DATA SCRAPING AND THE COMPUTER FRAUD AND ABUSE ACT: AN ANALYSIS OF WHEN UNWANTED DIGITAL ACCESS SHOULD IMPLICATE AN "ANTI-HACKING" STATUTE

Brent W. McDonough

I. INTRODUCTION

In 2017, hackers obtained the names, birth dates, Social Security numbers, and addresses of 143 million Americans during the Equifax data breach.[1] In 2013, every Yahoo account was hacked.[2] As more personal data is stored on Internet-connected devices, more personal data will be accessed in harmful, unexpected and unwanted ways.[3] While enhancing cybersecurity measures is an important step towards protection, total digital security will never be possible,[4] so there must be sufficient legal remedies in place as well. There are a number of data

1. *See* Seena Gressin, *The Equifax Data Breach: What to Do*, FEDERAL TRADE COMMISSION (Sept. 8, 2017), https:/À/Àwww.consumer.ftc.gov/Àblog/À2017/À09/Àequifax-data-breach-what-do. Essentially, the personal information of every American with a credit score was hacked. *Id.*

2. Elizabeth Weise, *Yahoo says 2013 hack hit all 3 billion user accounts, triple initial estimates*, USA TODAY (Oct. 3, 2017, 4:45 PM), https:/À/Àwww.usatoday.com/Àstory/Àtech/À2017/À10/À03/À3-billion-yahoo-users-breached-company-says/À729155001/À.

3. *See* Selena Larson, *Why Hacks Like Equifax Will Keep Happening*, CNN (Sept. 29, 2017, 8:49 AM), http:/À/Àmoney.cnn.com/À2017/À09/À29/Àtechnology/Àbusiness/Àequifax-hack-2017-cyberattacks/Àindex.html.

4. *See, e.g.*, Andrew McGill, *The Inevitability of Being Hacked*, THE ATLANTIC (Oct. 28, 2016), https:/À/Àwww.theatlantic.com/Àtechnology/Àarchive/À2016/À10/Àwe-built-a-fake-web-toaster-and-it-was-hacked-in-an-hour/À505571/À; Hayley Richardson, *Companies 'Must See Cyber Attacks As Inevitable'*, NEWSWEEK (Feb. 16, 2015, 1:07 PM), http:/À/Àwww.newsweek.com/Àcompanies-must-see-cyber-attacks-inevitable-307111.

protection laws in the United States,[5] but one of the most notable—and most controversial—is the Computer Fraud and Abuse Act ("CFAA").[6]

Enacted in 1986, the CFAA is best understood as a computer trespass law,[7] although its original purpose was narrowly to punish those who accessed a "federal interest computer" "without authorization."[8] The statute has been amended multiple times since its enactment in an effort to keep up with the Internet's evolving landscape.[9] While it now has a private right of action[10] and includes additional ways an individual or entity can criminally access another's computer and/or computer data, there is still considerable ambiguity in the statute's language and how it applies in the modern Internet era. Most notably Congress has left the phrase "without authorization" undefined and courts have failed to adopt a uniform understanding of when certain types of access are unauthorized so as to violate the CFAA, which is especially problematic in the context of civil lawsuits.[11]

The statute's broadest and most controversial provision, § 1030(a)(2)(C), makes liable anyone who "intentionally accesses a computer without authorization or exceeds authorized access, and thereby obtains information from any protected computer."[12] "Protected

5. *See*, *e.g.*, The Wiretap Act, 18 U.S.C. §§ 2510–2522 (1986); The Stored Communications Act, 18 U.S.C. §§ 2510–2522 (1986); The Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104–191, 110 Stat. 1936; The Fair Credit Reporting Act, 15 U.S.C. § 1681 et seq. (2016).

6. 18 U.S.C. § 1030. *See also* Tim Wu, *Fixing the Worst Law in Technology*, THE NEW YORKER (Mar. 18, 2013), https:/À/Àwww.newyorker.com/Ànews/Ànews-desk/Àfixing-the-worst-law-in-technology.

7. Orin S. Kerr, *Norms of Computer Trespass*, 116 COLUM. L. REV. 1143, 1143 (2016). The CFAA seeks to address "the unauthorized access and use of computers and computer networks." H. MARSHALL JARRETT & MICHAEL W. BAILEE, COMPUTER CRIME AND INTELLECTUAL PROPERTY SECTION CRIMINAL DIVISION, OFFICE OF LEGAL EDUCATION, PROSECUTING COMPUTER CRIMES 1 (2015).

8. *Id*. at 5. A "federal interest computer" essentially meant a computer owned by the Government or a financial institution like a bank. *Id.*

9. Congress amended 18 U.S.C. § 1030 in 1988, 1989, 1990, 1994, 1996, 2001, 2002, and 2008; *see also* JARRETT & BAILEE, *supra* note 7, at 2; Seth Rosenblatt, *Where did the CFAA come from, and where is it going?*, THE PARALLAX (Mar. 16, 2016), https:/À/Àwww.the-parallax.com/À2016/À03/À16/Àwhere-did-the-cfaa-come-from-and-where-is-it-going/À.

