

Forensics Tools for Social Network Security

Vishal Almeida, Andrew Karnbad, Palak Shah, and Steve Kim
Seidenberg School of CSIS, Pace University, White Plains, NY 10606 USA

Abstract

Social networking has exploded in terms of popularity for a variety of reasons that include personal connections, instant messaging, photo sharing, selling of merchandise, sharing of ideas and philosophical viewpoints. Where there is money to be made or the opportunity for one to take advantage of another there will always be a criminal element. How does one know that the friend they've just accepted an invitation from is really who they say they are? How does one know that the posting to their Facebook wall came from their friend or from someone who hacked into their friends account? Often times they do know but not always, and it is those times we strive to empower users to know the difference.

1. Introduction

This paper extends an earlier one [8] and takes into account several other concerns in both IT technology and the use of social networking. Here we discuss the use of numerous open-source tools and methods that, when combined, can give a more accurate assessment of a user's identity. Not necessarily that they are who they say they are but more that they are not attempting to hide their identity through various means. These could include location deception, origination from regions of the world with questionable objectives, the use of source IP obfuscation tools such as the TOR client or anonymous proxies.

This paper describes end user tools designed to protect the users of a Social Networking Site from security threats with a particular focus on Facebook. It also gives an overview of the technology behind the TOR client, its intended purpose as well as its unintended purpose, how it works to hide a user's source address and a description of technology which could identify the true identity of a TOR user. We will look at an open source tool designed to scour pre-defined web sites for the presence of e-mail addresses giving a snapshot of how a particular address may have been used on a variety of online sources. Lastly, through a tool developed on ASP.NET we can cull various online databases and present an assessment of an IP address to determine whether the IP address is an original or a proxy.

2. Methodology

The methods used in designing the tool and database include ASP.NET with a local flat-file database. It is intended that the data will either be pulled into a relational database on demand or at some predetermined interval. The intended end-user would be a law-enforcement agency or an end-user wishing to learn more about the origins and attributes of an IP or domain name.

3. Case Studies

The following case studies from literature search relate to privacy issues in social network sites.

3.1 Facebook Blaster

“The profiles of two teenagers, one male and one female, with fake names were created. Using a suitable software tool ‘Facebook blaster’, friend requests were sent massively. As a result, two networks of friends were created and access was granted to a significant amount of user's personal information. Both profiles received requests for friendship and personal chat by adults aged up to 53 years old. To this end, both profiles become members of popular groups. Specifically, the female profile became a member of famous pop star's fans group and the male one joined a sport's fan group. The purpose for joining these groups was to gain access to lists of other people's profile IDs that were already members. Both profiles joined application ‘Zoosk’ provided by Facebook for meeting new friends.” [6] It can be inferred from above case, that there was no protection at all from Facebook as well as the application inside. The intention of both the profiles was to attract young users to accept the friend request to join them to ‘Zoosk’ application to attract more and more people to join Facebook.

3.2 Use of personal information

The merchants and third party applications can take advantage of user's information for advertising without a user's knowledge. Many companies like Coke, Pepsi, Apple Computer and Proctor & Gamble sometimes use

Social Network Sites as their promotional tools. "Indeed, business can use the Social Ads tool in Facebook to place their advertisements to a target any particular group based by criteria related to location, sex, age, etc." [3] All Social Network Sites have privacy policies and by accepting the terms of the policy, the user has a right to keep some information private. The social network sites collect and store other personal data like personal interests, gender, age, education, occupation, home address and IP address to improve their website services. As a result, even if the users apply most privacy settings, they still do not have total control over the use of their personal information.

4. Facebook Security

4.1 Privacy fundamentals of Facebook

What does Facebook do to protect the security of user's information?

As per Facebook's security web page, Facebook has the industry standard and proprietary network monitoring tools constantly running in user's system to prevent security breaches and protect the security of user's data. In addition, Facebook always posts to a secure page when users are logging in and employs industry standard encryption.

