Stanford Law Review

Volume 70 February 2018

NOTE

Demystifying Hash Searches

Dennis Martin*

Abstract. A hash search is a very accurate, computationally efficient technique for testing whether a computer contains illicit material. Although police have been running hash searches for many years, case law is scarce regarding whether and to what extent the Fourth Amendment permits their use. Some commentators have argued that because hash searches reveal information concerning only the presence or absence of contraband, courts shouldn't consider them Fourth Amendment searches. Rather, courts should treat hash searches as a sort of digital dog sniff.

This Note disagrees. It argues first that even accepting the analogy to digital dog sniffs, hash searches nevertheless violate the Fourth Amendment under *Florida v. Jardines* whenever they are used to look for evidence outside the scope of a search warrant or other permissive mechanism. It then argues that there is no limiting principle that would permit the use of hash searches but not more sophisticated algorithms—algorithms that would constitute the modern equivalents of general warrants. Accordingly, it proposes a rule that covers not only the hash searches that are being used now but also the more sophisticated forensic techniques that will be used in the near future: Police conduct a Fourth Amendment search whenever they use an algorithm to perform a task that would be a search if conducted manually by a human.

* J.D. Candidate, Stanford Law School, 2018; M.S. in Computer Science Candidate, Stanford University, 2020. Thanks to David Sklansky and Robert Weisberg for their generous and thoughtful feedback. Thanks also to my editors and friends—Katherine Moy, Katie Kelsh, Rachel Neil, Dan Brenner, Will Orr, David Steinbach, Brian Baran, and Matt Getz—who made this Note better at every stage. It is dedicated to Marlene Cielec, who wanted very much to read it.

Table of Contents

Introduction

Suppose a government investigator executing a warrant to search your computer for evidence of tax fraud instead clicks through your hard drive file by file looking for pirated music. He's clearly exceeding the scope of his warrant in violation of the Fourth Amendment.[1] Now suppose he writes a computer program to do the exact same thing. Different result?

For many years, government investigators have used digital forensic software to conduct hash searches: a very accurate, very computationally efficient type of search that can be used not just for legitimate purposes but also to identify evidence of crimes outside the scope of a search warrant.[2] Still, many commentators argue that because these hash searches reveal information concerning only the presence or absence of illicit material, the Fourth Amendment does not prohibit their use.[3] They argue that we ought to treat hash searches as a sort of digital dog sniff.[4]

Courts, meanwhile, have been hesitant to apply the Fourth Amendment to algorithmic investigative techniques. Indeed, the Court only recently addressed, in *Riley v. California*, what limits the Fourth Amendment places on *human* searches of digital information.[5] By contrast, hash searches involve *algorithmic* searches of digital information. And hash searches are an appropriate place to begin to assess what sort of limits the Fourth Amendment imposes on algorithmic investigative techniques: Unlike many types of still-developing technological surveillance, hash searches are already being used, and their underlying technology is unlikely to change in the future.[6] And the technology behind hash searches is relatively easy to understand, even for laypeople.

It is increasingly important that courts weigh whether and how the Fourth Amendment governs algorithmic investigative techniques. Although some new technologies, like thermal imaging devices, give police investigative powers they've not previously had, algorithmic investigative techniques

1. *See, e.g.*, United States v. Carey, 172 F.3d 1268, 1270-71, 1273, 1276 (10th Cir. 1999); *see also* U.S. CONST. amend. IV ("The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.").

2. *See infra* Part I.

3. *See, e.g.*, Richard P. Salgado, Reply, *Fourth Amendment Search and the Power of the Hash*, 119 HARV. L. REV. F. 38, 43-46 (2006).

4. *See, e.g.*, *id.*

5. *See* 134 S. Ct. 2473, 2485 (2014).

6. *See, e.g.*, Salgado, *supra* note 3, at 38 (describing hashing as "an important fixture in forensic examinations" as of 2006).

typically mimic work that has historically been done by human officers.[7] Indeed, the very purpose of the technologies is to replace, and improve upon, human police work. But the Fourth Amendment should not be read to permit police to use computer programs to conduct investigations that would, if police conducted them manually, be illegal searches. Such a reading would allow law enforcement to shift its investigatory work onto algorithms and away from the Fourth Amendment.

This Note proceeds in five Parts. Part I explains the technology behind hash searches. Prominent commentators have described hashing algorithms as "complex"[8] and "complicated,"[9] and some courts have misunderstood how they function. Part I uses some simple examples to show that hash searches are not so arcane. Part II catalogs the various contexts in which courts have addressed hash searches, identifying points upon which courts agree and questions that remain open. Part III considers the argument that hash searches should be analyzed as digital dog sniffs. It argues that even if we accept this analogy, hash searches outside the scope of a warrant are nevertheless illegal searches under *Florida v. Jardines*.[10] Part IV argues, alternatively, that a reading of the Fourth Amendment permitting hash searches would also permit suspicionless algorithmic searches for ordinary evidence of criminal wrongdoing—twenty-first century general warrants. Finally, Part V argues that in light of these concerns, courts ought to adopt an affirmative framework for assessing their legality rather than a reactive one. This Note proposes such a framework: Police conduct a search when they use an algorithm to perform some task that would be a search if conducted by a human investigator.

I. Demystifying Hash Searches

Hash searches, like many concepts in computer science, can seem esoteric. Legal commentators have not helped: They've described hash searches as employing "complex mathematical algorithm[s]"[11] or "complicated mathematical operation[s]."[12] Some have suggested that judges are ill equipped to assess the legality of hash searches and other digital forensic techniques,

7. *See infra* Part V.B.

8. Salgado, *supra* note 3, at 38.

9. *See, e.g.*, Orin S. Kerr, *Searches and Seizures in a Digital World*, 119 HARV. L. REV. 531, 541 (2005).

10. 133 S. Ct. 1409 (2013).

11. Salgado, *supra* note 3, at 38.

12. *See, e.g.*, Kerr, *supra* note 9, at 541.

given those techniques' technical complexity.[13] And it is true that some courts have seemed to grasp only imprecisely how hash searches operate.[14] As this Part will show, however, hash searches are conceptually quite simple.

A. A Simple Hash Search

Before diving in, we need to distinguish between three different concepts, all of which relate to hashing: (1) a hash function, (2) a hash value, and (3) a hash set. A hash function is a mathematical process that takes some input, like a text file or an image, and outputs a hash value.[15] A hash value is a series of letters and numbers (what some courts have called a "digital fingerprint"[16]) assigned to a particular input.[17] And a hash set is a collection of inputs that are stored according to their hash values.[18] Examples will make these concepts clearer.

Suppose I write some simple hash function. It takes a string of text as an input and outputs the sum of the ordinal values of the text's constituent letters. If you feed my hash *function* the input "Ignatius," it outputs 100, which is thus the hash *value* for "Ignatius."

Figure 1

Input	I	G	N	A	T	I	U	S	**SUM**
Hash Value	9	7	14	1	20	9	21	19	**100**

I could write a similarly simple hash function for images. All digital images are made up of pixels, which are just tiny points of color situated in a two-dimensional array.[19] Each pixel is a composite of three component colors, red, green, and blue, each of which is assigned a value from 0 to 255.[20] One pixel,

13. *See, e.g., id.* at 578 ("Judges have little sense of how to distinguish a reasonable forensics process from an unreasonable one, however. The technical details are too complex and fluid.").

14. *See* Robyn Burrows, Comment, *Judicial Confusion and the Digital Drug Dog Sniff: Pragmatic Solutions Permitting Warrantless Hashing of Known Illegal Files*, 19 GEO. MASON L. REV. 255, 270-76 (2011).

15. *See* BRUCE SCHNEIER, APPLIED CRYPTOGRAPHY: PROTOCOLS, ALGORITHMS AND SOURCE CODE IN C, at 30 (20th anniversary ed. 2015); *see also* Salgado, *supra* note 3, at 39.

16. *See, e.g.*, United States v. Wellman, 663 F.3d 224, 226 n.2 (4th Cir. 2011) ("A 'hash value' is an alphanumeric string that serves to identify an individual digital file as a kind of 'digital fingerprint.'").

17. *See* SCHNEIER, *supra* note 15, at 30.

18. *See* Salgado, *supra* note 3, at 40 n.14.

19. *See* Code.org, *Images, Pixels and RGB*, YOUTUBE (Mar. 11, 2015), https://perma.cc/9BWD-ME65.

20. *Id.*

then, might be coded as (240, 0, 120); that is, it takes a red value of 240, a green value of 0, and a blue value of 120. And an image on your computer is just an array of these pixels. For example, an image with a resolution of 1280 x 1024 contains 1,310,720 pixels. So we could write a hash function that cycles through an image, pixel-by-pixel, and adds the red, green, and blue values of each pixel to some sum. Once the function reaches the last pixel in the image, it returns that sum, which is now the image's hash value, just like 100 is the hash value for "Ignatius" in Figure 1 above.

In any hash function, we look for two properties. First, a hash function must be *consistent*: Whenever we pass in a certain input, the function must always return the same output.[21] That is, "Ignatius" must return 100 every time it is inputted into the hash function described in Figure 1 above. This property is necessary for a hash function.[22] Second, we would like for a hash function to be *well distributed*.[23] That is, we'd like for it to return different outputs for different inputs as often as possible. When two inputs produce the same output, the hash function has generated a "collision."[24] Producing few collisions is not a necessary property of a hash function, but it helps distinguish good hash functions from bad ones.

We see, then, that our simple function is a valid hash function, just not a very good one: It behaves with perfect consistency but produces many collisions. Indeed, any other word whose letters sum to 100 will generate the same hash value, including "gauntlet" and "perturb" and many more besides.

Why do we care how often our hash function produces collisions? Because it affects the performance of our third concept: the hash *set*. Recall that a hash set is a collection of inputs stored according to their hash values. Another example will help clarify.

Suppose I write some computer program that allows you to check whether some word is contained in a standard English dictionary. So if you type in "groggily," my program will return "true," but if you enter some gibberish, my program will return "false." I'll begin by storing all the words in a standard dictionary somewhere, and then I'll need to design some method for searching all those words to see if any matches the one you enter. Consider two ways I could write such a program.

One method I could use would be to search the dictionary in alphabetical order, checking each word to see if it matches yours. If any word were to

21. *See* THOMAS H. CORMEN ET AL., INTRODUCTION TO ALGORITHMS 257 (3d ed. 2009) ("[A] hash function h must be deterministic in that a given input k should always produce the same output $h(k)$.").

22. *See id.*

23. *See id.* at 262.

24. *Id.* at 257.

match, my program would halt and return "true." But if my program reached the end of the dictionary without matching, it would return "false." So my program would take your word and check it first against "a," and then against "aardvark," and then against "aardwolf," and so on.

The weakness in such a method is obvious: My program would execute very quickly if you entered "abacus" but very slowly if you entered "zeitgeist." And the larger the set of words I was checking, the slower my program's average performance would be.[25]

A much better method would be one that takes your input word and goes immediately to the spot in the dictionary where that word should be. If the word were present, the program would return "true," and if absent, "false." But the program would take the same amount of time to execute regardless of what word you enter.

This is exactly what hash searches do. They take an input and run it through a function to generate a hash value; they then go to the space in a hash set denoted by that value to see whether there's some corresponding item in that space.[26] This is a fantastically powerful search technique: No matter how large our set grows, it will always take us the same (very short) amount of time to check whether that set contains some item.[27]

This is why it's important that hash functions be well distributed: When my program checks a space in its hash set, I want there to be only one item there. To return to my simple hash function, suppose you enter "honeycomb." The hash value for that word is 100, so my program goes to the 100th space in its hash set. But when it gets there, it sees that there are many, many words present: not just "honeycomb" but "gullibly" and "goutweed" and others. Now the search program must cycle through each word one by one to see whether any matches your input. This is undoubtedly a better method than the purely linear search described above: Only a subset of the full dictionary will be stored in this 100th bucket. But this method is still relatively inefficient. It would be much better if "honeycomb" were the only word stored in its bucket.

Well-distributed hash functions achieve something very near to a unique output for every input. Using SHA-1,[28] for example, it's easy to see how even

25. In computer science, such a search algorithm performs in linear time, which we annotate $O(n)$ (pronounced "big oh of en"). *See* CORMEN ET AL., *supra* note 21, at 44, 47-49. That is, the time it takes this algorithm to execute tends to grow proportionally to n, the number of words in my dictionary. *See id.*

26. *See id.* at 256.

