International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

COPY RIGHT

IJIEMR Transactions, online available on 27th Aug 2015. Link :http://www.ijiemr.org/downloads.php?vol=Volume-4&issue=ISSUE-04

Title: **PROTECTING THE USERS SENSITIVE INFORMATION FROM MELICIOUS ATTACKS BY ANALYZING AND IDENTIFYING THE SUSPECIOUS APPLICATIONS IN SOCIAL NETWORK BY USING IPSFAPP**

Volume 04, Issue 04, Pages: 530–538.
Paper Authors

B.GEETHA KUMARI, JAGETI PADMAVTHI
GNITS

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

PROTECTING THE USERS SENSITIVE INFORMATION FROM MELICIOUS ATTACKS BY ANALYZING AND IDENTIFYING THE SUSPECIOUS APPLICATIONS IN SOCIAL NETWORK BY USING IPSFAPP

[1]B.GEETHA KUMARI, [2]JAGETI PADMAVTHI
[1,2]Assistant professor,Dept of CSE, GNITS

ABSTRACT:

With 20 million installations per day, third-party applications are one of the main reasons for Facebook's popularity and dependence. Unfortunately, hackers have understood the potential of using applications to spread malware and spam. The problem is already important, as we have found that at least 13% of the applications in our data set are harmful. So far, the research community has focused on detecting harmful publications and campaigns. In this document, let's ask the question: given a Facebook application, can we determine if it is harmful? Our key contribution is the development of (Intrusion Protection System to Analyze Facebook Apps) IPSFApp-Facebook's strict application apps, probably the first tool focused on detecting malicious applications on Facebook. To develop IPSFApp, we used the information collected by observing the behavior of Facebook 111K applications viewed in 2.2 million users on Facebook. First of all, we identify a number of features that help us distinguish harmful applications from benign applications. For example, we have found that malicious applications often share names with other applications and generally require less permissions than benign applications. Second, by exploiting these distinctive features, we demonstrate that IPSFApp is able to detect malicious applications with 99.5% accuracy, without false positives and a high rate of true positive (95.9%). Finally, we explore the malicious Facebook application ecosystem and identify the mechanisms that these applications use to spread. Interestingly, we find that many applications conspire and support each other; In our data set, we found 1584 applications that allow viral propagation of 3723 other applications through its publications. In the long run, we believe that IPSFApp is a step towards creating an independent app review and classification system in order to alert Facebook users before installing applications.

INTRODUCTION:

What Is A Social Network?

Wikipedia defines a social network service as a service which “focuses on the building and verifying of online social networks for communities of people who share interests and activities, or who are interested in exploring the interests and activities of others, and which necessitates the use of software.” A report published by OCLC

provides the following definition of social networking sites: "Web sites primarily designed to facilitate interaction between users who share interests, attitudes and activities, such as Facebook, Mixi and MySpace."

What Can Social Networks Be Used For?

Social networks can provide a range of benefits to members of an organisation:

Support for learning: Social networks can enhance informal learning and support social connections within groups of learners and with those involved in the support of learning.

Support for members of an organisation: Social networks can potentially be used my all members of an organisation, and not just those involved in working with students. Social networks can help the development of communities of practice.

Engaging with others: Passive use of social networks can provide valuable business intelligence and feedback on institutional services (although this may give rise to ethical concerns).

Ease of access to information and applications: The ease of use of many social networking services can provide benefits to users by simplifying access to other tools and applications. The Facebook Platform provides an example of how a social networking service can be used as an environment for other tools.

Common interface: A possible benefit of social networks may be the common interface which spans work / social boundaries. Since such services are often used in a personal capacity the interface and the way the service works may be familiar, thus minimising training and support needed to exploit the services in a professional context. This can, however, also be a barrier to those who wish to have strict boundaries between work and social activities.

Examples of Social Networking Services:

Examples of popular social networking services include:

Facebook: Facebook is a social networking Web site that allows people to communicate with their friends and exchange information. In May 2007 Facebook launched the Facebook Platform which provides a framework for developers to create applications that interact with core Facebook features

MySpace: MySpace is a social networking Web site offering an interactive, user-submitted network of friends, personal profiles, blogs and groups, commonly used for sharing photos, music and videos..

Ning: An online platform for creating social Web sites and social networks aimed at users who want to create networks around specific interests or have limited technical skills.

Twitter: Twitter is an example of a micro-blogging service. Twitter can be used in a variety of ways including sharing brief information with users and providing support for one's peers.

Note that this brief list of popular social networking services omits popular social sharing services such as Flickr and YouTube.

