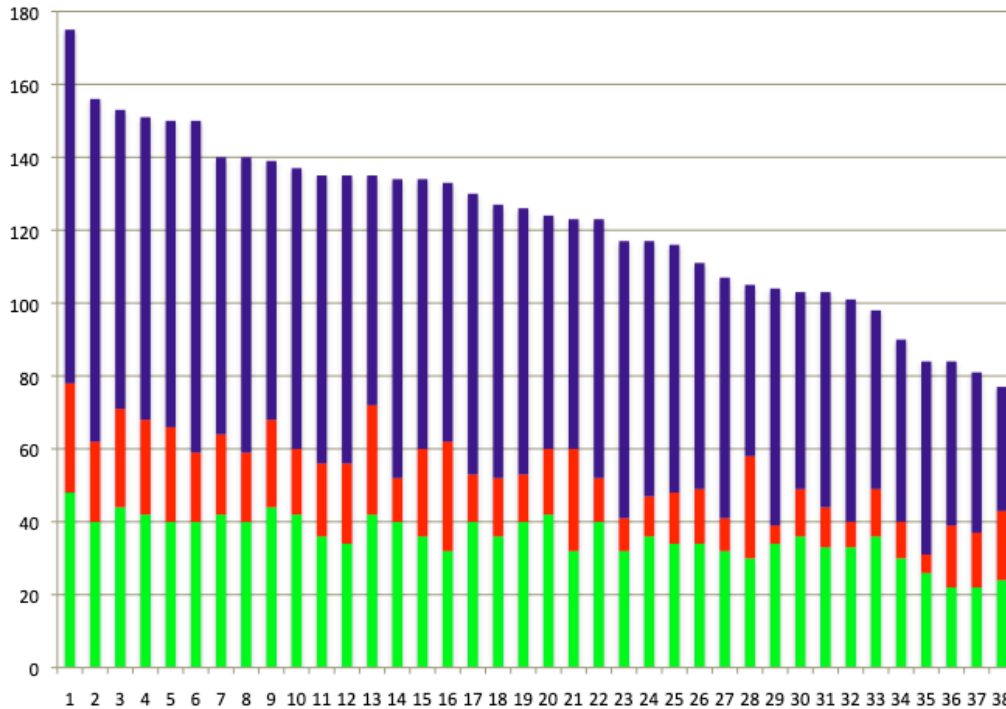


## COS 109 Final Exam, Fall 2022

I graded this myself. Maybe it's just my imagination, but did this class do better than last year's group, or maybe the exam was easier? In any case, the median was 125 and the quartiles were 138 and 103. For comparison, last year's median was 112, and the quartiles were 137 and 89; for Fall 2020, the numbers were 120, 132, 101. 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) If  $m$  and  $n$  are positive integers, how many 1-bits (that is, bits whose value is 1) are there in the binary representation of  $2^{m+n} - 1$ ?

**$m+n$ .** Powers of 2 have a single 1-bit; a number that is 1 less than a power of two is all ones. Pretty well done, perhaps because it was covered thoroughly in the Q/A.

- (b) The source code for the Linux kernel is about 1.3 GB of text files in C and the like, uncompressed. Which of these techniques would be most suitable for compressing it for faster downloading?

**GIF      JPG      Lempel-Ziv      MP3      MPEG      PNG      Zipf**

**Lempel-Ziv.** The others (except Zipf, a distractor) are for compressing images, not text. Also well done.

- (c) If Alice and Bob are discussing the significance of Etaoin Shrdlu, which of these is the most likely topic of their conversation?

**compression      error detection      machine translation      passwords      programming languages**

**Compression.** We talked about this in the context of letter frequencies in English, observing that one could compress text by using fewer bits for more common letters, which these are.

(d) Which one of these would be the most appropriate name for someone working at a Certificate Authority?

