Problem Set 6

1. You've seen how the RSA encryption scheme works, but why is it hard to break? In this problem, you will see that finding secret keys is as hard as finding the prime factorizations of integers. Since there is a general consensus in the crypto community that factoring numbers with a few hundred digits requires astronomical computing resources, we can therefore be sure it will take the same kind of overwhelming effort to find RSA secret keys of a few hundred digits. This means we can be confident the private RSA keys are not somehow related to the public keys.

For this problem, assume that $n = p \cdot q$ where p, q are both odd primes and that e is the public key and d is the secret key of the RSA protocol as described in Week 6 Notes. Let $x := e \cdot d - 1$.

a. Show that $\phi(n)$ divides x.

$\phi(n) = \phi(pq) = \phi(p)\phi(q) = (p-1)(q-1)$ since p, q are prime
$de = 1 \pmod{(p-1)(q-1)}$ from the definition of d
so $\phi(n) \mid (ed-1)$ from the definition of congruent
thus $\phi(n) \mid x$.

b. Conclude that 4 divides x.

$\phi(n) = \phi(p)\phi(q)$ since they are relatively prime

$= (p-1)(q-1)$ since they are prime

$p \wedge q$ are odd $\Rightarrow (p-1) \wedge (q-1)$ are even

so $\phi(n)$ is a multiple of 4

so $4 \mid \phi(n)$

$4 \mid \phi(n) \wedge \phi(n) \mid x \Rightarrow 4 \mid x$

c. Show that if $\gcd(r,n)=1$, then $r^x \equiv 1 \pmod{n}$

$\phi(n) \mid x$ from part a.

$\exists k \in \mathbb{N} \ni x = k\phi(n)$

$r^x = r^{k\phi(n)} = (r^{\phi(n)})^k$

$r^{\phi(n)} \equiv 1 \pmod{n}$ from Euler's theorem

so $r^x \equiv (1 \pmod{n})^k = 1 \pmod{n}$ //

A square root of m modulo n is a nonnegative integer $s<n$ such that $s^2 \equiv m \pmod{n}$. Here is a nice fact to know: when n is a product of two odd primes, then every number m s.t. $\gcd(m,n)=1$ has 4 square roots modulo n.

In particular, the number 1 has 4 square roots modulo n. The two trivial ones are 1 and $n-1$ (which is $\equiv -1 \pmod{n}$). The other two are called the nontrivial square roots of 1.

d. Since you know x, then for any integer, r, you can also compute the remainder, y, of $r^{x/2}$ divided by n. So $y^2 \equiv r^x \pmod{n}$. Now if r is relatively prime to n, then y will be a square root of 1 modulo n by part c.

Show that if y turns out to be a <u>nontrivial</u> root of 1 modulo n, then you can factor n. *Hint:* From the fact that $y^2-1=(y+1)(y-1)$, show that $y+1$ must be divisible by exactly one of q and p.

$y^2 \equiv r^x \pmod{n}$

$y^2 \equiv 1 \pmod{n}$ since $r^x \equiv 1 \pmod{n}$

$n \mid (y^2-1)$ by def of congruent in modulo

$pq \mid (y+1)(y-1)$

So either $p \mid (y+1)$ or $p \mid (y-1)$ and $q \mid (y+1)$ or $q \mid (y-1)$

$0 < y-1 < y+1 < n$ since y is non trivial

$0 < y+1 < pq \Rightarrow y+1$ is only divisible by <u>one</u> of $\{p,q\}$

So $\gcd(y+1, n) = p$ or q

Thus we can factor n.

e. It turns out at least half the positive integers $r < n$ that are relatively prime to n will yield y's in part (d) that are nontrivial roots of 1. Conclude that if, in addition to n and the public key, e, you also knew the secret key d, then you can be sure of being able to factor n.

Most r's will hold $\gcd(r, n) = 1$
Half of these will yield non-trivial roots (y in part d)
So if you choose $r < n$ at random, it won't take too long before you find a y, which in part 4. we showed can factor n.

2. The Massachusettes Turnpike Authority is concerned about the integrity of the new Zakim bridge. Their consulting architect has warned that the bridge may collapse if more than 1000 cars are on it at the same time. The authority has also been warned by their traffic consultants that the rate of accidents from cars speeding across bridges has been increasing.

