CMSC452 Final

1. This is an open-everything exam. You can use anything except ask another person. **Caution:** if you copy from the web or elsewhere mindlessly you will probably get it wrong.

2. There are 3 problems which add up to 60 points. There was a take-home part that was worth 40 points.

3. All problems have the next page or two blank. You may use the page a problem is on and the blank pages right after it for your answer.

4. The exam is Thursday May 13 8:00PM-10:15PM unless you have contacted me to make other arrangements. So the exam is 2 hours and 15 minutes

5. For each question show all of your work and use LaTeX or write VERY NEATLY. Clearly indicate your answers. No credit for illegible answers.

6. Please write out the following statement: *I pledge on my honor that I will not give or receive any unauthorized assistance on this examination.*
1. (20 points) Give an example of each of the following.

**NO PROOF REQUIRED**

(a) (7 points) A decidable set that is thought to NOT be in NP.

**SOLUTION**

The following are all correct:

- \( \overline{SAT} \), \( \overline{CLIQ} \), the complement of any NPC set. We do not thing NP is closed under complement.
- Any PSPACE complete set. QSAT, Generalized Geography, Regex-Equivalence.
- Any EXPSPACE or EXPTIME complete set: Chess, Go, trex-equivalence, WS1S, WS2S, S1S, S2S, Succint versions of NPC problems. Simlating a Turing Machine for \( n \) steps (\( n \) given in binary).

The following are INCORRECT:

- \( \Sigma^* \), \( \emptyset \), 2-SAT, 2-COL, PRIMES, \( \{ e : M_e \text{ has a prime number of states} \} \), any set that is in P.
- SAT, 3-SAT, TRAV (also called Travelling Salesman Problem), any set that is in NP.
- HALT, TOT (The set of all total Turing machines), any undecidable problem.
- Sets I never heard of and could not find on the web anywhere.
- Sets that are ill defined.

**END OF SOLUTION**

(b) (7 points) A set that is in NP, but whether it is in P or NP-complete is unknown.

**SEE SOLUTION TO 1c**
(c) (6 points) ANOTHER set that is in NP, but whether it is in P or NP-complete is unknown.
SOLUTION to both 1b and 1c

The following are **CORRECT**

- FACT which is the set:
  \[ \{(a, b) : \text{there is a factor of } a \text{ that is } \leq b\} \]
- Discrete Log, which needs to be written as a set.
- Graph Isomorphism which is \{\((G, H) : G \text{ and } H \text{ are isomorphic}\}\}.
- Compliment of FACT, DL, GI.

The following are **INCORRECT**

- Eulerian Circuite, 2SAT, 2COL, PRIMES, Any problem in P.
- CLIQ, SAT, any problem that is NPC.
- The set of parity games where player 0 is the winner. (WARNING- parit games are not much fun.)
- Circuit Minimization problems- these are \(\Sigma^p_2\) or \(\Pi^p_2\) depending on how they are phrased. Hence they are NOT in NP.
- WS1S- this is not in NP.
- The Sum of Squares problem: Given \(a_1, \ldots, a_n\) and \(b_1, \ldots, b_n\) determine if \(\sum_{i=1}^{n} a_i^2 < \sum_{i=1}^{n} b_i^2\). Note that the \(a_i\) and \(b_i\) are in binary so the input is very short. This problem is NOT KNOWN TO BE IN NP

• Diophantine Equation. This either refers to Hilbert’s tenth problem which is undecidable, or it is ill-defined.
• Sets I never heard of and could not find on the web anywhere.
• Sets that are ill defined.

END OF SOLUTION
2. (20 points) In this problem we modify our Turing Machines to they can (1) take elements of \( \mathbb{Z} \) as input and (2) have elements of \( \mathbb{Z} \) as output. (They can still output Y and N when used to decide sets.)

Let \( g \) be a COMPUTABLE function from \( \mathbb{Z} \) to \( \mathbb{Z} \) such that

\[
\cdots \leq g(-2) \leq g(-1) \leq g(0) \leq g(1) \leq g(2) \leq \cdots.
\]

Let

\[
Y = \{ \ldots, g(-2), g(-1), g(0), g(1), g(2), \ldots \}
\]

Is \( Y \) decidable?

If YES then give an algorithm that decides \( Y \).

If NO then prove that it is not.

Do this problem on this page and the next page.
SOLUTION

YES, IMAGE(\(g\)) is computable. There are cases

a) There exists \(a, b \in \mathbb{Z}\) such that, for all \(x \in \mathbb{Z}\), \(a \leq g(i) \leq b\). Then IMAGE(\(g\)) is finite, so computable.

b) There exists \(b\) such that, for all \(x \in \mathbb{Z}\), \(g(i) \leq b\), but \(\lim_{x \to -\infty} g(x) = -\infty\).

Let \(A\) be the FINITE SET \(\{g(0), g(1), \ldots\}\).

Here is the algorithm for IMAGE(\(g\)).

