CHAPTER 6
CONCLUSION

6.1 BACKGROUND

This chapter discusses the achievement of this research in terms of contribution and originality of work. In addition to these, limitations are also included so that it can be improved in the future work.

6.2 ACHIEVEMENTS

This research managed to achieve all of the three objectives. In the objective one, an understanding of the AIS algorithms with BIS are explained in each phase of the IMSM model. A method such as reviewing literatures and analysis of the theories and concepts help in achieving the said objective. For the objective two, the research manages to enhance the method of clustering spam messages by introducing a new algorithm named as HICNA while for the third objective; identifying the reliability and performance of the proposed algorithms and making comparison with existing method were conducted.

This research is about classifying or clustering SMS spam into defined groups. A number of achievements have been made through this research. The **first** contribution enables to show the mapping between the theory of BIS and AIS. Previous researches studied on the relation of BIS with the AIS but there was not many publications found that shows the mapping between these two theories. There are four types of AIS algorithms; Negative

Selection, Immune Network Theory, Clonal Selection, and Danger Theory. All of these theories have a relationship with the immune system and their roles are available in a certain level of the human immune system. By showing BIS and AIS mapping using diagram, it would give more understanding on how they should work and what their roles are in the human body.

The **second** contribution is the spam management model proposed in this research. Integrated Mobile Spam Model (IMSM) is the model which consists of three phases to manage spam in mobile devices. Detection, classification and severity determination are three phases in the IMSM model and each phase is related to one another. Detection is the process to identify the SMS messages into ham or spam. Then the spam messages are used in the second phase to classify them into several categories of spam. Lastly, the severity level of each spam message will be determined according to the defined category. The main characteristic of IMSM is that it uses AIS algorithms for all the three phases, offering one-single solution for managing spams.

The **third** contribution is the algorithm used for classification purposes. As this research focuses on the second phase which is classification or clustering, an enhanced algorithm called Hybrid Immune Clonal Network Algorithm (HICNA) was introduced to classify the spam messages into categories. This algorithm is a combination of two types of AIS algorithms (i.e. Clonal Selection and Immune Network Theory). There are no algorithms that have been developed for classifying spam messages using the combination of two

AIS algorithms from previous researches. In addition, this algorithm can also be applied for the detection phase and result shows that it is reasonable to be used.

The **fourth** contribution is by introducing 16 clusters or categories of spam messages with keywords in each cluster. Delaney et al., (2012) introduced 10 clusters of spam messages. Because of limitation in terms of keywords and meaning in each cluster, this research proposed a total of 16 clusters. Each cluster was enhanced by having its own keywords and meaning, addressing the limitation with previous clusters.

In terms of originality of work, series of experiments have been conducted such as detection and clustering using WEKA, detection of SMS messages using three features (i.e. length of messages, special characters and keywords) using five proposed algorithms and clustering using HICNA. Besides, this research also helps to further analyze the clustering SMS spam as this process has not been widely studied by researchers in mobile devices. In addition, this research can be used as a role base for mobile spam engine because it involves the process of detection and clustering spam messages.

6.3 LIMITATIONS

Although this research achieved the aforementioned objectives, there are limitations throughout. The first limitation is the dataset used for this research. The number of dataset for spam messages is not many and difficult to obtain. Hence, all spam messages need to be collected from different datasets and combined into one dataset for spam messages named FadhilahSpam dataset.

The second limitation is HICNA still depend on the expert judgment for the clustering process. After clustering the spam messages, results show either in True Positive (i.e. messages that correctly classified into groups) or False Positive (i.e. messages that incorrectly classified into groups). Depending on the result from expert judgement would effect the decision as different expert would judge the decision differently.

The third limitation is the mobile environment for detection and clustering is different. Thus, it needs different approaches although using the same AIS algorithm.

The fourth limitation is the time taken for detection and clustering is not stated to see the performance using different algorithms and approaches.

6.4 FUTURE WORKS

For the future work, this research will continue for severity determination to identify the level of danger in each spam messages according to the category of messages. Besides, the HICNA will be improved by automating the process of expert judgment and all spam messages can be categorized into identified group. It is hoped that this research inspires and help other researchers to study clustering spam messages using other methods and algorithms.

6.5 SUMMARY

Several achievements such as the mapping of AIS with BIS, spam management model named IMSM and a new algorithm for clustering spam messages named HICNA are hoped to provide contributions in research related to the computational area. However, there are some limitations that need to be improved for the future work.