Understanding crypter-as-a-service in a popular underground marketplace

DECISION: Accept as short

AUTHORS: Alejandro de la Cruz and Sergio Pastrana

Summary of Reviews

- Review 1: 0 (4)
- Review 2: -1 (3)
- Review 3: 2 (4)

Reviews

Review 1

Total score: 0

Overall evaluation: 0

Reviewer’s confidence: 4

Summary

The manuscript "Understanding crypter-as-a-service in a popular underground marketplace" presents an exploratory descriptive analysis of the cryptography and encryption sub-forum of the popular clear web forum HackFroums. The authors claim that their paper "provides the first study on an online underground market dedicated to crypter-as-a-service" (p. 1) and indicate that their study fills a gap in the literature around this particular service. Through a series of basic descriptive analyses of scraped forum data, the authors provide some general figures on the supply and demand in this market and the interaction between users and posts. In addition, they complement their analyses with a case study of a crypter that was apparently offered on the market as early as at least 2017 and "already fingerprinted by the AV industry" (p. 7) to examine the veracity of the claims made by sellers in the marketplace. Descriptive analyses of this form of CaaS are somewhat crude and the findings to advance the state of the art rather limited.

Major comments

1. After reading the manuscript, a fundamental question I have is what the research question is. The authors describe three contributions of their work in the introduction, but why are they important to science and/or practice? In the absence of a research question, assessing the appropriateness of the analyses and the relevance of the findings is a difficult task.

In relation to the above, I find it difficult to understand the rationale behind some of the analyses presented in the paper. For example, on p. 4 the authors state that "The last approach investigated in the generic analysis is to get the number of posts asking for help or for a specific product instead of selling a crypter", but they do not say why they investigate this aspect. On the same page, the authors state that they "also attempt to correlate this [forum] activity with real-world events, such as cyber incidents or the 2020 shutdown breakout", but again I don't understand what the purpose of such analysis is.

2. Regarding the case study in section 5, I have doubts about its validity. The authors state that they "got access to a released version [of ByteCrypter v3] from June 2022, already fingerprinted by the AV industry" and that "the updated, most recent version, allegedly remains undetected from AVs" (p. 7). Given that, as the authors themselves indicate, the stub has been fingerprinted, what kind of claims can be made from the results about the proportion of engines that flag the crypted versions of the binaries as malicious, and how reliable such claims are?

3. In Figure 4, the authors present the temporal distribution of their data, namely threads created and comments. Although the double axis makes it difficult to interpret the figure, there are two aspects that stand out to me.

The first and most obvious is the drop in activity in 2010. The authors explain this drop by the LulzSec data leak in early 2011, but if the drop in activity occurred in 2010, as the graph seems to indicate, the data leak would not explain this phenomenon. What alternative explanation can the authors offer for such a drop in this sub-forum activity?

In this regard, the authors state that they cannot "correlate [la spike en comments en 2011] with a significant event that might be related to this anomaly". How then should readers interpret this data?

The second aspect is the sharp decrease in comments (or should I say total absence of comments) observed in 2022 despite the fact that new threads are still being registered. Does this mean that dozens of new threads were created at that time and received no comments for almost a year? Is it possible that something went wrong during data collection?

4. In section 4.3 of the manuscript, the authors examine the "social network" of the forum data they have collected, but it is not clear to me how the authors have constructed the network. What do the nodes and edges represent? Is it a one-mode or two-mode network? The authors indicate that it is a directed network, how have they operationalized the direction?

5. There seem to be issues with some references.

On page 4, the authors indicate that "other studies show similar trends in other activities in this forum [2], [15]", it is not clear to me exactly what the authors are referring to with this phrase, but I have skimmed through those studies and found no evidence to support that claim (or my interpretation of it).

On the other hand, when the authors explain the metrics that they have chosen to describe their network in section 4.3, they reference two blog posts ([19, 20]). I think there are enough open access scientific texts on network analysis to rely on in blog posts.

