CONNECTICUT
LAW REVIEW

VOLUME 49 SEPTEMBER 2017 NUMBER 5

Note

Total Recall: Computers and the Warrant Clause

JOSHUA PERLDEINER

The inherent conflict between the requirement for limited-scope warrants and searches under the Fourth Amendment and the structure of computer file systems has yet to be resolved. Where searches must be, by law, constrained but, due to the realities of computer architecture, cannot be so confined, there is a tension that must be addressed. This Note proposes a solution whereby law enforcement would seek to use a form of blind search algorithm, allowing a program to do the searching of the computer system in question, but only returning hits where appropriate and justified by the scope of the warrant.

NOTE CONTENTS

Total Recall: Computers and the Warrant Clause

JOSHUA PERLDEINER*

INTRODUCTION

"The requirement that warrants shall particularly describe the things to be seized makes general searches under them impossible and prevents the seizure of one thing under a warrant describing another. As to what is to be taken, nothing is left to the discretion of the officer executing the warrant."[1] Of course, this imperative breaks down once computers are involved.[2] File structures are infinitely mutable and are often designed, consciously or unconsciously, by the end user of a computer.[3] Unlike the physical world of objects in space, it is possible for a file of nearly any size to be contained in any folder or subfolder.[4] The scope of the warrant granted to search that computer, therefore, is functionally infinite. In virtual space, we may very well find an "elephant in a bread box."[5]

It is this very liberality of scope that concerns us. Where there can be no boundaries, all searches must therefore be permitted. Certainly, the issuing magistrate may restrict the search criteria when warrants for digital

---

* University of Connecticut School of Law, J.D. 2017. Thanks to Judge Michael Sheldon, Dr. Albert Harper, and Evelyn O'Regan for their assistance on this subject.

[1] Marron v. United States, 275 U.S. 192, 196 (1927).

[2] *See*, *e.g.*, United States v. Burgess, 576 F.3d 1078, 1089–91 (10th Cir. 2009) (holding that a search warrant "authorizing a search of 'computer records' and 'items of personal property which would tend to show a conspiracy to sell drugs'" was not overbroad).

[3] The only cases where this is not true would be those in which the file structure was designed by some third party and the end-user was constrained to make use of that structure. Since the organization of files on a computer is an inherently personal and, in many ways, arbitrary matter, it is impossible to construct generalized norms about where and how files of certain types will be stored. Even best-case scenarios posit possibilities.

[4] *See* United States v. Hill, 459 F.3d 966, 978 (9th Cir. 2006) ("Images can be hidden in all manner of files, even word processing documents and spreadsheets. Criminals will do all they can to conceal contraband, including the simple expedient of changing the names and extensions of files to disguise their content from the casual observer."); United States v. Mann, 592 F.3d 779, 782 (7th Cir. 2010) ("Unlike a physical object that can be immediately identified as responsive to the warrant or not, computer files may be manipulated to hide their true contents.").

[5] *See* Miles v. Oklahoma, 742 P.2d 1150, 1151 (Okla. Crim. App. 1987) ("Just as it is patently beyond the scope of a warrant to search for an elephant in a breadbox or a shotgun in a snuff can, so also is it an unreasonable search to look for a handgun nine inches long inside an utility bill or bank statement envelope, in a flat notebook, or inside any other container too small to hold the weapon."); *see also* Horton v. California, 496 U.S. 128, 142 (1990) (holding that items seized from petitioner's home were discovered in lawful accordance with the "plain-view" doctrine).

media are issued.[6] Federal courts have been loath to make use of this tool.[7] This laissez-faire attitude toward digital searches where warrants are concerned permits any officer to conduct what amounts to a general rummaging search of any computer for which he receives a magistrate's signature. Computers of all stripes, from cell phones to desktops, act as repositories for personal data that may amount to a person's business and personal records in their entirety, and they can include conversations had decades ago, logs of passwords, business transactions, and anything else imaginable. Computers log everything.[8] Thus, it may be possible to reconstruct a person's life from his digital shadow. Surely there are reasons to see these vast repositories of information are protected from general rummaging by government agents.

As the problem was created by technology, perhaps technology can provide its solution. It is conceivable that a search protocol can be created which would permit limited access to a seized computer, providing, for example, all files that match certain criteria for police or other government agents to view. This solution would conceal the nonpertinent files while still presenting anything that meets the requirements of the warrant. If there are hits, it is likely they could be used to develop a new warrant with an expanded scope. I call this solution the *concealed search*, a way of peeling layers off the onion one at a time so as not to stray into impermissible territory inadvertently.

The following questions then arise: 1) is a concealed search desirable and, 2) is a concealed search viable? If the search is desirable but not yet viable, new methodology can be developed by computer scientists to address the problems. If it is feasible but not desirable, the inquiry can end there.

Whether or not a concealed search is desirable is a question of law. Would it help properly enforce the protections of the warrant clause? A separate desirability consideration exists: would it make searching computers easier or harder? Whether the search is possible can be asked in one of two ways: is it possible, under the law as it stands, for these

---

[6] At least theoretically, as was acknowledged in *United States v. Richards*, 659 F.3d 527, 538 (6th Cir. 2011) ("[T]he majority of federal courts have eschewed the use of a specific search protocol . . . . And there is no reason to so limit computer searches.").

[7] *See, e.g.*, United States v. Stabile, 633 F.3d 219, 238–39 (3d Cir. 2011); *Mann*, 592 F.3d at 785–86; *Burgess*, 576 F.3d at 1092–93; United States v. Cartier, 543 F.3d 442, 447–48 (8th Cir. 2008); United States v. Khanani, 502 F.3d 1281, 1290 (11th Cir. 2007); United States v. Hill, 459 F.3d 966, 977 (9th Cir. 2006); United States v. Adjani, 452 F.3d 1140, 1149–50 (9th Cir. 2006); United States v. Brooks, 427 F.3d 1246, 1251 (10th Cir. 2005); Guest v. Leis, 255 F.3d 325, 334 (6th Cir. 2001); United States v. Upham, 168 F.3d 532, 534 (1st Cir. 1999).

[8] For example, see the *Event Manager* on a PC. It is functionally impossible for a computer to make a change without logging it.

searches to be conducted? If so, is it technically possible, given the current limits of search algorithms?

I. IS A CONCEALED SEARCH DESIRABLE FROM THE VIEWPOINT OF THE FOURTH AMENDMENT TO THE UNITED STATES CONSTITUTION?

The Fourth Amendment provides that:

> 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.[9]

Warrantless searches are *per se* unreasonable, and thus violate the Fourth Amendment, save in a very narrow range of circumstances.[10] One of the restraining elements of the warrant clause is the requirement of *scope*.[11]

The particularity requirement of the warrant clause requires that the areas to be searched be particularly described[12]—and this requirement is honored in every circumstance except for those searches which peer into the digital realm.[13] Once a computer has been seized and a search warrant has issued, the entire machine may be searched as long as the searching agency has probable cause to believe there is evidence of a crime *somewhere in the computer*.[14]

---

[9] U.S. CONST. amend. IV.

[10] *See* Jones v. United States, 357 U.S. 493, 499 (1958) ("The exceptions to the rule that a search must rest upon a search warrant have been jealously and carefully drawn . . . .").