Opportunities and Challenges:

The popularity and ease of use of social networking services have excited

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

institutions with their potential in a variety of areas. However effective use of social networking services poses a number of challenges for institutions including long-term sustainability of the services; user concerns over use of social tools in a work or study context; a variety of technical issues and legal issues such as copyright, privacy, accessibility; etc. Institutions would be advised to consider carefully the implications before promoting significant use of such services.

What is networking?

Networking is the word basically relating to computers and their connectivity. It is very often used in the world of computers and their use in different connections. The term networking implies the link between two or more computers and their devices, with the vital purpose of sharing the data stored in the computers, with each other. The networks between the computing devices are very common these days due to the launch of various hardware and computer software which aid in making the activity much more convenient to build and use.

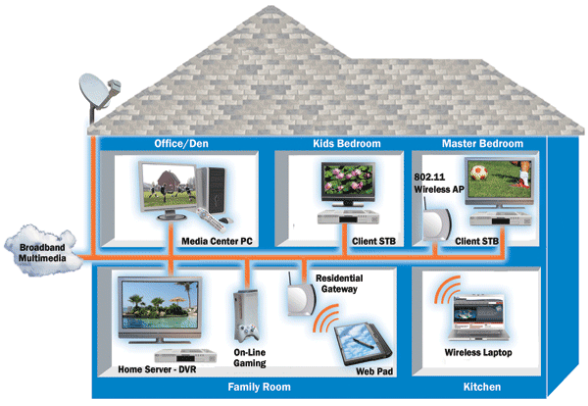

Fig: Structure of Networking between the different computers

How networking works?

General Network Techniques - When computers communicate on a network, they send out data packets without knowing if anyone is listening. Computers in a network all have a connection to the network and that is called to be connected to a network bus. What one computer sends out will reach all the other computers on the local network

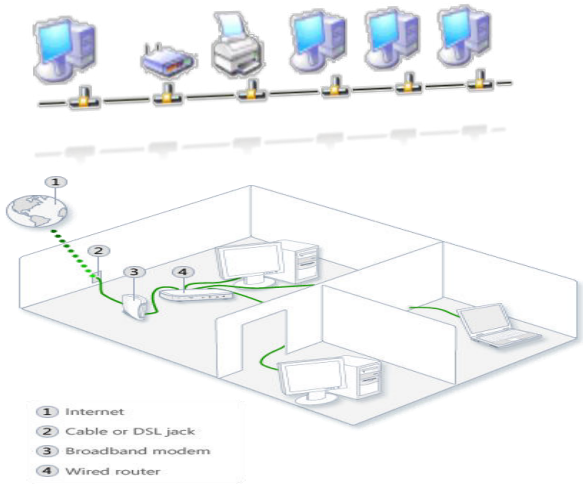

Fig: Above diagrams show the clear idea about the networking functions

For the different computers to be able to distinguish between each other, every computer has a unique ID called MAC-address (Media Access Control Address). This address is not only unique on your network but unique for all devices that can be hooked up to a network. The MAC-address is tied to the hardware and has nothing to do with IP-addresses. Since all computers on the network receives everything that is sent out from all other computers the MAC-addresses is primarily used by the computers to filter out incoming network traffic that is addressed to the individual computer.When a computer communicates with another computer on the network, it sends out both the other

computers MAC-address and the MAC-address of its own. In that way the receiving computer will not only recognize that this packet is for me but also, who sent this data packet so a return response can be sent to the sender.

On an Ethernet network as described here, all computers hear all network traffic since they are connected to the same bus. This network structure is called multi-drop.One problem with this network structure is that when you have, let say ten (10) computers on a network and they communicate frequently and due to that they sends out there data packets randomly, collisions occur when two or more computers sends data at the same time. When that happens data gets corrupted and has to be resent. On a network that is heavy loaded even the resent packets collide with other packets and have to be resent again. In reality this soon becomes a bandwidth problem. If several computers communicate with each other at high speed they may not be able to utilize more than 25% of the total network bandwidth since the rest of the bandwidth is used for resending previously corrupted packets. The way to minimize this problem is to use network switches.

Characteristics of Networking:

The following characteristics should be considered in network design and ongoing maintenance:

1) **Availability** is typically measured in a percentage based on the number of minutes that exist in a year. Therefore, uptime would be the number of minutes the network is available divided by the number of minutes in a year.
2) **Cost** includes the cost of the network components, their installation, and their ongoing maintenance.
3) **Reliability** defines the reliability of the network components and the connectivity between them. Mean time between failures (MTBF) is commonly used to measure reliability.
4) **Security** includes the protection of the network components and the data they contain and/or the data transmitted between them.
5) **Speed** includes how fast data is transmitted between network end points (the data rate).
6) **Scalability** defines how well the network can adapt to new growth, including new users, applications, and network components.
7) **Topology** describes the physical cabling layout and the logical way data moves between components.