10. 18 U.S.C. § 1030(g).

11. *Compare* hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099, 1109 (N.D. Cal. 2017) ("[W]hether 'access' to a publicly viewable site may be deemed 'without authorization' under the CFAA where the website host purports to revoke permission is not free from ambiguity."), *with* Craigslist Inc. v. 3Taps Inc., 942 F. Supp. 2d 962, 969–70 (N.D. Cal. 2013) (holding that an individual who continues to access data on another's website after that access was explicitly revoked violates the CFAA).

12. 18 U.S.C. § 1030(a)(2)(C).

computers" include every Internet-connected device in the US,[13] thus the scope of the statute depends on how courts interpret "authorization" or "authorized access." Clearly hacking into a stranger's password-protected email account would constitute access without authorization, but in cases that fall short of hacking, authorization to access a computer or data can be very dependent on the perspectives of those who are accessing that data and those whose data has been accessed. This is problematic because individuals and entities often use § 1030 (a)(2)(C) as a sword in civil actions and if they win—i.e. they successfully prove that the other party accessed their data without authorization—there is nothing stopping a prosecutor from then pursuing a criminal case, which can potentially convert things like an innocuous Terms of Service violation into a criminal offense.[14]

A situation that has proven particularly problematic and highlights the need for a clearer standard for unauthorized access is where one party "scrapes" data from another's website, meaning they use bots to automatically extract massive quantities of data. This Commentary analyzes how the CFAA should apply in such situations and discusses the decision in *hiQ Labs, Inc. v. LinkedIn Corp.*, 273 F. Supp. 3d 1099 (N.D. Cal. 2017), where the court found that LinkedIn had no authority to stop hiQ, a tech start-up, from "scraping" user data from LinkedIn because those users had made their profile pages publicly viewable. The court found that accessing data on a public webpage can never violate the CFAA, regardless of how and why that data is accessed. This Commentary argues that the court's analysis is too simplistic and proposes a new authorization standard.

II. WHAT IS DATA SCRAPING?

Data scraping "is the act of taking content from a website with the intent of using it for purposes outside the direct control of the site

13. Congress defines protected computer as any computer "used in or affecting interstate or foreign commerce or communication," which effectively means any computer connected to the Internet. *See* 18 U.S.C. § 1030(e)(2); JARRETT & BAILEE, *supra* note 7, at 4.

14. *See* Orin S. Kerr, *Vagueness Challenges to the Computer Fraud and Abuse Act*, 94 MINN. L. REV. 1561, 1587 (2010) [hereinafter *Vagueness Challenges*] (arguing that "federal prosecutors . . . try to exploit the breadth and ambiguity of the statute to bring prosecutions based on aggressive readings of the statute"); Parker Higgins, *Critical Fixes for the Computer Fraud and Abuse Act*, ELECTRONIC FRONTIER FOUNDATION (Jan. 29, 2013), https:/À/Àwww.eff.org/Àdeeplinks/À2013/À01/Àthese-are-critical-fixes-computer-fraud-and-abuse-act.

owner."[15] Unlike web indexing, a practice which companies like Google use to track down relevant information and links to be included in keyword search results,[16] data scraping entails retrieving data off of a third-party's website without that party's knowledge or permission.[17] Web indexing and data scraping both require the use of bots, but web indexing bots are often considered good, while data scraping bots are considered bad.[18] Website hosts try to protect themselves in two common ways. First, they implement technical barriers to block bad bots.[19] For example, LinkedIn has at least six "automated countermeasures" aimed at preventing data scraping.[20] Second, they try to create legally enforceable protections against scraping by prohibiting it in their Terms of Service or User Agreements.[21] Notably, however, the Ninth Circuit has held that Terms of Service violations alone will not implicate the CFAA.[22] The next section analyzes *hiQ* and discusses how the court distinguished precedent to create a bright line standard for access "without authorization."

15. Courtney Cleaves, *Web Scraping Protection: Everything You Wanted To Know (but were afraid to ask)*, DISTIL NETWORKS BLOG, https:/À/Àresources.distilnetworks.com/Àall-blog-posts/Àweb-scraping-everything-you-wanted-to-know-but-were-afraid-to-ask (last visited Nov. 20, 2018).

16. *How Search organizes information*, GOOGLE, https:/À/Àwww.google.com/Àsearch/Àhowsearchworks/Àcrawling-indexing/À (last visited Nov. 20, 2018).

17. *See* Cleaves, *supra* note 15.

18. *See id.*; *Stop Web Scraping*, DISTIL NETWORKS, https:/À/Àwww.distilnetworks.com/Àweb-scraping/À (last visited Nov. 20, 2018).