(a) Input \(y\)
(b) If \(y \in A\) then say YES.
(c) If \(y \notin A\) then compute \(g(0), g(-1), \ldots\), until you either hit \(y\), so say YES, or find an \(i\) such that \(g(i) < y\), so say NO.

c) There exists \(a\) such that, for all \(x \in \mathbb{Z}\), \(g(i) \geq a\), but \(\lim_{x \to \infty} g(x) = \infty\).

Let \(A\) be the FINITE SET \(\{g(0), g(-1), \ldots\}\).

(a) Input \(y\)
(b) If \(y \in A\) then say YES.
(c) If \(y \notin A\) then compute \(g(0), g(1), \ldots\), until you either hit \(y\), so say YES, or find an \(i\) such that \(y < g(i)\), so say NO.

d) Both
\(\lim_{x \to \infty} g(x) = \infty\) and
\[
\lim_{x \to -\infty} g(x) = -\infty.
\]

We leave this one to you.

**GRADING**

- If you recognized that there was a different cases and to them correctly, full credit, 20 points.
- If you did only one case (I think always the unbounded in both directions) then half credit, 10 points. **NOTE** These were fairly badly written, so a regrade request is likely to lose points.
- If you did something else or we didn’t understand what you did, then 0 points. **NOTE** These were really badly written, so a regrade request is likely to lose points.

**END OF SOLUTION**

3. (20 points)

For this problem you should assume \( n \) is a large power of 2, and \( \Sigma = \{ a, b \} \).

(a) (5 points) Let \( w \) be a string of length \( n \). Show that there is a Chomsky Normal Form CFG for \( \{ w \} \) with \( O(n) \) rules.

**SOLUTION**

\[
w = w_1 \cdots w_n.
\]

\[
S \rightarrow w_1 U_1
\]

\[
U_1 \rightarrow w_2 U_2 \quad \text{(So get } w_1 w_2 U_2)\]

\[
U_2 \rightarrow w_3 U_3 \quad \text{(So get } w_1 w_2 W_3 U_3)\]

\[
: \]

\[
U_{n-1} \rightarrow w_n U_n \quad \text{(So get } w_1 \cdots w_n U_n)\]
(b) (5 points) Let \( w = a^n \). Show that there is a Chomsky Normal Form CFG for \( \{w\} \) with \( O(\log n) \) rules.

**SOLUTION**

Let \( n = 2^k \).

\[ S \rightarrow A_1 A_1 \]  (So get \( A_1^2 \))

\[ A_1 \rightarrow A_2 A_2 \]  (So get \( A_2^2 \))

\[ A_2 \rightarrow A_3 A_3 \]  (So get \( A_3^2 \))

\[ \vdots \]

\[ A_{k-1} \rightarrow A_{k-2} A_{k-2} \]  (So get \( A_{k-2}^{2k-2} \))

\[ A_k \rightarrow A_{k-1} A_{k-1} \]  (So get \( A_{k-1}^{2k-1} \))

\[ A_{k+1} \rightarrow A_k A_k \]  (So get \( A_k^{2k} \))

\[ A_k \rightarrow a \]

**END OF SOLUTION**

(c) (10 points) Show that there exists a string \( w \) such that ANY Chomsky Normal Form CFG for \( \{w\} \) requires MUCH MORE than \( \log n \) rules.

**SOLUTION**

Let \( w \) be Kolm random of length \( n \).

So any description of \( w \) must be of length \( \geq n \).

Let \( G \) be a Chomsky Normal Form CFG that generates \( \{w\} \).

Assume it has \( R \) rules.

Hence it has at most \( 4R \) nonterminals (in the case that every rules uses diff nonterminals). So we can represent each nonterminal by \( \Theta(\log R) \) bits. Hence the entire CSG is of size \( \Theta(R \log R) \).
The CFG IS a description of \( w \). Hence we have a description of \( w \) of length \( \Theta(R \log R) \).

So

\[
R \log R \geq n
\]

Therefore \( R \geq \frac{n}{\log n} \).

**GRADING**

- If you said to take \( w \) to be Kolm random and gave some of the proof (they were usually badly written) then full credit, 10 points.
- If you gave a string \( w \), for example

\[
ab^2ab^4ab^8ab^{16}\ldots a^{\log n-1}
\]

(The \( \log n - 1 \) does not quite work to get you a string of length \( n \), but thats fine.)

and CLAIMED it REQUIRED a large CNF CFG because YOU could not come up with a better one you were told RESPECT! and got 5 points. (If you did this and you appeal you need to give me a CLEAN CLEAR PROOF that ANY CNF CFG takes MUCH MORE than \( \log n \) size. MUCH more DOES NOT mean \( 2\log n \) or \( \log n + 1 \).

- If you did not give a string or tell why such a string exists, if you gave a SET of strings and claimed it required a large CFG, if you took \( n = 8 \) or some other small number, 0 points.

- If you said that \( 2\log n \) or \( \log n + 1 \) or something like that was much bigger than \( \log n \) then 0 points.

**END OF SOLUTION**