Types of Networks:

Organizations of different structures, sizes, and budgets need different types of networks. Networks can be divided into one of two categories:

- peer-to-peer
- server-based networks

1. Peer-to-Peer Network:

A peer-to-peer network has no dedicated servers; instead, a number of workstations are connected together for the purpose of sharing information or devices. Peer-to-peer networks are designed to satisfy the networking needs of home

networks or of small companies that do not want to spend a lot of money on a dedicated server but still want to have the capability to share information or devices like in school, college, cyber cafe

2. Server-Based Networks:

In server-based network data files that will be used by all of the users are stored on the one server. With a server-based network, the network server stores a list of users who may use network resources and usually holds the resources as well. This will help by giving you a central point to set up permissions on the data files, and it will give you a central point from which to back up all of the data in case data loss should occur.

Network Communications:

- Computer networks use signals to transmit data, and protocols are the languages computers use to communicate.
- Protocols provide a variety of communications services to the computers on the network.
- Local area networks connect computers using a shared, half-duplex, baseband medium, and wide area networks link distant networks.
- Enterprise networks often consist of clients and servers on horizontal segments connected by a common backbone, while peer-to-peer networks consist of a small number of computers on a single LAN.

Advantages of Networking:

1. Easy Communication:

It is very easy to communicate through a network. People can communicate efficiently using a network with a group of people. They can enjoy the benefit of emails, instant messaging, telephony, video conferencing, chat rooms, etc.

2. **Ability to Share Files, Data and Information:**This is one of the major advantages of networking computers. People can find and share information and data because of networking. This is beneficial for large organizations to maintain their data in an organized manner and facilitate access for desired people.
3. **Sharing Hardware:**Another important advantage of networking is the ability to share hardware. For an example, a printer can be shared among the users in a network so that there's no need to have individual printers for each and every computer in the company. This will significantly reduce the cost of purchasing hardware.
4. **Sharing Software:**Users can share software within the network easily. Networkable versions of software are available at considerable savings compared to individually licensed version of the same software. Therefore large companies can reduce the cost of buying software by networking their computers.
5. **Security:**Sensitive files and programs on a network can be password protected. Then those files can only be accessed by the authorized users. This is another important advantage of networking when there are concerns about security issues. Also each and every user has their own

set of privileges to prevent those accessing restricted files and programs.

6. **Speed:**Sharing and transferring files within networks is very rapid, depending on the type of network. This will save time while maintaining the integrity of files.

EXISTING SYSTEM: So far, the research community has paid little attention to OSN apps specifically. Most research related to spam and malware on Facebook has focused on detecting malicious posts and social spam campaigns.

- Gao *et al.* analyzed posts on the walls of 3.5 million Facebook users and showed that 10% of links posted on Facebook walls are spam. They also presented techniques to identify compromised accounts and spam campaigns.
- Yang *et al.* and Benevenuto *et al.* developed techniques to identify accounts of spammers on Twitter. Others have proposed a honey-pot-based approach to detect spam accounts on OSNs.
- Yardi *et al.* analyzed behavioral patterns among spam accounts in Twitter.
- Chia *et al.*investigate risk signaling on the privacy intrusiveness of Facebook apps and conclude that current forms of community ratings are not reliable indicators of the privacy risks associated with an app.

There are disadvantages in existing system they are

- Existing system works concentrated only on classifying individual URLs or posts as spam, but not focused on identifying malicious applications that are the main source of spam on Facebook.
- Existing system works focused on accounts created by spammers instead of malicious application.
- Existing system provided only a high-level overview about threats to the Facebook graph and do not provide any analysis of the system.

PROPOSED SYSTEM: In this paper, we develop IPSFApp, a suite of efficient classification techniques for identifying whether an app is malicious or not. To build IPSFApp, we use data from MyPage-Keeper, a security app in Facebook.We find that malicious applications significantly differ from benign applications with respect to two classes of features: On-Demand Features and Aggregation-Based Features.We present two variants of our malicious app classifier— IPSFApp Lite and IPSFApp.IPSFApp Lite is a lightweight version that makes use of only the application features available on demand. Given a specific app ID, IPSFApp Lite crawls the on-demand features for that application and evaluates the application based on these features in real time.IPSFApp—a malicious app detector that utilizes our aggregation-based features in addition to the on-demand features. Advantages of our system are:

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

- The proposed work is arguably the first comprehensive study focusing on malicious Facebook apps that focuses on quantifying, profiling, and understanding malicious apps and synthesizes this information into an effective detection approach.
- Several features used by IPSFApp, such as the reputation of redirect URIs, the number of required permissions, and the use of different client IDs in app installation URLs, are robust to the evolution of hackers.