[11] *Id.*

[12] *See* Steele v. United States, 267 U.S. 498, 503 (1925) (holding that "[i]t is enough if the description is such that the officer with a search warrant can, with reasonable effort, ascertain and identify the place intended."); *see also* United States v. Hassell, 427 F.2d 348 (6th Cir. 1970); Irwin v. United States, 89 F.2d 678 (D.C. Cir. 1937) (search warrant described premises as occupied by defendant known as the "Humidor," at a designated address. The ground floor was occupied by a cigar store of that name, but the building contained four floors, all occupied by the defendant and a search was made of the entire building. The court found the warrant adequate for such a search); Carney v. United States, 79 F.2d 821 (6th Cir. 1935); Chiaravolloti v. United States, 60 F.2d 192 (7th Cir. 1932) (incorrect designation of premises as within city inconsequential); Rose v. United States, 45 F.2d 459 (8th Cir. 1930); Fall v. United States, 33 F.2d 71 (9th Cir. 1929); Giacalone v. United States, 13 F.2d 108 (9th Cir. 1926); United States v. Ortiz, 311 F. Supp. 880, 883 (D. Colo. 1970) ("Mountain cabins ordinarily have no specific address number and must be identified by a description of the physical features of the cabin and its general location. We do not deem it necessary for the warrant to incorporate the United States Geological Survey Map or to contain a legal description of the property."); United States v. Thomas, 216 F. Supp. 942 (N.D. Cal. 1963); United States v. Neadeau, 2 F.2d 148 (N.D. Wash. 1924); United States v. Chin On, 297 F. 531 (D. Mass 1924); Easley v. State, 459 S.W.2d 410 (Ark. 1970); Cole v. State, 237 So. 2d 443 (Miss. 1970).

[13] *See e.g.*, Nicole Friess, *When Rummaging Goes Digital: Fourth Amendment Particularity and Stored E-Mail Surveillance*, 90 NEB. L. REV. 971 (2012).

[14] *Id.* at 990.

The reason for the scope requirement for warrants is rooted in historical fact. Prior to the War of American Independence, the Crown created the hated instruments of the general warrant and the writ of assistance.[15] These tools gave their bearers the power to compel any subject to permit any type of intrusion;[16] even British attorneys wrote of the writs of assistance with scorn and derision.[17] When the Supreme Court warned against "general rummaging expeditions," it was with these precise rummagings in mind.[18]

Why, then, is a computer not equivalent to, say, a study? A warrant authorizing the search of a man's desk based upon probable cause to believe that his papers contain evidence of fraud would certainly be upheld as sufficiently specific to satisfy the particularity requirement.[19] One could hardly imagine that the owner of that desk would store every piece of paper he had ever signed, read, or touched in it. Yet, this is precisely the way our computers function.[20] The correct metaphorical equivalent to a search warrant for a computer is not a search warrant that permits a physical search of a suspect's office. It is a warrant that permits a search of a warehouse full of information, much of it private in nature, much of it personal correspondence.[21]

---

[15] Writs of assistance were granted by the British Parliament to officials who requested the writ; there was no requirement of justification or support. Customs officers were empowered to commandeer civil officials and all other persons to assist in the execution of the writ. Writs were not a general warrant, but rather were a standing court order to serve as identification that the person carrying them was a customs officer empowered to engage in searches. Carolyn N. Long, *The Origins of the Fourth Amendment*, 11 INSIGHTS ON L. & SOC'Y 4, 4 (2011).

[16] *Id.*

[17] James Otis Jr. called them the "worst instrument of arbitrary power, the most destructive of English liberty and the fundamental principles of law, that was ever found in an English law book." The contents of this speech only exist in the notes of John Adams, who was present at the time. JOHN ADAMS, WORKS, II, 124 note, 521-525; X, 246-249, 274-276.

[18] *E.g.*, Coolidge v. New Hampshire, 403 U.S. 443, 467 (1971) (noting the evils of general rummaging through a person's belongings).

[19] *See Searching and Seizing Computers and other Electronic Devices, in* HOMELAND SECURITY, LEGAL DIVISION HANDBOOK 12.

[20] Even data that is marked for deletion is not immediately discarded by a computer. The areas of memory that contain the data continue to retain it until new information requires the space. Only then is it overwritten, and even so, in some cases this overwriting is only partial. *See generally* A GUIDE TO UNDERSTANDING DATA REMANENCE IN AUTOMATED INFORMATION SYSTEMS, (Patrick R. Gallagher ed., 1991); GORDON HUGHS & TOM COUGHLIN, TUTORIAL ON DISK DRIVE DATA SANITIZATION (UCSD CENTER FOR MAGNETIC RECORDING RESEARCH).

[21] For example, during the prescient Second Circuit Judicial Conference of 1990, an expert from IBM described the capabilities of a computer to store information about every person in the country and make it instantly available. Judge George Pratt said in dismal tones: "The potential is mind-boggling, and the only tool we have to deal with these intrusions into privacy is the [F]ourth [A]mendment." *Fourth Amendment Symposium: The Supreme Court and Local Government Law: The 1988–89 Term*, 6 TOURO L. REV. 31, 53 (1990).

There is another implicit question in the searching of digital data: at what point is it seized? Does the seizure take place when the computer is copied for search, as it must be under current federal procedures, or only when it is physically impounded?[22] While suggestions have been offered, no judicial determination on this subject has yet been made.[23]

As of this instant, there are several lines of federal cases that uphold a limitless scope of search when search warrants issue for computers and the data therein.[24] However, there are also cases holding that the scope of a search and seizure warrant for a computer drive requires more oversight than simply permitting the agents of law enforcement to copy everything and search in every crevice.[25] This conflict as to how law enforcement should handle the digital environment is confusing and dangerous, to say the least. Either they are being too conservative, and therefore missing evidence that should rightly come to light, or they are being too intrusive, and being permitted to use evidence that was gathered in violation of the Fourth Amendment.[26]

One of the most recent Supreme Court cases on cell phone searches, *Riley v. California*, is on point here.[27] The Supreme Court held that a warrantless search of a cell phone pursuant to arrest violated the warrant requirement of the Fourth Amendment.[28] The scope of the reasonable warrantless search incident to arrest, therefore, does not include the contents of digital devices on the person of the arrestee.[29] This is a slim extension of the Fourth Amendment protection into the digital world.

---

[22] *See* COMPUTER CRIME & INTELLECTUAL PROP. SECTION, U.S. DEP'T OF JUSTICE, SEARCHING AND SEIZING COMPUTERS AND OBTAINING ELECTRONIC EVIDENCE IN CRIMINAL INVESTIGATIONS 77–78 (3d ed. 2009) ("Because examining a computer for evidence of crime is so time consuming, it will be infeasible in almost every case to do an on-site search of a computer or other storage media for evidence . . . . In many cases, rather than seize an entire computer for off-site review, agents can instead create a digital copy of the hard drive that is identical to the original in every respect.").

[23] Of particular interest is Orin S. Kerr's *Fourth Amendment Seizures of Computer Data*. 119 YALE L. J. 700 (2010).

[24] *See, e.g.*, United States v. Upham, 168 F.3d 532, 535–36 (1st Cir. 1999) (holding warrant authorizing seizure of all computer software, hardware, disks, and disk drives sufficiently particular, though it did not restrict items to be seized to items related to suspected crimes because computer equipment instrumental). *But see* United States v. Carey, 172 F.3d 1268, 1273 n.4 (10th Cir. 1999) (holding warrant authorizing search of suspect's computer for files pertaining to sale or distribution of controlled substances insufficiently particular to justify seizure of child pornography images from closed files on defendant's hard drive).

[25] *See, e.g.*, *Carey*, 172 F.3d at 1273 (holding that closed computer files were not in plain view and thus unlawfully seized).