Can user know who's viewing his/her profile or how often it's being viewed?

From Facebook's discussion forum of security, Facebook does not provide the ability to track who is viewing user's profile, or parts of user's profile, such as user's photos. Applications by outside developers cannot provide this functionality, either. Applications that claim to give you this ability will be removed from Facebook for violating policy. However, user can report applications that provide untrustworthy experiences by clicking the "Report Application" at the bottom of the application's "About" page, or by clicking "Report" at the bottom of any canvas page within the application.

Can user see who's viewed his/her profile?

There's a group or application claiming user can find out who has been viewing user's profile. Facebook does not provide applications or groups to allow people to track profile views or see statistics on how often a particular piece of content has been viewed and by whom. Few applications the user uses may ask for permission to access content from user's News Feed and Wall. Granting this permission does not allow applications to see who has viewed your profile. It simply allows applications to see which friends have interacted with posts, such as which friends liked or engaged with a particular wall post. [12] However, when you are not sure about the identity of any person who sends you a

friend request via Facebook, you can certainly ask that person to send an email through his Yahoo, Gmail or Hotmail account to disclose his IP address and though that the location of that person.

4.2 Recent Privacy issues of Facebook

According to TechCrunch [16] and numerous other online technophile sites - a prominent Facebook fan page has been hacked, defaced and, as a result, closed down. "The victim? Mark Zuckerberg. The defacement? This message, apparently: Let the hacking begin: If Facebook needs money, instead of going to the banks, why doesn't Facebook let its users invest in Facebook in a social way? Why not transform Facebook into a 'social business' the way Nobel Prize winner Muhammad Yunus described it? <http://bit.ly/fs6rT3> What do you think? #hackercup2011" [16]

In above case, for hacking Mark Zuckerberg's account, one reason could be possible that this account has many different passwords for different users. Many passwords for one account, is sufficient to deface a page or to steal personally identifiable information for hackers. May be Mark Zuckerberg or one of his colleagues have had their passwords guessed or stolen, or perhaps had been 'side-jacked' by a tool such as FireSheep [4] while using unsecured public WIFI hotspot. Vulnerability in Facebook's code allows unauthorized parties to post updates to pages, which could be used for the purposes of phishing, spam and malicious attack.

"So, if you are the administrator of a popular page on Facebook, it wouldn't do any harm to check that all is in order. You may also want to ensure that your public forums are regularly monitored just in case a similar incident occurs in the future, which might result in your Facebook fans receiving unauthorized updates." [17]

IT security and control firm, "Sophos", has published its Security Threat Report 2011 (Figure 1), which provides analysis for cybercrime during the last year and looking at IT security trends to watch in 2011. Unsurprisingly, the numbers of attacks increasing steadily throughout 2010 as malware, phishing and spam on social networks have all continued to rise in the past year. Sophos's Poll is based on number of spam, phishing or malware attacks the users has experienced via various social network sites.

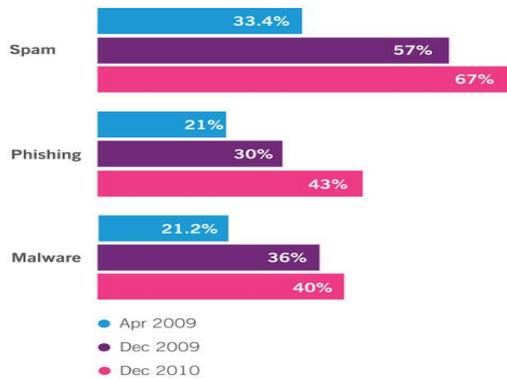


Figure 1: Security Threat Report 2011 [18]

"Rogue applications, click jacking, survey scams – all unheard of just a couple of years ago, are now popping up on a daily basis on social networks such as Facebook," said Graham Cluley, senior technology consultant at Sophos. [18]

