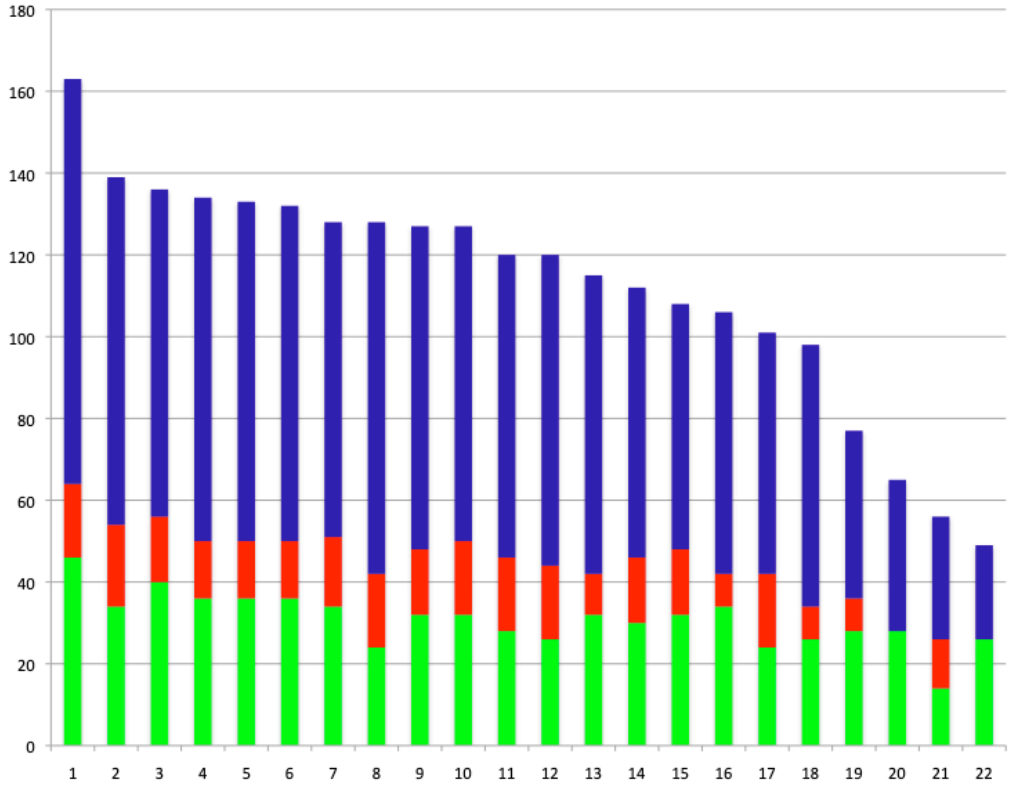
COS 109 Final Exam, Fall 2020

I graded this myself. The median was 120, the upper quartile was 132, and the lower was 101. The corresponding values for last year were 101, 120 and 74, all much lower. Is this because of an easier exam, smarter and more dedicated students, better instruction, the wonders of Zoom, or a change in the political outlook? Your guess is as good as mine. The colors in the graphs below 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) Whose picture will appear on the United Kingdom's new 50 pound note?

Alan Turing (and Queen Elizabeth on the obverse, as someone noted)

(b) In a conventional Windows or macOS file system, which of these is the most likely relationship between the number of files and the number of physical disk blocks?

many more files than blocks **many more blocks than files** **about the same number of files and blocks**

many more blocks than files. Blocks are typically a few KB, files are very large (think pictures or videos), and file with any content at all has at least one block.

(c) 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**

IPv4, pair of Unicode characters. DEADBEEF is 32 bits. Ethernet is 48, IPv6 is 128, 24 bits is not 32.

(d) If I use my cellphone camera to make a movie of daytime traffic on Nassau Street in Princeton, it has enough memory for about 20 minutes. If instead I make a movie of the night sky while looking for meteor showers, what will I likely discover about the length of movie I can make?

longer movie at night **shorter movie at night** **about the same** **no way to predict**

longer movie at night, because the mostly uniformly dark sky will compress much better.

