COS 109 Final Exam, Fall 2021    December 15-20, 2021

3 hours    180 points total

You may do this exam in any 3-hour period. When you are finished, return the exam to room 311 in the CS building, or scan it (or just the answers) and email it to bwk@princeton.edu. You must return it by 5:00pm EST on Monday December 20. I would appreciate it if you could tell me how you plan to return it so I know where and when to look, but it’s not a requirement.

Please PRINT your name here

Honor Pledge: “I pledge my honor that I have not violated the Honor Code during this examination.”

Please write the pledge in full and sign it:

This examination is open-book and open-note:
• you may use the textbook, course notes, your own notes, corrected problem sets and solutions, old exams and answer sheets from the course web page, lab instructions, etc.
• you may use a calculator.
• you may not use anything else; specifically, you may not use a computer, phone or tablet (except that you can use a calculator program on one of these, and you can use your computer to view course notes, answer sheets, etc.). No other Internet use is allowed.

There are 180 points for the questions; use the point values for each question to allocate your time (one point per minute). If you’re writing or calculating a lot on a question, you may be off on the wrong track.

Write your answers directly on these pages; use the back if necessary. In general, be brief, but if you need more space, attach extra pages and make sure your name is on every extra page. Please write legibly -- I can't grade it if I can't read it. [You can submit just the answers if that makes scanning easier.]

Good luck.

1. (50 pts)

2. (20 pts)

3. (110 pts)

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

(a) Alice wants to digitally sign a document and is wondering whether to use AES or RSA. How would you advise her?

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

(b) In October, the US Supreme Court finally decided the long-running dispute between Oracle and Google.

-- In whose favor did they decide? ______________________________

-- What was the basic issue at issue? (3-4 words is enough!) ____________________________

(c) How old is Cecilia, in base 10?

(d) 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

(e) In late 2020, Apple began selling computers with M1 processors that it designed itself and had fabricated by TSMC in Taiwan. Which of these companies would be most financially affected by Apple’s action?

- Amazon
- Facebook
- Google
- Intel
- Microsoft
- Netflix

(f) Add these two binary numbers:

```
0110101001.00111
1001010110.11001
-------------
```
(g) Suppose that I use SHA-1 to compute a cryptographic hash of a message that I am going to send to a friend. Then I change “bwk” to “BWK” in one place in the message and re-compute the hash. What is the relationship between the first cryptographic hash value and the second?

- the same
- 3 bits are different
- 3 bytes are different
- about half the bits differ
- every bit is different

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

Tim Berners-Lee  Bill Gates  Dennis Ritchie  Edward Snowden  Mark Zuckerberg

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

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

(j) What kind or level of programming language is being described in this excerpt from David Auerbach’s 2018 book *Bitwise: A Life in Code*? “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.”

(k) 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
- RAM
- ROM

(l) What ASCII character occurs most frequently in ordinary English text like a novel or a newspaper story? Hint: as we saw in class, it occurs nearly twice as often as the next most frequent character.

(m) Headline in early December: “Trump's social media site quietly admits it's based on Mastodon. Mastodon had threatened to sue Trump's 'Truth Social' site for allegedly violating its ________________ license.” What two- or three-word phrase belongs in the blank?

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

- FCC
- FTC
- ICANN
- ITU
- UNESCO
- USPTO
- WIPO
(o) Suppose that the Princeton admissions office wants to use machine learning to predict the GPA at graduation from Princeton of high-school students who have applied. Which one of these ML techniques would be the most suitable for predicting the ultimate GPA of current applicants? (This is not to suggest that such techniques would necessarily be useful or desirable.)

supervised learning  unsupervised learning  image recognition
clustering  recommendation system  sentiment analysis

(p) 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

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

AT&T  FCC  FTC  GCHQ  IETF  NIST  RTFM  WIPO  WTF

(r) 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?

(s) If I wanted to collect all the contents of all the disks on the laptops of the 31 students in COS 109 onto a single disk, which one of these is the smallest amount of space that would be sufficient?

