The Legality of Web Scraping: A Proposal

Tess Macapinlac*

TABLE OF CONTENTS

* J.D., May 2019, The George Washington University Law School; B.A., Mathematics, May 2015, Trinity University. Notes Editor, *Federal Communications Law Journal*, Vols. 70-71. Thank you to the staff of the Federal Communications Law Journal and Meredith Rose, FCLJ Journal Adjunct, for their patience, advice, and hard work.

I. INTRODUCTION

When people think of hacking, many may think of people using computers to break into government databases or city records, like in a scene from a television show like *Arrow*.[1] The scene often involves hurried typing, furrowed brows, instant results, and often, very few punishments for hacking. Hacking, which may feel like a modern innovation due to continual improvements in technology, has been infused into pop culture for years. Certainly, newer television shows like *Mr. Robot*, *Scorpion*, and *Blacklist* show hacking in a variety of lights and taking place in a variety of circumstances.[2] Even in the 1997 movie *Independence Day*, a satellite technician saves the world by hacking into an alien mothership.[3] Prior to that, *Jurassic Park* showed a juvenile hacker taking on a UNIX system to reactivate security measures in a dinosaur park gone berserk.[4]

Others may think of hacking as they see it in the news. For instance, the infamous Ashley Madison hack revealed the identities and contact information of the site's users, who frequented the Ashley Madison website with the intention of having discreet extramarital affairs.[5] The Home Depot hack is another infamous incident, where the credit card numbers of almost fifty million customers were revealed.[6] Reports show that over five thousand breaches occurred in 2017, compromising almost eight billion records.[7]

This image of hackers sitting in a dark room, bent over computers, furiously typing complicated computer code is the image that many people tend to associate with the term hacking.[8] Personal information revealed, secrets unleashed, and access to information a person was never supposed to

---

1. *See* Jessica Conditt, *High-Tech TV: How Realistic Is the Hacking in Prime-Time TV Shows?*, ENGADGET (Apr. 6, 2015), https://www.engadget.com/2015/04/06/high-tech-tv/ [https://perma.cc/URC6-MYLJ].

2. *See* Forbes Technology Council, *'Hackers' on TV: Popular Shows That Get Technology Right*, FORBES (Oct. 26, 2016, 9:00 AM), https://www.forbes.com/sites/forbestechcouncil/2016/10/27/hackers-on-tv-popular-shows-that-get-it-right/#7570307e6f05 [https://perma.cc/PFM7-3W54].

3. INDEPENDENCE DAY (Twentieth Century Fox 1996).

4. JURASSIC PARK (Universal Pictures 1993).

5. *See* Robert Hackett, *What to Know About the Ashley Madison Hack*, FORTUNE (Aug. 26, 2015), http://fortune.com/2015/08/26/ashley-madison-hack/ [https://perma.cc/92TD-VLGT].

6. Paul Gil, *The Greatest Computer Hacks*, LIFEWIRE, https://www.lifewire.com/the-greatest-computer-hacks-4060530 [https://perma.cc/CX9A-SJZ4] (last updated Nov. 8, 2017).

7. Daniel Solove, *Data Security Is Worsening:2017 Was the Worst Year Yet*, TEACHPRIVACY: PRIVACY + SECURITY BLOG (Feb. 16, 2018), https://teachprivacy.com/data-security-is-worsening-2017-was-the-worst-year-yet/ [https://perma.cc/7NMH-EDXT].

8. *See* Forbes Technology Council, *supra* note 2.

have are all ideas equally associated with hacking.[9] With so much personal data given over to companies and held in electronic formats,[10] people are right to be concerned with hackers and the damage they can do.

However, a lot of this hacking rides on the idea of secrecy. Whether it is information that is given to a company with the condition of confidentiality or unknown information relating to the computer system on an alien spaceship, hacking relies on the idea that the hacker isn't supposed to know or be able to get the information that they are taking.[11] Thus, the idea of hacking publicly available information does not fit into either of these categories. Companies promise to do their best to keep consumer information safe and private.[12] However, if certain information is public, then by definition, all people should have access, and none of it should be a secret. Regardless, a practice known as web scraping is considered hacking under the Computer Fraud and Abuse Act ("CFAA"), codified at 18 U.S.C. § 1030 (2012).[13] Web scraping is the act of pulling data from a website's output and saving it to a file or database.[14]

This Note focuses on the scraping of publicly available information and how this particular act should not be considered illegal under the CFAA. First, this Note will more thoroughly explore the technicalities and benefits of web scraping, as well as the relevant sections of the CFAA. Next, it will examine some of the prominent cases that have used the CFAA to prosecute web scraping. It will then examine why web scraping should not be punishable under the CFAA. It will go on to present a proposed amendment and the thought process that went into its language. Finally, it will consider the potential supporters and opponents of the proposed amendment. At its core, this Note is a proposal to add an amendment to the CFAA that would legalize the web scraping of publicly available websites.

---

9. *See Data Breaches 101: How They Happen, What Gets Stolen, and Where It All Goes*, TREND MICRO (Aug. 10, 2018), https://www.trendmicro.com/vinfo/us/security/news/cyber-attacks/data-breach-101 [https://perma.cc/3TEA-CXRT] (listing the types of information that have been involved in hacks throughout history).

10. *See id.*

11. *See id.* (explaining that hacking occurs when a cybercriminal looks for weaknesses in a company's security system, attacks the network, and extracts information).

12. *See*, *e.g.*, *Making Technology for Everyone Means Protecting Everyone Who Uses It*, GOOGLE, https://safety.google/ [https://perma.cc/L2ZQ-QYT7] ("We protect you online with industry-leading security."); *U.S. Online Privacy Statement*, AMERICAN EXPRESS, https://www.americanexpress.com/us/legal-disclosures/online-privacy-statement.html [https://perma.cc/S637-2YXF] ("At American Express, we are committed to safeguarding your privacy."); *Facebook's Commitment to Data Protection and Privacy in Compliance with the GDPR*, FACEBOOK, https://www.facebook.com/business/news/facebooks-commitment-to-data-protection-and-privacy-in-compliance-with-the-gdpr [https://perma.cc/V7X5-RXDJ] ("Facebook takes data protection and people's privacy very seriously . . . ").

13. 18 U.S.C. § 1030 (2012).

14. *See* Geoff Boeing & Paul Waddell, *New Insights into Rental Housing Markets Across the United States: Web Scraping and Analyzing Craigslist Rental Listings*, 37 J. PLAN. EDUC. & RES. 457, 459 (2016).

II. A BASIC UNDERSTANDING OF WEB SCRAPING AND THE CFAA

A. Web Scraping

The method in question is known as web scraping. A web scraper is a piece of computer code that translates into an automated bot.[15] This bot then accesses web pages, finds specific data, extracts it from the web page, and saves it on a computer or similar device.[16] A person can then access the data and use it for a variety of purposes, such as in research or business.[17] Web scraping is useful for anyone who needs a large amount of information from a large number of websites; while everything this kind of bot does can be done manually, the work is done faster and more efficiently by utilizing web scraping.[18] Web scraping is not an uncommon practice, as bots account for nearly a quarter of all Internet traffic, due to businesses, researchers, and others using web scraping for different reasons.[19]

Web scraping can be used for a variety of purposes. One of the most common examples is search engines, which use scraping to link users to pertinent webpages.[20] Since search engines play an important role in the online ecosystems for both users and companies alike, the stigma associated with search engine web scraping activities is situational and limited.[21] Academia is another field that may utilize web scraping. For instance, Geoff Boeing and Paul Waddell built their own web scraper to scrape data for their paper concerning rental housing markets.[22]

Another common form of web scraping takes place on budgeting apps, like Mint.[23] In order to use Mint, a user uploads authorization to access their different bank accounts.[24] The app then scrapes the account information so that users can track their budgeting and spending habits from a single app.[25]

