International Journal of Advance Research in Science and Engineering
Vol. No.4, Special Issue (01), September 2015
www.ijarse.com
IJARSE
ISSN 2319 - 8354

A CASE STUDY OF WEB CONTENT MINING IN HANDLING CYBERCRIME

B.Tirupathi Kumar[1], K.Chandra Sekharaiah[2], P.Mounitha[3]

Asst.Professor ,CSE Dept., Malla Reddy Institute of Technology, Hyderabad (India)
Professor, School of IT, JNTU Hyderabad (India)
Asst.Professor ,CSE Dept.MGIT, Hyderabad (India)

ABSTRACT

In this paper, we delineate how a web crawler tool is used for web content mining in the case of a website which is a culprit website involved in the three crimes of i.e. Sedition, State Emblem of India (Prohibition of Improper Use) Act 2005 violation and Cheating crime. The organization JNTUHJAC was affiliated to the online cheating 'Government of Telangana' (CGoT) and online seditious 'Government of Telangana' (SGoT). It has been operational in the JNTUHyderabad university during 2011-14. We present a few snapshots of the culprit website's home page from wayback machine, a web crawler tool for Internet archiving of websites. The snapshots capture evidence of the usage of cheating & seditious identity 'Government of Telangana' (to which the said JNTUHJAC organization claimed belongingness) by mining the archived website contents. We present how they serve as cyber intelligence alternatives for countering cybercrime.

Keywords: Cheating 'Government of Telangana' (CGoT), Seditious 'Government of Telangana' (SGoT), Fake 'Government of Telangana' (FGoT)

I. CYBERFORENSICS

Cyber forensics is the field of study w.r.t. forensic evidence related to/based on web data and resources. TrueBack, Cyber Check, F-DAC, F-RAN, Mobile Check, AdVik-CDR Analyser, NeSA, and WinLiFT etc. are various Cyber Forensics tools indigenously developed by C-DAC, Thiruvananthapuram, India towards latest methods of investigation by police. A cyber crime is one that violates a Law(e.g. Cyberlaw)/Code(e.g. Cr.P.C.)/Act/Constitution using web as a media. Often, cyber crimes are difficult for handling and are elusive to proof of evidence. Sometimes, however, cognizable evidence of cybercrimes cannot belie the cyber forensic experts. In this paper, cheating crime and sedition crime are presented as the cybercrimes of interest towards handling unlawful organized activity in JNTUH. The rest of the paper is organized as follows. Section 2 gives a brief of Internet Archiving and Way back Machine, a web crawler tool. Section 3 gives in brief the details of the Sedition Law and Cheating crime etc. which are perpetrated web-based by the culprit organization. Section 4 captures the snapshots of culprit website's homepage which is result of web content mining obtained by the web crawler tool wayback machine.

II.WEB ARCHIVING

The Internet Archive is a non-profit digital library. It is based in San Francisco. Its function is archiving. Secondly, it is an activist organization that advocates for a free and fair Internet. The Internet Archive allows the public to upload and download digital material to its data cluster. However, the bulk of its data is collected automatically by its web crawlers. Its web crawler works to preserve as much of the public web as possible. Its web archive, The Wayback Machine, contains over 150 billion web captures.

2.1 Wayback Machine: A Web Crawler Tool

The Internet Archive Wayback Machine is a tool to capture snapshots of webpages/sites online. It is an automated system to generate an archive that contains crawls/archives of websites. As of 1Jul2015, it contains 23 petabytes of data. Currently, it is growing at a rate of 50-60 terabytes per week. This eclipses the amount of text contained in the world's largest libraries, including the Library of Congress. In other words, the Internet Archive Wayback Machine is a service that allows people to visit archived versions of Web sites. Visitors to the Wayback Machine can type in a URL, select a date range, and then begin surfing on an archived version of the Web.

III. SEDITION & CHEATING

3.1 Sedition

The Sedition Law [4] penalizes any spoken or written words or visible representations, etc, which have the effect of bringing, or which attempt to bring into hatred or contempt or excite or attempt to excite disaffection towards "the government established by law".

3.2 Cheating

In [5], cheating could be a case of "Whoever cheats and thereby dishonestly induces the person deceived any property to any person, or to make, alter or destroy the whole or any part of a valuable security, or anything which is signed or sealed, and which is capable of being converted into a valuable security, shall be punished with imprisonment of either description for a term which may extend to seven years, and shall also be liable to fine." Sections 419, 425, 431 deal with personation and allied issues of cheating crime.

IV. CAPTURING THE SNAPSHOTS OF THE CULPRITS WEBSITE'S HOMEPAGE

2 Snapshots of the culprits website are shown below for the period 12Nov2011 – 1Aug2015. They are the results of web crawling. They are dated 13Feb2012, 18Apr2012, as per the automated web crawls by wayback machine. However, altogether, the web crawler tool saved the website 10 times. The 2 web crawls captured an image with a logo on the R.H.S. which contains Indian National Emblem. This is a cyber crime already reported in our earlier work in [1]. Also, we find mention of 'Government of Telangana' in the logo on the R.H.S. This is not only CGoT but also SGoT. This FGoT is misleading and cheating the nation. The logo is part of the image during the years 2011-14. As per the chargesheet in [5], one of the culprits is absconding . Hence, it cannot be ruled out that the CGoT/SGoT/FGoT works in the background/underground/parallelly now. In the last one year or so, the culprit website removed the image with the logo.

International Journal of Advance Research in Science and Engineering
Vol. No.4, Special Issue (01), September 2015
www.ijarse.com
IJARSE
ISSN 2319 - 8354

Snapshot 1: dt.13Feb2012

Snapshot 2: dt.18Apr2012

V. CONCLUSIONS

The culprit GoT is involved in 3 crimes altogether clearly. In this paper, sedition crime and cheating crime are highlighted. The web content mining results drawn from the wayback machine, a web crawler tool clearly prove the crimes. Over the years 2011-14, these two crimes were ignored in the sense that no FIR is registered despite complaint. Using Wayback Machine web archiving tool, the instances of web crawls are retrieved. The cyberforensic data shows that these 2 crimes were perpetrated against the nation. The 2 snapshots clearly show clues to prove the crimes of culprit GoT. Such organized crimes are threats to the national spirit and national integration in India.

If online culprit GoT is spared, it will set wrong precedence and such crimes tend to recur. It becomes a vicious cycle. Therefore, such a Cheating/Fake/Seditious Govt. should be prohibited and rooted out. In [3], we attempted to remedy the situation by making available Guidance and Counseling web page to make transparent to the general public the issues involving cybercrimes against the nation.

REFERENCES

[1] Usha Gayatri P., Neeraja S., Leela Poornima Ch., Chandra Sekharaiah K. and Yuvaraj M., "Exploring Cyber Intelligence Alternatives for Countering Cyber Crime", Proceedings of the 8th INDIACom; INDIACom-2014, International Conference on "Computing for Sustainable Global Development", 5-7March2014, Bharatiya Vidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi (INDIA).

[2] https://archives.org/web

[3] https://sites.google.com/site/chandraksekharaiah/india-against-corruption-jntu

[4] http://www.rmlnlu.ac.in/webj/sedition.pdf

[5] https://sites.google.com/site/chandraksekharaiah/miscellaneous333