COGITATIO

Media and Communication (ISSN: 2183-2439)
2015, Volume 3, Issue 2, Pages 42-52
Doi: 10.17645/mac.v3i2.270

Article

The Copyright Surveillance Industry

Mike Zajko

Department of Sociology, University of Alberta, Edmonton, T6G 2H4, Canada; E-Mail: zajko@ualberta

Submitted: 9 April 2015 | In Revised Form: 20 June 2015 | Accepted: 13 July 2015 |
Published: 30 September 2015

Abstract
Creative works are now increasingly distributed as digital "content" through the internet, and copyright law has created powerful incentives to monitor and control these flows. This paper analyzes the surveillance industry that has emerged as a result. Copyright surveillance systems identify copyright infringement online and identify persons to hold responsible for infringing acts. These practices have raised fundamental questions about the nature of identification and attribution on the internet, as well as the increasing use of algorithms to make legal distinctions. New technologies have threatened the profits of some media industries through copyright infringement, but also enabled profitable forms of mass copyright surveillance and enforcement. Rather than a system of perfect control, copyright enforcement continues to be selective and uneven, but its broad reach results in systemic harm and provides opportunities for exploitation. It is only by scrutinizing copyright surveillance practices and copyright enforcement measures that we can evaluate these consequences.

Keywords
algorithmic; copyright surveillance; copyright enforcement; identification; internet

Issue
This article is part of the special issue "Surveillance: Critical Analysis and Current Challenges", edited by James Schwoch (Northwestern University, USA), John Laprise (Independent Researcher) and Ivory Mills (Northwestern University, USA).

1. Introduction

Much of our daily lives now involve interacting with digital "content". The relationships we have with these digital goods are governed in part by intellectual property rights, and a new industry has developed to take advantage of this legal fact. The copyright surveillance industry monitors the distribution and use of copyrighted works, identifies instances of copyright infringement, and responds against allegedly infringing uses and individuals. Dedicated companies use automated methods to operate at enormous scale, scanning millions of hours of audio and video each day, and bringing suit against hundreds of thousands of individuals each year. The questions I am posing are: how are the data packets and digital fragments passing through our computer networks identified as copyrighted content? How are these digital flows traced to identifiable individuals, and how are persons held responsible for internet traffic? What are the consequences of these determinations for data flows as well as people?

In short then, my research questions are about identification based on digital traces. One set deals with identifying traffic flows and content as intellectual property, the second set deals with identifying people and holding them accountable for traffic flows. Traffic and content are identified through algorithmic comparisons to known "signatures" or characteristics. Individuals can be identified by comparing numeric identifiers (IP addresses) recorded by monitoring software to logs maintained by internet service providers. Both methods can result in misidentification and reduce the complexities of copyright law to opaque decisions made by automated systems. My paper concludes by analyzing copyright trolling, a specific kind of surveillance and enforcement that combines the two forms of identification (of content and persons) in a particularly exploitative manner. I argue that systemic harms result from

today's wide-ranging copyright regimes, although copyright's enforcement remains contingent and uneven. The internet has been seen as a threat to copyright, but copyright surveillance and enforcement technologies have also come a long way in the internet era, and are now the tools of a profitable industry.

Copyright is a convoluted body of law with strange consequences for a digitally-networked society. It justifies constraints on our behavior, but is also routinely violated as we go about our daily lives (Greenberg, 2014, pp. 82-83). Copyright allows people and institutions to claim a monopoly in the use of a piece of writing, or an image, or the tiniest fragment of recorded sound. It enforces scarcity by restricting copying—an act that is essential to human creativity (Cohen, 2012, ch. 4) and also what our computers and networks are designed to do best. Because of this, digital networks have threatened copyright, but they also allow for pervasive forms of copyright surveillance and enforcement that have become the business model of a dedicated industry. The sweeping scope of this industry has disrupted the traditional "equilibrium" of copyright's under-enforcement (Balganesh, 2013b; Lessig, 2001, pp. 249-250). Copyright can now be enforced against persons and actions that would have previously escaped copyright owners' attention, but this enforcement is uneven and inconsistent. Rather than seeking total control over the distribution of creative works, copyright enforcement is selective, tolerating some uses and intervening against others. This is because copyright depends on private actors bringing forward claims of infringement, and the pursuit of such claims is not always advantageous or desirable. On the other hand, some of the actors described below have built businesses dedicated to pursuing "profit-based litigation" (DeBriyn, 2012) and demanding monetary settlements from scores of alleged infringers.

