What network simulator questions do users ask? a large-scale study of stack overflow posts

Syful Islam[1], Yusuf Sulistyo Nugroho[2], Md. Javed Hossain[3]
[1]Nara Institute of Science and Technology, Japan
[2]Universitas Muhammadiyah Surakarta, Indonesia
[1,3]Noakhali Science and Technology University, Bangladesh

ABSTRACT

The use of network simulator as a modern tool in analyzing and predicting the behaviour of computer networks has grown to reduce the complexity of its accuracy measurement. This attracts researchers and practitioners to share problems and discuss them to improve the features. To communicate the related issues, users move to online question-answering platforms. Although recent studies have shown the popularity and benefits of adopting network simulation tools, the challenges users face in using the network simulator remain unknown. In this research paper, we examine 2,322 network simulator related stack overflow question posts to gain insights into the topics and challenges that users have discussed. We adopt the latent dirichlet allocation model to understand the topics discussed in stack overflow. We then investigate the popularity and difficulty of each topic. The results show that users use stack overflow as an implementation guideline for the network simulation model. We determine 8 discussion topics that are merged into 5 major categories. Simulation model configuration is the most useful topic for users. We also observe that target network protocol modification and network simulator installation are the most popular topics. Network simulator installation and target network protocol modification issues have been challenging for most users. The findings also highlight future research that suggests ways to help the network simulator community in the early stages to overcome the popular and difficult topics faced when using network simulation tools.

Corresponding Author:

Syful Islam
Laboratory of Software Engineering
Nara Institute of Science and Technology, Japan
Email: islam.syful.il4@is.naist.jp

1. INTRODUCTION

Since the complexity of communication networks have increased to measure the accuracy of system behavior in traditional analytical techniques [1], network simulators (NS) are used as a modern technique to analyze and predict the behavior of computer networks. In the implementation, NS allows the users to design, modify and test the networking protocols in a simulation mode and is modeled with devices, links and application. to report the performance of the tested networks. Today's application of simulation techniques has attracted researchers and practitioners to discuss and find a way to improve the features. To communicate the NS-related issues, update, users turn to an online question-answering platforms, such as stack overflow (SO), and to get help and advice from community about the technical problems they face. Stack overflow is an online

question-answering platform that accommodate the need of various users and discusses a wide-range of topics and programming languages.

A number of research on the quality and content of online question-answering platforms, such as stack overflow, has indicated its significance. Zagalsky et al. [2] reported that SO with R-tag has become a communication channels that have a relationship with the topics discussed in R software development community forum. Squire [3] showed that the measurement of quality, users' participation, and the effectiveness of responding time in the SO forum have been the key factors that caused developers to move from mailing list. Calefato et al. [4] proposed a guideline of making a successful question in SO forum, while Wang et al. [5] analyzed the important factors that impact the time of receiving an accepted answer in four Stack Exchange websites. SO discussion topics were also investigated in several studies. Beyer et al. [6] performed a large scale empirical study on SO to analyze topics and the current trends that developers discuss. Finally they automatically classify the posts into seven question categories. Other study on mobile-developers topics in SO was conducted by Rosen and Shihab [7].

Extending prior works that analyze the quality of SO questions and their topics, in this paper, we conduct a large scale empirical study on the topics covered by NS questions posted in SO. We also investigate the types of issues faced by the NS users by categorizing the keywords used in the questions. We further study the NS questions that are most difficult to answer by calculating the PD score [8]. Based on the analysis of 2,322 NS-related questions from SO, we find that the NS-related threads seekers use SO platform to discuss simulation model configuration, target network protocol modification, NS installation, simulation model performance measure and NS build error. Although simulation model configuration is the most useful topic amongst the NS users, it does not so popular. The most viewed NS-related discussion topic by the users is NS installation. Most discussions in Stack Overflow are initially triggered with a *How-to* type of questions. This indicates that NS users usually ask for an instruction, guidance or tutorial to solve their problems. However, although the number of responses is high, NS installation and target network protocol modification are the most difficult types of questions to get the appropriate answers.

Our findings show that NS users need a specific discussion forum to reach the maturity level of similar project-specific discussion platforms, such as the Eclipse forum. So that the most common issues, such as plug-ins and documentation, could be identified and suggesting to new users how they can address the issues that prevent their entry to the discussion forums[9]. In addition, our insights can be used for guidance to conduct future research on NS-related discussions in other channels and artefacts in the complex software development environment.

Paper Organization. The rest of this paper are organized is being as. Section 2. presents the research methodology. In detail, we describe the research questions, our procedures to collect data, and our online appendix. The study results are presented in section 3. to answer the formulated research questions. The implication of the study are then explained in section 4. section 5. and 6. describe the threat to validity and present the related works, respectively. Finally, we conclude the paper in section 7.

