

Automatic Fake Instant Message Detection

Rerngvit Yanggratoke^{1,2}, Abdullah Azfar^{1,3}

¹Royal Institute of Technology, Sweden. ²Helisinki University of Technology, Finland.
³Norwegian University of Science and Technology, Norway. {rerngvit, azfar} @kth.se



ABSTRACT

Fake Instant Messages (IM) have become a vital problem in messenger services. They are widespread and used by hostile attackers to distribute malicious messages such as Spam or Phishing messages. Detecting and notifying the IM users About the messages will mitigate the problem.

In this poster, we have proposed 4 new techniques of detecting the fake messages. The techniques include conversation language, time, phishing URL pattern, and Google Safe Browsing API. The ideas represented here are simple and implementable in any of the messenger services.

1. INTRODUCTION

Instant Messaging (IM) services are one of the most used web applications for last few years. People all over the world send and receive information through the IM services. A recent study shows that only mobile instant messaging in Europe had a rise to 26.7 million users in 2007 and projected to be 80 million users in 2013 [1]. Each country has a favorite IM service which has the highest penetration. According to statistics AIM has 36% market share in USA; GTalk has 35% market share in India and 26% market share in Japan and South Africa; Yahoo! Messenger has 50-70% market share in India, Indonesia and Saudi Arabia and ICQ has 44% market share in Germany and 56% market share in Russia [2].

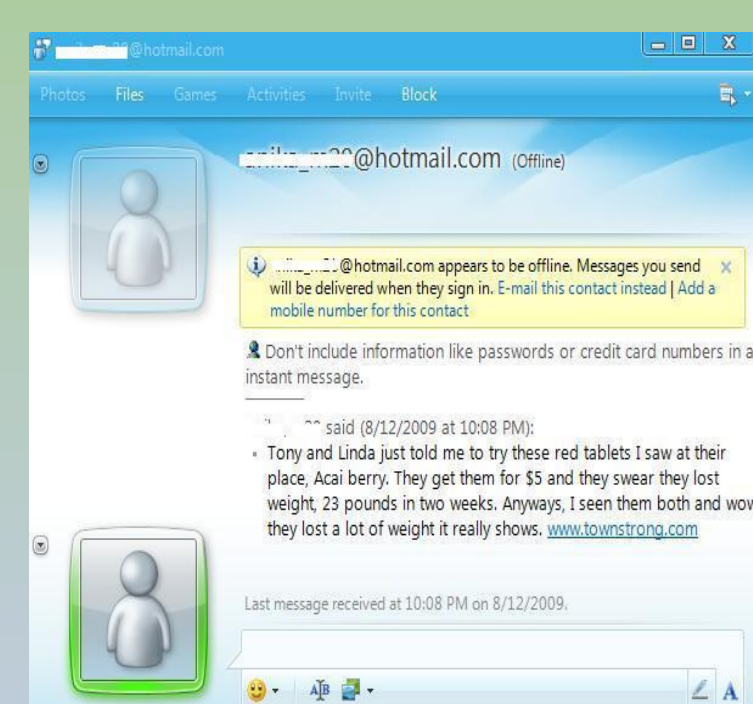


Figure 1: A fake message containing malicious URL

These statistics show how important the instant messenger services have become in our daily life. But, one problem that makes the instant messenger users' life horrible is the widespread of fake messages. As an illustration, IM users often receive a message not created by either party of the conversation. The message is completely a phishing, distributed by the hostile attacker. This message usually contains a link together with persuasive sentences to follow. Some users do not realize about the threat and follow the link accordingly. The result is Virus infection in the user machine or the machine becomes a part of the Botnet.

The motivation behind writing this poster is to find techniques for detecting these fake messages and notify the users about it. The detection of the fake Messages will help

millions of people to protect their computers or laptops from unwanted virus or phishing attack. The detection is done based on the conversation language, time, malicious URL Pattern and Google Safe browsing API.

The users of the instant messenger services will be directly benefited. It is a human nature that if they are affected with some virus by using some application then they try to avoid using that application again. Whenever a user receives a fake message and gets infected by a virus then that user often stops using the service. Instead, the user uses some other application. If proper detection techniques are applied then this scenario can be avoided.