19. *See* Appellant's Opening Brief at 7, hiQ Labs, Inc. v. LinkedIn Corp., No.17-1683 (N.D. Cal. Oct. 3, 2017).

20. The six technical countermeasures LinkedIn uses to protect itself against bots and scraping are its: (1) "FUSE system, which scans and imposes a limit on the activity that a user my initiate on the website;" (2) its "Quicksand system, which monitors patterns of access to LinkedIn's servers to look for non-human activity indicative of scraping;" (3) "Sentinel system, which scans, throttles, and at times blocks suspicious activity associated with specific Internet Protocol (or IP) addresses;" (4) "Org Block system, which blocks a manually-created list of IP addresses and contains a program to identify IP addresses used by large-scale scrapers;" (5) "Request Scoring systems, which monitor and restrict activity indicative of access by bots; and (6) "'robots.txt' file, which provides instructions to bots that attempt to access LinkedIn's servers and prohibits automated programs like those used by automated data scraper." *Id.*

21. *See*, *e.g.*, *Prohibited Software and Extensions*, LINKEDIN (Nov. 17, 2017), https:/À/Àwww.linkedin.com/Àhelp/Àlinkedin/Àanswer/À56347/Àprohibition-of-scraping-software?lang=en; *Automated Data Collection Terms*, FACEBOOK, (Apr. 15, 2010), https:/À/Àwww.facebook.com/Àapps/Àsitescrapingtosterms.php.

22. Facebook, Inc. v. Power Ventures, Inc., 844 F.3d 1058, 1067 (9th Cir. 2016).

RUTGERS UNIVERSITY LAW REVIEW

III. *HIQ LABS V. LINKEDIN*

Founded in 2002 and acquired by Microsoft in 2016, LinkedIn asserts itself as "the world's largest professional network with more than 546 million users in more than 200 countries and territories worldwide."[23] hiQ, on other hand, is a bit younger, much smaller, and specifically focuses on employee recruitment and retention software.[24] hiQ was founded in 2012 "to improve HR through data science."[25] hiQ and its team of engineers have developed two software tools—"Skill Mapper" and "Keeper"[26]—that help recruiters and HR departments evaluate the qualifications and career trajectory of current and prospective employees.[27] These tools are powered by the data hiQ scrapes from LinkedIn,[28] which includes profile updates,[29] articles viewed and/or "liked," influencers and companies "followed," and comments and posts written.[30]

23. *About*, LINKEDIN, https:/À/Àpress.linkedin.com/Àabout-linkedin (last visited Nov. 20, 2018).

24. HIQ LABS, INC., https:/À/Àwww.hiqlabs.com/Ànew-who-we-are (last visited Nov. 20, 2018).

25. *Id.*

26. *See Enterprise Solutions*, HIQ, https:/À/Àwww.hiqlabs.com/Ànew-index/À, (last visited Nov. 20, 2018); *see also* John E. Dunn, *LinkedIn Accused of Chilling Access to Information Online*, NAKED SECURITY BY SOHPOS (Nov. 8, 2018), https:/À/Ànakedsecurity.sophos.com/À2017/À12/À19/Àlinkedin-accused-of-chilling-access-to-information-online/À ("Keeper can be used by employers to detect staff that might be thinking about leaving while Skill Mapper summarizes the skills and status of current and future employees.").

27. *Enterprise Solutions*, *supra* note 26 ("We provide a crystal ball that helps you determine skills gaps or turnover risks months ahead of time, and a platform that shows you how and where to focus your efforts.").

28. hiQ Labs, Inc. v. LinkedIn Corp., 273 F. Supp. 3d 1099, 1104 (N.D. Cal. 2017) ("hiQ's [business] model is predicated entirely on access to data LinkedIn users have opted to publish publicly. hiQ relies on LinkedIn data because LinkedIn is the dominant player in the field of professional networking.").

29. Profile updates can include changing a profile picture, editing or adding descriptions under the work experiences users list on their LinkedIn profile page, and adding or removing skills. *See Adding, Editing, or Removing a Position in Your Profile's Experience Section*, LINKEDIN, https:/À/Àwww.linkedin.com/Àhelp/Àlinkedin/Àanswer/À1646/Àadding-editing-or-removing-a-position-in-your-profile-s-experience-section?lang=en, (last visited Nov. 20, 2018); *Adding and Removing Skills on Your Profile*, LINKEDIN, https:/À/Àwww.linkedin.com/Àhelp/Àlinkedin/Àanswer/À4976?query=skills, (last visited Nov. 20, 2018).