**Alice      Bob      Carol      David      Eve      Mallory      Trent**

**Trent, the trusted third party.**

(e) The NSA’s TAO (Tailored Access Operations) program surreptitiously “installs beacon devices into our targets’ electronic devices” before they are shipped to overseas purchasers, to enable network monitoring and other effects. What kind of attack is this an example of?

**DoS      IoT      MITM      MS-DOS      spear-phishing      Trojan horse**

**Man in the middle.** The devices are installed as the electronics passes from manufacturer to purchaser. Not a Trojan Horse, where the victims consciously take in an attractive offering.

(f) Braille represents each letter, number or other symbol as 3 rows and 2 columns of dots that are individually raised or not raised, like this:



How many different characters can there be in Braille?

**64 (or 63, if one excludes a pattern with no raised dots). 2^6. Mostly done right.**

(g) On Dec 11, 2022, ChatGPT reported that “We’re experiencing exceptionally high demand. Please hang tight as we work on scaling our system,” and was otherwise non-responsive. What kind of attack is this an example of?

**DoS      IoT      MITM      MS-DOS      spear-phishing      Trojan horse**

**DoS, a denial of service.** Their servers couldn’t keep up with the load.

(h) In *The Innovators*, Walter Isaacson quotes \_\_\_\_\_ as saying “It is possible to invent a single machine which can be used to compute any computable sequence.” Whose name belongs in the blank?

**Alan Turing.** Most people recognized this one.

(i) “These digital tokens use \_\_\_\_\_ technology, in which computers contribute to a shared ledger that can be used to track digital assets.” What word or phrase belongs in the blank?

**Blockchain.** Well done.

(j) Name one example of the kind of a digital asset mentioned in the previous question (part (i)).

**Bitcoin, Ethereum, NFT, ...** Also well done.

(k) Ransomware encrypts all the files on a victim’s computer, and the victim has to pay the bad guys for a password that will decrypt them. If you were (just hypothetically) a bad guy, which of these mechanisms would be most suitable for implementing your ransomware attack?

**AES      DES      MD5      Rot13      RSA      SHA-512      Tor**

**AES.** You need an essentially unbreakable encryption algorithm that runs fast enough (so not RSA). And Tor hasn’t anything to do with this.

(l) Circle all of the following that must be kept secret from an adversary to protect a message encrypted with AES:

**AES algorithm      AES key      AES source code      compiler      compiled AES code**

**AES key only.** The algorithm is public, as are versions of the source code (including the one that was passed around in class).

- (m) In November 2022, the Supreme Court agreed to hear the case *Jack Daniel's Properties v. VIP Products*. VIP sells dog chew toys shaped like a Jack Daniel's whiskey bottle with the label "Bad Spaniel's." The toy includes dog-related changes to the original label, like replacing alcohol content with "43% Poo by Vol." and replacing "Old No. 7" with "Old No. 2." What specific kind of intellectual property does this case deal with?

**Trademark.** Practically a textbook example.

- (n) What distinguished Princeton alum was recently named the university's new provost?

**Jeff Bezos '86      Christopher Eisgruber '83      Robert Kahn \*64**  
**Jennifer Rexford '91      Eric Schmidt '76      Peter Wendell '72**

**Jennifer Rexford.** I was glad to see that everyone got this Christmas gift.

And says Jen, "LoL. Just showed that to my mom, and she is still laughing!"

- (o) The US national debt is about 31 trillion dollars in December of 2022. How many *bytes* are needed to store this value in binary?

**6.**  $31 \times 10^{12}$  is about  $2^5 \times 2^{40}$ , or 45 bits. Round up. Most people got it.

- (p) Suppose that over the 12 days of Christmas I plan to give the following presents. Circle all those that *do not* use radio.

**Apple airpods      barcode reader      GPS receiver      prox card      smart watch      US passport**

**Barcode reader,** which uses a red laser.

- (q) Put these names into chronological order of when they made the contribution(s) that caused them to be mentioned in COS 109, by writing the numbers 1 through 5 on them.

