A Bird's Eye View of Knowledge Needs Related to Penetration Testing

Akond Rahman and Laurie Williams
aarahman@ncsu.eduandlawilli3@ncsu.edu
North Carolina State University
Raleigh, North Carolina, USA

CCS CONCEPTS

• **Security and privacy → Software security engineering**.

KEYWORDS

knowledge, penetration, security, stack exchange, testing

ACM Reference Format:
Akond Rahman and Laurie Williams. 2019. A Bird's Eye View of Knowledge Needs Related to Penetration Testing. In *Hot Topics in the Science of Security Symposium (HotSoS), April 1–3, 2019, Nashville, TN, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3314058.3317294

1 INTRODUCTION

According to the National Institute of Standards and Technology (NIST), penetration testing is an assessment conducted on software systems to identify vulnerabilities that could be exploited by adversaries [1]. Despite the importance of penetration testing in software security, practitioners search for strategies and guidance on how to get started in the domain of penetration testing. We hypothesize that practitioners have knowledge needs related to penetration testing, which can be synthesized using penetration testing-related questions posted on questions and answer (Q&A) websites. A systematic investigation can identify the knowledge needs of practitioners related to penetration testing, helping the cyber-security community in advancing the field of cyber-security education. *The goal of this paper is to help cyber-security researchers in advancing the field of cyber-security education by analyzing penetration testing-related questions posted by practitioners.*

We answer two research questions:

- **RQ1**: What are knowledge needs of practitioners related to penetration testing?
- **RQ2**: How frequently are practitioners' knowledge needs related to penetration testing viewed? How frequently are the identified knowledge needs answered?

We conduct a systematic investigation with 548 questions posted on a security-related Q&A website called 'Information Security Stack Exchange' [2]. We apply card sorting [3] to identify the knowledge needs. Next, we quantify the frequency of the identified knowledge needs.

Our contribution is a list of knowledge needs related to penetration testing.

Table 1: Selection of Questions for Analysis

Initial question count	51,056
Criteria-1 (Ques. tagged as 'penetration-test')	1,111
Criteria-2 (Ques. with > 0 views)	1,111
Criteria-3 (Ques. with score > 0)	835
Criteria-4 (Ques. that are relevant)	548
Final question count	548

2 METHODOLOGY

We conduct an empirical study using questions collected from the Information Security Stack Exchange, a Q&A website where information security-related questions are posted by users. The Q&A website allows tags to specify the category to which the question belongs to. A question includes a title and a body, which respectively summarizes and provides a description of the need of the question provider. We use the tag 'penetration-test' to identify questions that are related to penetration testing. Following Rahman et al. [1]'s advice we apply a filtering criteria to obtain questions needed for analysis. The final count of questions are displayed in the 'Final question count' row of Table 1. We use the title and body of each question to determine knowledge needs by applying qualitative analysis described below.

RQ1: What are knowledge needs of practitioners related to penetration testing?: Similar to Rahman et al. [2], we use card sorting, a qualitative analysis technique to identify the knowledge needs from the penetration testing-related questions. Card sorting is a qualitative analysis technique to identify categories from textual artifacts [3].

RQ2: How frequently are practitioners' knowledge needs related to penetration testing viewed? How frequently are the identified knowledge needs answered?

By capturing the view count of each identified category, we can capture the interest of practitioners who are registered and not registered to Information Security Stack Exchange. We use the metric 'view count per question (VQ)', to answer RQ2 using Equation 1:

[1] https://nvd.nist.gov/800-53/Rev4/control/CA-8

[2] https://security.stackexchange.com/

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
HotSoS, April 1–3, 2019, Nashville, TN, USA
© 2019 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-7147-6/19/04…$15.00
https://doi.org/10.1145/3314058.3317294

Table 2: Ans. to RQ1: Five Knowledge Needs Related to Penetration Testing are Presented with Brief Descriptions.

