International Journal of Cyber-Security and Digital Forensics (IJCSDF) 2(1): 79-87
The Society of Digital Information and Wireless Communications, 2013 (ISSN: 2305-0012)

Investigating the Number of Users and Months to Make TULUNGAN Effective Against Self-Promoting Users

Alexis V. Pantola, Susan Pancho-Festin, Florante Salvador
De La Salle University, 2401 Taft Avenue Manila - Philippines,
University of the Philippines, Diliman, Quezon City, Philippines
pong.pantola@delasalle.ph, susan.pancho@up.edu.ph, bong.salvador@delasalle.ph

ABSTRACT

Tulungan is a reputation system for collaborative web filtering that is both consensus-independent and self-promoting resistant. Simulation results show that Tulungan requires less than 50% good users in order to become effective against malicious and self-promoting users (i.e., bad users). Being effective means that it is able to give high reputation values to good users relative to bad users. In addition, the effectiveness of Tulungan is also confirmed by comparing the number of correct URL categorizations that it made against the wrong ones. This is in contrast with other reputation systems that require at least 50% good users in order to be effective against malicious users and even a greater percentage of good users to work against self-promoting users.

Although previous studies show the consensus-independent and self-promoting resistant properties of Tulungan, the simulation covers only a fixed number of total users and months. This paper presents additional simulation involving Tulungan that confirms the number of users and months needed for it to become effective. Results show that the reputation values of good users suffer if it involves less than 200 users and 2 simulated months.

KEYWORDS

Reputation System, Web Filtering.

1 INTRODUCTION

Collaborative web systems such as Untangle [1] and Wikipedia [2] allow anyone with Internet access to contribute information (e.g., website categories in Untangle and articles in Wikipedia). Such paradigm creates a scalable system since source of contributions are not limited to few individuals [3]. However, this approach is prone to issues such as inaccurate contents. These can be caused by malicious contributors who intentionally provide erroneous contribution [4].

Reputation systems are developed to reduce the effects of malicious users. It utilize information gathered from transactions in a web system to assess contributors and contributions [5]. Transactions can be in the form of URL categorization in collaborative web filtering systems.

In most reputation systems, contributions are rated by users. Raters validate the contributions made by contributors whether they are accurate or not. In this paper, a user that is both a good contributor (i.e., contributes accurate contents) and rater (i.e., gives a positive rating to accurate contents and negative rating to inaccurate ones) is referred to as a good user. A user who intentional does the opposite (i.e., deliberately making incorrect contribution and rating) is referred to as a malicious user.

Generally, reputation systems are effective when the number of good users is more than their malicious counterpart. These reputation systems are considered as consensus-dependent [6]. Consensus-dependent systems rely heavily on the number of good users to influence the accuracy of contributions. If the number of good users is less than 50% of the total population of users, these systems favor the malicious users (i.e., malicious users are given higher reputation relative to their good counterpart). Such scenario allows inaccurate contents provided by malicious users to be considered as correct.

There are even cases when malicious users are still favored by reputation systems even if they are outnumbered by their good counterpart. This happens when malicious users employ attacks such as self-promotion [6]. Self-promotion happens when a malicious user rates his own contribution as positive even if said contribution is incorrect. In this paper, malicious users who perform self-promotion are referred to as self-promoting users.

Tulungan is a reputation system that is designed for collaborative web filtering. Simulation shows that is effective even if the percentage of good users is less than malicious users. This makes it consensus-independent. Furthermore, Tulungan is also proven to be self-promoting resistant since a simulation involving a majority number of self-promoting users shows that Tulungan can still produce more correct URL categorizations than wrong ones.

This paper further investigates the consensus-independent and self-promoting resistant properties of Tulungan by verifying the total number of users and number of months needed for it to become effective.

2 REVIEW OF RELATED WORK

Reputation systems can be classified into two, namely content-driven and user-driven.

2.1 Content-Driven Reputation System

A content-driven reputation system depends on the content of a contribution to determine its correctness. For example, WikiTrust [10, 11, 12] relies on the contents of Wikipedia pages to determine their accuracy. The stability of a content is equated to its credibility. It assumes that the less frequent an entry of a Wikipedia page is changed, the more credible it is since reviewers find it as already accurate and does not require any correction. However, such an assumption may lead to a wrong conclusion since a “non-edit” to an entry does not necessarily imply that said entry is correct. There are cases when authors are “lazy” in reviewing and editing an entire article and focus only in entries that interest them thereby leaving other entries unedited. For example, a biography entry in Wikipedia is proven inaccurate even if it is not edited for 132 days [13]. If the credibility of this entry is measured using WikiTrust, it is possible that it will be incorrectly labeled as accurate.