[26] This is certainly the view that proponents of the so-called "mosaic theory" hold. Orin S. Kerr, *The Mosaic Theory of the Fourth Amendment*, 111 MICH. L. REV. 311, 353–54 (2012).

[27] 134 S. Ct. 2473 (2014).

[28] *Id.* at 2478–79.

[29] *Id.*

II. IS A CONCEALED SEARCH EASIER OR HARDER ON LAW ENFORCEMENT?

This question can be asked from one of two viewpoints: first, would the use of a concealed search algorithm be more taxing on law enforcement? And second, would the use of a concealed search algorithm allow law enforcement more consistently to avoid the pitfalls of garnering evidence that could not, then, be admitted in a court of law? These are complex questions that are comprised of a multitude of sub-issues and concerns, and thus must be addressed in turn.

The first concern is that of an increase in the difficulty of the actual process of searching the data contained on a computer. This is an operational concern, one that is precisely focused on the user interface of any system which is ultimately designed. It is conceivable that a system might exist that would be much worse to use than simply being permitted to search freely. It is conceivable that a system might be created that would be much more user friendly, returning results in a much faster manner than a manual search. Because this is a question of implementation, it cannot be resolved at this stage of the analysis.

The second concern is whether or not the use of a concealed search algorithm would avoid the pitfalls of gathering evidence that could later not be presented to a tribunal. Obviously, resolving the circuit split one way or another would make it easier to predict whether or not certain digitally obtained evidence was admissible.[30] There are three potential paradigms that we might see: 1) the current paradigm, where there is no clarity as to whether there is a limit to the scope of digital search warrants, 2) a paradigm in which there is a clear limit on scope for digital search warrants, and 3) a paradigm where there is an unlimited scope for any search warrant issued for the contents of a computer or other digital storage medium.

A. *No Clarity*

If the paradigm of today continues into the future, there will be a lack of clarity in the law as to what is permissible to seize and where it is permissible to search on the hard-disk drive ("HDD") of a computer.[31] In some jurisdictions, scope will remain a concern, while it will not in others. The resulting confusion will cause law enforcement no end of problems—law enforcement has, as a result, taken a conservative approach to

---

[30] Indeed, it may be true that within each circuit there is little question of admissibility. However, state rules can and do conflict with the federal admissibility issue wherever evidence law is not descended directly from a Constitutional mandate. *See* Brandon L. Garrett, *Constitutional Law and the Law of Evidence*, 101 CORNELL L. REV. 1, 59–62 (2015) (explaining the "conflict between constitutional rights and the complex and tradition-bound body of state and federal evidence law").

[31] *Supra* note 23 and accompanying text.

gathering information from disks and other digital sources.[32] Essentially, under this rule, law enforcement would be best served by assuming the strictest reading of the rule possible.[33]

The strictest reading of this rule is similar to that found under the limited scope paradigm, and may be more limited in fact than whatever rule the courts may eventually issue. In order to be certain that the evidence gathered is admissible, it will always be safer to hedge on the side of caution. Therefore, as long as the issue remains unsettled, the wise law enforcement agency will approach the analysis as though under the subsequent paradigm, here listed as Section B.

B. *Limited Scope*

The limited scope paradigm proposed by this Note in Section I may present a number of inherent problems.[34] It will be more difficult, under such a constitutional mandate, for law enforcement agencies to collect evidence. It is as great an imposition as the imposition of the warrant requirement in the first place, requiring essentially the extension of all the concepts behind the warrant requirement into a world that has hitherto been unregulated by them. It would certainly throw state systems that do not currently require specificity for digital media into massive disarray.[35]

However, with order there comes predictability. In this paradigm, law enforcement agencies will know what is acceptable and what is not. Questions would be resolved in favor of a stricter interpretation, and the use of a concealed search algorithm would save law enforcement much agonizing over how evidence is to be gathered.

In this setting, pre-approved search terms and criteria would be set by the warrant. It might even be prudent to issue a ruling that any evidence gathered in accordance with the concealed search software that makes use of the pre-approved search criteria is presumptively within the scope of the warrant, subject to some kind of challenge.[36] The absence of a concealed search algorithm under this hypothetical configuration of the law would make gathering digital evidence much more difficult.

---

[32] *See generally* LEGAL DIVISION HANDBOOK, *supra* note 19.

[33] *Id.*

[34] *Supra* Section I.

[35] As a Federal Constitutional mandate, such a paradigm would supplant the rules on warrant specificity in any jurisdiction that currently does not require one.

[36] Some jurisdictions already require digital searches to comport with a certain scope requirement. *See, e.g.*, United States v. Grimmett, 439 F.3d 1263 (10th Cir. 2006) (stating that it was not overbroad to authorize a search warrant for images under file extensions .jpg, .mpg, and .bmp). These jurisdictions would see enormous benefit from such a rule. *See generally* David J. S. Ziff, *Fourth Amendment Limitations on the Execution of Computer Searches Conducted Pursuant to a Warrant*, 105 COLUM. L. REV. 841 (2005).

C. *Unlimited Scope*

In a paradigm where the scope of a search warrant is unlimited when it comes to digital media, there would obviously be no need for concealed search software. Indeed, it would be needlessly cumbersome, burdening law enforcement in their quest for the truth.[37]

Out of the three paradigms, law enforcement would benefit from the concealed search software in two cases. It is not even necessary to posit a presumption of being within the scope when using such a search engine; the mere application of such a procedure would tend to restrict law enforcement to searching in files and places that they were permitted to search. Of course, the converse danger exists: that the concealed search software will fail to turn up files and folders that are permissible under the warrant as issued. This is a technological problem, and I will address that in Section IV, *infra*.

III. IS THE USE OF A CONCEALED SEARCH PERMISSIBLE UNDER CURRENT LAW?

Currently, the law does not forbid the use of the concealed search as long as the warrant requirement is met. It certainly would be an additional, if unnecessary, step for law enforcement to undertake. However, there is nothing preventing the use of a concealed search algorithm on a cloned device.

When computers are seized for search, the general rule is that they are cloned by the agency which seized them so they may be searched off site and to prevent any degradation or loss of seized data.[38]

A. *Is the Use of a Concealed Search Technically Possible Given the Current State of Search Algorithms?*

In order to answer this question satisfactorily, one must understand a great deal about computer architecture,[39] the structure of data,[40] how

---

[37] Although, again, we should note that the desire to seek out what actually occurred is only one of the many factors in determining how law enforcement should proceed. As always, this must be balanced with a healthy respect for the rights of the public. In this hypothetical, the courts or the legislature has determined that this idea of privacy in one's computer is not as important as the social good to be gained from allowing law enforcement to search the entire contents of any machine for which they have a warrant.

[38] The computers are generally, then, returned. The clone is kept for indexing and searching for an indefinite period. *See* OFFICE OF LEGAL EDUC., EXEC. OFFICE FOR U.S. ATTORNEYS, SEARCHING AND SEIZING COMPUTERS AND OBTAINING ELECTRONIC EVIDENCE IN CRIMINAL INVESTIGATIONS 76–79 (2009) (instructing law enforcement and U.S. Attorneys that searching computers must almost always be done off-site, and the solution to this is to image or clone the system so it may be examined). There is no legal requirement that the cloned system be destroyed.

searching works,[41] and the recent advancements in the field of optimized searching. Only then can I address the meat of the question at hand, namely: can this ambitious project realistically be done?

