Phish Testing is the Enemy of Phishing Defense

5 Steps to Effective Phishing Simulation Programs

COFENSE.COM
© Cofense 2019. All rights reserved.
It is an uncomfortable truth that no matter how good your perimeter controls, malicious emails are still making their way to the inbox. Indeed, recent Cofense research observed that 90% of all phishing threats (Malware, Credential Phish, Business Email Compromise and Scams) identified by the Cofense Phishing Defense Center were found in environments using Secure Email Gateways.

As organizations begin to accept this uncomfortable truth, it is natural for them to want to attempt to understand their risk posture – answering the question “How vulnerable are my users to phishing threats?” To do this they turn to ‘phishing tests’ – the human equivalent of a penetration test, where the recipient passes or fails. Unfortunately, unlike machines and code, humans can’t be patched to close known vulnerabilities. Therefore, the results from phishing ‘tests’ cannot be effectively operationalized to reliably measure reduce phishing risk to the organization. More significant is the harm that phishing ‘tests’ can do to an organization’s ability to enlist users in phishing defense.

All phishing awareness campaigns and activities are designed to reduce risk. This guide aims to provide best practice guidance on how phishing simulation programs can measurably demonstrate reduction in risk by promoting desired behavior when a real attack campaign hits.
1. **BE OPEN & TRANSPARENT ABOUT THE PROGRAM**

When embarking on phishing simulation programs it is essential to be transparent with your users that the program exists; the reasons for its existence; and the desired outcomes.

When phishing simulation programs are approached as ‘tests,’ there is a desire to keep them secret, with the mistaken belief that telling users about the ‘tests’ will influence the results and provide an inaccurate picture of risk. However, even in the most openly communicated programs, end users do not know when they will receive simulations, and what form they will take. When users are aware that they will be receiving phishing simulations – just as they sometimes receive real phishing emails – they will be more vigilant when reading all emails, thus providing further opportunity for risk reduction.

<table>
<thead>
<tr>
<th><strong>Phish Testing</strong></th>
<th><strong>Phish Behavioral Conditioning</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Focused on failure</td>
<td>Focused on desired behavior</td>
</tr>
<tr>
<td>Infrequent – often once or twice per year</td>
<td>Ongoing – regular scenarios addressing evolving threat actor tactics</td>
</tr>
<tr>
<td>Compliance driven</td>
<td>Risk driven - focused on real organizational threats</td>
</tr>
<tr>
<td>Impossible to measure true impact on improved security posture</td>
<td>Demonstrable impact on improved security posture</td>
</tr>
</tbody>
</table>
It is essential that end users understand why the phishing simulation program exists. When phishing ‘tests’ are conducted, a pass or fail mentality can dominate, negatively impacting the perception of the program. When users understand that the program is part of broader training initiatives, where the goal is to help users identify emerging threats and actively participate in the defense of the organization, the perception can shift. Rather than feeling “the organization is trying to trick me,” the mindset becomes more positive, driven by the innate human desire to help. It is this positive sentiment that can be harnessed and nurtured by Security Awareness professionals to increase program engagement and overall program success.

Also remember that the knowledge learned through the program is transferable - users can apply it to their personal email too. This is especially important for organizations that allow their users to access personal email accounts on corporate devices or allow corporate email accounts to be used for personal purposes.

Appropriate definition of desired program outcomes will ensure that success can be effectively measured and communicated. We’ll explore these outcomes in the next section.
2. **DEFINE THE DESIRED OUTCOMES**

As we’ve already discussed, phishing ‘tests’ lead to simplistic outcomes of pass or fail. However, with an overarching goal of better defense against phishing threats, Security Awareness teams must cast a net wider in their definition of what constitutes a successful simulation program.

The phishing threat landscape continues to evolve at a rapid pace. New phishing threat tactics and techniques emerge on an almost daily basis. The primary goal of phishing simulation programs should be to keep the risks of phishing front and center in users’ minds and condition them to be able to recognize these new and emerging threat actor tactics. In the same way an Olympic athlete trains to compete, your end users should be training to ensure they perform to their best ability in a real attack situation.