2. RESEARCH METHOD

This section presents our research questions, data collection process, and the replication packages in an online appendix.

2.1. Research questions

This study aims to extract insights into NS-related discussion characteristics on the SO platform. To achieve this goal, we have created three research questions to guide our research. We present these questions with motivation. Network simulators (NS) have become a high demand for network engineers and researchers [10] . Different type of users will have various NS-related problems that require a different area of expertise. For example, some users require specific expertise in the tcl scripting, but others could have problems on network protocols, or design features. Thus, the difficulties faced by users are likely to differ. Since users get the benefit from question-answering platforms to communicate issues, the objective of this research question is to understand the most useful and popular NS topics that are frequently faced by NS community. In addition, identifying widely discussed NS topics is the first step in highlighting issues that are gaining more attention.

RQ$_2$: What types of questions do users face?

Taking the results from RQ_1, we then set out to empirically study the types of question that were asked

by users. Previous study [7] shows that users ask the questions in different types (i.e. how, what, why). Similar to the approach of prior study [7], this analysis is performed to identify the nature of difficulties encountered while using NS.

RQ₃: What topics are the most difficult to answer?

The key motivation of this *RQ* is to investigate the topics that are difficult to answer. Finding the topics that are hard to answer will help the users to get more attention from the NS community. Furthermore, it highlights the topics that require better support (tools/framework/ official documentation) for addressing NS usage related difficulty.

2.2. Data Collection

Figure 1 outlines the methodology for collecting the data which is described is being as. We initially downloaded the latest SO data dump (July 2008 to December 2019) that is publicly available on the SO Torrent [11] . The SO data contains all the Q&A with the metadata (creation date, favourite count, views, and score). The initial collected dataset contains 46,947,633 posts, where 39.83% (18,699,426) are question posts and 60.17% (28,248,207) are answer posts spanning from July 2008 to December 2019.

Step 1: Filter using simulator tag. SO posts are typically tagged by relevant tags to improve visibility of the posts. Following the similar approach that was used in previous study [7] , we collect the initial NS dataset by filtering posts that contain simulator as a tag word. The output of this step is 1,407 NS posts.

Step 2: Discover relevant tags. In this step, we extract the co-occurring tags with simulator from 1,407 posts to discover relevant tags. One major risk of discovering relevant tag is the possibility of introducing noise in the main dataset. For example, JNS java network simulator (JNS)for implementing ns2 is a relavant post that contains tag words Java along with simulator and ns2. Therefore, to mitigate this problem, we group and aggregate the target tags through semi-automatic process. In detail, the process includes string search using manual verification of newly explored tags. In addition, we validate the tags by applying tag relevance threshold (TRT) and tag significance threshold (TST) as metrics to validate tags:

$$TRT_{tag} = \frac{tag\_posts}{Sum(tag\_posts)} \tag{1}$$

$$TST_{tag} = \frac{tag\_posts}{popular\_tag\_posts} \tag{2}$$

where tag_posts is the number of NS posts for the tag, Sum(tag_posts) is the total number of posts for the tag, and popular_tag_posts means number of NS posts for most popular tag. For instance, omnet ++ is a tag word that occur only 11 times as a co-occurring tag in simulator tagged post while the total number of question posts on SO is 1406. Therefore, we also included such kind of tags in the final tag set. Thus, the output of step 2 is manually validated 4 tags (see Table 1) as being representative of NS discussions.

Step 3: Collect tagged posts. After getting the NS-related tag sets, we utilize those tags to identify and extract posts to create the final NS post dataset. The output of step 3 is 2,322 posts from SO being used as final dataset in the subsequent sections.

Step 4: Extract post title and preprocess for LDA. In this step, we apply a filter to remove irrelevant information. For topic modeling we only focus on title of the question posts since body of post can introduce noise to our analysis. Using the similar approach to prior studies [7] to extract the title of the posts, we performed pre-processing of the data. This includes removal of emails, newline characters, stop words using regular expression [12] and python NLTK [13]. We subsequently build a bigram model using Gensim [14] and lemmatize the words to map the original words. The output of step 4 is NS post title corpus which is used as the input of LDA (Latent Dirichlet Allocation) model.

