Instegogram: A New Threat and Its Limits for Liability

Jennifer Deutsch & Daniel Garrie*

Social media networks represent the largest, most dynamic risk to organizational security and allocating liability.[1] Recently, a new risk was developed combining digital image steganography and social media into the corporate environment. Last year, researchers demonstrated that "malware hidden in images posted to social media sites can be used for command and control (C2) channels."[2] While neither steganography nor social media are new, it is novel to combine both as a tool for malware distribution.[3] This article considers whether a company can be held liable to a third-party for unknowingly posting an image with embedded malware.

Generally, steganography involves placing a hidden message within a transport medium, in such a way that the casual observer is not aware that a message had been sent. Digital image steganography is the practice of hiding code inside images. While the hidden code slightly alters the original image's appearance, changing its color tone and

* *Jennifer Deutsch* is a consultant for Law & Forensics and is a staff editor for the Journal of Law & Cyber Warfare.
Daniel Garrie is the Editor-in-Chief for the Journal of Law & Cyber Warfare. He is also the Managing Partner of Law and Forensics, a Forensic Neutral for JAMS, and a Partner with ZEK.

[1] *What is Social Media Security*, ZEROFOX, https://www.zerofox.com/social-media-security (last visited Apr 17, 2017).

[2] Amanda Rossseau, Daniel Grant & Hyrum Anderson, *Instegogram: Leveraging Instagram for C2 via Image Steganography*, ENDGAME (2016), https://www.endgame.com/blog/instegogram-leveraging-instagram-c2-image-steganography (last visited Jan 28, 2017).

[3] *Id.*

making it appear more pixilated than a clean version,[4] the difference is subtle and easily goes unnoticed.[5]

Recently, malicious actors have successfully used digital image steganography to conceal the command and control operations required to operate malware.[6] Inherently, steganography with malware C2 channels embedded is an appealing tool to malicious actors because of its stealthiness.[7] Moreover, "malicious actors have leveraged these social media platforms to bolster their existing operations" because the online content is dynamic.[8] Websites constantly update and with each update comes a new opportunity for infiltration.[9]

A new scheme known as "Instegogram" "mirror[s] the utilization of social networks for C2, while exploring the

---

[4] *Id.*

[5] Rene Millman, *Huge Malvertising Campaign Uses Steganography to Hide malware in plain sight*, SC MAGAZINE UK (2016), https://www.scmagazineuk.com/huge-malvertising-campaign-uses-steganography-to-hide-malware-in-plain-sight/article/530879/ (last visited Feb 8, 2017).

[6] Rossseau, *supra* note 2.

[7] Verine Etsebeth, *Malware Attacks: Corporate Responsibility and Liability*, EMERALD GROUP PUBLISHING (2007), http://www.emeraldgrouppublishing.com/learning/management_thinking/articles/pdf/malware.pdf. (last visited Jan 2017).

[8] *Id.*

[9] Kacy Zurkus, *Social media, the gateway for malware*, CSO ONLINE (2016), http://www.csoonline.com/article/3106292/social-networking/social-media-the-gateway-for-malware.html (last visited Jan 17, 2017); *See also* Jai Vijayan, *Attack Uses Image Steganography For Stealthy Malware Ops On Instagram*, DARK READING (2016), http://www.darkreading.com/endpoint/attack-uses-image-steganography-for-stealthy-malware-ops-on-instagram/d/d-id/1327170 (last visited Apr 28, 2017).

feasibility of using steganography on a particular site - Instagram."[10] Under this scheme:

> once the remote system is compromised, encoded images can be posted from the command machine using Instagram's API. The remote system will download the image, decode it, execute the encoded commands, encode the results in another image, and post back to Instagram.[11]

Although Instegrogram was originally created for academic purposes,[12] its potential use as part of a malware attack poses the question of who would be liable for such an attack.

Under the Communications Decency Act (CDA), companies that offer web-hosting services are typically shielded from liability for most content that customers or malicious users place on the websites they host.[13] The CDA grants immunity to providers of interactive computer services from liability arising from content created by others.[14] The protections extend to individuals who operate websites and web forums to which other individuals can

---

[10] Rossseau, *supra* note 2.
[11] *Id.*
[12] *Id.*
[13] 47 U.S.C. § 230 (1998); *See also The Legal Implications of Social Networking: The Basics (Part One)*, INFOLAWGROUP LLP (2015), http://www.infolawgroup.com/2011/06/articles/social-networking/the-legal-implications-of-social-networking-the-basics-part-one/ (last visited Mar 3, 2017).
[14] Melissa Landau Steinman & Mikhia Hawkins, *When Marketing Through Social Media, Legal Risks Can Go Viral Venable*, VENABLE LLP (2010), https://www.venable.com/files/publication/b4f467b9-0666-4b36-b021-351540962d65/presentation/publicationattachment/019f4e5f-d6f8-4eeb-af43-40a4323b9ff1/social_media_white_paper.pdf (last visited Feb 17, 2017).