1 MB  1 GB  1 TB  1 PB  1 EB  1 ZB  1 YB

(t) A few years ago, the web page for CS 152 at Harvard said “Almost every computation in the world (at a rough guess, 15,000,000,000,000,000,000 instructions every second) was written in some programming language.” If the computers doing these computations are all typical laptops like yours, *very roughly* how many computers would that be?
(u) The ACLU says “Stingrays […] mimic cell phone towers and send out signals to trick cell phones in the area into transmitting their locations and identifying information” rather than sending it to base stations. What kind of attack is this most like?

- man in the middle
- ransomware
- spear-phishing
- Trojan horse
- war driving

(v) Modern computers can efficiently process integers of several sizes, usually 1, 2, 4, 8, and sometimes 16 bytes long. Which of these is the least number of bytes that could be used for storing a binary number representing the current population of New Jersey?

- 1
- 2
- 4
- 8
- 16
- none are big enough

(w) OIT notice, 12/7/21: “The outage that’s currently impacting AWS’s US-EAST-1 region is having a significant impact on the performance of Blackboard and Canvas as well as third-party tools that are integrated with Canvas.” What service is AWS providing to Blackboard, Canvas, et al? A couple of words is enough.

(x) Suppose that Google upgrades the cameras that it uses for Street View from 12 megapixels to 30 megapixels. If Google uses 100 PB to store its existing images for the USA, how much space will it need to store the new images?

(y) 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?
2. (20 points) Understanding Programs

(a) The following Python code is supposed to simulate flipping a fair coin \textit{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 \texttt{random.random()} produces a new random floating-point value between 0 and 1. The \texttt{print} statement is syntactically correct as well.)

\begin{verbatim}
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
\end{verbatim}

(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?
(c) Suppose that version 2.0 of the Toy machine includes an instruction called \texttt{REM}, 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 \texttt{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?

\begin{verbatim}
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     2                divide accumulator value by 2, 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
\end{verbatim}

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

(e) How does the running time grow as a function of or in proportion to \( N \), the number of input numbers?
3. (110 points, 5 each) Miscellaneous

(a) An April 2021 article in the *London Review of Books* says “a ________________ is a small sequence of letters and numbers that a website generates and deposits in your browser.”

(i) What word belongs in the blank?

(ii) The article goes on to say “a pixel is a tiny, transparent image, and you can't see it on your screen. […] Why would someone want to show you an invisible image? What does the pixel do?” Very briefly but clearly, answer the author’s question.

(b) Some years ago, a COS109 student told me that the instructions for Lab 3 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.

(c) The number 401b30e3b8b5d629635a5c613c68791e has 32 hexadecimal digits.

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

<table>
<thead>
<tr>
<th>AES key</th>
<th>Ethernet address</th>
<th>IPv4 address</th>
<th>IPv6 address</th>
<th>MD5 hash</th>
<th>SHA-1 hash</th>
</tr>
</thead>
</table>

(ii) Briefly explain why this is not a prime number.
(d) On 12/3/14, YouTube described an overflow problem in their programming: “We never thought that a video
would be watched in numbers greater than a 32-bit integer (2,147,483,647 views)”.

(i) If the value 2,147,483,647 is stored as a 32-bit integer, what is its leftmost bit: 0 or 1?

(ii) What is its value in hexadecimal?

(iii) When this number is incremented by 1, what is the resulting value in hexadecimal?

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

BA0BAB COFFEE DECODE EFFACE FACADE F00D1E OFF1CE

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

(ii) Which one has the largest numeric value?

(iii) If instead they are interpreted at 24-bit RGB colors, which one has the least amount of green?

(f) A news story says that there are probably about 100 billion planets in the Milky Way.

(i) If astronomers want to give each planet a unique number, using the smallest possible number of bits, how
many bits would that number have?

(ii) How many bytes would such numbers occupy?
(g) Princeton logs all your Internet connections, including source IP address, the IP address you visit, your Ethernet address, and the Unix standard time (the number of seconds since 1970) at the beginning of the connection and at the end of it.