Step 5: LDA topic modeling. As illustrated in Algorithm 1, in this step, we identify the NS topics using SO post title. To obtain the topic names, we use the LDA technique [15] , which was also used by previous studies [7, 16-18]. In this study, we apply the Mallet model of the LDA technique [19] to create group of NS posts in our dataset based on the keywords exists in the title of posts. To obtain the optimal number of topics k, we perform the modeling process in several iterations. First, we run the LDA for range (0-50) with 3 step size increment. Second, we choose the sub-optimal range (4-20) based on the coherence score [20]. Third, we again run the model for sub-optimal range with 1 step size increment and thus optimally come-up with 8 topics.
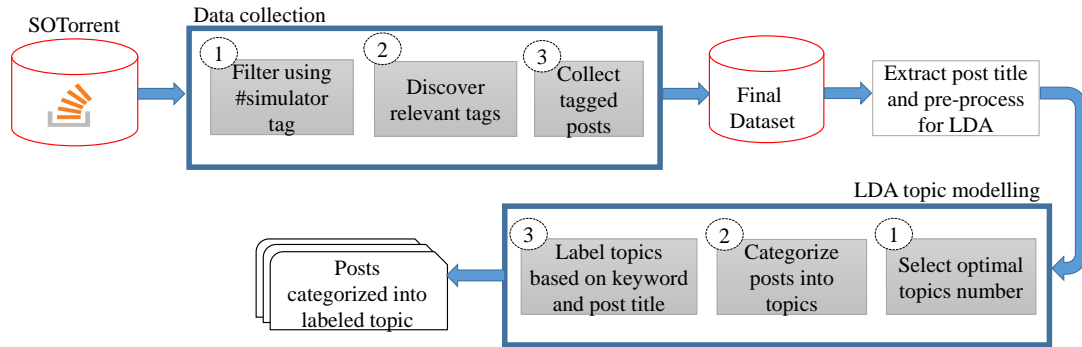
Figure 1. Overview of the methodology of NS study

Table 1. The tag list used to identify and extract NS related posts. The TRT and TST values are presented in percentages.

Filtered tag	#Initial posts	#Final posts	TRT (%)	TST (%)
simulator	1,407	1,407	100.00	100.00
ns2	15	599	2.50	1.06
ns-3	11	316	3.48	0.78
omnet++	11	1,406	0.78	0.78

Finally, we run the model with topic number k=8 and obtain 8 NS topics with their associated keywords (20 keywords per topic).

Algorithm 1: NS topic modeling using LDA

Input: NS_{qposts}=Stack Overflow NS question posts obtained in step-3
Result:Suggested NS topics (k) and keywords Method:
$NS_{all\_post\_title}$=Extract titles from NS_{qposts};
NS_{post_title}=Preprocess the $NS_{all\_post\_title}$ to remove noise;
For:run LDA topic modeling on NS_{post_title} using a custom range ($N_1 - N_2$) iterate with step size 3;
compute coherence score for each topic number;
select sub-optimal topic range ($O_1 - O_2$) based on coherence score;
For:run LDA topic modeling on NS_{post_title} using sub-optimal range ($O_1 - O_2$) iterate with step size 1;
compute coherence score for each topic number;
select optimal topic number k based on coherence score;
run the LDA topic modeling on NS_{post_title} for the optimal topic number k
return Suggested NS topics (k) and keywords

2.3. Online Appendix

We Publish the replication package. It contains (1) the NS dataset, (2) python codes, and (3) results of the study. The package is available at https://github.com/syful-is/Network/simulator.

3. RESULT AND DISCUSSION

This section describes the analyses of SO posts and topics to answer the research questions. In details, we present each research question alongside with the approach and the results.

3.1. *RQ₁: What kind of topics presented in the network-simulators related discussion?*
Approach: To answer this *RQ*, we apply the LDA topic modeling for identifying topics based on the title of NS post as described in section 2. We label the topic names based on the keywords suggested by LDA and by a manual reading of the top 25 question posts for each topic. Manual analysis of topic keywords and question posts reveals that some topics have similar meanings and ask similar types of questions, such as keywords related to simulation model and network model configuration.

Table 2. Top 5 discussion topics that relate to network simulators

Topic name	Top 10 keywords	#Posts
Simulation model configuration	vein, node, packet, network, message,send, omnet, vehicle, create, node	999
Target network protocol modification	file, implement, find, add, function, route, set, parameter, code, module	527
NS installation	installation, error, omnet, omnetpp, variable, window, unable, read, problem, ubuntu	278
Simulation model performance measurement	simulation, run, time, calculate, work, delay, distance, end, through-put, result	276
NS build error	omnet, inet, make, project, build, link, error, library, fail, command	242
sum		2,322

Table 3. Usefulness and popularity of the top 5 discussion topics. The score, favorite and view counts are presented in average

Topic name	Score	Favorite	View
Simulation model configuration	0.37	0.28	420.52
NS build error	0.32	0.15	477.67
Simulation model performance measurement	0.32	0.21	488.59
Target network protocol modification	0.31	0.14	634.86
NS installation	0.18	0.09	602.08

While these are different topics, they relate to simulation model configuration. Therefore, some topics are merged and grouped into the same topic name. Thus, we obtain 5 final topic names from 8 topics suggested by LDA topic modeling.In addition to the results of this *RQ*, we will also look at the most popular NS topics among users. To identify the most useful and popular topics, we use three different metrics (i.e., score, favorite, and views) that were also used in previous studies [7, 8, 21, 22]. We used the SO tour [23] as reference for definition of the metrics.

