AN ITERATIVE ALGORITHM FOR TRUST AND REPUTATION MANAGEMENT

Hu Hao

Jan 27, 2015

NANYANG TECHNOLOGICAL UNIVERSITY

References:

- 1. Ayday, E.; Hanseung Lee; Fekri, F., "An iterative algorithm for trust and reputation management," *Information Theory, 2009. ISIT 2009. IEEE International Symposium on* , vol., no., pp.2051,2055, June 28 2009-July 3 2009
- 2. Ayday, E.; Hanseung Lee; Fekri, F., "Trust management and adversary detection for delay tolerant networks," *MILITARY COMMUNICATIONS CONFERENCE, 2010 - MILCOM 2010* , vol., no., pp.1788,1793, Oct. 31 2010-Nov. 3 2010
- 3. Ayday, E.; Fekri, F., "Iterative Trust and Reputation Management Using Belief Propagation," *Dependable and Secure Computing, IEEE Transactions on* , vol.9, no.3, pp.375,386, May-June 2012

NANYANG TECHNOLOGICAL UNIVERSITY

ITRM (iterative trust reputation mechanism)

Background:

- In the environments of online communities, web services, ad-hoc networks, P2P computing and e-commerce communities, the recipient of the service has no choice but to rely on the reputation of the service provider based on the latter's prior performance.

Goals:

- The scheme is robust in filtering out the peers who provide unreliable ratings.

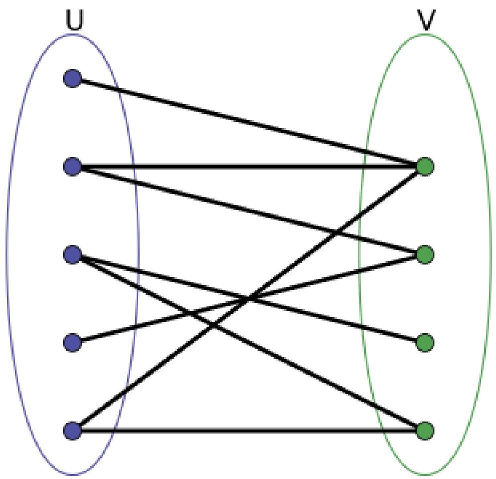
Adversary:

- Bad-mouthing: malicious raters collude and attack the service providers with the highest reputation by giving low ratings
- Ballot-stuffing: malicious raters collude to increase the reputation value of peers with low reputations

NANYANG TECHNOLOGICAL UNIVERSITY

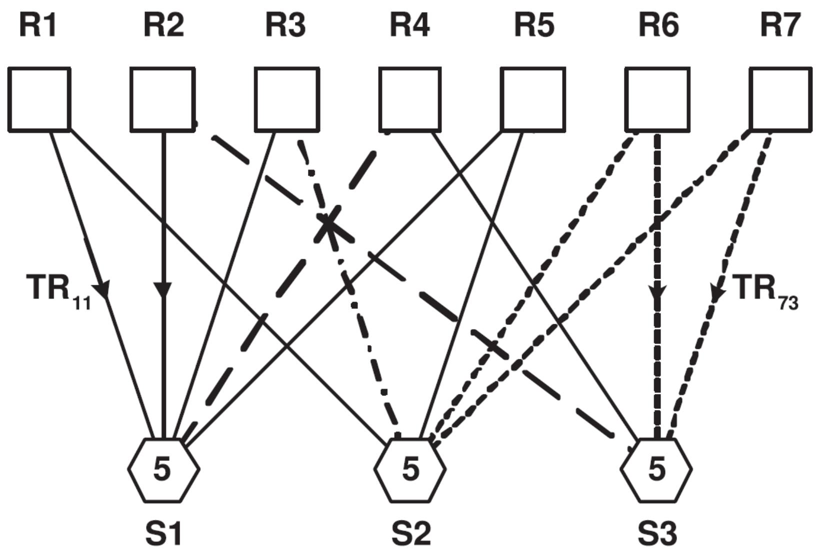
Bipartite Graph

- In the mathematical field of graph theory, a **bipartite graph** (or **bigraph**) is a graph whose vertices can be divided into two disjoint sets U and V such that every edge connects a vertex in U to one in V.



Illustrative example of ITRM

Every check-vertex has some opinion of what the value of each bit-vertex should be.