4.3 Protecting yourself

Understanding the Social Networking Sites Structure and de-anonymization attack to exploit user's information

All social networking sites have same basic structure. Each user in the network has a profile which contains some sensitive information like photographs, relationship status, etc. One of the important technical components of a social network is its website, and the underlying web application. The web application provides the main functionalities of the social network. This functionality usually has features that allow a web visitor to become a member, to view other user profiles, and to join groups. To become a member of a social networking site, users can sign up at the website, which only requires a valid e-mail address for verification purposes. Since all social networks have millions of users (Table 1), most popular social networks include features that allow users to be organized in groups. Through this feature, users of a social network can easily find other users with the same hobbies, studying in same university, etc.

Name of social network	# users	Focus	Alexa traffic rank [1]	Supports groups
Facebook	400,000,000+	general audience, worldwide	2	✓
MySpace	260,000,000+	music, worldwide	11	✓
Friendster	90,000,000+	general audience, worldwide	111	✓
LinkedIn	50,000,000+	business, worldwide	53	✓
StudiVZ	15,000,000+	students, Germany	179	✓
Xing	8,000,000+	business, Europe	285	✓
Biggadda	5,500,000+	teenage audience, India	3,082	✓
Kiwibox	2,500,000+	teenage audience, worldwide	74,568	✓

Table 1: Popular social network sites [1]

In general, two different types of groups can exist: (1) Public groups, which allow all members of the social network to join when they wish to join without any permission. Some social networks even allow non-group members to list the members of public groups (e.g., Facebook) (2) Closed groups, which require some authorization before a member is allowed to join. This means that a group administrator or moderator needs to manually approve each membership request. The web applications for the most popular social networks depend on hyperlinks and HTTP GET parameters to implement the communication between a user and the actual web application. For example, (1) [http://www.facebook.com/ajax/profile/picture/upload.php?id=\[userID\]](http://www.facebook.com/ajax/profile/picture/upload.php?id=[userID]) and (2) [http://www.facebook.com/group.php?gid=\[groupID\]&v=info&ref=nf](http://www.facebook.com/group.php?gid=[groupID]&v=info&ref=nf); are some of the examples for different web applications that represent two groups of hyperlinks. The first link (1) is used to tell the web application of Facebook to display the currently logged-in user's "home" area. Since the hyperlink for this operation is the same for every user of the social network, the links of this type are *static hyperlinks*. The second link (2) sends a request to the web application that the user with the ID wants to upload a new profile picture. This link contains the userID and so it is a *dynamic hyperlink*. This type of links explicitly contains information about a user since the link is unique for each user of the social network. Through the web application, these hyperlinks provide the "internal" state keeping and communication between the web application and the user's web browser. Since web browsers are just the interpretation of links, they simply add the URLs of all visited web pages to the browsing history of a user. Since the important information is already encoded in the URL itself, whether the website is using security-enhancing protocols: HTTPS for protecting the actual content (however, as per Facebook blog, if browser is using a secure connection ("HTTPS") to communicate with the website it ensures that the information you send remains private.) it does not matter. "From an attacker's point of view, this behavior is interesting, since it enables the attacker to identify groups a user has visited, and even potentially identify a specific user. That is, if the attacker is able to determine which pages are in the victim's browsing history (i.e., he can compute the function for pages loaded via dynamic hyperlinks), he can use this information to de-anonymize a user." [5]

The de-anonymize attack can also be applied through History Stealing. In this attack, a browser implements the function, which implicitly checks whether a target URL is in the browsing history or not. An attacker can create an HTML page with links to find web pages of interest and use background image tags in the a: visited style information. Since images can be referenced with

URLs, the user's browser will then access these URLs if a target web page is in the browser.

Another way of de-anonymize attack is to use client-side scripting as JavaScript to generate and go over through a list of target links and programmatically check for the: visited style to see if a link is present in web browser. To enumerate Facebook's group members, the attacker can extract the group IDs from the group directory, and then try to enumerate the members for each group. Facebook permits each logged-in user to search within the member lists of arbitrary groups. This search can be used to filter the member lists to only show members whose first and/or last name fully matches the search token.

