Assignment –4

Discussion on Hash Collisions

Impact on Performance

- **Hash Collisions**: In the Rabin-Karp algorithm, a hash collision occurs when two different substrings have the same hash value. This leads to false positives, where the hash values match but the actual substrings do not.
- **Performance**: While hash collisions do not affect the worst-case time complexity of O(m * n) in the worst case, they can increase the number of character comparisons in practice. Frequent collisions can degrade the algorithm's performance, making it closer to the naive approach in efficiency.

Handling Hash Collisions

- **Verification Step**: The algorithm handles hash collisions by performing a character-by-character comparison only when the hash values match. This additional step ensures that false positives are identified and discarded.
- **Choosing a Good Hash Function**: Using a large prime number (e.g., q = 101) for modulus operation helps in distributing hash values more uniformly and reducing collisions.
- **Double Hashing**: An advanced technique involves using two different hash functions. If both hash values match, then a character-by-character comparison is performed.

By incorporating these strategies, the Rabin-Karp algorithm maintains its average-case efficiency and handles hash collisions effectively, ensuring reliable substring search performance.