CS 151 Complexity Theory		**Spring 2004**
	Problem Set 7	
Out: May 20		*Due:* **May 27**

Reminder: you are encouraged to work in groups of two or three; however you must turn in your own write-up and note with whom you worked. You may consult the course materials and text (Papadimitriou). Please attempt all problems.

1. Consider the following generic setup: out of all 2^n strings in $\{0,1\}^n$, some subset $B \subseteq \{0,1\}^n$ of them are "bad" (for some application). You don't know B directly, but you do have an efficient way to recognize a bad string when you see one. That is, there is a small Boolean circuit C with n inputs for which $C(x) = 1$ if and only if $x \in B$. A natural thing to want to do is to estimate the number of bad strings. We can formulate this as the task of deciding the following promise problem LARGESET:

   - Input: circuit C with n inputs, integer k
   - YES instances: those pairs (C, k) for which $|\{x : C(x) = 1\}| \geq 3(2^k)$
   - NO instances: those pairs (C, k) for which $|\{x : C(x) = 1\}| \leq \frac{1}{3}(2^k)$

   In this problem you will show that LARGESET is in **AM**.

   (a) For a $k \times n$ matrix A with 0/1 entries and a vector $b \in \{0,1\}^k$, define the function $h_{A,b}(x) : \{0,1\}^n \to \{0,1\}^k$ by $h_{A,b}(x) = Ax + b$ (where all arithmetic is performed modulo 2). Prove that for all $x \in \{0,1\}^n$ and $y \in \{0,1\}^k$,

   $$\Pr_{A,b}[h_{A,b}(x) = y] = 2^{-k}$$

   and that for all $x_1, x_2 \in \{0,1\}^n$, $x_1 \neq x_2$, and $y_1, y_2 \in \{0,1\}^k$,

   $$\Pr[h_{A,b}(x_1) = y_1 \wedge h_{A,b}(x_2) = y_2] = 2^{-2k}.$$

   This shows that the family of functions $H = \{h_{A,b}\}$ is a *2-universal* family of hash functions from n bits to k bits. The following is a consequence (that you can verify using Chebyshev's Inequality, but you need not prove for this problem set): for each fixed $y \in \{0,1\}^k$,

   $$\Pr_{A,b}[\exists x \in B \;\; h_{A,b}(x) = y] \geq 1 - \frac{2^k}{|B|}.$$

   (b) Using part (a), give a 2-round **AM** protocol for LARGESET.

2. Recall that a *clique* in an undirected graph $G = (V, E)$ is a subset $V' \subseteq V$ with edges between every pair of vertices in V'. We know that the language

   $$\text{CLIQUE} = \{(G, k) : G \text{ has a clique of size } k\}$$

is **NP**-complete. You will show that there is some constant $\delta > 0$ for which CLIQUE is **NP**-hard to approximate to within N^δ in the following sense: if there is an N^δ-approximation algorithm for CLIQUE, then **NP** = **ZPP**. Here N is the length of the input (G, k).

The PCP Theorem implies that there is some constant $\epsilon > 0$ for which given a 3-CNF formula ϕ it is **NP**-hard to distinguish between the following two cases:

YES : ϕ is satisfiable
NO : every assignment to ϕ satisfies at most a $(1 - \epsilon)$ fraction of the clauses

Below you will describe a *randomized* transformation from an instance ϕ into a graph G whose intended effect is that a YES instance produces a graph with a large clique, while a NO instance produces a graph with only a very small clique. Here n is the number of variables in ϕ.

(a) Suppose ϕ is a NO instance, and consider the following probabilistic experiment: pick $\log_2 n$ clauses from ϕ uniformly at random, take their conjunction, and call this CNF ϕ_1; repeat n^3 times to get CNFs $\phi_1, \phi_2, \ldots, \phi_{n^3}$. Show that for a fixed assignment A:

$$\Pr[A \text{ satisfies at least } n^{3-\epsilon} \text{ of the } \phi_i] < e^{-n^2}.$$

Hint: What is the probability that A satisfies a given ϕ_i? What is the expected number of ϕ_i satisfied by A? You may want to use the fact that $(1 - \epsilon)^{1/\epsilon} \le 1/e$ for $1 > \epsilon > 0$, and the Chernoff bound: if X is the sum of independent 0/1 random variables with expected value $\mathrm{E}[X] = \mu$, then $\Pr[X > 2\mu] \le e^{-\mu/3}$.

(b) Argue that the above randomized procedure produces from ϕ a collection of 3-CNFs $\phi_1, \phi_2, \ldots, \phi_{n^3}$ for which

i. ϕ is a YES instance $\Rightarrow$ $\Pr[\exists$ assignment A simultaneously satisfying all of the $\phi_i] = 1$, and

ii. ϕ is a NO instance $\Rightarrow$ $\Pr[$no assignment satisfies more than $n^{3-\epsilon}$ of the $\phi_i] \ge 1/2$.

(c) Describe an efficient deterministic procedure to construct a graph G from the collection of 3-CNFs in part (b) for which

i. $\exists$ assignment A simultaneously satisfying all of the $\phi_i \Rightarrow G$ has a clique of size n^3, and

ii. no assignment satisfies more than $n^{3-\epsilon}$ of the $\phi_i \Rightarrow$ no clique in G has size greater than $n^{3-\epsilon}$.

(d) Prove that there exists a constant $\delta > 0$ for which an N^δ-approximation algorithm for CLIQUE implies that **NP** = **ZPP**, where N is the length of the input.