freely post content.[15] This encompasses almost any online service (*e.g.*, Google, Facebook, Instagram) that publishes information provided by a third-party or information content provider.[16] As such, commerce platforms and employers that provide or enable computer access for multiple users on their computer networks/servers to access the Internet may qualify for immunity.[17] Congress recognized that websites that display third-party content may have an infinite number of users generating an enormous amount of potentially harmful content,[18] and holding website operators liable for that content "would have an obvious chilling effect," in light of the difficulty of screening posts for potential issues.[19]

One crucial exception to CDA immunity is that no protection exists if the website controls the information content.[20] Thus, the CDA provides no immunity to an Internet service provider, like Yahoo, when it, rather than a third party, is the "Information Content Provider."[21] That is, if the service provider is "responsible, in whole or in part, for the creation or development of the offending content,"[22] its actions fall outside the CDA's protections. Consequently, courts have held that a service provider is not immune from

---

[15] *Donato v. Moldow*, 374 N.J.Super. 475, 865 A.2d 711 (N.J. Super. Ct. App. Div. 2005).
[16] *See Id.*; 47 U.S.C. § 230(c)(1) (1998).
[17] *Delfino v. Agilent Techs., Inc.*, 145 Cal. App. 4th 790, 52 Cal. Rptr. 3d 376 (Cal. App. Dep't Super. Ct. 2006); *Miller v. Fed. Express Corp.*, 6 N.E.3d 1006 (Ind. Ct. App. 2014).
[18] *Doe v. Backpage.com, LLC.*, 817 F.3d 12 (1st Cir. 2016).
[19] *Zeran v. America Online, Inc.*, 129 F.3d 327, 331 (4th Cir. 1997).
[20] *Malware Risks and Mitigation Report*, BITS (Jun. 2011), https://www.nist.gov/sites/default/files/documents/itl/BITS-Malware-Report-Jun2011.pdf.
[21] *Fair Housing Council of San Fernando Valley v. Roommates.com, LLC.*, 521 F.3d, 1157, 1164 (9th Cir. 2008).
[22] 47 U.S.C. § 230(f)(3) (1998).

suit where the provider itself creates or helps to develop, rather than merely publishes, the unlawful content.[23] Arguably, sponsored ads, which are the advertiser's content posted by the respective networks, have no effect on immunity. Such features merely reflect choices about what content can appear on the website and in what form, and thus are editorial choices that fall within the purview of traditional publisher functions.[24] Likewise, a company that uses a social media network to create the picture, or develop information, "controls" that information and would not be immune. Even when a company is contractually permitted to advertise on a social network, only the social network remains immune.

Whether the CDA protections extend to damage caused by malware is still largely an open question of law.[25] However, a 2003 Third Circuit case supports its application to malware.[26] In *Green v. America Online*, the court held that a malicious program constituted "information" for purposes of the CDA, even though it was not communicated in the traditional sense.[27] The court concluded that the service provider, AOL, could not be held liable for the victim's computer receiving a signal from their service if it was sent by a third party with no role in the program.[28] Thus, the Court shielded web hosts if their hosted sites are a source of malware without the host's knowledge. Under the Court's

---

[23] *Roomates.com*, 521 F.3d at 1168-69; *See also Anthony v. Yahoo! Inc.*, 421 F.Supp.2d 1257, 1262-63 (N.D. Cal. 2006).
[24] *Reit v. Yelp! Inc.*, 29 Misc.3d 713 (N.Y. Sup. Ct. 2010); *See Obado v. Magedson*, 612 Fed.Appx. 90 (3d Cir. 2015).
[25] BITS, *supra* note 20.
[26] *Green v. America Online*, 318 F.3d 465 (3d. Cir. 2003) (holding that AOL not liable for user-posted virus placed into AOL chatroom).
[27] *Id.*
[28] *Id.*

interpretation in *Green*, a web service company cannot be held liable for third-party malware on a hosted website.

Thus, it is probable that companies would be liable for any third-party damage resulting from an Instegogram attack for which they did know or should have known the digital image was infected. As no statutory immunities exist to shield social media users, a company would be liable for any resulting damage caused by a criminal hacker's embedded C2 infrastructure. But is it even likely that a content creator could have known there was embedded malware? Digital image steganography rests on an employee's ability to notice minute differences in an image,[29] so Instegogram liability would rely on the limitations of the human eye to perceive minute changes in color and light intensity, small distortions that could reasonably go unnoticed.[30]

Of course, terms of service agreements may provide further protections. Among other things, a term of service agreement may allocate responsibility for an Instegogram attack with provisions waiving liability for viruses and any resulting third-party damage. Depending on the language, such a waiver may limit, prevent, or allocate any damages resulting from an Instegogram attack.

The increasing use of social media in the corporate environment poses new avenues for liability. As shown,

---

[29] *See* Rossseau, *supra* note 2 (describing digital image steganography's ability to be invisible to a person, "if Eve intercepts the image in transit, she is oblivious to the fact that the stego image contains any message at all since the image appears to be totally legitimate both digitally and to the human eye.").

[30] Joann Kennedy, *Use Offense to Inform Defense. Find Flaws Before the Bad Guys Do Sans Penetration Testing*, SANS (2004), https://cyber-defense.sans.org/resources/papers/gsec/steganography-corporate-environment-106511.

social media networks are generally immune from the consequences of an Instegogram attack. By contrast, the content creator is left exposed to potential liability for an Instegogram attack.