2.2 User-Driven Reputation System

A user-driven reputation system relies on user rating to measure the accuracy of contents. They are highly utilized in websites that allow Internet users to contribute contents and rate these contributions. The following are some user-driven reputation systems for collaborative web systems: rating system of Epinions [14], feedback forum of eBay [15,16,17,18,19], karma system of Reddit [20], social recommendation of Digg [21,22], moderation system of Slashdot [23,24,25], Rater-rating reputation system [26].

2.3 Comparison of Algorithms

Table 1 provides a comparison of the different reputations systems identified in this section. It is based from the study presented in [27]. The target characteristics of the reputation system considered in this study are shown in the last row. For easier comparison, table cells that are shaded are those characteristics that match the target.

WikiTrust satisfies all the targeted characteristics except it is a content-driven reputation system. The rating system of Epinions, feedback forum of eBay, karma system of Reddit, and Digg’s social recommendation are all susceptible to self-promotion and are consensus-dependent.

Epinion employs Eroyalties credits to encourage users to provide high quality contribution. However, such credits do not prevent malicious users from performing self-promotion in order to gain Eroyalties credits even if they did not provide accurate reviews.

The feedback forum of eBay accepts only feedback from users who were involved in a particular transaction (i.e., seller or buyer). This reduces but does not eliminate the chances of self-promotion being consummated effectively by malicious users.

The moderation system of Slashdot satisfies all the targeted characteristics except for its dependence on a content manager. This makes it relatively not scalable compared to reputation systems that do not employ the services of content managers.

Similar to Slashdot's moderation system, Rater Rating also missed on a single targeted characteristic. It is not consensus-independent. However, just like the moderation system of Slashdot, it is not susceptible to self-promotion.

Table 1. Comparison of Different Systems

System	User-Driven	needs a Content Manager	Self-Promoting	Consensus-Independent
WikiTrust	no	no	.	yes
Rating System of Epinions	yes	no	●	no
Feedback Forum of eBay	yes	no	●	no
Karma System of Reddit	yes	no	●	no
Digg's Social Recommendation	yes	no	●	no
Slashdot's Moderation System	yes	yes	.	yes
Rater Rating	yes	no	.	no
Target Characteristics	yes	no	.	yes

●-susceptible to attack
.-not susceptible or minimal susceptibility

3 THE TULUNGAN REPUTATION SYSTEM

This section presents a summary of the algorithm used by Tulungan as presented in the paper [6].

The algorithm consists of three phases: initialization, contribution and rating, and computation (refer to Algorithm 1). The initialization phase is executed everytime a new user and/or URL category are introduced in the reputation system. The last two phases are repeated every month.

Algorithm 1. Tulungan Reputation System

Initialization Phase:

1. Initialize the contribution reputation φ_c, rating reputation φ_r, and level α of all URL categories s in $\mathbf{S}$.
2. Initialize the contributor reputation ρ_c and rater reputation ρ_r of user u and add it in $\mathbf{U}$.

Contribution and Rating Phase:

3. Allow all users u to add contribution c in $\mathbf{C}$.
4. Determine potential raters p and add them in $\mathbf{P}$.
5. Allow all users u that are potential raters to add rating group g in $\mathbf{G}$ and rating r in $\mathbf{R}$.

Computation Phase:

6. Update the rating reputation φ_r of all URL categories s in $\mathbf{S}$.
7. Update the contribution reputation φ_c of all URL categories s in $\mathbf{S}$.
8. Update the level α of all URL categories s in $\mathbf{S}$.
9. Update the contributor reputation ρ_c of all users u in $\mathbf{U}$.
10. Update the rater reputation ρ_r of all users u in $\mathbf{U}$.
11. Update the overall reputation ρ_o of all users u in $\mathbf{U}$.

The initialization phase sets the contribution reputation and rating reputation values of all URL categories as well as the contributor and rater reputation values of users. The contributor and rater reputation values will start with a value close to zero. This means that all new users are treated equal (i.e., Tulungan is not aware whether they are good, malicious, or self-promoting).