Ways to Facebook Security

When day by day, the numbers of scamming and hacking incidents are increasing through Facebook site, Facebook is increasing a number of ways to help its users to protect them and their accounts to make their experience on Facebook more secure.

* Facebook is launching one-time passwords to make it secure to use public computers at places like hotels, cafes or airports. If the user has any doubts about security of the computer he is using at some place, while accessing Facebook, Facebook can text user a one-time password to use instead of his regular password. The user needs to text "otp" to 32665 on his mobile phone (U.S. only), and he will immediately receive a password that can be used only once and expires in 20 minutes. In order to access this feature, the user needs a mobile phone number in his account.

* The ability to sign out of Facebook remotely is now available to everyone. These session controls can be useful if the user logs into Facebook from a friend's phone or computer and then forget to sign out. From his Account Settings, he can check whether he is still logged in on other devices and remotely log out.

"Under the Account Security section of your Account Settings page you'll see all of your active sessions, along with information about each session. In the unlikely event that someone accesses your account without your permission, you can also shut down the unauthorized login before resetting your password and taking other steps to secure your account and computer.

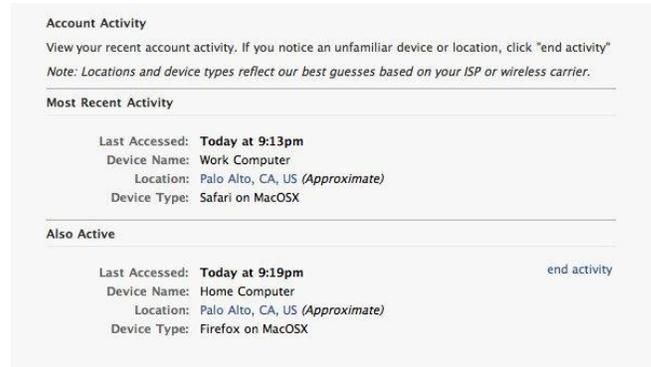


Figure 2: Facebook account activity [2]

When people log in to Facebook they are prompted to keep their security information updated. If you lose access to your account, this information helps us verify who you are and get you back into your account quickly. You can also update your security information at any time from https://www.facebook.com/update_security_info.php [2].

4.4 FireSheep

On October 24, 2010, Toorcon 12 announced the release of FireSheep, which can be incorporated in the Firefox browser. "When logging into a website you usually start by submitting your username and password. The server then checks to see if an account matching this information exists and if so, replies back to you with a "cookie" which is used by your browser for all subsequent requests. It is extremely common for websites to protect your password by encrypting the initial login, but surprisingly uncommon for websites to encrypt everything else. This leaves the cookie (and the user) vulnerable. HTTP session hijacking (sometimes called "side-jacking") is when an attacker gets a hold of a user's cookie, allowing them to do anything the user can do on a particular website. On an open wireless network, cookies are basically shouted through the air, making these attacks extremely easy." [4]

After installing FireSheep a user can connect to any open Wi-Fi network and click the "Start Capturing" button. As soon as anyone on the network visits an unsecure website through open Wi-Fi connection known to FireSheep, their name and photo will be displayed and with a simple double click a stranger can instantly be logged in as that user.

5 Public IP Addressing

All traffic traversing the Internet must have a public IP address. Private IP addressing, as defined by RFC 1918 [13] cannot be used as a source address on the public Internet and will be blocked at the entry point by all ISPs. This leaves us with a unique fingerprint which

can provide evidence of the last point of exit used to access the Internet. This is not necessarily the point of entry as discussed later in this paper but a starting point.

5.1 Geographic location

Equally important in the determination of trustworthiness is the country of origin of an IP address. IP addresses, much like telephone numbers, are handed out by one of five Regional Internet Registries (RIR) (Table 2) in a coordinated fashion to ISPs around the globe and to entities which can prove that they require a subset of public IP addresses.

