new random floating-point value between 0 and 1. The **`print`** statement is syntactically correct as well.)

```
i = 1
heads = 0
print("heads =", heads, "tails =", tails)
while i < 1000:
   r = random.random()  # random number r >= 0, < 1.0
   if r >= 0.5:
      heads = heads + 1
   else:
      tails = 1

tails = 0
move print(…) to the end  (missed by many)
while i <= 1000:          (missed by many)
   ...
```

```
else:
   tails = tails + 1
i = i + 1                  (missed by many)
```

(b) If you want to simulate an unbalanced coin that comes up heads 3/4 of the time and tails 1/4 of the time, what change(s) would you make to the program above to achieve this, after it has been corrected?

`r >= 0.25  (or r <= 0.75)` but not `r >= 0.75`

(c) Suppose that version 2.0 of the Toy machine includes an instruction called **REM** (for remainder), which divides its operand into the value in the accumulator and leaves the remainder in the accumulator. For example, if the accumulator contains 17, the instruction **REM 5** will leave 2 in the accumulator. What does the following program print when given the sequence of input numbers **3 1 4 1 5 9 2 6 5 4 0**?

```
TOP    GET                 get a number from keyboard into accumulator
       IFZERO   BOT        if accumulator value is zero, go to instruction BOT
       STORE    TEMP       store accumulator value in location TEMP
       REM      3          divide accumulator value by 3, put remainder in accumulator
       IFZERO   TOP        if accumulator is zero, go to instruction TOP
       LOAD     TEMP       load accumulator with value from location TEMP
       PRINT               print value in accumulator
       GOTO     TOP        go to instruction labeled TOP
BOT    STOP
TEMP   0                   when execution begins, this location will contain 0
```

1 4 1 5 2 5 4

(d) Briefly, in at most half a dozen words, what computation is this program performing? *Do NOT just repeat the instructions in words.*

Prints numbers not divisible by 3.

(e) How does the running time of the program grow as a function of or in proportion to **n**, the number of input numbers?

log n n n log n n^2 n^3 2^n depends on sizes of input numbers

n. Does just about the same thing for each input number.

(f) John von Neumann said “We introduce an order (the conditional transfer order) which will, depending on the sign of a given number, cause the proper one of two routines to be executed.” What Toy machine instruction would this order correspond most closely to?

IFPOS. Not **IFZERO**, a popular wrong answer. That doesn’t depend on the sign.

3. (105 points, 5 each) Miscellaneous

(a) A guide to online privacy says “That [single] cookie, (1) placed on your computer when you visit most web sites, will contain information about (2) the names of the sites you visit, (3) how often you visit them, (4) what you click on, and (5) how long you stay on each site.” I have marked the five assertions about cookies in the quoted sentence. For each, is it sort of accurate, or is it basically misleading or wrong?

(1) [✓ **accurate**] **wrong**

(2) **accurate** [✓ **wrong**]

(3) **accurate** [✓ **wrong**]

(4) accurate [✓ wrong]

(5) accurate [✓ wrong]

Cookies usually contain some number that identifies you when you return, but not the analytics that this mentions, and in any case never anything that didn't originate on the server.

(b) Many years ago, a COS109 student told me that the instructions for one of the labs would not print properly when she used Safari. I could see nothing obviously wrong, so I started removing parts of the file (originally about 6,400 lines) to find smaller versions that still exhibited the problem. Describe a *systematic general procedure* for efficiently locating such a problem, when you have no clue at all about what the problem is. (It turned out to be a missing `</a>`.) ***Be brief!*** I'm looking for an idea, not an essay. 10-15 words should be plenty.

Binary search. Pretty much everyone got this.

(c) Canadian postal codes (the equivalent of US ZIP codes) have the form **`LDL DLD`**, where each **`L`** is a letter and each **`D`** is a digit, as in **`N3B 0A2`**, where my sister lives. Only 20 of the 26 possible letters are used.

(i) If postal codes are encoded in binary with each letter and digit encoded separately, how many bits are required for a postal code?

27. Each L takes 5 and each D takes 4.

(ii) If instead each possible code is assigned its own unique binary value by enumerating all possible codes, how many bits are required for a postal code?

23. There are 8 million combinations. Both parts were well done.

(d) Ken Auletta's *Googled: The End of the World as We Know It* (2009) says that Google stores "two dozen or so tetabits (about 24 quadrillion bits)."

(i) As you know by now, there is no such thing as a tetabit. Assuming that "quadrillion" is correct, what word should have been used instead of tetabit?

Petabit (since quadrillion is right, it's 10^15).

(ii) How many ***petabytes*** did Google store at that time?