The contribution and rating phase allows users to give contribution. As an example, a user can contribute that www.google.com is NOT a gambling site or www.nba.com is a Sports and

Recreation site. Figure 1 shows a sample page in giving a contribution.

Figure 1. Contribution Page of Tulungan

Aside from providing contributions, the potential raters are determined in the second phase. Potential raters are provided with a set (or several sets) of contributions to rate. Each set is composed of three URLs as shown in the sample page in Figure 2.

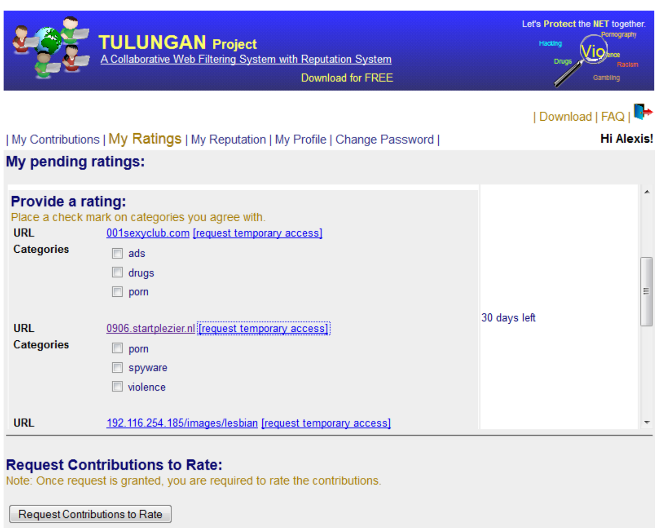


Figure 2. Rating Page of Tulungan

Rating a set of three URLs is significant since this set is composed of one unknown URL and two control URLs. The unknown URL, as the name implies, is the URL with unknown categories. On the other hand, the categories of the two control URLs are already known by Tulungan.

When a user rates the three URLs, Tulungan can verify the "seriousness" of the rating of the unknown URL by checking if the rating of the two control URLs are correct. If they are correct, Tulungan assumes that the user is "serious" in providing ratings to the three URLs. This approach is adopted from reCAPTCHA [28].

The unknown as well as the two control URLs are presented in a way that raters have no idea which is the unknown URL. This is very critical to give Tulungan a higher chance of detecting incorrect rating. If raters are given a clue on which are the unknown URL and the control URLs, users (i.e., malicious users) may intentionally provide correct ratings on the control URLs and wrong rating on the unknown URL.

The computation phase calculates the new contribution and rating reputation of URL categories as well as the contributor and rater reputation values of users. This is essential to measure the credibility of users in providing contributions as well as the accuracy of URL categories. The reputation values that were computed in this phase.

4 EVALUATION OF THE REPUTATION SYSTEM

4.1 Evaluation Set-up

The simulation follows the set-up described in [29]. However, instead of varying the percentage of good users, the percentage is fixed at 20%. This percentage is selected since this is the minimum percentage of good users needed to make Tulungan effective against both malicious and self-promoting users.

Instead of varying the percentage of good users, the total no. of users and total no. of months used in the simulation are varied.

4.1.1 Varying Number of Total Users

These simulations verify if the total number of users has a bearing in the result presented in [29]. Take note that in [29], the total number of users is fixed at 500. In these simulations, the total number of users varies. In addition, aside from the percentage of good users which is fixed at 20%, the number of simulated months is also fixed at 12 months.

Simulations under this category start with 100 total users. The users contribute and rate for 12 simulated months. The average of reputation values per user type and correct and wrong URL categorizations is performed after the 12th month.

The process is repeated with 200 total users. This is further repeated with an increment of 100 in the total number of users until it reaches 1000 total users.

Through this approach, the number of months needed before the effectiveness of Tulungan is observed can be determined.

4.1.1 Varying Number of Months

These simulations verify if the number of months has a bearing in the result presented in [29]. Take note that in [29], the number of months used in the simulations is 12 months.

In these simulations, the number of months varies. However, the percentage of good users is fixed. Similar to the simulations discussed in the previous section, the percentage of good users in these simulations is fixed at 20%. Aside from this, the total number of users is also fixed at 500 users.

Similar to the previous section, the simulations take note of the average reputation values of the good users and their bad counterpart as well as the number of correct and wrong URL categorizations at the end of every month. The simulations run for 12 months.

Through this approach, the number of months needed before the effectiveness of a reputation system is observed can be determined.