ARIN	American Registry for Internet Numbers	North America, Canada, many Caribbean & Atlantic Islands
LACNIC	Latin American and Caribbean Internet Access	Latin America, South America, & some Caribbean Islands
RIPE NCC	Réseaux IP Européens Network Coordination Centre	Europe, Middle East, and parts of Central Asia
AfriNIC	African Network Information Center	Africa
APNIC	Asia Pacific Network Information Centre	Asia Pacific region

Table 2: Regional Internet Registries [10]

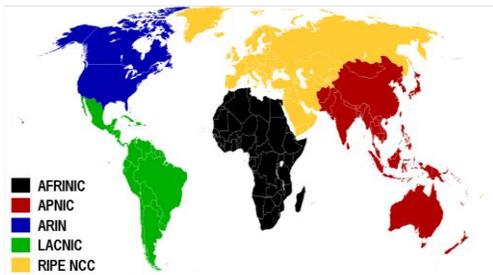


Figure3: Global depiction of RIR's [11]

5.2 Manual location determination

A benefit of this system of addressing is that we can look up an IP address of a random website in Beijing China www.bj17909.com by pinging it in a Windows command window (Figure 4). We see that the DNS resolution has returned a 32-bit public IP address 210.51.4.26. Any number of IP address lookup tools freely available will indicate that the IP address was assigned by APNIC and the country of origin is China (Figure 4). This is not to say that this site is nefarious

but it is originating from country at the top of many lists which track Internet fraud.

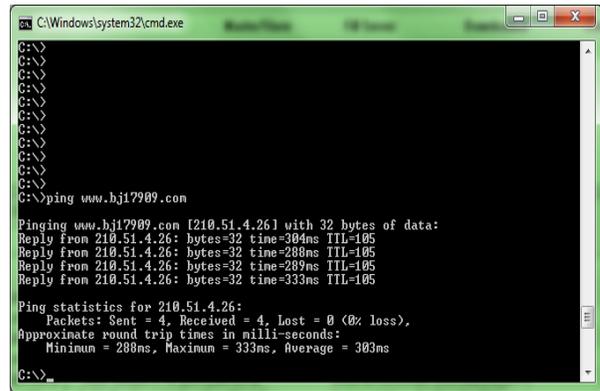


Figure 4: DNS resolution.

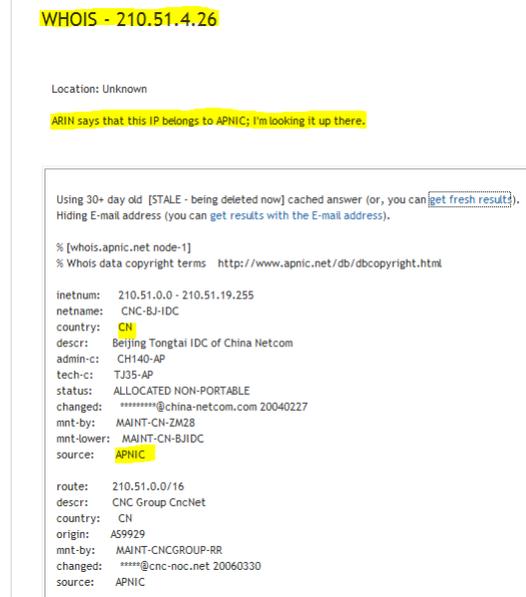


Figure 5: WHOIS lookup [9]

By cross-referencing IP addresses and domains to a list of geographically suspect sources the trust-ability factor can be enhanced. This type of profiling, while not conclusive, can produce a telling story when according to Symantec, “More malware is coming out of China than from any other country.” [15] At first glance the chart below would indicate that the United States was ahead of China but since most malware is sent using US hosted servers the source is skewed. The report took into account the source IP’s of the device connecting to stateside server.