(e) In whose collected works would you most likely find the sequence 1, 2, 3, …, ut, re, mi, fa, sol, la?

Babbage **Bach** **Goethe** **Hertz** **Leibnitz** **Mahler** **Newton**

Leibnitz, whose use of musical notes for the extra hex digits was discussed in class one day.

(f) Most modern computers have multiple "cores," that is, two or more individual CPUs on a single chip, like the Intel chips in current PCs and Macs. Assuming that all the potential processing power can be perfectly utilized, how does that processing power increase in proportion to **n**, the number of CPUs on a chip?

logarithmic **linear** **n log n** **quadratic** **cubic** **exponential**

linear

(g) Google's search index contains "hundreds of billions of web pages" and is over 100,000,000 gigabytes in size. If Google stores 200 billion web pages, what is the approximate size of an average web page?

500 KB. 100 x 10^15 bytes / 200 x 10^9 pages

(h) My brother turned 75 last month but said he felt much younger: "75 is the new 50, at least in base _____." What base was he using?

15. In base 15, 50 is 5 * 15^1

(i) From a new cyber-security bill in the US House of Representatives: "_____ devices have at least one transducer (sensor or actuator) for interacting directly with the physical world, have at least one network interface, and are not conventional Information Technology devices, such as smartphones and laptops." (*H.R. 1688*, 11/20/20) Which of these belongs in the blank?

5G **Bluetooth** **Cyber-warfare** **Drone weapons** **Integrated circuit**
Internet of Things **Radio frequency identification** **Two-factor authentication**

Internet of Things

(j) Alice says "A crypto algorithm is much more likely to be secure if everyone knows exactly how it works." Bob says "Nonsense! The only way to make a cryptographic algorithm really secure is to keep how it works a secret." Eve says "The real security has to be in the key you use." Who is right?

none of them **only Alice** **only Bob** **only Eve** **both Alice & Eve** **both Bob & Eve**

Alice and Eve. We talked about the public process of creating AES, and how security by obscurity doesn't work.

(k) A Duo "push" is a 6-decimal-digit number that I have to type to log in to Princeton systems. What's the least number of bits necessary to encode this number to send it to OIT for validation?

20 bits. 2^20 is somewhat over one million (10^6). A fair number of people came up with 17 bits, which would handle numbers only up to 2^17, or about 131,000.

(l) On November 18, 2020, the _____ decided to allocate 45MHz of spectrum to Wi-Fi use, taking it away from a little-used vehicle-to-vehicle automobile safety technology. Which of these entities

made this decision?

Auto Alliance **Congress** **FCC** **FTC** **ITU** **NHTSA** **Trump administration**

FCC, which is in charge of spectrum allocation.

(m) Which pair of these are most closely related in purpose? Circle the two acronyms.

CCPA **DMCA** **EULA** **GCHQ** **GDPR** **MIME** **MVNO** **RTFM**

CCPA, GDPR. Both are laws intended to give consumers the right to control how much data can be collected about them by online sites.

(n) A recent technical white paper says "5G phone systems should be able to support 1 million connected devices per square kilometer." How many supported devices could you fit into a square meter?

1. A square kilometer is 1 million square meters. Missed by a fair number of people.

(o) Which one of these people is generally credited with first suggesting that computers could do more than just compute numerical values?

Charles Babbage **Anthony Babington** **Tim Berners-Lee** **Vannevar Bush**

Grace Hopper **Ada Lovelace** **Alan Turing** **John von Neumann**

Ada Lovelace.

(p) What decimal *integer* is the infinitely long binary number `1101100.1111111111`... closest to?

109. Convert 1101101 to decimal.

(q) When Nicki in the office next door to me visits Amazon.com, the first web page she sees says "Hello, Nicki", but when I visit Amazon.com, the page (otherwise identical) says only "Hello". *In a word or two*, what technical mechanism probably underlies this difference?

Cookies

(r) If **n** and **m** are integers, how many 1-bits (that is, bits whose value is 1) are there in the binary representation of $4^n \times 4^m$?