The link to the crypter reference that the authors select for their case study is "reported as unsafe" by Microsoft, so I don't think it is appropriate to include it in the reference list of a possible publication.

Finally, I wonder how reliable the figures on share of devices covered by Windows OS are that the authors reference on p. 8 ([30, 31]).

Minor comments

6. On p. 4 the authors mine text to describe the most frequently used terms in the sub-forum. Although a word count is useful to see the most common words, this analysis is often complemented with a term frequency-inverse document frequency (TF-IDF) analysis to identify which terms are the most important. Have the authors considered such an analysis? If so, why did they discard it?

On the other hand, I think the authors should refrain from making categorical statements based on counts of words, such as “a key takeaway is that users are mostly concerned with the product sold, its effectiveness and the type of crypter” (p. 4). I do not believe that these types of conclusions can be drawn from these analyses.

7. It should be noted that the list of keywords used to request help or a specific product on p. 4 is not exhaustive, so it is possible that the proportion of messages classified as help-seeking is underrepresented.

8. On p. 4, the authors further examine the top 100 posts by number of visits. Could the authors reflect on what kind of insight would result from such analysis and what it would mean for the enire crypter-as-a-service econsystem in the sub-forum? For example, what implications would it have for the analysis if older posts had more views?

9. On p. 5, the authors state that “two thirds of the users only reply in one post, this suggests that either they purchased the crypter and met their expectations [...]”. Isn't it possible that users also responded because the product did not meet their expectations?

10. Considering that the authors conducted an exploratory and descriptive study, I was expecting an interesting discussion that would shed light on the relevance of the findings in the broader context of CaaS. Instead, the discussion is brief and superficial. I miss some discussion on how the ecosystem that the authors call "crypters-as-a-service" (but which could be called "obfuscation-as-a-service" in accordance with the literature), differs or aligns with the broader CaaS ecosystem. Can the authors provide additional context for these findings?

11. On p. 2, the authors “describe the three most common roles involved [in the crypter-as-a-service model]”, but how do they know that this is how cybercriminals organize around this as-a-service model? What evidence do they have?

12. Due to time constraints, I was unable to verify priority in discovery claims in the manuscript (such as the one the authors claim on p. 1 or the one they attribute to Motoyama on p. 7), but given that the authors claim that their paper provides the first study on an online underground market dedicated to "crypter-as-a-service", I would suggest a literature review on "obfuscation-as-a-service", which as of April 22, 2024, yields 31 hits on Google Scholar.

13. Although in general I was able to follow the text without problems, it was difficult in some parts to understand what the authors meant. I would recommend that the authors check the grammar and spelling thoroughly. In some cases, the grammar and spelling are correct but key information is missing. On other occasions, it is a matter of style. For example, I found it very difficult to understand the paragraphs on p. 6 where the authors explain the metrics of in- and out-degree and eigenvector centrality.

14. On the same page 6, the authors state that “there is no actual relation between the degree and popularity of users”, as shown by “a preliminary study”. Could the authors clarify what they mean here? Is it a preliminary study published elsewhere (in which case we would need a reference), or is it a pilot study by the authors (in which case the term "pilot study" might be clearer), or perhaps something else?

The authors then conclude that the “crypter marketplace is an isolated environment” (p. 6). Is it not possible that the crypter sub-forum is simply less popular within the entire forum?

15. It is not clear how the “ranking in the forum” (p. 6) is calculated.

16. On page 4, the authors refer to India as an occidental country. This may be a matter of perspective, but I believe it is commonly referred to as an oriental country.

17. Although I appreciate the authors' effort to show a screenshot of the interactive graph of the forum data, Figure 6 is indecipherable. Instead of the screenshot, I suggest that the authors redesign a figure that allows to visualize the network clearly. For example, it would be important that it allows to understand the structure of the network and, if it includes text, that the font size is large enough to be legible. The interactive network is a great addition but is not a key takeaway by itself.