Both to lighten traffic and to discourage speeding, the Authority has decided to make the bridge one-way and to put tolls at both ends of the bridge. So cars will pay tolls both on entering and exiting the bridge, but the tolls will be different. In particular, a car will pay $3 to enter onto the bridge and will pay $2 to exit. To be sure that there are never too many cars on the bridge, the Authority will let a car onto the bridge only if the difference between the amount of money currently at the entry toll booth minus the amount at the exit toll booth is strictly less than a certain threshold amount of $To.

The consultants have decided to model this scenario with a state machine whose states are triples of natural numbers, (A, B, C), where

- A is an amount of money at the entry booth
- B is an amount of money at the exit booth
- C is a number of cars on the bridge

Any state with C > 1000 is called a collapsed state, which the Authority dearly hopes to avoid. There will be no transition out of a collapsed state. Since the toll booth collectors may need to start off with some amount of money in order to make change, and there may be some number of "official" cars already on the bridge when it is opened to the public, the consultants must be ready to analyze the system started at <u>any</u> state. So let A_0 be the initial number of dollars at the entrance toll booth, B_0 the initial number of dollars at the exit toll booth, and C_0 the number of official cars on the bridge when it is opened. The authority will be careful to ensure that C_0 is not

large enough to cause a collapse. You should assume that even official cars pay tolls on exiting or entering the bridge after the bridge is opened.

a. Give a mathematical model of the Authority's system for letting cars on and off the bridge by specifying a transition relation between states of the form (A, B, C) above.

1. Car enters $(A, B, C) \rightarrow (A+3, B, C+1)$ for $(A-B) < T_0 \wedge C < 1000$
2. Car exits $(A, B, C) \rightarrow (A, B+2, C-1)$ for $0 < C \leq 1000$.

b. Characterize each of the following derived variables:

A, B, A+B, A-B, 3C-A, 2A-3B, B+3C, 2A-3B-6C, 2A-2B-3C

as one of the following: constant, strictly increasing, strictly decreasing, weakly decreasing but not constant, weakly increasing but not constant, none of the above, and briefly explain your reasoning.

Derived Variable	Δ under transition i	Δ ii	Classification
A	>0	0	WI
B	0	>0	WI
A+B	>0	>0	SI
A-B	3	-2	N
3C-A	0	-3	WD
2A-3B	6	-6	N
B+3C	3	-1	N
2A-3B-6C	0	0	C
2A-2B-3C	3	-1	N

The Authority has asked their engineering consultants to determine T and to verify that this policy will keep the number of cars from exceeding 1000. The consultants reason that if A_0 is the initial number of dollars at the entrance toll booth, B_0 is the initial number of dollars at the exit toll booth, and C_0 is the number of official cars on the bridge when it is opened, then an additional $1000 - C_0$ cars can be allowed on the bridge, so as long as $A - B$ has not increased by $3(1000 - C_0)$ there shouldn't be more than 1000 cars on the bridge. So they recommend defining

$$T_0 ::= 3(1000 - C_0) + (A_0 - B_0).$$

c. Use the results of part (b) to define a simple predicate, P, on states of the transition system which is satisfied by the start state, that is $P(A_0, B_0, C_0)$ holds, is not satisfied by any collapsed state, and is an invariant of the system. Verify that the P you define has these properties.

$$P: \underbrace{[2A - 3B - 6C = \underbrace{2A_0 - 3B_0 - 6C_0}_{D_0}]}_{\text{I}} \wedge \underbrace{[C \leq 1000]}_{\text{II}}$$

Verification:

Transition i: I is constant, so it holds (see b.)

We must also satisfy the condition that $A - B < T_0$

Suppose $A - B \geq T_0$

Then $2A' - 3B' - 6C' = D_0$ ($C' \leq 1000$ for this transition to be valid)

$6C' = 2A' - 3B' - D_0$

$6C' = 2A + 6 - 3B - D_0 = 2(A - B) - B - D_0 + 6$

$6C' \geq 2T_0 - B - D_0 + 6 \geq 2T_0 - B_0 - D_0 + 6$ since B is WI

$6C' \geq 6(1000 - C_0) + 2(A - B_0) - B_0 - D_0 + 6 = 6000 + 2A_0 - 3B_0 - 6C_0 - D_0 + 6$

$6C' \geq 6006 + D_0 - D_0$

$C' \geq 1001 \Rightarrow C' > 1000 \Rightarrow \Leftarrow$, so II assures $A - B < T_0$ is satisfied. Transition ii: I still constant II, C and A-B decrease.

d. A clever MIT intern working for the turnpike authority agrees that the Turnpike's bridge management policy will be 'safe': the bridge will not collapse. But she worms that the policy will be a 'deadlock' - a situation in which traffic can't move on the bridge even though it has not collapsed. Explain more precisely in terms of system transitions what the intern means, and briefly, but clearly, justify her claim.

