CSCI 3110 Hash function (2)

- **Factors to consider when analyzing hashing methods**
  - **Load factor :** $\alpha = \frac{N}{tableSize}$
  - **Size of the hash table**
  - **Successful search or not**

- **Comparing the four collision resolution approaches:**
  - Best case O(1)
  - Worse base O(n)

	Linear Probing	**α=1/2**	**α=2/3**	**α=7/8**
average case	$\frac{1}{2}\left[1+\frac{1}{1-\alpha}\right]$ for a successful search	1.5	2	4.5
	$\frac{1}{2}\left[1+\frac{1}{(1-\alpha)^2}\right]$ for an unsuccessful search	4.5	8	40.5

Quadratic Probing and Double Hashing	**α=1/2**	**α=2/3**	**α=7/8**
$\frac{-\log e(1-\alpha)}{\alpha}$, for a successful search	1.38	1.65	3.1
$\frac{1}{1-\alpha}$, for an unsuccessful search	2	3	8

Separate Chaining	**α=1/2**	**α=2/3**	**α=7/8**
$1+\frac{\alpha}{2}$, for a successful search	1.25	1.33	1.43
α, for an unsuccessful search	0.5	0.67	0.875

Conclusions and Discussions:

- Typically, α of a hash table should be kept below 2/3.
- Empirical comparisons of the four collision resolution methods show
  When α is 0.5, all four systems are about the same. As α approaches 1, separate chaining is the clear winner
- Criteria for good hashing function
  - Easy and fast to compute
  - Scatter the data evenly throughout the hash table
    - The calculation of the hash function should involve the entire search key

        - If the hash function uses modulo arithmetic, the base should be prime → the choice of table size as a prime number safeguard against many subtle kinds of patterns in the data.
- Comparing hashing implementation and balanced tree implementation of table:
    - if α can be kept small, then hashing is a better approach than other methods in terms of insertion/deletion/retrieval operations. Otherwise, a balanced binary tree implementation is more reliable (guaranteed lower bound performance)
    - Operations that make hashing a less efficient implementation than balanced search tree implementation:
        - Traverse in sorted order of search key → hash table does not support ordering at all!
        - Retrieval of record with the largest/smallest search key
        - Range query