**Tony Hoare      John von Neumann      Guido van Rossum      Bjarne Stroustrup      Ken Thompson**  
**Von Neumann, Hoare, Thompson, Stroustrup, van Rossum**

- (r) RSA-250, the largest currently solved factoring challenge number, has 250 decimal digits and is 829 bits long. About how many bits long would you expect RSA-500 to be?

**1658,** plus or minus a couple: basically twice as long.

- (s) Knowing that binary search doesn't work on unsorted data, a zealous programmer modifies a binary search function by adding statements at the beginning to verify that the data values being searched are in order. What is the likely running time of the revised function, expressed in terms of  $n$ , the number of data items?

**$\log n$        $n$        $n \log n$        $n^2$        $2^n$**

**$n$ .** The key word is *verify*. You can check for correct order just by going through the numbers. The most popular wrong answers were  $\log n$  (deceived by "binary search"?) and  $n \log n$  (deceived by "unsorted"?).

- (t) "Though \_\_\_\_\_ software is designed to be shared freely among coders and companies, this sharing is governed by licenses designed to ensure that it is used in ways to benefit the wider community of programmers." What word or phrase belongs in the blank?

**Open source.** "free" isn't specific enough.

- (u) We counted the number of people in the classroom one day by an algorithm that paired standing students; one

of each pair then sat down while the other retained the combined counts. Suppose that we want to do the same thing in a crowd of a million people, but instead of pairs, at each stage people will combine in groups of 4 at a time. If done perfectly, about how many stages will this procedure require to produce a final count?

**10. It's the log base 4 of 1 million, or equivalently, the number of times you can divide 1 million by four.**

- (v) Suppose that one afternoon during an interminable COS 109 lecture you use Safari to visit Amazon, Facebook and Google from your laptop, and Amazon sends you a cookie. Which of the following sites will receive that cookie later that day when you visit Amazon from your phone?

**only Amazon      Amazon+Facebook      Amazon+Google      all three of them      none of them**

**None.** The cookies are on the laptop, and not on the phone.

- (w) In the *Jeopardy* category "Let's Get I.T. On," the clue was "This programming language isn't a little worse than B minus; it's a 1980s improvement of a language called C." What is the language?

**C++.** Most people got this freebie.

- (x) "Companies make cordless mice that use radio signals instead of Bluetooth to communicate with the computer." (*NY Times*, 7/23/05) At best this is sloppy; at worst it's actively misleading. In no more than 5 or 6 words, identify and correct the imprecise statement or technical misinformation conveyed by this quote.

**Bluetooth is radio.**

- (y) If I were a high-school guidance counselor who wanted to use a computer to generate plausible college recommendation letters for my students, which one of these AI techniques or systems would be the most directly useful?

**ChatGPT      DALL-E2      deep learning      neural network      recommendation engine  
reinforcement learning      supervised learning      unsupervised learning**

**ChatGPT.** Another freebie.

## 2. (30 points) Understanding Programs

- (a) [6] The Python function `pow(m, n)` is supposed to raise a positive integer `m` to a positive integer power `n` by repeated multiplication; that is, it computes  $m^n$ . For example, `pow(2, 3)` should return 8 and `pow(3, 2)` should return 9. Unfortunately, this version has three small errors. Fix the errors: either rewrite `pow` or state clearly what the errors are and how you are fixing them. (This is a question about correct logic; don't worry about syntactic trivia, but make your corrected code clear.)

```
def pow(m, n):
    p = 1
    i = 1
    while i < n:
        p = p + m
        i = i + 1
    return m

def pow(m, n):
    p = 1
    i = 1
    while i <= n: # test has to be <= (or set i=0)
        p = p * m # multiplication!
        i = i + 1
    return p # the product that has been accumulated
```

There are other equivalent versions, for example run the loop from `i=0` to `i < n`. But you can't start with `p=m`, since then it won't work for `n=0`.