1. It's a power of 2.

(s) Google Maps is usually very responsive: at any particular resolution, the map moves smoothly as it is dragged in any direction, though occasionally there is a noticeable pause until some area is filled in. Which one of these techniques contributes most to its responsiveness?

caching **client-server** **cloud computing** **compilation** **compression** **conditional execution**

caching

(t) In the normal hexadecimal representation of colors, which of these colors does the value **`F00F00`** most closely approximate? (**`0`** is the digit zero.)

red **green** **blue** **cyan** **yellow** **magenta** **black** **white**

red. F00F00 is a lot of red, very little green, no blue.

(u) If I want to create a new top-level Internet domain called `.bwk`, analogous to `.biz`, `.info`, etc., which one of these would have to authorize its creation?

DNS **ICANN** **ITU** **registrar** **root server** **TLA** **W3C** **WIPO**

ICANN

(v) I have about 1000,000 files on my computer, of which about 1,000 are Word .doc and .docx files. How many times do I have to run Word to compute the total number of bytes in all of those .doc and .docx files?

none **once** **1,000 times** **1000,000 times** **no way to tell in advance**

none. The sizes are in the folders/directories, so you don't need to look at the contents.

(w) Which one of these Princeton alums was an inventor on an important controversial patent?

Jeff Bezos '86 **Bob Kahn *64** **Brian Kernighan *69**

Eric Schmidt '76 **Brad Smith '81** **Meg Whitman '77**

Jeff Bezos, 1-click. Almost everyone got this freebie.

(x) Suppose you are converting a decimal number **n** to its binary representation by hand, by repeated division by two. Which of these expressions best describes how the amount of work you have to do depends on **n**?

log n **n** **n log n** **n^2** **2^n** **doesn't depend on n** **no way to tell**

log n. Practically the definition of how to compute a log base 2.

(y) 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 mixing upper and lower case letters in `qwerty` followed by exactly two decimal digits?

6400. 2^6 * 10 * 10. Many people added 100 to 64, rather than multiplying. This was meant to be an easy variant of a midterm question.

2. (20 points) Understanding Programs

(a) [10 pts] The following Python code is supposed to print a 3-column table that shows each integer from 1 to 100 inclusive, together with its square and its cube; there should also be a line at the beginning of the table that labels the three columns. Sadly, the program doesn't work. Fix the errors by rewriting the code or clearly showing the changes you would make. (This is a question about correct logic, not syntax.)

```
n = 0
while n < 100:
    print("n     n squared      n cubed")
    print(n, n*n, n*3)

print("n     n squared      n cubed")
n = 1
while n <= 100:
    print(n, n*n, n*n*n)
    n = n + 1
```

Among the common errors, failing to put the heading outside the loop, and getting the loop limits wrong in various ways.

(b) [6 pts] Suppose that the Toy machine is augmented with a new instruction `CHSIGN` that changes the sign of the value in the accumulator. That is, if the accumulator value is positive it becomes negative, and if it is negative it becomes positive; zero is unchanged. Here is a small program that uses the `CHSIGN` instruction, with reminders about what the instructions do.

```
MORE  GET            get a number from user, place it in accumulator
```

```
        IFZERO END      if accumulator value is 0, go to END
        IFPOS  MORE     if accumulator value is >= 0, go to MORE
        CHSIGN          change sign of value in accumulator
        PRINT           print value in accumulator
        IFPOS  MORE
END     STOP
```

If this program is given the sequence of input numbers **2 1 –7 3 –8 –4 –6 5 9 0**, what does it print?

7 8 4 6. It ignores positive numbers, but prints the absolute value of negative numbers. Mostly well done, I was pleased to see.

(c) The Python function **`weird`** takes three input values and returns one computed result. What name would you give to the function that would convey much more clearly what it computes?

```
def weird(v1, v2, v3):
   m = v1
   if v2 > m:
      m = v2
   if v3 > m:
      m = v3
   return m
```

max or any variant thereof.

(d) Modify the implementation of **`weird`** in any way that preserves its API and correct operation. You only need to clearly indicate what you would change.