— The average score of the NS question posts. According to SO tour, members are allowed to up-vote posts that are considered useful to users. This votes are summarized as the score. We use this score as one of the metric to measure the usefulness of the post topics.

— The average number of posts marked as favourite by SO users. This metric is used to measure the usefulness of the post topics.

— The average number of views of the post by both unregistered and registered users. According to the SO Tour, if a question post is viewed by many users, this post is considered popular among them. Therefore, this metric shows the popularity of the topic.

Results: LDA topic modelling on SO posts suggest that users mainly discuss 5 NS-related topics. Table 2 shows the 5 topic names, number of posts for each topic, and the top 10 associated-keywords obtained from LDA topic modeling. We find that, simulation model configuration is the most common topics discussed by users, followed by target network protocol modification. The other three topics discussed by users are NS installation, simulation model performance measurement, and build error.

In the second part of analysis, we examine the NS topics usefulness and popularity among users. Table 3 shows that simulation model configuration is the most useful topic based on the average score and favorite count of the posts. This indicates that, users find the posts related to this topic as most useful. In addition, we find that based on view count of posts, target network protocol modification and NS installation are the top two most popular topics among the users. This indicates that, posts related to this topics are most commonly searched by NS users on SO.

Table 4. Different issues of the top 5 topics faced by developers. The values are presented in percentage

Topic name	how-to	what	why	others
NS installation	78.06	4.68	1.43	15.83
NS build error	70.25	7.85	0.00	21.90
Target network protocol modification	64.71	8.92	1.33	25.05
Simulation model configuration	56.66	12.02	0.90	30.43
Simulation model performance measurement	55.80	9.05	2.17	32.97

RQ$_1$: Summary

LDA topic modelling on SO posts suggest that users mainly discuss 5 NS-related topics. We find that Simulation Model Configuration is the most useful topic to users. In addition, we observe that target network protocol modification and NS installation issues are the most popular topics among the users.

3.2. *RQ$_2$: What types of questions do users face?*

Approach: To examine the questions faced by the users, we apply the same approach as previous studies to identify the types of posts in SO [7, 8, 24]. We utilize two steps to obtain the results. First, we manually investigate 30 random sample questions using keywords (i.e., how, what, and why). We observed that some question asked for instruction without using 'how-to' keyword. For example 'Is there anyway to change coverage in wireless node?' is a 'how-to' types NS question post. Therefore, after manual investigation, we also include 'Is there anyway' as search string to identify 'how-to' questions. In the same way, we append a keyword list to classify the post into question (i.e., how, what, and why) types. Finally, we apply the keyword list in the post title and body to obtain classification of the questions. The categories used in NS question post labelling are as follows:

— How - is a question type asks for instructions to perform a task. For example, *"how to draw xgraph in satcom in ns2 simulation"*. This question asks for instruction on xgraph feature in ns2 simulation.

— What - is a question type ask for information that are more abstract, conceptual in nature, asking for decision help, or ask on non-functional requirements. For example, *"What network simulation model should I use for simulating the behavior of an ad-hoc network in OMNET++"*. This type of question is asking on the network simulation model for predicting behavior of an adhoc network in OMNET++.

— Why - is a question post that ask for review, reason, or cause for something. For example, *"Why is the following tcl script for NS2 gives error for procedure implementation?"*. This question asks for clarifying why an error has happened.

— Others - is a question post that can't be classified by keyword search in the title and body of the post. For example, *"#includes in OMNeT++ Unit Tests"*.

Results: Table 4 shows that most NS posts of each topic ask for an instruction to perform their specific tasks. This is indicated by high percentage of how-to type of question, ranging from 55.80% to 78.06%. The NS Installation topic has the highest how-to type of question 78.06%, showing a necessity for rich resources of guidance to install and manage the NS tools. The Simulation Model Configuration has the highest What type of post 12.02%. This suggests the necessity of general information about supported features on simulation model configuration of the NS. Finally, the Simulation Model Performance Measure has the highest why type of post 2.17%. This suggests the necessity of discussion forums and improved documentation on simulation model performance measurements issues.