The poster is divided into 3 sections. Section 2 will explain the design of the system by elaborating how each of the technique works. Section 3 will give a discussion of the techniques. And, Section 4 will conclude this poster.

2. DESIGN

The fake message detection can be done by using some filtering mechanisms at the receivers end as illustrated in

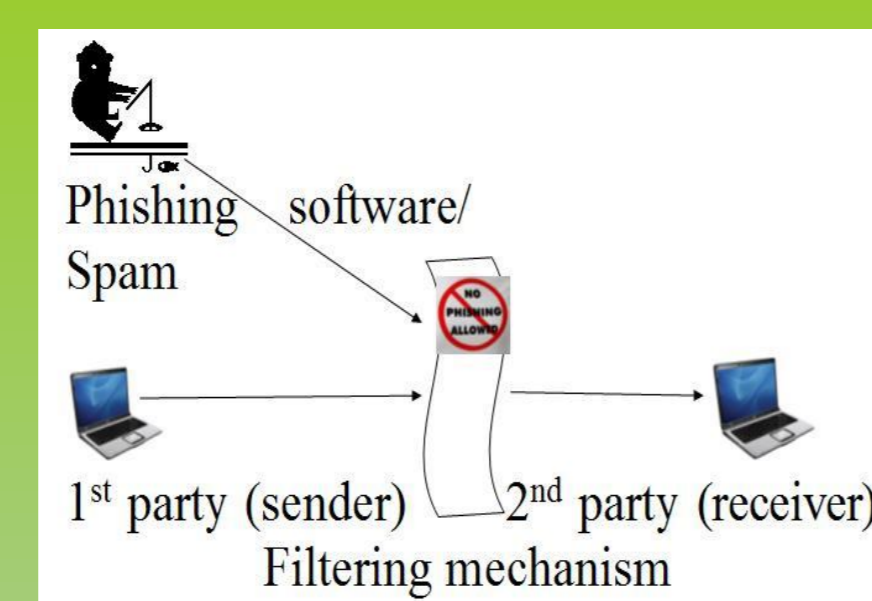


Figure 2: Fake messages Detection at the receiver side

2.1 Detection based on conversation language

Most of the time it is seen that the two parties involved in conversation uses the same language each and every time they have a chat. Suppose two people are having a chat in Bangla language, it is much likely that they will have a chat in Bangla language next time also.

To practically implement the detection we can utilize a chat history. The chat history can be saved in any instant messenger services. By analyzing the chat history and calculating the most used language, one can determine the language normally communicated between two parties. This can be done based on the Unicode of the characters [3].

The filter can work as follow:

If the new message's language is in the most used language then it will be considered as a legitimate message, otherwise it will be marked as a fake message.

For an ongoing conversation, it becomes even easier to detect a fake message. For example, if the chat is going on in Bangla language and suddenly one message arrives in other language then it is likely to be a fake message. The conversation is shown in Figure 3.