When an end user receives a suspicious email, there are a number of broad behaviors they can exhibit:

<table>
<thead>
<tr>
<th>Reporting</th>
<th>Response</th>
<th>Behavior</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did Not Report</td>
<td>Opened Attachment/Clicked Link (did not report)</td>
<td>Undesirable</td>
<td>User viewed the phishing email then clicked the link or opened the attachment</td>
</tr>
<tr>
<td>No Action</td>
<td>Viewed Email (did not report)</td>
<td>Neutral</td>
<td>User viewed the phishing email, and perhaps deleted it, but did not report it</td>
</tr>
<tr>
<td>Reported</td>
<td>Opened Attachment/Clicked Link Then Reported</td>
<td>Desirable</td>
<td>Opened Attachment/Clicked Link Then Reported</td>
</tr>
<tr>
<td></td>
<td>Reported Only</td>
<td>Very Desirable</td>
<td>Reported Only</td>
</tr>
</tbody>
</table>
Unsurprisingly, the most undesired behavior is that the user read the email and then clicked the link or opened the attachment – both actions that can lead to compromise in a real attack situation. Neutral behavior is when the user reads the email and perhaps recognizes it as a threat and deletes it. Many initially feel that this should be a desirable behavior. However, when a phishing attack evades technical controls and is delivered to the end user, Security Operations teams are blind to the threat. The action of a user deleting a suspicious email provides protection to them but does not help to protect the organization overall. By reporting the suspicious email, Security Operations teams can get much needed visibility, and respond appropriately.

Therefore, the desired outcome in phishing simulation programs should be to encourage end user reporting of suspicious emails. Simulation reporting rate is an unambiguous indicator of program success – after all, clicking a ‘Report Phishing’ button is a conscious and deliberate action. It may also surprise some that an instance where a user falls susceptible to a phish, but subsequently reports it, is a desired behavior. By promoting reporting of simulation emails as the most desired outcome, organizations can address other challenges such as users creating rules to auto-delete simulation emails sent by a known platform. No longer does the rule automatically produce desired behavior and no longer are your metrics negatively impacted. By creating a culture of reporting and fostering a safe environment in which users can admit to mistakes, Security Operations teams can respond much faster and prevent an initial compromise becoming a headline-grabbing breach.
3. MEASURE THE RIGHT METRICS

When phishing ‘tests’ are conducted, measurement is driven by the pass or fail mindset, and a fixation is placed upon susceptibility, or click rate. However, click rate is an inherently unreliable metric. How do you compare it over time? Do you send the same simulation again some months later? If so, the testing environment is already different – users have already seen the email once and how often do phishing threat actors send the same email to the same recipient repeatedly? Do you send a different simulation email? Will the motivators be the same? How likely is it that the recipients will have the same external factors potentially affecting their actions – such as workload or personal circumstances – and will they respond to the emotional triggers in the same way each time?

Many other factors affect the reliability of click rate as an isolated metric to measure actual risk. Did the recipient actually receive the email into their inbox? Did they actually see the email with everything else that was going on when they received it? Just because they did not click a link, or open an attachment, was that a conscious decision that was made? All of these factors impact your ability to know with a degree of certainty that you are truly improving the security posture of your organization when relying on click rate as a primary metric. Indeed, NIST agrees with this assertion. Referencing their publication ‘User Context: An Explanatory Variable in Phishing Susceptibility’ they state:
Given the variety of phishing premises and user contexts, no amount of training will consistently reduce click rates to zero, but the findings helped better understand the user’s role in early detection, combined with technological solutions, and determined that awareness training and reporting should be fully supported and even incentivized in the workforce.

