Predictive Models in Cybercrime Investigation: An Application of Data Mining Techniques

A. S. N. Murthy, Indian Police Service, Karnataka, India
Vishnuprasad Nagadevara, Indian Institute of Management Bangalore, India
Rahul De’, Indian Institute of Management Bangalore, India

ABSTRACT

With increased access to computers across the world, cybercrime is becoming a major challenge to law enforcement agencies. Cybercrime investigation in India is in its infancy and there has been limited success in prosecuting the offenders; therefore, a need to understand and strengthen the existing investigation methods and systems for controlling cybercrimes is greatly needed. This study identifies important factors that will enable law enforcement agencies to reach the first step in effective prosecution, namely charge-sheeting of the cybercrime cases. Data on 300 cybercrime cases covering a number of demographic, technical and other variables related to cybercrime was analyzed using data mining techniques to identify and prioritize various factors leading to filing of the charge-sheet. These factors and the respective priority rankings are used to suggest various policy measures for improving the success rate of prosecution of cybercrimes.

Keywords: Charge-Sheeting, Classification Trees, Cybercrime, Data Mining, Neural Nets

INTRODUCTION

With the growing penetration of the Internet and Information and Communication Technologies (ICT) in countries around the world, and, in particular, in developing countries, it is expected that along with the benefits that accrue there will also be problems. These problems have to do with both the technology as well as the use of the technology. Criminal and mischievous elements around the world are intent on using or abusing the open facilities of the ICT technologies for their benefit or entertainment. The central problem of law enforcement and regulations is to establish the framework within which abuses with and through the use of ICT is checked.

A problem of law enforcement and regulations with regard to ICT spread is that the many features and facilities of the technologies are growing and evolving. Few still know all the possible ways in which the technologies will be used. It is clear that the Internet provides a lot of information that is beneficial, but it is not obvious as to what manner the information may be used. Law enforcement and regulatory...
agencies around the world are struggling to define what policies to adopt to both restrict abuse and to enhance use of ICT.

In a developing country like India, the penetration rates of the Internet and of ICT are low but growing. Table 1 shows some data on Internet and ICT penetration in India as compared to two developed countries—Sweden and USA. Though the numbers are low for India, for all the indicators, owing to the large population base, the problems of usage and enforcement are very high. For instance, the number of Internet subscribers in India is 10.36 million as on December 2007 and another 57.83 million have access to Internet through various other media (TRAI, 2008).

The concern for monitoring and controlling crimes committed using ICT is called Cybercrime. That this is of growing importance was reflected in a recent conference (in September 2007) organized by the Indian Central Bureau of Investigation (CBI) in association with the software industry. Participants from 37 countries attended the conference. The main thrust of the conference was to have effective international cooperation in dealing with cybercrimes which covers legal standardization, standardization of procedural obligations, capacity building, and partnership with other agencies. This conference underscored the fact that cybercrime has become a major issue for law enforcement agencies in recent years.

**OBJECTIVES AND SCOPE**

The broad goal of this study is to understand and strengthen existing investigation methods, systems and practices for controlling cybercrime. A specific objective of this study is to identify the enabling factors which will facilitate charge-sheeting of cybercrime cases in India. The scope of study is limited to analyzing cybercrimes reported to law enforcement agencies in five states of India, namely Karnataka, Tamil Nadu, Andhra Pradesh, Kerala and Delhi.

The rest of the paper begins with a section on the definition of cybercrime, which is followed by a review of cybercrime prevention efforts from around the world and also a review of academic literature on cybercrime. This is followed by a discussion of the methodology of data collection and analysis. In the next sections the results are presented along with an analysis of the findings.

**Background: What is Cybercrime?**

It is extremely difficult to determine when the first crime involving a computer actually occurred. In 1801 Joseph Jacquard, a textile manufacturer in France, designed the forerunner of the computer punched card. This device allowed repetition of a series of steps in the weaving of special fabrics. So concerned were Jacquard’s employees with the threat to their traditional employment and livelihood that they resorted to sabotage in order to discourage Mr. Jacquard from further use of this technology. This could be considered as the first “computer” crime to have been committed (United Nations, 1994).

**Cybercrime Definition**

Even after many years, a global definition of computer crime has not been arrived at. The

<table>
<thead>
<tr>
<th>Country</th>
<th>Per capita income (US $)*</th>
<th>Internet per 100 persons**</th>
<th>PCs per 100 persons**</th>
<th>Mobile users per 100 persons**</th>
<th>Landline users per 100 persons**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sweden</td>
<td>46,060</td>
<td>76.97</td>
<td>83.49</td>
<td>105.92</td>
<td>59.52</td>
</tr>
<tr>
<td>USA</td>
<td>46,040</td>
<td>69.1</td>
<td>76.22</td>
<td>77.40</td>
<td>57.15</td>
</tr>
<tr>
<td>India</td>
<td>950</td>
<td>5.44</td>
<td>1.54</td>
<td>14.83</td>
<td>3.64</td>
</tr>
</tbody>
</table>

Table 1. Internet and ICT penetration in India and two developed countries. Source: UN E-Government Survey, 2008
Exploration of Adoption of Service Innovations through Technology Road-Mapping: Case of Location Based Services
Tugrul Daim, Robert Harmon and Haluk Demirkan (2012). Technological Applications and Advancements in Service Science, Management, and Engineering (pp. 152-172).
www.igi-global.com/chapter/exploration-adoption-service-innovations-through/66291?camid=4v1a