Check vertices(rater-peer)

Bit vertices (service provider)

$$WR_{ij} = w_{ij} \bullet TR_{ij}$$

$$w_{ij} = \lambda^{t-t_{ij}}$$ Age-factor

ν Iteration times

$$TR_j^{\nu} = \frac{\sum_{i \in A} R_i \times WR_{ij}^{\nu}}{\sum_{i \in A} R_i \times w_{ij}(t)}$$

NANYANG TECHNOLOGICAL UNIVERSITY

Then, we compute the inconsistency factor of each check-vertex *i,* using values of bit vertex, B is the set of bit-vertex which *i* has connect to

$$C_i^{\nu} = \left[1 \Big/ \sum_{j \in B} \hat{\lambda}^{t - t_{ij}} \right] \sum_{j \in B} d\left(TR_{ij}^{\nu-1}, TR_j^{\nu-1}\right)$$

d(,) is the distance metric used to measure the inconsistency

$$d\left(TR_{ij}^{\nu-1}, TR_j^{\nu-1}\right) = \left| TR_{ij}^{\nu-1} - TR_j^{\nu-1} \right| \hat{\lambda}^{t - t_{ij}}$$

Check vertex *i* with highest inconsistency, place it in the blacklist if the inconsistency is greater than threshold τ

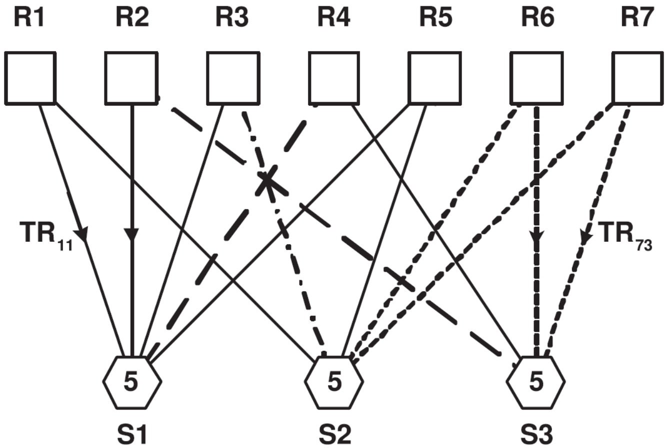
The iteration stops if there is no vertex with inconsistency greater than τ

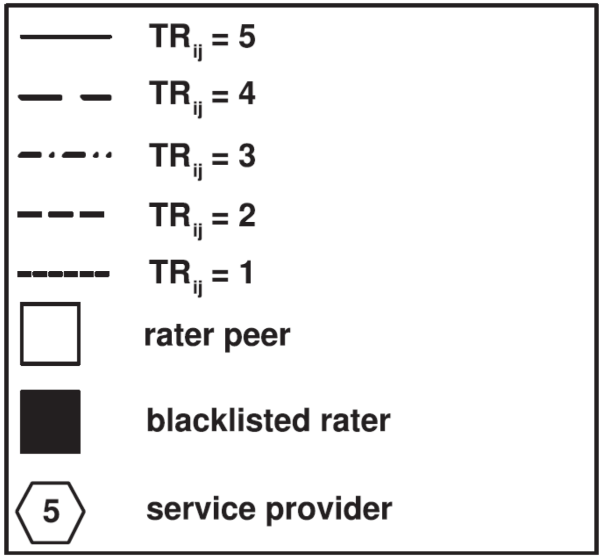
NANYANG TECHNOLOGICAL UNIVERSITY

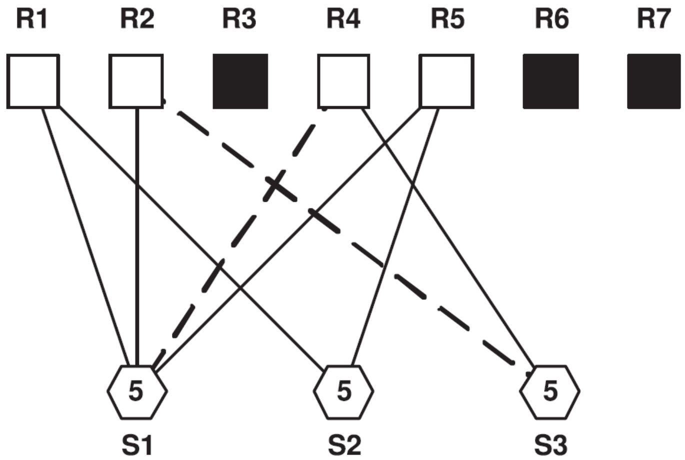
Example

$\tau = 0.7$

$TR_i = 5$







Iteration	TR_1	TR_2	TR_3
0	4.8	3	2.75
1	4.8	3.5	3.33
2	4.8	4.33	4.5
3	4.75	5	4.5

Iteration	C1	C2	C3	C4	C5	C6	C7
0	1.1	.72	.10	1.52	1.1	1.87	1.87
1	.85	.43	.35	1.23	.85	2.42	-
2	.43	.35	.77	.65	.43	-	-
3	.12	.38	-	.63	.12	-	-

NANYANG TECHNOLOGICAL UNIVERSITY

From the example ,we can see:

1. ITRM gives better estimation of TR_j's compared to the weighted averaging method (corresponded to the zero iteration)

2. Rater 3, although honest, is also blacklisted at the third iteration, it's reasonable when a honest but faulty rater's rating have a large deviation from the other honest raters.