(b) [2] Once it has been properly fixed, how does the running time of this algorithm depend on  $m$ ?  
 $\log m$      $m$      $m \log m$      $m^2$      $m^3$      $2^m$     independent of  $m$   
**independent of  $m$ .**

(c) [2] Once it has been properly fixed, how does the running time of this algorithm depend on  $n$ ?  
 $\log n$      $n$      $n \log n$      $n^2$      $n^3$      $2^n$     independent of  $n$   
 **$n$ . The loop goes around  $n$  times.**

(d) [1] Once fixed, does this function work properly when  $m$  is not an integer? **Yes** or **No**  
**yes**

(e) [1] Once fixed, does this function work properly when  $n$  is not an integer? **Yes** or **No**  
**no**

(f) [4] Here's a Python function, with various parts identified by line numbers; the line numbers are not part of the function.

```
1:     def AbsoluteValue(v):
2:         if v >= 0:
3:             return v
4:         else:
5:             return -v
```

(i) Which part is the API for this function? Identify the line or lines by number.

**Line 1**

(ii) Which part is the implementation? Identify the line or lines by number.

**Lines 2-5**

(g) [6] The Python function `random.randint(1,100)` produces an endless sequence of random integers between 1 and 100 inclusive; any number is as likely as any other, so over a long enough period, any number will occur about as often as any other number. If the following Python loop is executed, approximately how many lines of each type of output would you expect to see?

```
for i in range(0,10000):
    num = random.randint(1,100)
    if num > 60:
        print("big")
    elif num > 30:
        print("middling")
    elif num > 5:
        print("small")
```

**big:            4000**

**middling:      3000**

**small:          2500**

**Each test peels off some portion of the numbers; only 9500 are printed in total.**

(h) [6] Suppose that the Toy machine version 2.0 has a new instruction **rshift N**, which shifts the contents of the accumulator **N** bit positions to the right (and discarding the bits that “fall off the end”). What does this program print when given the number **37** as input?

```

      GET           get a number from user, place it in accumulator
TOP   PRINT        print content of accumulator
      IFZERO  DONE  if accumulator content is zero, go to location DONE
      RSHIFT  1     shift accumulator content one bit to the right
      GOTO    TOP   go to instruction labeled TOP
DONE  STOP        stop execution

```

37 18 9 4 2 1 0

It’s just dividing by two each time around the loop. The output includes the initial value and the final zero. People had more trouble with this one than I expected.

(i) [2] How does the running time of this program depend on the size of the input number **N** that it is given?

logarithmic    linear     $N \log N$     quadratic    cubic    exponential    independent of **N**

logarithmic. It’s dividing by two each time around the loop!!

### 3. (100 points, 5 each) Miscellaneous

(a) In November 2022, the SI units were updated by adding names for the new biggest and smallest units: zetta and yotta are now followed by **ronna** and **quetta** as the largest numbers, and there are now matching **ronto** and **quecto** as the smallest.

(i) How many quectograms are there in a quettagram, expressed as a power of ten?

10<sup>60</sup>

(ii) What power of two is closest to this number, the number of quectograms in a quettagram?

2<sup>200</sup>

(b) *Base64 encoding* is a technique that represents arbitrary binary data in a printable form. It’s similar to hexadecimal, but uses 6-bit chunks instead of 4: each possible 6-bit input combination is encoded with a unique 8-bit ASCII letter (a-z, A-Z) or digit (0-9) or other character (+, /) so the result of the encoding is a sequence of ASCII characters that is longer than the input sequence of bits.

(i) If an IPv4 packet is 3000 bytes long, how long is the Base64 encoding of the packet?

4000 bytes. Every 6 bits becomes 8.

(ii) If instead of Base64, we write the packet contents in hexadecimal, how many hex digits would it take to write out the packet contents?

6000. Each byte takes two hex digits.

(c) When I create an online grocery order at Shoprite, every time I add an item to my cart, nearly two dozen trackers try to monitor me. Fortunately my defenses block them all (or so I believe).

