Final Exam
COS 116 Spring 2012: The Computational Universe

Name: ANSWER KEY

Email: @princeton.edu

This is an in-class exam. No collaboration is allowed. There are 8 questions on 7 pages. You may use the back sides of the pages if you run out of room on the front.

Write and sign the honor pledge:

"I pledge my honor that I have not violated the Honor Code during this examination."

1. Hardness of different problems: (2pts each part)

Arrange the following problems in order of increasing difficulty for computers to solve, from 1=easiest to 8=hardest. For the first, describe how to solve it. For each of the the rest, say whether or not you are certain if it is harder than the previous one, and explain why. (For example, for #2 explain whether it is harder than #1 and if so why; and so on for higher numbers.)

Problems:
(A) I give you the whole Facebook friend network for Princeton University students, and also a "clique" of 10 Princeton students for which I claim that each one of the 10 is a friend the other 9. Verify that these 10 do indeed form a clique.

(B) I give you a boolean expression that looks like this:

$(A + B + C) \cdot (\sim D + F + G) \cdot (\sim A + G + \sim L) \cdot (\sim B + P + Z) \cdot$ ...etc

with 26 different variables (the letters of the alphabet) and where "~A" mean "*not* A", for example. Tell me if there is a variable assignment that causes this expression to be true.

(C) I give you a English language message represented as bits, encoded with a one-time pad based on random bits, but I do not give you the one-time pad. Figure out the message anyway, using statistics of English letter and word frequencies.

(D) I give you an arbitrary program and an arbitrary input. Tell me whether the program will execute the STOP instruction when running on that input.

(E) Look through the employee records at Princeton University and report the salary of the highest paid employee.

(F) Look through the employee records at Princeton University and print out a list, ordered by social security number.

(G) I give you a map of the 400 biggest cities in the USA, showing the distances between them. Tell me the shortest route that visits each city exactly once.

(H) I tell P and Q are two very large prime numbers but do not tell you what they are. Instead I tell you R, which is equal to the product (P x Q). Figure out P and Q.

#1 (easiest):
A: Can be verified by checking 9 friends for each of the 10, so 90 steps total.

#2:
E: Can take one sweep over N employees (say N=a couple thousand) to find the highest salary.

#3:
F: Need to sort N employees. In class we discussed merge sort taking N^2 steps. While there are somewhat faster approaches they are still slower than N steps (problem E).

#4:
B: We do not know any faster way than to try exponential (2^N) combinations and while this is slow for N=26 we could contemplate doing it (something like 64 million steps). This problem is NP complete but the input is not too large.

#5:
H: We do not know how to factor products of very large prime numbers without trying all smaller primes up to the square root, which requires figuring out all those smaller primes. It is possible someone could figure out a way to do it but our current system of e-commerce assumes this is very difficult (implying that it is a safe bet).

#6:
G: This is the traveling salesperson problem (TSP), which is a classic NP-complete problem. For N=400 we cannot realistically contemplate solving it.

#7:
C: This problem is impossible to solve if the one-time pad is not known.

#8 (hardest):
D: This is the halting problem, which is impossible to solve.

2. Computer architecture and communication: (3pts each part)
Briefly explain (two or three sentences) each of the following concepts:

(a) memory hierarchy
Method to accelerate memory access by putting most frequently used information in faster memory (e.g. cache) and least frequently used information in slowest memory (e.g. disk). There can be many layers (e.g. L1 cache, L2, etc).

(b) wave division multiplexing
Way to squeeze many signals simultaneously through a single fiber optic cable by splitting into different wavelengths of light (and thereby increase bandwidth).

(c) instruction pointer
Special register in the CPU that says where in memory to look for the next instruction in the program.

(d) packet
A message sent over a network like the internet is broken up into smaller fixed-size units (packets) each sent independently; and then the message is reassembled from packets at the receiver.

3. Messages and circuits:
Here is a message M encoded as E with a one time pad P as described in lecture where 0 means "*don't flip*" and 1 means "*flip*" (written in groups of 7 bits):

(a) What are the bits of the decoded message?

E: 1011011 1110001 0010011 1011101 1011110
P: 0010011 0111000 1011110 0010010 0010011
M: 1001000 1001001 1001101 1001111 1001101 ← *your answer*
(3pts)

(b) Suppose this string of bits represents a sequence of integers, each written as 7 bits. What is the highest integer value that could be encoded in 7 bits?

2^7-1 = 127
(2pts)

Suppose those 7-bit numbers actually represented English characters, using this ASCII format:

48	0	49	1	50	2	51	3	52	4	53	5	54	6	55	7
56	8	57	9	58	:	59	;	60	<	61	=	62	>	63	?
64	@	65	A	66	B	67	C	68	D	69	E	70	F	71	G
72	H	73	I	74	J	75	K	76	L	77	M	78	N	79	O
80	P	81	Q	82	R	83	S	84	T	85	U	86	V	87	W
88	X	89	Y	90	Z	91	[	92	\	93	]	94	^	95	_

(c) What does the message say?
HIMOM ← *your answer*
(2pts)

(d) Write the truth table for a boolean expression that computes one bit of M based on the corresponding bits from P and E.

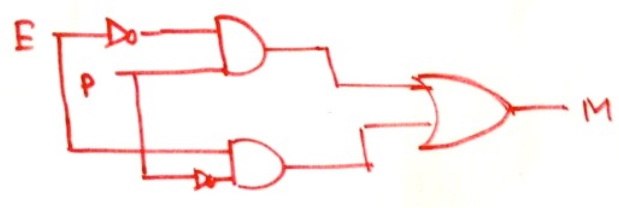
P	E	M
0	0	0
0	1	1
1	0	1
1	1	0

(3pts)

(e) Write a boolean expression for (d).