RQ$_2$: Summary

Results show that users mainly ask how-to type of questions, followed by what and why, respectively. In addition, we find that NS Installation is the most dominant topic in asking for an instruction (i.e., how-to). This indicates the necessity of providing a guidance to reliably install NS tools.
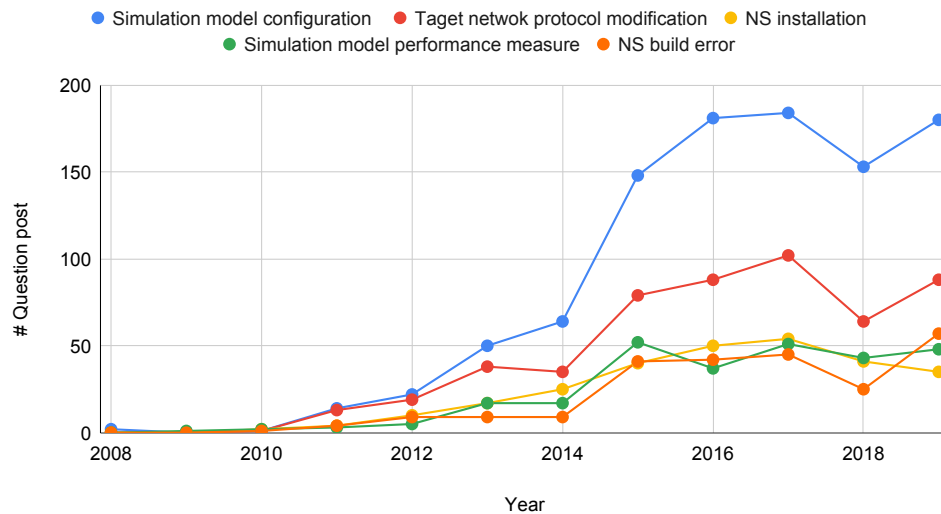
Figure 2. NS topic categories evolution over time

Table 5. Difficulty measure of NS topics. The of number answers, accepted answers and comments are presented in average while the PD score is presented in percentage

Topic name	answers	accepted answers	comments	PD (%)
NS installation	0.99	0.35	1.71	0.057
Target network protocol modification	0.92	0.38	1.19	0.060
NS build error	0.93	0.35	1.27	0.073
Simulation model performance measurement	0.96	0.36	1.22	0.074
Simulation model configuration	0.90	0.35	1.26	0.095

3.3. *RQ₃: What topics are the most difficult to answer?*

Approach: To answer this *RQ*, we investigate the difficulty of the NS topic by utilizing four metrics which are also used in the previous studies [7, 8]. For the first three metrics, we collect the metadata of each topic, that are, answer count, accepted answer count, and comments count to compute the average values. For the fourth metric, that is, the Probability of Difficulty (PD) score, we extract the answer count and view count of each topic. We then calculate the average (i.e., avg. answer count and avg. view count) to find the PD score, formulated as:

$$PD\ score\ \ =\ \frac{Average\ Answer\ Count}{Average\ View\ Count} \times 100\% \tag{3}$$

In general, a high number of views on a topic but a small number of answers indicates that only a small number of people can answer the topic's questions. Therefore, we adopt the PD score to measure the difficulty of the topic. The lower the PD score, the harder it is to answer questions in NS-related discussion topics.

Results: Table 5 shows the difficulty measure of NS topics. We find that NS topics in this analysis do not significantly differ in term of the avg. answer count (0.9-0.99), avg. accepted answer count (0.35-0.38) and avg. comment count (1.19-1.71). Hence, we consider the PD score to determine the most difficult topic to answer. As described in Table 5, NS Installation is the most difficult topic faced by users, followed by Target Network Protocol modification. Although NS build error and Simulation model performance measurement are not as difficult as the top two topics, they have similar level of difficulty according to our analysis. Finally, simulation model configuration is the least difficult topics to answer, accounting for 0.095% of PD score in the result.

RQ$_3$: Summary

`NS Installation` and `Target Network Protocol modification` are the top two most difficult NS-related topics to answer, with PD score 0.057% and 0.060%, respectively.

4.   IMPLICATIONS

The results of this study will help the NS users better understand and focus on the most pressing NS issues. This section describes how our results can help practitioners, researchers, and guides to software development projects.

4.1.   Practitioners

According to Table 5, the most popular topic on how to install NS has less probability to receive accepted answer. The NS community can benefit from these findings to devise better tutorials (i.e., video tutorial, documentation) to reduce the barrier of NS usage. Our findings can also help NS community to prioritize the task by considering the areas of the difficult NS topics while performing experiments. NS installation, build error, simulation model configuration are the topics with highest share without accepted answers. NS developers can take these issues into account to improve the user experience. In addition, Figure 2 shows the trend of NS topic categories evolution over time. This trend hints the necessity of dedicated online blog for each NS that can help to create large community and improve NS usage experience.

4.2.   Researchers

Our large-scale empirical research provides an overall view of NS-related topics being discussed on the SO platform. We found five NS-related topics discussed in SO, that are, Simulation Model Configuration, Target Network Protocol modification, NS Installation, Simulation Model performance measure, and NS Build error. We also focus the most popular and difficult NS topics. Therefore, we encourage researchers to develop techniques to help NS users answer this difficult question.