30. *See* Shaun Nichols, *hiQ prevails /À LinkedIn must allow scraping /À Of your page info*, THE REGISTER (Aug. 14, 2017, 11:28 PM GMT), https:/À/Àwww.theregister.co.uk/À2017/À08/À14/Àhiqlinkedinbotsscraping/À; *Drawing the Line with Public Web Data*, HIQ,

Interestingly, hiQ scraped LinkedIn's data for five years before LinkedIn sent hiQ a Cease-and-Desist Letter, demanding hiQ stop its data scraping.[31] LinkedIn also implemented IP blocks, which would allow its servers to recognize hiQ's scraping bots and prevent them from extracting LinkedIn profile data.[32] However, hiQ was able to circumvent those blocks multiple times and continue scraping.[33] Somewhat surprisingly, hiQ beat LinkedIn to court and filed a preliminary injunction in the Northern District of California, claiming, *inter alia*, that "LinkedIn's actions constitute[d] unfair business practices under Cal. Bus. & Prof. Code §§ 17200 *et seq*."[34] Regarding the preliminary injunction, the court ruled in favor of hiQ, finding hiQ would likely go out of business if it did not enjoin LinkedIn from preventing hiQ's scraping. The court found that hiQ's continued existence outweighed the privacy interests of LinkedIn users in their personal data.[35]

Because this was a preliminary injunction, the court did not decide the case on its merits; however, it did present arguments on the relevant legal claims.[36] While there were a number of claims to address, the key issue was "[w]hether hiQ's continued access to the LinkedIn public profiles violate[d] the CFAA. . . ."[37] In relevant part, the court held that hiQ had raised serious questions as to the merits of LinkedIn's CFAA claim.[38] The underlying rationale behind the court's holding was that the CFAA is ambiguous and courts should not find violations where an entity accesses publicly available data, regardless of whether a website host has

https:/À/Àstatic1.squarespace.com/Àstatic/À5803b57737c581885cbd0667/Àt/À59c424aa80bd5edc54f3e437/À1506026666468/ÀPublicWebData.pdf, (last visited Nov. 20, 2018).

31. *See hiQ* 273 F. Supp. 3d at 1107; Cease-And-Desist Letter from Abhishek Bajoria, Senior Litigation Counsel, LinkedIn Corp., to Mary Weidick, hiQ Labs, Inc. (May 23, 2017), https:/À/Àstatic1.squarespace.com/Àstatic/À5803b57737c581885cbd0667/Àt/À59721e45725e2539a60bb195/À1500651078233/ÀLetter+from+LinkedIn+to+HiQ+Labs.pdf.

32. *hiQ*, 273 F. Supp. 3d at 1103.

33. *See id*.

34. *Id.* ("hiQ has raised serious questions as to whether LinkedIn, in blocking hiQ's access to public data, possibly as a means of limiting competition, violates state law."). hiQ also brought promissory estoppel and free speech claims. *Id*.

35. *Id*. at 1119. The court reasoned:

> [T]he actual privacy interests of LinkedIn users in their *public* data are at best uncertain. It is likely that those who opt for the public view setting expect their public profile will be subject to searches, date mining, aggregation, and analysis. On the other hand, conferring on private entities such as LinkedIn, the blanket authority to block viewers from accessing information publicly available on its website for any reason, backed by sanctions of the CFAA, could pose an ominous threat to public discourse and the free flow of information. . . . *Id*.

36. *Id.* at 1105.

37. *Id*. at 1108.

38. *See generally id.* at 1103.

explicitly stated that it does not want its data accessed in a certain way or by a certain entity or individual.[39]

In concluding that the CFAA does not apply to publicly accessible data, like LinkedIn profile data, the court had to distinguish Ninth Circuit precedent, which arguably stands for the propositions that (1) access "without authorization" under the CFAA is unambiguous; (2) companies like LinkedIn have the authority to revoke a data scraper's access to their website; and (3) any violation of that revocation implicates the CFAA.[40] The Northern District reasoned that the Ninth Circuit's decisions in *Facebook, Inc. v. Power Ventures, Inc.* and *United States v. Nosal*, both of which were binding precedent, were not directly on point because neither specifically dealt with a situation where the CFAA was being used to restrict access to otherwise publicly viewable data.[41] The court reasoned that, while a literal interpretation of the CFAA suggests "where authorization has been revoked by the website host, that 'access' can be said to be 'without authorization,'"[42] Congressional intent in the context of the CFAA's application to scraping publicly viewable data was "not free from ambiguity."[43] The court stated that Congress could not have "intended to police traffic to publicly available websites on the Internet [because] the Internet did not exist in 1984."[44] Rather, the CFAA was enacted to deter and punish "'hacking' or 'trespass' onto private, often password-protected mainframe computers."[45] Despite the fact that Congress has expanded the scope of the CFAA over the years and LinkedIn's CFAA claim constitutes a civil action under 18 U.S.C. § 1030(g), the court stated "[the] construction of the CFAA must take into account the fact the statute may be enforced criminally and that its

39. *See id*. at 1113–15.