27. Unlike our linear-time algorithm above, our hash search performs on average in constant time, denoted $O(1)$ ("big oh of one"). *See id.*

28. SHA-1 (Secure Hash Algorithm 1) is a cryptographic hash function developed in part by the National Security Agency. *See* SCHNEIER, *supra* note 15, at 442; Tim Fisher, *What Is*

footnote continued on next page

minor changes to our input generate dramatically different outputs:[29]

Figure 2

Input	Hash Value
We the People	6f43ee071387dcdf2a260d275960074c93878b19
We the people	dd884fb00446d3e8eec52c94e006f08094a0c9de

Using one of these well-distributed hash functions, my search program performs much more efficiently.[30] So if you were to search a set of three-word phrases contained in the Constitution for "We the People," my program would go immediately to the space in our set denoted by the value "6f43ee071387dcdf2a260d275960074c93878b19," observe that the space is occupied, and return "true."

B. How Law Enforcement Uses Hashing

Police conduct a hash search when they use an algorithm to check whether the hash value of a given item, like an image or MP3, is contained within a certain hash set. There are several ways police can execute a hash search. Typically, the algorithm that executes a hash search is just one of many tools offered by a piece of digital forensic software, like Forensic Toolkit[31] or EnCase.[32] Police can sometimes conduct a hash search of a suspect's files using digital forensic software installed on computers located at the stationhouse. They

SHA-1?: Definition of SHA-1 and How It's Used to Verify Data, LIFEWIRE (updated Oct. 30, 2017), https://perma.cc/T4WR-5NBR.

29. For an SHA-1 hash value generator, see SHA1 ONLINE, https://perma.cc/G3RE-6AUT (archived Nov. 6, 2017). The degree of difference matters. Suppose you have two hash values. The first hash value corresponds to some known piece of text. You don't know the text to which the second value corresponds, but you observe that the second value is very similar to the first. If a minor change to some text produced only a minor change to its corresponding hash value, you could infer that the second (unknown) piece of text is similar to the first. That would make hash values backward computable and thus less secure. *See infra* note 50.

30. It is worth noting that the hash search only performs more efficiently than the linear search when we're searching large datasets. To return to my dictionary example, the difference between checking each word in alphabetical order (a linear search) and going immediately to the space where the word I'm looking for should be (a hash search) is trivial if the dictionary contains only a few words. But large datasets are the ones we're concerned about in the digital forensics context.

31. *See* ACCESSDATA, FORENSIC TOOLKIT (FTK): USER GUIDE 25, 29-30 (2016), https://perma.cc/C9C7-JGUR; *see also infra* text accompanying notes 117-23.

32. *See* Guidance Software, EnCase Forensic 2 (n.d.), https://perma.cc/4HZL-ATJJ; *see also infra* text accompanying note 88.

might do so if, for example, an email provider forwards them an email containing an allegedly illegal attachment[33] or if a suspect is sharing potentially illicit material on a public site.[34] Alternatively, they might get a search warrant and then use digital forensic software to search a suspect's computer.[35] Searching a computer using forensic software, however, is so time consuming as to make impractical conducting such a search at the scene of an investigation.[36] As the Department of Justice (DOJ) acknowledges, "It may take days or weeks to find the specific information described in the warrant because computer storage devices can contain extraordinary amounts of information."[37] Accordingly, the DOJ recommends that investigators request permission, in their warrant applications, to conduct their forensic searches off-site.[38]

Police can conduct an off-site search for digital evidence in two ways. They can seize the physical computer and remove it to the stationhouse.[39] Alternatively, police can make a digital copy of the computer's hard drive.[40] This copy, called an "image copy," "duplicates every bit and byte on the target drive including all files, the slack space, Master File Table, and metadata in exactly the order they appear on the original."[41] Police can confirm that the original hard drive matches the image copy by running each through a hash function: If the outputted hash values match, then the hard drive and the image copy are identical. Police can then, back at the stationhouse, take as long as they need to thoroughly search the image copy without depriving the suspect of his property.

Once police have seized a suspect's computer or created an image copy of it, they might have various reasons to use a hash search. They might use hash values to show that data introducing at trial is the same as the data they seized from the defendant. They might also use hash values to quickly exclude files guaranteed not to contain evidence, such as the files constituting a computer's

33. *See infra* Part II.B.

34. *See infra* Part II.A.

35. *See* COMPUT. CRIME & INTELLECTUAL PROP. SECTION, CRIMINAL DIV., U.S. DEP'T OF JUSTICE, SEARCHING AND SEIZING COMPUTERS AND OBTAINING ELECTRONIC EVIDENCE IN CRIMINAL INVESTIGATIONS 79-82 (3d ed. 2009), https://perma.cc/5ECT-J4JR [hereinafter CCIPS MANUAL].

36. *See id.* at 77.

37. *Id.*

38. *See id.* at 76-78.

39. *See, e.g.*, United States v. Hay, 231 F.3d 630, 632, 637 (9th Cir. 2000) (holding that the seizure of an entire computer was reasonable "because of the time, expertise, and controlled environment required for a proper analysis").

40. *See* CCIPS MANUAL, *supra* note 35, at 78.

41. *Id.* (quoting United States v. Vilar, No. S305CR621KMK, 2007 WL 1075041, at *35 n.22 (S.D.N.Y. Apr. 4, 2007)).

operating system. They might use hash values to locate evidence for which they have a warrant.[42] And they might sometimes use hash values to identify evidence of crimes outside the scope of their warrant.

1. Preserving evidence

Let's start with a benign use of hashing: preserving evidence for trial. Before she can introduce evidence at trial, the proponent of the evidence must show that it is authentic; that is, the proponent "must produce evidence sufficient to support a finding that the item is what the proponent claims it is."[43] Prosecutors often use hash values to authenticate digital evidence.[44]

After creating an image copy of a hard drive, investigators will typically generate a hash value for each file on the drive. Because each file generates a unique hash value[45] and even a minor change in the file will cause a significant change in its hash value,[46] this process creates a sort of digital chain of custody. Suppose law enforcement officers seize and copy a suspect's hard drive on January 1. If they seek to introduce files from that hard drive against the now-defendant during a criminal trial on August 8, they should be able to show that the hash values for those files on August 8 are the same as they were on January 1.

Courts already allow the use of hash value comparison to confirm that evidence offered at trial is the same as evidence seized from the defendant.[47] Indeed, so reliable are hash value comparisons that some defendants have argued (unsuccessfully) that digital evidence should be inadmissible unless validated by a hash value comparison.[48]

---

42. *See infra* note 53.

43. *See* FED. R. EVID. 901(a).

44. *See* CCIPS MANUAL, *supra* note 35, at 199 ("[P]rosecutors may consider using the 'hash value' or similar forensic identifier assigned to the data on the drive to authenticate a copy of that data as a forensically sound copy of the previously admitted hard drive."); *see also* FED. R. EVID. 901(b)(4) (providing that an item offered as evidence can be authenticated using its "distinctive characteristics").

45. *See infra* notes 74-75 and accompanying text.

46. *See supra* note 29 and accompanying text; *see also, e.g.*, *supra* Figure 2.

47. *See, e.g.*, United States v. Glassgow, 682 F.3d 1107, 1110 (8th Cir. 2012) (noting that the SHA-1 value of the evidence admitted at trial matched the value of the files found on the defendant's computer).

48. *See, e.g.*, United States v. Stewart, 839 F. Supp. 2d 914, 931 (E.D. Mich. 2012) (rejecting the defendant's contention that the government conducted an unfair investigation because it failed to publish hash values publicly before analyzing his hard drives); United States v. Hock Chee Koo, 770 F. Supp. 2d 1115, 1123-24 (D. Or. 2011) (rejecting the defendants' argument that digital evidence seized by the government was inadmissible because the program used to copy his files did not generate hash values but noting that

footnote continued on next page

2. Excluding files known to be noncontraband

Digital investigators also perform hash searches in order to efficiently eliminate from their search set files they know to be noncontraband.[49] Say, for example, an investigator wants to exclude from the hard drive she is searching all the files that constitute the Windows operating system. As her program generates hash values for each file for evidence preservation purposes, it can also check those values against a hash set containing all standard Windows files.

Suppose some file on the suspect's computer generates the hash value "1ec656e9986aee222067f37a0a610fc7e8fa0787."[50] The investigator's forensics program will go to the space in her program's hash set of Windows files denoted by that hash value; if that space is occupied, her program flags the file as "safe," and she knows that she doesn't need to look at that file during her investigation. The National Software Reference Library collects many such hash sets (such as those containing copies of Microsoft's software) and provides them to federal, state, and local law enforcement agencies for exactly this purpose.[51]

3. Searching for evidence

Finally, law enforcement officers can use hash searches to look for evidence of crime. This use of hash searches follows naturally from the preceding two. Consider a generic digital forensic investigation. The investigator has already generated hash values for all files on the computer to preserve them for use at trial. She's also already checked those values against hash sets containing files known to be harmless. It is now very simple for her to check those values against hash sets comprising files known to contain

---

the defendants could argue to the jury that the copy introduced into evidence was not the same as the original).

49. *See, e.g.*, ACCESSDATA, *supra* note 31, at 29.

50. One other property of (good) hash functions is that their values work in only one direction: That is, given some hash value, you can't work backward to derive its input. *See* SCHNEIER, *supra* note 15, at 29-31. This is why an online company that cares about security will store its users' passwords as hash values rather than in plain English: If its password database is compromised, the passwords can't be easily derived. *See* Paul Ducklin, *Serious Security: How to Store Your Users' Passwords Safely*, NAKED SECURITY (Nov. 20, 2013), https://perma.cc/RTE8-S7EP (providing recommendations for safely storing users' passwords). This also means that you can hide indecipherable messages in the text of your Note.

51. *See National Software Reference Library (NSRL): Library Contents*, NAT'L INST. STANDARDS & TECH., https://perma.cc/T4XR-UJQU (archived Nov. 7, 2017); *see also* Neil C. Rowe, *Testing the National Software Reference Library*, 9 DIGITAL INVESTIGATION S131, S131-32, S137 & tbl.5 (2012) (evaluating the library's coverage by vendor as of August 2011).

illegal material, like pirated media or child pornography.[52] Recall that the great strength of hash searches is that they allow a user to search through a massive dataset quickly—so quickly, in fact, that the additional cost of checking the suspect's files against a database of illicit material is near zero.[53]

No wonder, then, that hash searches for illicit material seem so attractive. They require hardly any additional work by the digital investigator, they extend the length of the investigation by only a minimal amount of time, and they identify evidence of serious crimes that would otherwise often go undiscovered. Given how efficient and effective they are, it is important to determine whether and under what circumstances they are legal.

II. How Courts Have Handled Hash Searches

Federal courts have primarily confronted hash searches for evidence of crime in three contexts. First, many federal courts have addressed hash searches of information exposed to the public on peer-to-peer (or P2P) networks.[54] Defendants in these cases have challenged hash searches either as searches in violation of the Fourth Amendment or as providing an insufficient basis for probable cause. Nevertheless, all courts to address hash searches of publicly exposed information have found them both legal and reliable. Second, several courts have encountered hash searches used by email providers to filter messages for child pornography.[55] The primary issue in these cases, however, has been not whether hash searches violate the Fourth Amendment but whether the email providers are acting as government entities or agents.

Third is the context with which this Note is concerned: hash searches by government investigators of private information stored on digital storage devices.[56] Far fewer courts have addressed this use of hashing. And they have disagreed about whether and when hash searches are permitted.

---

52. This is not a simplification. In one of the more prominent cases to address hash searches, *United States v. Mann*, the investigating officer engaged two filters: one that identified safe files and one that identified those known to contain child pornography or other illicit material. *See* 592 F.3d 779, 781 (7th Cir. 2010). The first was called KFF Ignorable (where "KFF" stands for "Known File Filter"); the second, KFF Alert. *See id.*

53. *See supra* notes 26-27 and accompanying text. Of course, a magistrate might issue a warrant for police to use a hash search to look for illicit material. But the magistrate will only issue the warrant upon a showing of probable cause that the computer contains evidence of a crime. *See* U.S. CONST. amend. IV ("[N]o Warrants shall issue, but upon probable cause . . . ."). Here, we are concerned with the scenario where police use a hash search to search for evidence of a crime for which they have no reason to suspect the computer's owner.

54. *See infra* Part II.A.

55. *See infra* Part II.B.

56. *See infra* Part II.C.

