Privacy-Preserving Search In Hamming Space

Alice Wang

28 April 2015

Where Errors Occur

- The substring matching error is determined by the error of equality tests with the randomized protocol based on primes:
  - FRR(i.e., x=y but output x!=y): 0%; FAR: $\left(\frac{\ln D^2}{D}\right)^t for\ D \geq 9$, e.g., 1.3% for D=1000 and t=1; 0.019% for D=1000 and t=2
  - Solution: increases the number of primes t and uses prime match only when the segment size if sufficiently large

- Beyond 32 bits (default in matlab), a binary string is converted into a large decimal only in approximation rather than in exact
  - E.g., for two 625-bit binary strings bin_x and bin_y
    - pdist2(bin_x, bin_y, 'hamming')*625 = 11 (Hamming distance)
    - bi2de(bin_x) = 1.3609e+188 (large decimal after conversion)
    - bi2de(bin_y) = 1.3609e+188 (large decimal after conversion)
  - Solution: set the maximum size of each segment to be 32 bits. That is, there should be a lower bound for the number of partitions $l \geq \lfloor D/31 \rfloor + 1$

Simulation I

D = 10000; N = 130;

Q1 ground truth ID (Hamming distance)

1(0) 92(60) 97(129) 95(146) 96(186) 100(187) 93(191) 98(216) 91(245) 94(253) 99(295)

Q2 ground truth (Hamming distance)

2(0) 104(51) 106(55) 103(58) 108(112) 101(116) 107(164) 105(175) 110(181) 109(259) 102(294)

Q3 ground truth (Hamming distance)

3(0) 112(31) 114(34) 116(81) 111(109) 117(137) 118(151) 119(180) 120(193) 113(226) 115(260)

Q4 ground truth (Hamming distance)

4(0) 127(47) 124(49) 130(74) 128(91) 129(120) 125(171) 126(191) 122(253) 121(288) 123(289)

- Elapsed time is 0.205712 seconds.

Proposed algorithm: s>=3 prime match, t=5, l = $\lfloor D/31 \rfloor$+1 = 323, k=10:40:300

Q1 NN ID found (NN radius k):

1(10) 92(90) 97(130) 95(170) 93(210) 96(210) 100(210) 91(250) 98(250) 94(290) 0(0)

Q2 NN ID found (NN radius k):

2(10) 103(90) 104(90) 106(90) 101(130) 108(130) 107(170) 105(210) 110(210) 109(290) 0(0)

Q3 NN ID found (NN radius k):

3(10) 112(50) 114(50) 116(90) 111(130) 117(170) 118(170) 119(210) 120(210) 113(250) 115(290)

Q4 NN ID found (NN radius k):