4.2. Simulation Results

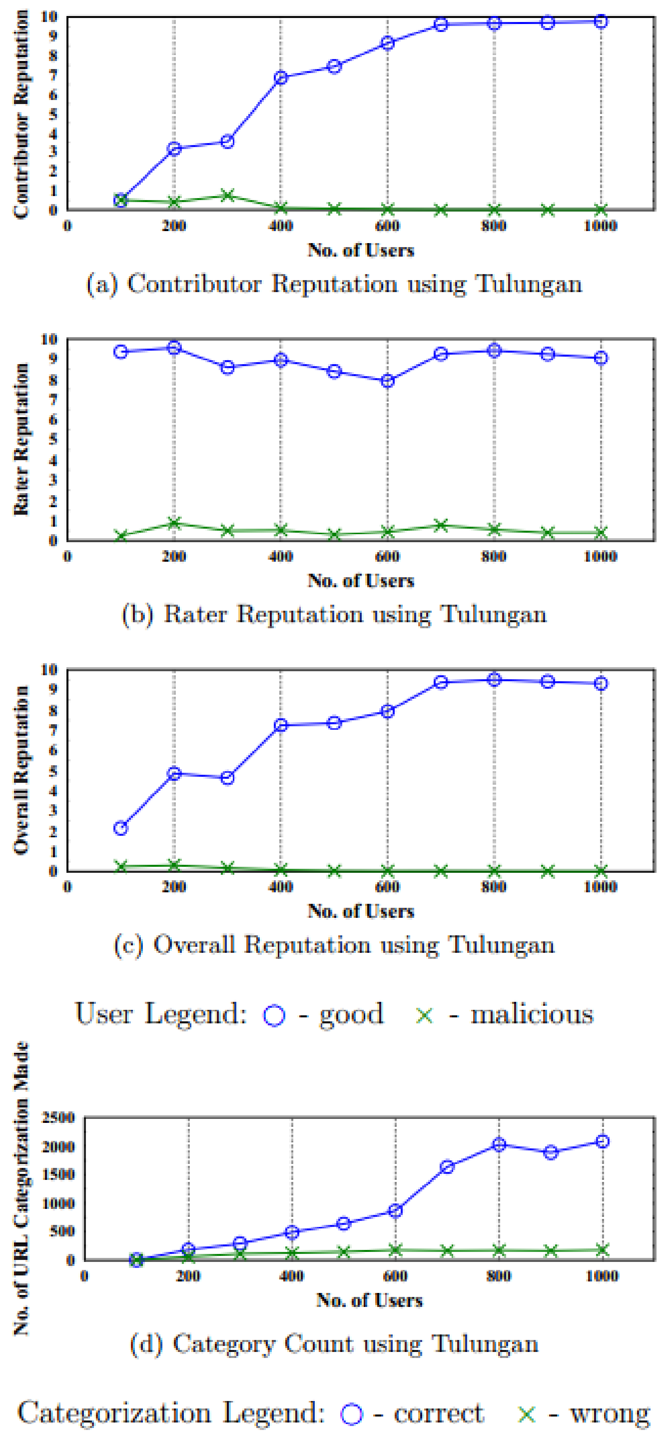


Figure 3. Good versus Malicious Users: Varying Number of Total Users

4.2.1 Good vs. Malicious Users (Varying Number of Total Users)

Figure 3 shows that only 200 users are needed for Tulungan to be effective. When there are 200 users, the contributor reputation of good users reach already a value of 3 while malicious users have less than 1. A much better result can be seen with the rater reputation. Good users almost reach a value of 10 for their rater reputation, while malicious users reach only 1.

Similarly, the number of correct URL categorization is greater than the wrong ones when there are at least 200 users.

At 100 users, the contributor reputation of both good and malicious users are approximately the same (i.e., less than 1). This may be attributed to the insufficient number of contributions that are actually worth rating. This can be verified with the number of correct URL categorizations made with 100 total users. As seen in the graph, there is almost a negligible number of correct URL categorizations.

It should be noted that even if there are insufficient number of contributions that are worth rating, the rater reputation of good users is still high even if the total users is 100. This may be attributed to the fact that Tulungan does not solely rely on the number of rating, but also relies heavily on the control URLs. Since the good users provide correct rating on the control URLs, it is expected that their rater reputation will be high.