Copyright surveillance is an international business, and the copyright enforcement actions that follow an identification of infringement are often carried out without state involvement. But state-backed legal regimes remain in the background, with their threats of liability and sovereign violence. The internet sometimes still seems like a lawless place that frustrates state controls, but it consists of physical networks based in territories and jurisdictions. It is these networks' territorial basis that allows the state-backed monopolies of copyright to have any meaningful effect. This also makes copyright and the industries it supports vulnerable to legal and political reforms.

2. Surveilling Digital Flows for Intellectual Property

Much of the data circulating through our networks can be claimed as the intellectual property of some person or legal entity. Any unauthorized use or reproduction of this data can therefore be a violation of "intellectual property rights". The rightsholders of this digital content are often part of massive industries (most notably the music and film industries) that have turned their attention to the internet since the 1990s. As a consequence, a new industry has developed to offer copyright surveillance and enforcement as a service. This is an industry that depends on the fact of infringement to support its existence (Lobato & Thomas, 2013), even as it ostensibly fights to stop it.

Copyright was originally developed to regulate publishers and booksellers dealing in unauthorized copies, but today all of our computers make and circulate copies of cultural goods. Before home computers and the internet, a great deal of copying and circulation also took place on a regular basis. People made photocopies, VHS and cassette recordings, sang popular lyrics, and repurposed melodies. But this behavior largely escaped the notice of copyright owners. It was ephemeral, dispersed, impossible to track and difficult to control. The internet is different. Its traffic is visible by default, and content can be accessed around the world. The amount of internet traffic that arguably infringes copyright has become so large that human intervention cannot possibly keep up with it.

Fortunately for copyright owners, we now live in an age of algorithmic surveillance and algorithmic enforcement (Depoorter & Walker, 2013, pp. 333-335). Because algorithms are poor at adjudicating the intricacies of copyright law, these systems regularly generate false positives (see Depoorter & Walker, 2013, p. 335; Katyal, 2009, pp. 414-415). But they are effective enough to serve the interests of copyright owners, and are having a massive effect on the availability of online content. Algorithmically-selected links are now hauled off the web by the million, and with the assistance of internet service providers (ISPs), thousands of allegedly infringing individuals can be identified and threatened by the agents of copyright owners (copyright surveillance companies and law firms).

Below, I outline the global scale of the copyright surveillance and enforcement industry and analyze some of its common practices. This analysis is furthered by internal emails and documents from MediaDefender—once one of the industry's most notable companies. While MediaDefender surveilled peer-to-peer networks, algorithmic copyright surveillance and enforcement is increasingly built-in to internet services, with YouTube's Content ID system being the most notable example. My paper discusses the case of a video caught by Content ID's extra-judicial copyright enforcement system, before closing with a more recent trend, in which BitTorrent surveillance services have been used to drive profit-seeking copyright lawsuits in several countries. But first, I relate my interest in identification and data flows to previous work in the fields of internet and surveillance studies, where social theory has taken different approaches to related problems.

3. Digital Identification and Social Theory

Surveillance scholars have long been interested in what Clarke (1988) called "dataveillance": the surveillance of data generated by persons, and the tracking of persons through data (see Elmer, 2004, pp. 36-39). In one influential contribution to surveillance studies, Haggerty and Eriscson (2000, pp. 611-614) discuss how persons and their bodies are transformed into data, generating so-called "data doubles" that are then used as the basis of discriminations among populations. But a great deal of copyright surveillance is not interested in monitoring persons or populations. In these cases, the targets of surveillance and intervention are traffic flows—not people. Surveillance companies and their monitoring algorithms are often unconcerned about the identities of the persons that can be linked to, or held responsible for this traffic. The goal is to discriminate among content and act against anything identified as infringement, rather than against the infringing party.

