Protect both Integrity and Confidentiality in Outsourcing Collaborative Filtering Computations

Balázs Pejó, Qiang Tang, Husen Wang

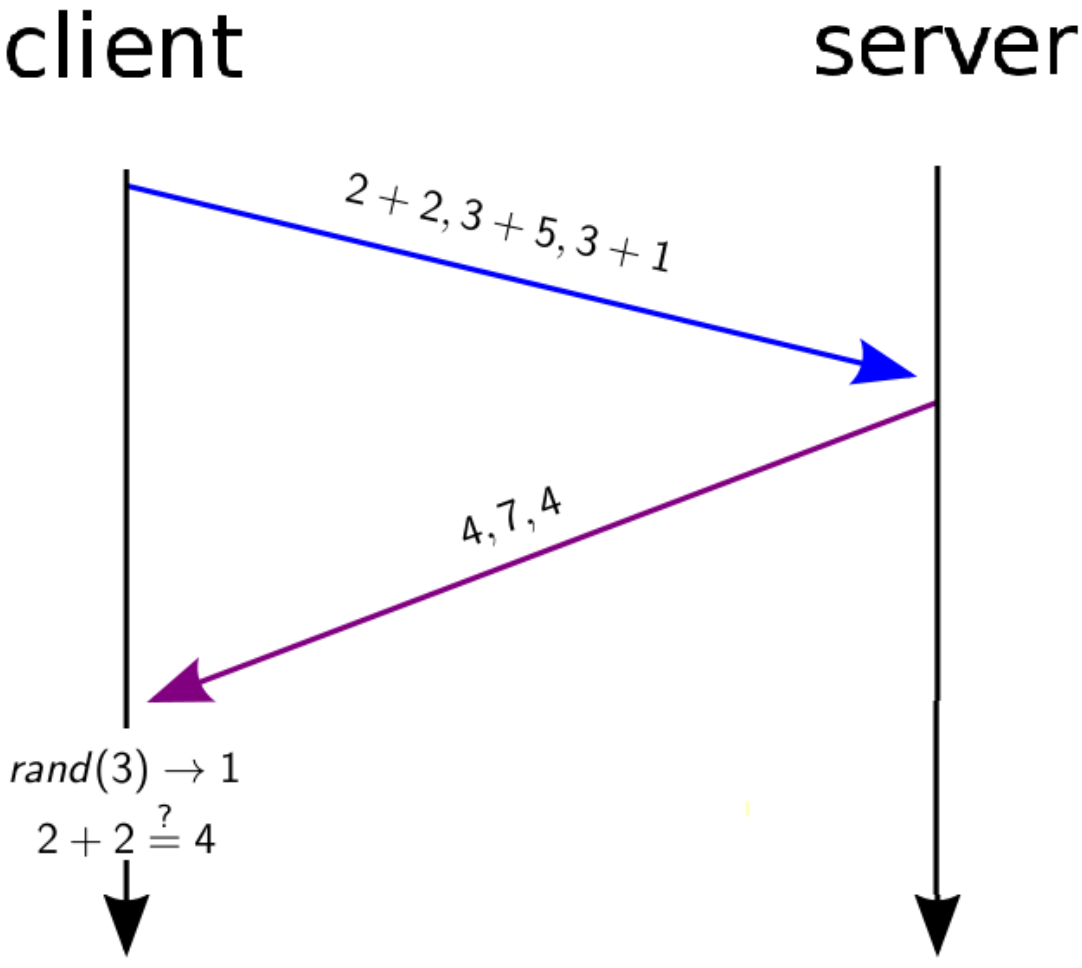
27th Jun 2016
IEEE CLOUD, San Francisco

uni.lu
UNIVERSITÉ DU LUXEMBOURG

SnT
securityandtrust.lu

Outline

Outsourcing



Weighted Slope One

- $r_{u,i}$: Rating value from user u for item i.
- $q_{u,i}$: Indicator, whether user u rated item i or not.