4.2.2 Good vs. Malicious Users (Varying Number of Months)

Figure 4 shows that 9 months are needed to make Tulungan completely effective. Although, only two months are needed to make the rater and overall reputation of good users higher than malicious users, and 7 months to make the number of correct URL categorizations higher than the wrong ones, the contributor reputation of good users becomes higher only compared to their malicious counterpart after 9 months.

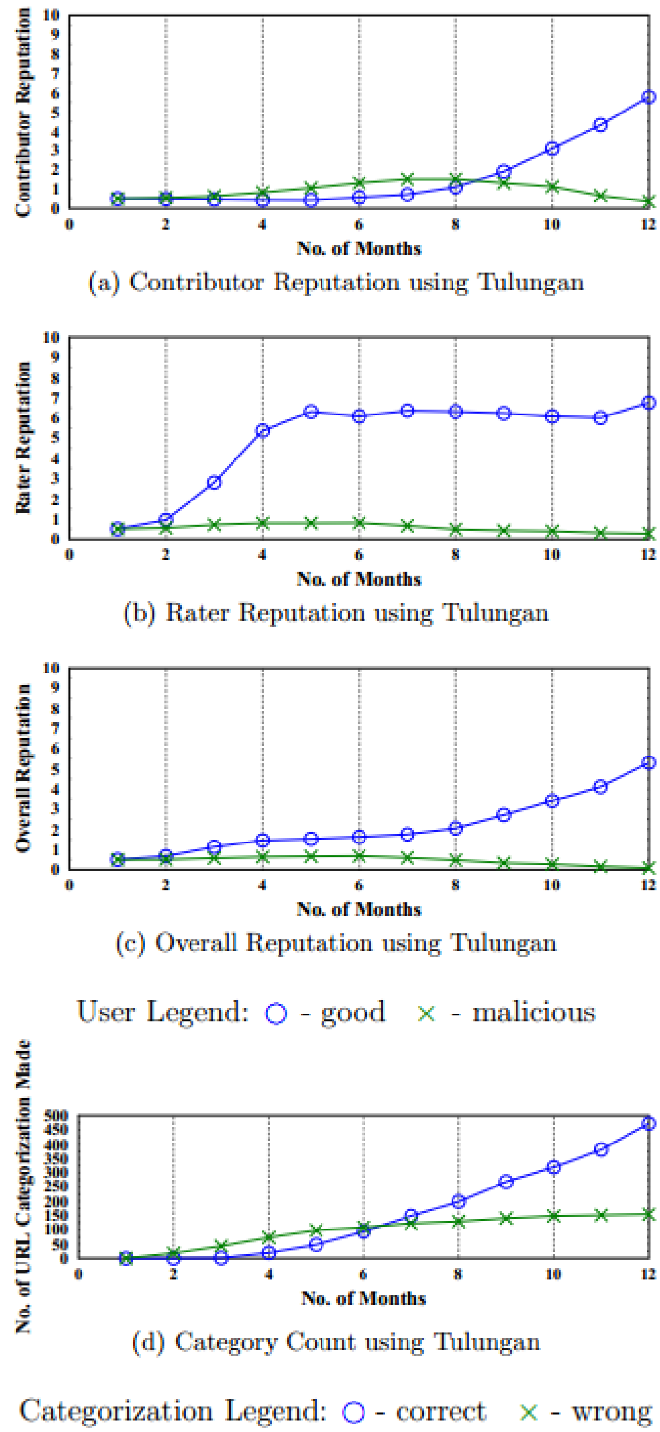


Figure 4. Good versus Malicious Users: Varying Number of Months

As discussed in the previous section, the rater reputation of good users increases faster due to the use of control URLs. The contributor reputation on the other hand requires a momentum to build before a significant increase can be seen. Prior to the 9th month, the total increase on the contributor reputation of good users is less than 1, while from the 9th month to the 12th month, an increase of 1 to 2 contributor reputation value can be seen every month.