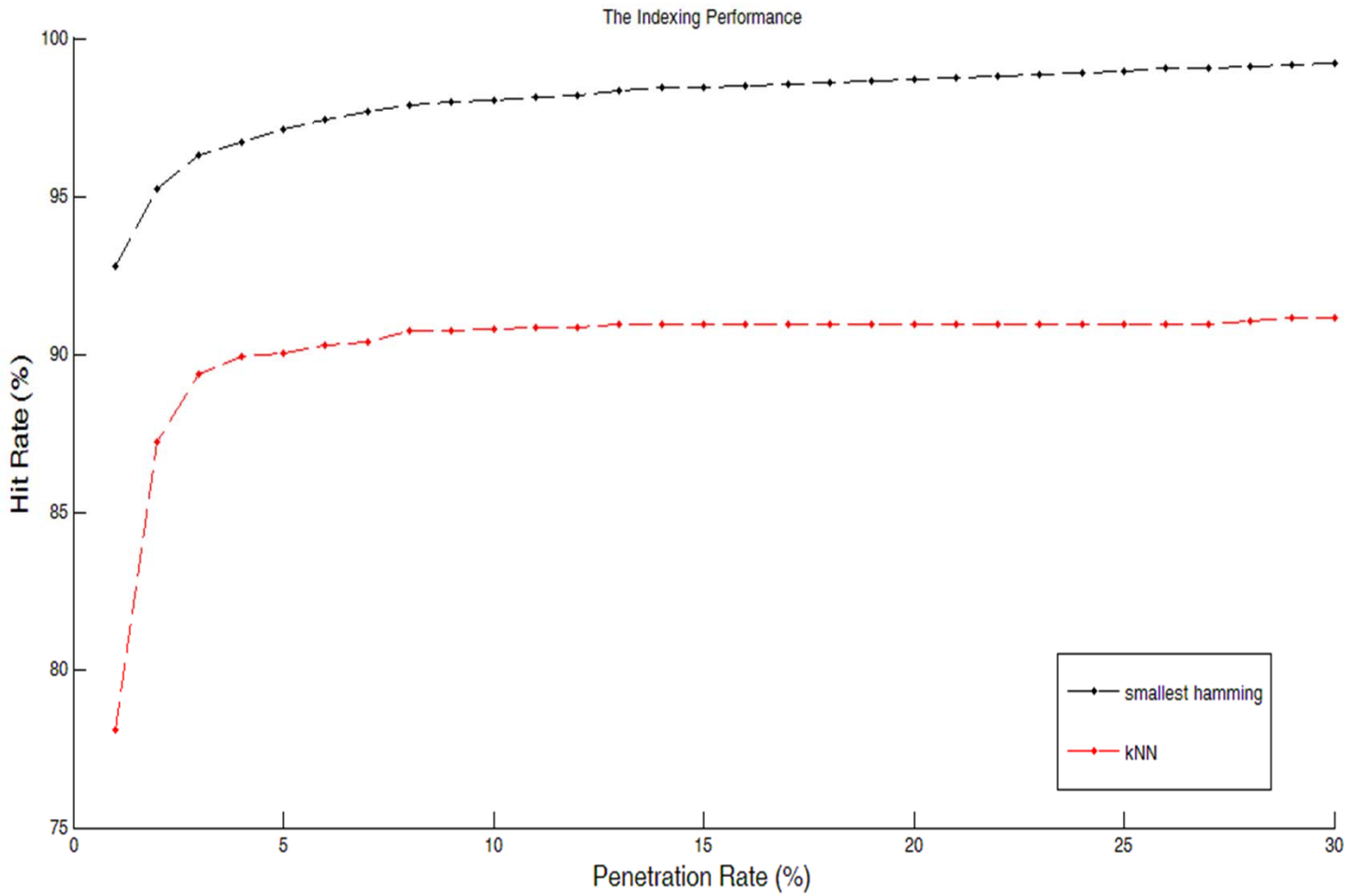
4(10) 124(50) 127(50) 130(90) 128(130) 129(130) 125(210) 126(210) 121(290) 122(290) 123(290)

- Elapsed time is 310.382668 seconds.

Within the same radius k, the instance similarities are distinguishable in our method!