---

15. *Id.*
16. *Id.*
17. *Id.*
18. *Id.*
19. *See* Tiffany Hu & Aaron Rubin, *Data for the Taking: Using the Computer Fraud and Abuse Act to Combat Web Scraping*, SOCIALLY AWARE (July 21, 2014), https://www.sociallyawareblog.com/2014/07/21/data-for-the-taking-using-the-cfaa-to-combat-web-scraping/ [https://perma.cc/X4DW-C2E2].
20. Jeffrey Kenneth Hirshey, *Symbiotic Relationships: Pragmatic Acceptance of Data Scraping*, 29 BERKLEY TECH. L.J. 897, 898 (2014).
21. *See id.*
22. Boeing & Waddell, *supra* note 14.
23. Mary Wisniewski, *Is It Time to End Screen Scraping?*, AM. BANKER (Nov. 7, 2014, 3:55 PM), https://www.americanbanker.com/news/is-it-time-to-end-screen-scraping [https://perma.cc/7YNH-LNL8].
24. *See How It Works*, MINT, https://www.mint.com/how-mint-works/security [https://perma.cc/6WXZ-ZCNT] ("We use [your login user names and passwords] to establish a secure connection with your financial institution or credit card company. This enables us to download and categorize your transaction information securely and automatically.").
25. *See* Wisniewski, *supra* note 23.

Even journalists utilize web scraping for a variety of online investigations, to the point where both lawyers and journalists make suggestions on how to do so in an ethical and legal fashion.[26] Journalist Nael Shiab points to his own career as an example of web scraping for journalistic purposes.[27] However, many cases involving web scraping and the CFAA focus on businesses that use web scraping as a part of their business models, rather than search engines or academia.[28]

B. The Computer Fraud and Abuse Act

In 1984, Congress passed the Comprehensive Crime Control Act, codified in 18 U.S.C. §1030, in order to combat the growing threat of computer crime and hacking.[29] Over the next two years, Congress continued to investigate issues presented by computer crimes and how federal statutes could tackle such crimes.[30] In order to address such issues, Congress held hearings on potential bills focused on computer crimes, which, in 1986, culminated in Congress passing the Computer Fraud and Abuse Act, which amended various parts of 18 U.S.C. §1030.[31]

At the time the CFAA was brought into effect, the government and various financial institutions were the primary entities that used computers and thus were most vulnerable to hackers.[32] As such, the CFAA was designed with classified information and credit or financial information in mind.[33] Eventually, as computers and the Internet became widely used by civilians, definitions in the CFAA were expanded to cover computers that could be involved in interstate commerce, which implicated any computer connected to the Internet.[34]

---

26. *See generally* Rachel Goodman, *Tips for Data Journalism in the Shadow of an Overbroad Anti-Hacking Law*, AM. CIVIL LIBERTIES UNION (Oct. 13, 2017, 1:00 PM), https://www.aclu.org/blog/privacy-technology/internet-privacy/tips-data-journalism-shadow-overbroad-anti-hacking-law [https://perma.cc/4T37-ETKE] (providing advice for journalists who want to avoid liability when webscraping); *see generally* Nael Shiab, *On the Ethics of Web Scraping and Data Journalism*, GLOBAL INVESTIGATIVE JOURNALISM NETWORK (Aug. 12, 2015), https://gijn.org/2015/08/12/on-the-ethics-of-web-scraping-and-data-journalism/ [https://perma.cc/6WFM-KJXS] (information about web scraping and ethics).

27. Nael Shiab, *Web Scraping: A Journalist's Guide*, GLOBAL INVESTIGATIVE JOURNALISM NETWORK (Aug. 11, 2015), https://gijn.org/2015/08/11/web-scraping-a-journalists-guide/ [https://perma.cc/6WFM-KJXS]] ("For example, I created [a web scraper] to compare the alcohol prices between Quebec and Ontario.").

28. *See* hiQ Labs, Inc. v. LinkedIn, Inc., 273 F. Supp. 3d 1099 (N.D. Cal. 2017) (order granting preliminary injunction); Craigslist Inc. v. 3Taps, 964 F. Supp. 2d 1178 (N.D. Cal. Aug. 16, 2013); Cvent, Inc. v. Eventbrite, Inc., 739 F. Supp. 2d 927 (E.D. Va. Sept. 15, 2010).

29. *See* 18 U.S.C. § 1030; H. MARSHALL JARRETT ET. AL., PROSECUTING COMPUTER CRIMES 1 (2d ed. 2015).

30. JARRETT ET. AL., *supra* note 29.

31. *Id.*

32. *See* 18 U.S.C. § 1030(a)(2)(A); 18 U.S.C. § 1030(a)(2)(B); *hiQ Labs, Inc.*, 273 F. Supp. 3d at 1109 (order granting preliminary injunction); JARRETT ET. AL., *supra* note 29, at 1.

33. *See* 18 U.S.C. § 1030(a)(2)(A); 18 U.S.C. § 1030(a)(2)(B); JARRETT ET. AL., *supra* note 29, at 1.

34. JARRETT ET. AL., *supra* note 29, at 2.

In the context of the CFAA, hacking occurs when a person "intentionally accesses a computer without authorization or exceeds authorized access."[35] In this case, the relevant hacking occurs when a person accesses a "protected computer."[36] For these purposes, a "protected computer" is defined to include a computer "which is used in or affecting interstate or foreign commerce or communication."[37] Notably, the user does not have to use the computer for interstate or foreign commerce or communication; rather, the computer simply has to be capable of doing so.[38] Thus, any computer connected to the Internet could be considered a protected computer under the CFAA.[39]

The CFAA defines "exceeds authorized access" to mean "to access a computer with authorization and to use such access to obtain or alter information in the computer that the accesser is not entitled to so obtain or alter."[40] The CFAA does not define the term "without authorization," but experts interpret the term as referring to a person who is an outsider of the institution, like a hacker, as opposed to an insider, who would have access in the first place.[41]

The CFAA has long been criticized for its hefty punishments and vague definitions.[42] The CFAA's vague definitions can make a simple act like lying about your age online fall under the definition of hacking.[43] Additionally, a single act can violate different parts of the CFAA, resulting in a compounded sentence for a single act.[44] In fact, some CFAA violations carry more severe punishments than an aggravated assault charge.[45]

The criticism surrounding the CFAA reached a high point during the criminal case against Aaron Swartz. Swartz, known for helping to launch Reddit, broke into an electrical closet at the Massachusetts Institute of Technology, wired his laptop into MIT's system, and proceeded to download academic articles from the online database JSTOR.[46] In the District Court for the District of Massachusetts, Swartz was tried for eleven violations of the CFAA,[47] as well as wire fraud, which could have led to thirty-five years in

---

35. 18 U.S.C. § 1030(a)(2).
36. *Id.* § 1030(a)(2)(C).
37. *Id.* § 1030(e)(2)(B).
38. JARRETT ET. AL., *supra* note 29, at 4.
39. *Id.*
40. 18 U.S.C. § 1030(e)(6).
41. JARRETT ET. AL., *supra* note 29, at 5.
42. *See* Seth Rosenblatt, *Where Did The CFAA Come From, and Where is It Going?*, THE PARALLAX (Mar. 16, 2016), https://www.the-parallax.com/2016/03/16/where-did-the-cfaa-come-from-and-where-is-it-going/ [https://perma.cc/A2AT-7HHH].
43. *Id.*
44. *See* Sam Gustin, *Aaron Swartz's Father Praises 'Aaron's Law' Proposal*, TIME (June 27, 2013), http://business.time.com/2013/06/27/aaron-swartzs-father-praises-aarons-law-proposal/ [https://perma.cc/WQL2-LYNK].
45. Rosenblatt, *supra* note 42.
46. Nick Bilton, *Internet Activist Charged in M.I.T. Data Theft*, N.Y. TIMES: BITS (July 19, 2011, 12:54 PM), https://bits.blogs.nytimes.com/2011/07/19/reddit-co-founder-charged-with-data-theft/?_r=0 [https://perma.cc/Y7DN-HNUV].
47. Rosenblatt, *supra* note 42.