There will reach a state in which no more transitions are possible because of the discrepancy between the two toll prices

Justification:

Let C_{tot} represent the total number cars that have successfully crossed the bridge.

$A - B \geq C_{tot} + A_0 - B_0$ since $A-B$ grows by 1 for each completed crossing

$\geq 3000 + A_0 - B_0$ when ≥ 3000 cars have crossed

$\geq 3000 - C_0 + A_0 - B_0$

$= T_0$

This invalidates the $A - B < T_0$ condition, blocking all transition (i)'s

Then only cars can exit, (ii) and after they leave the bridge $C=0$, and no more transitions can be made, leaving the system in deadlock.

3. Vertices u, v in a digraph are said to be <u>unconnected</u> when there is no path $u \to v$ or $v \to u$. The following procedure can apply to any digraph, G:

Pick two vertices $u, v \Rightarrow$ either

1. There is an edge (u,v) of G and there is a path $u \to v$ which does not include this edge; in this case delete this edge (u,v), or.
2. u and v are unconnected; in this case, add the edge (u,v).

Repeat these operations until it is no longer to find vertices u, v to which an operation applies.

This procedure can be modeled as a state machine. The start state is G, and the states are all possible digraphs with the same vertices as G. The final states are the digraphs on which no operation is possible.

a. For any state, G, let e be the number of edges, and p its number of pairs of unconnected vertices. Define a decreasing $n \in \mathbb{N}$ valued derived variable that is a function of e and p. Conclude that the procedure terminates started on any finite digraph, G.

$n = e + 2p$

In transition 1. an edge is deleted so e decreases. The conditions for transition 1 $\Rightarrow$ that the deletion of this edge will not increase p

In transition 2. one edge is created so e increases by 1, and p is decreased by at least 2,

so in both cases n decreases.

Conclusion of termination. 1. will cease when only necessary edges are left

2. will stop when all pairs are connected.

b. Prove that the set of final states reachable from DAG start states are path graphs.

Proof

First we will show that all path graphs represent final states.

Consider a path graph P.

Transition (1.) is impossible since there is only one directed path from any arbitrary vertex to another in a path graph.

Transition (2.) is impossible since all vertices are connected in a path graph.

Thus no operations can be performed on P, and P is a final state.

Next we will show that all final states are path graphs are final states by showing that all non-path DAG's are not final states

Consider an arbitrary DAG G. G is not a path graph.

Case 1 (All vertices are connected):

Then there must exist a vertex, v, that has edges to multiple vertices by in-paths, out-paths, or both (meaning multiple connections both ways). Since all vertices are connected, transition (1.) can be performed by removing the edge connecting v to the second vertex it has an edge to (or from) since the two vertices v has paths to are already connected.

Case 2 (Some vertices are unconnected)

Operation (2.) can be performed.

So an operation can be performed on G, making it not a final state.

Thus all non-path graphs are non-final states

So all final states are path graphs

C. Prove that the property of being a DAG is an invariant of this procedure.

Proof

By the definition of invariant, graphs whose start states are DAGs must remain DAGs under all transitions for DAG to be an invariant property.

Consider an arbitrary DAG G.

Case 1 (the conditions for transition 1 are met):

Then we reach the next state by deleting some edge of G.

Deleting an edge on a graph with no cycles (since G is a DAG this holds true) will not create any cycles.

So G has no cycles after transition 1

Case 2 (the conditions for transition 2 are met):

Adding an edge (u, v) creates a cycle between two vertices, say u & v, only if a path from $v \to u$ already exists, since a cycle is a positive length path beginning and ending at the same vertex (u, in this case).

Transition 2 adds an edge only between unconnected vertices, so the condition to create a cycle is not satisfied, and no cycles are created by transition 2.

Case 3 (G is a final state):

Trivial, nothing changes

So by the definition of a DAG, G remains a DAG after any valid transition since it remains free of any cycles.

So the DAG property is invariant. //

d. Prove that if G is a DAG, the procedure terminates with a line graph whose path is a topological sort of the partial order defined by G.
Hint: Strengthen the DAG invariant in the previous part.