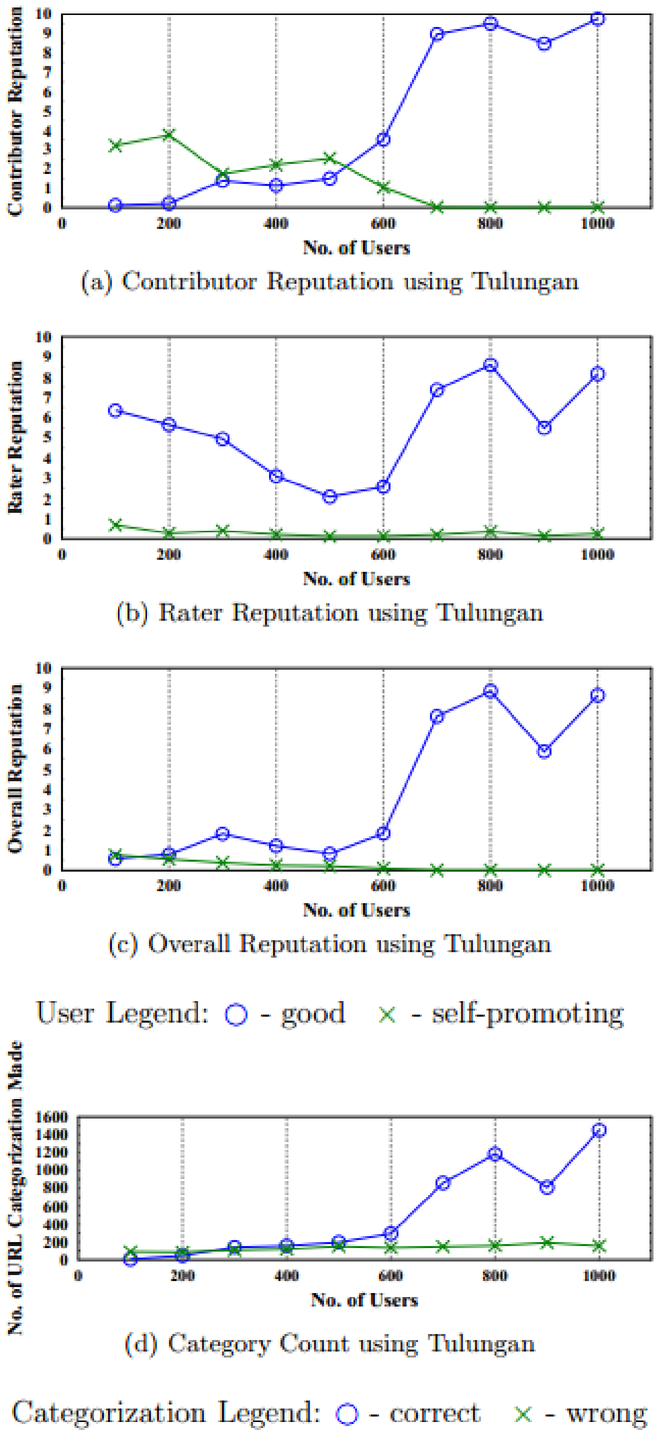


Figure 5. Good versus Self-Promoting Users: Varying Number of Total Users

4.2.3 Good vs. Self-Promoting Users (Varying Number of Total Users)

Figure 5 shows that 600 total users are needed by Tulungan in order to be completely effective against self-promoting users.

Consistent with the previous results, Tulungan is able to limit the rater and overall reputation values of bad users (i.e., self-promoting users) to 1 or less regardless of the total number of users. However, the contributor reputation of self-promoting users peaked at 4 when there are 200 total users. Good users are able to overtake self-promoting users in terms of contributor reputation when there are 600 total users.

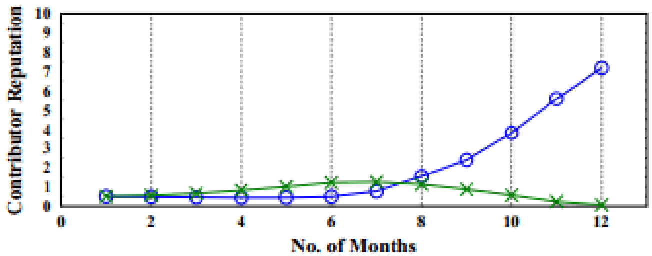
However, it should be noted that the number of wrong URL categorizations are limited to less than 200 by Tulungan regardless of the total number of users. This is less than 15% of the highest number of correct URL categorizations achieved by the reputation system.