$P\overline{E} + \overline{P}E$ (3pts)

(f) Draw a circuit that computes (d). (3pts)



(g) How would make a circuit that does this for all 35 bits? (One or two sentences and/or small diagram.)

Best answer: just make 35 copies of the circuit in (f). (2pts)

(h) How much slower would the circuit in (g) produce its output than the circuit in (f)?

Same speed. (2pts)

(i) Is the circuit in (g) considered *sequential*? Briefly, how do you know?

No it is not sequential because the outputs do not loop back into inputs anywhere. (2pts)

4. Computer security:

Briefly explain (two or three sentences) each of the following concepts: (2pts each part)

(a) trojan horse (in computers)

Program pretends to be a different program to trick users into typing password or other sensitive information (like bank info).

(b) CAPTCHA

A programmatic test to determine if the user is human. Based on tasks that are hard for computers, such as reading messed up characters in an image. Used to get free email accounts, for example.

(c) tipping point

The point at which the spread of a virus (or idea, transmitted virally) takes off exponentially.

(d) distributed denial of service

An attack wherein a bot net simultaneously accesses a single web site from thousands of different computers, thereby preventing anyone else from being able to access that web site.

(e) virus vs. worm

Virus requires human intervention (for example sent as email attachment that user opens) whereas worm spreads from computer to computer exploiting weakness in software that is already running.

5. Graphics: (2pts each)
(a) In the RGB color model, roughly what combination of colors gives you orange?
R=255 G=165 (anything in the middle) B=0

(b) What combination gives you black?
R=0,G=0,B=0

(c) Briefly, what does the "ambient" term in Phong illumination account for?
It is a cheap, simple approximation to account for lighting that generally scatters everywhere in the scene.

(d) What about the "specular" term?
Highlights on glossy surfaces, for example a reflection of the light source in a mirror surface.

(e) In broad terms what are the main differences between how a rasterizing renderer and a ray tracer work?
Rasterize: project objects in scene into the image plane, and draw pixels for them.
Ray trace: start from camera and work backwards into the scene bouncing off objects.

(f) What are some visual effects that can be seen in one of the renderers and not in the other?

Ray tracer can easily show reflections, shadows and transparent objects like glass.

6. Algorithms
Here is an algorithm invented by the Greek mathematician Euclid:

```
while b ≠ 0
{
    if a > b
       a ← a − b
    else
       b ← b − a
}
print a
```

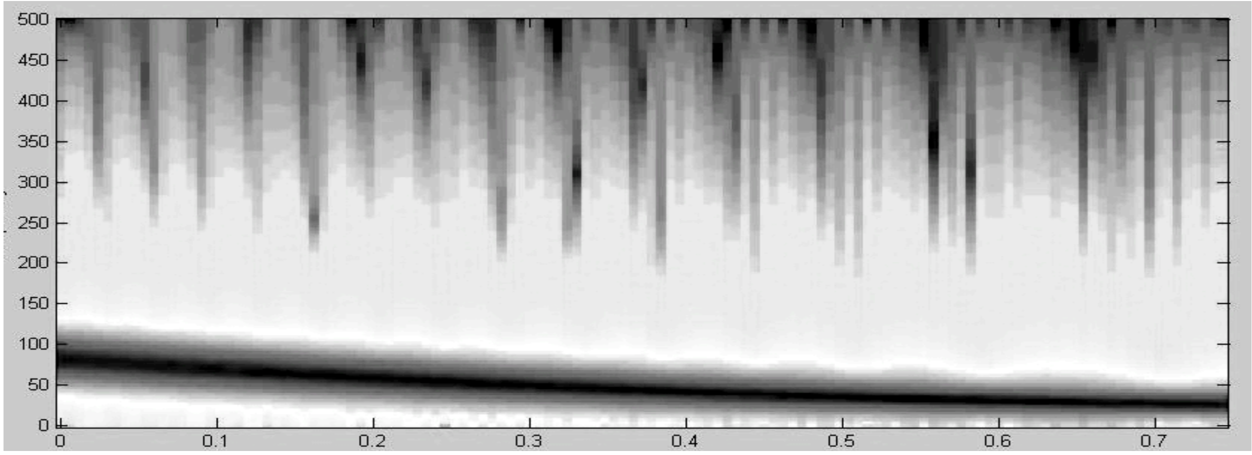
(a) What does this algorithm do for if a starts at 78 and b starts at 12?

Prints a=6. (4pts)

(b) What does this algorithm do for other inputs in general?

Prints greatest common divisor of a and b.
(4pts)

7. Digital audio: (4pts each part)
Here is a spectrogram of a digital audio signal:



(a) What do you think are the units on the two axes on the spectrogram?
Horizontal (X) axis: time (sec) Vertical (Y) axis: frequency (hz)

(b) What would you expect this signal to sound like to your ears?

Strong low tone gets deeper during the interval. Higher pitch chirps repeating periodically.

8. SPAM (3pts each part)

(a) A naive way to filter spam email is to write a set of rules that describe spam emails, e.g. contain the words "mortgage" or "viagra" or written in all capitals. Describe a simpler and more effective way to filter spam email.
See lecture or lab for how to computer SPAM score based on corpus of examples.

(b) What are advantages of your improved method over the naive method.

Adapts to your examples without having to enumerate a bunch of rules.

(c) Will it work for all spam email? Explain briefly.

No spammers are always figuring out new tricks to get around spam filters.

(d) How might a spammer "poison" your spam filter?

Send many spam emails that have ham words in them to ruin the statistics of your learned filter.