Technical Disclosure Commons

Defensive Publications Series

February 06, 2019

Countering phishing attacks by automated responses

Brian Shucker

Brian Brewington

Countering phishing attacks by automated responses

ABSTRACT

This disclosure describes techniques to automatically respond to phishing attacks received via communication channels such as email, chat, SMS, etc. A large number of automated responses require the attacker to differentiate between automated responses and human responses which can require substantial effort. In this manner, the described techniques actively disincentivize phishing attacks by causing an increase in the cost for an attacker to conduct such attacks.

KEYWORDS

- phishing
- automatic response
- active defense
- passive defense
- spam filter
- bot
- virtual assistant

BACKGROUND

Phishing attacks have many direct and indirect costs. Attackers use various channels such as email, SMS, chat/messaging platforms, etc. to find targets that respond to the bait, e.g., click on a malicious link provided via the channel. For example, the link may be to a webpage that attempts to gather confidential information from targets. The attacker then uses the obtained confidential information to access user information, e.g., financial information, private

information such as email, photos, and other content stored in online services, etc. and to launch other attacks against the target. For example, a phishing email may include content that is designed to be perceived as being from a bank and require the user to click a link and verify account details.

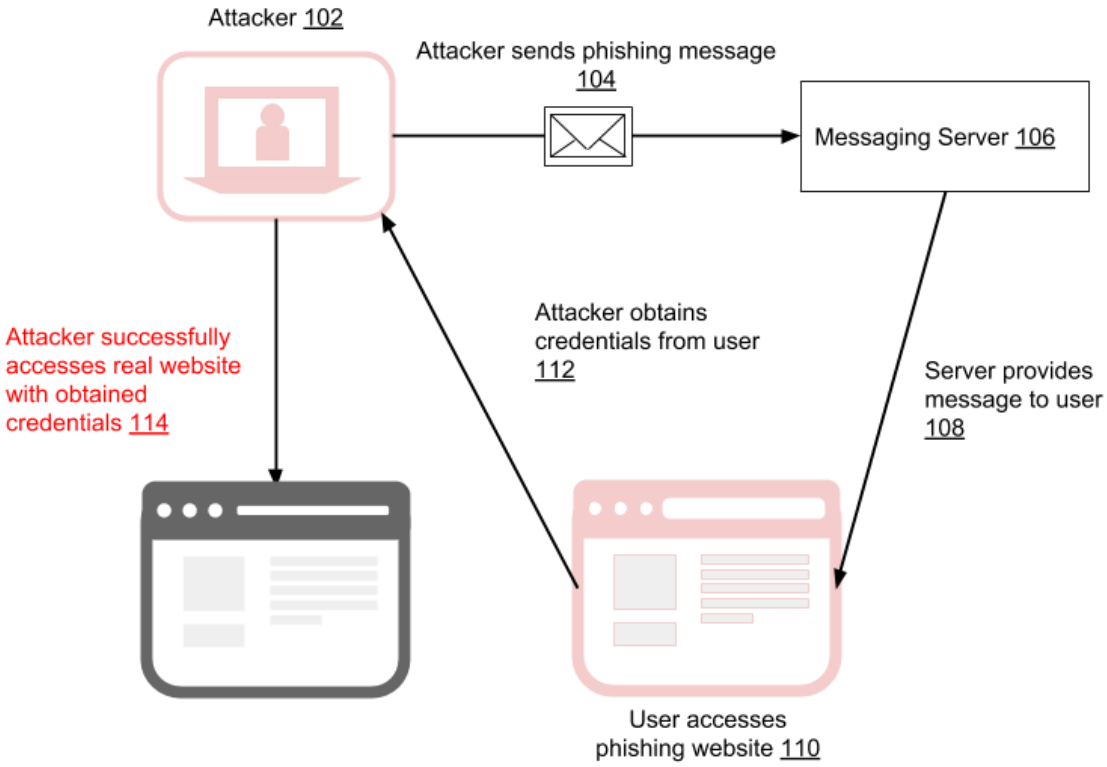

Fig. 1: Example phishing attack

Fig. 1 illustrates an example of a successful phishing attack. An attacker (102) sends a large number of phishing messages (104) to various targets via email, messaging platforms, SMS, etc. A messaging server (106) receives and forward the messages to users (108). While many of the messages may be detected by a spam filter and therefore filtered, a fraction of the messages that are not detected as phishing messages are still provided to users. When a user accesses the phishing website (110), e.g., designed to mimic the look-and-feel of a real website,

and provides credentials (112), the attacker can use such credentials to successfully access the real website (114).

Reducing phishing attacks improves security, reduces cost, and increases user trust. Passive approaches such as spam filters that are able to intercept a large number of phishing messages and warn or prevent users from clicking on links in such messages. Such passive defense is only able to protect users that have the defense system in place.