Raters' trustworthiness

- Beta distribution:

prior to first time-slot, for each rater-peer *i*, the R_i value is set to 0.5 ($\alpha_i = 1$ and $\beta_i = 1$).

If rater-peer is blacklisted, R_i is decreased by setting:

$$\beta_i(t+1) = \lambda\beta_i(t) + (C_i + 1 - \tau)^\delta$$

Otherwise, R_i is increased by setting:

$$\alpha_i(t+1) = \lambda\alpha_i(t) + 1$$

How to choose the threshold τ?

- τ-eliminate-optimal scheme:

we declare a reputation scheme to be τ-eliminate-optimal if it can eliminate all the malicious raters whose inconsistency exceeds the threshold τ.

Lemma 1: Let Θ_j be the number of unique raters for the j^{th} SP. Then, a sufficient condition for the inconsistency C_i, at the first iteration, to exceed the threshold τ for all malicious raters is given by

$$\sum_{r\in\Lambda} \Psi_r \geq (\hat{b}m + b\tau) \qquad (2)$$

Here, $\Psi_r = \frac{mX+n\Theta_r\lambda^Q}{X+\Theta_r\lambda^Q}$ for $r \in \Lambda$, where Λ is the index set of the set Γ.

Given $C_i \geq \tau$ for a malicious rater i, for a τ-eliminate-optimal scheme, we require that the inconsistency of the malicious rater exceeds the inconsistencies of all of the honest raters.

NANYANG TECHNOLOGICAL UNIVERSITY

How to choose the threshold τ?

Lemma 2: (τ-eliminate-optimal condition): Let d_t be the total number of outgoing edges from an honest rater in t elapsed time-slots. Then, provided that Lemma 1 is met, ITRM would be a τ-eliminate-optimal scheme if the condition

$$\frac{\mu}{d_t} > 1 - \frac{\Theta \lambda^Q \Delta}{D} \quad (3)$$

is satisfied with high probability at the t^{th} time-slot.

NANYANG TECHNOLOGICAL UNIVERSITY

Parameters

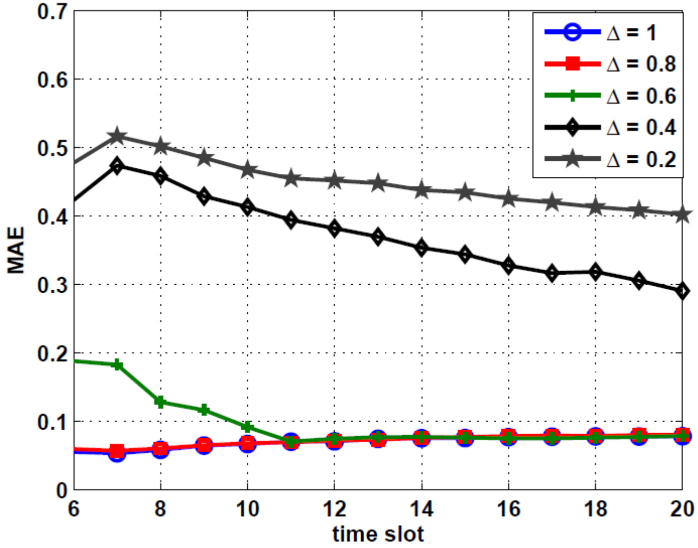
D	Number of malicious raters
H	Number of honest raters
N	Number of service providers
m	Rating given by an honest rater
n	Rating given by a malicious rater
X	Total number of malicious rates TR_{ij} per a victim SP
d	Total number of newly generated outgoing edges, per time-slot, by an honest rater
b	Total number of newly generated outgoing edges, per time-slot, by a malicious rater
$\hat{b}$	Total number of newly generated attacking edges, per time-slot, by a malicious rater
Δ	$\hat{b}/b$ (i.e., fraction of attacking edges per time-slot)
μ	Total number of un-attacked SPs rated by an honest rater