40. *See id*. at 1109; Facebook, Inc. v. Power Ventures, Inc., 844 F.3d 1058, 1067 (9th Cir. 2016) ("[A] defendant can run afoul of the CFAA when he or she has no permission to access a computer or when such permission has been revoked explicitly. Once permission has been revoked, technological gamesmanship or the enlisting of a third party to aid in access will not excuse liability."); United States v. Nosal, 844 F.3d 1024, 1028 (9th Cir. 2016).

41. Writing for the court, Judge Chen reasoned that neither *Nosal* nor *Facebook* "confronted the precise issue presented here: whether visiting and collecting information from a publicly available website may be deemed 'access' to a computer 'without authorization' within the meaning of the CFAA where the owner of the web site has selectively revoked permission." *See hiQ*, 273 F. Supp. 3d at 1109.

42. *See id*. (citing Craigslist Inc. v. 3Taps Inc., 942 F. Supp. 2d 962 (N.D. Cal. 2013).

43. *See id*.

44. *See id*.

45. *Id*. (citing H.R. REP. NO. 98–894, 1984 U.S.C.C.A.N. 3689, 3691–92, 3695–97 (1984); S. REP. NO. 99–432, 1986 U.S.C.C.A.N. 2479, 2480 (1986)).

interpretation would apply uniformly to criminal as well as civil enforcement."[46]

The court extensively cited and then adopted the approach to access "without authorization" taken by Professor Orin Kerr in his law review article, *Norms of Computer Trespass*.[47] Kerr argues that authorization and access to digital content under the CFAA should be understood using similar norms that guide trespass law.[48] To better understand digital trespass, Kerr argues that lawmakers and courts must inquire into: (1) the nature of the space; (2) the means of access; and (3) the context of access.[49] Kerr concludes that the nature of the Internet is inherent openness; thus, for purposes of the CFAA, "[t]he authorization line should be deemed crossed only when access is gained by bypassing an authentication requirement . . . . such as a password gate."[50] Accessing data on a publicly available webpage or profile page can never violate the CFAA under this analysis, regardless of expectations of privacy, Terms of Service, or nature of the access. In adopting this approach to access "without authorization," the court reasoned that because hiQ's data scraping bots did not bypass password-protection or encryption, hiQ could not have violated the CFAA.[51]

The next section argues that Kerr's bright line rule to CFAA authorization, and the court's adoption of that rule, is too simplistic. Courts should be given more discretion in determining when access is unauthorized and Congress should amend the statute to account for that judicial discretion.

46. *See id*. at 1110 n.7 (citations omitted).

47. *See id*. at 1111; *see generally* Kerr, *supra* note 7, at 1148–50. For a biography of Professor Orin Kerr, see *Orin Kerr*, USC GOULD SCHOOL OF LAW, http:/À/Àgould.usc.edu/Àfaculty/À?id=73523 (last updated Aug. 30, 2018).

48. *See* Kerr, *supra* note 7, at 1146 ("This Essay offers a framework to distinguish between authorized and unauthorized access to a computer. It argues that concepts of authorization rest on trespass norms.").

49. *See id*. at 1150–53.

50. *See id*. at 1161. Kerr elaborates:

> [C]ourts should adopt presumptively open norms for the Web . . . . Limited efforts to regulate access such as terms of use, hidden addresses, cookies, and IP blocks should be construed as merely speed bumps rather than virtual barriers . . . . The authorization line should be deemed crossed only when access is gained by bypassing an authentication requirement. *Id*.

51. *See hiQ*, 273 F. Supp. 3d at 1113–15.

IV. A NEW STANDARD FOR ACCESS "WITHOUT AUTHORIZATION"

On the one hand, decisions like *hiQ* and scholars like Orin Kerr argue that there should never be a CFAA violation in the context of publicly available data.[52] On the other hand, decisions like *Facebook* suggest that entities owning or hosting data should be able to use the CFAA to shield themselves from unwanted access.[53] Both arguments are valid, but neither allows courts to sufficiently balance concerns about digital privacy and cybersecurity with the risk of giving website hosts total discretion in deciding who gets to access their data.[54]

As the court in *hiQ* accurately points out, "[c]ontext matters."[55] Clearly, hacking into a password-protected computer would constitute accessing that computer "without authorization."[56] In situations that are not outright hacking, Congress should amend the CFAA in a way that allows courts to adopt a rebuttable presumption of authorization when

52. *See id*. at 1112 ("Where a website or computer owner has imposed a password authentication system to regulate access, it makes sense to apply a plain meaning reading of 'access' 'without authorization' such that 'a defendant can run afoul of the CFAA when he or she has no permission to access a computer or when such permission has been revoked explicitly.' But, as noted above, in the context of a publicly viewable web page open to all on the Internet, the 'plainness' of the meaning of 'access' 'without authorization' is less obvious. Context matters."). *See also* Jonathan Mayer, *The "Narrow" Interpretation of the Computer Fraud and Abuse Act: A User Guide for Applying* United States v. Nosal, 84 GEO. WASH. L. REV. 1644, 1654 (2016) ("[T]he inquiry is whether a defendant had authorized access to *any* information or service within the computer system. If the defendant did, then she is not susceptible to *without* authorization liability.").