Figure 6: Malware source map [7]

Top Ten Countries by Count (Perpetrators)

1. United States 66.1%
2. United Kingdom 10.5%
3. Nigeria 7.5%
4. Canada 3.1%
5. China 1.6%
6. South Africa 0.7%
7. Ghana 0.6%
8. Spain 0.6%
9. Italy 0.5%
10. Romania 0.5%

5.3 Method of determining location from IP address (program)

Once an IP address of the user who visits our website or profile or who sent us an email or message, there are tremendous numbers of websites, which provide the facility to get the latitude, longitude, city and country information of that IP. For example, ip2location.com is a geo IP solution, which provides free source code in PHP and ASP.NET to help you to identify visitor's geographical location, i.e. country, region and city and through that latitude, longitude, ZIP code, time zone, connection speed, ISP and domain name etc. using a proprietary IP address lookup database and technology without invading the Internet user's privacy. The example of ASP.NET code in Figure 6.

```
using System.Collections.Generic;
using System.Linq;
using System.Web;
using System.Web.UI;
using System.Web.UI.WebControls;
using System.Data.SqlClient;
namespace ip2Proxy
{
    public partial class _Default : System.Web.UI.Page
    {
        protected void Page_Load(object sender, EventArgs e)
        {
            string ipaddress;
            long ipno;
            string[] arripaddress;
            int i;

            if (Request.Form.Count > 0)
            {
                // get the IP address from the form
                ipaddress = Request.Form["ipaddress"];
                ipaddress = ipaddress.Replace("\r\n", "");
                ipaddress = ipaddress.Replace(" ", "");
            }
        }
    }
}
```

```
arripaddress = ipaddress.Split(',');
if (arripaddress.Length != 0)
{
    Response.Write("<p>");
    Response.Write("<h1><u>IP2Proxy&#153; Lookup
Results</u></h1>");
    Response.Write("<table border = 1>");
    // display header
    Response.Write("<tr>");
    Response.Write("<td align=center>IP Address</td>");
    Response.Write("<td align=center>Country Code</td>");
    Response.Write("<td align=center>Country Name</td>");
    Response.Write("</tr>");
    for (i = 0; i <= arripaddress.Length - 1; i++)
    {
        if (arripaddress[i] != "")
        {
            SqlDataReader reader;
            // select MS-SQL database using DSNless connection
            SqlConnection sqlConn = new SqlConnection(@"Server=KUSHAL-
PC\SQLEXPRESS;Database=IP2Proxy;Trusted_Connection=True;");
            // query string to lookup the country by matching Anonymous IP address
            SqlCommand sqlCmd = new SqlCommand("SELECT
COUNTRY_CODE,COUNTRY_NAME FROM IP2PROXY WHERE
IP_ADDRESS=" + arripaddress[i] + "", sqlConn);
            sqlCmd.Connection.Open();
            // execute the query
            reader =
            sqlCmd.ExecuteReader(System.Data.CommandBehavior.CloseConnection);
            // display results
            if (reader.Read())
            {
                Response.Write("<tr>");
                Response.Write("<td align=center>" + arripaddress[i] + "</td>");
                Response.Write("<td align=center>" + reader.GetString(0) + "</td>");
                Response.Write("<td align=center>" + reader.GetString(1) + "</td>");
                Response.Write("</tr>");
            }
            else
            {
                Response.Write("<tr>");
                Response.Write("<td align=center>" + arripaddress[i] + "</td>");
                Response.Write("<td align=center></td>");
                Response.Write("<td align=center></td>");
                Response.Write("</tr>");
            }
            sqlCmd.Connection.Close();
        }
    }
    Response.Write("</table>\n");
    Response.Write("<p>\n");
}
else
{
    Response.Write("Please enter ip address.");
}
}
```

Figure 7: ASP.NET code

Once you incorporate above ASP.NET page with necessary HTML code, you can run the webpage into web directory, configure the web directory with IIS and set the SQL database on your local machine; you can browse ip2Proxy.aspx using http protocol [14]. We are extending this functionality by adding sample data about known proxy IP, anonymous IP and TOR node IP into the database and by calling this database, the user will know about the location of the IP address. The IP address which has no accompanying country name and country code, is the anonymous IP address. In real world, using above functionality one can build a website for the location look up of an IP address by incorporating various database sources as back end of the result. The sample result is in Figure 7.



