1. Consider a hash table with 26 buckets and 3 slots. Assume that there are 12 distinct keys and each key begins with a capital letter. If we have the hash function h(k) map each key into a bucket using its leading letter and correspond the letters A to Z to the numbers 0 to 25. Please answer the followings questions :
2. Compute the loading factor of this table and explain what is loading factor.
3. Given the keys D, B, GH, GG, B2, C3, C2, C1, A2, B7, C6, GA, please convert them into home buckets and complete the hash table. In addition, you should indicate which keys are **synonyms**, and which keys cause **overflow**.
4. Please propose a hash application. TAs provide the hash application in cryptography as below. Your answer should have the same integrity as TAs’ example.

|  |
| --- |
| [雜湊於密碼學的應用 : 信息驗證]  Hash函數H將可變長度的數據M作為輸入，產生固定長度的Hash值h=H(M)。一個好的Hash函數具有以下特點 : 對於大的輸入集合使用該函數，產生的輸出結果近乎均勻分布(Uniform distribution)且看起來像隨機數。簡而言之，Hash函數的首要目標是保證數據完整性。對於M，任何一點改變都會改變Hash的結果。  信息驗證是用來驗證信息完整性的一種機制。信息認證確保收到的數據確實和發送前一樣(意即在傳輸過程中沒有被修改、插入、刪除或重放)，以下圖為例 :    [解釋] Alice目前想要傳送一封信息給Bob，但為了保證傳送過去的信息是沒有在中途被經過修改的，所以會把要傳送的信息(Data)進行Hash函數處理，得到H(Data)後接在Data後方傳給Bob。當Bob接收到信息後，會將Data的部分取出來做一次Hash函數，再和接在後方的H(Data)比對，若相等則代表信息沒有在途中遭受修改。  Reference : Cryptography and Network Security Principles and Practice, 7th Edition Chapter11 – William Stallings |

1. When implementing a hash table, one of the commonly used collision resolution methods is *open addressing*. Assume that *k* is the key, *m* is the size of the hash table, and each table entry can hold one key.
2. The simplest open addressing method is called linear probing, where if a slot is already occupied, the following slots are “probed” sequentially. Assume an auxiliary hash function *h’(k)=k* mod *m*. Please give the hash function *h(k,i)* of linear probing with the given auxiliary function, where *i* is the number of times the hash table has been probed with this key (and thus *i* always starts from 0).

|  |
| --- |
| Definition of Linear Probing :  *h(k,i)=(h’(k)+i) mod m* |

1. Linear probing is easy to implement but suffers from the problem of primary clustering. A more complex open addressing method which can mitigate this problem is double hashing, using a hash function in the form of *h(k,i)=(h1(k)+ih2(k))* mod *m*, with two auxiliary hash functions *h1(k)* and *h2(k)*. The number *i* is the number of times the hash table has been probed with the key *k*. Assume two auxiliary hash functions *h1(k)=k* mod 16, *h2(k)=1+(k mod 15)*. Show the final hash table content after inserting all of the following keys to an initially empty hash table in the given order: {16,3,35,67,51,1,15,31,19,17}.
2. Prof. Alpha proposes a different double hashing method, using two auxiliary hash functions *h1(k)=k* mod *16* and *h2(k)=2(k mod 8)*. Assume the table size m is 16. Please give an example to illustrate why Prof. Alpha’s method is problematic.
3. Consider a hash function *h(k)=k%D*, where D is to be defined. We wish to find a proper D value using minimal number of attempts, while each attempt requires supplying the function with k and observing the result of h(k). Indicate how this may be achieved in the following two cases.
4. D is known to be a prime number in the range [10, 20].
5. D is of the form 2k, where k is an integer in [1, 5].

5. Suppose that you are to design a Bloom filter with minimum P(u) with n=100000, m=5000, and u=1000.

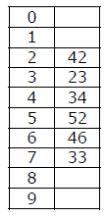
a. Use any of the results obtained in the textbook, and determine the number

of hash functions required. Show your computations.

b. What is the probability, P(u), of a filter error with the required number of

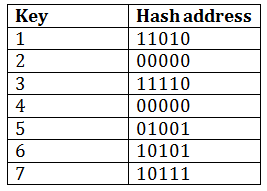
hash functions computed above?

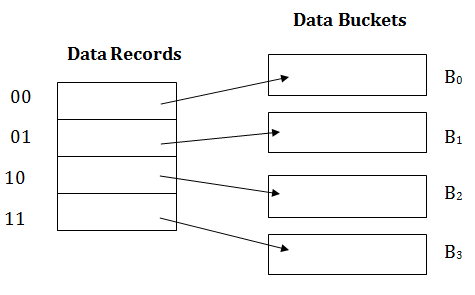
1. A hash table of length 10 uses open addressing with the hash function h(k)=k mod 10, and linear probing. After inserting 6 values into an empty hash table, the table is as shown below.



Which one of the following choices gives a possible order in which the key values could have been inserted in the table?  
(A) 46, 42, 34, 52, 23, 33  
(B) 34, 42, 23, 52, 33, 46  
(C) 46, 34, 42, 23, 52, 33  
(D) 42, 46, 33, 23, 34, 52

1. Which one of the following hash functions on integers will distribute keys most uniformly over 10 buckets numbered 0 to 9 for i ranging from 0 to 2020?   
   (A) h(i) =i^2 mod 10  
   (B) h(i) =i^3 mod 10  
   (C) h(i) = (11 ∗ i^2) mod 10  
   (D) h(i) = (12 ∗ i) mod 10
2. The dynamic hashing is used to overcome the problem of static hashing like bucket overflow without resulting in poor performance. Please answer the questions below.
3. Please discuss the strengths and weaknesses of dynamic hashing.
4. Consider the following grouping of keys mapped into buckets, depending on the prefix of their hash address. Please fill up the data buckets.





1. If we need to insert key 9 with hash address 10001 into the above structure, what is the result of the insertion. (Please draw the similar structure as the second figure)