作业三：

1. 利用 AIS 31 的理论，请论证下 DRBG_Hash 怎么保证前向安全性和增强后向安全性的？

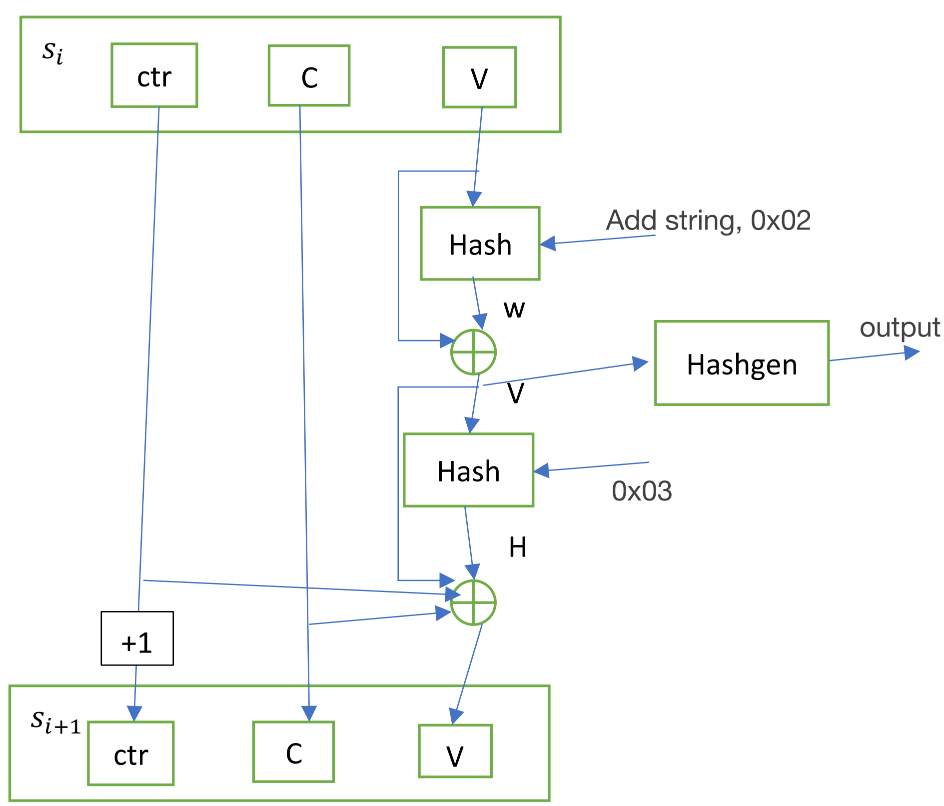
如下图所示，我们可以仿照 AIS 31 利用哈希函数来保证安全性的例子，画出如下图所示的 Hash_DRBG 结构。（由于使用了 Microsoft Word 的插入图形做图，箭头效果不尽如人意，勉强看吧。）

解答：

从图中我们可以看出，这个设计，仍然使用了两个哈希函数来保证其安全性。

首先来论证，如何保证前向安全性。前向安全性是指知道当前的输出，无法预测随机数发生器之后的输出。如图，假设我们知道的当前的 output（图中右端 Hashgen 的输出），显然我们无法 Hashgen 函数之前的 V 值，同时由于我们无法知道 C 值和 ctr 的值，我们将无法进行后续的计算计算出下一个 output。所以这能保证前向安全性，主要通过 Hashgen 函数的单向性保证。

其次，我们来论证，如何保证增强的后向安全性。增强的后向安全性指的是，如果已知输出和中间状态，无法知道以前的输出。具体到图上，就是说，如果我们已知了中间状态s_i这个状态中的三个状态量 $C, V, \text{reseed\_counter}$, 以及这个状态下的输出 output，我们无法推测之前的状态$s_j (j < i)$的输出是多少。很显然，我们知道了上述数据，我们可以计算出随机数发生器之后的输出，但是由于第二个 hash 函数以及之后的五个数据异或的操作，即使我们可以知道前一个状态s_{i-1}的 ctr 和 C 值，我们也无法根据当前的 V 倒推计算出 H，V 的中间值，w 等的值，所以我们无法知道前一个状态的输出和 V 值。这就保证了增强后向安全性。



2. 未处理 01 序列中 0 的概率 p0=0.4，1 的概率 p1=0.6，经过 4 级（非重叠）异或链后处理，即 XOR-4，处理后序列的 01 概率是多少？再分析下，XOR 后处理是否能够减弱序列之间的相关性？请给出一个理论上的阐述或证明。

解答：

（1）我们可视序列中出现的序列中的每一个比特为独立的随机变量。由于选用的后处理方法为非重叠的异或链后处理。所以问题可建模成，已知随机变量X_0, X_1, X_2, X_3独立同分布，服从参数为 0.6 的 0-1 分布，求随机变量$Y = X_0 \oplus X_1 \oplus X_2 \oplus X_3$的分布。

而这个问题，可利用堆积引理轻松解决。堆积引理的内容如下

- 定义对于$\{0,1\}$上的随机变量$X_i, p_i = P(X_i = 0)$
- 定义X_i的偏差为：$\epsilon_i = p_i - \frac{1}{2}$
- 设$X_{i_1}, X_{i_2}, \cdots, X_{i_k}$是独立的随机变量，$\epsilon_{i_j}$表示随机变量$X_{i_j}$的偏差（$j = 1,2,\cdots,k$），那么对变量$Y = X_{i_1} \oplus X_{i_2} \oplus \cdots \oplus X_{i_k}$的偏差，则有：$\epsilon_{i_1,i_2,\cdots,i_k} = 2^{k-1}\prod_{j=1}^{k}\epsilon_{i_j}$.