(i) If IPv4 addresses are used, how many bytes would be required to store this information for one connection, in the most straightforward and conventional representation?

(ii) If IPv6 were used instead of IPv4, how many bytes would be required?

(h) 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 _________________  IPv6 _________________

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

(i) Morse code uses combinations of one to five dots and/or dashes to represent letters, digits, and punctuation marks. For example, E is a single dot (·), A is dot-dash (· -), and Q is dash-dot-dot-dot-dot (--- · · · ·). Suppose you are designing a new version of a Morse-like code, in which every character will consist of some combination of exactly 6 dots and/or dashes. Describe briefly how you would systematically assign upper case letters and digits to combinations of 6 dots and dashes. Write down enough of your characters, or explain how you would create them, so clearly that there is no ambiguity about your design.
(j) 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

(k) This partial Unix directory listing shows size, 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
```

(l) 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 US phone reports its number and its position to an accuracy of one yard or meter 1,000 times/day. *Very roughly* how many terabytes of tracking information are uploaded by all these phones every day in total? Be precise about your assumptions about how information is represented.
(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 prototype binary machine that would handle numeric values up to at least one million (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 million?

(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 million?

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

(n) A deep-space communications system reports on the health of a piece of equipment by sending a continuous stream of status reports at 1 bit per second. 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.
(o) 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 December 2018, a new Mersenne prime was discovered, the largest known so far: $2^{82,589,933} - 1$. It has 24,862,048 digits in its decimal representation.

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

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

(p) McKinsey reports (11/13/21) that the value of the autonomous vehicles sector went from $10$ billion in 2011 to $320$ billion in 2021. Assume (unrealistically) that this improvement was a smooth exponential process.

(i) What was the growth rate of value per month during this time?

(ii) If the sector continues this exponential rate of progress, in what year will the value be $10$ trillion?

(q) A Slashdot story last year said “There are only 4 billion floating-point numbers, so you can test them all.” The story didn’t say what kind of test, but suppose it is simply to add each possible pair of numbers and see if each sum is correct.

(i) How many such tests would there be? An expression as an integer power of 2 is all that’s needed.

(ii) Double-precision floating-point numbers are 8 bytes long. How many tests would be needed to do the same tests for all double-precision numbers? Again, a simple expression is all that’s needed.
(r) Supercomputers with lots of processors are often 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?

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

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

(s) A \textit{NY Times} headline (1/16/13) says “City police plan to put GPS devices in pill bottles to find drugstore thieves.” Naturally the story is light on technical details but it does say that “when a decoy bottle is lifted from a special base it begins to emit a tracking signal.” Assess whether these statements derived from the article are likely to be correct, or are unlikely to be correct.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Likely</th>
<th>Unlikely</th>
</tr>
</thead>
<tbody>
<tr>
<td>The locations of stolen pill bottles could be monitored by GPS satellites</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A GPS device in a decoy bottle could send its location to a GPS satellite</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The device could receive enough power from GPS satellites to send a signal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A GPS device in a decoy could send its location to cellphone base stations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spectrum space for the devices would be allocated by the FDA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
(i) 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 $N$?

(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 $N$?

(u) Random quickies.

<table>
<thead>
<tr>
<th>Question</th>
<th>True</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bitcoin can’t be used to pay off ransomware demands because it’s anonymous</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>Your laptop’s Ethernet address changes as you walk from Friend 008 to the Dinky station</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>A single parity bit can correct a single-bit error in a single byte</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>Bitcoin is an example of a cryptocurrency with a relatively stable dollar value</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>Rot26 encryption is easier to decrypt than Rot13 is</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>A lossless compression algorithm will make some inputs larger, not smaller</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>Most DNS queries will access a root server</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>Every DNS query accesses a registrar</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>If you use HTTPS to access a web site, your ISP does not know which site it is</td>
<td>true</td>
<td>false</td>
</tr>
<tr>
<td>If you use Tor to access a web site, your ISP does not know which site it is</td>
<td>true</td>
<td>false</td>
</tr>
</tbody>
</table>