Introduction to Algorithms	**Problem Set 7**
CS 482 Spring 2004	**Due April 2, 2004**

Please hand in each problem on a separate sheet with your name and NetID on each.

Reading: Chapter 6. In particular, the NP-completeness proofs should follow the framework in Sections 6.1–6.4. The overall structure you should use in these proofs is in Section 6.4: To prove that a problem X is NP-complete, first show that X is in $\mathcal{NP}$; then choose a problem Y that is known to be NP-complete (it can be any of the ones described in the chapter); and show $Y \leq_P X$.

(1) *(This is Problem 3 at the end of Chapter 6.)* Consider a set $A = \{a_1, \ldots, a_n\}$ and a collection $B_1, B_2, \ldots, B_m$ of subsets of A. (That is, $B_i \subseteq A$ for each i.)

We say that a set $H \subseteq A$ is a *hitting set* for the collection $B_1, B_2, \ldots, B_m$ if H contains at least one element from each B_i — that is, if $H \cap B_i$ is not empty for each i. (So H "hits" all the sets B_i.)

We now define the *Hitting Set* problem as follows. We are given a set $A = \{a_1, \ldots, a_n\}$, a collection $B_1, B_2, \ldots, B_m$ of subsets of A, and a number k. We are asked: is there a hitting set $H \subseteq A$ for $B_1, B_2, \ldots, B_m$ so that the size of H is at most k?

Prove that *Hitting Set* is NP-complete.

(2) *(This is a modified version of Problem 15 at the end of Chapter 6.)* Some friends of yours maintain a popular news and discussion site on the Web, and the traffic has reached a level where they want to begin differentiating their visitors into paying and non-paying customers. A standard way to do this is to make all the content on the site available to customers who pay a monthly subscription fee; meanwhile, visitors who don't subscribe can still view a subset of the pages (all the while being bombarded with ads asking them to become subscribers).

Here are two simple ways to control access for non-subscribers: you could (1) designate a fixed subset of pages as viewable by non-subscribers, or (2) allow any page in principle to be viewable, but specify a maximum number of pages that can be viewed by a non-subscriber in a single session. (We'll assume the site is able to track the path followed by a visitor through the site.)

Your friends are experimenting with a way of restricting access that is different from and more subtle than either of these two options. They want non-subscribers to be able to sample different sections of the Web site, so they designate certain subsets of the pages as constituting particular *zones* — for example, there can be a zone for pages on politics, a zone for pages on music, and so forth. It's possible for a page to belong to more than one zone. Now, as a non-subscribing user passes through the site, the access policy allows him or her to visit one page from each zone, but an attempt by the user to access a second page from the same zone later in the browsing session will be disallowed. (Instead, the user will be directed to an ad suggesting that he or she become a subscriber.)

More formally, we can model the site as a directed graph $G = (V, E)$, in which the nodes represent Web pages and the edges represent directed hyperlinks. There is a distinguished *entry node* $s \in V$, and there are zones $Z_1, \ldots, Z_k \subseteq V$. A path P taken by a non-subscriber is restricted to include at most one node from each zone Z_i.

One issue with this more complicated access policy is that it gets difficult to answer even basic questions about reachability, including: is it possible for a non-subscriber to visit a given node t? More precisely, we define the *Evasive Path* problem as follows: Given G, $Z_1, \ldots, Z_k$, $s \in V$, and a *destination node* $t \in V$, is there an s-t path in G that includes at most one node from each zone Z_i? Prove that *Evasive Path* is NP-complete.

(3) One thing that's not always apparent when thinking about traditional "continuous math" problems is the way discrete, combinatorial issues of the kind we're studying here can creep into what look like standard calculus questions.

Consider, for example, the traditional problem of minimizing a one-variable function like $f(x) = 3 + x - 3x^2$ over an interval like $x \in [0, 1]$. The derivative has a zero at $x = 1/6$, but this in fact is a maximum of the function, not a minimum; to get the minimum, one has to heed the standard warning to check the values on the boundary of the interval as well. (The minimum is in fact achieved on the boundary, at $x = 1$.)

Checking the boundary isn't such a problem when you have a function in one variable; but suppose we're now dealing with the problem of minimizing a function in n variables $x_1, x_2, \ldots, x_n$ over the unit cube, where each of $x_1, x_2, \ldots, x_n \in [0, 1]$. The minimum may be achieved on the interior of the cube, but it may be achieved on the boundary; and this latter prospect is rather daunting, since the boundary consists of 2^n "corners" (where each x_i is equal to either 0 or 1), as well as various pieces of other dimensions as well. Calculus books tend to get suspiciously vague around here, when trying to describe how to handle multivariable minimization problems in the face of this complexity.

It turns out there's a reason for this: minimizing an n-variable function over the unit cube in n dimensions is as hard as an NP-complete problem. To make this concrete, let's consider the special case of polynomials with integer coefficients over n variables $x_1, x_2, \ldots, x_n$. To review some terminology, we say a *monomial* is a product of a real-number coefficient c and each variable x_i raised to some integer power a_i; we can write this as $cx_1^{a_1}x_2^{a_2}\cdots x_n^{a_n}$. (For example, $2x_1^2x_2x_3^4$ is a monomial.) A *polynomial* is then a sum of a finite set of monomials. (For example $2x_1^2x_2x_3^4 + x_1x_3 - 6x_2^2x_3^2$ is a polynomial.)

We define the *Multivariable Polynomial Minimization* problem as follows: given a polynomial in n variables with integer coefficients, and given an integer bound B, is there a choice of real numbers $x_1, x_2, \ldots, x_n \in [0, 1]$ that cause the polynomial to achieve a value that is $\leq B$?

Choose a problem Y from the book or from lecture that is known to be NP-complete and show that

$$Y \leq_P \textit{Multivariable Polynomial Minimization}.$$