Many organizations set arbitrary targets on what click rate should be and strive for a linear reduction over time to meet them. However, a simple linear decline in click rate over time is the sign of an unhealthy phishing simulation program. As we’ve mentioned previously, the phishing threat landscape is fast evolving. Emerging tactics and techniques change what phishing looks like. Motivators change, threats become more complex and challenging to identify. It’s essential that phishing simulation programs mirror this evolution. In other words, it’s entirely normal (and desired) for click rate to fluctuate in a healthy program. Some simulation campaigns will result in higher click rate – particularly those that involve novel tactics and techniques or target specific groups of individuals. To measure program effectiveness and understand the impact it is having on reduction of risk, it is necessary to look beyond click rate and measure reporting and resilience.

Resilience to phishing is measured by dividing the number of users who report a simulation by the number of users who clicked. Strive for a number greater than 1 – that is, at least an equal number reporting to those clicking. For example, if 100 users click a link in a simulation, but 50 report it, the resiliency score is 0.5. Cofense customers across different industries demonstrate different levels of resilience – for example, customers in the Energy sector have an average resilience greater than 5 – i.e. 5 users report the email for every 1 user who clicks.
Imagine that behavior in a real attack situation – Security Operations teams are given visibility of threats they were likely otherwise blind to and Security Awareness teams can justifiably promote the fact that their programs have truly reduced risk.

So, rather than just measuring click rate, measure reporting and calculate resilience. Set initial targets around the desired resiliency rate, greater than 1. That way, you can vary the complexity of your simulations and introduce evolving tactics and techniques without getting hung up on click rate. The click rate can vary, but if resilience remains within target and users are still exhibiting the desired behaviors to give your Security Operations teams visibility of threats they otherwise wouldn’t see – you’re golden. As your program develops, don’t just benchmark your performance against others in your industry – strive to meet and exceed the resilience of the highest performing industries.

Don’t overlook other metrics such as time to report – i.e. how quickly the user reports the simulation email after receiving it. By lowering this metric, you’re directly impacting the ‘mean time to detect’ or even ‘mean time to remediate metrics that are so critical to your Security Operations teams.
For metrics to be truly reliable, you need to maximize the chance that the simulation message will be seen and thus interacted with. Many factors influence this, such as inline email controls and recipient workload and travel schedules. For example, you may want to send a simulation on a specific day or at a certain time to ensure that more recipients interact with it. But what about recipients who are traveling in different time zones? What about users who are out of the office for a few days and are likely to return to a full inbox where the simulation email will be deeply buried? It’s therefore essential that the simulation platform is able to deliver emails directly to the user inbox, only when they are active. This provides reliable reporting of mail delivery and knowledge that users received the simulation email at the best time, thus maximizing the educational experience.

When you have reliable simulation metrics focused on reporting of suspicious emails and resilience to phishing, you can combine them with broader metrics for a more complete picture on reduction of risk. Look to include the number of non-simulation emails that are reported to your Security Operations teams. Measure the number and percentage of these emails that are categorized as malicious. This provides rich insight into how well phishing simulation programs are actively helping the organization to better defend against phishing attacks.
4. **FOCUS ON THE RIGHT THREATS**

Most organizations are limited in the number of simulation campaigns that they can run in any year. Every single one of them must count and have the biggest potential impact on reduction of risk. All phishing simulation platforms of note provide extensive libraries of email templates upon which simulation campaigns can be based. Rather than picking at random, carefully consider your template choices.

Some vendors strongly advocate the use of randomization within simulation programs. This involves sending each recipient in the simulation campaign a different email, the reason being that you don’t want one recipient to ‘warn’ their colleagues about the ‘test.’ When the outcome of the program is to better defend against phishing threats, rather than ‘testing,’ advocating randomization is bad advice. It demonstrates a lack of understanding of how mature organizations increase their resilience to phishing attacks. To determine whether randomization is right for your organization, ask yourself this question:

“When a real phishing campaign hits, do I want my users a) warning each other about the threat, telling their colleagues not to click, or b) remaining silent?”