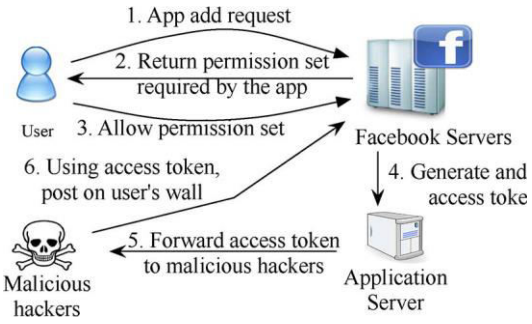

Fig: System Architecture

IMPLEMENTATION: Every implementation is having its own uses. We discussed about the implementation of opinion mining in this paper. They are:

Data collection: The data collection component has two subcomponents: the collection of facebook apps with URLs and crawling for URL redirections. Whenever this component obtains a facebook app with a URL, it executes a crawling thread that follows all redirections of the URL and looks up the corresponding IP addresses. The crawling thread appends these retrieved URL and IP chains to the tweet information and pushes it into a queue. As we have seen, our crawler cannot reach malicious landing URLs when they use conditional redirections to evade crawlers. However, because our detection system does not rely on the features of landing URLs, it works independently of such crawler evasions.

Feature extraction: The feature extraction component has three subcomponents: grouping of identical domains, finding entry point URLs, and extracting feature vectors. To classify a post, MyPageKeeper evaluates every embedded URL in the post. Our key novelty lies in considering only the social context (e.g., the text message in the post, and the number of Likes on it) for the classification of the URL and the related post. Furthermore, we use the fact that we are observing more than one user, which can help us detect an epidemic spread. It detects Presence of Spam keywords like ‘FREE’, ‘DEAL’ and ‘HURRY’.

Training: The training component has two subcomponents: retrieval of account statuses and training of the classifier. Because we use an offline supervised learning algorithm, the feature vectors for training are relatively older than feature vectors for classification. To label the training vectors, we use the account status; URLs from suspended accounts are considered malicious whereas URLs from active accounts are considered benign. We periodically update our classifier using labeled training vectors.

Classification: The classification component executes our classifier using input feature vectors to classify suspicious URLs. When the classifier returns a number

of malicious feature vectors, this component flags the corresponding URLs information as suspicious. The classification module uses a Machine Learning classifier based on Support Vector Machines, but also utilizes several local and external white lists and blacklists that help speed up the process and increase the over-all accuracy. The classification module receives a URL and the related social context features extracted in the previous step.These URLs, detected as suspicious, will be delivered to security experts or more sophisticated dynamic analysis environments for an in-depth investigation.

Detecting Suspicious: The Detecting Suspicious and notification module notifies all users who have social malware posts in their wall or news feed. The user can currently specify the notification mechanism, which can be a combination of emailing the user or posting a comment on the suspect posts.

CONCLUSION: Applications present convenient means for hackers to spread malicious content on Facebook. However, little is understood about the characteristics of malicious apps and how they operate. In this paper, using a large corpus of malicious Facebook apps observed over a 9-month period, we showed thatmalicious apps differ significantly from benign apps with respect to several features. For example, malicious apps aremuchmore likely to share names with other apps, and they typically request fewer permissions than benign apps. Leveraging our observations, we developed IPSFApp, an accurate classifier for detecting malicious Facebook applications.Most interestingly, we highlighted the emergence of app-nets—large groups of tightly connected applications that promote each other. We will continue to dig deeper into this ecosystem of malicious apps on Facebook, and we hope that Facebook will benefit from our recommendations for reducing the menace of hackers on their platform.

REFERENCES:

1. C. Pring, “100 social media statistics for 2012,” 2012 [Online]. Available: http://thesocialskinny.com/100-social-media-statistics-for-2012/
2. Facebook, Palo Alto, CA, USA, “Facebook Opengraph API,” [Online]. Available: http://developers.facebook.com/docs/reference/api/
3. “Wiki: Facebook platform,” 2014 [Online]. Available: http://en.wikipedia.org/wiki/Facebook_Platform
4. “Pr0file stalker: Rogue Facebook application,” 2012 [Online]. Available: https://apps.facebook.com/mypagekeeper/?status=scam_report_fb_survey_scam_pr0file_viewer_2012_4_4
5. “Whiich cartoon character are you—Facebook survey scam,” 2012 [Online]. Available: https://apps.facebook.com/mypagekeeper/?status=scam_report_fb_survey_scam_whiich_cartoon_character_are_you_2012_03_30
6. G. Cluley, “The Pink Facebook rogue application and survey scam,” 2012 [Online]. Available:

http://nakedsecurity.sophos.com/2012/02/27/pink-facebook-survey-scam/
7. D. Goldman, "Facebook tops 900 million users," 2012 [Online]. Available: http://money.cnn.com/2012/04/23/technology/facebookq1/index.htm
8. R. Naraine, "Hackers selling $25 toolkit to create malicious Facebook apps,"