53. *See*, *e.g.*, *Facebook*, 844 F.3d at 1068 ("Power deliberately disregarded the cease and desist letter and accessed Facebook's computers without authorization to do so. It circumvented IP barriers that further demonstrated that Facebook had rescinded permission for Power to access Facebook's computers. We therefore hold that, after receiving written notification from Facebook on December 1, 2008, Power accessed Facebook's computers 'without authorization' within the meaning of the CFAA and is liable under that statute."); Craigslist Inc. v. 3Taps Inc., 942 F. Supp. 2d 962, 969–70 (N.D. Cal. 2013) ("Assuming that the CFAA encompasses information generally available to the public such as Craigslist's website, Defendants' continued use of Craigslist after the clear statements regarding authorization in the cease and desist letters and the technological measures to block them constitutes unauthorized access under the statute").

54. Giving website hosts too much discretion could allow them to effectively criminalize innocuous activity through Terms of Service or User Agreements. As the court points out in *United States v. Drew*, 259 F.R.D. 449, 464 (C.D. Cal. 2009), "if a website's terms of service controls what is 'authorized' and what is 'exceeding authorization' - which in turn governs whether an individual's accessing information or services on the website is criminal or not, section 1030(a)(2)(C) would be unacceptably vague because it is unclear whether any or all violations of terms of service will render the access unauthorized, or whether only certain ones will."

55. *See hiQ*, 273 F. Supp. 3d at 1112.

56. *See, e.g.*, *supra* note 53.

data is already publicly viewable or available. Those seeking to enforce CFAA liability—whether it be a federal prosecutor or private litigant—should be able to rebut the presumption of authorization to publicly available data by showing that defendant's access was objectively unreasonable. To determine unreasonableness courts should consider (1) the means of access; (2) the validity and enforceability of contractual, technical, and legal countermeasures to prohibit or prevent certain types of access; (3) the merits of access; and (4) the protection of privacy and data integrity. These considerations should be balanced against each other to determine whether access was authorized.

A. Means of Access

As mentioned above, when data is accessed in circumvention of password-protection, encryption, or even a locked door, the circumventor can be said to have accessed that data "without authorization" under the CFAA.[57] On the other hand, simply opening up an Internet browser, searching for an individual on Google, and following a link to his or her LinkedIn profile should not implicate the CFAA.[58] Data scraping falls in between access through Internet browsing and access through the circumvention of an authentication gate.[59] Thus, when an individual or entity employs data scraping as a means of accessing data, courts should use heightened scrutiny in determining whether that access was authorized, regardless of whether or not the data was otherwise publicly available. That does not mean every instance of data scraping is offensive and that data scraping alone should implicate the CFAA; however, courts should not simply throw a case out because the data was otherwise publicly viewable. Data scraping can be very harmful when carried out recklessly or maliciously.[60]

B. Countermeasures to Access

While Terms of Service violations alone generally will not implicate the CFAA,[61] courts have found CFAA violations where an individual or

57. *See, e.g.*, *supra* note 53.

58. *See generally hiQ*, 273 F. Supp. 3d at 1111 (discussing authorization in the context of the "open nature of the Web.")

59. *See supra* Section II.

60. *See e.g.,* Goldman, Eric, *QVC Can't Stop Web Scraping*, FORBES (Mar. 24, 2015, 12:15PM), https:/À/Àwww.forbes.com/Àsites/Àericgoldman/À2015/À03/À24/Àqvc-cant-stop-web-scraping/À#3dbebf6c3ca3 ("Resultly's automated scraper overloaded QVC's servers, causing outages that allegedly cost QVC $2M in revenue.").

61. *See supra* note 51.

entity has employed some kind of "technological gamesmanship" to circumvent a technical block, whether that be an IP block or other technical countermeasures.[62] When an entity like LinkedIn has programmed a number of technical countermeasures aimed at blocking data scraping bots and harmful intrusions, that should trigger heightened judicial scrutiny of the type of access at issue. Organizations like LinkedIn should not have to program their anti-bot countermeasures to protect against some bots, but allow the bots of their competitors and other scrapers to extract as much data as they want. While continued access to data when faced with an IP block or a Cease-and-Desist Letter, without more, should not automatically implicate the CFAA,[63] the fact that they exist should further enhance a court's scrutiny of the access, essentially to then consider the merits of the access and the degree of harm caused by the access.