1. *Computer Architecture*

Computers store their data on drives, encoded in magnetic regions.[42] There are two general types of drives: platter (or hard-disk drive)[43] and solid state (or solid-state drive).[44] HDDs, or traditional drives, operate by placing a disk that has two magnetized sides in the center of a housing or enclosure.[45] This platter spins in place and contains rings of data, much like a record.[46] The data is encoded on the platter by magnetizing the material, just as you would press grooves into a record.[47] There is no attempt to measure the valence of the magnetic field: regions are either magnetized (and therefore read as a "1") or not (and therefore read as a "0"), meaning that all data must be encoded in binary.[48]

---

[39] Both the physical makeup of a computer, as well as the structure of the operating system, are important.

[40] Obviously, different operating systems have different file structures and different means of formatting mean different challenges for search engines. However, it is important to note that all modern Windows and Mac OS computers come with auto-indexing agents. In essence, they are, already made to be searched. GARY SHELLY & MISTY VERMAAT, DISCOVERING COMPUTERS, FUNDAMENTALS 259 (2008).

[41] The science of search optimization is a lively sub-field of computer science. *See, e.g.*, Rocío L. Cecchini et al., *Multiobjective Evolutionary Alogrithms for Context-Based Search*, 61 J. OF THE AM. SOC'Y FOR INFO. SCI. & TECH. 1258 (2010); Andrew R. Conn & Sébastien Le Digabel, *Use of Quadratic Models with Mesh-adaptive Direct Search for Constrained Black Box Optimization*, 28 OPTIMIZATION METHODS AND SOFTWARE 139 (2013).

[42] REMZI ARPACI-DUSSEAU & ANDREA ARPACI-DUSSEAU, OPERATING SYSTEMS: THREE EASY PIECES 2 (2015).

[43] Hard disk drives stand in stark contrast to the now-defunct floppy-disk drive (FDD). Floppy disks are similar to hard disks in theory, but are far less sturdy and more suspect to a number of hardware failures. The use of floppy disks declined due to the ubiquity of flash memory, which is the type of storage utilized in an SSD.

[44] Solid-state drives include not only the drives that are inside some computers, but also the very common USB "flash" drive. REMZI APPACI-DUSSEAU & ANDREA APPACI-DUSSEAU, *supra* note 42, at 2. These drives are typified by having no moving components (hence, solid state) and by containing so-called flash memory (generally of the NAND variety, discussed *infra*), which is charged (or "flashed") during the writing process and then remains charged until it is overwritten. *See* Chris Preimesberger, *Data Storage: NAND Flash Memory: 25 Years of Invention, Development*, EWEEK (Apr. 11, 2012), www.eweek.com/c/a/Data-Storage/NAND-Flash-Memory-25-Years-of-Invention-Development-684048 [https://perma.cc/LM74-4AYT] (stating that charging NAND memory earned the name "flashing" because it was similar to a camera flash).

[45] V.S. Semenov, *Magnetic Field of Thin Film Heads for Longitudinal Recording on Hard Disks*, TECHNICAL TOOLS IN CONTROL 1735, 1735–36 (Aug. 1, 2011).

[46] *Id.* at 1736.

[47] *Id.* at 1736–37.

[48] SHAN X. WANG & ALEX M. TARATORIN, MAGNETIC INFORMATION STORAGE TECHNOLOGY 4 (1999).

An arm, known as the read-write head, moves across the surface of the platter.[49] Each platter requires two read-write arms—one for the top surface, and one for the bottom.[50] It is possible to have an HDD that stacks multiple storage platters, with each platter being given its own pair of arms.[51] Unlike a record player, the read-write arm does not physically interact with the platter. Rather, it cruises along a few nanometers above the storage medium and only interacts with the platter via a magnetic field.[52]

The second, and more modern, type of drive is the solid-state drive, known as an SSD. SSDs are composed of banks of electrical cells, arranged into grids.[53] These grids are sectioned off to make "pages," which are separated into "blocks."[54] Each individual cell is made up of what is called "flash" memory—specifically NAND flash, which is itself comprised of floating transistors.[55] Once NAND flash memory is charged, it retains that charge even when it is not powered up, making it a nonvolatile memory storage device.[56] Instead of reading and writing data using a physical arm that has to move over the surface of a platter, SSDs check the charge of the NAND gates.[57] This is faster than the read-write capacity of an HDD by several orders of magnitude.[58]

There are some drawbacks to SSDs, however. The most glaring of these is that the electrical components tend to degrade after many read-write cycles and become slower.[59] Erasing data is also more difficult on an

---

[49] *Id.*

[50] *Id.* at 6–8.

[51] *Id.* at 7.

[52] *Id.* It is for this reason that you should never move an HDD while the computer is still running or without properly shutting off the drive. When a drive is properly powered down, the read-write arms retract from the platters. If they are simply killed without going through the proper shut-down process, the read-write arms are still engaged. Any movement of the drive at this point may cause the arms to come into contact with the platters and not only destroy memory, but actually ruin the drive itself.

[53] K. ESHGHI & R. MICHELONI, INSIDE SOLID STATE DRIVES, SSD ARCHITECTURE AND PCI EXPRESS INTERFACE 19–25 (2013).

[54] *Id.*

[55] *Id.*

[56] *Id.*

[57] *Id.*

[58] HDDs can access or write to memory at a speed of 2,000–7,000 microseconds. SSDs can perform the same functions in 25-100 microseconds (for reading) or 250-1500 microseconds (for writing). Joel Hruska, *How Do SSDs Work?*, EXTREMETECH (Feb. 26, 2016), www.extremetech.com/extreme/210492-extremetech-explains-how-do-ssds-work [https://perma.cc/7MNH-D27M].

[59] This raises an interesting question about seizures. Because moving data on an SSD does degrade (albeit minutely) the quality of the drive, would the *movement of information* on an SSD be considered an interference with property rights and therefore require a separate authorization by warrant? Furthermore, can we calculate how many angels can dance on the head of a pin? In justice to the Scholastics, that question was never seriously contemplated in the Middle Ages, and was merely a Victorian slander.

SSD. As you will see in the Data Structure section below, information is rarely purged from a storage device. Rather, when information on a drive is no longer needed, it is flagged for overwrite. I will address the unique problems with SSD overwriting in the Data Structure section, where it more properly belongs, but bear in mind that there is always a sacrifice to be made when choosing between a magnetic-storage medium (e.g., an HDD) and a flash-memory medium (e.g., an SSD).

2. *Data Structure*

Computers operate on what is called an assembly language ("asm").[60] These languages are symbolic of the machine code instructions that the processor itself follows. Most high-level computing languages (C++, Perl, etc.) are metaphorical and must be translated into assembly (by a program called a compiler) before they can be executed.[61]

The graphical user interface ("GUI") that we see when we use a computer is merely a metaphorical representation of the binary data stored on the main drive. The relationship of the stored data (the ones and zeros) to the GUI is a symbolic one. The way data is stored on a drive is known as a *file system*.[62] File systems are generally divided into *directories*, which we can analogize to the folders that are visible in a GUI. These generally include indices of what they contain so the file system (and the GUI browser) can locate and retrieve their files.

The most common file systems are the New Technology File System ("NTFS")[63] and the Hierarchical File System ("HFS").[64] Most modern file

---

[60] *See, e.g.*, PAUL A. CARTER, PC ASSEMBLY LANGUAGE 11 (2006), http://pacman128.github.io/static/pcasm-book.pdf [https://perma.cc/CQV4-6NTS] (explaining how CPUs have their own assembly languages, where each statement in the assembly language directly represents an instruction in the machine language).

[61] *See id.* (stating that compilers convert high-level programming languages into low-level ones, and noting that high-level language statements are more complex than the statements made in assembly languages).

