Ques 2 - Birthday Attack

Algorithm:

For d digits, there are a maximum of 2^d possible hashes.
For a plaintext p in P, we have p -> {0, 1}^d.
If the number of plaintexts in P is greater than 2^d, there will always be at least two strings x and y such that hash(x) = hash(y).

For plain text, I have assumed that the strings can have a-z, A-Z, 0-9. This gives me 62 characters.

Using the birthday paradox, I have worked with a set of at least N = 2^((d+1)//2). Note that because of this (d+1)//2, the size of set generated randomly will be same for 2n and 2n-1, and so the memory used in their runs is going to be the same.

For getting the length L of the strings to be used (to reduce memory usage, I have worked with the least bound I was able to find), I have used: 62^L > N

That is, the smallest length which gives me a string set of size greater than the one I found using the birthday paradox.

Now, I generate a random set of N strings of length L. Check if hashes of two strings come out to be equal (for first d bits). If yes, report the 2 strings, else repeat.

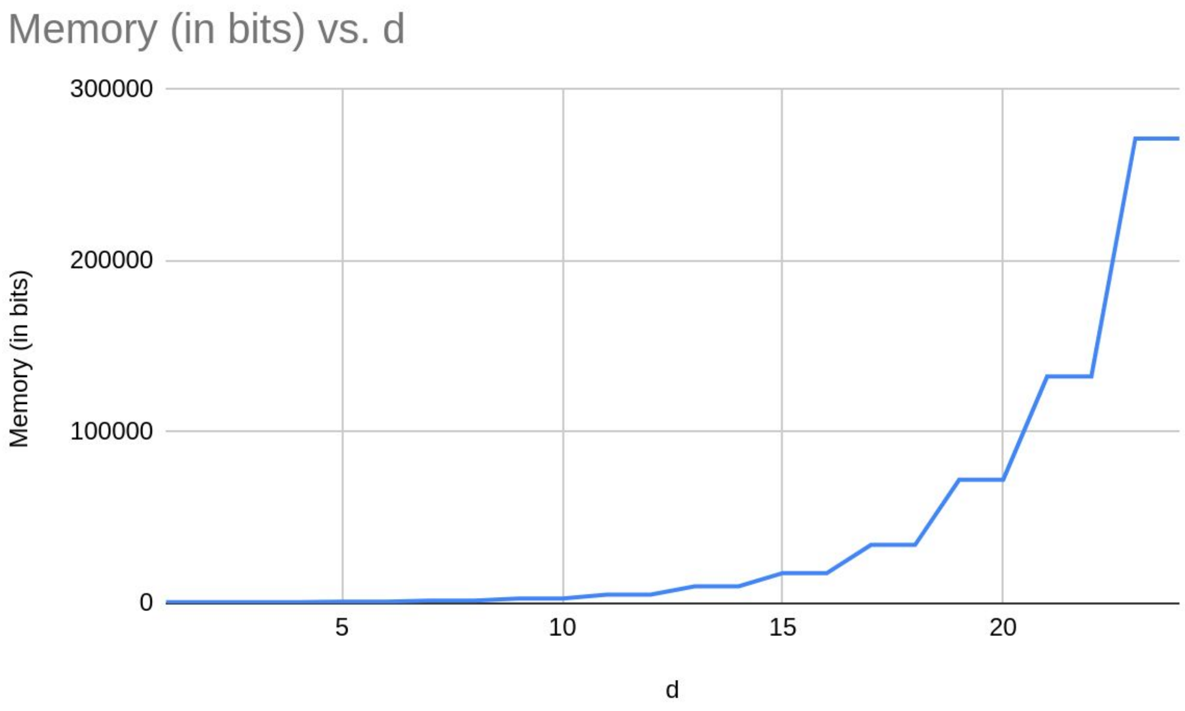
For the purpose of reproducibility, I have seeded the random generator with 0.