C. Merits of Access

Courts should next consider the merits of the access with heightened scrutiny. This will naturally require enhanced discretion on the part of the judiciary so as not to punish those who use data scraping, or other means of access, in a safe, socially beneficial way. One significant issue raised by proponents of the approach taken by the court in *hiQ* is that a broad interpretation of the CFAA will have a chilling effect on security researchers and potentially criminalize their work, which entails discovering security flaws in computer networks.[64] Some estimates suggest that data breaches and hacking incidents will cost businesses 8 trillion dollars over the next four years and many of these businesses do not have the in-house resources or personnel to effectively minimize the cybersecurity risks they face.[65] Businesses commonly conduct "bug bounty" programs that look to outside researchers for help in identifying

62. *See Facebook*, 844 F.3d at 1067–68. For a description of additional types of technical countermeasures to prevent data scraping, see Appellant's Opening Brief at 7–8, hiQ Labs, Inc. v. LinkedIn Corp., No.17-1683 (9th Cir. Oct. 3, 2017).

63. *See supra* note 52.

64. *See* Brief of Amici Curiae Electronic Frontier Foundation, Duckduckgo, and Internet Archive In Support Of Plaintiff-Appellee at 21 n.21, hiQ Labs, Inc. v. LinkedIn Corp., No. 17-16783 (9th Cir. Nov. 27, 2017); Rosenblatt, *supra* note 9 ("Although the CFAA has been amended eight times since 1986 to address newer threats, its core remains an obstacle to today's security researchers and coders.").

65. *See* Jason J. Hogg, *Cyber hacks driving 'bug bounty' jobs and programs in corporate America*, FOX BUSINESS (Mar. 7, 2018), https:/À/Àwww.foxbusiness.com/Àfeatures/Àcyber-hacks-driving-bug-bounty-jobs-and-programs-in-corporate-america.

cybersecurity vulnerabilities.[66] This Commentary supports the need for a "security researcher" exception given the increasing vulnerabilities posed by hacking incidents like the Equifax data breach.[67] However, researchers who actually have alternative, malicious motives should not be able to abuse this exception and use it as a shield against CFAA incrimination simply because they call themselves security researchers.

In addition, Google and other search engines commonly use "crawling" bots to index web content for key word search results.[68] The use of crawling bots is distinguishable from data scraping bots and, moreover, web indexing is incredibly crucial to the efficacy and efficiency of search engines.[69] Additional examples of objectively reasonable uses of bots abound[70] and this Commentary does not explore each one, but it is important for courts to dig into the merits of access and determine whether that access promotes a social benefit worth protecting.

D. Privacy Concerns and Maintaining Data Integrity

In viewing access to public data with heightened scrutiny, courts must dig into, not only the merits of access, but also the privacy concerns directly or indirectly related to that access and the affront to data integrity caused by that access. If websites with public information like LinkedIn are not allowed to implement measures to block data scrapers like hiQ, how can they continue to protect the integrity of the data they host? LinkedIn asserts in its Appellate Brief: "Rather than putting in the effort to build its own business, hiQ expropriates member data from LinkedIn's servers on a massive scale, and then turns around and sells that data to companies that wish to furtively monitor their employees."[71] While innovation is certainly a social benefit, courts need to determine whether a business like hiQ's, which is entirely built off of and maintained using scraped data, is worth protecting.

66. *See id*.

67. *See* Gressin, *supra* note 1.

68. *See* Benoit Bernard, *Web Scraping and Crawling Are Perfectly Legal, Right?*, BEN BERNARD BLOG (Apr. 18, 2017), https:/À/Àbenbernardblog.com/Àweb-scraping-and-crawling-are-perfectly-legal-right/À.

69. *See id*.

70. *See generally* Cindy Cohn & Marcia Hofmann, *Rebooting Computer Crime Law Part 2: Protect Tinkerers, Security Researchers, Innovators, and Privacy Seekers*, ELECTRONIC FRONTIER FOUNDATION (Feb. 4, 2013), https:/À/Àwww.eff.org/Àdeeplinks/À2013/À02/Àrebooting-computer-crime-law-part-2-protect-tinkerers-security-researchers.

71. *See* Appellant's Opening Brief at 1, hiQ Labs, Inc. v. LinkedIn Corp., No.17-1683 (9th Cir. Oct. 3, 2017).

Courts must also consider how access impacts the privacy expectations of others associated with the data being accessed. This was a major concern in another CFAA case decided by the Northern District of California, *Craigslist Inc. v. 3Taps Inc.*, where individuals posted various ads for housing on Craigslist and then a third-party company, defendant 3Taps, scraped those ads and republished them on their own site.[72] In some instances, individuals who posted ads and sold or rented a housing unit on Craigslist continued to receive unwanted and unexpected communications and solicitations because their ads were still posted on 3Taps' website without their knowledge.[73] The CFAA should be able to aid in protecting against this type of privacy intrusion.

