Форма 5: разбор

1:

> Возьмем хеш-функцию h(x) = (kx + m) mod p, где p, k и m рандомятся. Возможны ли ситуации, когда хеш получается неравномерно распределенным? Покажите такие p, k и m, что из всех чисел от 0 до p существует ОЧЕНЬ МНОГО (O(p)) чисел y таких, что h(x) никогда не равно y. Советуем подумать про делимости.

Можно заметить, что если k, m, p все четные, то у нас не бывает нечетных остатков у выражения (kx + m) mod p. Это значит, что если все параметры зарандомятся четными (а это происходит с вероятностью примерно 1/8), то из p значений хеш-функции актуальными будут только p/2, что жутко неэффективно.

Для того, чтобы таких проблем с делимостью не было, обычно выбирают простые числа, и тогда распределение близко к равномерному.

2:

> Возьмем для строки полиномиальный хеш с базой 10. То есть h(s + c) = h(s) * 10 + (c - 'A' + 1), где (...) вернет номер символа в алфавите. Тогда какой будет хеш от строки ABACABA? Как вы думаете, почему с такой базой может быть удобно дебажить полиномиальные хеши?

Можно посчитать вручную, что хеш от ABACABA = 1213121. Это специальная хитрая идея, которая позволяет достаточно удобно дебагать хеши - вместо того, чтобы смотреть на значения по простому модулю, у вас каждая буква заменяется на соответствующую цифру. Только это не работает, если букв больше чем от A до J, но для дебага идеально.