7 Cybersecurity (fms27)

Some naïve teenagers want to attack a web application that does not salt its users' passwords, but simply hashes them. Hoping they will someday obtain the file of digests on the dark web, they want to precompute a compressed lookup table that will let them quickly crack all lowercase alphanumeric passwords of up to 15 characters once they get the file. They implement precomputed hash chains, but they ignore the possibility of collisions because they do not understand the issue. Their table must fit in their available disk space s. Each digest is 16 bytes long.

Please use the following symbol names and use sensible approximations where needed.

l_a	length of alphabet (charset size) for targeted passwords	$26 + 10 = 36$
l_d	length of one digest, in bytes	16
l_p	length of password, in bytes (max len to be explored)	15
s	storage space available for the compressed table, in bytes	8×10^{12}
t_h	time to compute a hash on attackers' CPU, in seconds	10^{-5}

(a) Under the attackers' incorrect assumption that hash collisions may be ignored:

(i) Derive formulae for the maximum number n_c of chains in the table, and for the minimum length l_c (in passwords) of each chain. Also calculate numerical values for these quantities. *(4 results.)* [4 marks]

(ii) Give a clear and full description of the algorithm for recovering the password that is the preimage of a given digest d. Pseudocode is not required for full marks, but will be rewarded if it adds clarity and precision. [5 marks]

(iii) Derive a formula for the minimum number n_h of hashes to be computed to build the compressed table, and calculate its numeric value. *(2 results.)* [2 marks]

(iv) Use the n_h value from Part (a)(iii) to calculate the time t_s needed to compute the whole table sequentially, assuming one hash computation takes time t_h on the attackers' CPU. Then imagine parallelising the work in three ways: moving from CPU to GPU, using a bank of GPUs, or using a distributed pool of collaborators. Give a reasonable speedup factor for each. Compute how long it would take to build the compressed table if all three speedups were adopted. *(5 results.)* [5 marks]

(b) Briefly explain what hash collisions are and why they matter. How would the presence of collisions invalidate the answer to Part (a)(iii)? [4 marks]