Change > to >= in either or both places, change variable names, change order of tests, … Anything that preserves the behavior. Changing "return" to "print" changes the behavior, so that's not right.

3. (110 points, 5 each) Miscellaneous

(a) In *Click Here to Kill Everybody*, Bruce Schneier says "If 100 systems are interacting with each other, that's about 5,000 interactions and 5,000 potential vulnerabilities resulting from those interactions. If 300 are all interacting with each other, that's 45,000 interactions."

(i) If 1,000 systems interact with each other, ***approximately*** many interactions would Schneier report?

500,000.

(ii) How does the number of such interactions grow as a function of or in proportion to the number of systems?

N^2. A standard example of a quadratic process. The clue is that each system interacts with each other.

(b) The US postal service encodes a lot of address information in "Intelligent Mail" barcodes like the one in the picture below. There are 65 vertical bars.

(i) How many different possible addresses can this encoding represent? Just give an expression.

4^65 or 2^130. I think this was covered in the Q/A; I hope the attendees benefited.

(ii) What is the closest power of 10 to this number?

10^39. Easiest seen from the fact that 2^100 is about 10^30; count up from there.

(c) Suppose that you were to fill your bedroom (let's say about 10' x 12' x 8' and ignore furniture) with 256 GB SD cards. ***Approximately*** how many exabytes of storage could you fit in? Explain your answer quantitatively.

5 EB. 10 x 12 x 8 is about 10^3 cubic feet. Nearly 2000 cubic inches in a cubic foot, so 2x10^6 cubic inches. About 10 cards / cu in, so 20 x 10^6 cards x 250 x 10^9 bytes = 5000 x 10^15 = 5 x 10^18. Some people had trouble with the various powers of 10, and there were some 2-dimensional rooms in evidence, but on balance this was gratifyingly well done.

(d) Suppose that you want to build a computer network with some redundancy: every router is connected to ***exactly two*** other routers.

(i) Sketch what your networks would look like for 4, 5, and N routers.

A ring of N routers. No artistic talent required.

(ii) How does the total number of connections between routers grow in proportion to N, the number of routers?

log N **N** **N log N** **N^2** **N^3** **2^N**

N. Each new router has to be connected to exactly two others.

(e) Circle the best answers.

"AI Winter" refers to an artificial intelligence funding cycle that begins every January	**true**	**false**
In supervised learning, a human monitors each decision made by an ML system	**true**	**false**
Deep learning systems are based on multiple layers of artificial neurons	**true**	**false**
Expert systems are the technology behind today's most successful chess programs	**true**	**false**
Machine-learning spam detection systems are usually based on unsupervised learning	**true**	**false**

(f) Somewhat surprisingly, with 23 people in a room (remember the days when we used to meet in classrooms?) the odds are about 50% that two people will have the same birthday; with 80 people, it's 99.99%. Suppose that each person writes his or her name and birthday on an index card.

(i) Describe an ***efficient*** algorithm to determine whether any two people have the same birthday. (Don't worry about multiple duplicates or triples.) *Be brief*; two or three short sentences ought to be enough.

Sort, look for adjacent duplicates. I really wanted more than how Quicksort works.

(ii) If there are N index cards, how does the time that your algorithm takes vary in proportion to N?

n log n

(g) "USB flash drives, left in a company parking lot, adorned with the company logo, are picked up by curious employees, who put them in their computers and open what looks like an innocuous Word document. In fact, once run, it is software that collects passwords and other confidential information and sends it to the attackers."

(i) What specific kind of attack is this an example of?

denial of service **man in the middle** **phishing** **scareware** **spyware** **Trojan horse**

Trojan horse.

(i) What programming language would be most suitable for creating this attack?

Visual Basic. We did this in class at some length.

(h) Many years ago when disk space was a precious resource, various companies offered software that automatically compressed all files on a hard disk, typically gaining a factor of two. Suppose a (hypothetical) startup called squeezit.com makes a better offer: send them any disk at all and they will losslessly compress it by a factor of two and send it back. Is this claim credible (yes, no, maybe), and why do you say so? If it isn't clear a priori, what experiment(s) would you perform to verify or disprove their claims? ***Be brief.*** I'm looking

for an idea, not an essay.