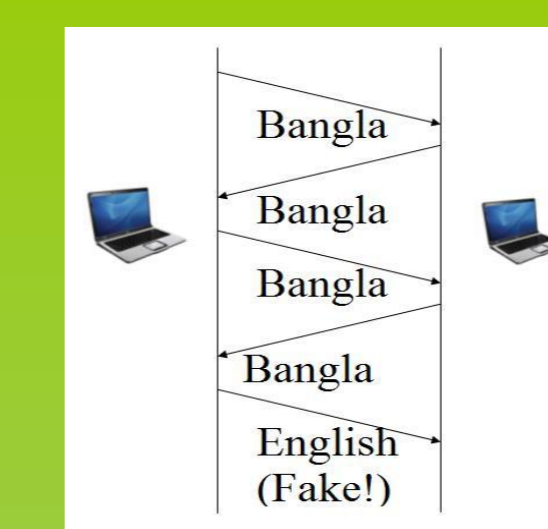


Figure 3: Detection of fake message based on conversation language

2.2 Detection based on time of Conversation

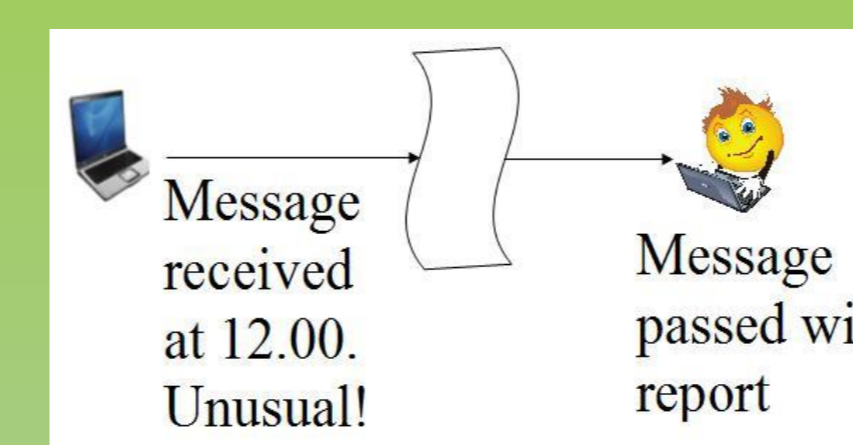


Figure 4: Detection based on conversation time

Two parties often have a chat at a particular time of the day. For example, two friends often have a chat in the evening whereas the business chats take place during office time. Whatever the topic of the chat is, it is assumable from the chat history of two parties when is the time they will usually have a chat. As an illustration, from the chat history, if it is found that most of the times two parties have chat in the evening and suddenly today one party is sending a message in the afternoon, it is most likely to be a fake message. This message will be marked as fake and the receiver will be notified about it. This is illustrated in Figure 4. It is up to the receiver now whether he will accept the message or not.

2.3 Detection based on phishing URL pattern

There are 4 characteristics of phishing URL [4]. They are- IP-Based URL, URL containing domain just registered, number of domains, and number of dot in the URL. Because phishing messages also contain such URL, we can adopt the techniques to detect the phishing message as well.

The filter can work as follow:

If the message contains URL, the URL will be checked against these characteristics. If the URL contains these characteristics to some thresholds, the message will be reported as fake.

2.4 Detection based on Google Safe Browsing API

Google Safe Browsing API is "an experimental API that enables client applications to check URLs against Google's constantly updated blacklists of suspected phishing and malware pages" [5]. As the fake message often contains link to such phishing and malware pages, we can use the API to detect such messages.

The filter can work as follow:

If the message contains URL, the URL will be checked against the Google Safe Browsing API. Then, if the URL is reported as malicious URL, report the message as fake.

3. DISCUSSION

Each of the techniques proposed here has some drawbacks. Two parties can willingly change their conversation language. In that case the messages will be falsely marked as fake. The parties can start having chat at any time of the day. So, fake message detection based on time of chat can also create false positives.

The rest two techniques rely on the message containing the URL, the characteristics of the URL, or the Google Safe Browsing API database. An attacker can send a fake message without the URL as well. As a result, the message will not be detected.

Some malicious URLs may not reflect the characteristics or may not reside in the Google Safe Browsing API database. A message contains such URLs will pass through the filter as well.

Even though these weaknesses remain in the proposals, the techniques are helpful to the users. If a suspicious message is found then an alerting notification is sent to the receiver. The receiver can then decide whether to receive the message or to ignore it.

4. CONCLUSION

This poster has delineated techniques to detect fake Instant messages. We have also pointed out some drawbacks of these techniques. We hope that our proposed techniques will help users to identify which messages are fake and which are not. And finally, the proposed techniques are only for alerting people about fake and spamming messages, they do not ensure any remedy for these messages.

References

- [1] <http://www.multilingual-search.com/forrester-research-inc-data-on-mobile-instant-messaging-im-in-europe/25/01/2008>
- [2] <http://marcosblog.com/2008/09/01/instant-messaging-global-market-share-data>
- [3] F. Haneef, F. Latif, M.S.H. Khiyal. "Unicode Aided language Identification across Multiple Scripts and Heterogeneous Data." Information Technology Journal 6(4), pages 534-540, 2007.
- [4] I. Fette, N. Sadeh, A. Tomasic 2007. "Learning to detect phishing emails." In Proceedings of the 16th international Conference on World Wide Web, Banff, Alberta, Canada, Pages 649-656, 2007.
- [5] <http://code.google.com/apis/safebrowsing>