Chapter 9: Conclusions and Future Work

9.1 Conclusions

The work done in this thesis was interdisciplinary. It used the latest advancement in AI to solve challenges in web security. The thesis proposed and provided web security solutions for various platforms like computers, mobiles, etc. It uses various ML technologies like conventional ML, deep learning, and FL to solve some intriguing challenges in web security.

Chapter 2 proposed various data collection and pre-processing techniques for ML-based study of malicious webpages and Android hybrid apps. This chapter also introduced and described MalCrawler, which was developed as part of this research to seek and crawl malicious webpages. The datasets discussed in Chapter 2 were further analyzed and visualized in Chapters 3. Chapter 4 described the research carried out with conventional ML techniques in attribute selection and classification for identifying malicious webpages. Chapter 5 and Chapter 6 used deep learning for the classification of malicious webpages. While the work in Chapter 5 used structured data as input for learning, Chapter 6 utilized unstructured raw web content as input. The latter used both unsupervised learning with autoencoders and supervised learning with DNN on the raw web content. Chapter 7 and 8 were dedicated to web security on the mobile platform. Chapter 7 used centralized ML to analyze hybrid apps on the Android platform. Based on the ML model developed and its analysis, vulnerable apps were identified on the Google Play store, and various recommendations were made to improve the security of the hybrid apps ecosystem. Also, two Android apps were developed as part of this study to understand hybrid apps behavior and to monitor them. Chapter 8 used FL for proposing a privacy preserving mobile web security solution. It is the first known attempt to use the novel FL technology in web security, and this work has laid a foundation for the next generation of mobile security solutions using FL. Notable achievements of this thesis can be summarized as:

- MalCrawler, developed as part of this thesis, provided an effective means of crawling and collecting malicious websites. It performed well as a focused crawler and could successfully crawl malicious websites overcoming evasion techniques employed by such sites.

- Malicious webpages dataset and hybrid apps dataset were compiled. Both these datasets can support future research in ML based web security solutions.

- The thesis provides a detailed insight into the attributes that can be used for the classification of malicious webpages. It explored various conventional ML and deep learning methods for the classification of webpages, including using structured and unstructured raw web content as input. It developed efficient ML models for classifying malicious webpages, which surpassed previous similar solutions in classification results.

- As part of the thesis, hybrid apps security on the mobile platform is analyzed using ML. Based on this analysis, suggestions are provided for improving the security architecture of these apps. Also, the ML-based solution proposed could detect vulnerable hybrid apps on the Google Play store. Further, an app named 'WebView Tool' [195] was developed to understand the Android WebView architecture, and a background app named 'WebView Monitor' [191] was designed to report security breaches by hybrid apps.

- The thesis proposes a FL-based web security solution for the mobile platform. The proposed solution could overcome limitations of centralized ML-based security solutions, like privacy, poor data diversity, insufficient user data or unavailability of the latest data, etc. Moreover, the solution is self-evolving, i.e., it learns continuously to handle newer unseen threat vectors.

- Solutions proposed and developed in the thesis are ready for deployment in live scenarios.

9.2 Future Directions

The work done in this thesis has opened up scope for future research in this interdisciplinary field. Scope for future work includes, but not limited to, the undermentioned ideas:

- The work done in this thesis was limited to the field of web security. The future scope can be widened to include application security and network security as well. Accordingly, research can be pursued to design Anti-Virus, Network Intrusion Detection applications, etc., on the lines of ML and FL-based solutions proposed in this thesis.

- MalCrawler was developed in this thesis to collect malicious webpages from the Internet. However, this technology, with slight modifications, can be used by search engines to avoid malicious webpages. Thus, search engines can use this technology to improve the safety of users browsing the Internet. Also, security agencies can use MalCrawler for searching the unindexed Internet, the Darknet, which is the playground for various criminal activities. Also, MalCrawler's current design offers scope for improvement, which may be taken up as part of further research. It may be improved further to incorporate the following features: parallelization of crawler instances, adaptive crawl speed to emulate various user responses, web response for CAPTCHA and login modules, and support for cloud deployment and scalability.

- The conventional ML and deep learning models, which were developed as part of this thesis for predicting malicious webpages, may be used for making a Browser Security Plugin. This plugin will help users avoid malicious websites by dynamically predicting malicious pages before they are rendered on the browser. Also, the use of *'Tensorflow.js*' [150] can be explored for training and running these models directly on the web browser. *'Tensorflow.js'* is a new ML framework based on JavaScript, which can be run on browsers. Using this, both model training and prediction can be executed on the browser's environment. It

will be indeed interesting to see how these security solutions perform on the browser using this JavaScript-based framework.

- In this thesis security analysis of Android hybrid apps was carried out using ML. A similar study may be carried out for the iOS platform. iOS has an equivalent of the WebView component of Android; it is called 'UIWebView' [196]. Thus, analysis of hybrid apps on the iOS platform using 'UIWebView' can be carried out on similar lines. Further, this analysis of hybrid apps had used the SVM classifier. A similar analysis using deep learning may improve results, and thus, may be planned as part of the future scope.

- A FL-based cross-device web security solution was developed for this thesis. This solution had used deep learning for local training on the clients. As part of the future scope, unsupervised learning techniques may be explored, as it would reduce the dependence on any third-party app for data labeling. Future plans may also include extending this FL model to iOS devices, to provide a seamless solution that will work for both the popular mobile platforms (Android & iOS). Furthermore, a plan to build an 'Android Total Security Suite' using FL may be explored, which will provide a comprehensive security shield to mobile devices from all threats. Lastly, the possibility of developing a privacy-preserving FL-based browser security plugin may also be explored.