DESCRIPTION

This disclosure describes techniques to automatically respond to phishing messages sent via communication channels such as email, chat, SMS, etc. using a bot. The automatic responses actively increase the cost of phishing attacks.

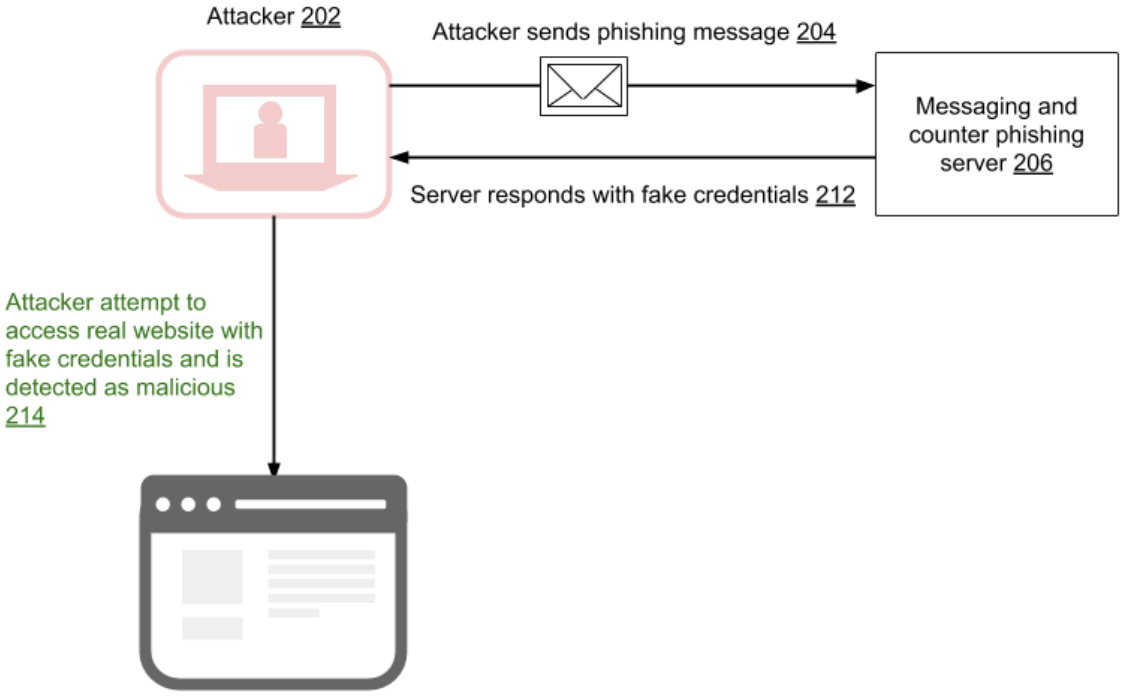

Fig. 2: Countering phishing attacks with automatic responses

Fig. 2 illustrates an example of countering phishing attacks, using techniques described herein. When an attacker (202) sends a phishing message (204), a messaging server (106) (which may implement counter phishing techniques, or coordinate with other servers) detects that the message is a phishing message, e.g., using techniques similar to spam filters, based on user input, etc. Upon detection of the phishing message, the server responds (212) with fake credentials for the real website. When the attacker attempts to access the real website with such credentials (214), the real website can detect the attack based on the credentials and take mitigating action.

The automated response can be generated by using techniques such as those used to implement chat bots or virtual assistants. The responses are provided in a manner that makes it difficult for the attacker to separate the automated responses from human responses.

For example, the counter phishing server can respond to a web form provided on the attacker website with fake information (e.g., login credentials for the real website). Such responses may be provided automatically for each instance of the phishing message, thus flooding the attacker with responses. This increases the cost for the attacker to separate any user-provided real credentials (from users that received the message) from automated responses.

Further, the fake information can be generated such that the information allows tracing attempted logins back to a phishing attack. For example, a cryptographic means of generating invalid credentials can be implemented to provide fake credentials to the attacker. For example, if the real website is a bank, and the login information includes an account number, the counter phishing server may be set up to generate fake account numbers (e.g., using cryptographic means provided by the bank) that cannot be distinguished from real account numbers. The real website can detect attempts to use such credentials and implement measures to mitigate such attacks.

The techniques allow phishing attempts to be detected before real users respond. Further, by increasing the cost to conduct a successful attack, the techniques improve security for all users, regardless of whether they have passive defense systems in place. The techniques are implemented with user permission to access message data to detect phishing attacks, and to generate and send automated responses to detected phishing messages. The techniques can be utilized by email providers, chat/message platforms, cybersecurity providers, internet service providers (ISPs), or other parties.

CONCLUSION

This disclosure describes techniques to automatically respond to phishing attacks received via communication channels such as email, chat, SMS, etc. A large number of automated responses require the attacker to differentiate between automated responses and human responses which can require substantial effort. In this manner, the described techniques actively disincentivize phishing attacks by causing an increase in the cost for an attacker to conduct such attacks.