prison and a million dollar fine.[48] However, the charges were never resolved, as Swartz took his life prior to any resolution.[49] Several House representatives accused the prosecution of "act[ing] too aggressively" in their suit against Swartz.[50]

The extreme charges and tragic ending in Swartz's case resulted in Aaron's Law, a bill introduced by Rep. Zoe Lofgren of California and Senator Ron Wyden of Oregon in 2013.[51] This bill focused on three revisions of the CFAA: (1) a violation of terms of service could not be prosecuted under the CFAA, (2) eliminating redundancies from various sections of the CFAA, and (3) rewriting CFAA penalties to be proportionate to the crime.[52] This bill received praise from various entities, such as the Electronic Frontier Foundation.[53] However, the bill was stalled in Congress and never passed.[54] The case against Swartz remains a hallmark CFAA case.[55]

Several federal courts have ruled on cases involving the CFAA, with particular focus on authorization.[56] While the following cases have not involved web scraping, they show courts' evolving opinions on authorization.[57] This examination begins in the Northern District of California, with the case *United States. v. Nosal* ("*Nosal I*"). David Nosal was a high-level executive at a renowned recruitment firm, Korn/Ferry International ("KFI").[58] Nosal left KFI to open a competing firm.[59] However, two people at KFI helped set up Nosal's competing firm by using their employee credentials to obtain trade secrets and other valuable information

---

48. Ryan Singel, *Feds Charge Activist as Hacker for Downloading Millions of Academic Articles*, WIRED (July 19, 2011, 2:55 PM), https://www.wired.com/2011/07/swartz-arrest/ [https://perma.cc/8CEH-B8WK].

49. Brendan Sasso & Jennifer Martinez, *Lawmakers Slam DOJ Prosecution of Swartz as 'Ridiculous, Absurd'*, THE HILL (Jan. 15, 2013, 10:52 PM), http://thehill.com/policy/technology/277353-lawmakers-blast-trumped-up-doj-prosecution-of-internet-activist [https://perma.cc/2VMC-4M35].

50. *Id.*

51. Aaron's Law Act of 2013, H.R. 2454, 113th Cong. (2013); *see* Gustin, *supra* note 44.

52. Aaron's Law Act of 2013, H.R. 2454, 113th Cong. (2013); Aaron's Law Act of 2013, S. 1196, 113th Cong. (2013).

53. Cindy Cohn, *Aaron's Law Reintroduced: CFAA Didn't Fix Itself*, ELEC. FRONTIER FOUND. (Apr. 29, 2015), https://www.eff.org/deeplinks/2015/04/aarons-law-reintroduced-cfaa-didnt-fix-itself?page=12 [https://perma.cc/R3WV-GCTN].

54. Ruth Reader, *3 Years After Aaron Swartz's Death, Here's What Happened to Aaron's Law*, MIC (Jan. 11, 2016), https://mic.com/articles/132299/3-years-after-aaron-swartz-s-death-here-s-what-s-happened-to-aaron-s-law#.XhPbkciIR [https://perma.cc/BKU8-CG7P].

55. Kim Zetter, *The Most Controversial Hacking Cases of the Past Decade*, WIRED (Oct. 26, 2015, 7:00 AM), https://www.wired.com/2015/10/cfaa-computer-fraud-abuse-act-most-controversial-computer-hacking-cases/ [https://perma.cc/B2HG-7NR4].

56. *See* LVRC Holdings LLC v. Brekka, 581 F.3d 1127 (9th Cir. 2009); United States v. Nosal, No. CR 08-00237 MHP, 2010 WL 934257 (N.D. Cal. Jan. 6, 2010) [hereinafter *Nosal II*]; United States v. Nosal, NO. CR 08-00237 MHP, 2009 WL 981336 (N.D. Cal. Apr. 13, 2009) [hereinafter *Nosal I*].

57. *See LVRC Holdings LLC*, 581 F.3d 1127; *Nosal II*, 2010 WL 934257; *Nosal I*, 2009 WL 981336.

58. *Nosal I*, 2009 WL 981336, at *1.

59. *Id.*

and send them to Nosal.[60] Nosal, along with one of the employees who helped him, were indicted on several counts, including eight violations of the CFAA.[61]

Nosal moved to dismiss the CFAA charges based on the argument that the CFAA's discussion of the terms "without authorization" and "exceeds authorized access" does not cover misuse or misappropriation of information that was lawfully obtained.[62] The court found this argument unconvincing, noting that the judicial view of "authorization" under the CFAA was expanding and that because the case was still in its pleading stage, the evidence presented by the government was sufficient to allow the charges.[63] The motion to dismiss the CFAA charges was denied.[64]

After this came the case *LVRC Holdings LLC v. Brekka*.[65] Christopher Brekka, an employee of LVRC Holdings ("LVRC"), emailed a number of LVRC documents from his LVRC computer to his and his wife's personal email accounts.[66] Brekka later stopped working for LVRC, but because he had emailed the documents to personal email accounts, the documents remained on his computer.[67] He also continued to access personal documents using his former employee credentials.[68] LVRC filed suit, claiming that Brekka had violated the CFAA both when he emailed documents to himself and when he accessed LVRC records after his employment there ceased.[69] The district court granted summary judgment for Brekka, and LVRC appealed the decision.[70] The Ninth Circuit Court of Appeals interpreted "without authorization" and "exceeds authorized access" more narrowly than the court in *Nosal I* and affirmed the decision of the district court.[71]

After the appellate decision from the *Brekka* case was announced, Nosal filed another motion for dismissal, claiming that the court's earlier denial to dismiss the CFAA charges was now contradictory to its decision in *Brekka*.[72] In this case ("*Nosal II*"), the court examined the definition of "exceeds authorized access" and decided that the CFAA phrase "obtain or alter information in the computer that the accesser is not entitled so to obtain or alter" cannot be expanded to cover a corporation's policies regarding

---

60. *Id.*
61. *Id.* at *1-2.
62. *Id.* at *4.
63. *Id.* at *6-7.
64. *Id.* at *7.
65. *LVRC Holdings LLC*, 581 F.3d 1127.
66. *Id.* at 1129-30.
67. *Id.* at 1130.
68. *Id.*
69. *Id.*
70. *Id.*
71. *See id.* at 1133-35.
72. *See* United States v. Nosal, No. CR 08-00237 MHP, 2010 WL 934257, at *1 (N.D. Cal. Jan. 6, 2010) [hereinafter *Nosal II*].

misappropriation of information.[73] Thus, the court dismissed five of the CFAA charges.[74]

The government appealed *Nosal II*.[75] The court found that the government's proposed interpretation of "exceeds authorized access" would "transform the CFAA from an anti-hacking statute into an expansive misappropriation statute."[76] The court found the district court's narrow interpretation more convincing and affirmed.[77] These are just a few of the cases that have spoken on the meaning of authority in the CFAA, and they demonstrate the confusion behind the language of the CFAA.