The consequences of this kind of copyright enforcement are not limited to some separate realm of data or "pure virtuality" (Haggerty & Ericson, 2000, p. 611). While the distinction between online and offline, or digital and physical can still have its uses, it is the problematic link between them that deserves more focus. Increasingly, the trend in social theory has been away from "digital dualism" (particularly the dichotomy of "real" and "virtual") and towards an appreciation of how our reality is constituted through digital technologies and embodied experience, or the relationship between bodies and code (Jurgenson, 2012; Wellman & Haythornthwaite, 2002). YouTube videos do not reside and circulate in cyberspace. They are part of our world, and when a video is blocked there can be very real and material consequences for the persons involved (as seen in the Lansdowne Library case discussed below).

The other kind of copyright surveillance discussed herein does indeed target people through data, but approaches its problem from the opposite direction typically of interest to surveillance studies. Rather than "abstracting human bodies from their territorial settings and separating them into a series of discrete flows" (Haggerty & Ericson, 2000, p. 606), the process I am interested in here is how these digital flows are attributed to human bodies, or how people can be identified by "suturing or coupling of pieces of information in disjunctive time and scattered spaces" (Monahan, 2009, p. 158). The specific example of this process considered herein is "copyright trolling", in which an IP address linked to file-sharing activity must be translated into a street address and a particular resident (typically, the internet subscriber). This translation cannot be achieved simply through technical means—there is no method that an outside observer can use to independently pin the IP address (used to route packets of information) to an individual residence. While copyright surveillance companies can monitor file-sharing traffic and record the IP addresses of the devices involved, they must secure the compliance of an ISP to correlate these digital addresses with subscribers' street addresses. Typically, this compliance is achieved under the weight of the law governing the territory in question.

These concerns of copyright surveillance are related to two fundamental problems of our networked age. The first problem is control over the cultural objects and the information circulating through our networks (see Poster, 2006, p. 186). In other words, how internet controls can be achieved, and to what effect. The second problem is holding individuals accountable for data traffic, or how digital records can be sutured together to identify a person "behind" the internet traffic (see Poster, 2006, pp. 113-116). While these questions are now fundamental concerns for a variety of actors, it is important to understand how copyright surveillance and enforcement companies have taken to answering them, given the systemic harms that copyright regimes are capable of producing (see Cohen, 2012, ch. 4).

4. Copyright Surveillance at a Glance

As described above, copyright surveillance has two basic targets: content and persons (see Table 1 below). The vast majority of copyright surveillance does not aim to identify infringing individuals. Instead, algorithmic surveillance is used on a massive scale to identify copyrighted content by comparing digital fragments to particular "signatures". This involves systematically monitoring file-sharing protocols or "crawling" websites. The algorithms tasked with this surveillance look for certain file names or other characteristics of the content being distributed online, and compare these to a database of known copyrighted content. What happens when the algorithm detects potential infringement depends on the party conducting the surveillance (or more likely, the party paying for it as a service). Assuming that this party is a copyright owner, they may do nothing at all. Knowing what is being downloaded or shared across the internet can be useful information. Companies that provide copyright surveillance often promote their services as a way to gather market information or "business intelligence" (Lobato & Thomas, 2013). However, my main interest is copyright surveillance that is geared toward intervention. This can include having the content removed, making it more difficult to access, or targeting the persons allegedly infringing copyright. Individuals can be targeted for lawsuits, or be subject to private enforcement regimes like the US Copyright Alert System (Zimmerman, 2014), which (like Canada's "notice-and-notice" system, see Tarantino, 2012) notifies internet subscribers when their IP address has been linked to infringing activity.

Table 1. Two basic types of copyright surveillance and associated enforcement.

