Accepted Manuscript

Investigating the deceptive information in twitter spam

Chao Chen, Sheng Wen, Jun Zhang, Yang Xiang, Jonathan Oliver, Abdulhameed Alelaiwi, Mohammad Mehedi Hassan

PII: S0167-739X(16)30154-6
DOI: http://dx.doi.org/10.1016/j.future.2016.05.036
Reference: FUTURE 3060

To appear in: *Future Generation Computer Systems*

Received date: 15 November 2015
Revised date: 1 May 2016
Accepted date: 25 May 2016

Please cite this article as: C. Chen, S. Wen, J. Zhang, Y. Xiang, J. Oliver, A. Alelaiwi, M.M. Hassan, Investigating the deceptive information in twitter spam, *Future Generation Computer Systems* (2016), http://dx.doi.org/10.1016/j.future.2016.05.036

Investigating the Deceptive Information in Twitter Spam

Chao Chen[†], Sheng Wen[†], Jun Zhang[†], Yang Xiang[*†], Jonathan Oliver[§], Abdulhameed Alelaiwi[‡], Mohammad Mehedi Hassan[‡]

[†] *School of Information Technology, Deakin University*
221 BurwoodHwy, Burwood, Vic 3125, Australia

[§] *Trend Micro, Australia*
606 St Kilda Road, Melbourne, Vic 3004, Australia

[‡] *College of Computer and Information Sciences, King Saud University*
Riyadh 11543, Saudi Arabia

Abstract

Online Social Networks (OSNs) such as Facebook and Twitter have become popular communication and information sharing tools for hundreds of millions of individuals in recent years. OSNs not only make people's life more connected, but also attract the interest of spammers. Twitter spam generally contains deceptive information, such as "free voucher" and "weight loss advertisement" to attract the interest of victims. A comprehensive analysis on the deceptive information will be of great benefit to the detection of Twitter spam. This paper presents a study of deceptive information in Twitter spam. The analysis is based on a collection of over 550 million tweets with around 6% spam. We find that various deceptive content of spam performs differently in luring victims to malicious sites. We also find the regional response rate to various Twitter spam outbreaks vary greatly. These two factors can contribute to improve the performance of spam detection techniques.

Keywords: Online Social Network, Big Data, Twitter Spam Analysis

1. Introduction

Twitter has now become one of the largest online social network sites. Over 500 million registered users spend vast time making friends with people