A. Hash Searches of Publicly Exposed Information

One strategy law enforcement officers commonly use to combat child pornography is to search for illegal images and movies shared over peer-to-peer networks.[57] Officers typically begin by keyword searching for a term they know to be associated with child pornography, like "Lolitaguy"[58] or "pedo collection."[59] They collect the search results and compare the hash values from those results to those of files known to contain child pornography.[60] For any matches, they record the sharing user's IP address, which is publicly displayed by the peer-to-peer program.[61] To help expedite this process, various nonprofits and law enforcement agencies have developed tools to automate key components of these investigations, like identifying IP addresses and comparing hash values.[62] These tools access only information already visible to any member of the public using the peer-to-peer network; they simply access that information more efficiently.

Once officers match child pornography to an IP address, they typically subpoena the user's internet service provider for the user's name and address.[63] They then apply for a warrant to seize that individual's computer and search it for child pornography.[64]

Defendants have typically challenged evidence seized as a product of these peer-to-peer network investigations on two grounds. First, they've argued that law enforcement officers conduct impermissible Fourth Amendment searches

57. A peer-to-peer network allows computers to communicate with one another directly. *See* MGM v. Grokster, Ltd., 545 U.S. 913, 919-20 (2005). These networks are an alternative to traditional computer networks that use a client-server model, where client computers communicate with one another indirectly through a central server. *See id.* at 920. Peer-to-peer networks were popularized by Napster and similar services that let users share music files. *See id.* at 919-20, 924-25.

58. *See* United States v. Borowy, 595 F.3d 1045, 1046 (9th Cir. 2010) (per curiam).

59. *See* United States v. Chiaradio, 684 F.3d 265, 271 (1st Cir. 2012).

60. *See, e.g.*, *id.*

61. *See, e.g.*, *id.*

62. *See, e.g.*, United States v. Dreyer, 804 F.3d 1266, 1270 (9th Cir. 2015) (en banc) (describing the RoundUp tool developed by the Internet Crimes Against Children Task Force, a collection of federal, state, and local law enforcement officers); *Chiaradio*, 684 F.3d at 271 (describing the "enhanced peer-to-peer software," or "EP2P," developed by the Federal Bureau of Investigation (FBI)); United States v. Dodson, 960 F. Supp. 2d 689, 692-93, 696-97 (W.D. Tex. 2013) (describing Child Rescue Coalition's Child Protection System (CPS) software). For background on the CPS software, see *The Solution*, CHILD RESCUE COALITION, https://perma.cc/VM88-B87Q (archived Jan. 16, 2018).

63. For a good example of an affidavit detailing the full investigatory process from initial peer-to-peer search through subpoena, see United States v. Miknevich, 638 F.3d 178, 180-81 (3d Cir. 2011).

64. *See, e.g.*, *Chiaradio*, 684 F.3d at 271.

when they identify child pornography located on a suspect's personal computer. And second, they've argued that the fact that a given file's hash value matches that of a known piece of child pornography is an inadequate basis for probable cause. Both arguments have proved consistent losers.

When defendants have argued that these investigations are illegal searches, courts have responded that users of peer-to-peer networks have no reasonable expectation of privacy in files they expose to the public.[65] The Ninth Circuit, for example, held that a defendant had no reasonable expectation of privacy in his shared files because those files were "entirely exposed to public view; anyone with access to [the peer-to-peer network] could download and view his files without hindrance."[66] It also rejected the defendant's contention that the government's use of forensic software to conduct hash value comparisons changed this analysis.[67] "In this context," the court held, "the hash-mark analysis functioned simply as a sorting mechanism to prevent the government from having to sift, one by one, through [the defendant's] already publically exposed files."[68]

Courts have been similarly dismissive of claims that hash value comparisons are inadequate to establish probable cause. For example, in *United States v. Cartier*, the defendant argued that it's possible for two different files to have the same hash value and that no law enforcement officer, prior to getting a warrant, had actually seen child pornography on his computer; rather, police had only seen the matching hash values.[69] Thus, the argument went, hash value comparison couldn't have been enough for probable cause.[70] The district court

65. This is just an extension of the ordinary rule that although the government conducts an impermissible search when it intrudes upon a person's reasonable expectations of privacy, *see* United States v. Jones, 565 U.S. 400, 406 (2012), a person has no reasonable expectation of privacy in "objects, activities, or statements that he exposes to the 'plain view' of outsiders," California v. Ciraolo, 476 U.S. 207, 215 (1986) (quoting Katz v. United States, 389 U.S. 347, 361 (1967) (Harlan, J., concurring)).

66. United States v. Borowy, 595 F.3d 1045, 1048 (9th Cir. 2010) (per curiam); *see also* United States v. Stults, 575 F.3d 834, 843 (8th Cir. 2009) ("We hold that Stults had no reasonable expectation of privacy in files that the FBI retrieved from his personal computer where Stults admittedly installed and used LimeWire to make his files accessible to others for file sharing."); United States v. Perrine, 518 F.3d 1196, 1205 (10th Cir. 2008) (holding that a defendant who had installed peer-to-peer software had no reasonable expectation of privacy in his ISP subscriber information).

67. *See Borowy*, 595 F.3d at 1048 (per curiam).

68. *Id.*; *see also* United States v. Dodson, 960 F. Supp. 2d 689, 696-97 (W.D. Tex. 2013) ("Additionally, the software's use of root hash values to locate similar files is not something unique to the CPS program itself[;] rather, as explained previously, this information is readily available to *all* [peer-to-peer network] users.").

69. *See* 543 F.3d 442, 446 (8th Cir. 2008).

70. *See id.*

disagreed, and the Eighth Circuit affirmed.[71] Similarly, in *United States v. Miknevich*, the Third Circuit held that the combination of a matching hash value and a highly descriptive file name was enough to support a probable cause finding.[72] This was true even though the magistrate issuing the warrant never viewed the suspicious file.[73] In general, law enforcement officers aver that a hash value is "akin to a digital fingerprint" that's "more than 99.9999% reliable."[74] And courts tend to agree that this is sufficient for probable cause, regardless whether police ever view the image that generated the hash value.[75]

B. Hash Searches of Content Shared or Stored Online

A second context in which courts have encountered hash searches is when internet companies—including email providers, cloud storage companies, and social media companies—filter content for illicit material. Under 18 U.S.C. § 2258A, companies that provide electronic communication or digital storage services are required to report child pornography of which they have "actual knowledge" to the National Center for Missing and Exploited Children (NCMEC).[76] Although the statute doesn't require them to, some companies take additional steps to proactively identify child pornography shared or stored by their users.[77] For example, to identify emails that contain child pornography, some email providers compute the hash values of all attachments their users send or receive.[78] They then compare those hash values to a database of known child pornography and forward any matches to NCMEC.[79]

---

71. *See id.*

72. *See* 638 F.3d 178, 184-85 (3d Cir. 2011).

73. *See id.* at 183.

74. United States v. Collins, 753 F. Supp. 2d 804, 806 n.3 (S.D. Iowa 2009); *accord* United States v. Wellman, 663 F.3d 224, 226 n.2 (4th Cir. 2011) ("A 'hash value' is an alphanumeric string that serves to identify an individual digital file as a kind of 'digital fingerprint.'"); *Miknevich*, 638 F.3d at 185 (noting that the hash value was described in the affidavit as a "digital fingerprint"); United States v. Naylor, 99 F. Supp. 3d 638, 639 (S.D. W. Va. 2015) ("A hash value is essentially a 'digital fingerprint' unique to a particular file."); United States v. Bershchansky, 958 F. Supp. 2d 354, 357 n.3 (E.D.N.Y. 2013) (describing the hash value as a "fingerprint or digital signature" and noting that the affidavit claimed "99.9999 percent certainty" (quoting a law enforcement officer's affidavit)), *aff'd*, 788 F.3d 102 (2d Cir. 2015).

75. *See, e.g.*, *Miknevich*, 638 F.3d at 183-85.

76. *See* 18 U.S.C. § 2258A(a) (2016).

77. *See, e.g.*, United States v. Ackerman, 831 F.3d 1292, 1294 (10th Cir. 2016) (discussing AOL's screening system); *see also* Riva Richmond, *Facebook's New Way to Combat Child Pornography*, N.Y. TIMES: GADGETWISE (May 19, 2011, 6:00 AM), https://perma.cc/U54Q-K22G.

78. *See, e.g., id.*

79. *See, e.g., id.*

Email users who've been prosecuted for transacting in child pornography have frequently argued that their email providers act as government agents in screening and reporting their emails and that those providers' actions are therefore constrained by the Fourth Amendment.[80] Most courts, however, have disagreed: "A reporting requirement, standing alone, does not transform an Internet service provider into a government agent whenever it chooses to scan files sent on its network for child pornography."[81] And although several courts have held that NCMEC acts as a government entity or agent under § 2258A, they've only determined that NCMEC conducts a search when it opens and visually observes the contents of emails forwarded to it;[82] those courts haven't addressed whether NCMEC's confirming the email provider's hash search with one of its own would be a Fourth Amendment search.

C. Hash Searches of Private Information by Government Actors

Unlike the previous two scenarios, few courts have addressed whether law enforcement officers violate the Fourth Amendment when they use hashing tools to sort through private information. Indeed, federal courts have published opinions on this issue only five times.[83] Some courts that have addressed hash searches have shown confusion in distinguishing generic hashing software, which is a staple of any digital forensic investigation, from optional hashing filters designed to flag specific types of evidence like child pornography. But courts that have understood and addressed the issue directly

80. *See, e.g.*, United States v. Stevenson, 727 F.3d 826, 830 (8th Cir. 2013); *see also* Skinner v. Ry. Labor Execs. Ass'n, 489 U.S. 602, 614 (1989) ("Although the Fourth Amendment does not apply to a search or seizure, even an arbitrary one, effected by a private party on his own initiative, the Amendment protects against such intrusions if the private party acted as an instrument or agent of the Government.").

81. *E.g.*, *Stevenson*, 727 F.3d at 830; *see, e.g.*, United States v. Richardson, 607 F.3d 357, 364, 367 (4th Cir. 2010); United States v. Keith, 980 F. Supp. 2d 33, 40 (D. Mass. 2013); *see also Ackerman*, 831 F.3d at 1295 (treating AOL as a private party without considering whether it might be acting as a government agent). *But see* United States v. DiTomasso, 56 F. Supp. 3d 584, 597 (S.D.N.Y. 2014). The court in *DiTomasso* held that the defendant had consented to AOL's searching his emails as a government agent because AOL had disclosed its intent to cooperate with law enforcement in its terms of service. *Id.* Admittedly, though, it is not entirely clear whether the court in *DiTomasso* meant to decide that AOL was a government agent or only that the defendant would have no claim even if AOL were a government agent.

82. *See Ackerman*, 831 F.3d at 1308; *Keith*, 980 F. Supp. 2d at 41.

83. To identify all reported opinions on this issue, I narrowed the Westlaw universe to reported opinions issued by federal courts in criminal cases and searched for the term "hash," which returned 376 criminal cases as of January 1, 2018. I read through each case that involved digital investigations as opposed to, for example, hash oil. In only five of those cases did courts address Fourth Amendment challenges to hash searches used by government actors to search private information.

have all been troubled by hash searches for evidence outside the scope of a warrant, though none has yet held the practice outright unconstitutional.

1. *United States v. Crist*

The Middle District of Pennsylvania was the first court to directly address whether hash searches of private information implicate the Fourth Amendment.[84] While Robert Crist was being evicted from his home, his possessions were placed outside without his knowledge or consent.[85] Soon after, Seth Hipple claimed Crist's computer from the curb.[86] As Hipple was rummaging through the computer's files, he found child pornography; he then turned the computer over to law enforcement.[87] Without getting a warrant, state investigators analyzed the computer using EnCase, a type of digital forensics software; they generated hash values for all files on the computer and then compared those values to the hash values of known or suspected child pornography.[88] From this hash value analysis, investigators identified five videos containing known child pornography and another 171 containing suspected child pornography.[89] Crist was later charged with knowingly receiving and possessing digital images and video files containing child pornography.[90]

Crist moved to suppress.[91] In response, the government argued that in using the EnCase program, it "'simply ran hash values on the computer,' which does not constitute a search within the meaning of the Fourth Amendment."[92]

84. United States v. Crist, 627 F. Supp. 2d 575 (M.D. Pa. 2008).

85. *See id.* at 577.

86. *See id.* Jeremy Sell, who'd been hired by Crist's landlord to remove Crist's possessions from the house he'd been renting, had called Hipple to let him know that Crist's computer would be out on the curb. *Id.*

87. *See id.*

88. *Id.* at 576, 578.

89. *Id.*

90. *See id.* at 579.

91. *See id.* at 576.