(i) What programming language are the trackers most likely to be written in?

JavaScript

(ii) Name two tools that you or I could use to block explicit trackers.

Ghostery, NoScript, ... Not as well posed as I thought, so graded generously.

- (d) “\_\_\_\_\_ stinks!”, says a top secret PowerPoint slide produced by \_\_\_\_\_ and revealed to the world in 2013 by \_\_\_\_\_. Fill in the blanks with the appropriate names.

**Tor, NSA, Snowden**

- (e) “Morse decided to puzzle his brain no more on how 23 in base 10 could be expressed in base 5.” (From the 1976 Inspector Morse novel *Last Seen Wearing*, by Colin Dexter.)

(i) What is 23 base 10 expressed in base 5?

**43.** (It’s surprising that someone as smart as Morse couldn’t figure this out.)

(ii) What is 23 base 10 expressed in base 2?

**10111**

(iii) What is 23 base 10 expressed in base 16?

**17**

- (f) Princeton’s new Stadium Drive parking garage has spaces for 1,560 cars. Suppose that a license-plate reader records the plate number as text, the arrival time and the departure time for each car.

(i) Estimate approximately how many bytes would be needed to store the plate number, arrival time, and departure time for any given car, reasonably compactly.

**15 bytes?** Maybe 6-7 for the plate and 4 each for time in and time out.

(ii) Estimate approximately how many megabytes of disk space would be needed to store all this data for one year, making sensible assumptions about the amount of traffic in and out of the garage.

**4-5 MB?** 15 B/car \* 1000 cars/day \* 300 days/year. Some residual confusions about bits and bytes.

The question wasn’t very clearly posed, so graded generously.

- (g) Joe College has 1,000 files on his computer, of which 50 are correctly labeled Word .docx files and 25 are correctly labeled Excel .xlsx files.

(i) How many times does Joe have to run Word to compute the total number of bytes in all of those .docx and .xlsx files?

**Zero.** All the information about sizes is in the directory, as beaten to death in problem sets and previous exams.

(ii) How many times does he have to run Word and Excel to determine whether the largest Word file is larger than the largest Excel file?

Word \_\_\_\_\_ Excel \_\_\_\_\_

**Zero.** The question does say “correctly labeled”, so there is no cheating in the names.

- (h) *Supreme Conflict*, a 2008 book on the Supreme Court, describes how before each session each of the nine justices shakes hands with each of the others.

(i) What is the total number of handshakes?

**36 (or 72).** Not 81; people rarely shake hands with themselves.

(ii) If there are N people doing handshakes, how does the total number of handshakes grow in proportion to N?

**N^2.** Both parts of this question were covered in the Q/A, and at other times.

- (i) Netflix still distributes physical DVDs, but the number of users has been shrinking steadily, from about 16 million in 2010 to about 1 million today in 2022.

(i) If this decline has been a smooth exponential process, what is the percentage rate of decrease *per month*?

**2%.** Rule of 72: The number was cut in half 4 times in 12 years, so 3 years == 36 months to cut in half once, so  $72/36 = 2\%/month$

(ii) If the decline continues at the same rate in the future, in what year will there be only 1,000 users left? (Clearly this is a very over-simplified model, so excess precision is not appropriate.)

**2052.** It takes 10 halvings to get from 1 M to 1000; each takes 3 years, so 30 years. I was really hoping to see people use the Rule of 72, since it's a good approximation when the numbers are clearly approximate anyway. Dragging out the calculator to use its exponent and log functions is overkill and leads to specious precision.

- (j) In December 2022, TSMC announced plans to build two new integrated circuit fabrication plants in Arizona, at a cost of \$40B! The first plant was going to use a 6 nm technology but will now use 4 nm; the second plant will open later and use 3 nm technology. (I have simplified the numbers a bit.)