No. Reductio ad absurdum: keep sending back the result of the previous submission. There was a fair amount of waffling and "maybe it works".

(i) Suppose I want to use ***cryptography*** to convince you that I did not change the exam after the Q/A session. Describe a procedure that uses cryptography that could have guaranteed that any changes made after the session would be detectable. You might or might not find some of these helpful: AES, CAS, DES, RSA, SHA-3, Tor. ***Be brief***. I'm looking for an idea, not an essay.

Compute and publish a secure hash of the exam PDF (e.g., with SHA-3) ***before*** the Q/A. Run the same algorithm on the PDF of the exam when you receive it. Any change will result in a different hash.

(j) Late in 2020, Apple announced its new M1 laptops, which are based on ARM architecture instead of Intel X86. Assess the likelihood of each of the following statements.

Apple will have to pay license fees to ARM Holdings	**likely**	**unlikely**
Apple will need to write a new assembler for the M1	**likely**	**unlikely**
Compiled M1 programs will run unchanged on old Macs	**likely**	**unlikely**
Apple could simulate the M1 on existing Macs	**likely**	**unlikely**
Apple could simulate Intel X86 software on M1 Macs	**likely**	**unlikely**

(k) A while ago the *New York Review* predicted that the number of Internet-connected devices would grow from 12.5 billion in 2010 to 50 billion by 2020.

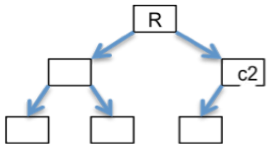
(i) Assuming that this is a smooth exponential growth, what is the growth rate per month of the number of Internet-connected devices?

1.2%. It quadruples in 10 years so it doubles in 5 years; 72/60 is about 1.2%.

(ii) In what year will there be 50 trillion connected devices if growth continues at the same rate?

2070. It's doubling in 5 years, so there are 10 doublings in 50 years, taking 50B to 50T.

(l) A particular computer network is organized as a *balanced tree*: the root computer **R** is directly connected to two other computers, each of which is in turn connected directly to two other computers, and so on, with no duplicates. "Balanced" means that the connections are made so that as much as possible each computer has exactly two children. For instance, if another computer were added to the network below, it would be added as the right child of **c2**.



(i) Which of these terms best describes how the number of connecting wires will grow in proportion to **n**, the number of computers on the network?

logarithmic **linear** **n log n** **quadratic** **cubic** **exponential**

linear. Each computer is connected to one, two or three others.

(ii) Which of these terms best describes how the maximum distance from any computer to any other computer in the network will grow in proportion to **n**?

logarithmic **linear** **n log n** **quadratic** **cubic** **exponential**

logarithmic. The path length from root to any leaf is log n, so log n from any leaf to any other leaf.

(m) Assume that there are at most 8 billion people on earth.

(i) Approximately how many Ethernet addresses could each person have? A simple expression is fine; don't do arithmetic.

2^48 / 2^33 = 2^15. 8 billion is about 2^33. Almost everyone wrote something correct for this pair.

(ii) Approximately how many IPv6 addresses could each person have? A simple expression is fine; again, no arithmetic is needed.

2^128 / 2^33 = 2^95

(n) UTF-8 is a variable-length encoding of Unicode that represents ASCII characters as one byte and other characters as two or more bytes. It uses one or more of the leftmost bits of each byte to indicate the length of the character: 1-byte characters are represented as 0xxxxxxx, and 2-byte characters are represented as 110xxxxx 10xxxxxx, where each x is either 0 or 1. (This encoding saves space if the text is mostly ASCII.)

(i) How many one-byte characters can this format encode? (An expression is ok.)

128 (2^7). Quite well done on average.

(ii) How many two-byte characters can this format encode? (An expression is ok.)

2048 (2^11)

(o) For each of the following "facts" about cryptographic systems, circle the correct answer.