2011 [Online]. Available: http://zd.net/g28HxI
9. HackTrix, "Stay away from malicious Facebook apps," 2013 [Online]. Available: http://bit.ly/b6gWn5
10. M. S. Rahman, T.-K. Huang, H. V. Madhyastha, and M. Faloutsos, "Efficient and scalable socware detection in online social networks," in *Proc. USENIX Security*, 2012, p. 32.
11. H. Gao *et al.*, "Detecting and characterizing social spam campaigns," in *Proc. IMC*, 2010, pp. 35–47.
12. H. Gao, Y. Chen, K. Lee, D. Palsetia, and A. Choudhary, "Towards online spam filtering in social networks," in *Proc. NDSS*, 2012.
13. P. Chia, Y. Yamamoto, and N. Asokan, "Is this app safe? A large scale study on application permissions and risk signals," in *Proc. WWW*, 2012, pp. 311–320.

"WhatApp? (beta)—A Stanford Center for Internet and Society Website with support from the Rose Foundation," [Online]. Available: https://whatapp.org/facebook/
14. "MyPageKeeper," [Online]. Available: https://www.facebook.com/apps/application.php?id=167087893342260
15. Facebook, Palo Alto, CA, USA, "Facebook platform policies," [Online]. Available: https://developers.facebook.com/policy
16. Facebook, Palo Alto, CA, USA, "Application authentication flow using OAuth 2.0," [Online]. Available: http://developers.facebook. com/docs/authentication/
17. "11 million bulk email addresses for sale—Sale price $90," [Online]. Available: http://www.allhomebased.com/BulkEmailAddresses.htm
18. E. Protalinski, "Facebook kills app directory, wants users to search for apps," 2011 [Online]. Available: http://zd.net/MkBY9k
19. SocialBakers, "SocialBakers: The recipe for socialmarketing success," [Online]. Available: http://www.socialbakers.com/
20. "Selenium—Web browser automation," [Online]. Available:http://seleniumhq.org/
21. "bit.ly API," 2012 [Online]. Available:

http://code.google.com/p/bitlyapi/wiki/ApiDocumentation

22. Facebook, Palo Alto, CA, USA, "Permissions reference," [Online]. Available: https://developers.facebook.com/docs/authentication/permissions/
23. Facebook, Palo Alto, CA, USA, "Facebook developers," [Online]. Available: https://developers.facebook.com/docs/appsonfacebook/tutorial/
24. "Web-of-Trust," [Online]. Available: http://www.mywot.com/
25. F. J. Damerau, "A technique for computer detection and correction of spelling errors," *Commun. ACM*, vol. 7, no. 3, pp. 171–176,Mar. 1964.
26. C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *Trans. Intell. Syst. Technol.*, vol. 2, no. 3, 2011, Art. no. 27.
27. J. Ma, L. K. Saul, S. Savage, and G. M. Voelker, "Beyond blacklists: Learning to detect malicious Web sites from suspicious URLs," in *Proc. KDD*, 2009, pp. 1245–1254.
28. A. Le, A.Markopoulou, and M. Faloutsos, "PhishDef: URL names say it

    all," in *Proc. IEEE INFOCOM*, 2011, pp. 191–195.
29. C. Wueest, "Fast-flux Facebook application scams," 2014 [Online]. Available: http://www.symantec.com/connect/blogs/fast-fluxfacebook-application-scams
30. "Longest path problem," 2014 [Online]. Available: http://en.wikipedia.org/wiki/Longest_path_problem

Author 1

Name:**B.geetha kumari**, M.Tech (cse)passout in 2009. Assistant professor in GNITS
Mail.id:Geetha.bapr07@gmail.com
3 yrs experience in mallaredddy engineering college for women.2.5 yrs experience in GNITS

Author 2:

AGETI PADMAVTHI
M.tech cse Assistant professor in GNITS
jageti.padmavathi4@gmail.com
Gnits 7 years and Bsit 2 years= total 9 years