Computational Stage

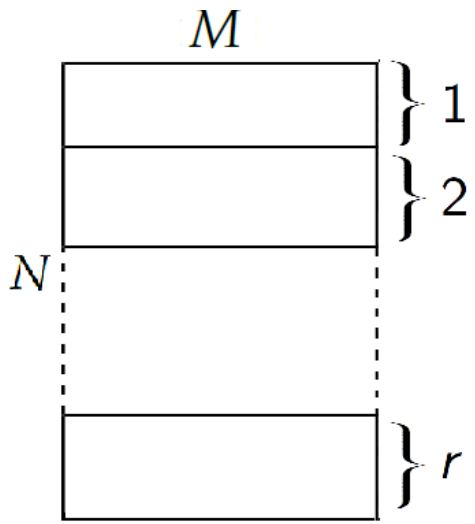
$$\phi_{i,j} = \sum_u q_{u,i} q_{u,j} \qquad \delta_{i,j} = \sum_u q_{u,i} q_{u,j} (r_{u,i} - r_{u,j})$$

Prediction Stage

$$p_{u,i} = \frac{\sum_j \delta_{i,j} + r_{u,j} \phi_{i,j}}{\sum_j \phi_{i,j}}$$

Section 2

1. Introduction
2. Splitting Approach
3. Auxiliary Data Approach
4. 2 Server Setting



Strategies

- Server: Randomly selects γ blocks, and sets ζ percent of the corresponding output random.
- Client: In every block, it randomly chooses θ values to verify.
- Cheating rate: $\rho = \frac{\gamma\zeta}{r}$
- Verification cost: $r\theta$
- Detection rate:

$$P_d = 1 - \left(\frac{\binom{\frac{M(M-1)(1-\zeta)}{2}}{\theta}}{\binom{\frac{M(M-1)}{2}}{\theta}} \right)^{\gamma}$$

Detection rates

$r = 1$

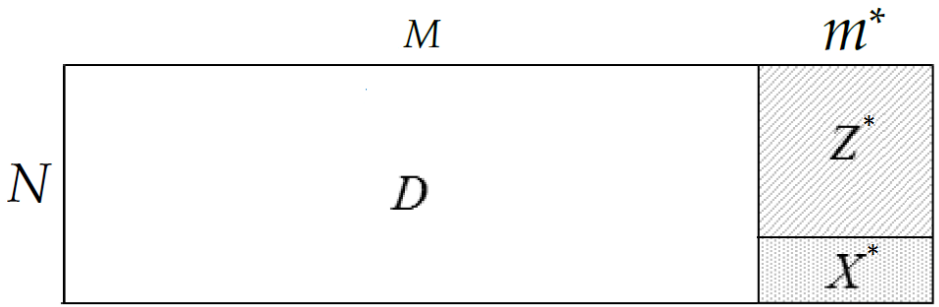
θ \ ρ	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-6}	2^{-9}
10	0.9990	0.9437	0.7369	0.4755	0.1457	0.0194
20	1.0000	0.9968	0.9308	0.7249	0.2702	0.0383
40	1.0000	1.0000	0.9952	0.9243	0.4674	0.0752
100	1.0000	1.0000	1.0000	0.9984	0.7930	0.1776
200	1.0000	1.0000	1.0000	1.0000	0.9571	0.3236

$r = 10$

θ \ ρ	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-6}	2^{-9}
1	1.0000	0.9954	0.8594	0.6240	0.1239	0.0035
2	1.0000	1.0000	0.9802	0.8594	0.2757	0.0125
4	1.0000	1.0000	0.9996	0.9802	0.4923	0.0412
10	1.0000	1.0000	1.0000	0.9999	0.8171	0.1537
20	1.0000	1.0000	1.0000	1.0000	0.9666	0.3194

Section 3

1. Introduction
2. Splitting Approach
3. Auxiliary Data Approach
4. 2 Server Setting



Strategies

- Server: Randomly sets ρ percent of output values to random.
- Client: Verifies the values corresponding to m^* which is $\approx \frac{m^{*2}}{2}$
- Detection rate:

$$P_d = 1 - \frac{\binom{\frac{(M+m^*)(M+m^*-1)}{2} - \frac{m^*(m^*-1)}{2}}{\frac{\rho(M+m^*)(M+m^*-1)}{2}}}{\binom{\frac{(M+m^*)(M+m^*-1)}{2}}{\frac{\rho(M+m^*)(M+m^*-1)}{2}}}$$

Detection rates

m^* \ ρ	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-6}	2^{-9}
10	1.0000	1.0000	0.9975	0.9452	0.5077	0.0842
20	1.0000	1.0000	1.0000	1.0000	0.9498	0.3103
40	1.0000	1.0000	1.0000	1.0000	1.0000	0.7824