Let the invariant be that G' is a refinement of $G \; \forall \; G \to G'$
Thus they have the same domains, codomains, and all relations hold.
So the only possible line graph is a topological sort of the original graph.

4. a. Give an example of a stable match between 3 boys and 3 girls where no person gets their first choice.

Boy	Choices	Girl	Choices
1	a, b, c	a	2, 3, 1
2	b, c, a	b	3, 1, 2
3	c, a, b	c	1, 2, 3

Matches: (1, b), (2, c), (3, a)
This was constructed by giving each person their second choice, and making every first choice a reciprocal last choice.

4.b. Describe a simple procedure to determine whether or not a stable marriage problem has a unique solution, that is, only one possible stable marriage assignment.

Match everyone using the mating algorithm twice.

1. First normally
2. Second with the gender sets swapped.

If the solutions are identical, the problem has a unique solution.

The mating algorithm matches each male with his optimal female mate, and each female with her pessimal spouse. If the same solution is yielded by each matching, then we have found the only stable solution, since optimal = pessimal (state) $\Rightarrow$ only (state)

5. A Harvard BS graduates and starts with an annual salary of $140,000, with a $25,000 raise guaranteed every year. An MIT SB graduate starts with $100,000, with a guaranteed 15% raise every year. Assume the bank rate is a fixed 3% per year. That is, the bank will pay $1.03 a year from now if you deposit $1.00 today.

a. Suppose both graduates retire after the same number of years. Use the fact that $x = o((1+\epsilon)^x)$ to explain why an MIT SB must come out ahead if they work for enough years.

Let H and M be the respective total incomes of the graduates, who work for n years.

$$H = \sum_{y=0}^{n} (140{,}000 + 25{,}000y)\left(\frac{1}{1.03}\right)^y$$

$$M = \sum_{y=0}^{n} (100{,}000(1.15)^y)\left(\frac{1}{1.03}\right)^y$$

The bracket sections show the differences in the sums. Since H's bracketed section grows linearly and M's grows exponentially, $\exists n \to \forall m > n$ $M > H$.

b. Suppose both graduates retire after n years. For which values of n is the MIT graduate's salary package better than the Harvard grad's?

$n \geq 15$.

See ps6p5b.py

6. Recall the book stacking challenge from the course notes where you have an unlimited supply of books to stack.

a. What if instead of all books weighing the same, you have a book that weighs 1 lb, then $\frac{1}{2}$, $\frac{1}{4}$, ... Say you have n such books, and also that you have a duplicate of the lightest book. How far out can you stack the books? Note all books are still the same size just different weights.

Note, I let book length = 1

$n=1$ $r=0$ (COM at edge for the furthest stack) (Heaviest book placed on bottom)

$r = \sum_{i=1}^{1} m_i x_i = 1(0)$ (book is centered at the edge) overhang = $\frac{1}{2}$ book

$n=2$. We are placing a book of equal weight below the above stacks (of 1). The COM of the top book is at the edge of the bottom book (since that is how it was stacked on the table)

$r = \sum_{i=1}^{2} m_i x_i = m(x_1 + (x_1 - \frac{1}{2})) = 0 \quad 2x_0 = \frac{1}{2} \quad x_0 = \frac{1}{4}$

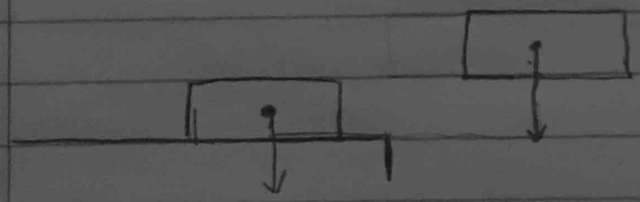
So the top book overhang increases by $\frac{1}{4}$

$n = m \in \mathbb{N}$ $\sum_{i=2}^{m} m_i = m_1$. So we can treat this the same as stacking two books of equal mass. Thus the result will be the same as for $n=2$, the overhang will increase by $\frac{1}{4}$.

So **for n books, the overhang = $\sum_{i=0}^{n} \frac{1}{4}$**

b. What if you had to stack such that the lightest books were on bottom of the stack and the heaviest books were on top of the stacks. How far out can you stack an infinite number of books where each book is twice as heavy as the book below it? (Infinitely or finitely far).