Figure 8. Output for location of an IP address

5.4 SOCKS Proxy

It is very important to know about the particular IP address, whether it is a known proxy or completely anonymous. Now a days there are many arrangements to make the original IP address as proxy for several purposes. For example, SOCKS is a networking proxy protocol which provides the facility to enable hosts on one side of a SOCKS server to gain full access to hosts, on the other side of the SOCKS server without the need for direct IP-reachability. The SOCKS server authenticates and authorizes requests, establishes a proxy connection, and exchanges data between hosts.

6 The Onion Router

The Onion Router or TOR [19] is an open-source routing system designed to allow users to surf the Internet with anonymity. It does so via a complex network of OR's (onion routers) which consists of a minimum of three nodes. These nodes include the entry node, the middleman, and the exit node. Each of these OR'S removes the header of the TOR cell and sends the packet off to the next OR. The last node, the exit OR, removes the last header, decrypts the packet and places the payload in "normal" TCP packet and sends it out the last router. The beauty of these technique is that no one is any the wiser since the packet seems to be coming from this last innocuous exit point. Even the OR's along the way can't identify the complete packet path since the header is changed at each trip along the way. This method of obfuscating ones true source of entry onto the Internet was intended by the TOR creators to allow users to browse, post, etch without fear of persecution by any number of entities. A person in a censored country, someone reporting a wrong-doing, etc. Companies who wish to perform research on another company or product without alerting the world that they are doing so. Not an illegal activity in any

sense, just a means of keeping one's private business, private. Naturally this method of anonymity can and is used for illegal purposes as it can easily hide (many times over) a user's true source IP address and identity. TOR could be used to participate in fraudulent e-commerce activities, espionage, blackmail, sexually-related crimes, and so on. The list of possibilities for improper use is likely longer than for legitimate use.

6.1 TOR Detection

An anonymous paper entitled *Approximating a Global Passive Adversary against TOR* was submitted to SePTIS in 2008 [19] which purport to have a method with which the true identity of a TOR entry point can be determined. Through the use of a LinkWidth, a bandwidth estimation technique, and a TOR network edge router under their control they were able to manipulate bandwidth and observe these fluctuations through the Internet back to the original host in many cases. While the specific mechanisms in place and technology in use exceeds the bounds of this Capstone project, the findings of the submission warrant further investigation by others who seek to harness this ability. This technology or ability would be particularly those in law enforcement who are better enabled in terms of resources and positioning themselves appropriately in the flow of traffic.

6.2 TOR Status

When a workstation is participating on the TOR network it becomes a node on the network as described in the TOR project section. As such, it becomes crucial to understand if traffic is flowing from one of these known nodes. Granted, you would only know if an IP address had exited out of a known, current, node and not the true source of the packet however this piece of information is only an indication of the trustworthiness of the identity of the sender. At the time of this writing, there were 2395 routers or nodes participating on the TOR network [19]. TOR status sites such as <http://torstatus.blutmagie.de> query the TOR network and produce near real-time results on current node information. As shown in figure 8. Statistics include the router name which is what the end-user chose as their nodes name when installing the TOR client. Also an indication of the amount of bandwidth available, the uptime of the node on the TOR network, and the hostname resolved from a public DNS server. This information is available via GUI as shown in Figure 8 or may be downloaded into a CSV or comma separated value format. Our prototype can pull this information from the TOR status server itself or it can become a node on the TOR network which would produce similar, if not better, results.