根据堆积引理，我们知道对于$X_i(i = 1,2,3,4)$，有$\epsilon_i = p_i - \frac{1}{2} = -0.1$, 故$\epsilon_Y = 2^{4-1} \times (-0.1)^4 = 0.0008$. 故处理后序列$P_0 = 0.5008, P_1 = 0.4902$.

（2）序列之间的相关性。

问题可描述成如下：

假设X_1, X_2是同一个随机数发生器输出的连续两比特，Y_1, Y_2是另一个完全相同的随机数发生器输出的连续 2 比特。其中该随机数发生器由于某些原因，会导致连续输出的比特之间存在相关性，亦即X_1, X_2存在相关性，Y_1，Y_2之间也存在相关性。且有$E(X_1) = E(X_2) = E(Y_1) = E(Y_2) = \mu, cov(X_1, X_2) = cov(Y_1, Y_2) = c, corr(X_1, X_2) = corr(Y_1, Y_2) = \rho$。其中，cov 表示协方差，而 corr 表示相关系数。

我们的目的就是证明，将这两个比特序列进行异或之后得到的新比特序列$(X_i \oplus Y_i)$的相关性减弱了。即证明$|corr(X_1 \oplus X_2, Y_1 \oplus Y_2)| < |corr(X_1, X_2)| = |\rho|$.

$Proof$ 计算得到，$E(X_1 \oplus X_2) = E(Y_1 \oplus Y_2)$

$$
\begin{aligned}
& (0 \oplus 0) \times P(X_1 = 0, X_2 = 0) \\
& +(1 \oplus 1) \times P(X_1 = 1, X_2 = 1) \\
& +(1 \oplus 0) \times P(X_1 = 1, X_2 = 0) \\
& +(0 \oplus 1) \times P(X_1 = 0, X_2 = 1) \\
&= P(X_1 = 0, X_2 = 1) + P(X_1 = 1, X_2 = 0) \\
&= E(X_1) + E(X_2) - 2E(X_1X_2) \\
&= E(X_1) + E(X_2) - 2(EX_1 \cdot EX_2) - 2cov(X_1, X_2) \\
&= 2\mu - 2\mu^2 - 2c
\end{aligned}
$$

那么不难计算得到$cov(X_1 \oplus X_2, Y_1 \oplus Y_2) = 4(c^2 + 2c\left(\mu - \frac{1}{2}\right)^2)$，根据协方差和相关系数的公式，我们知道只需要将协方差除以两个随机变量的标准差之积即可。

在这里，我们做一个近似。我们知道服从 0-1 分布的随机变量 Z, 若$E(Z) = p$趋近于$\frac{1}{2}$，那么σ_Z也将趋近于$\frac{1}{2}$。而当前讨论的情况恰好符合这一情形。

另外，也可以证明，在随机变量Z,W的期望趋近于$\frac{1}{2}$时，我们有，$corr(Z,W) = \left(\frac{cov(Z,W)}{\sigma_Z\sigma_W}\right) \approx 4cov(Z,W)$，在本问题中，我们将采取上述两个近似。

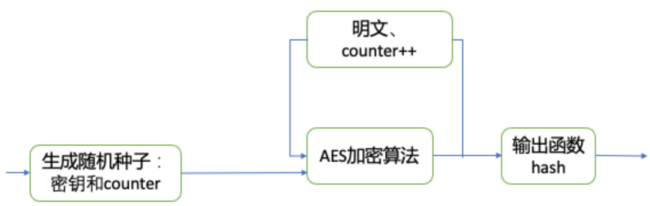
那么有$corr(X_1 \oplus X_2, Y_1 \oplus Y_2) \approx 4cov(X_1 \oplus X_2, Y_1 \oplus Y_2) = 16\left(c^2 + 2c\left(\mu - \frac{1}{2}\right)^2\right) \approx \rho^2 + 8\rho\left(\mu - \frac{1}{2}\right)^2 \approx \rho^2$。前面两个约等号使用的近似已经在前文中说过了，而最后一个约等号的解释如下：当μ趋近于 1/2 时，$\left(\mu - \frac{1}{2}\right)^2$也将趋近于 0，故也是可以忽略的项。

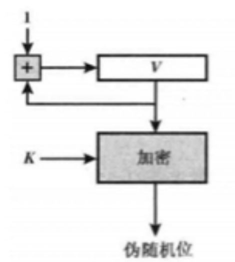
而由于ρ的取值在 0 到 1 之间，所以我们有ρ^2的绝对值小于ρ的绝对值。得证。

3. 请利用学到的设计和检测知识，对第一次课你设计的 RNG 进行分析和改进。

解答：
分析：

我企图设计的 RNG 是基于 AES 算法的 CTR 模式的 DRNG。以下左图是我设计的草图：





(a) CTR模式

而引用课件中基于分组密码的 PRNG 机制中的结构图如上右图。

对比二者，可以发现如下几个不同之处：

1. 加密的内容不同，前者加密时是用明文和计数器的值进行字符串连接。而后者加密时，明文是可选项，即若输入有明文时，将明文和计数器的值异或之后存在计数器中，再对计数器中的内容进行加密。
2. 后者没有关于随机数种子的来源的描述
3. 前者在输出之前进行了 hash 操作，原意是将输出与内部状态进行隔离。防止敌手得到算法的内部状态，但是这是没有必要的，因为加密算法已经保证了这一点，无需再使用 hash 函数进行保证。加了 hash 函数反而造成了不必要的计算，从而浪费资源。

改进：

将 hash 函数删除。