Target	Means of Identification	Possible Interventions	Examples
Content	Comparing file properties to a known signature	Takedown notices, automated filtering, traffic disruption	DMCA takedowns, YouTube Content ID, file interdiction (MediaDefender)
Persons	Recording IP addresses and reconciling these with an ISP's logs	Deterrent/educational notices, degraded internet service, lawsuits	HADOPI, Canadian notice-and-notice, US Copyright Alert System, RIAA file-sharing lawsuits, copyright trolling

Copyright surveillance is almost entirely the domain of private industry. While the French internet copyright regime (HADOPI) created a government agency dedicated to enforcement, the system relies on a private company to monitor the country's internet traffic (see Bridy, 2011, pp. 733-735). The copyright surveillance industry is modest in size (a large monitoring and enforcement company might have a few dozen employees), but it monitors an enormous scope of online activity and facilitates sweeping legal interventions. Some copyright owners employ small surveillance and enforcement firms and achieve massive reach by leveraging algorithmic methods (Farivar, 2012). Monitoring the public or quasi-public internet for copyrighted content can in theory be achieved at scale by anyone, including academics (Chothia, Cova, Novakovic, & Toro, 2013; Zhang, Dhungel, Wu, & Ross, 2011). Copyright surveillance companies do use specialized software, but they generally do not enjoy any privileged access to internet traffic.

The first of these copyright surveillance companies were founded in 1999 and 2000, during the rapid rise of the file-sharing service Napster (Doan, 2000) and the accompanying legal campaign to stop internet piracy. For several years this campaign by the music industry generated lawsuits against tens of thousands of US individuals accused of online infringement. Over the course of the mid to late 2000s these efforts were largely abandoned. Today they can be recognized as a failed attempt to criminalize widespread and normalized behavior (Bachmann & Jaishankar, 2011; Harris, 2012). However, copyright surveillance has continued towards other ends, such as targeting web services and search engines. In recent years, mass file-sharing lawsuits have resurfaced, but these have generally been oriented towards generating revenue for minor copyright owners rather than deterring infringement of major creative works.

Deeper insight into the copyright surveillance industry was made possible in 2007, when six months of internal files and emails from US-based MediaDefender appeared online (see Roth, 2008; Zetter, 2007). At the time of this supposed hack of the company, MediaDefender was one of the more notable firms in the industry, having been purchased for $43 million in 2005 (Mennecke, 2005). The company was working to expand into a number of business opportunities, including helping to identify individuals sharing child pornography, and its own video download service. However, the majority of MediaDefender's business activity involved "protecting" particular titles for copyright owners by monitoring several file-sharing networks for newly released or soon-to-be released titles. When these files were found, downloads could be disrupted through various means. These included flooding file-sharing networks with "decoy" or "spoof" versions (which appear genuine, but are instead unplayable, limited to promotional content, or redirect to an approved source. See Anderson, 2007; Katyal, 2003, pp. 356-358).

MediaDefender did not generally collect evidence or IP addresses for use in litigation, but it did record and share data on file-sharing for other purposes. These included answering queries from copyright owners about the amount of file-sharing in a particular country or region. In two e-mail exchanges, rightsholders asked the company about the popularity of individual songs being considered for release as singles. MediaDefender's clients included some of the world's biggest copyright owners (Universal, Paramount, Sony BMG). One "small monitoring contract" with a major record label paid $10,000 a month to monitor three file sharing networks for the presence of particular files. Different levels of protection, for different lengths of time, were offered for between $5,000 and $15,000 per title (Anderson, 2007). In one email exchange during 2007, it was estimated that the company of approximately 60 employees was working on around 3,000 projects at once, with company servers pushing out around 3 billion decoy and spoof files a day.

MediaDefender's fortunes declined following the compromise of its files in 2007, and the company eventually went out of business. But the information disclosed about its operations can still tell us several things. First, even a small firm can monitor and intervene against file sharing on a massive scale. MediaDefender's methodology combined algorithmic discrimination with human judgment, but it was the algorithms that enabled its broad scope. The following section elaborates on how this algorithmic copyright enforcement has evolved since MediaDefender's heyday, through built-in systems such as YouTube's Content ID. Afterwards, I will turn to the topic of copyright surveillance for the purposes of personal identification.

5. Caught in the YouTube Vortex

In 2012 the Lansdowne Public Library and its Teen Advisory Board in Pennsylvania made a video promoting reading and uploaded it to YouTube. The video parodied Michael Jackson's "Beat It" (Read It, 2012), featuring teens dancing and singing about reading. In less than three days the video was identified as potentially infringing copyright and taken down from YouTube. In their efforts to restore the video over the following year, library staff would need to navigate the tangles of copyright law, content ownership, and algorithmic enforcement.