V. APPLYING THESE STANDARDS TO *HIQ V. LINKEDIN*

Under the factors laid out above, the Ninth Circuit should find that hiQ has violated the CFAA, but that the only appropriate remedies are economic damages and/or injunctive relief. Given the public nature of the data that hiQ has scraped, their authorization to access that data should be presumed.[74] However, in rebutting that presumption, LinkedIn would be able to establish that hiQ's use of data scraping bots, and LinkedIn's prohibition of data scraping in its User Agreement, its implementation of multiple IP blocks, and its issuance of a Cease-and-Desist Letter heighten the court's scrutiny of hiQ's access.[75]

The court would then consider the merits of hiQ's access and find that they do not outweigh LinkedIn's right to protect its data. While innovation is generally socially beneficial, hiQ's business is entirely predicated on the data it scrapes from LinkedIn.[76] The same applies to the *Craigslist* case, where the defendant's business was predicated on scraping and reposting ads from Craigslist.[77] Protective measures implemented by companies like LinkedIn and Craigslist should be legally enforceable. If LinkedIn or Craigslist wants another entity to have access to its data for commercial purposes, they should be able to contract with each other to do so. LinkedIn already has the ability to contract with and sell its data to third party advertisers;[78] they should be allowed, at their

72. *See Craigslist*., 942 F. Supp. 2d at 966–68.
73. *See id*.
74. *See hiQ*, 273 F. Supp. 3d at 1113.
75. *See id.* at 1104.
76. *See id.*
77. *See Craigslist*., 942 F. Supp. 2d at 966–68.
78. *See LinkedIn Marketing Solutions*, LINKEDIN, https:/À/Àbusiness.linkedin.com/Àmarketing-solutions/Àads, (last visited Nov. 24, 2018).

discretion, to do the same with companies like hiQ. Although privacy concerns regarding hiQ's access tip in hiQ's favor because LinkedIn has a similar tool in its Recruiter platform and the timing of LinkedIn's IP blocks and Cease-And-Desist Letter coincided with the launch of that platform,[79] hiQ's access to LinkedIn's data has questionable merits and LinkedIn should have the right to protect itself and maintain the integrity of its data. Thus, given the use of data scraping, LinkedIn's implementation of legal and technical countermeasures, and an analysis that shows the merits of hiQ's access should not outweigh LinkedIn's ability to restrict that access, hiQ should be found to have accessed LinkedIn's data "without authorization" under the CFAA.

An important part of the reasoning in *hiQ* was the court's comparison of digital world trespass to physical world trespass, which the court used to show that hiQ's access to LinkedIn data did not amount to trespass.[80] While an analogy to physical world trespass is helpful in understanding digital trespass, the court's reasoning in coming to its conclusion was faulty. An accurate comparison demonstrates that hiQ's access actually was equivalent to trespass or at least an unwanted intrusion that LinkedIn should be able to protect itself against.

The court in *hiQ* stated that prohibiting hiQ from scraping LinkedIn's data would be like prohibiting a person from viewing a sign on a storefront window that was otherwise viewable by the public on the sidewalk.[81] It is true that a store could not prohibit an individual on the sidewalk from looking at a sign on its storefront window, but hiQ's scraping goes beyond what would be mere viewing of a storefront window. A more suitable analogy between the physical world and data scraping demonstrates that hiQ's activity equates to trespass or an unwarranted intrusion.

Say there are two retailers—Retailer A and Retailer B—selling similar products with stores next to each other. Retailer B is new and really needs insight into pricing models, customer spending habits, customer demographics, and customer affinities in order to remain competitive and attract new business. The problem is that Retailer B is completely new and has no preexisting customer base or industry expertise. Thus, part of its business strategy is to have one of its employees sit in Retailer A's store the entire day, every single day and record the characteristics and profiles of every individual who walks into

79. *See hiQ*, 273 F. Supp. at 1106–07, 1117–18.
80. *See id*. at 1111–13.

81. *See id*. at 1113 n.9.

Retailer A. The employee also records every change in Retailer A's prices and every new product launched. Retailer A has asked Retailer B to cease its practices and Retailer A's customers have no idea that everything the customers do and buy is being recorded by Retailer B. While of course, Retailer A does not want Retailer B to be imprisoned, it does want to protect the integrity of their business and the privacy of their customers. Whether Retailer B's activities constitute a full-fledged trespass is not necessarily clear, but Retailer B's activities are certainly an unwarranted intrusion and Retailer A should be allowed to restrict Retailer B's ability to conduct its clandestine and invasive activities.

V. CONCLUSION

The variety of cases implicating the CFAA reveal how the statute, as written, fails to take into account the nuances of modern Internet usage. This Commentary's proposed standard should be palatable both for those who oppose the court in *hiQ's* argument that the Internet is inherently open and thus the CFAA should effectively only protect against hacking, and those that believe a website owner or host should have total discretion to decide when access is and is not authorized. Hopefully, both sides would recognize that a rebuttable presumption of openness and a relatively high burden of demonstrating that access was objectively unreasonable would be an improvement upon the statute in its current form. Given the complexity of modern Internet usage, courts need a standard that enables a balancing framework within which they can keep in check those who violate digital privacy and cybersecurity interests, but exercise restraint so as not to over-criminalize innocuous activity.