(i) If a 12-inch wafer using 6 nm technology has 200 chips, how many would it have with the 4 nm technology, if nothing else changes?

**450.** The ratio of linear dimension (the "technology") is  $6/4$  or 1.5, so the area ratio is  $1.5^2$ , or 2.25.

(ii) If a 12-inch wafer using the 6 nm technology has 200 chips, how many would it have with the 3 nm technology, if nothing else changes?

**800.** Same reasoning, with a linear ratio of 2, so a factor of 4. Neither part was very well done, though we have talked about area problems repeatedly and they showed up frequently on problem sets and previous exams.

- (k) A network address translator (NAT) maps internal IP addresses from a specified range into a single external IP address and vice versa.

(i) A home wi-fi router acts as a NAT. It usually uses the IPv4 address range 192.168.0.0 to 192.168.255.255 internally, assigning a unique IP address in this range to each device in the home. In principle, how many devices could this range support? Express it as a power of 2.

**$2^{16}$ .** There are 16 bits in the range.

(ii) A large corporation might use a similar mechanism but with the internal IPv4 address range 10.0.0.0 to 10.255.255.255. How many devices would this address range support? Express it as a power of 2.

**$2^{24}$ .**

- (l) The late John Lions, author of *A Commentary on the Unix Operating System*, once said of the source code listing of 6th Edition Unix, "The whole documentation is not unreasonably transportable in a student's briefcase." For each of the following, would it be reasonably transportable in your backpack?

A hundred terabytes of laptop SSD disks	<b><u>yes</u></b>	<b>no</b>
A terabyte of magnetic core memory from the 1960s	<b>yes</b>	<b><u>no</u></b>
A listing on paper of a C implementation of the AES algorithm	<b><u>yes</u></b>	<b>no</b>
A listing on paper of a C implementation of the TCP/IP protocol	<b><u>yes</u></b>	<b>no</b>
10 kilometers of bare fiber optic cable	<b><u>yes</u></b>	<b>no</b>

All of these were passed around in class in some form. The SSD in your laptop is the same size as the memory chips that were passed around; a hundred of them wouldn't be very much at all. The core memory was only



2KB, so half a billion of them would be a very large pile indeed. I passed around the AES listing, which is only 3 or 4 pages long, and the fat book with all of TCP/IP is no bigger than most books. The fiber optic spool was 9,000 meters.

- (m) An article about the NYC marathon says “Every competitor will wear a shoe with a chip that will record their progress, and can send e-mail updates every five kilometers to spectators who subscribe to the service.” For each of the following inferences that a non-technical reader might make from this quotation, assess whether they are likely to be correct or likely to be incorrect. (A marathon is about 26 miles or 45 km long.)

the chip has enough memory to store at least a dozen time measurements	<b>correct</b>	<u>incorrect</u>
the chip uniquely identifies the runner who wears it	<u>correct</u>	<b>incorrect</b>
the chip uses GPS to determine how far the runner has run so far	<b>correct</b>	<u>incorrect</u>
the chip sends e-mail messages to a server	<b>correct</b>	<u>incorrect</u>
the chip’s memory determines the maximum number of e-mail subscribers	<b>correct</b>	<u>incorrect</u>

The chip is just like your prox; it doesn’t do anything except broadcast its number when near a sensor.

- (n) The hex value **00FFFF** can be interpreted as an RGB color. Suppose that instead this value is interpreted simply as a 24-bit integer, stored in a variable **v**, and incremented by 1 with the Python statement **v = v+1**.  
 (i) What is the resulting value of **v** in hexadecimal?

**010000**.

- (ii) What color is the resulting value closest to?

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

**black.** We talked a couple of times about how minimal values of colors were basically black.

- (o) If I use my phone to send mail to a friend in England, as the mail goes from me to his laptop, which of these mechanisms (A) is almost sure to be used? (B) might be used but need not be? (C) is very unlikely to be used? Circle the best answer.