[62] *See Using Disks and Other Storage Media: Filesystems*, LINUX SYSTEMS ADMINISTRATORS GUIDE, http://www.tldp.org/LDP/sag/html/filesystems.html [https://perma.cc/MJ7R-L5HM] (last visited Mar. 16, 2017) ("A *filesystem* is the methods and data structures that an operating system uses to keep track of files on a disk or partition; that is, the way the files are organized on the disk." (emphasis in original)).

[63] NTFS is the file system used by all Windows NT-based operating systems. HELEN CUSTER, INSIDE THE WINDOWS NT FILE SYSTEM (Microsoft Press, ed. 1994). Both the Mac OS X kernel and Linux kernels have open-source drivers for NTFS file systems.

[64] HFS or HFS+ is the file system currently used by the Mac OS. *See* Susie Ochs, *Apple Previews Apple File System to Replace HFS+ in 2017,* MACWORLD (June 13, 2016 4:24 PM), http://www.macworld.com/article/3083406/storage/apple-previews-apple-file-system-to-replace-hfs-in-2017.html [https://perma.cc/55X4-M6HE] ("Apple's operating systems have used the HFS+ file system for more than 18 years . . . .").

systems keep journals to record changes in the storage of data.[65] In NTFS, this is known as the $LogFile.[66] In HFS systems, there are catalog and volume header files, which contain indexical information of where everything is stored.[67] These indices keep track of where information is located on the HDD or SSD and allow the operating system ("OS") to locate files and "remember" where they are by making reference to the index file where the record is stored.

Each file is also provided with metadata that is associated with it in the file system.[68] Metadata is data about other data; it includes, in an NTFS system, the log file (journaling changes to other data), the volume descriptor (containing information about the current volume),[69] the attribute definition table (defining certain character strings used by the OS), the root directory (pointing to the "core" or "home" folder of the NTFS system), the cluster allocation bitmap (which keeps track of the status of data in the system, namely whether there is data in certain regions or not), the volume boot code (which can be the actual code or a pointer to guide the system to finding it), etc.[70] The directories or *folders* also contain metadata. However, unlike the old directories of the FAT system, NTFS directories do not contain the metadata of their constituent files.[71]

Rather, in an NTFS system, which comprise most modern Windows-based computers, the files themselves each contain their own metadata.[72] These are the header (low-level data used by NTFS, which includes a number of pointers), and the standard information attribute (fundamental properties such as date/time stamps, read-only, hidden, system, volume label, directory, and archive tags).[73] These attributes are also recorded in an HFS+ system, though they are stored in separate arrays and data trees.[74]

---

[65] *Journaling Filesystems*, LINUX INFORMATION PROJECT, www.linfo.org/journaling_filesystem.html [https://perma.cc/2GKL-PLL4] (last visited Mar. 16, 2017).

[66] *How NTFS Works*, MICROSOFT (last updated Mar. 28, 2003) https://technet.microsoft.com/en-us/library/cc781134(v=ws.10).aspx [https://perma.cc/447Q-35VF].

[67] *See HFS+*, NTFS.COM, http://ntfs.com/hfs.htm (last visited Mar. 16, 2017) (noting that the volume headers and catalog files store data and contain records).

[68] *See generally* ARPACI-DUSSEAU & ARPACI-DUSSEAU, *supra* note 42.

[69] Volumes are partitioned regions of drives. *NTFS System (Metadata) Files*, PC GUIDE, www.pcguide.com/ref/hdd/file/ntfs/archFiles-c.html [https://perma.cc/53KT-HWXD] (last visited Mar. 16, 2017).

[70] *Id.*

[71] *NTFS Directories (Folders)*, PC GUIDE, www.pcguide.com/ref/hdd/file/ntfs/files.htm [https://perma.cc/3LZG-Q3GP] (last visited Mar. 16, 2017).

[72] *Id.*

[73] *Id.*

[74] *See HFS+,* NTFS.COM, www.ntfs.com/hfs.htm [https://perma.cc/KVY6-TM6G] (last visited Mar. 16, 2017) ("HFS+ uses B-trees to store most volume metadata.").

Files most commonly contain other metadata beyond these simple tags, though that is entirely dependent on the point of origin.[75] For example, modern digital cameras and cell phones often record the geolocation coordinates where pictures were taken.[76] The extent and nature of this secondary type of metadata is highly variable, but is still embedded in the file, even though users may not be able to see it.[77]

The primary metadata that defines the date of the file's creation and where it is stored cannot be scrubbed from either HFS+ or NTFS. The systems *require* this information in order to operate. If it were removed, the file would effectively be deleted. However, secondary metadata can easily be scrubbed from any and all files with the use of a text editor.[78] The block of text that precedes the file proper is almost always the secondary metadata section, which can regularly be erased without consequence.

3. *The Search Function*

Every modern operating system has an integrated search index. These search functions index the contents of the machine and perform several different searches on the indices in order to find the files that users are looking for. For example, the MacOS search engine, known as Spotlight, functions by keeping a logfile of all creations, modifications, and deletions.[79] As changes are made to stored data, the system kernel[80] sends

---

[75] *See* Karen Coyle, *Understanding Metadata and Its Purpose*, 31 J. OF ACAD. LIBRARIANSHIP 160, 160–63 (2005) (describing the "Creative Commons software," which adds "descriptive" metadata to academic works).

[76] *See* Karan Haider, *How to Turn off Geotagging in Android,* UBERGIZMO (May 4, 2015), http://www.ubergizmo.com/how-to/turn-off-geotagging-android/ [https://perma.cc/VW6G-5LXX] (explaining that most modern smartphones "geotag" the photos they take with location data); Kyle Schurman, *Geotagging Cameras*, LIFEWIRE (last updated Feb. 18, 2017), https://www.lifewire.com/geotagging-cameras-492865 [https://perma.cc/JZQ7-XKPQ] ("Some cameras have a built-in GPS unit, which allows the geotagging to be an automatic process.").

[77] Eileen B. Libby, *What Lurks Within: Hidden Metadata in Electronic Documents Can Win or Lose Your Case*, CENTER FOR PROFESSIONAL RESPONSIBILITY (Apr. 2007), http://www.americanbar.org/content/dam/aba/administrative/professional_responsibility/what_lurks_within.authcheckdam.pdf [https://perma.cc/8HJE-6MH8].

[78] Since all files are collections of data, it is possible to view that data as a textual representation. Opening an image file with a text editor will, for example, give you an insane cloud of letters, numbers, and symbols to look at.

[79] *See Guides and Sample Code*: *How Does Spotlight Work?,* APPLE INC., https://developer.apple.com/library/content/documentation/Carbon/Conceptual/MetadataIntro/Concepts/HowDoesItWork.html [https://perma.cc/CM5K-45B8] (last visited Mar. 16, 2017) ("Every time a file is created, modified, or deleted, the kernel notifies the Spotlight engine that it needs to update the system store for changed file.").

[80] The system kernel is the central portion of an operating system, which has control over all computing functions. *Kernel Definition*, LINUX INFORMATION PROJECT, www.linfo.org/kernel.html [https://perma.cc/WM5M-GMFA] (last visited Mar. 16, 2017). This is in contrast to a shell, which is the external portion of an operating system, usually one that interacts with the users. *Id.*

an alert to Spotlight.[81] Spotlight makes use of a plugin to then import the metadata of the file that has been altered.[82] The imported metadata is then stored in a separate file, in addition to the HFS+ journal.[83] When Spotlight is running, users can search for files that have the requested metadata in their systems.[84]

In order for our hypothetical concealed search to operate, it would also need access to the logfiles, indices, and search tables of the host computer. In order to keep law enforcement agents from accessing the computer outside the context of the concealed search system, and to enable ease of use, the simplest method for doing this would be to clone the target computer, then run a separate diagnostic operating system built specifically for this purpose. This specialized OS would essentially be providing the search functions, while cutting off all accessibility to the files within.