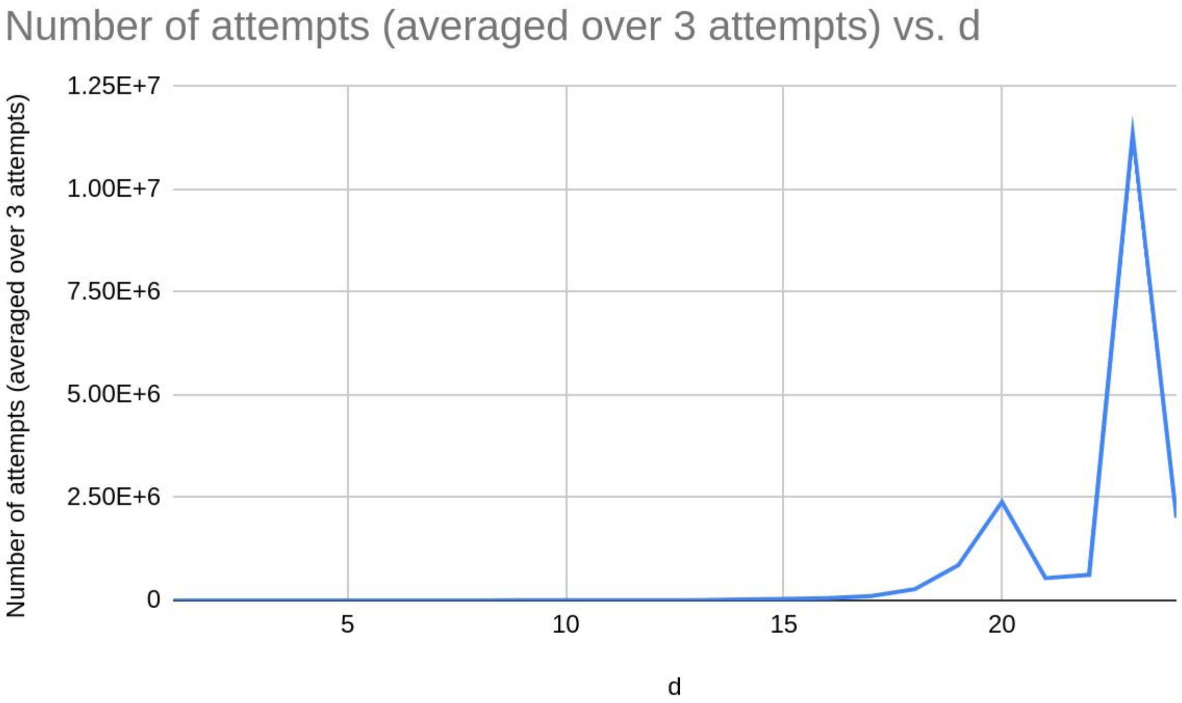
Results:

The values of d, memory and avg attempts are as below:

d	String 1	String 2	Memory (in bits)	No of attempts (averaged over 3 runs)
1	2	y	768	2
2	A	c	768	4
3	6	Y	768	13
4	5	n	768	14
5	Gi	NZ	1024	36
6	Ue	NO	1024	64
7	S7	FQ	1536	75
8	2S	hO	1536	179

9	IB	tw	2752	699
10	G6	po	2752	959
11	qiP	gJ6	5120	2537
12	zvr	7pL	5120	3755
13	XOO	emT	9984	5243
14	9v4	2ML	9984	25210
15	zpx	x5Z	17728	34960
16	P4t	LG4	17728	51151
17	2K8k	U5SH	34176	103260
18	pzW9	hhhz	34176	270993
19	xLAn	WZ8q	72192	854385
20	ex5X	qANj	72192	2396724
21	OTCl	xR19	132480	541913
22	xCnT	63ul	132480	621259
23	ECwDT	aiTtg	271424	11330112
24	ifjVb	AiDUW	271424	2008469




Number of attempts (averaged over 3 attempts) vs. d
Number of attempts (averaged over 3 attempts)
1.25E+7
1.00E+7
7.50E+6
5.00E+6
2.50E+6
0
5
10
15
20
d