4.3.   Software Development Projects

The findings of this study recommend for software development projects to consider preparing project-specific discussion forums. As observed in the study, the number of NS-related discussions in SO have increased over time, but the attention from the community is not high. This leads to the most problem discussed remains unsolved. The results indicate that considering to prepare project-specific discussion forums is important. Sharing the project-specific problems in the project-related community forums will increase the probability of getting the solutions. This will also enable developers to provide a community-related information or announcement. As reported by Tantisuwankul et al. [25] that software projects tend to adopt communication channel for both capturing new knowledge and updating existing knowledge, and since the information or announcement that specifically relates to the software projects is important to share amongst the community, providing an NS-specific discussion forum is necessary.

5.   THREAT TO VALIDITY

There are several threats that may affect the validity of the NS study. This section describes the threats to validity in detail.

5.1.   Construct validity

The threats to construct validity may emerge in our experiments. During the SO question extraction phase, we may miss some NS questions due to the tag-based extraction technique. Since the number of this issue is small, thus, the impact of the missing questions is not significant.

5.2.   External validity

The threats to external validity may appear in data preparation phase. We conducted an empirical study of 2,322 NS questions from Stack Overflow, but could not generalize the results to other question-answering online platforms.

5.3. Reliability

We mitigate the threats to reliability by preparing online appendix of the dataset and scripts. This online appendix is described in Section 2.3..

6. RELATED WORK

This section describes the NS-related work. First, we review some prior research on NS and its implementation on network related research domain. Second, we discuss SO-based case studies and studies on topic modeling.

6.1. Network Simulator

There are several studies on NS. Prior works compared NS tools to subjugate barrier to select the suitable one that support users objective [10] and the usage on wireless networks [26-29]. In another paper, Campanile et al. conducted a case study to demonstrate the effectiveness of network simulator in real applications, and modeling studies [27]. Comparative studies on wireless network simulators were also conducted by Lessman et al. [30] and Korkalainen et al. [31] to help other users to quickly identify which simulator is most suitable for their needs. There are also some previous research [32-36] to utilize NS as the tools to perform simulation work for different wireless network scenarios and intrusion detection. As NS implementations increase in academia and industry, we investigate the problems that users are facing. The results of this study provide the research community with insights to understand areas that require more attention.

6.2. Stack Overflow

The SO data have also been analyzed in several studies. In a study by Rosen and Shihab [7], the authors has summarized mobile-related questions from SO to identify specific issues on various mobile platforms The SO dataset was also used to understand the challenges chatbot developers [22]. Mahajan et al. [37] proposed a recommendation system to fix run-time exception by utilizing SO dataset. Riccardo et al. [38] utilized SO dataset in PostFinder system to support software developers with suitable code snippets. Cai et al. [39] proposed a API recommendation method that also depends on SO dataset. Uddin et al. [40], proposed an automated system to mine the API usage. This study also utilize SO as a primary data source. As far as we know, no research has been conducted on SO NS-related posts. Our study complements previous work at SO by analyzing NS-related posts. We collected and categorized NS topics and investigated the popularity and difficulty of the topics. We believe that our research sheds spot light on the areas where NS users are facing challenges.

7. CONCLUSION

To understand the characteristics of NS issues discussed by users, we conducted a large-scale empirical study on 2,322 NS questions posted in SO. In our study, we analyze (i) the types of discussion topics and their popularity, (ii) types of questions that frequently faced by the users, and (iii) the difficulty of topics shared in SO. The results of our study have shown that simulation model configuration is the most common discussed and useful topic amongst the users, while target network protocol modification and NS installation are become the most popular NS-related topics in SO. NS users are frequently ask for an instruction of NS installation by posting a how-to type of question. This suggests the importance of providing improved NS installation document. Furthermore, the findings show that the most difficult NS related questions posted in SO are NS installation and target network protocol modification. Based on this study, we also shows the increase of NS related discussion in Stack Overflow. Therefore, there are many open issues in future work, such as a comprehensive understanding of the evolution of NS-related discussions and further study of NS topics on other online discussion platforms.

REFERENCES

[1] V. Stocker, G. Smaragdakis, W. Lehr, and S. Bauer, "The Growing Complexity of Content Delivery Networks: Challenges and implications for the Internet Ecosystem," *Telecommunications Policy*, vol. 41, no. 10, 1003–1016, 2017.

[2] A. Zagalsky, D. M. German, M. A. Storey, C. G. Teshima, and G. Poo-Caamano, "How the r community creates and curates knowledge: An extended study of stack overflow and mailing lists," *Empirical Software Engineering*, vol. 23, no. 2, pp. 953–986, Apr. 2018.

[3]   M. Squire, "Should we move to stack overflow?: Measuring the utility of social media for developer support," *in Proceedings of the 37th International Conference on Software Engineering* - Volume 2, ser. ICSE '15. Piscataway, NJ, USA: IEEE Press, 2015, pp. 219–228.