This custom OS only needs four functions: 1) the ability to import and read all files as text, 2) an indexing function, 3) a search function for a mass of text, and 4) an import function to allow the files which match the specified judicially approved searches to be exported to another environment where they can be opened and examined. Of course, there is the constant problem of evolving security measures, which plagues all computerized security schemes, from DRM[85] on down, but these will be addressed *infra* in Section D.

This would create a sort of black box, which is good *only* for finding and exporting the files that match the criteria approved on the warrant. Data could be fed into the black box in the form of cloned systems, read and dismantled into discrete text files, and then searched for appropriate metadata tags. In this endeavor, headers and secondary metadata would undoubtedly help.[86] However, it is possible, with the use of good

---

[81] *Guides and Sample Code: How Does Spotlight Work?*, *supra* note 79.

[82] *Id.*

[83] *Id*.

[84] *See Guides and Sample Code: Spotlight Metadata Attributes,* APPLE INC., https://developer.apple.com/library/content/documentation/Carbon/Conceptual/MetadataIntro/Concepts/SpotlightAttrs.html#//apple_ref/doc/uid/TP40001851-CJBEJBHH [https://perma.cc/8E2B-FX3X] (last visited Mar. 16, 2017) (noting that users can search Spotlight based on certain "metadata attributes").

[85] Digital Rights Management, being the most commonly attacked type of security, is that code installed in programs or documents that prevents people from running, using, or viewing it unless they have the proper authorization. *See, e.g.*, Julia Layton, *How Digital Rights Management Works*, HOWSTUFFWORKS TECH, http://computer.howstuffworks.com/drm.htm [https://perma.cc/6Z55-BZTW] (last visited Mar. 16, 2017) (explaining how DRM refers to any scheme used to prevent unauthorized access to copyrighted material). Essentially, these are antipiracy measures. Like every computerized security measure, it is only ever a matter of time before one becomes obsolete and is broken by the antisecurity public at large.

[86] Many files, in addition to the secondary metadata blocks, contain headers that tell their executing program certain information. For example, graphics files often have a header that inform viewing programs of their total number of colors, etc. These headers are extremely useful for determining file types even in the absence of traditional file endings. For example, a user might change

exemplars, to "match" different document types based solely on their content.[87]

Search functions are generally measured by comparing two axes common to all problems of pattern recognition and information retrieval: precision and recall.[88] Precision, also called predictive value, represents the fraction of retrieved data that is relevant.[89] Recall, also called sensitivity, is the fraction of relevant instances that the system retrieves.[90] This is essentially a problem of false positives and false negatives. In a system containing documents *a* through *z*, where *a*, *b*, *c*, *d*, *q*, and *x* are relevant to the search terms, a high-precision system will tend to return results that include a higher percentage of the relevant terms—thus, a result of *a*, *b*, *c*, *d*, *e*, *f*, and *x* would be a fairly precise result. However, a search result can be precise without having good recall. If the search instead returned *a* and *x*, it would be a very precise search but would have very poor recall. A search with low precision and high recall might include *a*, *b*, *c*, *d*, *e*, *f*, *g*, *h*, *i*, *j*, *q*, *x*, *s*, *t*, *u*, *v*. The result contains all of the relevant answers, but they are lost in a sea of noise, i.e. irrelevant returns.

One could even conceptualize the problems of recall and precision as precisely those problems, which are in legal tension here: the level of recall is the necessary scope of the search (large enough so that all relevant items are returned); the level of precision represents the remaining privacy rights as to the irrelevant information. The intersection of these two concerns is exactly what the law must find—the search function must have enough recall to find every single relevant item contained on the target system, while maximizing precision in order to protect as much "irrelevant" data as possible. Where most search functions seek to harmonize the levels of precision and recall,[91] here we must have one hundred percent recall and maximized precision.

---

a .jpg file so that it ended with a .doc ending, but if opened as text, the concealed search system would check the file header against all known types and recognize the .jpg-style header, indicating that an image had been disguised as a Word file.

[87] This would not work on encrypted documents, but the trouble with encryption is a wholly separate question that will plague the law whether or not the concealed search system is adopted.

[88] *See e.g.*, Rocio L. Cecchini et al., *Multiobjective Evolutionary Algorithms for Context-Based Search*, 61 J. AM. SOC'Y FOR INFO. SCI. & TECH 1258, 1259 (2010) (explaining that two objectives, precision and recall, can be used to define the effectiveness of a search query).

[89] *See, e.g.*, *Precision-Recall*, SCIKIT LEARN, http://scikit-learn.org/stable/auto_examples/ /model_selection//plot_precision_recall.html [https://perma.cc/H5ML-JXB8] (last visited Mar. 16, 2017) ("[P]recision is a measure of result relevancy . . . . A system with high precision . . . return[s] very few results, but most of its predicted labels are correct . . . .").

[90] *See, e.g.*, *id.* ("[R]ecall is a measure of how many truly relevant results are returned.").

[91] The classical search function optimization is called the F-measure or F1 score. This is measured at 2 * ((precision*recall)/(precision+recall)). *See, e.g.*, Powers, D.M.W., *Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation,* 2 J. OF MACHINE LEARNING TECH. 37 (2011).

The function therefore, of r = 100% where p is maximized, is the one that we must strive to find. Recall must be at one hundred percent to enable law enforcement to search every possible file that falls within the scope of the warrant. However, the more precise the search can be within that caveat or restriction, the more protective we can be of individual rights to privacy.[92] This, then, is how we can define our problem. We will have solved it when we have a concealed search capable of fulfilling that role.

4. *Types of Searches*

Not all search functions are created equal. There are a number of basic types of searches, each of which will be discussed below. All of them require at least a basic index in order to function. Therefore, we should first be concerned with the task of indexing a host machine's contents. In order to build an index, we need some way of tokenizing the contents of the cloned machine.[93] The process of tokenization is an exercise in and of itself, but at the very simplest level we could choose a number of basic units as "tokens" based on the common signatures in the file types we are concerned with.[94] Tokens are merely syntactic units, or tags that permit the sorting of data. Thus, tagging an image with the tokens "image," "building," "public building," would attach three tokens to it and allow those tokens to be searched.

Construction of the index will be the most complex part of the procedure. In order to construct an index that is helpful when looking at data that is not stored as pure text (images, sound, video), there will need to be a way to determine the content of those files. If the scope of the warrant simply permits law enforcement to seize "all video files," or "all

---

[92] This means essentially that we can, ourselves, define the scope of the search. Though the physical metaphor serves some value, it can only stretch to a certain extent. In reality, the recall/precision problem is a much more representative understanding of "scope" when it comes to computing. Files do not hide in places with size, they are instead either *relevant* or *irrelevant* to the search.

[93] *See generally* CHRISTOPHER D. MANNING ET AL., AN INTRODUCTION TO INFORMATION RETRIEVAL (Cambridge University Press, ed. 2009) (describing the ways in which search engines work).