92. *See id.* at 581 (citation omitted) (quoting the digital forensic examiner's testimony). The government first assumed that running a hash value analysis was not a search governed by the Fourth Amendment. *See* Supplemental Brief Opposing Pre-trial Motion to Suppress at 6-8, *Crist*, 627 F. Supp. 2d 575 (No. 1:07-cr-00211-YK), 2008 WL 6855527. And because that hash value analysis, in combination with the information conveyed by Hipple, made police "substantially certain" that the remaining files on the computer contained child pornography, their subsequent search (opening additional files) did not exceed the scope of Hipple's private search and therefore did not violate the Fourth Amendment. *See id.* at 8-11 (quoting United States v. Runyan, 275 F.3d 449, 463 (5th Cir. 2001)); *see also* United States v. Jacobsen, 466 U.S. 109, 115 (1984) ("[A]dditional invasions of [an individual's] privacy by the Government agent must be

footnote continued on next page

The district court disagreed. It explained that "[b]y subjecting the entire computer to a hash value analysis[,] every file, internet history, picture, and 'buddy list' became available for Government review."[93] And this was not made constitutionally permissible by the fact that investigators "didn't look at any files" and "simply accessed the computer."[94] Thus, the court held that warrantless "'running of hash values' is a search protected by the Fourth Amendment."[95]

2. *United States v. Comprehensive Drug Testing*

Next to address hash searches was the Ninth Circuit. In 2002, the federal government was investigating the use of steroids by professional baseball players.[96] At the time, Comprehensive Drug Testing, Inc. (CDT) administered Major League Baseball's drug-testing program; it maintained electronic records of players tested as well as their test results.[97] During the government's investigation, federal agents learned of ten players who had tested positive in CDT's program, and they obtained a warrant to seize those players' records.[98] When they executed the warrant, however, the agents seized and reviewed the records for hundreds of additional players.[99] They did so despite the fact that the warrant required "computer personnel" to conduct the initial review of the seized data, segregate data outside the warrant's scope, and return that extraneous data to CDT.[100]

The Ninth Circuit described the situation as "an obvious case of deliberate overreaching by the government in an effort to seize data as to which it lacked probable cause."[101] It therefore ordered the government to return the data it had seized.[102] The Ninth Circuit concluded its per curiam opinion with a general admonition: "The process of segregating electronic data that is seizable

---

tested by the degree to which they exceeded the scope of the private search."); *Runyan*, 275 F.3d at 464-65 (applying the private search reconstruction doctrine to digital storage media).

93. *Crist*, 627 F. Supp. 2d at 585.

94. *Id.* (quoting the hearing transcript).

95. *See id.* at 585-87.

96. *See* United States v. Comprehensive Drug Testing, Inc. (*Comprehensive Drug Testing II*), 621 F.3d 1162, 1166 (9th Cir. 2010) (en banc) (per curiam).

97. *See id.*

98. *Id.*

99. *Id.*

100. *See id.* at 1171.

101. *Id.* at 1172.

102. *See id.* at 1174.

from that which is not must not become a vehicle for the government to gain access to data which it has no probable cause to collect."[103]

But a five-judge concurrence in that case[104] has left more of an impression on other courts' decisions about hash searches.[105] That concurrence recommended to magistrate judges several restrictions they ought to impose on warrants for electronic data.[106] One of these restrictions concerned the government's "sophisticated hashing tools"[107]: "These and similar search tools should not be used without specific authorization in the warrant, and such permission should only be given if there is probable cause to believe that [well-known illegal files those tools identify] can be found on the electronic medium to be seized."[108]

3. *In re United States's Application for a Search Warrant*

Subsequently, a magistrate judge in the Western District of Washington relied on the *Comprehensive Drug Testing* concurrence to deny the government's warrant application.[109] The government proposed in its affidavit to use hash values during its search of the digital devices of a man suspected of copyright infringement and trafficking in counterfeit software.[110] The magistrate noted that hash values can be helpful in segregating files irrelevant to the government's investigation.[111] But "they can also be used to search and find evidence outside the scope of the warrant automatically and systematically."[112] He therefore required the government to add the following language to its

103. *Id.* at 1177.

104. *See id.* at 1178 (Kozinski, C.J., concurring).

105. *See, e.g.*, *In re* U.S.'s Application for a Search Warrant to Seize & Search Elec. Devices from Edward Cunnius, 770 F. Supp. 2d 1138, 1148-49 (W.D. Wash. 2011) (applying several of the concurrence's recommendations).

106. *See Comprehensive Drug Testing II*, 621 F.3d at 1179-80 (Kozinski, C.J., concurring). The recommended restrictions are: (1) the government must "waive reliance upon the plain view doctrine"; (2) "[s]egregation and redaction of electronic data must be done either by specialized personnel or an independent third party"; (3) "[w]arrants and subpoenas must disclose the actual risks of destruction of information as well as prior efforts to seize that information in other judicial fora"; (4) "[t]he government's search protocol must be designed to uncover only the information for which it has probable cause"; and (5) "[t]he government must destroy or, if the recipient may lawfully possess it, return non-responsive data." *Id.* at 1180.

107. *See id.* at 1179.

108. *Id.*

109. *See In re U.S.'s Application for a Search Warrant*, 770 F. Supp. 2d at 1139, 1148-49.

110. *See id.* at 1139, 1152.

111. *See id.* at 1152.

112. *Id.*

application: "[T]hese methodologies, techniques and protocols will not include the use of 'hash value' libraries to search the electronically stored information for items that are not set forth in the items authorized to be seized . . . [by] this warrant."[113]

4. *United States v. Mann*

Soon after *Comprehensive Drug Testing*, the Seventh Circuit also addressed hash searches. Matthew Mann was an Indiana lifeguard instructor who installed a video camera in the women's locker room at the pool where he worked.[114] One of the female lifeguard students discovered the camera, rewound the tape, and identified Mann.[115] She turned the tape over to police, who got a warrant to seize Mann's computers and to search them for voyeuristic content.[116]

Police used software called Forensic Toolkit (FTK) to investigate Mann's computers.[117] Like the EnCase software in *Crist*,[118] FTK first generates a hash value for each file on the computer.[119] And again like EnCase, FTK offers several optional filters for segregating certain types of files.[120] The investigating officer in *Mann* used two of those filters: KFF Alert and KFF Ignorable.[121] KFF Ignorable is designed to isolate standard files known to be irrelevant to an investigation, like those constituting a computer's operating system.[122] KFF Alert, however, flags files known to contain illicit material, including many containing child pornography.[123] In Mann's case, FTK identified four KFF Alert files and flagged an additional 677 thumbnails.[124] The investigating officer proceeded to open those files and found "many, many images of child pornography."[125] Mann challenged both the use of FTK and the opening of individual files as having exceeded the scope of the warrant.[126]

---

113. *Id.*

114. United States v. Mann, 592 F.3d 779, 780 (7th Cir. 2010).

115. *Id.*

116. *See id.* at 780-81.

117. *Id.* at 781.

118. *See supra* text accompanying note 88.

119. *See Mann*, 592 F.3d at 781.

120. *See id.*; *supra* text accompanying note 88.

121. *See Mann*, 592 F.3d at 781.

122. *See* ACCESSDATA, *supra* note 31, at 29.

123. *See Mann*, 592 F.3d at 781.

124. *Id.*

125. *Id.*

126. *See id.* at 781-82.

The Seventh Circuit held that there was "no reason to believe that [the investigator] exceeded the scope of the warrant by employing the FTK software without more."[127] The court reasoned that Mann could have had voyeuristic images anywhere on his computers.[128] And the FTK software had been useful in indexing and cataloging his files into a more easily searchable format.[129] But the court did not address whether use of the KFF Alert filter specifically—as opposed to FTK generally—constituted a search.

The court did hold, however, that the investigator had exceeded the scope of the warrant when he opened and visually inspected the four KFF Alert files.[130] That filter had compared Mann's images to those in a database of known child pornographic images, whereas the investigator's warrant authorized him to look only for images that Mann himself had captured through his locker room camera.[131] Thus he should have known that any flagged files would fall outside the scope of the warrant.[132] Accordingly, the court suppressed those images.[133]

The *Mann* court also addressed whether magistrates in the Seventh Circuit ought to adopt the Ninth Circuit's recommendations from *Comprehensive Drug Testing*.[134] Its answer was clear: "We . . . reject Mann's suggestion that we take our cue from the more comprehensive rules regarding computer searches recently outlined by the Ninth Circuit."[135] Nevertheless, as the next case shows, not all district courts within the Seventh Circuit have been as untroubled by hash searches as the *Mann* court.

---

127. *Id.* at 784.

128. *Id.*

129. *See id.*

130. *See id.* at 784-85.

131. *See id.* at 784.

132. *See id.*

133. *See id.* at 785. The court nevertheless held the flagged files to be "severable" from the remaining files and thus upheld Mann's conviction. *Id.*

134. At the time the Seventh Circuit decided *Mann*, these recommendations were part of the en banc Ninth Circuit's majority opinion. *See* United States v. Comprehensive Drug Testing, Inc. (*Comprehensive Drug Testing I*), 579 F.3d 989, 1006 (9th Cir. 2009) (en banc), *revised and superseded per curiam by* 621 F.3d 1162 (9th Cir. 2010) (en banc). The Ninth Circuit subsequently reissued that opinion, and the recommendations were relegated to a five-judge concurrence. *See Comprehensive Drug Testing II*, 621 F.3d 1162, 1179-80 (9th Cir. 2010) (en banc) (Kozinski, C.J., concurring).

135. *Id.* at 785 (citing *Comprehensive Drug Testing I*, 579 F.3d at 993, 1000).

5. *United States v. Schlingloff*

The fifth case to consider hash searches was decided in the Central District of Illinois, where *Mann* is binding precedent.[136] The Seventh Circuit's opinion in *Mann*, however, had not clearly distinguished between FTK software, which can be used to catalog a computer's contents, and the KFF Alert tool, an optional component of FTK software that can be used to identify only contraband.[137] If, for example, the investigating officer should have known that any KFF Alert files would fall outside the scope of his warrant, why wasn't his using the KFF Alert filter in the first place an impermissible search? It's possible that the court didn't answer this question because it didn't recognize that the KFF Alert filter was optional and that police had affirmatively chosen to apply it. In any case, the *Schlingloff* court was forced to grapple with this ambiguity.

The facts of *Schlingloff* are straightforward. Law enforcement got a warrant to search a residence in Rock Island, Illinois for evidence of passport fraud and harboring an alien; the warrant's affidavit indicated that there was reason to believe that computers found in the residence would contain evidence of those crimes.[138] Christopher Schlingloff's computer was seized and searched using FTK software—the same software at issue in *Mann*.[139] The investigating officer enabled the KFF Alert filter, and the hash search flagged two child pornography files, which the officer briefly opened.[140] Schlingloff was later prosecuted for possession of child pornography, and he moved to suppress the files.[141]

The court began by discarding Schlingloff's argument that the government's use of the FTK software alone had exceeded the scope of its warrant; that argument had already been rejected by the *Mann* court.[142] But Schlingloff also argued that even if the government's use of FTK was permissible, nevertheless its use of the KFF Alert filter in a case not involving child pornography exceeded the warrant's scope.[143]

The court acknowledged that there was a difference between the FTK software, used to catalog evidence, and the KFF filter, used to search only for

136. United States v. Schlingloff, 901 F. Supp. 2d 1101 (C.D. Ill. 2012), *appeal dismissed per stipulation*, No. 12-3661 (7th Cir. Mar. 4, 2013).

137. *See id.* at 1103-04.

138. *See id.* at 1102.

139. *See id.* at 1102-03; *supra* text accompanying notes 117-23.

140. *See Schlingloff*, 901 F. Supp. 2d at 1102.

141. *See id.* at 1103.

142. *See id.*

143. *See id.* at 1103-04.

evidence of a crime outside the warrant's scope.[144] It explained that "in light of the admitted ability to confine the FTK search by not enabling the KFF filter for child pornography alerts," the government's investigator had taken an "affirmative additional step to enable the KFF alerts that would identify known child pornography files as part of his search for evidence of passport fraud or identity theft."[145] Because the subject matter of the search warrant "bore no resemblance to child pornography, it [was] difficult to construe" the investigator's use of KFF's targeted hash search as "anything other than a deliberate expansion of the scope of the warrant."[146]

Still, despite its clear discomfort with the KFF filter, the court didn't go so far as to hold that the warrantless use of that filter alone violated the Fourth Amendment.[147] Instead, three factors "in conjunction" with one another were enough to convince the court that the government had exceeded its warrant: (1) the officer had "affirmatively enabl[ed] the KFF filter" in (2) "a non-pornography case," and the officer had (3) "open[ed] the files once alerted to their presence."[148]

D. Takeaways

From these cases we can derive several conclusions about how courts have treated hash searches. To begin, courts have consistently acknowledged the evidentiary significance of hash value analysis: When the hash values of two files match, those files are almost certainly the same.[149] Additionally, courts haven't identified any legal problems with the use of hash searches by private parties or by the government when the information being searched is publicly available.[150]

Courts have disagreed, however, on whether a hash search, either in the absence of a warrant or to look for evidence outside the scope of a warrant, poses a Fourth Amendment problem. The *Crist* court, for example, held that the mere use of hash search software, without visual inspection, was a search.[151] The *Mann* court, in turn, held that police did not exceed the scope of their warrant when they used hash search software; rather, they exceeded their

144. *See id.* at 1104-05.