III. WEB SCRAPING CASES

Several cases that have come before federal courts demonstrate these courts' attempts to determine what the CFAA says regarding hacking and web scraping. In the case *Craigslist v. 3Taps*, a company called 3Taps scraped Craigslist pages in order to aggregate and republish Craigslist's ads.[78] Despite Craigslist's various efforts to stop 3Taps from doing so, 3Taps continued scraping the website, and Craigslist filed suit against 3Taps, claiming that this was a form of hacking under the CFAA.[79]

One course of action that Craigslist took was to block all IP addresses associated with 3Taps, so that no one using those IP addresses could access Craigslist.[80] 3Taps was able to bypass that barrier through technological means in order access Craigslist.[81] The United States District Court for the Northern District of California found that because 3Taps was banned from accessing the site and then maneuvered around that ban, the web scraping that followed the ban was hacking under the CFAA.[82] 3Taps' access had been revoked, and so further scraping constituted unlawful access.[83] Without that ban and barrier, 3Taps' actions may not have been considered hacking under the CFAA.[84]

In the case *Cvent, Inc. v. Eventbrite, Inc.*, a company called Eventbrite[85] scraped the Cvent website to get information about events, which Eventbrite then made available on its own site, as a competitor of Cvent.[86] However, the

---

73. *Id.* at *7.
74. *Id.* at *8 .
75. United States v. Nosal, 676 F.3d 854, 856 (9th Cir. 2012) [hereinafter *Nosal III*].
76. *Id.* at 857.
77. *Id.* at 864.
78. Craigslist Inc. v. 3Taps, 964 F. Supp. 2d 1178, 1180 (N.D. Cal. Aug. 16, 2013).
79. *Id.* at 1180-81.
80. *Id.* at 1181.
81. *Id.*
82. *Id.* at 1185.
83. *Id.*
84. *See id.*
85. Cvent, Inc. v. Eventbrite, Inc., 739 F. Supp. 2d 927, 930 (E.D. Va. Sept. 15, 2010) (Eventbrite "maintains an online event planning, sales, and registration service hosted on its website, www.eventbrite.com.").
86. *Cvent, Inc.*, 739 F. Supp. 2d at 930.

District Court for the Eastern District of Virginia found that this web scraping did not constitute hacking, as the Cvent website was accessible "without requiring any login, password, or other individualized grant of access."[87]

Finally, the *hiQ Labs, Inc. v. LinkedIn, Inc.* case is currently pending in the District Court for the Northern District of California[88] and may have great bearing on this issue. A company called hiQ scraped public profiles on LinkedIn, a networking website.[89] hiQ provided employers with aggregated information from these public profiles and indicated what skills employees had, as well as which employees could potentially be recruited to work elsewhere.[90]

LinkedIn sent hiQ a cease-and-desist letter, stating that hiQ's access to LinkedIn had been restricted and that further scraping could result in a suit under the CFAA.[91] In response, hiQ filed suit seeking an affirmative declaratory ruling of rights regarding its ability to access publicly available LinkedIn profiles.[92] hiQ also filed a request for an order to allow it to continue the web scraping while the suit proceeded.[93] The court allowed hiQ to continue web scraping while the suit proceeded, as it found that the public nature of the profiles made the CFAA's relevance to the case questionable.[94] Additionally, the court weighed the need for hiQ to continue to operate while the suit was ongoing.[95] Because scraping LinkedIn profiles was such an integral part of the operation, the court allowed hiQ to continue its practice for the duration of the suit.[96]

These cases represent a small sample of web scraping suits brought under the CFAA. They show that the legislative world is ready for the law to make a firm decision on web scraping and its relationship to the CFAA.

IV. WEB SCRAPING SHOULD NOT BE UNDER CFAA JURISDICTION

A. Public Information Is Not Subject to Hacking

First, public information, such as profiles that are freely accessible to the public, cannot be "hacked" in any traditional sense and thus should not be under the CFAA's jurisdiction. Looking to the legislative history, the court in *hiQ* pointed out that the CFAA was not created to police access to publicly

87. *Id.* at 932-34.
88. hiQ Labs, Inc. v. LinkedIn, Inc., 273 F. Supp. 3d 1099 (N.D. Cal. 2017) (order granting preliminary injunction).
89. *See id.* at 1104.
90. *Id.*
91. *Id.*
92. *Id.*
93. *See id.*
94. *Id.* at 1120.
95. *Id.* at 1105.
96. *Id.* at 1120.

available websites on the Internet, as the Internet did not exist yet.[97] Rather, it was meant to prevent hackers from getting into "private, often password-protected mainframe computers."[98] Indeed, aside from protected computers, as defined earlier, the CFAA points directly to computers related to financial institutions and departments or agencies of the United States government.[99] The people who wrote these laws could not have fathomed the widespread nature of computers and the Internet today. The CFAA needs to be brought into today's world, with all of its modern technologies.

Others compare the CFAA to a trespass statute that applies the Internet, instead of the physical world. In his article *Norms of Computer Trespass*, Orin Kerr states it succinctly: "Unauthorized access statutes are computer trespass statutes."[100] He points out that the Internet and the websites on it are the equivalent of an "open public square in the physical world."[101] In the physical world, a locked door with a limited number of keys would be akin to a website that requires a password or other authorization for access.[102] But a public website, being akin to a public square, is open for anyone to enter and observe and learn as they will.[103] There is no need for a key or additional authorization to witness the happenings in the public square.[104]

Kerr's analysis, cited in the court's order in the *hiQ* case, points towards a need for clarification on this topic.[105] Kerr's public square[106] versus locked door argument[107] echoes the sentiment that the CFAA should not cover information that is publicly available on the Internet. Clarification of the CFAA could also include language that would allow the amendment to fall in line with current case law. For instance, if the CFAA were amended to fall in line with Kerr's thinking, then the final verdict in the *Cvent* case would be preserved—the web scraping would not be considered hacking under the CFAA because Cvent's site did not require a login or password for access.[108]

A change to the CFAA's language could also prevent, in Kerr's hypothetical,[109] the potential legality of a person stealing a key, using that key to enter a private home, and learn and observe what is going on in that home at will. In the online world, this would be equivalent to a hacker stealing a password and using that password to access an otherwise inaccessible

---

97. *Id.* at 1109.
98. *Id.*
99. *See* 18 U.S.C. § 1030(a)(2)(A); 18 U.S.C. § 1030(a)(2)(B).
100. Orin Kerr, *Norms of Trespass*, 116 COLUM. L. REV. 1143, 1153 (2016).
101. *Id.* at 1163.
102. *Id.* at 1153.
103. *Id.* at 1163.
104. *See id.* at 1163 ("A person who connects a web server to the Internet agrees to let everyone access the computer much like one who sells his wares at a public fair agrees to let everyone see what is for sale.").
105. hiQ Labs, Inc. v. LinkedIn, Inc., 273 F. Supp. 3d 1099, 1112 (N.D. Cal. 2017).
106. Kerr, *supra* note 100, at 1163.
107. *Id.* at 1153.
108. *See* Cvent, Inc. v. Eventbrite, Inc., 739 F. Supp. 2d 927, 932 (E.D. Va. Sept. 15, 2010).
109. Kerr, *supra* note 100, at 1153.

website. With careful writing of new CFAA language, this action would not be legal, which fits well with Kerr's parallel of the public Internet as the online version of "an open public square in the physical world."[110]

The *hiQ* order fits very well with the concept of publicly available information not being subject to the CFAA. Notably, hiQ only scraped LinkedIn profiles that were public.[111] As the court in *hiQ* notes, LinkedIn users can limit the people to whom a profile is made public.[112] For instance, a profile can be made public to direct connections on LinkedIn, all LinkedIn members, or the entire public.[113] hiQ did not scrape private profiles; it only scraped data from users who chose to make their information public to the entire Internet.[114] To use Kerr's analogy, hiQ just spent time in the open public square.[115] hiQ did not try to get into any locked doors to get its information.[116]

The CFAA needs to be changed in order to bring the law up to speed with some aspects of the Internet. As people put more of their information and lives on the Internet, it is only natural for people and companies to begin raising privacy concerns. Changes to the CFAA would help both people and companies address those privacy concerns. Companies like hiQ would have a bright-line rule to follow for web scraping practices and business models. Companies like LinkedIn could use that same bright-line rule to educate their consumers about their privacy options and the implications of public information and make technical adjustments to their websites to protect consumer privacy. Consumers, in turn, could better understand the privacy implications of their online information and make better-informed decisions for themselves.