If your answer is a), great job! Randomization is undesired, as you want your simulation programs to condition users to behave correctly in a real attack situation. If your answer is b), we’d respectfully suggest you re-evaluate the goal of your programs, as they are not currently optimized to stop real phishing attacks.
Phishing simulation programs are not one-size fits all. They should be based upon evolving threat actor tactics impacting your industry or location and on the specific threats your organization is facing. When threat reporting becomes the primary outcome of simulations, that behavior becomes second nature when real threats land in the inbox. The culture of reporting – the promotion of a mindset of ‘see something, say something’ – provides visibility into the type of phishing threats that are bypassing your technical controls, such as a Secure Email Gateway, and are being delivered to user inboxes.

These reported emails provide a valuable source of intelligence to influence your phishing simulation programs. By working with Security Operations, Security Awareness can increase the relevance of phishing simulation programs by ensuring they have a high degree of focus on the type of threats actually received. There is little to be gained in running multiple attachment-based scenarios if the majority of threats you receive are credential phish. This collaboration between Security Awareness and Security Operations provides benefits on both sides. Not only can Security Awareness teams demonstrate that their programs are relevant and improving security posture, Security Operations teams work safe in the knowledge that phishing simulation campaigns are focusing on the biggest threats. This leads to a reduction in SOC workload by having more risk-aware end users who report threats, increase visibility, and enable fast response. This ultimately reduces the mean time to remediate, as you’ll never have a zero click rate in any type of phishing campaign.
5. **ANALYZE, COMMUNICATE, REPEAT**

To keep pace with the evolution of the phishing threat landscape, phishing simulation programs cannot be one-and-done exercises. Simulations should be ongoing, addressing emerging threats. Results of these programs should be studied to ensure they remain relevant and help the organization better defend against real attacks.

When analyzing program results, look for areas where improvements in resilience can be made. In many cases, a typical bell curve could be applied to user behavior in phishing simulations:

- **Will Click Often, Never Reports**
- **Occasional Clicks, Very Low Reporting**
- **Rarely Clicks, Infrequent Reporting**
- **Never Clicks, Always Reports**
The greatest improvements are gained by focusing on those users who demonstrate an ability and willingness to do the right thing, even if it is infrequent. This enables the organization to not only reduce the risk of errant clicks, but importantly maximizes the visibility of threats to drive faster response. Programs should also aim to accommodate the outliers in the curve – such as appropriate rewards and recognition for those that always demonstrate the desired behavior – to encourage others to do the same. Consider appropriate measures to reduce risk associated with those that click often and never report. Depending upon your organizational risk appetite, this can range from targeted simulations to repeat clickers to implementation of increased controls.

It is increasingly common for executive boards to follow the results of phishing simulation programs. Some public organizations publish the results of these programs in quarterly and annual filings, tying them to bonuses. Therefore it’s critical that the metrics used are relevant and can be easily communicated and understood in the context of the organization, with appropriate comparison and benchmarking to industry peers and broader verticals.
Broaden the scope of the metrics you report to support holistic phishing defense activities:

- Categorization and analysis of identified threats: understand the magnitude specific to your organization and where to focus your efforts to improve recognition and reporting.

- Resiliency trend over time: in addition to specific comparisons, understand your overall resiliency over time – it should get better as your program matures.

- What types of phish are entering the environment?

- What strains of malware and exploits do we need to ensure our endpoints are not vulnerable to?

- How many threats have we identified that we were otherwise blind to?

- How quickly are we able to analyze user reported emails?

- Are we improving our mean time to respond?

And remember, when you reach the nirvana of your desired resiliency rate or your resiliency rate outstrips that of other organizations, don’t pump the brakes on your program. Keep going – threat actors will continue to innovate, so continue to tune your vital human intelligence engine to identify attacks and stop them in their tracks.

\(^1\)Kristen K Greene et al “User Context: An Explanatory Variable in Phishing Susceptibility” (NIST)