145. *Id.* at 1105.

146. *Id.*

147. *See id.* at 1106 ("The use of the KFF alerts alone may not move this case beyond the scope of the warrant . . . .").

148. *See id.*

149. *See supra* Part I.A.1; *see also supra* note 74 and accompanying text.

150. *See supra* Parts II.A-.B.

151. *See* United States v. Crist, 627 F. Supp. 2d 575, 585 (M.D. Pa. 2008).

warrant's scope only when they viewed the files the hash search software had identified.[152] And while the *Schlingloff* court suggested that police exceeded their warrant's scope when they used hash search software to look for child pornography during a passport fraud investigation, the court ultimately relied on their having done so in conjunction with opening and viewing the files.[153] Meanwhile, *Comprehensive Drug Testing II* and *In re United States's Application for a Search Warrant* suggest that magistrates should prohibit police from using hash searches unless explicitly authorized.[154] But neither case states that using hash searches without explicit authorization would violate the Fourth Amendment or specifies whether hash searches are problematic even if police never look at the files those hash searches identify.

Not only have these courts disagreed as to whether and when the use of a hash search exceeds the scope of a warrant, but none has proposed a framework for thinking about how the Fourth Amendment governs hash searches. I turn now to one framework commentators have proposed: treating hash searches as digital dog sniffs.

III. Hash Searches as Digital Dog Sniffs

In 2006, Richard Salgado authored what remains the definitive legal analysis of hash searches: *Fourth Amendment Search and the Power of the Hash.*[155] His article has been frequently cited by both courts and commentators.[156] In that article, he proposed analyzing hash searches for child pornography under the Supreme Court's dog sniff cases.[157] Like the drugs at issue in those cases, child pornography is illegal to possess.[158] And like the dogs in those cases, hash

152. *See* United States v. Mann, 592 F.3d 779, 784-85 (7th Cir. 2010).

153. *See Schlingloff*, 901 F. Supp. 2d at 1106.

154. *See Comprehensive Drug Testing II*, 621 F.3d 1162, 1179 (9th Cir. 2010) (en banc) (Kozinski, C.J., concurring); *In re* U.S.'s Application for a Search Warrant to Seize & Search Elec. Devices from Edward Cunnius, 770 F. Supp. 2d 1138, 1152-53 (W.D. Wash. 2011).

155. Salgado, *supra* note 3.

156. *See, e.g.*, United States v. Ackerman, 831 F.3d 1292, 1294 (10th Cir. 2016); United States v. Ganias, 824 F.3d 199, 235 (2d Cir.) (en banc) (Chin, J., dissenting), *cert. denied*, 137 S. Ct. 569 (2016); Timothy C. MacDonnell, *Orwellian Ramifications: The Contraband Exception to the Fourth Amendment*, 41 U. MEM. L. REV. 299, 345 n.331 (2010); Burrows, *supra* note 14, at 258 n.26; Michael Mestitz, Note, *Unpacking Digital Containers: Extending* Riley*'s Reasoning to Digital Files and Subfolders*, 69 STAN. L. REV. 321, 352 n.174 (2017).

157. *See* Salgado, *supra* note 3, at 44-46.

158. *See id.*; *see also, e.g.*, 18 U.S.C. §§ 2252(a)(4), 2252A(a)(5) (2016) (prohibiting possession of child pornography within federal or tribal jurisdictions, on federal property, or with a link to interstate or foreign commerce); CAL. PENAL CODE §§ 311.1-.12 (West 2017) (defining child pornography offenses).

searches can be trained (that is, programmed) to reveal only illicit material.[159] Thus, Salgado suggested that these cases seem to "allow for the routine use by government of hash-based contraband detection in any search of a digital storage device, regardless of the scope of the search authority."[160]

This Part first summarizes the Court's binary search cases. It then describes in greater detail the argument by Salgado and others that the use of hash searches for contraband like child pornography is equivalent to a digital dog sniff. Finally, it explains that even if courts were to treat hash searches as digital dog sniffs, they would nevertheless be unconstitutional under the Court's reasoning in *Florida v. Jardines*.[161]

A. Binary Search Doctrine

The Court's binary search doctrine stems from *United States v. Place*.[162] In that case, the Court held that "[a] 'canine sniff' by a well-trained narcotics detection dog" of a suspect's luggage located in a public place "did not constitute a 'search' within the meaning of the Fourth Amendment."[163] This was true for two reasons. First, unlike an officer's rummaging through the luggage, the dog sniff didn't "expose noncontraband items"; it was thus "much less intrusive than a typical search."[164] And second, "the information obtained [was] limited" because "the sniff disclose[d] only the presence or absence of narcotics."[165] Because of these two factors, the Court described the dog sniff as "*sui generis*," stating that it was "aware of no other investigative procedure that is so limited both in the manner in which the information is obtained and in the content of the information revealed."[166]

It might be tempting to argue that the Court's description of dog sniffs as sui generis limits *Place*'s rule to canines. The Court, however, waited only a year before acknowledging that dog sniffs weren't unique. Citing *Place*, it held in *United States v. Jacobsen* that "[a] chemical test that merely discloses whether or not a particular substance is cocaine does not compromise any legitimate interest in privacy."[167] Justice Kagan recently restated the point that dog sniffs

159. *See* Salgado, *supra* note 3, at 44-46.

160. *See id.* at 45-46.

161. 133 S. Ct. 1409 (2013).

162. 462 U.S. 696 (1983).

163. *Id.* at 707.

164. *Id.*

165. *Id.*

166. *Id.*

167. *See* 466 U.S. 109, 123 (1984).

are not, in fact, sui generis.[168] Comparing dog sniffs to binoculars, she wrote that a drug-sniffing dog is just another "specialized device for discovering objects not in plain view."[169]

B. Digital Dog Sniffs

The analogy from drug-sniffing dogs to hash searches for digital contraband is appealing for several reasons. To take the most prominent example of digital contraband, investigators have frequently used hash searches to look for child pornography.[170] Like *Place*'s dog sniff, police can use targeted hash searches that reveal only the presence or absence of child pornography.[171] And like illicit drugs, "Congress has decided . . . to treat the interest in 'privately' possessing" child pornography "as illegitimate."[172] Thus, the argument goes, because the hash search "reveals no information other than the location of [material] that no individual has any right to possess," it "does not violate the Fourth Amendment."[173]

It's true that hash searches for child pornography reveal no information other than the presence or absence of illicit material. Hash searches, like dog sniffs, provide information in binary: Either *yes*, the hash value of some file on a suspect's computer matches the hash value of some known piece of child pornography, or *no*, it does not. And even though accuracy is relevant not to whether a given technique is a search but only to whether it's sufficient to establish probable cause,[174] it's also true that hash searches are highly accurate. The odds of two files producing the same hash value are "infinitesimally small,"[175] and police frequently avow that hash value comparisons are "more than 99.9999% reliable."[176] Dogs, by contrast, generate very high rates of false

168. *See* Florida v. Jardines, 133 S. Ct. 1409, 1418 (2013) (Kagan, J., concurring).

169. *Id.*

170. *See supra* Parts II.A-.C.

171. One example is the KFF Alert filter at issue in *United States v. Mann*. *See* 592 F.3d 779, 781 (7th Cir. 2010).

172. *See Jacobsen*, 466 U.S. at 123; *see also* 18 U.S.C. §§ 2252(a)(4), 2252A(a)(5) (2016).

173. *See* Illinois v. Caballes, 543 U.S. 405, 410 (2005); *see also* Salgado, *supra* note 3, at 44-46.

174. *Cf.* Florida v. Harris, 133 S. Ct. 1050, 1058 (2013) (discussing the role a drug-sniffing dog's accuracy plays in the probable cause determination). Nothing in *Harris*'s probable cause analysis suggests that there is a point at which a dog's drug sniff is so inaccurate as to constitute a search. *See infra* note 230.

175. Salgado, *supra* note 3, at 39 n.6 (citing SCHNEIER, *supra* note 15, at 429 (2d ed. 1996)).

176. *See* United States v. Collins, 753 F. Supp. 2d 804, 806 n.3 (S.D. Iowa 2009); *supra* note 74 and accompanying text.

positives.[177] In terms of accuracy, then, hash searches are like dog sniffs but even better.

Many Fourth Amendment commentators have for these reasons found much promise in hash searches.[178] They seem to be both minimally intrusive and maximally accurate. As the next Subpart demonstrates, however, even if we accept the analogy to dog sniffs, hash searches violate the Fourth Amendment.

C. Digital Dog Sniffs Under *Florida v. Jardines*

Recall the case with which we're primarily concerned: when police have a warrant (or exception[179]) to search for evidence of one crime, such as tax fraud, and they use hash filters to search for evidence, such as child pornography, of some unrelated crime. Under *Jardines*, this use of hashing is an impermissible search.

The ability of police to use a drug-sniffing dog is not without limits. In *Jardines*, the government argued that *Place* and *Jacobsen* meant that "investigation by a forensic narcotics dog by definition cannot implicate any legitimate privacy interest."[180] The Court disagreed, holding that "[t]he government's use of trained police dogs to investigate the home and its immediate surroundings [was] a 'search' within the meaning of the Fourth Amendment."[181] This was because the police brought the dog onto a "constitutionally protected area"—the home's curtilage—"through an unlicensed physical intrusion."[182]

---

177. *See Caballes*, 543 U.S. at 412 (Souter, J., dissenting) (noting that according to one study, "dogs in artificial testing situations return false positives anywhere from 12.5% to 60% of the time"); *see also* United States v. Bentley, 795 F.3d 630, 632, 635 (7th Cir. 2015) (upholding a finding of probable cause to search a vehicle based in part on an alert from a drug-sniffing dog that alerted "93% of the time he [was] called to do an open-air sniff of a vehicle" but was accurate only 59.5% of the time).

178. *See, e.g.*, Salgado, *supra* note 3, at 46; Burrows, *supra* note 14, at 256-57; *see also* Note, *Data Mining, Dog Sniffs, and the Fourth Amendment*, 128 HARV. L. REV. 691, 708-10 (2014) (arguing for applying the limits on dog sniffs to crime-sniffing algorithms that mine large datasets).

179. *See* United States v. Karo, 468 U.S. 705, 717 (1984) ("Warrantless searches are presumptively unreasonable, though the Court has recognized a few limited exceptions to this general rule."). The *Karo* Court cited three examples: *United States v. Ross*, 456 U.S. 798 (1982) (automobiles); *Schneckloth v. Bustamonte*, 412 U.S. 218 (1973) (consent); and *Warden v. Hayden*, 387 U.S. 294 (1967) (exigent circumstances).

180. *See* Florida v. Jardines, 133 S. Ct. 1409, 1417 (2013); *see also* Brief for the State of Florida at 11, *Jardines*, 133 S. Ct. 1409 (No. 11-564), 2012 WL 1594294.

181. *Jardines*, 133 S. Ct. at 1417-18.

182. *See id.* at 1414-15. The curtilage is "the area 'immediately surrounding and associated with the home.'" *Id.* at 1414 (quoting Oliver v. United States, 466 U.S. 170, 180 (1984)).