TCP/IP	<u>almost sure</u>	<b>might be</b>	<b>very unlikely</b>
Ping	<b>almost sure</b>	<b>might be</b>	<u>very unlikely</u>
Fiber-optic cable	<u>almost sure</u>	<b>might be</b>	<b>very unlikely</b>
NAT	<b>almost sure</b>	<u>might be</u>	<b>very unlikely</b>
Wi-fi base station	<u>almost sure</u>	<b>might be</b>	<b>very unlikely</b>

Ping is just a diagnostic tool. I think that wi-fi would be used for a laptop at the other end with high probability, but I accepted “might be” as well, since it’s not certain.

- (p) A *NY Times* article about E-ZPass, the electronic highway toll system, says, “A list of valid and invalid tag numbers is sent every day to computer drives in every toll booth. As a vehicle drives through an E-ZPass lane, a high-speed optical reader almost instantly identifies the tag mounted to a dashboard or windshield and matches it against the list to see if the holder has enough money set aside to pay the toll.” Identify three technical “facts” in this quotation that are almost surely wrong, or at least badly misleading. **Be brief**— a few well-chosen words should be adequate for each.

No need to send both valid and invalid tags; valid alone is fine, and of course the set of invalid tags might not even be defined. I very much doubt that there are drives in toll booths; that would be a logistics nightmare,

costly, and unreliable. And every day? E-ZPass is a form of RFID, not optical. The comparison against the list would be done on a server, and it's not likely that they would check for enough money, just send a bill later. Bad reporting all round.

(q) Refer to the ASCII chart on the cover page of the exam.

(i) If `ch` is a variable that contains the ASCII character `[` (left bracket), how many bits must be changed to convert `ch` into the ASCII character `{` (left brace)?

**1 bit.** 5B and 7B, so you just have to compare 101 and 111.

(ii) If `ch` is a variable that contains an arbitrary ASCII character, explain in at most half a dozen words what this test is trying to determine. **DO NOT** just repeat the code in words.

```
if ch >= 65 and ch <= 90 ...
```


**Is `ch` an upper-case letter?** Convert 65 base 10 to 41 base 16 (which is A) and 90 base 10 to 5A base 16 (which is Z). Somewhere along the way realize that you're looking at upper-case letters.

(r) To coordinate their romantic activities, Alice and Bob naturally use public-key cryptography to exchange encrypted email. Suppose that Eve learns Alice's private key. What can Eve now do?

Eve can convince Bob that she (Eve) is really Alice	<u>true</u>	false
Eve can convince Mallory that she (Eve) is really Alice	<u>true</u>	false
Eve can convince Alice that she (Eve) is really Bob	true	<u>false</u>
Eve can read an encrypted message from Bob to Alice	<u>true</u>	false
Eve can read an encrypted message from Alice to Bob	true	<u>false</u>

If Eve learns Alice's private key, she is Alice for all things cryptographic. But she is not Bob.

(s) [10 pts] Random quickies: Circle the best answers.

The GDPR applies primarily to residents of Germany	true	<u>false</u>
Trans-oceanic Internet traffic is transmitted with communications satellites	true	<u>false</u>
The Turing machine predates von Neumann's <i>Johnniac</i> computer at IAS	<u>true</u>	false
"An IP address is like a zip code: it tells where your computer is located"	true	<u>false</u>
End-to-end encryptions prevents the NSA from knowing what your browser connects to	true	<u>false</u>
<code>/* You are not expected to understand this */</code> comes from Unix kernel source code	<u>true</u>	false
I would have to purchase the domain kernighan2024.com directly from ICANN	true	<u>false</u>
"The S&P500 returned 9% annually so your investment doubled in value in 8 years"	<u>true</u>	false
Alan Turing was the first recipient of the ACM Turing Award	true	<u>false</u>
The Cuneiform character zum  whose hex representation is 1236E, fits in 2 bytes	true	<u>false</u>