It was unclear who had been responsible for the takedown in the first place, since YouTube's takedown system is automated, but operates under the direction of copyright owners. The system initially referred the library to Warner/Chappell Music (Mengers, 2013), but Jackson's music has been transferred to Sony/ATV. The librarian who had filmed the video filled out the forms to appeal the decision and sought licensing from Sony (Schwartz, 2012). She also made personal appeals to Sony, which included travelling to New York and trying to enter Sony's offices. At one point, Sony claimed that they wanted the lyrics in the video changed. Later the company allowed to video to be put online, but only on the library's website and not on any other site, and only for a limited time period (Mengers, 2012). After national news media began covering the story, Sony moved to allow the video to be re-instated (Schwartz, 2012).

Was the video infringing? Was it fair use (see Schwartz, 2012)? Because of the legal uncertainties and gray areas of copyright law, these legal distinctions can only be made by a court (see Katyal, 2009, pp. 411-412; Lee, 2008). But the absence of a court's judgment did not prevent an algorithmic judgment. Months later, the same YouTube video had its audio muted through an automated enforcement action. Once again, an algorithm had been tripped, silencing the library and its teens. Sony denied being behind the muting. According to the Library's director, Sony claimed that they did not have the power to restore the audio, and that the content had been caught in the "YouTube vortex" (New Media Rights, 2013). The library phoned YouTube but could not speak to a human being (Mengers, 2013). One of the librarians eventually submitted a claim through YouTube's online appeal process, but she needed the help of a lawyer to craft a fair use argument which would be effective in having the audio restored (Mengers, 2013; New Media Rights, 2013). What eventually turned out to be a copyright success story required exceptional efforts on the part of library staff, as well as legal help to properly engage with YouTube's enforcement regime and appeal its algorithms.

6. Scan and Notice

Online copyright enforcement is generally meant to deny or restrict the availability of content. Denial can be achieved either by directly disrupting access (as in MediaDefender's interdiction efforts), through built-in enforcement regimes such as YouTube's (see below), or by using existing copyright laws to issue what are known as "takedown notices" for content (Lobato & Thomas, 2013, p. 615). Millions of pieces of content are targeted by such notices every week, which can be effective wherever ISPs and online services are required to take them seriously. The processing of takedown notices is crucial for these companies to maintain "safe harbor" protection under the copyright laws of the US, EU, and Canada (among other nations, see Fernández-Díez, 2014; Tarantino, 2012). Safe harbor protects companies providing internet services against liability for infringement carried out by their users. However, this protection often only applies to companies as long as they remain unaware that their users are infringing copyright (Fernández-Díez, 2014, pp. 67-69). Under safe harbor, service providers have an incentive to limit what they know about their users' activities. When a legitimate takedown notice arrives informing them of infringement on their service, service providers are obliged to take action.

As a consequence, copyright surveillance companies have been algorithmically flagging infringement across the web and sending a growing deluge of notices to major content hosting platforms. Online service providers have to decide whether each notice is legitimate and should be complied with. Google processed a weekly average of between seven and nine million URLs in late 2014 (Google, 2015), with each request from a major copyright owner typically listing thousands of URLs for removal. Just as the employees of copyright surveillance companies use automated scanning and algorithmic discrimination to create these lists, Google uses its proprietary blend of algorithms and human review to decide which takedown notices should be complied with, and which should be rejected (Google rejected less than 1% of these notices in 2013, see Google, 2014, p. 13).

YouTube (owned by Google) maintains a similar system for handling takedown notices, but also operates its own automated system for identifying infringing content, known as Content ID. This proactive system of identifying infringement was developed while YouTube was embroiled in a billion-dollar lawsuit with Viacom (which accused YouTube of not taking action against videos that it knew were infringing, see Zimmerman, 2014, pp. 264-265). Rightsholders provide YouTube with "reference files" of their content, and the site scans each uploaded video looking for a match. If Content ID matches an uploaded video to one of its 25 million or so active reference files, the copyright owner can choose to block the video (or mute its audio), show ads, or track its viewership (Google, 2014).