Although police generally enjoy an implied license to conduct a knock and talk, the police in *Jardines* exceeded that license when they brought a trained dog to the door to execute a sniff.[183] The Court noted that "[t]he scope of a license—express or implied—is limited not only to a particular area but also to a specific purpose."[184] Thus, the knocker on a front door invites "the visitor to approach the home by the front path, knock promptly, wait briefly to be received, and then (absent invitation to linger longer) leave."[185] This implied license does not, however, permit "a trained police dog to explore the area around the home in hopes of discovering incriminating evidence."[186] Because that purpose—discovering incriminating evidence—is outside the scope of the implied license, a police officer whose dog performs a drug sniff thereby conducts a search subject to the protections of the Fourth Amendment.[187]

The same limitation applies to hash searches. Like an implied license to approach a person's front door, a warrant is limited "not only to a particular area but also to a specific purpose."[188] In the context of computer searches, this means that police can search only the places identified in the warrant (the computer or other digital storage device), and for only a specific purpose—to find evidence of the crime identified in the warrant. Just as police conducting a dog sniff take action inconsistent with the purpose of their implied license to approach a person's door, so too do police conducting a hash search for evidence unrelated to the crime identified in their warrant take action inconsistent with the purpose of their warrant. Thus, like the officers in *Jardines*, police who conduct a hash search for evidence outside the scope of their warrant conduct an illegal search.[189]

There are two counterarguments. First, it may be that the *Jardines* rule is limited to the home and does not extend to computers. Second, it may be that the rule is limited to implied licenses and does not extend to warrants. Neither argument withstands scrutiny.

The argument that the *Jardines* rule is limited to the home is easily disposed of. It is true that the Court has frequently emphasized that "the right of a man

183. *See id.* at 1415-17.

184. *Id.* at 1416.

185. *Id.* at 1415.

186. *Id.* at 1416.

187. *See id.* at 1413, 1416-18.

188. *See id.* at 1416; *see also infra* text accompanying notes 193-95.

189. *See* Horton v. California, 496 U.S. 128, 140 (1990) ("If the scope of the search exceeds that permitted by the terms of a validly issued warrant or the character of the relevant exception from the warrant requirement, the subsequent seizure is unconstitutional without more."); *see also* Mincey v. Arizona, 437 U.S. 385, 393 (1978) ("[A] warrantless search must be 'strictly circumscribed by the exigencies which justify its initiation' . . . ." (quoting Terry v. Ohio, 392 U.S. 1, 26 (1968)).

to retreat into his own home and there be free from unreasonable government intrusion" stands "[a]t the very core" of the Fourth Amendment.[190] Nevertheless, in *Jardines* itself the Court extended its reasoning outside the home. After stating that "[t]he scope of a license—express or implied—is limited not only to a particular area but also to a specific purpose," the Court elaborated on that point with an example not from the home but from a traffic stop.[191] It explained: "Consent at a traffic stop to an officer's checking out an anonymous tip that there is a body in the trunk does not permit the officer to rummage through the trunk for narcotics."[192]

Nor is the *Jardines* rule limited to implied licenses. To the contrary, the Court has held that the actions police take pursuant to a warrant must also fall within the purpose of that warrant. In *Wilson v. Layne*, federal marshals and local police officers brought a reporter and photographer from the Washington Post with them to arrest a suspect, pursuant to a warrant, in what they thought was his home.[193] The Court observed that although every police action within a home need not "be explicitly authorized by the text of the warrant," "the Fourth Amendment does require that police actions in execution of a warrant be related to the objectives of the authorized intrusion."[194] Because the police's bringing members of the media with them did not aid their executing the warrant, that action violated the Fourth Amendment.[195] The same would presumably be true if police brought a drug-sniffing dog with them on a search or arrest not implicating drugs.

Not only is *Jardines* not exclusively concerned with homes or implied licenses; *Jardines* doesn't even make new law. Rather, it reiterates a well-established principle of the Fourth Amendment: When police take action that is objectively inconsistent with the purpose of some mechanism that permits them to conduct a search, they exceed the scope of that permissive mechanism[196] and thereby conduct an illegal search in violation of the Fourth Amendment.[197]

---

190. *See, e.g.*, Kyllo v. United States, 533 U.S. 27, 31 (2001) (quoting Silverman v. United States, 365 U.S. 505, 511 (1961)).

191. *See Jardines*, 133 S. Ct. at 1416.

192. *Id.*

193. 526 U.S. 603, 606-07 (1999), *abrogated in other part by* Pearson v. Callahan, 555 U.S. 223 (2009).

194. *Id.* at 611.

195. *Id.* at 614.

196. By permissive mechanism, I mean some justification—either a warrant or an exception to the warrant requirement—for police action that would in the absence of that justification constitute an unreasonable search or seizure.

197. *See, e.g.*, *Wilson*, 526 U.S. at 611, 614; Arizona v. Hicks, 480 U.S. 321, 325 (1987).

Consider the implied license at issue in *Jardines*. It permits police to do something the Fourth Amendment would otherwise prohibit: enter a suspect's curtilage to ask him questions about an investigation.[198] Other permissive mechanisms would also allow police to enter the suspect's curtilage. If the suspect were a felon in flight, police could enter his curtilage.[199] Or if the suspect gave his permission.[200] Or if police had a warrant.[201]

Indeed, this is just what it means to say that warrantless searches are presumptively unreasonable: Without a warrant or some exception to the warrant requirement, the search violates the Fourth Amendment.[202] An implied license is no different from any other exception to the warrant requirement.

And the test for determining whether police exceeded the scope of their implied license in *Jardines* is the same test courts use to determine whether police exceed the scope of their permissive mechanisms elsewhere. The *Jardines* Court made clear that its test for whether an officer's conduct exceeds the scope of her implied license is, consistent with the rest of Fourth Amendment doctrine, an objective one.[203] Because the *Jardines* officer's "behavior objectively reveal[ed] a purpose to conduct a search, which is not what anyone would think he had license to do," it was a search within the meaning of the Fourth Amendment.[204]

Compare *Jardines* to *Arizona v. Hicks*.[205] In that case, a bullet was fired through the floor of an apartment, injuring someone in the apartment below.[206] Police entered the apartment from which the bullet came to search for the shooter and weapons; they found and seized three weapons.[207] But while they were looking for those weapons, one officer noticed expensive-

198. *See Jardines*, 133 S. Ct. at 1414-16.

199. *See* Warden v. Hayden, 387 U.S. 294, 298 (1967).

200. *See* Schneckloth v. Bustamonte, 412 U.S. 218, 219 (1973).

201. *See* Payton v. New York, 445 U.S. 573, 586 (1980).

202. *See supra* note 179.

203. *See Jardines*, 133 S. Ct. at 1416-17. The Court rejected Florida's argument that looking to the officer's purpose is inconsistent with *Ashcroft v. al-Kidd*, 563 U.S. 731 (2011), and *Whren v. United States*, 517 U.S. 806 (1996), explaining that those cases held that an objectively reasonable stop or search is not vitiated by an unrelated purpose. *See Jardines*, 133 S. Ct. at 1416-17. In *Jardines*, by contrast, the question was whether the police conduct was objectively reasonable in the first place. *See id.* That question "depend[ed] upon the purpose for which [the police] entered." *Id.* And their purpose was "objectively reveal[ed]" by "their behavior." *Id.* at 1417.

204. *Jardines*, 133 S. Ct. at 1416-17.

205. 480 U.S. 321 (1987).

206. *Id.* at 323.

207. *Id.*

looking stereo equipment.[208] He moved a turntable to check the equipment's serial numbers.[209] When he phoned that serial number in to headquarters, he determined that the equipment was stolen.[210]

The Court held that moving the turntable to view and copy its serial number was an illegal search.[211] The intrusion into the apartment had been justified by exigent circumstances.[212] But by "taking action . . . unrelated to the objectives of the authorized intrusion," the police exceeded the scope of their permissive mechanism and thereby violated the Fourth Amendment.[213]

Thus, in *Hicks*, just as in *Jardines*, the Court asked whether officers' behavior, viewed objectively, was related to the goals or purpose of the permissive mechanism that allowed them to enter a constitutionally protected space. In *Hicks*, that mechanism was exigent circumstances, while in *Jardines*, it was the implied license. But the test in both cases was the same. It is also the same for consent searches (recall the Court's trunk-search example from *Jardines*).[214] And as the Court made clear in *Wilson*, the test is the same for searches pursuant to a warrant as well.[215]

So too, then, with hash searches. Consider the officers in *Schlingloff*.[216] Their warrant authorized them to search for evidence of passport fraud and harboring an alien.[217] When they enabled the KFF Alert filter, which identifies child pornography,[218] they took action that, viewed objectively, was inconsistent with the purpose of their warrant. They therefore exceeded the scope of that warrant, violating the Fourth Amendment. This is true whether we think of hash searches like dogs, as in *Jardines*, or like humans, as in *Hicks* or *Wilson*.

IV. Hash Searches as a Gateway to General Warrants

The previous Part showed that hash searches, when used to uncover evidence of a crime unrelated to the one identified in a warrant, are prohibited

208. *Id.*

209. *Id.*

210. *Id.*

211. *See id.* at 324-26.

212. *See id.* at 325.

213. *See id.*

214. *See* Florida v. Jardines, 133 S. Ct. 1409, 1416 (2013).

215. *See* Wilson v. Layne, 526 U.S. 603, 611 (1999), *abrogated in other part by* Pearson v. Callahan, 555 U.S. 223 (2009).

216. *See supra* text accompanying notes 136-48.

217. *See* United States v. Schlingloff, 901 F. Supp. 2d 1101, 1102 (C.D. Ill. 2012).

218. *See id.*

by the Fourth Amendment. This argument, however, does not answer the question whether hash searches of private data for criminal evidence are always Fourth Amendment searches. Although police may violate the Fourth Amendment when they conduct a dog sniff on a home's front porch without the homeowner's consent, they do not violate the Fourth Amendment when they conduct a dog sniff in a public place.[219] Maybe there are similar contexts in which the government has not sought a warrant but may nevertheless use hash searches to determine whether private data contains criminal evidence without violating the Fourth Amendment.

Consider, for example, the hash searches email providers use to screen attachments for child pornography.[220] Suppose the government imposed its own screen between email users and providers. If the government conducted a hash search on each email attachment sent or received and then sought a warrant if an attachment's hash value matched that of a file in a database of known contraband, would such a program violate the Fourth Amendment? If, like a dog sniff, a hash search is not a Fourth Amendment search (though the use of either to uncover evidence outside the scope of a warrant, license, or other permissive mechanism might sometimes violate the Fourth Amendment), then this program would seem to be permissible.

This Part shows that a Fourth Amendment that permits hash searches of private data for criminal evidence would also permit algorithmic searches for evidence of all manner of criminal wrongdoing. Such algorithmic searches would constitute the modern equivalent of general warrants. Because general warrants are prohibited by the Fourth Amendment,[221] hash searches of private data for criminal evidence, absent a warrant, must not be constitutionally permissible.

A. Probabilistic Algorithms

Recall that when investigators conduct a hash search for child pornography, their forensic software generates a hash value for each file on the hard drive and then compares that hash value to a database of known child pornography images.[222] If the file's hash value matches that of a file in the

219. *See* United States v. Place, 462 U.S. 696, 707 (1983).

220. *See supra* text accompanying notes 76-79.

221. *See* Riley v. California, 134 S. Ct. 2473, 2494 (2014) ("Our cases have recognized that the Fourth Amendment was the founding generation's response to the reviled 'general warrants' and 'writs of assistance' of the colonial era, which allowed British officers to rummage through homes in an unrestrained search for evidence of criminal activity.").

222. *See supra* Part I.B.3.

database, the software flags the file.[223] The information returned is binary—"safe" or "dangerous"—and extremely accurate.[224]

Now consider a slightly different piece of software. This software has been designed not to determine whether files match known child pornography images through hash value comparisons but to give a probabilistic assessment of how likely each file is to contain child pornography. The software does so by examining each image, pixel by pixel, and looking for patterns similar to those present in files known to contain illicit images.

This software exists and is being used. FTK (the software used in *United States v. Mann*[225]) provides investigators with a tool called Explicit Image Detection (EID).[226] This search algorithm "has been trained on a library of approximately 30,000 actual pornographic images."[227] The algorithm "looks for flesh tone colors" and other patterns indicative of child pornography; each image is then "automatically scored from 0 to 100" based on its "potential to be pornographic."[228]

There does not seem to be any reason the Fourth Amendment would treat hash searches and probabilistic algorithms differently. Software like EID differs from a hash search in that it is much more likely to produce false positives. A hash search can return only one of two probabilities: 0% or 100% (that is, the hash search signals either that two files match or that they do not match, but never that they *might* match). EID, meanwhile, can return any number in between.[229] But this has no bearing on whether using EID is a search: The accuracy of an investigative technique is relevant to whether that technique is sufficient to establish probable cause, not to whether it is a search in the first place.[230] EID, or any other technology that returns only a

223. *See supra* note 52 and accompanying text.

224. *See supra* note 74 and accompanying text (collecting cases in which courts have described hash values as "akin to a digital fingerprint" and highly accurate).