Category (Count)	Description	Example
Starting Point (236)	The questions that belong to this category relate to the need for general advice on how to conduct penetration testing in their organization or for a specific platform, such as web services, or for career purposes. Our finding suggest that practitioners frequently search for concrete guidelines on how to learn and apply penetration testing.	*"What are good resources for hands-on practice on network penetration testing?"*
Attack Simulation (229)	While conducting vulnerability exploits, practitioners face challenges and ask for help in Q&A websites. Examples of vulnerability exploits include SQL injection simulation, cross site forging, and password attacks. For this particular knowledge need, practitioners have the resources to exploit a particular vulnerability, but they fail to successfully exploit the vulnerability due to specific challenges for example, not understanding output of the tool, which was used to simulate the attack. The prevalence of this particular knowledge need underlines the importance of synthesizing knowledge on how to exploit vulnerabilities and conduct successful attacks.	*"How can I simulate a man-in-the-middle attack in an android emulator?"*
Best Practices (43)	Practitioners ask for advice on the best way to perform penetration testing. Practitioners' knowledge needs include using an appropriate tool or technique to perform a certain attack. As another example, practitioners also ask about best strategies to create a penetration testing process in an organization.	*"What are best practices to implement a secure CAPTCHA?"*
Legal concerns (32)	Before conducting penetration testing, practitioners ask about legal concerns i.e. if conducting penetration testing will lead to law breaking.	*"What are the penal consequences for a passive or active scan on a WebApp with no damage?"*
Ethics (8)	We observe practitioners to be concerned about the ethical issues related to penetration testing, even if performing penetration testing does not lead to law breaking.	*"What are ethical hacking requirements on a banking institution?"*

$$\text{VQ} = \frac{\text{total view count for questions related to a knowledge need}}{\text{total questions related to a knowledge need}} \quad (1)$$

We answer the second part of RQ2 by calculating the answer to question ratio (AQ) using Equation 2:

$$\text{AQ} = \frac{\text{total answer count for questions related to a knowledge need}}{\text{total questions related to a knowledge need}} \quad (2)$$

3 FINDINGS

We present the answers for our two research questions in this section.

Answer to RQ1: What are the knowledge needs of practitioners related to penetration testing?: We identify five knowledge needs related to penetration testing. We describe each knowledge need with brief descriptions in Table 2:

Answer to RQ2: How frequently are practitioners' knowledge needs related to penetration testing viewed? How frequently are the identified knowledge needs answered?

We provide detailed distribution of the view count per question (VQ), and answer per question (AQ) values in Table 3. We observe the knowledge need with the highest and lowest view count is respectively, starting point and legal concerns. The answer count per question is highest and lowest respectively for the knowledge need legal concerns and attack simulation.

Summary We summarize our findings as following:

- We identify five knowledge needs related to penetration testing: starting point, attack simulation, best practices, legal concerns, and ethics.
- The knowledge need that is most frequently viewed is legal concerns, whereas the knowledge need that is least frequently viewed is ethics.

Table 3: Ans. to RQ2: VQ and AQ for Five Knowledge Needs

Knowledge need	AQ	VQ
Attack simulation	1.7	3581.9
Ethics	2.0	362.1
Starting point	2.1	3818.8
Best practices	2.4	1028.8
Legal concerns	2.9	5130.1

- The knowledge need for which we observe the least amount of answers posted is attack simulation, which indicates limited availability of knowledge resources to conduct software and system attacks.

4 CONCLUSION

We have conducted an empirical analysis with 548 questions collected from Information Security Stack Exchange, a Q&A website, to observe what knowledge needs are asked by practitioners related to penetration testing. We identify five knowledge needs namely, starting point, attack simulation, best practices, legal concerns, and ethics. We observe the knowledge needs that have lowest amount of answers is attack simulation suggesting limited availability of knowledge resources to conduct software and system attacks. Based on our findings we advocate for enhanced cyber-security education in academia and industry, which will incorporate discussion on application, ethical issues, and legality of penetration testing. We also advocate for future research that synthesizes what concrete steps need to exploit software and system vulnerabilities.

REFERENCES

[1] Akond Rahman, Effat Farhana, and Nasif Imtiaz. 2019. Snakes in Paradise?: Insecure Python-related Coding Practices in Stack Overflow. In *2019 IEEE/ACM 16th International Conference on Mining Software Repositories (MSR)*. To appear.

[2] Akond Rahman, Asif Partho, Patrick Morrison, and Laurie Williams. 2018. What Questions Do Programmers Ask About Configuration As Code?. In *Proceedings of the 4th International Workshop on Rapid Continuous Software Engineering (RCoSE '18)*. ACM, New York, NY, USA, 16–22.

[3] T. Zimmermann. 2016. Card-sorting: From text to themes. In *Perspectives on Data Science for Software Engineering*, Tim Menzies, Laurie Williams, and Thomas Zimmermann (Eds.). Morgan Kaufmann, Boston, 137 – 141.