[94] For example, if we are dealing primarily with image files, we would want to tokenize the most common units used to write those files, particularly in the metadata. Of course, the optimal black box search device would tokenize *every potential combination of import*. While this sounds like a lot of work, expert systems can be "trained" by exposure to documents and input from "trainers" to help them learn more quickly. STUART J. RUSSEL & PETER NORVIG, ARTIFICIAL INTELLIGENCE A MODERN APPROACH 22 (1995). Indeed, pattern recognition of the type that law enforcement is likely to be interested in (namely the ability to parse data in pictures and determine precisely what they are pictures of before returning them as relevant) is generally the type that modern neural network or "deep learning" artificial intelligences are being trained to do. *Id.* at 24. It would speed the process and the accuracy of a concealed search system immeasurably to be provided with an artificial intelligence of the deep learning type. Indeed, it may actually be required, when one takes into consideration the amount of image-data that law enforcement agents are likely to be searching in the future.

image files," that hardly makes the concealed search technique worthwhile going forward. The amount of effort it would take to utilize it would not, then, be commensurate with the amount of protection gained. Thus, in addition to simple tokenizing, we will need to begin real analysis of data content.

Tokenizing will obviously still be important, particularly for files that are formatted primarily in text. However, when text is contained in an image, without image recognition capability (such as optical character recognition), it is unreturnable as a token.[95] Thus, we need not have only the easily created tokens that emerge from the raw data, but second- and third-order tokens that can be assembled by having our hypothetical system perform pattern recognition passes on actual files.[96] This is addressed in the machine learning section, *infra*.[97]

Once the files have been indexed in some way, any number of search techniques can be applied to this index. Each of these techniques embodies an algorithmic way to search the indexed data.[98] By identifying our issue as the conjunction between absolute recall and a precision as high as is practicable, we actually also place the problem squarely in the indexing phase of the search. Search techniques used to peer through indexed data vary, but we can throw out any that have less than perfect recall.

The simplest search technique is to use brute force.[99] This takes the proffered search terms and compares each term to every lexical entry in the index one by one. When the terms match, the file which is indexed to that entry is returned as a "hit."[100] Obviously, this is the slowest possible way to search any amount of data. There are no shortcuts. It can be made slightly faster by spawning a parallel process for every term. Thus, if the search terms are "image: yes," "child," "cocaine," and "scales," the general term image would be applied to the three specific search terms and three processes would be spawned. One by one, the data entries in the index would be compared; if the token is not linked to an image file, it is discarded. If it is linked to an image file, process compares all tokens of that file, looking for "child." Process two does the same, crawling at its own rate along the index, and when it reaches an image file it searches for tokens containing "cocaine," and so on.

---

[95] *See id.* at 22 (evaluating the use of tokens and tokenization).

[96] *See id.* (second- and third-order tokens are tokens describing the content of first-order tokens).

[97] *Infra* Section III.A.6.

[98] Essentially, each major type of search function is its own mathematical formula, following a set order of permutations to be performed on the data.

[99] This is also called a linear or sequential search. *Sorting and Searching*, *in* DONALD KNUTH, THE ART OF COMPUTER PROGRAMMING 3 (1998).

[100] *Id.*

Computers are very fast, but a search of this type would be prohibitively slow.[101] Thus, computer scientists have developed a number of different search techniques that increase search speed: binary searches, insertion sorts, shell sorts, quicksorts, comparison sorts, and so on.[102]

Sorted data can be searched with a binary process.[103] "Sorted" data has already been transformed at least once: a sorted array must have categorized its contents into a continuum.[104] This is easiest to see with an index that contains numerical data.

Index Number	Contents
0	4
1	7
2	16
3	20
4	37
5	38
6	43

In this table, we can see that the data points (which may have been extracted in the order of 43, 16, 37, 20, 38, 4, 7) have been ordered by application of some kind of sorting algorithm already. In a binary search, the search would begin at entry 3, neatly dividing the searchable data in half. If entry 3 does not match the search criteria, the search function will be able to tell, as a logical necessity, that the desired information is either greater or less than 20 (which is the content of the indexical entry). Therefore, the algorithm will be instructed to check the upper or lower half of the index, and will do so by leaping to the center of the remaining entries, dividing the index a second time, and so on, until like Zeno's arrow, we reach our asymptote where the data is located.

So, in our example, if we are looking for entry 38, our binary search function will begin at entry 3 and see a data point of "20." It will then know that, because our search term is higher than 20, it will have to check the upper bounds (index entries 4-6). The next step the search function will

---

[101] Compared to other types of searches. Of course, the actual physical speed of the search will depend on the hardware being used. It should be noted that the process of indexing the data will take *far longer* than any of the proposed search techniques here.

[102] There are a number of other potential search techniques and new techniques are being developed all the time. The field of pattern recognition is a rapidly expanding one, particularly with big players such as Google working to increase the effectiveness of their pattern recognition algorithms all the time. *See generally* THOMAS NIEMANN, SORTING AND SEARCHING ALGORITHMS: A COOKBOOK.

[103] *Id.* at 5.

[104] *Id.* at 4.

take is to investigate index entry 5 (the midpoint of 4-6) and discover that the search term matches the entry. It has located our data.

This need not be limited to numerical entries. It is possible to transform and sort tokens that are written as tags by, for example, listing them in alphabetical order. This kind of index transformation goes into the pre-search efforts. The search itself merely parses the index according to a logical sequence.

The sorting techniques (insertion, shell, quick) are all steeped in the complex minutiae of computer science.[105] Insertion sorts perform data transformations on the index by grabbing, then deleting the indexical entries and then rewriting them at the top of the index. Thus, to reach our optimally sorted index listed above using the hypothetical entry order (43, 16, 37, 20, 38, 4, 7) an insert sort would check to see if the first two elements are in the correct order. Since they are not, 16 would be deleted and rewritten as the first entry. This process is repeated for each entry until they are, finally, arranged in order.[106]

Other sorting operations can be used, but it suffices to demonstrate this one to make the point. Sorting and searching are relatively simple elements of a search function. It is in the indexing process that things get complicated. Most of our barriers will arise when the question is asked "how do we tokenize this data?"

5. *Major Barriers*

The first and most obvious barrier that was discussed *supra* is the deletion of metadata. However, there are a number of other problems to confront as well: encryption, off-site storage (i.e. "cloud" storage), and image data. We will address these problems one by one.

Secondary metadata can be a rich source of information about files. It provides a number of ready-made tokens for indexing and thus would tend to speed up the indexing process as well as the subsequent search of that indexed data. Deletion of this metadata would make searches more laborious for both the software and for law enforcement agents working with it. It is possible that such deletion, in the context of a given search, may actually give rise to a reasonable necessity of expanding the scope of the warrant and permitting the return of, say, all image files.[107] However, the simple deletion of identifying information *should not* establish a

---

[105] *See generally id.*

[106] *Id.* at 8–9.

[107] Of course, this would be a legal argument that would be have to be upheld. However, it is reasonable to assume that, granted the expected return of certain types of data, when the indexing process discovers that metadata has been deleted wholesale from a specific type of file, it may allow an inference of wrongdoing. Indeed, it would be equivalent to executing a search warrant and discovering something in the process of execution that leads to probable cause for a second search warrant.

presumption of wrongdoing. Rather, it should be viewed in context with all the other information available and may present one factor towards a finding of probable cause.

The issue of encryption is a thorny one. Whether or not the concealed search is adopted, encryption will remain a problem. No search system will be able to properly index an encrypted file because its contents will be illegible except as random characters.[108] In order to address encrypted files, it would be necessary that this decryption software be used *before* the indexing procedure begins. Of course, the search system could flag those documents that it identifies as encrypted[109] and place them to one side. They could be exported, decrypted by some other process, and then reinserted for examination, or they could simply be set aside and subjected to new searches on their own.