B. Web Scraping Allows Competition in the Marketplace

Web scraping allows smaller companies to compete with other players in the online marketplace. In a world where technology moves fast and information moves faster, it seems that whoever works the most efficiently has the advantage.[117] Scraping allows people to automate tedious work and spend their time on other pursuits.[118] This speed may give businesses an opportunity to efficiently compete with the leaders in their respective fields.[119] This argument helped carry the day in the court's analysis of the *hiQ* order.[120]

---

110. *Id.* at 1163.
111. hiQ Labs, Inc. v. LinkedIn, Inc., 273 F. Supp. 3d 1099, 1104 (N.D. Cal. 2017).
112. *Id.*
113. *Id.*
114. *Id.*
115. Kerr, *supra* note 100, at 1163.
116. *See generally id.* at 1163 (explaining the locked door analogy).
117. Kimberly Amadeo, *What Is Competitive Advantage? Three Strategies that Work*, THE BALANCE, https://www.thebalance.com/what-is-competitive-advantage-3-strategies-that-work-3305828 [https://perma.cc/HV44-GU3K] (last updated Mar. 19, 2019).
118. Boeing & Waddell, *supra* note 14, at 459.
119. Amadeo, *supra* note 117.
120. hiQ Labs, Inc. v. LinkedIn, Inc., 273 F. Supp. 3d 1099, 1118 (N.D. Cal. 2017).

The court determined that there are serious concerns that LinkedIn, as the main entity in the online professional networking market, may be using its significant weight for anticompetitive purposes.[121]

Clarifying the CFAA could provide a solid compromise between LinkedIn and hiQ. If additional language were phrased correctly, then under the CFAA, hiQ could legally scrape the profiles that users have actively chosen to make public to all users on the Internet. LinkedIn could educate users to what a completely public profile might mean, particularly in regard to companies like hiQ. By educating consumers about this aspect of their profiles, LinkedIn may be able to encourage, in a less anticompetitive fashion, users to keep their profiles public only to those within the LinkedIn network. By giving consumers a fuller picture of the online professional marketing landscape, LinkedIn could help users more fully understand who views their profiles and has access to their information.

However, the *Craigslist* case could present serious issues for such changes to the CFAA, as well as its potential interpretations. The court in the *Craigslist* case found that 3Taps' measures to get around the IP block made the ensuing web scraping constitute hacking under the CFAA.[122] This IP address block was a part of the ban that Craigslist enacted against 3Taps.[123] Depending on the language of the CFAA clarification, such bans may be permissible, particularly if the CFAA is clarified in line with Kerr's trespass analysis.[124] Under Kerr's analysis, passwords, logins, and other similar access or authorization mechanisms are acceptable methods of preventing access.[125]

A ban may not be an acceptable method of preventing access. The problem with a ban is that it could hinder competition, leaving the main entity in the industry with power and providing smaller companies with few avenues to remedy the situation. Companies could potentially ban rivals from visiting their site. For instance, Kerr uses the example of a news website sending letters to reporters for other news agencies to stop viewing their website.[126] By leaving open the option of banning some entities from viewing their websites, companies would not only place themselves in the crosshairs of antitrust suits, they could also create a gray area in the definition of "public."

Kerr argues that if companies want to truly limit the people who view their websites, they should actively use a password for all viewers to clearly state that the information on the sites is not, in fact, public information.[127] While this would clear up the idea of public versus private information in the CFAA's language, this may not be the most feasible option for companies.

---

121. *Id.*

122. Craigslist Inc. v. 3Taps, 964 F. Supp. 2d 1178, 1184-86 (N.D. Cal. Aug. 16, 2013).

123. *Id.* at 1180-81.

124. *See generally* Kerr, *supra* note 100, at 1153-61.

125. *See id.* at 1171-73.

126. Timothy B. Lee, *LinkedIn: It's Illegal to Scrape Our Website Without Permission*, ARS TECHNICA (July 31, 2017, 8:00 AM), https://arstechnica.com/tech-policy/2017/07/linkedin-its-illegal-to-scrape-our-website-without-permission/ [https://perma.cc/7UKL-7RFH].

127. *Id.*

However, banning companies from utilizing what is, to all other Internet users, completely public information is anti-competitiveness at its finest. So, while changes to the CFAA could have great implications for marketplace competition, the changes must be written to prevent any loopholes that would allow companies to get around the intent of this new language by introducing bans against all of their competitors.

Another potential obstacle is the common industry practice known as robots.txt. This is a text file embedded in a website's directory that instructs scraping bots on how to scrape available pages.[128] Search engines utilize web scraping to gather information that will show up on the search engine page.[129] A robots.txt notes which pages a bot can and cannot scrape.[130] Web scraping bots from other parties would also run into the robots.txt and be held to the same restrictions as web scrapers from search engines.[131]

However, it is unlikely that companies would use the robots.txt to block web scrapers focused on business from accessing pages. Presumably, a company would want web scraping bots from search engines to have access to public pages in order to increase the number of pages appearing in various searches and by extension the likelihood of a person using a search engine clicking on those web pages.

It seems doubtful that companies would sacrifice potential search engine hits to hinder another company from scraping pages for business or research purposes. The CFAA could likely be amended to include language that makes the robots.txt permissible, because all scraping bots, whether they come from search engines or competing companies, are not allowed to see certain pages.[132]

C. Proportionality, Vagueness, and Alternative Legal Avenues

The CFAA has been criticized for enacting punishments that are not proportional to the crime at hand.[133] This disproportionate punishment stems from the CFAA's redundancies.[134] A single crime can violate multiple sections of the CFAA, resulting in hefty prison sentences and fines.[135] A common example of this is the Swartz case, where his violations of the CFAA could have resulted in thirty-five years in jail and a million dollar fine.[136] The CFAA allows people or entities to file suit if they have suffered damage or

---

128. Stephan Spencer, *A Deeper Look at Robots.txt*, SEARCH ENGINE Land (Apr. 16, 2009, 8:00 AM), https://searchengineland.com/a-deeper-look-at-robotstxt-17573 [https://perma.cc/Q8Y5-2HD2].

129. *See* Hirshey, *supra* note 20, at 898.

130. *See* Spencer, *supra* note 128.

131. *See id.*

132. *See* Spencer, *supra* note 128.

133. *See* Gustin, *supra* note 44.

134. Zoe Lofgren & Ron Wyden, *Introducing Aaron's Law, A Desperately Needed Reform of the Computer Fraud and Abuse Act*, WIRED (June 20, 2013, 9:30AM), https://www.wired.com/2013/06/aarons-law-is-finally-here/ [https://perma.cc/6EFN-3BC6].

135. *Id.*

136. Singel, *supra* note 48.

loss by violation of the CFAA.[137] However, the CFAA defines "loss" to mean "any reasonable cost to any victim, including . . . restoring the data, program, system, or information to its condition prior to the offense, and any revenue lost, cost incurred, or other consequential damages incurred because of interruption of service."[138]

Some complain that this gives prosecutors too much leeway that can lead to excessive punishments.[139] Additionally, because so few cybercrimes have been prosecuted, judges have no precedent to look to, which forces them to look directly to the CFAA's loose standards for guidance regarding punishment.[140] To an extent, some courts have addressed the CFAA's vagueness. In *Nosal I* and *II*, as well as *Brekka*, the courts grappled with expansive and narrow interpretations of the CFAA.[141] In both *Brekka* and *Nosal II*, the court eventually settled on a narrower interpretation of definitions in the CFAA.[142] In *Nosal II*, the court noted that expanding the CFAA to cover misappropriation could mean that "describing yourself as 'tall, dark and handsome,' when you're actually short and homely, will earn you a handsome orange jumpsuit."[143] Courts are rightfully wary of interpreting this vague language, which continues to present issues in current, and likely future, cases. Congress needs to clarify some of that vague language and openly state its intent, so lawyers and judges do not have to guess at what the legislature meant by its choice of words.