225. *See* 592 F.3d 779, 781 (7th Cir. 2010); *supra* text accompanying notes 117-23.

226. AccessData, Inc., FTK 3: Explicit Image Detection (2009), https://perma.cc/49R3-NGAW.

227. *Id.*

228. *Id.*

229. *See id.* (depicting a range of probability scores, including 0, 11, 56, and 99).

230. *Cf.* Florida v. Harris, 133 S. Ct. 1050, 1058 (2013). In *Harris*, the defendant challenged a dog's positive alert as insufficient to establish probable cause: Because the dog had not been trained to alert to the type of drugs police found on the defendant, its alert must have been a false positive. *See id.* at 1053-54, 1058-59. The Court declined to establish an accuracy threshold or to require certain credentials, holding instead that courts should examine "all the facts surrounding a dog's alert" to determine whether those facts are sufficient to establish probable cause for a search. *See id.* at 1058. The Court never suggested that inaccuracy would somehow transform a dog's sniff into a Fourth

footnote continued on next page

probabilistic assessment about whether a file contains contraband, is thus likely to be just as constitutionally permissible as hash searches.

B. General Crime Detection Algorithms

Consider now a different piece of software similar to EID. Instead of providing a probabilistic assessment for each individual file, this software gives a probabilistic assessment of the likelihood that a computer as a whole contains evidence of a given crime.[231] So for example, the software might notice that "al Qaeda" occurs unusually often in your personal correspondence and that among the financial records stored on your computer are several shipping receipts to locations in the Middle East. The software returns a 36% chance that you've been providing material support for terrorism. Police get a warrant to search your computer for evidence of that crime. When they read the letters, it turns out you've merely been writing to your son, who's deployed with the Marines.

Again, police departments are already using software to predict specific types of crime.[232] As these algorithms become more sophisticated, it is easy to imagine a general crime detection algorithm that can scan a computer for evidence of all manner of crimes. As with the probabilistic algorithm in Part IV.A, although such a crime detection algorithm would be more powerful than a hash search, that algorithm would not be any more intrusive than a hash search. The Fourth Amendment is thus likely to treat both algorithms the same.

It is helpful to think about how such a crime detection algorithm might function relative to an algorithm performing a hash search. Both algorithms scan through the files on your computer one by one. The hash search algorithm asks, for each file: Does the hash value of this file match the hash value of some known piece of contraband? If so, the algorithm flags that file as "dangerous"; if not, it is marked "safe." The crime detection algorithm performs

---

Amendment search; rather, a positive alert from a highly inaccurate dog would simply be less likely to establish probable cause. *See id.*

231. *See generally* Note, *supra* note 178 (proposing a constitutional framework for analyzing algorithmic risk assessments).

232. *See* Andrew Guthrie Ferguson, *Big Data and Predictive Reasonable Suspicion*, 163 U. PA. L. REV. 327, 369-73 (2015); Elizabeth E. Joh, *Policing by Numbers: Big Data and the Fourth Amendment*, 89 WASH. L. REV. 35, 42-55 (2014); Jeremy Gorner, *With Violence Up, Chicago Police Focus on a List of Likeliest to Kill, Be Killed*, CHI. TRIB. (July 22, 2016, 3:54 PM), https://perma.cc/B92B-QBVB (describing the Chicago Police Department's "strategic subject list," which tracks residents "most likely to commit or be targeted by violence"); Justin Jouvenal, *Police Are Using Software to Predict Crime. Is It a "Holy Grail" or Biased Against Minorities?*, WASH. POST (Nov. 17, 2016), https://perma.cc/W7BW-7LMC (describing software used to predict property crimes).

similarly. It also looks at each file, but it asks: Does this file make it more likely that the owner of this computer committed some crime? If so, the algorithm adjusts some value it uses to track the overall probability that the owner of this computer committed that crime. At the end of the scan, it displays that probability to the investigator. And each of these algorithms scales easily. The hash search algorithm can ask a series of questions about each file: Is it child pornography? Pirated music? A computer virus? Just the same, the crime detection algorithm can track the probabilities of an unlimited number of crimes: Does this file increase the likelihood that this person committed tax evasion? Material support for terrorism? Narcotics trafficking?

Neither of these algorithms is obviously more invasive of the computer owner's privacy than the other. The hash search algorithm returns binary information about each individual file: Does that file match a piece of known contraband, or does it not? The crime detection algorithm returns probabilistic information about the computer as a whole: What's the likelihood it contains evidence of a given crime?[233] In neither case do investigators visually inspect individual files; rather, they only ever see the information the software displays after the algorithm has executed. Under a column header labeled "tax evasion," the general crime detection algorithm might list "31%." This information does not seem to be any more private than that revealed by a hash search algorithm, which flags a file "safe," or EID software, which returns a similar, though image-by-image, probabilistic assessment for only one particular crime.

233. There is a potential distinguishing principle between the two algorithms: The binary search doctrine is limited to technology that provides binary information about the presence or absence of *contraband*, not evidence of criminal wrongdoing more generally. The Court, however, has never precisely defined "contraband." And for Fourth Amendment purposes—including dog sniffs—the Court often treats "contraband" and "evidence of a crime" as essentially the same. In *Harris*, for example, the Court noted that the test for determining whether a dog's alert establishes probable cause is "whether all the facts surrounding a dog's alert, viewed through the lens of common sense, would make a reasonably prudent person think that a search would reveal *contraband or evidence of a crime*." *See* 133 S. Ct. at 1058 (emphasis added). Moreover, the Court has had no reason to clarify whether the binary search doctrine distinguishes between contraband and evidence more generally because no technology has been able to noninvasively provide binary information about the presence or absence of noncontraband evidence. Consider the genesis of the binary search doctrine in *United States v. Place*, 462 U.S. 696 (1983). Police used a dog to sniff a suspect's luggage, and the Court determined that its doing so wasn't a search. *See id.* at 707. That dog could only have indicated whether the luggage contained evidence of drug possession. *Id.* Suppose instead the dog could have indicated whether the bag contained evidence of tax evasion; would that have made the dog's sniff a search? The example seems absurd. But the sniff would still have been characterized by the two factors that made it not a search in *Place*: It would still be "limited both in the manner in which the information is obtained and in the content of the information revealed." *Id.*

C. Suspicionless Searches

The above analysis shows that a Fourth Amendment permitting hash searches likely also permits not only probabilistic algorithms like EID but also general crime detection algorithms. Use of these algorithms, however, would allow police to conduct suspicionless investigations into citizens' private data for evidence of ordinary criminal wrongdoing. Because the use of such algorithms would approximate the use of general warrants—but with a much greater potential to scale—the Fourth Amendment is unlikely to permit this practice. It is therefore equally unlikely to permit hash searches.

The Fourth Amendment "generally bars officials from undertaking a search or seizure absent individualized suspicion."[234] This is true both of searches and seizures pursuant to a warrant, which must "*particularly* describ[e] the place to be searched, and the persons or things to be seized,"[235] and of those searches and seizures that do not require a warrant but that must not be "unreasonable."[236] The Framers adopted this individualized suspicion requirement in response to the British use of general warrants—"warrants not grounded upon a sworn oath of a specific infraction by a particular individual, and thus not limited in scope and application."[237] Although the Court has recognized a subset of cases in which searches lacking individualized suspicion are nevertheless constitutionally permissible—"special needs" searches[238]—the Court has never approved a search "whose primary purpose was to detect evidence of ordinary criminal wrongdoing."[239]

Still, even if suspicionless *searches* for ordinary criminal wrongdoing are illegal, it follows that suspicionless algorithmic investigations are illegal only if they are also searches. But it makes sense to treat them as searches. This is because they raise the same concerns as the general warrants that motivated

234. Chandler v. Miller, 520 U.S. 305, 308 (1997).

235. U.S. CONST. amend. IV (emphasis added); *see also* Maryland v. King, 133 S. Ct. 1958, 1981 (2013) (Scalia, J., dissenting).

236. *See Chandler*, 520 U.S. at 308 (quoting U.S. CONST. amend. IV); *see also King*, 133 S. Ct. at 1981 (Scalia, J., dissenting).

237. *See King*, 133 S. Ct. at 1980-81 (Scalia, J., dissenting).

238. Vernonia Sch. Dist. 47J v. Acton *ex rel.* Acton, 515 U.S. 646, 653 (1995) (quoting Griffin v. Wisconsin, 483 U.S. 868, 873 (1987)); *see, e.g.*, *id.* at 646, 648, 664-65 (random drug testing of public school student-athletes); New York v. Burger, 482 U.S. 691, 702-04 (1987) (administrative searches of closely regulated businesses); United States v. Martinez-Fuerte, 428 U.S. 543, 545, 562 (1976) (border searches).

239. *See King*, 133 S. Ct. at 1981-82 (Scalia, J., dissenting) (quoting City of Indianapolis v. Edmond, 531 U.S. 32, 38 (2000)).

the Fourth Amendment: They have the potential to be used in ways "not limited in scope and application."[240]

An example will make this clear. Recall the hash searches email providers use to screen attachments for child pornography.[241] Suppose the government installed its own software screen between email users and providers, similar to how the government sometimes installs a pen register on a suspect's phone line.[242] That software could consist of a hash search for child pornography, like the one used by some email providers.[243] But it could also consist of general crime detection algorithms described above. By running that crime detection algorithm on every email and every attachment that passes through an email provider's servers, the government could investigate an unprecedented range of people and crimes, all without any individualized suspicion.

Although it's true that police wouldn't visually inspect each email and attachment, that should not be enough to immunize these algorithms against the Fourth Amendment.[244] After all, the Fourth Amendment does not only protect your right to keep intimate information away from police eyes; it also protects you from investigations into crimes for which police have no particularized reason to suspect you. This latter protection undergirds the Fourth Amendment's prohibition against general warrants and other suspicionless programs designed to "detect evidence of ordinary criminal wrongdoing."[245]

---

240. *See id.* at 1980; *see also* Riley v. California, 134 S. Ct. 2473, 2494 (2014); Boyd v. United States, 116 U.S. 616, 625-26 (1886).

241. *See supra* Part II.B.

242. By this I mean that the two investigatory tools would be functionally similar in that they would collect information about a user's communications with the knowledge of the service provider but not the user. The third-party doctrine, which permits the government's use of pen registers, *see* Smith v. Maryland, 442 U.S. 735, 742-46 (1979), is beyond the scope of this Note—though its continuing vitality may be in doubt, *see* Transcript of Oral Argument, Carpenter v. United States, No. 16-402 (U.S. Nov. 29, 2017) (debating the third-party doctrine in the context of cell site location information). In any case, the third-party doctrine likely does not permit the government to visually inspect the contents of emails without a warrant. *See* United States v. Warshak, 631 F.3d 266, 287-88 (6th Cir. 2010).

243. *See supra* Part II.B.

244. In *Kyllo v. United States*, the Court held that "obtaining by sense-enhancing technology any information regarding the interior of [a] home that could not otherwise have been obtained without physical 'intrusion into a constitutionally protected area' constitutes a search." 533 U.S. 27, 34 (2001) (citation omitted) (quoting Silverman v. United States, 365 U.S. 505, 512 (1961)). Similarly, the police should not be permitted to conduct what would otherwise be a search of an email by using a "sense-enhancing technology" that permits them to extract information from an email without visually inspecting it.

245. *See* City of Indianapolis v. Edmond, 531 U.S. 32, 37-38 (2000).

Algorithmic investigatory tools implicate this concern in a particularly powerful way. As several justices observed during oral argument in *United States v. Jones*,[246] if the government were permitted to use GPS to track suspects without a warrant, "then there is nothing to prevent the police or the government from monitoring 24 hours a day the public movement of every citizen of the United States."[247] Even if it doesn't violate the Fourth Amendment for government personnel to conduct in-person visual surveillance, computers change that calculus in two ways. First, although "memories are fallible, computers aren't."[248] By using computers, the government can amass much more information, for a longer period of time, than it was previously able to. And second, with rare exceptions, "no one . . . sends human beings to follow people 24 hours a day," but "with the machines, you can."[249]

That is, one important check against suspicionless searches is that they are relatively expensive: Because most people haven't committed a crime, it would take an excessive amount of resources to, for example, physically surveil a broad swath of the citizenry. Instead, police focus their resources on suspects for whom they have individualized suspicion because those investigations are more likely to prevent or resolve crimes. Computers, however, remove this resource constraint and make it possible for police to conduct suspicionless investigations on a massive scale.[250]

Thus, not only do algorithmic investigative tools allow the government to conduct suspicionless searches, but they also allow the government to conduct those searches on a much larger scale than has historically been achievable. The Fourth Amendment, which was adopted to protect against general warrants, should not be read to permit these suspicionless searches, regardless whether they take the form of hash searches (which are already in use) or general crime detection algorithms (which are still being developed).