The algorithms behind Content ID have been re-

fined over the years, and its appeals process has been elaborated and extended (La Rosa, 2014; Zimmerman, 2014, p. 272). Still, Content ID's proactive orientation exceeds the requirements of US law, and Google announced that in implementing it the company "goes above and beyond [its] legal responsibilities" (King, 2007). Google has created an extensive copyright monitoring and enforcement system that operates without court involvement, in part to keep the company from facing another massive lawsuit by rightsholders. In its effort to proactively police copyright, YouTube processes a staggering amount of video through Content ID (Google, 2014), and the system has helped channel over a billion dollars in advertising revenue to copyright owners (La Rosa, 2014). But those caught on the wrong side of YouTube's judgments, as in the Lansdowne Library case (see also Tarantino, 2012) have had to suffer the costs, without the transparency and due process that a court could provide (Zimmerman, 2014, p. 273).

Built-in monitoring and copyright enforcement systems are increasingly the norm for popular media-sharing websites (La Rosa, 2014). With growing numbers of people creating and distributing content online (Poster, 2006, pp. 244-249), these private copyright enforcement regimes are having a major effect in controlling the distribution of cultural goods. Algorithmic judgments may not carry the same weight as court orders, but they are effectively the law of these digital domains (see Lessig, 2006). However, some companies have combined the use of algorithmic surveillance and discrimination with the enforcement powers of the courts. They do so in order to link identifications of copyright infringement to individual persons. An entire business model has developed in recent years around identifying individuals tied to copyright infringement and compelling them to pay large penalties. The result, known as copyright trolling, might be the most exploitative use of copyright enforcement in the digitally-networked era.

7. Lawsuits and BitTorrent Trolls

Identifying persons is a relatively minor concern for the type of copyright surveillance described earlier: what matters is whether internet traffic includes copyrighted content, and whether it can be controlled. However, in the early 2000s many major US copyright owners felt they could achieve control through deterrence—by identifying and suing thousands of individuals accused of sharing songs. Their efforts failed (DeBriyn, 2012, pp. 84–85), and for a time copyright owners' lawsuits shifted from individuals to institutions (like YouTube and The Pirate Bay) that allegedly facilitated infringement. But by 2010, a new approach took hold among some of the more marginal copyright owners and their lawyers. Courts once again saw thousands of persons targeted in infringement suits, and judges were asked to help identify these defendants on the basis of IP addresses.

The actors bringing these sorts of suits are often described as "copyright trolls", but there are disagreements about just what distinguishes a troll from a more legitimate plaintiff. Trolling operations vary and are legally opportunistic, and it has proven difficult to define copyright trolls in a way that captures more than a portion of such operations (see Sag, 2015). Because of this, I avoid labeling any specific companies as copyright trolls. Instead (and largely in agreement with Sag, 2015), I refer to copyright trolling as a practice—one that threatens large numbers of individuals with copyright infringement claims, with the primary goal of profiting from settlements (or default judgments) rather than proceeding to trial on the merits of a case (see Curran, 2013).

While major copyright owners can engage in trolling, they generally prefer not to. This is typically the domain of smaller companies that do not receive large profits through sales and licensing, and see settlements as an easy way of generating revenue from individuals who are not paying for their works. Trolling is "profit-based litigation" (DeBriyn, 2012, p. 86), and to be successful it depends on accused infringers fearing the price of statutory damages and settling for smaller amounts. In the large subset of copyright trolling cases dealing with pornography, the pressure on defendants can be amplified by the fear of being publicly associated with pornography titles (Curran, 2013).

Copyright trolling is a strategy that depends on linking internet traffic to particular individuals, which is where copyright surveillance companies come into play. These companies monitor online traffic, record the IP addresses involved in sharing certain files, and hand the list to a law firm. The law firm then undertakes the next step by approaching the ISP that assigned the IP addresses, and having the ISP consult its logs to determine which address was assigned to which subscriber at a given time. Frequently, this requires a court to compel the ISP to disclose the subscriber's information (Anderson, 2010). The copyright surveillance company does not enter into the process again unless the plaintiff is forced to further substantiate the claim of infringement before a court.