Of course, web scraping should not be legal in all circumstances. Private profiles, or profiles that require a password login for access are not web pages that are scrapable under the principles outlined above.[144] Notably, making web scraping legal in certain circumstances under the CFAA would not make web scraping legal in every circumstance. There is no reason that a company could not prohibit web scraping without prior consent in its Terms of Use.

Under this method, web scraping without permission from the scraped company would result in a breach of contract; the scraped company could file a civil suit against the web scraping party. This suit could result in the payment of damages, which is a far more reasonable punishment than a

---

137. 18 U.S.C. § 1030(g).

138. *Id.* § 1030(e)(11).

139. Katie Bo Williams, *Judges Struggle with Cyber Crime Punishment*, THE HILL (Jan. 9, 2016, 9:54 AM), http://thehill.com/policy/cybersecurity/265285-judges-struggle-with-cyber-crime-punishment [https://perma.cc/KL6H-QB7J].

140. *Id.*

141. LVRC Holdings LLC v. Brekka, 581 F.3d 1127, 1131-35 (9th Cir. 2009); United States v. Nosal, No. CR 08-00237 MHP, 2010 WL 934257, at *3-6 (N.D. Cal. Jan. 6, 2010) [hereinafter *Nosal II*].

142. *See* United States v. Nosal, 676 F.3d 854, 857, 862-63 (9th Cir. 2012) [hereinafter *Nosal III*]; *see also LVRC Holdings LLC*, 581 F.3d at 1133-35.

143. *Nosal III*, 676 F.3d at 862.

144. *See supra* notes 109-110 and accompanying text.

potential sentence of "not more than ten years," which could be combined with a fine for a first-time offender of the CFAA.[145]

V. PROPOSAL: AMENDING THE COMPUTER FRAUD AND ABUSE ACT

As a solution to the CFAA's improper jurisdiction over web scraping, this Note proposes an amendment to the Computer Fraud and Abuse Act.[146] The proposed amendment would add a subsection to 18 U.S.C. § 1030 (e) and clarify that web scraping a publicly accessible website is not a form of hacking.

Specifically, the proposed amendment would define the ambiguous term "without authorization"[147] to mean "(1) to access a publicly available computer, website, or online service that has been effectively locked through means of a password or a similar mechanism without authorization and use such access to obtain or alter information in the computer that the accesser is not entitled to so obtain or alter, or (2) to access a private computer that a person was not given authorization to access and use such access to obtain or alter information in the computer that the accesser is not entitled to so obtain or alter. (3) The term 'a password or similar mechanism' does not include bans or blocks of particular entities."

Much of the language mirrors that of the CFAA's definition of the term "exceeds authorized access," which is done purposefully.[148] By mirroring the language of the "exceeds authorized access" definition, lawyers and courts would be able to better understand the nuanced differences between the two terms. This amendment would be an appropriate addition to the CFAA, as web scraping should not be considered to be a form of hacking under the CFAA.

The purpose of the proposed amendment is to clearly state that web scraping is not a form of hacking. Notably, by including that the computer, website, or online service is "locked through means of a password or a similar mechanism," the definition does not allow the people controlling web scrapers to scrape whatever they please. This definition limits them to web pages that are available to any member of the public, or more specifically, any page that is accessible without a login or a password is fair game for a web scraper.

Ultimately, this proposed amendment focuses on the importance of not considering web scraping to be a form of hacking or unauthorized access under the CFAA. This matters to the businesses that rely on web scraping for their livelihoods; even the presiding judge in the *hiQ* case allowed hiQ to continue web scraping during the course of the case, because without it, hiQ

145. 18 U.S.C. § 1030(e)(2)(B).
146. *Id.* at § 1030.
147. *Id.* at § 1030(a)(1).
148. *See id.* at § 1030(e)(6).

would no longer be able to function and would cease to exist.[149] The CFAA, as it currently stands, can make a person who web scrapes face multiple years in prison.[150] A small business operation should not face felony charges under the CFAA for web scraping, particularly when that same method is used to operate commonly known and widely used search engines like Google.[151] While the act of web scraping in certain circumstances may be punishable, it should not necessitate the level of punishment associated with a violation of the CFAA. This proposed amendment attempts to both clarify an ambiguous phrase in the CFAA and remove the unnecessarily harsh punishments associated with web scraping.

Subsection one of the proposed amendment focuses on computers, websites, and online services that are available to the public without the need for a login, password, or similar locking mechanism. This cuts to the core of the driving force behind the proposed amendment. Web scraping publicly available websites should not be considered access "without authorization."[152] If the information on the website is available to the entirety of the public without needing a password, then it should be scrapable under the CFAA.

Subsection two focuses on both public and private computers, websites, and online services that are effectively locked through a password or login mechanism. Under subsection two of the proposed amendment, scraping a website that has been effectively locked through a password may be illegal under the CFAA if the person scraping the site was not the person given the authority to access the website. This is to ensure that under the CFAA and this proposed amendment web scrapers would only be able to scrape websites that are publicly available to everyone. This is meant to protect the companies that offer passwords and logins as a means of privacy and the consumers who choose to take advantage of these privacy protections. The term "effectively" also forces companies to take actual steps toward reasonable protection, rather than simply complying by using the password "password" to get protection under the proposed amendment.

For instance, if LinkedIn were the website being scraped, only publicly available profiles could be scraped under the proposed amendment to the CFAA. If a scraper had stolen the login information of a LinkedIn user and used that information to access private profiles, that scraper could be convicted under the CFAA, since that information was not "given" to him personally. Should a scraper make their own LinkedIn profile and use their own login information to access and scrape private profiles, that would violate an existing section of the CFAA regarding a person who "exceeds authorized access."[153] This proposed amendment is not intended to make web scraping legal under all circumstances; it is simply intended to clearly state that web

149. hiQ Labs, Inc. v. LinkedIn, Inc., 273 F. Supp. 3d 1099, 1104 (N.D. Cal. 2017) (order granting preliminary injunction).

150. *See* 18 U.S.C. § 1030(c).

151. *See* Hirshey, *supra* note 20, at 898.

152. *See* 18 U.S.C. § 1030(a)(1).

153. *See id.* at § 1030(e)(6).

scraping websites available to the public should not be considered hacking under the CFAA.

Noticeably, subsection three of the proposed amendment states that only passwords, logins, and other similar access or authorization mechanisms are acceptable barriers to access. Something like the blocking of a particular IP address, like what happened in the *Craigslist* case, would not be a permissible access barrier under the proposed amendment.[154] Should such a case take place under the proposed amendment, the blocking of the IP address would not trigger the protection of the CFAA.

A. Supporters

1. Scraping Businesses and Academics

Some of the biggest supporters of the proposed amendment would be businesses and academic researchers who use web scraping in their respective lines of work. Their business models rely on the ability to do this, and the removal of the harsh punishments associated with the CFAA would make their businesses less risky. Academic researchers who use web scraping to access data for various research projects would likely want to do so without the potential for a federally mandated punishment.[155]

Some academics may be hesitant to pursue particular projects if they require web scraping. This proposed amendment may remove some of that hesitation and lead to ground breaking studies and innovation. Both parties have a lot to gain from this proposed amendment and could use their business clients or academic circles to garner support for the proposed amendment.