3 PB. Divide 24 by 8.

(iii) If (hypothetically) Google's storage use doubles every 12 months, roughly how many ***zettabytes*** would Google store in 2029, 20 years after the book was published?

3 ZB. It's a factor of a million (20 doublings). No fractional parts in this simple view of the world.

(e) The number **`401b30e3b8b5d629635a5c613cdb791e`** has 32 hexadecimal digits.

(i) Which of these could it be? Circle all that are possible.

[✓ **AES key**] **Ethernet address** **IPv4 address** [✓ **IPv6 address**] **SHA-256 hash**

The others are the wrong length.

(ii) ***Briefly*** explain why this number is not a prime number.

Its numeric value in binary ends in 0, so it's a multiple of 2. Saying that **`E`** is 14 is not quite right, but I was generous.

(f) Here are some hexadecimal values that sort of spell words; the character **`0`** is a zero and **`1`** is a one.

`BEDDED` **`C0FFEE`** **`D00DAD`** **`EFFACE`** **`FACADE`** **`F00D1E`** **`0FF1CE`**

(i) If they are interpreted as ordinary 24-bit integer values, which one has the smallest numeric value?

`0FF1CE` starts with 0; none of the others do. No arithmetic needed.

(ii) Which one has the largest numeric value?

`FACADE` starts with F; none of the others do. No arithmetic needed.

(iii) If instead they are interpreted as 24-bit RGB colors, which one has the largest amount of blue?

`COFFEE.` EE is bigger than ED, the only other contender. No arithmetic needed.

Notice a common theme?

(g) The first half of the first byte of an IP packet contains the version number of the protocol.

(i) Write out the bit patterns that one might most reasonably expect for IPv4 and IPv6.

IPv4 _______ **0100** ____________ IPv6 _______ **0110** _______________

(ii) What is the largest version number that this scheme allows for, in decimal?

15 (1111)

(h) A recent story about Google's management structure says (paraphrased) "Imagine N employees making decisions by everyone talking to everyone else."

(i) What is the complexity of this decision-making process, as a function of N, the number of employees?

N^2.

(ii) Suppose instead that there is a traditional hierarchical structure where each manager has at most 10 direct reports. If Google has roughly 100,000 non-management employees, how many levels of management would there be above them?

5. It's log base 10 of 100,000 but just as easy to write down the numbers.

(i) As data travels across the Internet, it is subjected to a fair amount of processing. For each of the following statements, circle the most appropriate answer.

IP packets have serial numbers to ensure that they are processed in the right order **true** [✓ **false**]

IP packets that arrive out of order have to be resent **true** [✓ **false**]

a long TCP message is broken into multiple IP packets [✓ **true**] **false**

Ethernet packets are reassembled into IP packets at each router along the way [✓ **true**] **false**

If an IP packet is damaged in transit, error correction bits will restore it **true** [✓ **false**]

(j) This partial Unix directory listing shows size in bytes, modification date and time, and filename for five files. Exactly which pair(s) of files do I have to compare byte by byte to determine whether or not they have identical contents?

```
347   Oct 29 16:04   f1.doc
354   Oct 28 16:05   f1.docx
354   Apr 22 20:03   f1copy.docx
355   Sep 20 08:51   f2.txt
354   Aug 20 08:51   f3.xls
```

Compare the three files with the same size; .xls might be a lie.

(k) The *NY Times* (12/10/18) says that companies are continuously tracking 200 million US cell phones many times per day per phone. Suppose that an average phone reports its number and its position to an accuracy of one yard or meter 1,000 times/day. Assuming that there is a reasonable attempt to be efficient, estimate ***very roughly*** how many terabytes of tracking information are uploaded by all these phones every day in total. Be precise and clear about your assumptions about how information is represented.

4 TB? Figure 10 bytes for a phone number and 10 bytes for a position in some encoding.

(l) Google Maps shows real-time traffic information for most roads. For each of the following statements, assess how likely it is to be basically correct about how traffic information works.

Cars are counted by analyzing images from satellite cameras	likely [✓ unlikely]
Roadside cameras provide most of the traffic data	likely [✓ unlikely]
The driver's cellphone can report the car's position	[✓ likely] unlikely
Cars can report their position by uploading it to a GPS satellite	likely [✓ unlikely]
A car must have a dedicated GPS unit to report its position accurately	likely [✓ unlikely]

(m) Charles Babbage's mechanical computers used decimal arithmetic. Each digit of a number was represented by a wheel with 10 values around its circumference; thus a 12-digit number would require 12 wheels. Imagine that Babbage had taken an early version of COS109, realized the advantages of binary representation, and wanted to build a binary mechanical computer that would handle numeric values up to at least one trillion (decimal).

(i) If Babbage were to use binary wheels (only two values on each wheel) instead of decimal, how many wheels would he need to handle decimal numbers up to one trillion?