4.2.4 Good vs. Self-Promoting Users (Varying Number of Months)

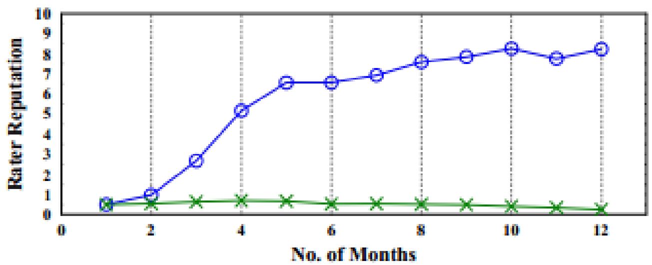
Figure 6 shows that 8 months are needed to make Tulungan completely effective. The result is very similar to the Good vs. Malicious Users (Varying Number of Months) simulation.

Only two months are needed to make the rater and overall reputation of good users higher than malicious users, and 7 months to make the number of correct URL categorizations higher than the wrong ones, the contributor reputation of good users becomes higher only than their malicious counterpart after 8 months.

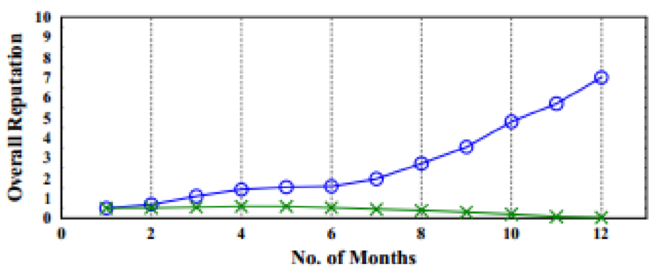
Consistent with the previous sections, the rater reputation of good users increases faster due to the use of control URLs. The good users require 8 months before a relatively rapid increase in contributor reputation can be observed.



(a) Contributor Reputation using Tulungan

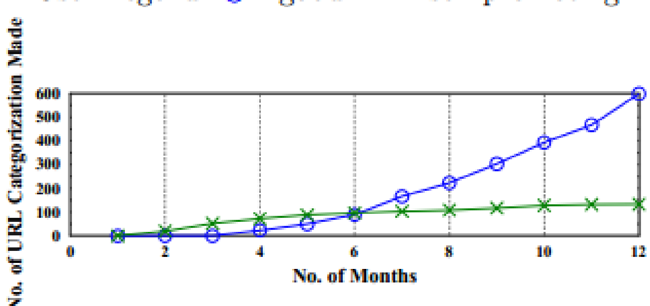


(b) Rater Reputation using Tulungan



(c) Overall Reputation using Tulungan

User Legend: ○ - good × - self-promoting



(d) Category Count using Tulungan

Categorization Legend: ○ - correct × - wrong

Figure 6. Good versus Self-Promoting Users: Varying Number of Months

5 CONCLUSION AND FUTURE WORK

The simulation results show that Tulungan is effective given a sufficient number of total users and number of months even if the percentage of good users is fixed at 20%.

Good users can easily increase their rater reputation as long as there are at least 200 users and 2 simulated months. However, Tulungan requires 600 users and 8 months before the contributor reputation of good users become higher than self-promoting users.

It should be noted that regardless of the number of total users and simulated months, the number of wrong URL categorizations is kept to a minimum relative to the correct URL categorizations.

Although Tulungan is self-promoting resistant, further study can be performed in order to assess its effectiveness against other reputation attacks such as slandering, whitewashing, orchestration, and denial of service.

6 REFERENCES

1. Untangle - Multi-functional firewall Software - Open Source Content Filter and Spam Filter [Online]. URL http://www.untangle.com. Last accessed March 2011
2. Wikipedia, the free Encyclopedia [Online]. URL http://www.wikipedia.org. Last accessed January 2011
3. M. Ebner, J. Zechner, Proceedings of I-KNOW 06, 6th International Conference on Knowlegde Management, ACM Press (2006)
4. T. Chesney, First Monday 11(11) (2006). URL http://www.firstmonday.org/issues/issue11_11/chesney/index.html
5. J. Kennes, A. Schiff, The Value of a Reputation System. Industrial Organization 0301011, Econ-WPA (2003). URL http://ideas.repec.org/p/wpa/wuwpio/0301011.html
6. A.V. Pantola, S. Pancho-Festin, F. Salvador, Science Diliman 23(2) (2012)
7. J.R. Douceur, In Proceedings for the 1st International Workshop on Peer-to-Peer Systems (IPTPS) pp. 251–260 (2002)
8. J. Zittrain, The Future of the Internet And How to Stop It. 36 (Yale University Press, 2008). URL http://www.amazon.com/Future-Internet-How-Stop/dp/0300151241
9. Some eBay Users Abuse Auction Site's Feedback System, Professor Finds [Online] (2007). URL http://www.physorg.com/news87832472.html. Haas School of Business, UC Berkeley
10. B.T. Adler, L. de Alfaro, in Proceedings of the 16th International Conference on World Wide Web