[4]   F. Calefato, F. Lanubile, and N. Novielli, "How to ask for technical help? evidence-based guidelines for writing questions on stack overflow," *Information and Software Technology*, vol. 94, pp. 186–207, Feb. 2018.

[5]   S. Wang, T.-H. Chen, and A. E. Hassan, "Understanding the factors for fast answers in technical qa websites," *Empirical Software Engineering*, vol. 23, no. 3, pp. 1552–1593, Jun. 2018.

[6]   S. Beyer, C. Macho, M. Di Penta, and M. Pinzger, "What kind of questions do developers ask on stack overflow? a comparison of automated approaches to classify posts into question categories," *Empirical Software Engineering*, vol. 25, no. 3, pp. 2258–2301, 2020.

[7]   C. Rosen and E. Shihab, "What are mobile developers asking about? a large scale study using stack overflow," *Empirical Software Engineering*, vol. 21, no. 3, pp. 1192–1223, 2016.

[8]   X. L. Yang, D. Lo, X. Xia, Z.-Y. Wan, and J.-L. Sun, "What security questions do developers ask? a large-scale study of stack overflow posts," *Journal of Computer Science and Technology*, vol. 31, no. 5, pp. 910–924, 2016.

[9]   N. Kahani, M. Bagherzadeh, J. Dingel, and J. R. Cordy, "The problems with eclipse modeling tools: A topic analysis of eclipse forums," *in Proceedings of the ACM/IEEE 19th International Conference on Model Driven Engineering Languages and Systems*, ser. MODELS '16, 2016, pp. 227–237.

[10] M. H. Kabir, S. Islam, M. J. Hossain, and S. Hossain, "Detail comparison of network simulators," *International Journal of Scientific and Engineering Research*, vol. 5, no. 10, pp. 203–218, 2014.

[11] S. Baltes, L. Dumani, C. Treude, and S. Diehl, "Sotorrent: reconstructing and analyzing the evolution of stack overflow posts," in *Proceedings of the 15th International Conference on Mining Software Repositories*, MSR 2018, Gothenburg, Sweden, May 28-29, 2018, A. Zaidman, Y. Kamei, and E. Hill, Eds. ACM, 2018, pp. 319–330. [Online]. Available: https://doi.org/10.1145/3196398.3196430.

[12] "Regular expression operations" *accessed 02-01-2020.* [Online]. Available: https://docs.python.org/3/library/re.html.

[13] "Python NLTK" *accessed 02-01-2020.* [Online]. Available: https://www.nltk.org/.

[14] "Gensim model" *accessed 02-01-2020.* [Online]. Available: https://radimrehurek.com/gensim/

[15] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, pp. 993–1022, Jan 2003.

[16] H. Zhang, S. Wang, T.-H. Chen, and A. E. Hassan, "Reading answers on stack overflow: Not enough!" *IEEE Transactions on Software Engineering*, 2019.

[17] S. Liu, R.-Y. Zhang, and T. Kishimoto, "Analysis and prospect of clinical psychology based on topic models: hot research topics and scientific trends in the latest decades," *Psychology, Health Medicine*, pp. 1–13, 2020.

[18] S. Choi and J. Seo, "An exploratory study of the research on caregiver depression: Using bibliometrics and lda topic modeling," *Issues in Mental Health Nursing*, 2020, pp. 1–10.

[19] A. McCallum, "AK McCallum, S Thrun, T Mitchell," *"Mechine Learning,"* vol. 39, no. 2, pp. 103-134, 2020.

[20] S. Boussaadi, H. Aliane, A. Cerist, and P. O. Abdeldjalil, "Modeling of scientists profiles base d on lda."

[21] M. Zahedi, R. N. Rajapakse, and M. A. Babar, "Mining questions asked about continuous software engineering: A case study of stack overflow," in *Proceedings of the Evaluation and Assessment in Software Engineering*, pp. 41–50, 2020.

[22] A. Abdellatif, D. Costa, K. Badran, R. Abdalkareem, and E. Shihab, "Challenges in chatbot development: A study of stack overflow posts," in *Proceedings of the 17th International Conference on Mining Software Repositories*, pp. 174–185, 2020.

[23] "SO Tour" accessed 10-02-2020. [Online]. Available: https://stackoverflow.com/tour

[24] C. Treude, O. Barzilay, and M.-A. Storey, "How do programmers ask and answer questions on the web?(nier track)," in *Proceedings of the 33rd international conference on software engineering*, pp. 804–807, 2011.

[25] J. Tantisuwankul, Y. S. Nugroho, R. G. Kula, H. Hata, A. Rungsawang, P. Leelaprute, and K. Matsumoto, "A topological analysis of communication channels for knowledge sharing in contemporary github projects," *Journal of Systems and Software*, vol. 158, p. 110416, 2019.