Comparison with the (no) splitting method (e.g. $r = 1$)

m^* \ ρ	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-6}	2^{-9}
$m^* = 20$	1.0000	1.0000	1.0000	1.0000	0.9498	0.3103
$\theta = 200$	1.0000	1.0000	1.0000	1.0000	0.9571	0.3236

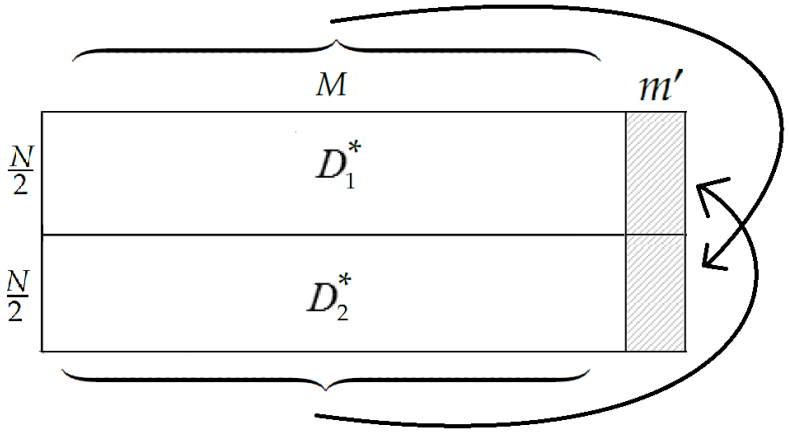
Verification Cost in seconds

via Splitting	via Auxiliary Data
0.0570	0.0054

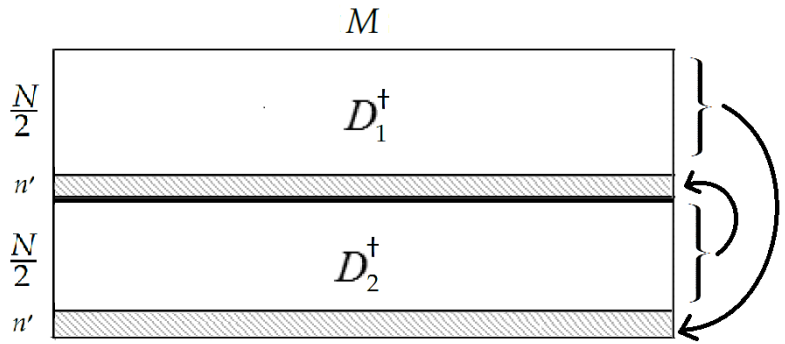
Section 4

Verification method

Computational Stage



Prediction Stage



Detection probabilities

Computational s. det.: $P_d = 1 - f(\rho)$

$$f(\rho) = \frac{\binom{\frac{(M+m')(M+m'-1)}{2} - m'(m'-1)}{\rho\frac{(M+m')(M+m'-1)}{2}}}{\binom{\frac{(M+m')(M+m'-1)}{2}}{\rho\frac{(M+m')(M+m'-1)}{2}}}$$

Prediction s. det: $P_d = 1 - g(\rho)$

$$g(\rho) = \frac{\binom{(1-d)M(\frac{N}{2} - n')}{\rho(1-d)M(\frac{N}{2} + n')}}{\binom{(1-d)M(\frac{N}{2} + n')}{\rho(1-d)M(\frac{N}{2} + n')}}$$

Detection rates

Computational Stage

m' \ ρ	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-6}	2^{-9}
2	0.7500	0.4375	0.2344	0.1211	0.0310	0.0039
4	0.9998	0.9683	0.7986	0.5390	0.1722	0.0232
10	1.0000	1.0000	1.0000	0.9970	0.7576	0.1613
20	1.0000	1.0000	1.0000	1.0000	0.9975	0.5243

Prediction Stage

n' \ ρ	2^{-9}	2^{-10}	2^{-11}	2^{-12}	2^{-13}	2^{-14}
1	1.0000	1.0000	1.0000	0.9998	0.9862	0.8827
2	1.0000	1.0000	1.0000	1.0000	0.9998	0.9862
4	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998

Conclusion

- Auxiliary data verification method outperforms the splitting approach.
- Proposed a new, more efficient verification method using two servers which can be used for the prediction stage too.