(ACM, New York, NY, USA, 2007), WWW'07, pp. 261–270. DOI http://doi.acm.org/10.1145/1242572.1242608. URL http://doi.acm.org/10.1145/1242572.1242608
11. B.T. Adler, K. Chatterjee, L. de Alfaro, M. Faella, I. Pye, V. Raman, in Proceedings of the 4th International Symposium on Wikis (ACM, New York, NY, USA, 2008), WikiSym '08, pp. 26:1–26:12. DOI http://doi.acm.org/10.1145/1822258.1822293. URL http://doi.acm.org/10.1145/1822258.1822293
12. B.T. Adler, L. de Alfaro, I. Pye, V. Raman, in Proceedings of the 4th International Symposium on Wikis (ACM, New York, NY, USA, 2008), WikiSym '08, pp. 15:1–15:10. DOI http://doi.acm.org/10.1145/1822258.1822279. URL http://doi.acm.org/10.1145/1822258.1822279
13. T. Cross, First Monday 11(9) (2006). URL http://firstmonday.org/issues/issue11_9/cross/index.html
14. Epinions.com - FAQs [Online]. URL http://www99.epinions.com/help/faq/. Last accessed June 2011
15. P. Resnick, K. Kuwabara, R. Zeckhauser, E. Friedman, Commun. ACM 43, 45 (2000). DOI http://doi.acm.org/10.1145/355112.355122. URL http://doi.acm.org/10.1145/355112.355122
16. Feedback Forum - How Feedback works [Online]. URL http://pages.ebay.ph/services/forum/feedback.html. Last accessed March 2011
17. Feedback - 5 Simple Steps to eBay Feedback [Online]. URL http://pages.ebay.ph/ebayexplained/feedback.html. Last accessed March 2011
18. E. Yao, R. Fang, B.R. Dineen, X. Yao, Journal of Business Research 62(12), 1281 (2009)
19. D. Steiner, Auction Bytes. The Independent Trade Publication for Online Merchants (2003). URL http://www.auctionbytes.com/cab/abu/y203/m01/abu0087/s02
20. Reddit.com: Help [Online]. URL http://www.reddit.com/help/faq. Last accessed June 2011
21. Digg [Online] (2012). URL http://www.digg.com
22. K. Lerman, A. Galstyan, in Proceedings of the first workshop on Online social networks (ACM, New York, NY, USA, 2008), WOSP '08, pp. 7–12. DOI http://doi.acm.org/10.1145/1397735.1397738. URL http://doi.acm.org/10.1145/1397735.1397738
23. Slashdot - FAQ [Online]. URL http://slashdot.org/faq. Last accessed June 2011
24. C. Lampe, Ratings Use in an Online Discussion System: The Slashdot Case. Ph.D. thesis, University of Michigan (2006)
25. C. Lampe, P. Resnick, in Proceedings of the SIGCHI conference on Human factors in computing systems (ACM, New York, NY, USA, 2004), CHI '04, pp. 543–550. DOI http://doi.acm.org/10.1145/985692.985761. URL http://doi.acm.org/10.1145/985692.985761
26. A.V. Pantola, S. Pancho-Festin, F. Salvador, in Proceedings of the 3rd international conference on Security of information and networks (ACM, New York, NY, USA, 2010), SIN '10, pp. 71–80. DOI http://doi.acm.org/10.1145/ 1854099.1854116. URL http://doi.acm.org/10.1145/1854099.1854116
27. K. Hoffman, D. Zage, C. NitaRotaru, ACM Comput. Surv. 42, 1:1 (2009). DOI http://doi.acm.org/10.1145/1592451.1592452. URL http://doi.acm.org/10.1145/1592451.1592452
28. reCAPTCHA - Stop Spam, Read Books[Online]. URL http://www.google.com/recaptcha/. Last accessed March 2011
29. Pantola, A. V., Pancho-Festin, S., & Salvador, F. (2013a, 3). TULUNGAN: A Self-Promoting Resistant Reputation System for Collaborative Web Filtering Systems. Second International Conference on Cyber Security, Cyber Warfare and Digital Forensic.