Chapter 9
Online User Interaction Traits in Web-Based Social Biometrics

Madeena Sultana
University of Calgary, Canada

Padma Polash Paul
University of Calgary, Canada

Marina Gavrilova
University of Calgary, Canada

ABSTRACT

During the Internet era, millions of users are using Web-based Social Networking Sites (SNSs) such as MySpace, Facebook, and Twitter for communication needs. Social networking platforms are now considered a source of big data because of real-time activities of a large number of users. In addition to idiosyncratic personal characteristics, web-based social data may include person-to-person communication, profiles, patterns, and spatio-temporal information. However, analysis of social interaction-based data has not been studied from the perspective of person identification. In this chapter, the authors introduce for the first time the concept of using interaction-based features from online social networking platforms as a novel biometric. They introduce the concept of social behavioral biometric from SNSs to aid the identification process. Analysis of these novel biometric features and their potential use in various security and authentication applications are also presented. Such applications would pave the way for new directions in biometric research.

INTRODUCTION

Biometric identification is defined as “an automated process of recognizing persons from their characteristics” (Jain et al., 2004). Traditionally, two types of biometrics: physiological and behavioural, has been considered (Jain et al., 2004). Physiological biometrics relies on physical attributes of individuals, such as face, palm, fingerprint, iris, ear etc. Behavioural biometrics are based on human activities dictated by person’s behaviour, such as handwriting, gait, signature and voice, which also afford discriminability among individuals. Behavioural biometrics are more vola-
Online User Interaction Traits in Web-Based Social Biometrics

Tile to changes one undergoes through the lifetime, but have the advantages of being non-intrusive and cost effective over physiological biometrics (Yampolskiy & Govindaraju, 2008). It is gaining an increasing demand for person identification and verification purposes to reduce security threats, especially in the cyber world (Sourin, 2006, Saeed & Nagashima, 2012).

In this chapter, we first classify the existing behavioural biometrics into two main categories: machine-independent human behaviour and machine-dependent human behaviour. Machine independent behavioural biometrics are well established in the community, with most of the early works falling under this category. For example, person identification from his/her unique signature (Monwar & Gavriloa, 2008, Saeed, 2014), gait or walking style (Bazazian & Gavriloa, 2012), or from a piece of drawing (Al-Zubi et al., 2003) are being studied over last decade. The latter category is emerging compared to the former and more demanding field of research due to increasing number of security threats in virtual domain. Because of the wide range of applications of computing devices and software, users can interact with machine in many ways. We classify machine dependent human behaviours further into the following three sub-categories based on the type of communication with the machine: interaction-based, style-based, and intelligence-based.

First subcategory is based on human interaction with input device that does not consider any knowledge or intelligence of the user. Some examples are keystroke dynamics (Bakelman et al., 2013), mouse dynamics (Jorgensen & Yu, 2011), touchscreen interaction (Bo et al., 2013, Frank et al., 2012) etc. Second subcategory takes into account the style or preference of a user during interaction with the computer. Coding style of a programmer (Spafford & Weeber, 1993), browsing style (Olejnik & Castelluccia, 2013) or handshaking style (Guo et al., 2013) would fall into this category. Finally, the third subdivision includes human intelligence, knowledge, and skills into account during interaction with software such as game playing strategy (Yampolskiy & Govindaraju, 2010), car driving skill (Igarashi et al., 2004), hobby or habits (Jiang et al., 2013) etc.

However, human activities with machine, especially with the computing devices, are not limited to programming, computing, gaming or simply typing. Now, in the era of social networking, our identity as well as everyday activities has been naturally extended into virtual world. According to the statistical report (statisticbrain, 2014), a popular social networking site Twitter has around 5.5 billion of active registered users who produces 58 million tweets per day and 9,100 tweets per second. In addition, 135,000 new users are signing up to Twitter every day. Such statistics demonstrate that activities in online social networking sites are now a part of daily life of millions of humans. Our first hypothesis is that, similarly to a physical world, behavioral patterns and habits are present in the daily activities of virtual world users. Patterns can be found in different online social activities such as: tweets, status updates, ‘likes’, URLs, photo and video tagging, media sharing, uploads, comments, instant messages, communications, and so on. Our second hypothesis is that a person can be identified based on his activities and information which accumulates through online social networking platforms. Therefore, in this chapter we are introducing the fourth subcategory of machine dependent behavioral biometrics: web-based social biometrics. We also investigate the feasibility of using users’ web-based social networking activities as novel behavioral biometrics to identify a person in virtual domain. Some identified application domains of the proposed web-based social biometrics are shown in Figure 1. Some of them have been investigated from author authorship attribution (Gray et al., 1997) point of view but rarely been studied from biometric identification perspective.

In this chapter, we introduce web-based social data as auxiliary behavioral biometric features. The concept of how these features can be used for