V. Treating Computers as We Treat Humans

Even if hash searches, when used without or beyond the scope of a warrant to analyze private information for evidence of criminal wrongdoing, are illegal because they would permit the equivalent of general warrants, so what? Determining that hash searches are illegal, or that EID is illegal, isn't

246. 565 U.S. 400 (2012).

247. Transcript of Oral Argument at 12-13, *Jones*, 565 U.S. 400 (No. 10-1259), 2011 WL 5360051 (Breyer, J.); *see also id.* at 9-10 (Roberts, C.J.); *id.* at 10-11 (Alito, J.).

248. *See id.* at 13 (Breyer, J.).

249. *See id.*

250. *See, e.g.*, Joh, *supra* note 232, at 60-62.

particularly helpful: Those techniques can (and likely soon will) be replaced with other algorithmic investigative techniques no less effective at uncovering evidence. Thus, a rule limited to one particular algorithm isn't much use. What we need instead is a framework for thinking about algorithmic search techniques more generally. The lessons we draw from hash searches can help us derive such a framework.

A. Choosing Between a Proactive and a Reactive Approach

At the outset, it is worth asking whether a Fourth Amendment framework for algorithmic searches is appropriate: Should courts instead address new technologies on a case-by-case basis? Orin Kerr, for example, has argued that courts should respond to changing technology through *equilibrium adjustment*.[251] Equilibrium adjustment is a "correction mechanism" whereby courts "adjust the level of Fourth Amendment protection to try to restore the prior equilibrium" whenever "new tools and new practices threaten to expand or contract police power in a significant way."[252] This approach is supported by two arguments. As a matter of institutional competence, courts are poorly equipped to assess the Fourth Amendment implications of new technologies before they've stabilized.[253] And as a normative matter, courts should attempt to preserve the same balance between the privacy rights of citizens and the investigatory powers of police as existed at the Founding.[254]

To start, Kerr's institutional competence argument does not compel a reactive approach in the context of algorithmic searches. Kerr argues that it is more appropriate for judges to assess new technologies reactively than to do so proactively due to the "dramatic mismatch [that] exists between the difficulty and complexity" of modern investigative techniques and the scant "empirical evidence judges have about what rules work."[255] Consequently, judges should wait until a new technology reaches a "reasonably stable state" before determining its legality under the Fourth Amendment.[256]

---

251. Orin S. Kerr, *An Equilibrium-Adjustment Theory of the Fourth Amendment*, 125 HARV. L. REV. 476, 480 (2011).

252. *See id.*

253. *See id.* at 535-37.

254. *See id.* at 481-82 ("Equilibrium-adjustment maintains fidelity to the Fourth Amendment in the face of rapid change by allowing judges to maintain the balance struck by the Fourth Amendment."); *see also* Kyllo v. United States, 533 U.S. 27, 34 (2001) (adopting a rule in part because it "assures preservation of that degree of privacy against government that existed when the Fourth Amendment was adopted").

255. *See* Kerr, *supra* note 251, at 535.

256. *See id.* at 539, 542.

There are several reasons courts should be leery of this argument, at least with respect to algorithmic investigative techniques. As an initial matter, these technologies are often not so complex as is sometimes suggested. For example, there is, as Part I above showed, no reason to think that courts lack the institutional competence to assess hash searches. And as demonstrated in Part IV above, courts can often extend the lessons they take from stable algorithms (like hash searches) to algorithms still under development (like the general crime detection algorithm). Moreover, the fact that courts might not yet understand a newly developed technology should not give the government carte blanche to use it. Rather, courts should expect the government to show that a new technology complies with the Fourth Amendment before admitting evidence produced by that technology; that approach would help courts develop institutional competence where they lack it.

As for Kerr's normative argument, although it may be desirable for courts to seek to maintain a balance between privacy rights and police powers, that goal calls for a proactive rather than reactive approach in the context of algorithmic searches. That is because algorithmic investigative techniques are not a one-time technological advancement; by the time these techniques stabilize, police powers and privacy rights will have become significantly imbalanced.

To take one example, Kerr suggests automobiles as a technology to which the Court has applied equilibrium adjustment.[257] Kerr notes that, before the advent of the automobile, "there appear to have been few limits on the police power to stop carriages and buggies to investigate crimes."[258] But "when the widespread introduction of cars threatened to dramatically facilitate crime," the Court determined that cars were entitled to no more protection than buggies: Police can search a car without a warrant so long as they have probable cause to believe that it contains evidence of a crime.[259] And this technology is "stable" in terms of the privacy—that is, physical protection from police surveillance (as opposed to legal protection under the Fourth Amendment)—it affords or denies individuals[260]: Cars today offer essentially the same privacy as cars in 1925, when the Court first adopted the automobile exception.[261]

257. *See id.* at 507.

258. *Id.*

259. *See id.* at 507-08.

260. *See id.* at 539.

261. *See* Carroll v. United States, 267 U.S. 132, 146-47 (1925).

By contrast, computing power has been increasing exponentially for over fifty years.[262] Artificial intelligence (AI) and machine learning—two branches of computer science employing predictive algorithms—are not one-time developments; rather, advances in these fields are ongoing and have been for many years.[263] Thus, Stephen Hawking and other equally accomplished scientists can make wildly indeterminate statements about AI, such as that it will be "either the best, or the worst thing, ever to happen to humanity,"[264] because none of us knows what AI will look like in ten years. Algorithmic investigative techniques will come to rely increasingly not on the hash algorithms described here but on AI and machine learning—technologies that, as of yet, show no signs of stabilizing. Instead of reacting belatedly to already-obsolete advancements in these technologies, judges should adopt an affirmative framework for thinking about algorithmic investigative techniques.

B. A Simple Rule for Algorithmic Investigative Techniques

The above analysis of hash searches suggests such a proactive framework. Hash searches are troubling because they are designed and executed by humans and because they perform a task that would be a search if performed by humans. The courts that have encountered hash searches have acknowledged this concern but haven't identified a Fourth Amendment framework for responding to it.[265] The concern itself, however, suggests a rule: Police conduct

262. In 1965, Gordon Moore noted that the number of transistors in an integrated circuit was doubling every year. *See* Thomas L. Friedman, Opinion, *Moore's Law Turns 50*, N.Y. TIMES (May 13, 2015), https://perma.cc/9EA4-93X3. In 1975—after he'd been proved right—Moore predicted that the number would continue doubling every two years. *See id.* This rate of exponential increase became known as Moore's Law. *Id.* But it's unclear how long Moore's Law will continue to hold. *See After Moore's Law*, ECONOMIST: TECH. Q. (Mar. 12, 2016), https://perma.cc/A49J-D8YW (noting that "[t]he pace of advance has been slowing" and describing ways that chipmakers are continuing to try to improve performance).

263. While advancements in AI are nothing new, they show no signs of slowing down. *See, e.g.*, Will Knight, *5 Big Predictions for Artificial Intelligence in 2017*, MIT TECH. REV. (Jan. 4, 2017), https://perma.cc/BD9D-56PH; John Markoff, *The Rapid Advance of Artificial Intelligence*, N.Y. TIMES (Oct. 14, 2013), https://perma.cc/CP67-682Y; Robert McMillan, *AI Has Arrived, and That Really Worries the World's Brightest Minds*, WIRED (Jan. 16, 2015, 6:30 AM), https://perma.cc/UY36-SSJ3. Machine learning has received less attention in popular media, but its advancements continue apace. *See, e.g.*, Alex Hern, *Google Says Machine Learning Is the Future. So I Tried It Myself*, GUARDIAN (June 28, 2016, 3:00 AM EDT), https://perma.cc/QJ8Q-U24C ("The world is quietly being reshaped by machine learning.").

264. *See* Alex Hern, *Stephen Hawking: AI Will Be "Either Best or Worst Thing" for Humanity*, GUARDIAN (Oct. 19, 2016, 4:05 PM EDT), https://perma.cc/265H-WG2H (quoting Stephen Hawking).

265. *See supra* Part II.C.

a search when they use a computer to perform a task that would be a search if conducted manually by a human.

This rule, violation of which would serve as a sufficient but not necessary condition for determining that warrantless use of a new technology presumptively violates the Fourth Amendment, is easy to apply in the vast majority of cases. As this Note has demonstrated, technologies that commentators describe as "complex"[266] or "complicated"[267] can often be explained and comprehended quite easily.[268] And almost all algorithmic investigative techniques can be performed, if less efficiently, by humans. Indeed, an algorithm is just "a step-by-step procedure for solving a problem or accomplishing some end."[269] Just as humans manually use algorithms to solve Rubik's Cubes,[270] they could, given the time and inclination, manually replicate a hash search. It therefore makes sense to treat algorithmic searches the same as human searches.

To be sure, and unlike Kerr's equilibrium adjustment framework, this rule wouldn't cover all technological advancements. Certain types of technology simply can't be replicated by humans. Indeed, this limitation is aptly illustrated by the disagreement between Justices Scalia and Alito in *United States v. Jones* over whether "a very tiny constable" could have replicated the work of a GPS transmitter.[271] But this implies only that the rule should be adopted as a sufficient, not a necessary, means of determining that warrantless use of a given investigative technology presumptively constitutes an unreasonable search. Other technologies that aren't replicable by humans can remain covered by a reactive rule; because these technologies more often represent discrete rather than continuing advancements (like the GPS in *Jones*), it may be the case that little is lost by taking a reactive posture to their adoption by law enforcement.

While the rule I've proposed doesn't cover all possible technological advancements, it works particularly well for investigative algorithms, which are likely to rely increasingly heavily on AI and machine learning technolo-

266. *See, e.g.*, Salgado, *supra* note 3, at 38.

267. *See, e.g.*, Kerr, *supra* note 9, at 541.

268. *See supra* Part I.

269. *Algorithm*, MERRIAM-WEBSTER, https://perma.cc/R6ZX-EN9G (archived Nov. 8, 2017).

270. *See, e.g.*, Chris Durnford, *How to Solve the Rubik's Cube*, RUBIK'S (June 19, 2014), https://perma.cc/X5QR-GMJQ; Rakshith MG, *A List of Every Rubik's Cube Algorithm You Will Ever Need*, HOBBYLARK, https://perma.cc/8CCE-9PST (last updated Dec. 31, 2017).

271. *Compare* 565 U.S. 400, 406 n.3 (2012) (suggesting that "a constable's concealing himself in the target's coach in order to track its movements" "is not far afield"), *with id.* at 420 n.3 (Alito, J., concurring in the judgment) (suggesting that this "would have required either a gigantic coach, a very tiny constable, or both").

gies. After all, the "intelligence" AI seeks to replicate and improve upon is that of humans; when police employ AI in service of an investigation, they're simulating the work a human investigator would historically have performed. Similarly, machine learning technologies typically replicate human learning processes. Just as a doctor gets better at recognizing melanoma by looking at lots of examples of that cancer, a machine learning algorithm designed to identify melanoma goes through the same learning process.[272]

Because they often mimic human intelligence and human learning, AI and machine learning algorithms are more similar to human decisionmaking processes than to many other technologies. Moreover, these technologies are not discrete advancements but are still developing.[273] Consequently, investigative algorithms using these technologies are both (a) better suited to a rule treating them the same as human searches and (b) more in need of this affirmative framework. Courts wrestling with these technologies' privacy implications should insist that law enforcement explain how its algorithmic technologies work, as this Note has done with hash searches. They should then ask themselves whether the same algorithm, if performed by a human investigator, would be a search.

Conclusion

Not all investigative techniques replicate work traditionally done by humans. But for an important subset of investigative technologies—algorithmic search techniques—the very purpose of the technologies is to replace, and improve upon, human police work. In one sense, these algorithmic searches are less invasive than traditional searches because a piece of software, rather than a set of human eyes, inspects a person's private data. But algorithmic searches also make possible a twenty-first century version of general warrants—suspicionless searches for ordinary evidence of criminal wrongdoing—on an unprecedented scale. Courts should guard carefully against this possibility and prohibit algorithms from going where police are forbidden to tread.

272. *See* Taylor Kubota, *Deep Learning Algorithm Does as Well as Dermatologists in Identifying Skin Cancer*, STAN. NEWS (Jan. 25, 2017), https://perma.cc/W82D-RUM9 (describing an algorithm trained on a database of 130,000 skin disease images).

273. *See supra* notes 263-64 and accompanying text.