40. 10^12 is 2^40

(ii) If he were to use hexadecimal wheels instead (16 digits on each wheel), how many wheels would he need for decimal numbers up to one trillion?

10. Each wheel takes care of 4 bits.

(iii) What hexadecimal value would appear on these wheels when displaying the largest possible number?

FFFFFFFFFF. The largest possible number is not 1 trillion!

(n) A deep-space communications system reports on the health of a piece of equipment by sending a continuous bit stream of status reports. There are three possible status values: OK, High and Low. 98% of the time, the status is OK, while High and Low each occur only 1% of the time. Give an encoding of the three values into three different bit patterns that will minimize the average number of bits sent over a long period of time. Your encoding does not have to use the same number of bits for each status, but there must be no ambiguity about how to decode a sequence of values as they arrive at the receiver.

0 10 11. Note the "no ambiguity" part, which was missed by many.

(o) Suppose I want a sorted list of the names of all students whose birthday is in January. One algorithm is to sort all students by name, then select those born in January. A second algorithm is to select the students born in January, then sort the result. Assuming there are 6,000 students, is the first algorithm likely to be faster than the second, slower, or about the same, and why? You must make your case by quantitative reasoning.

Second. First is 6000 log 6000 + 6000; second is 6000 + 500 log 500

(p) A *Mersenne prime* is a prime number of the form $2^n - 1$ where n itself is prime, for example $31 = 2^5 - 1$. In October 2024, a new Mersenne prime was discovered: $2^{136279841} - 1$. It's the largest known so far, and it has 41,024,320 digits in its decimal representation.

(i) If it is written out in binary, how many binary digits does it have?

136279841

(ii) How many of those binary digits are zero?

None

(q) Suppose that part of the file system on a computer looks like the following diagram, where "`D`" implies directory and "`F`" implies ordinary file. Names indented to the same level are in the same directory, so, for example, directory `D1` contains `F1.TXT`, `F2.DOC`, `D2` and `D4`.

```
D1
    F1.TXT
    F2.DOC
    D2
        F3.DOC
        F4.XLS
        D3
            F5.DOC
    D4
        F6
        F7.DOC
        D5
            F5.DOC
```

(i) If a program like Finder or File Explorer starts at the root of this structure, which items must the program read to find all of the `.DOC` files in directory `D1` and its descendants?

D1 D2 D3 D4 D5. You have to read all the directories, but none of the files.

(ii) Which items must be read to determine if `F4.XLS` is an Excel spreadsheet?

D1 D2 F4.XLS. You have to read the file to be sure of its contents.

(r) Supercomputers with lots of processors are sometimes organized as a "mesh" where each processor is connected to its nearest horizontal and vertical neighbors on a rectangular grid. Suppose that there are **N** processors, each processor is an identical rectangular box, and the boxes fill a large room from floor to ceiling.

(i) How many connections to neighbors does a typical processor have?

6. Not 4 – these are 3-dimensional packings.

(ii) How does the total number of connections among all processors grow in proportion to **N**?

Linear.

(iii) If technology improves so that the current length, width and height of each processor can each be shrunk by a factor of four, about how many processors would now fit in the room?

64. 4 x 4 x 4

(s) Suppose, not unrealistically, that **N** high-tech companies are involved in a bunch of lawsuits.

(i) If each company sues each other company, how does the number of lawsuits grow in proportion to or as a function of **N**?

N^2.

(ii) Companies may also band together in groups of various sizes to sue companies that are not in the group; for instance if **N** were 4, we might have A suing B, C and D; A and B suing C and D; A, B and C suing D; and so on. If all possible combinations of companies initiate such suits, how does the number of possible lawsuits grow in proportion to **N**?

2^N.

(t) Random quickies (10 points).

Bitcoin can't be used to pay off ransomware demands because it's anonymous	**true**	[✓ **false**]
Your laptop's Ethernet address changes as you walk from Friend 008 to the Dinky station	**true**	[✓ **false**]
A single parity bit can correct a single-bit error in a single byte	**true**	[✓ **false**]

Bitcoin is an example of a cryptocurrency with a relatively stable dollar value **true** [✓ **false**]

In WW2, Navajo code talkers relied on security by obscurity [✓ **true**] **false**

A lossless compression algorithm will make some inputs larger, not smaller [✓ **true**] **false**

The first trans-Atlantic fiber-optic cable was laid in the 1860s **true** [✓ **false**]

Every DNS query accesses a registrar **true** [✓ **false**]

If you use HTTPS to access a web site, your ISP does not know which site it is **true** [✓ **false**]

If you use Tor to access a web site, your ISP does not know which site it is [✓ **true**] **false**