Finitely far.
Place the heaviest book on top of the system, and suppose that it is placed ∋ its center of mass is just one (much less than ∞) book length past the edge of the table. Since this book is double the mass of the books below it it is ≥ in mass to the sum of the masses of the books below it (see geometric series).
So in order to put the center of mass of the system at the edge of the table, the books below must have a center of mass 1 book length to the left of the table's edge, leaving a gap of 1 full book length between the edges of the adjacent books.
So length < 1 full book ⇒ less than ∞.



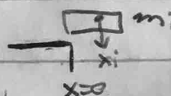
C. What if the books were Harmonically weighted: $1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}$. and the heaviest book had to be on top. Would it be possible for the top of the stack to be arbitrarily far past the edge of the table?

Yes.

Let $R = \sum_{i=1}^{n} x_i m_i$ be the center of mass of the stack of books.

Define the coordinate system as follows $x=0$ is at the edge of the table. x_i is the center of mass (middle since they are of uniform density) of book i.

Using this system a stack will balance iff $R \leq 0$



Suppose we want to place the top book at some arbitrary distance δ from the edge of the table.

Then $R \leq 0$ for the stack to be stable

$\sum_{i=1}^{n} x_i m_i \leq 0$ for a stack of n books

$\sum_{i=1}^{n-1} x_i m_i + 1(\delta) \leq 0$ (the top book has weight 1)

$\sum_{i=1}^{n-1} x_i m_i \leq \delta$

$\sum_{i=1}^{n-1} x_i \left(\frac{1}{i+1}\right) \leq \delta$

$\sum_{i=1}^{n-1} \left(\frac{1}{i+1}\right)$ is a harmonic series starting at $\frac{1}{2}$

Since it grows $\sim \ln(n)$, we can make the left hand of the inequality large enough to satisfy the above equation without putting the books too far behind (namely $-\frac{1}{2}$(book length)) the edge simply by stacking enough books.

The top of the stack can be arbitrarily far from the edge of the table

7. Use the integral method to find upper and lower bounds for the following summation that differ at most by 0.05.

$$\sum_{i=1}^{\infty} \frac{1}{i^3}$$

$$\sum_{i=1}^{\infty} \frac{1}{i^3} = 1 + \frac{1}{8} + \frac{1}{27} + \sum_{i=4}^{\infty} \frac{1}{i^3}$$

$$\int_4^{\infty} \frac{1}{x^3}\,dx \leq \sum_{i=4}^{\infty} \frac{1}{i^3} \leq \int_4^{\infty} \frac{1}{(x-1)^3}\,dx$$

$$\frac{-1}{2x^2}\Big|_4^{\infty} \leq \sum_{i=4}^{\infty} \frac{1}{i^3} \leq \frac{-1}{2(x-1)^2}\Big|_4^{\infty}$$

$$0 + \frac{1}{2(4^2)} = \frac{1}{32} \leq \sum_{i=4}^{\infty} \frac{1}{i^3} \leq 0 + \frac{1}{2(3)^2} = \frac{1}{18}$$

So $\boxed{1.19 \leq \sum_{i=1}^{\infty} \frac{1}{i^3} \leq 1.22}$

8. a. Given that $f(x) = O(g(x))$, prove that $f(x)^2 = O(g(x)^2)$

Proof

$\exists c \geq 0 \wedge x_0 \ni \forall x \geq x_0, |f(x)| \leq c\, g(x)$ from the def of O

Then $\forall x \geq x_0 \quad |f(x)^2| \leq (cg(x))^2 = c^2 g(x)^2$ assuming $g \geq 0$

So $\exists d \geq 0$ (namely c^2) $\wedge$ x_0 (namely x_0) $\ni \forall x \geq x_0, |f(x)^2| \leq d\, g(x)^2$

So from the def of O, $f(x)^2 = O(g(x)^2)$ //

b. Let $f(x) ::= 2x$ and $g(x) ::= x$, so $f(x) = O(g(x))$. Prove that $2^{g(x)} = o(2^{f(x)})$, so $2^{f(x)} \neq O(2^{g(x)})$.

Proof

$$\lim_{x\to\infty} \frac{2^{g(x)}}{2^{f(x)}} = \lim_{x\to\infty} \frac{2^x}{2^{2x}} = \lim_{x\to\infty} \frac{2^x}{2^x 2^x} = \lim_{x\to\infty} \frac{1}{2^x} = 0$$

So, by def of little oh, $2^{g(x)} = o(2^{f(x)})$

then $2^{f(x)} \neq O(2^{g(x)})$ from Lemma 7.8 //