[26] F. Wilhelmi, M. Carrascosa, C. Cano, A. Jonsson, V. Ram, and B. Bellalta, "Usage of network simulators

in machine-learning-assisted 5g/6g networks," *arXiv preprint arXiv:2005.08281*, 2020.

[27] L. Campanile, M. Gribaudo, M. Iacono, F. Marulli, and M. Mastroianni, "Computer network simulation with ns-3: A systematic literature review," *Electronics*, vol. 9, no. 2, p. 272, 2020.

[28] T. K. Bhatia, R. K. Ramachandran, R. Doss, and L. Pan, "A review of simulators used for vanets: The case-study of vehicular mobility generators," in *2020 7th International Conference on Signal Processing and Integrated Networks (SPIN). IEEE*, pp. 234–239, 2020.

[29] P. Mercati, R. Ayoub, and T. Rosing, "Reliot: Reliability simulator for iot networks," in *Internet of Things-ICIOT 2020: 5th International Conference, Held as Part of the Services Conference Federation*, , Proceedings, vol. 12405, p. 63 2020.

[30] J. Lessmann, P. Janacik, L. Lachev, and D. Orfanus, "Comparative study of wireless network simulators," in *Seventh International Conference on Networking (icn 2008). IEEE*, pp. 517–523, 2008.

[31] M. Korkalainen, M. Sallinen, N. K¨arkk¨ainen, and P. Tukeva, "Survey of wireless sensor networks simulation tools for demanding applications," in *2009 Fifth international conference on networking and services. IEEE*, pp. 102–106, 2009.

[32] M. A. Saad, M. H. Ali, S. Alani, A. H. Ali, and Y. A. Hussein, "Performance evaluation improvement of energy consumption in ad-hoc wireless network," *Int. J. Adv. Sci. Technol*, vol. 29, no. 3, pp. 4128–4137, 2020.

[33] V. Anupriya, S. Sangavi, P. Shivani, and P. V. Dharany, "An efficient approach for vehicle traffic monitoring by collaborating vehicular mobility module and network simulator 3," in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE*, pp. 379–384, 2020.

[34] S. Godala, and R. P. V. Vaddella, "A study on intrusion detection system in wireless sensor networks," *International Journal of Communication Networks and Information Security*, vol. 12, no. 1, pp. 127–141, 2020.

[35] S. Islam, R. Kuri, M. H. Kabir, and M. J. Hossain, "Exploring congestion control mechanism of tcp vari-ants over wired wireless networks," *International Journal of Scientific Engineering Research*, vol. 9, no. 9, 2018.

[36] R. Kuri, S. Islam, M. J. Hossain, and M. H. Kabir, "Simulation-based comparative analysis on the effect of black-hole attack and rushing attack on the mobile ad-hoc network," *International Journal of Applied Engineering Research*, vol. 14, no. 10, pp. 2383–2387, 2019.

[37] S. Mahajan, N. Abolhassani, and M. R. Prasad, "Recommending stack overflow posts for fixing runtime exceptions using failure scenario matching," in *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, pp. 1052–1064, 2020.

[38] R. Rubei, C. Di Sipio, P. T. Nguyen, J. Di Rocco, and D. Di Ruscio, "Postfinder: Mining stack overflow posts to support software developers," *Information and Software Technology*, vol. 127, p. 106367, 2020.

[39] L. Cai, H. Wang, Q. Huang, X. Xia, Z. Xing, and D. Lo, "Biker: a tool for bi-information source based api method recommendation," in *Proceedings of the 2019 27th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, pp. 1075–1079, 2019.

[40] G. Uddin, F. Khomh and C. K. Roy, "Mining api usage scenarios from stack overflow," *Information and Software Technology*, vol. 122, p. 106277, 2020.

BIOGRAPHY OF AUTHORS

Syful Islam is an assistant professor at the Department of Computer Science and Telecommunication Engineering, Noakhali Science and Technology University in Bangladesh. Currently he is a Ph.D student at the Nara Institute of Science and Technology, Japan. He obtained Master Degree from the same institute. His research interests include mining Stack Overflow and software ecosystems. Further info on his homepage :https://syful-is.github.io/

Yusuf Sulistyo Nugroho is a lecturer and researcher at the Department of Informatics Engineering in Universitas Muhammadiyah Surakarta, Indonesia. He received the Ph.D degree from Nara Institute of Science and Technology, Japan in 2020. His research interests include mining software repositories and software ecosystems.
Further info on his homepage :https://yusufsn.github.io/

Md. Javed Hossain is an associate professor at the Department of Computer Science and Telecommunication Engineering, Noakhali Science and Technology University, Bangladesh. His research interests are Fuzzy Logics, Signal Processing in Wireless Communication, FinFET Nano Devices.
Further info on his homepage :https://nstu.edu.bd/faculty-member/md-javed-hossain-bdr991