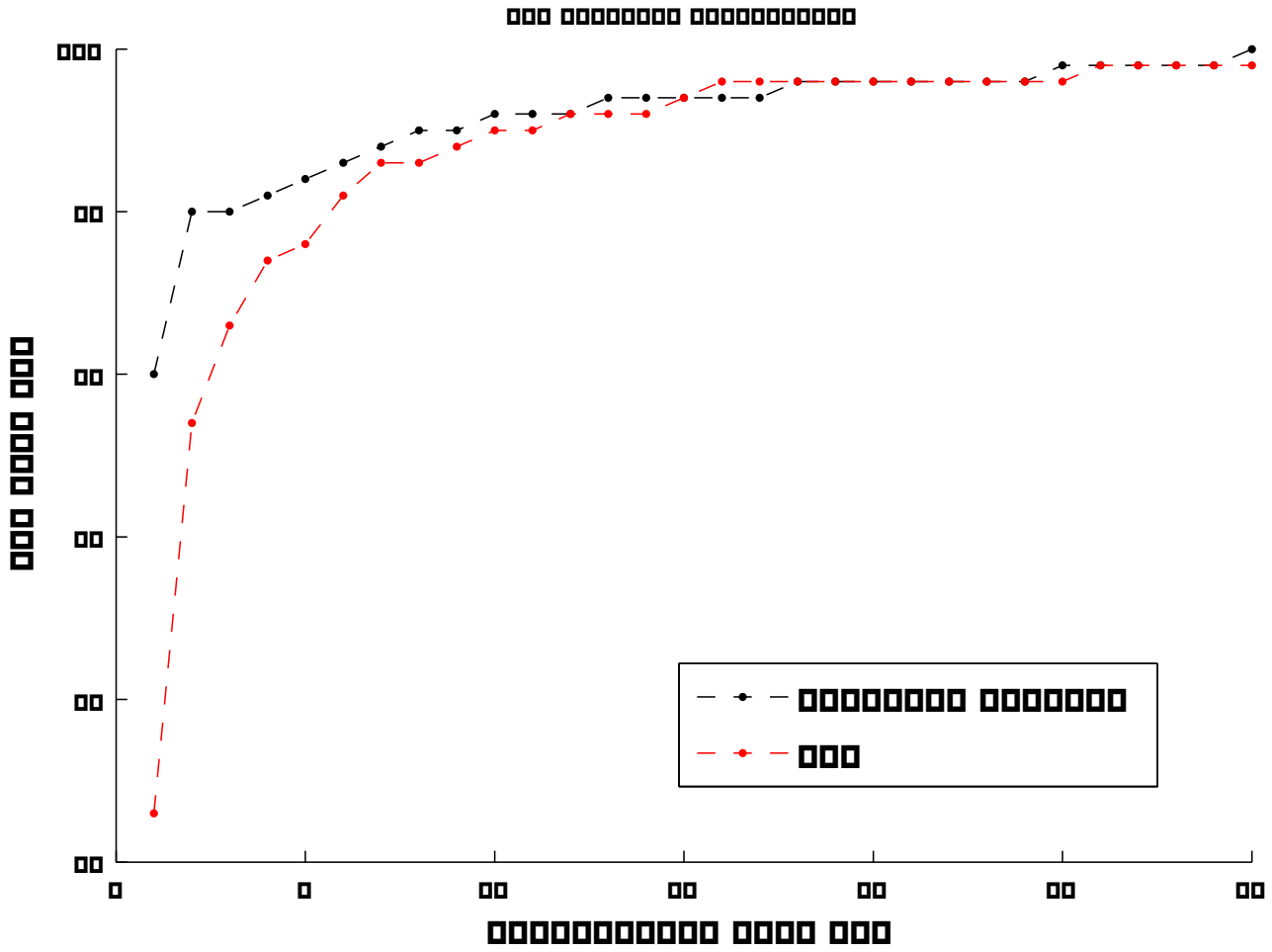
Face Experiments

- FERET LSSC @ 448bits, K = [100, 150]



Face Experiments

- K = [50, 80, 110, 120, 125, 130, 135, 150]



Simulation II

- Simulating 2048-bit IrisCodes each with a 2048-bit validity mask

$$d = \frac{|X \oplus Y \cap M_X \cap M_Y|}{|M_X \cap M_Y|} = \frac{|(X \cap M_X) \oplus (Y \cap M_Y)|}{|M_X \cap M_Y|}$$

$$= \frac{1}{w} \times H(X', Y')$$

So, $H(X', Y') \leq k = wd$

- Multi-Party Security Computation (MPC) by garbled circuits to evaluate the weight $w = |M_X \cap M_Y|$

Simulation II Results

>> knnMain_simu_iris_probeset: D = 2048, N = 130

Q1 ground truth (Hamming distance)

1(61) 94(113) 97(119) 96(131) 95(146) 91(152) 98(233) 99(242) 100(321)

Q2 ground truth (Hamming distance)

2(91) 101(158) 110(163) 102(181) 107(199) 103(204)

Q3 ground truth (Hamming distance)

118(107) 3(113) 113(116) 119(150) 120(162) 116(225) 115(266) 111(274) 112(282) 114(347)

Q4 ground truth (Hamming distance)

4(64) 124(130) 121(156) 128(214) 130(218) 127(237) 126(297)

- Elapsed time is 0.131673 seconds.

Proposed algorithm: s>=3 prime match, t=5, l = 67, d = 0.05:0.015:0.2

Q1 NN ID found (NN radius k):

1(97) 94(126) 97(126) 95(155) 96(155) 91(184) 98(242) 99(271) 100(329)

Q2 NN ID found (NN radius k):

2(92) 101(175) 110(175) 102(203) 103(230) 107(230)

Q3 NN ID found (NN radius k):

3(126) 113(126) 118(126) 119(154) 120(183) 116(241) 115(270) 111(299) 112(299) 114(386)

Q4 NN ID found (NN radius k):

4(95) 124(152) 121(180) 127(237) 128(237) 130(237) 126(294)

- Elapsed time is 694.422704 seconds.