18. In Figure 4, the authors state that in 2020 there is a “slight increase in the [forum] activity, potentially derived from the lockdown effects [4]”. The increase does indeed appear to be slight. Is this a significant increase or could it be due to chance?

19. In Table 2, the authors include the category “Users that create”. Do the authors actually mean “users that *only* create [threads]?

20. On p. 8, the authors state that “we show that all the products reviewed are Windows and .NET, [...]”. I may have missed it, but where exactly do they show that?

Conclusion

In my opinion, the authors still need to address some important issues before the article is ready for publication. For this reason, I recommend major revisions.

Review 2

Total score: -1

Overall evaluation: -1

Reviewer’s confidence: 3

This paper analyses crypter-as-a-service activities on HackForums.

The authors crawl HackForums' 'Cryptography and Encryption' board, "to get all the threads announcing crypters". It is not clear how this criterion was instrumented -- through a text search for 'crypter'? Some other method? This is a fairly important missing detail. At the same time, Section 3.2 describes ordinary adversarial crawling techniques and safeguards that don't seem necessary to describe here.

The authors provide four data-driven analyses of the resulting data: (1) the textual data shows nothing particularly unexpected, confirming that commentary mostly covers the products, their marketing and their effectiveness, and requests for products. (2) an activity-over-time analysis finds an odd historical drop in posting which may be HackForums-wide. (3) Manual analysis of adverts reveals some centralisation of the market and that approximately 1/4 of posts used image 'pamphlets' to advertise crypters, as well as different categories of product. (4) Social network analysis of the board which turns up very little -- the authors list the interactive graph as an output here, but given they themselves find little to extract from it, and they note several limitations with its implementation, this seems of dubious further value.

The authors also carry out a small experiment with an old version of one of the top-selling crypter products, running crypted versions of both a Monero wallet and the Windows calculator application through VirusTotal. They find that the outdated crypter's output binaries are detected by many AV products, as could be expected.

This paper contains a number of grammatical errors, and needs careful proofreading by a native English speaker. At some points it was quite difficult to understand what the authors were trying to say.

Overall, I think the concept of an analysis of the crypter-as-a-service ecosystem is a valid one, and could make for a good paper for a venue like WACCO, but this work is currently underdeveloped. At the end of the paper I am not sure what I now know about the crypter ecosystem that I did not know at the end of Section 2. The authors should try and refine the presentation of their data-driven analyses so that the significance of findings is clearer, and try to connect what they are learning from the forum to operational concerns or research challenges. Elements such as the ByteCrypter case study seemed set up without a clear research question in mind -- the authors had previously described that AV systems detect crypter stubs, and it's not clear what was learnt by confirming this in a single case.

Review 3

Total score: 2

Overall evaluation: 2

Reviewer’s confidence: 4

This paper examens the ‘crypter-as-a-service’ market through the lens of underground forums. The authors present numerous interesting findings stemming from analyses of top posts and social connections in a 'crypter' supply and demand platform.

Strengths

+ Well-structured and readable paper, that empirically investigates the crypter-as-a-service ecosystem, leveraging measurements of underground forums.

+ This is a good data analysis paper that uses the crawled data effectively to investigate a number of patterns theoretically grounded in earlier work.

Weaknesses

+ The authors leave some of the potential explanations for their findings to the reader.

+ The analysis is restricted to one underground forum, yet the underground economy is a larger body of these forums. This leaves one to wonder how representative the results are.

Comments for the authors

I really appreciate the work that went in to this paper. The paper is structured well, and its contributions are clear. The conceptual basis is solid, and the results show interesting patterns. My only suggestion would be to the authors to elaborate on the potential explanations of their findings - why do we observe these patterns? Next, I would highly appreciate a section where the authors position the used dataset in the ecosystem of underground forum, to speak to the question of the representativity of their analysis.