In the case of off-site storage, that is well beyond the bounds of this inquiry. Off-site or "cloud" storage is, at its most extreme, a type of storage wherein the data is not held on the seized device but rather must be accessed by that device while the data itself is stored in some server room somewhere.[110] There is no protection for this kind of data because the user has already shared it with a third party—that is, whatever party is holding the data for them in their server room. There may be a societal expectation that this data is private, but it is not one that the courts have been prepared to recognize.[111] Therefore, there is no need for the concealed search when data is stored off site. There is, in fact, no need for a search warrant at all.[112]

Image data is the hardest of our problems to address. Unlike textual data, it is incomprehensible to computers without some kind of interpretation software. Computers, currently, do not generally "see" images the way we do.[113] This presents a problem for indexing—how can

---

[108] Encrypted files will also tend not to follow the common file templates that were discussed *supra*, so it would be impossible to, say, compare the headers with common header types.

[109] Probably by flagging every document that it cannot otherwise identify.

[110] This means that, yes, all data is being held somewhere. Any time you access data "online," it has a physical location in a server room somewhere *owned by some third party.* Those bits of data are in fact magnetic charges on some company's disk platter.

[111] This falls under the same heading as cell phone data, such as texts. *See e.g.*, *In re:* William E. Sharp/Kentucky State Police, Ky. Op. Atty. Gen. 11-ORD-144 (Ky. A.G.) (2011) WL 4478524.

[112] *See e.g.*, Smith v. Maryland, 442 U.S. 735, 743–44 (1979); 28 C.F.R. § 59.4(a) (2016) ("A search warrant should not be used to obtain documentary materials believed to be in the private possession of a disinterested third party unless it appears that the use of a subpoena, summons, request, or other less intrusive means of obtaining materials sought, and the application for the warrant has been authorized as provided in paragraph (a)(2) of this section."); *see also*, United States v. Graham, 824 F.3d 421, 427 (4th Cir. 2016) ("Under the third-party doctrine, an individual can claim 'no legitimate expectation of privacy' in information that he has voluntarily turned over to a third party.").

[113] That is changing. Google has been leading the charge when it comes to image recognition, but something as simple seeming as Facebook's face-recognition ability actually signals a sea change in the ways computers can interpret visual data. Optical character recognition is another way that computers

we tokenize data that is presented as a "block" of incomprehensible data? Well, for one, it would certainly be possible to count general colors present in an image based on the text data we can retrieve from it. So, for example, if law enforcement agents are looking for photographs taken at night, the search could return all images with a certain percentage of dark colors.[114] However, there are even more interesting and useful ways of having computers examine images, to the point where they can actually interpret their contents.[115]

6. *Neural Networks and Image Data*

Google and other major computing companies have been working to create what are called neural networks.[116] Neural networks are generally trained for a period of time by exposing them to the types of things they are meant to locate.[117] After many billions of iterations of training, they become adept at "looking at" images. That is, they can "see" and even *recreate* images, and not by a sort of rote memory where they reproduce the precise pixels of the original image. Google, in fact, has created a neural network that is capable of imagining wholly new images out of parts it has seen elsewhere.[118]

As this technology matures, it will become even more effective. As it is, image recognition software is good enough that it has been used by law enforcement authorities in the past and it will likely continue to be used in the future with greater accuracy. If a system such as this were added to the concealed search system, it would allow the indexing of images to a very precise degree: by analyzing images and tagging them with tokens based on what is contained therein, it would allow the search to determine whether they were relevant or not by comparing those tokens to the tokens

---

can "see" things, in that case (generally) being alphanumeric characters contained in .pdf files. *See* ADOBE, ADOBE ACROBAT, https://acrobat.adobe.com/us/en/acrobat/how-to/ocr-software-convert-pdf-to-text.html [https://perma.cc/9A3U-TN7U] (last accessed Oct. 30, 2016).

[114] This could conceivably be achieved by several means. The simplest would be to tokenize the color table used in the image. Image files, in order to minimize space, are generally compressed in such a manner as to refer to a central color table contained within the file. If that color table has primarily dark tones, or if the dark tones are referenced more often within the file, we may assume that the picture itself is dark with a very high degree of certainty. *See generally* CHARLES POYNTON, DIGITAL VIDEO AND HDTV: ALGORITHMS AND INTERFACES (2003); INDEXED COLOR AND PALETTES, A FEW SCANNING TIPS, www.scantips.com/palettes.html [https://perma.cc/WC37-YS3Z] (last accessed Dec. 21, 2016).

[115] *Infra* Section F.

[116] Neural networks are the latest and most promising type of artificial intelligence research, replacing Markov Chain decision makers, expert systems, and Bayesian networks.

[117] *E.g.,* MATHWORKS, NEURAL NETWORK TOOLBOX, https://www.mathworks.com/products/neural-network.html [https://perma.cc/C5LA-L2KP] (June 11, 2017).

[118] These can be very unnerving. *E.g.*, DEEP DREAM GENERATOR http://deepdreamgenerator.com/ [https://perma.cc/SF5K-5JFV] (last accessed Mar. 20, 2017).

approved by the warrant, and to return or not return them, as the case may be.

CONCLUSION

As we have seen, the technological potential to create such a system currently exists. The very legal conundrum of scope when dealing with digital information must be unhooked from the metaphor of the physical world. The scope metaphor does not apply to the digital world. Instead, we must replace that metaphor with a new understanding: the confluence of recall and precision.

Even if the concealed search is not adopted, the current understanding of digital searches must change. We are not talking about a world where there is no scope; we are not talking about a world where there is a scope that is somehow relatable to the physical. We need a totally new paradigm of understanding, one that requires us to restructure the manner in which we think about searches altogether.

Once that understanding is widespread, it will become clear that there should be some limit to imprecise searches when executing search warrants of devices that store data. Whether that limit is enforced by the concealed search or some other, so far unthought-of solution, remains to be seen.

APPENDIX I: SAMPLE DIGITAL SEARCH WARRANT APPLICATION

Digital Search Warrant Application
Police Case Number
Agency Name

Application for a Digital Search Warrant
To: A Judge of the Superior Court
The undersigned hereby applies for a warrant for the seizure and search of the below-described systems on the basis of the facts set forth in the:
affidavit below affidavit(s) attached

Date Signed (Prosecuting Authority)
Type/Print Name

Affidavit
The undersigned affiant, being duly sworn, deposes and says:

1. Neque porro quisquam est qui dolorem ipsum quia dolor sit amet.
2. Consectetur, adipisci velit.
3. This affiant witnessed the named defendant engage in a hand-to-hand drug sale in the aforesaid parking lot.
4. The above-listed confidential informant, having informed your affiant that the defendant frequently wears a GoPro bodycam that transmits one image through his buttonhole every 15 seconds, and that these images are stored on the below-named system, has probable cause to believe evidence of a crime will be found there, namely: photographs of the drug sale.

Description of System

1. Dell PC, serial number: *SAMPLE SERIAL NUMBER*
2. Windows 10 Operating System, NTFS

Requested Search Terms:
(file types)
image, text

(tokens)
cocaine, night, meeting (boolean add to cocaine only), parking lot, cars

(metadata)
stored during the period of 12/20-12/22
geolocation tag (parking lot)
file created by: gopro camera

Include All Encrypted Files? n

Date Signed
Subscribed and sworn before me on (date) Signed

Finding
The foregoing Application for a search and seizure warrant of digital information, and affidavit(s) attached to said Application, having been submitted to and considered by the undersigned, the undersigned finds from said affidavit(s) that there is probable cause to believe the evidence of a crime will be returned by this search and, therefore, that probable cause exists for the issuance of a warrant for the seizure and search of the above-named system.

Date and Signature