NANYANG TECHNOLOGICAL UNIVERSITY

Simulation

$$MAE = | \mathrm{TR}_j - \overline{\mathrm{TR}}_j |$$

Where $\overline{\mathrm{TR}}_j$ is the actual value of the reputation



W=D/(D+H)=0.1 (10% malicious peers)

Fig. 3: MAE performance of ITRM versus time for bad mouthing when $W = 0.10$ and varying Δ

NANYANG TECHNOLOGICAL UNIVERSITY

Simulation

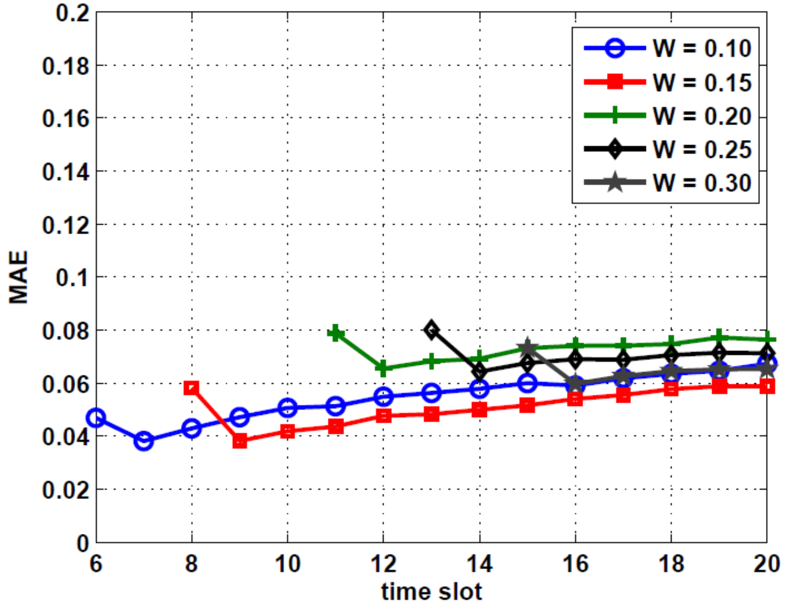


Fig. 4: MAE performance of ITRM versus time for bad mouthing and varying W

NANYANG TECHNOLOGICAL UNIVERSITY

Simulation(comparisons)

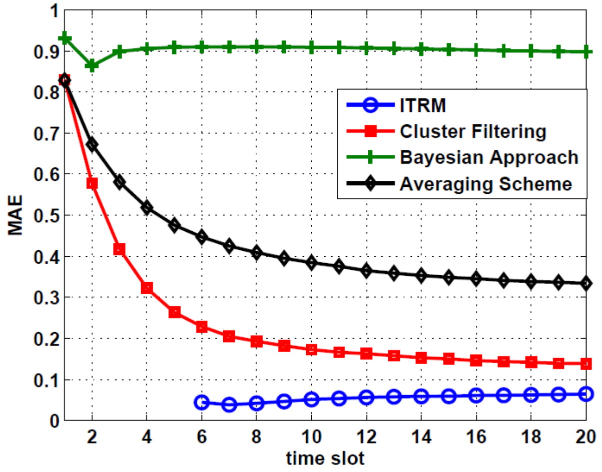


Fig. 5: MAE performance of various schemes for bad-mouthing when $W = 0.10$

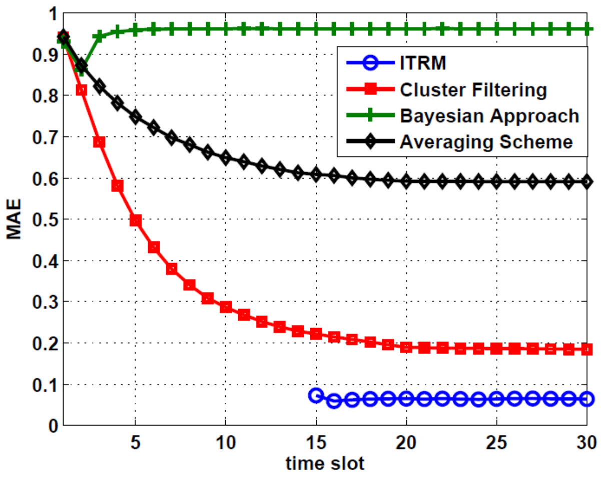


Fig. 6: MAE performance of various schemes for bad-mouthing when $W = 0.30$

NANYANG TECHNOLOGICAL UNIVERSITY

Discussion:

1. How to establish a distributed model?

2. What if the malicious raters turn good?

3. New comer attack?

NANYANG TECHNOLOGICAL UNIVERSITY