Some people may say that academics have nothing to fear, as it's unlikely that the government would go after a researcher who scrapes websites purely for the purposes of a study. Indeed, in *Nosal II*, when the government pushed for an expanded interpretation of "exceeds authorized access," the government claimed that it would not prosecute minor violations of the CFAA.[156]

But are people willing to risk their freedom relying on the one-time promise of a few lawyers from one state's Attorney General's office? One person may believe that their violation of the CFAA is minor, while another person, perhaps a prosecutor or a judge, may disagree. Even in court, a person cannot say, "This is a minor violation of the CFAA, and at least one lawyer from one Attorney General's office in one state claimed that they would not prosecute minor violations of the CFAA, so this case should be dismissed." First and foremost, that person is actively admitting that their actions were in violation of federal law, which is highly unadvisable. Next, a judge may find

---

154. *See* Craigslist Inc. v. 3Taps, 964 F. Supp. 2d 1178, 1186 (N.D. Cal. Aug. 16, 2013)..

155. *See* Casey Fiesler, *Law and Ethics of Scraping: What HiQ v LinkedIn Could Mean for Researchers Violating TOS*, MEDIUM (Aug. 15, 2017), https://medium.com/@cfiesler/law-ethics-of-scraping-what-hiq-v-linkedin-could-mean-for-researchers-violating-tos-787bd3322540 [https://perma.cc/7Y6K-Y3PE].

156. United States v. Nosal, 676 F.3d 854, 857, 862 (9th Cir. 2012) [hereinafter *Nosal III*].

that the lawyer's claims are not something that people can rely upon, or that the violation is not minor, or that it does not matter that the violation is minor because it remains a violation. Businesses and academics need clear-cut definitions showing that web scraping is legal under the CFAA. Otherwise, "[t]he difference between puffery and prosecution may depend on whether you happen to be someone an AUSA [Assistant United States Attorney] has reason to go after."[157]

2. Legislators and Policy Groups

The community of supporters in Congress for the proposed amendment may include Senator Ron Wyden of Oregon and Representative Zoe Lofgren of California.[158] They presented Aaron's Law to the Senate and the House, respectively.[159] Both are familiar with the issue of vagueness in the CFAA and may support a bill that looks to clarify the CFAA. Their support could draw media and public attention to problematic parts of the CFAA.

The supporting members of Congress could draw on the narrow scope of the proposed amendment as a strength for its success. Short, simple bills could be seen as low-hanging fruits that are easy to get through Congress.[160] Some may consider this a weakness, since if it faces opposition in Congress and compromise is required, there may not be enough substance in this proposed amendment that could be compromised. A solution could be adding this proposed amendment to a larger package of bills that have more substance and therefore, more space for potential compromises.

The Electronic Frontier Foundation ("EFF") is an organization that was also part of the community that followed Aaron Swartz's case and supported Aaron's Law.[161] The EFF could help rally support from the online and technologically inclined community, as it has done in the past for Aaron's

---

157. *Id.*

158. *See Wyden, Lofgren Paul Introduce Bipartisan, Bicameral Aaron's Law to Reform Abused Computer Fraud and Abuse Act*, U.S. SENATOR RON WYDEN OF OR. (Apr. 21, 2015), https://www.wyden.senate.gov/news/press-releases/wyden-lofgren-paul-introduce-bipartisan-bicameral-aarons-law-to-reform-abused-computer-fraud-and-abuse-act- [https://perma.cc/6JHB-BRYE].

159. Aaron's Law Act of 2013, H.R. 2454, 113th Cong.; Aaron's Law Act of 2013, S. 1196, 113th Cong.

160. *See*, *e.g.,* Burgess Everett, *Not The Onion: Congress Set to Pass Bills*, POLITICO (Aug. 1, 2018), https://www.politico.com/story/2018/08/01/congress-republicans-bills-agenda-753296 [https://perma.cc/7GNB-LLDF] (noting that Congress is passing smaller appropriations bills with little resistance, rather than a larger spending bill).

161. *See* April Glaser, *Aaron Swartz's Work, Computer Crime Law, and 'The Internet's Own Boy'*, ELEC. FRONTIER FOUND. (Aug. 27, 2014), https://www.eff.org/deeplinks/2014/08/aaron-swarts-work-internets-own-boy [https://perma.cc/SCP8-C9MJ]; Kurt Opsahl & Trevor Timm, *Aaron's Law Introduced: Now is the Time to Reform the CFAA*, ELEC. FRONTIER FOUND. (June 20, 2013), https://www.eff.org/deeplinks/2013/06/aarons-law-introduced-now-time-reform-cfaa [https://perma.cc/EQP2-SBAZ].

Law.[162] The EFF has called for CFAA reform many times in the past.[163] Perhaps the EFF would support the proposed amendment.

The EFF has a long relationship with scholars, such as Orin Kerr, as well as other potentially interested parties, such as the American Civil Liberties Union, the Center for Democracy and Technology, and Stanford's Center for Internet and Society.[164] The EFF's past work could be useful in rallying that base again and could provide support in a multitude of ways. When writing opinions or orders, judges can cite to academic articles as support for a particular ruling.[165] Interested parties could also write amicus briefs expressing the position of various players in the online ecosystem and how such rulings could affect these parties.[166]

Supporting members of Congress could call on other interested members or entities like the EFF and discuss the possibility of adding additional amendments to the CFAA, as additional amendments may provide more room for compromise with opposing congresspeople. For instance, the EFF has long been a proponent of reforming the CFAA's excessive and redundant criminalization.[167] That, among other things, could be added to the proposed amendment to add additional material for inevitable compromises. However, such additions would have to be added carefully, as no one likes to see important policy implications treated like potential bargaining chips.

3. Online and Technology Communities

Some parties may also turn to the public and online communities for support. For instance, when working on a draft of Aaron's Law, Representative Lofgren sought feedback on Reddit, which has significant ties

---

162. *See* Adi Kamdar, *Calling All Engineers and Technologists: We Need Your Help to Reform the CFAA*, ELEC. FRONTIER FOUND. (Apr. 9, 2013), https://www.eff.org/deeplinks/2013/06/aarons-law-introduced-now-time-reform-cfaa [https://perma.cc/G3SL-J8UA].

163. *See*, *e.g.*, *id.*; Kurt Opsahl & Trevor Timm, *Aaron's Law Introduced: Now is the Time to Reform the CFAA*, ELEC. FRONTIER FOUND. (June 20, 2013), https://www.eff.org/deeplinks/2013/06/aarons-law-introduced-now-time-reform-cfaa [https://perma.cc/EQP2-SBAZ].

164. *See* Cindy Cohen et al., *EFF's Initial Improvements to Aaron's Law for Computer Crime Reform*, ELEC. FRONTIER FOUND. (Jan. 17, 2013), https://www.eff.org/deeplinks/2013/01/effs-initial-improvements-aarons-law-computer-crime-reform [https://perma.cc/5C53-FP45].

165. *See, e.g.,* hiQ Labs, Inc. v. LinkedIn, Inc., 273 F. Supp. 3d 1099, 1111-13 (N.D. Cal. 2017) (order granting preliminary injunction) (court examination of an academic article).

166. Interested parties can join together to write a single amicus brief. *See*, *e.g.*, Brief of Amici Curiae Electronic Frontier Foundation et al. in Support of Plaintiff-Appellee, hiQ Labs, Inc., No. 3:17-CV-03301-EMC, 2017 WL 5757674 (9th Cir. 2017) (No. 17-16783).

167. Lee Tien, *Why the CFAA's Excessive Criminalization Needs Reform*, ELEC. FRONTIER FOUND. (Apr. 2, 2013), https://www.eff.org/deeplinks/2013/04/cfaas-excessive-criminalization [https://perma.cc/7YPD-LN7R].

to Swartz.[168] Additionally, rallying the technologically inclined community may be helpful in ensuring that the proposed amendment does not accidentally disallow some technologies that are commonly used and are also not intended to be the targets of the CFAA.

For instance, technologists or engineers could assist with wording in order to ensure that robots.txt are not expressly prohibited under the proposed amendment. These technologists could come from the interested entities' experts, as well as other interested parties. For instance, there is an entire Reddit thread dedicated to web scraping.[169] Between the Reddit ties to Aaron Swartz and the community consisting entirely of people who are interested in or actively choose to do web scraping, there is a large group of people who would be affected by a proposal like this and who may be waiting for an opportunity to discuss the realities of a proposed amendment.

