

CIS4517/5517 Quiz 5:

1. (3 points) What are the three major sources in cloud computing that cause security problems?

Loss of control

Lack of trust

Multi-tenancy

2. (3 points) True or false: determine whether the following statement is true or false. If you think it's false, please provide a brief justification.

In the Amazon cloud, two instances that run from distinct accounts at roughly the same time are often assigned to the same physical machine, this is called sequential placement locality.

False,

parallel placement locality

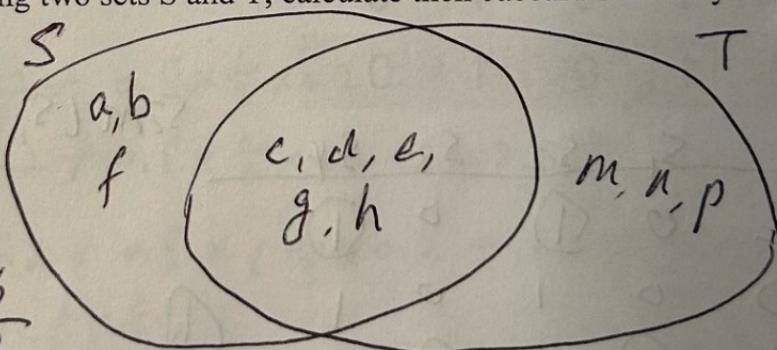
3. (6 points) Given the following two sets S and T, calculate their Jaccard similarity and Jaccard distance.

$$S = \{a, b, c, d, e, f, g, h\}$$

$$T = \{c, d, e, g, h, m, n, p\}$$

$$\text{sim} = \frac{|S \cap T|}{|S \cup T|} = \frac{5}{11}$$

$$\text{dist} = 1 - \text{sim} = \frac{6}{11}$$



4. (6 points) List all 12-shingles in the following phrase "Cloud Security" where the whitespace is considered as a valid character.

Cloud Securi
loud securit
oud security

5. (12 points) Given the following input matrix of four documents S1, S2, S3, and S4,

- Calculate the similarity between documents S1 and S3.
- Using the min-hashing approach, find the signature for random permutation {CDAEBFG}

Please show your work.

	S₁	S₂	S₃	S₄
A	1	0	1	0
B	1	0	0	1
C	0	1	0	1
D	0	1	0	1
E	0	1	0	1
F	1	0	1	0
G	1	0	1	0

$$\text{a) } \begin{array}{l} \text{Sim}(S_1, S_3) = \\ S_1 \text{ AND } S_3 = \end{array}$$

1 1 0 0 0 1 1
1 0 0 0 0 1 1

1 0 0 0 0 1 1

(3 ones)

$$S_1 \text{ OR } S_3 = \begin{array}{l} 1 1 0 0 0 1 1 \\ 1 0 0 0 0 1 1 \\ ----- \\ 1 1 0 0 0 1 1 \end{array}$$

$$\text{sim}(S_1, S_3) = \frac{3}{4} \quad (4 \text{ ones})$$

	<u>S₁</u>	<u>S₂</u>	<u>S₃</u>	<u>S₄</u>
1 C	0	1	0	1
2 D	0	1	0	1
3 A	1	0	1	0
4 E	0	1	0	1
5 B	1	0	0	1
6 F	1	0	1	0
7 G	1	0	1	0

$$\text{b) signature: } 3 \ 1 \ 3 \ 1$$

sig: 3 1 3 1