CHAPTER 1

INTRODUCTION

1.1 Background

Cybercrime had always been a concern to Internet users and businesses. It was defined as criminal action that uses or targets computers [1]. There were many kinds of cyber attacks such as viruses, spam emails, Trojan, phising, and carders. The purpose of these attacks ranged from stealing confidential information to cyber terrorism. Unfortunately the frequency of cybercrime activities had increased over the years [2]. Advanced technologies were used by the cybercriminals to ease their actions. One of these was the one that was called bot. [3] defined bots as autonomous programs that do some actions automatically, and botnets as networks of autonomous programs that act on instructions. Botnets provided convenience for the cybercriminals that they did not have to do the work by themselves and for the rapid rate of actions. Botmasters (designers of botnet) were reported to make a lot of money by offering their botnet services [1].

Web spambot was a specific variant of bot which was used to spread spam contents on the web [4]. According to [5], the global spam rate in 2010 averaged 89.1% and spambots were responsible for the 88.2% of all spam. Such high rate of spam created problems for Internet users and businesses. Recently spam also targets community websites such as social networking sites, blogs, wikis, and online forums [6]. A website full of spam annoyed its visitors. To make things worse, some of those spam contents were designed for phising purpose which was harmful to the users. Internet businesses
also suffered from spam that their precious resources were wasted by unsolicited contents. Moreover it also damaged the businesses’ reputation.

A number of solutions were proposed to fight spam, either by prevention or detection. Unfortunately most of those techniques were focusing on the spam content, not on the source [7], which were not very effective as the spammers could easily change or modify the content to bypass the security.

This research proposed a solution to detect bots by observing web navigation behavior. The technique used aimed to stop spam dissemination from its source directly, the bots, which some other spam filtering techniques did not do.

### 1.2 Scope

The scope of this research was to create a program that can differentiate between human and bots by observing web navigation behavior. The program recorded both web activities and their session IDs, hence essentially recording the web navigation behavior.

Link obfuscation method was incorporated to work together with the program by creating decoy links. The thesis would explore some classifiers such as Naive Bayes, Support Vector Machines, and k-Nearest Neighbor, to optimize the detection of a bot session. The result was to be measured through the false positive and false negative of the bot/human classification. The program was to be implemented in a web application built in PHP. The author would create PHP files to be included into it. After implementing the technique, the website would have the ability to:

- Generate decoy links
• Record users navigation
• Derive actions from users’ navigation records
• Determine which session is generated by bot and which one is generated by human
• List suspicious IP addresses and usernames which are allegedly used by bot sessions

1.3 Aims and Benefits

This research aimed to implement a technique that can effectively limit bots’ activities on the web which would reduce the number of spam they generate.

The proposed solution should give benefit to Internet users, that they would receive less unsolicited contents. It could also help businesses to reduce the damage (economic and reputation) caused by spam.

1.4 Structures

This thesis would be structured in the following way

Chapter 1 INTRODUCTION

Brief explanation about the thesis background, scope, and aim & benefits.
Chapter 2  THEORETICAL FOUNDATION

Description of theories and related works that would be used as basis for creating the thesis.

Chapter 3  PROBLEM ANALYSIS

Discussion of the problem to be solved in this thesis.

Chapter 4  SOLUTION DESIGN

Explanation of the proposed solution’s design, including diagrams.

Chapter 5  RESULTS

Presentation of the experimental results from the research in tables and diagrams.

Chapter 6  EVALUATION

Evaluation on the results achieved by this thesis, as well as the obstacles and limitation of research.

Chapter 7  CONCLUSION AND RECOMMENDATION

Conclusion of the thesis and recommendations for future works.