14.3

From Section 13.2 we recall that a multiset multicover means we are given the same universe U as in set cover, but this time a set of *multisets* where each set specifies the number of copies of each element contained within. We denote $M(S,e)$ to mean the multiplicity of e in S. Note that we have the same constraint as in set multicover where each element e must be covered r_e times.

Since set multicover is just a special case of multiset multicover (where $M(S,e) = 1$ for all e,S) we only need to consider multiset multicover. Let's look at the LP relaxation:

$$\text{Minimize } \sum c(S)x_S \text{ subject to:}$$
$$\sum_{S \ni e} M(S,e)x_S \geq r_e \quad \forall e$$
$$x_S \geq 0 \quad \forall S$$

As we did with our algorithm for set cover, let's consider the optimal solution $\vec{x}^*$ for this LP. We define $y_S = \lfloor x_S^* \rfloor$, $p_S = x_S^* - y_S$, and $r_e' = r_e - \sum_{S \ni e} y_S$. Now we apply randomized rounding using our p_i as the probabilities, so:

$$x_{S_i} = \begin{cases} 1 & \text{with probability } p_i \\ 0 & \text{with probability } 1 - p_i \end{cases}$$

We repeat this process $t = 8 \log n$ times(the 8 will be nice for our analysis later) to get partial solutions $C_1, \ldots, C_t$. Now, we consider the multiset C' which is formed by taking each element of S y_S times for every multiset S. To form our final output multiset C, we take two copies of this multiset C' and then one copy of each of the C_i. Now we just need to prove that with high probability this multiset C is also a multicover, and also that the expected cost of C is within $O(\log n)$ of OPT. To prove the latter, let's start by denoting OPT_{LP} as the optimum value of the LP relaxation and realizing that clearly $OPT_{LP} \leq OPT$. We can split our output C up into its components consisting of the two copies of C' and then the elements C_i obtained by randomized rounding. To analyze the first component, note that $c(C') \leq OPT_{LP}$ because we know that $y_S \leq x_S^*$ for all sets S by how we defined the y_S. We also realize that

$$E[c(C_i)] = \sum p_S \cdot c(S) \leq OPT_{LP}$$

which is also true just by definition (this time the definition of the p_i). Therefore,

$$\begin{aligned} E[c(C)] &= 2 \cdot c(C') + \sum_{i=1}^{t} E[c(C_i)] \\ &\leq 2 \cdot OPT_{LP} + \sum_{i=1}^{t} OPT_{LP} \\ &= (2+t)OPT_{LP} \\ &= O(\log n)OPT_{LP} \quad \text{because } t = 8 \log n \\ &\leq O(\log n)OPT \end{aligned}$$

which is exactly what we wanted. Now that we have the cost in good shape, we need to consider the probability that C is actually a multicover. This analysis is actually fairly similar to the version for set cover that we did in class - we start out by looking at the element e_i for some fixed i. For this element, we want to consider r'_{e_i}. There are two cases: either $r'_{e_i} \leq \frac{r_{e_i}}{2}$ or $r'_{e_i} > \frac{r_{e_i}}{2}$. In the first case, note that we can write out $r'_{e_i} = r_{e_i} - \sum y_S$ and rearrange it to see that it implies that $2\sum y_S \geq r_{e_i}$. Based on the definition of C', we know that this means the element e_i was covered completely by the two copies of C' that we incorporated into our solution C, so we are fine.

The second case is a little more complicated. We first label all the sets containing e_i as $S_1, \ldots, S_k$. We will also want to consider their corresponding probabilities $p_1, \ldots, p_k$. We can now define kt random variables as follows:

$$X_{lj} = \begin{cases} M(S_l, e_i) & \text{if } e_i \text{ was covered by } S_l \text{ in } C_j \\ 0 & \text{otherwise} \end{cases}$$

We let $X = \sum_{l=1}^k \sum_{j=1}^t X_{lj}$ and note that this is equal to the number of times e_i was covered by all the C_i. Therefore, for this element e_i, the probability that it was not covered r'_{e_i} times is $\Pr[X < r'_{e_i}]$. We now denote $\mu = E[X]$ and note that by all our definitions we have that $\mu \geq tr'_{e_i}$. We can also note that the variables X_{lj} all take values in $[0, r_{e_i}]$ by definition. For our probability analysis, we will need to use a slightly modified version of the Chernoff bound that states that for $X = \sum_{i=1}^k X_i$ where the X_i are i.i.d. and take values in the range $[0, M]$, we have that

$$\Pr[X \leq (1-\epsilon)\mu] \leq e^{-\frac{\mu\epsilon^2}{2M}}$$

where $\mu = E[X]$ (note that this bound can be obtained by considering the convexity of the K-L divergence $D(\cdot||\cdot)$ and can also be found in the Wikipedia article on Chernoff bounds). Anyway, using this bound and all our assumptions/definitions we find that

$$\begin{aligned} \Pr[X < r'_{e_i}] &= \Pr[X < 1 - (1 - \frac{r'_{e_i}}{\mu})\mu] \quad (\mu = E[X]) \\ &\leq e^{-\frac{\mu(1-\frac{r'_{e_i}}{\mu})^2}{2r_{e_i}}} \quad \text{by our Chernoff bound} \\ &\leq e^{\frac{-tr'_{e_i}(1-\frac{1}{t})^2}{2r_{e_i}}} \quad \text{because } \mu \geq tr_{e_i} \end{aligned}$$

To continue simplifying this equation, we now use the fact that $r'_{e_i} \geq \frac{r_e}{2}$ (remember we are still in the second case). This implies (skipping some of the algebra) that

$$\begin{aligned} \Pr[X < r'_{e_i}] &\leq e^{\frac{-t+2-\frac{1}{t}}{4}} \\ &\leq e^{1/2} \cdot e^{-t/4} \\ &= e^{1/2} \cdot e^{-2\log n} \\ &= \frac{e^{1/2}}{n^2}. \end{aligned}$$

Note that this was only for a fixed i, so we are not quite done. We now consider (again, skipping over some simple algebraic steps)

$$\begin{aligned}
\Pr[C \text{ is a multicover}] &= 1 - \Pr[\text{there exists an } i \text{ such that } e_i \text{ is not covered in } C] \\
&= 1 - n \cdot \frac{e^{1/2}}{n^2} \\
&= 1 - \frac{e^{1/2}}{n}.
\end{aligned}$$

Of course this is not that nice a guarantee, but we can use tricks similar to those we did in class in editing the number of times t we generate our partial solutions to show that C is a multicover with overwhelming probability.