Trent can have two different RSA public keys with the same private key	**true**	**false**
The RSA algorithm resulted from a worldwide competition run by NIST	**true**	**false**
RSA is based on the difficulty of multiplying large prime numbers	**true**	**false**
If Alice loses her private key, a certificate authority can regenerate it for her	**true**	**false**
Eve can easily find two different documents with the same SHA-3 hash	**true**	**false**

(p) Parity involves adding extra bits to help with error detection and perhaps correction.

(i) Write out a 4-bit binary encoding of the numbers 0 through 7 where the leftmost bit is a parity bit chosen so that each number has odd parity.

`1000 0001 0010 1011 0100 1101 1110 0111`

(ii) Give one example of a 4-bit binary pattern that your encoding will detect as an error.

`0000` or any other 4-bit pattern not in the list above.

(q) An IPv4 address is a 32-bit integer.

(i) If the IPv4 address **`15.127.255.254`** is stored in a 32-bit integer variable **`v`** and incremented by the statement **`v = v+2`**, what is the resulting value, also expressed in dotted decimal notation?

`15.128.0.0`. Did we do something like this in the Q/A? Not as well done as I had hoped.

(ii) What is that resulting value expressed in hexadecimal?

`0F800000`

(r) The *NY Times* described a system that permits parents to monitor teenagers who are driving the family car, by visiting a web site that displays the current car location on a map. Assuming that the system is implemented in the most technically feasible and sensible way, assess the likelihood of the following statements:

The position of the car could be monitored by satellite imaging	**likely**	**unlikely**

A GPS receiver in the car could broadcast the car's location to a GPS satellite	**likely**	**unlikely**
A GPS-enabled cellphone in the car could report its location to a cellphone base station	**likely**	**unlikely**
A cellphone can only report its location when a conversation is in progress	**likely**	**unlikely**
RFID would be a viable alternative to a cellphone-based location system	**likely**	**unlikely**

(s) A guide to online privacy says "That [single] cookie, (1) placed on your computer when you visit most websites, 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 probably wrong?

(1)	**accurate**	**wrong**
(2)	**accurate**	**wrong**
(3)	**accurate**	**wrong**
(4)	**accurate**	**wrong**
(5)	**accurate**	**wrong**

My picture is that a cookie might be placed on your computer, but that cookie almost surely contains only some id number so you can be identified upon return. With very high probability, the cookie does not contain any of the other information in assertions (2) through (5).

(t) Quickies (a word or two).

"Everybody knows that APIs are not copyrightable," said Supreme Court Justice Sonya Sotomayor '76, during oral arguments in the case _____**Oracle**_____ vs _____**Google**_____.

In November 2020, the Federal Trade Commission accused _____**Zoom**_____ of lying when it claimed to provide end-to-end encryption for its services.

"Senator, we run _____**ads**_____," said Mark Zuckerberg to Orrin Hatch at a Senate hearing in 2018.

"The _____**Enigma**_____, which looks like a typewriter, consisted of a keyboard and wheels which scrambled messages." (*Reuters*, 12/4/20)

"Files stored in the _____**cache**_____ are typically reused across multiple sites instead of having the browser re-download each file for every page/tab load." (News story about the latest version of Chrome, 12/5/20.)

(u) [10 pts] True/false questions. Circle the best answer.

In WW2, Navajo code talkers relied on security by obscurity	**true**	**false**
An optical telegraph and a wireless Ethernet are both vulnerable to eavesdropping	**true**	**false**
The first trans-Atlantic fiber optic cable was laid in the 1860s	**true**	**false**
A tech company must have one profitable year before it can file for an IPO	**true**	**false**
Alan Turing was knighted for his wartime work on decoding the Enigma machine	**true**	**false**

It is possible for there to be more web pages than IPv4 addresses	**true**	**false**
Canvas fingerprinting is a biometric authentication technology	**true**	**false**
A two-factor device is used to do prime testing for the RSA algorithm	**true**	**false**
“Cloud computing” refers to carrying Internet traffic by satellite	**true**	**false**
DNS requires that the server for the web address www.yahoo.fr be located in France	**true**	**false**