B. Opponents

1. Scraped Businesses

The companies that are being scraped would be opponents of the proposed amendment. Naturally, these would be the companies that are presumably losing business to other businesses that use web scrapers to utilize the scraped company's information and potentially undercut their profits. However, it is worthwhile to note that this amendment would not necessarily allow web scraping to occur anywhere on any website. Private profiles would still be protected under the proposed amendment, and so the scraping of private web pages would be illegal under the proposed amendment. Additionally, these scraped companies could prohibit web scraping in their Terms of Use and sue scraping companies for breach of contract. The goal of this amendment is not to make the Internet and all information on it a free-for-all; one goal is simply to prevent web scrapers from having to face the harsh penalties associated with the CFAA.

These scraped companies may also argue that technological measures like the blocking of IP addresses should be enough to trigger the protections of the CFAA. Adding a provision to the Terms of Use that the company reserves the right to block IP addresses could add another level of damages to a breach of contract suit against a web scraping party. Additionally, the company may have protections under 17 U.S.C. § 1201 (2012), which protects against a person "circumvent[ing] a technological measure that effectively controls access to a work protected under this title."[170] However, this protection only comes into effect if the information in question is

---

168. Zoe Lofgren, *I'm Rep Zoe Lofgren, Here Is a Modified Draft Version of Aaron's Law Reflecting the Internet's Input*, REDDIT: AMA (Feb. 1, 2013), https://www.reddit.com/r/IAmA/comments/17pisv/im_rep_zoe_lofgren_here_is_a_modified_draft/ [https://perma.cc/W89L-X2RL].

169. r/scrapinghub, https://www.reddit.com/r/scrapinghub/ [https://perma.cc/DDU8-E3EJ] (last visited Dec. 1, 2017).

170. 17 U.S.C. § 1201(a)(1)(A) (2012).

copyrighted.[171] Other statutes protect an entity against a multitude of potential harms that redundantly reappear in the CFAA, making the CFAA ripe for reform.[172] Another potential solution is for companies to reserve the rights to utilize a robot.txt in their Terms of Service.

2. Consumer Privacy Advocates

Another party that might oppose the proposed amendment is consumer privacy advocates, who may voice concern that this win for businesses and researchers would come at the cost of consumer privacy. The cases presented earlier reflect companies doing fairly responsible business; their focus on web scraping data related to event days and times, advertisements, and employment.[173] Another company may not be so reasonable. In an article concerning the *hiQ* order and ethics, Casey Fiesler points out that the context of privacy must be considered.[174] For instance, a widely publicized event urging people to attend may not have particularly high privacy implications. A dating profile that was made public may have far higher privacy concerns regarding personal matters like sexual orientation.[175] Data collection can present a serious privacy concern, and web scrapers may not be taking that into account.[176]

While this proposed amendment cannot force web scrapers to behave ethically, companies and consumers have other methods of recourse. Particularly private online interactions, such as those on dating websites, should be made completely private, either by the user's choice or the company's online construction. However, some consider that privacy by default may be the right choice for consumers, who can then open up their websites and profiles at their own discretion.[177] Companies could also release PSAs regarding public online profiles and what other companies can do with public information.

Should the proposed amendment become law, companies at high risk for scraping could also set up banner notifications warning their users about these new laws and their implications for public profiles and directing users to the websites' privacy settings. This would place choice in the consumers' hands; given the information, consumers could then make their own informed decisions regarding their online presence and privacy. By understanding the risks and taking the necessary steps to ensure their own privacy, consumers and companies could work to protect their privacy interests. Alternatively,

---

171. *Id.*

172. *See* 18 U.S.C. § 2511 (2012) (protects against the interception of electronic communications); 18 U.S.C. § 1832 (2012) (protects against the theft of trade secrets); 18 U.S.C. § 2701 (2012) (protection against the unlawful access to stored communication).

173. *See supra* notes 78-145 and accompanying text.

174. Fiesler, *supra* note 155.

175. *Id.*

176. *Id.*

177. *See*, *e.g.*, Sam Pfeifle, *"Privacy by Default" May Be Big Post-Regulation Issue*, INT'L ASS'N OF PRIVACY PROF'LS (Sept. 30, 2013), https://iapp.org/news/a/privacy-by-default-may-be-big-post-regulation-issue/ [https://perma.cc/2FY9-98H6].

companies at high risk for scraping could change all settings to private and put up a similar banner notification informing consumers of the change. In this situation, all consumers would be protected by default and then would be able choose to open their information up to others at whatever level they choose.

VI. CONCLUSION

Pop culture has played a bigger role in cybersecurity and the public perception of hacking than many may think. The 1983 movie *War Games* involved a teenager hacking into the North American Aerospace Defense Command and nearly starting World War III.[178] After watching the film, President Reagan brought the plot up with his national security advisors to see if this was something that could happen in real life.[179] This began the cybersecurity and hacking laws that we know of today.[180] Some television shows show hacking in a more realistic, modern light.[181] Other classics push hacking as the solution to protect the world from aliens or a few people from the human error that puts them in a dinosaur's path.[182]

Yet hacking has become bigger and more harmful than pop culture could have imagined. These breaches are very real and can reveal the personal information of people around the world.[183] The legal definition of hacking[184] may not be nearly as glamorous sounding, as invasive, or as detrimental to the public. Some hackers are prosecuted on a federal level, as though they had attempted to access a government computer without authorization.[185]

People who utilize web scraping are some of these parties who are prosecuted at a high level and face the high punishments associated with the CFAA.[186] The CFAA was originally meant for people who tried to unlawfully access computers belonging to financial institutions or the United States Government.[187] Web scrapers who access web pages that are available to the general public are not the community that the CFAA was originally intended to target. There was no way for lawmakers at that time to even conceive of the idea of web scrapers. How could they possibly build a law for something they never knew was possible?

---

178. WAR GAMES (MGM Studios Inc. 1983).

179. Fred Kaplan, *'WarGames' and Cybersecurity's Debt to a Hollywood Hack*, N.Y. TIMES (Feb, 19, 2016), https://www.nytimes.com/2016/02/21/movies/wargames-and-cybersecuritys-debt-to-a-hollywood-hack.html [https://perma.cc/D6F7-NYEA].

180. *See id.*

181. Forbes Technology Council, *supra* note 2.

182. *See* INDEPENDENCE DAY (Twentieth Century Fox 1996); *see also* JURASSIC PARK (Universal Pictures 1993).

183. *See* Jeff John Roberts, *Here are 10 of the Biggest Corporate Hacks in History*, FORTUNE (June 22, 2017), http://fortune.com/2017/06/22/cybersecurity-hacks-history/ [https://perma.cc/V3SA-XGSD].

184. *See supra* note 35 and accompanying text.

185. *See* 18 U.S.C. § 1030(a)(3).

186. *See id.*

187. *See id.* at § 1030(a)(2)(A); *see also id.* at § 1030(a)(2)(B).

Thus, it is up to today's world to update the Computer Fraud and Abuse Act to protect those who the CFAA can harm due to ambiguous language and the failure to update as technologies and practices have changed. Between the history, the rises in technology, and the recent and current court cases surrounding the CFAA and the act of web scraping, the time is right for the legislation to reflect the significant technological changes that have occurred since 1986.

The above proposed amendment is the first step in modernizing a potentially outdated CFAA to reflect the practices of today's technologically advanced world. The amendment not only acknowledges that information generally available to the public is truly available to all people and entities, even business competitors; it also protects those who want to work and grow in this world without facing federal punishments as a result. It is time for Congress to clarify the CFAA and relaunch it as a modern law in line with a modern world.