Copyright trolling (sometimes called "speculative invoicing") is often thought of as a particularly American practice, since statutory damages in the US can be up to $150,000 per work infringed, and the average cost of defending a copyright infringement case through trial (excluding judgment and awards) ranges between $384,000 and $2 million, depending on the size of the copyright claim (Am. Intellectual Law Ass'n, of the Economic Prop. Report Survey 2011, cited in Balganesh, 2013a, p. 2280; Depoorter & Walker, 2013). This makes settling for between $1500 and $5000 a more attractive option, which has led many commen-

tators to liken the process to extortion or "legal ransom" (Curran, 2013). However, the legal strategy of copyright trolling has also seen extensive use in the UK (*Golden Eye [International] Ltd. & Anor v Telefonica UK Ltd*., 2012) and may have been pioneered in Germany (Lobato & Thomas, 2013, p. 618; Roettgers, 2011). In 2010, tens of thousands of individuals had been sued in this manner in the US (Anderson, 2010). By 2011 the number exceeded 200,000 (Ernesto, 2011) and was possibly much higher in Germany (Roettgers, 2011). By 2014, copyright trolling cases made up the majority of copyright cases filed in several US federal court districts (Sag, 2015). No one knows just how many individuals have settled in these cases or how much money has been collected in total, although millions of dollars have clearly been paid to different trolling operations.

While some trolling operations have specialized in the copying of images, news articles, and audio samples (Curran, 2013; Polonsky, 2012), my focus is on trolling cases that target file-sharing on BitTorrent. The BitTorrent protocol rose to popularity in the mid-2000s as a way of distributing and sharing large files, which made it ideal for videos. Since this protocol operates as a distributed system and has no central point of control, it cannot be shut down by court order in the same way as earlier file-sharing systems such as Napster and Kazaa. Traffic on BitTorrent is highly visible however, since each downloaded file is received as many small pieces from numerous users (all of whom constitute a "swarm"), and each of these contributors can be identified by an IP address. BitTorrent activity can be monitored through a number of means (Chothia et al., 2013), but in essence, it is by joining a swarm that one can record the IP addresses of all those who are also participating in it.

Just as the activities and IP addresses of downloaders and uploaders are largely visible on BitTorrent, so are the activities of copyright surveillance companies (Chothia et al., 2013; Ernesto, 2012). However, we know little about their methods, since these have rarely been submitted as evidence and examined in court. Copyright trolling cases, by and large, do not proceed to trial. This might be because the costs of litigating a case exceed what might be recovered as a settlement, but also because of the risk of an unfavorable judgment against the troll. The information obtained through monitoring BitTorrent is relevant primarily for the "discovery phase" of a suit, where a court order is sought to compel an ISP to identify its subscribers. With the identities of alleged infringers in hand, a troll can then proceed to demand settlements from them, and the information used to make these demands need never be assessed as evidence in a court of law. In a typical copyright trolling case, the subscriber's name and home address is all that is needed to send out a settlement letter (demanding payment of a few thousand dollars to make the suit disappear). However, depending on the plaintiff and the defendant's actions, further investigations can be carried out. In some US copyright trolling cases, defendants arguing their innocence have undertaken polygraph tests or had their computers searched by forensic examiners (Malibu Media LLC, 2014). Defendants in these cases must decide how far they are willing to go to demonstrate their innocence (versus paying the settlement), and plaintiffs must decide how far they are willing to go to pursue a settlement or judgment (houstonlawy3r, 2013).

Ultimately, the copyright troll business model depends on legal regimes and judges that can facilitate these sorts of actions. In the US, judges have by and large granted the court orders sought to identify subscribers, but legal decisions since 2013 have limited the ability of trolling cases to sweep up thousands of individuals at once (houstonlawy3r, 2013; Ren, 2013; Sag, 2015). The most egregious trolling practices have also faced the threat of legal sanctions in US courts (Haslach, 2013). In a significant UK case, a judge granted a court order to identify suspected infringers, but imposed conditions on the manner in which settlement offers could be made (*Golden Eye ([nternational] Ltd. & Anor v Telefonica UK Ltd*., 2012). Similarly, an attempt to identify thousands of subscribers in Canada was met with reservations from a judge who, citing privacy concerns and the "spectre" of the copyright troll, imposed conditions that would limit the opportunities to profit from settlement demands (*Voltage Pictures LLC v. John Doe and Jane Doe*, 2014). The case was subsequently cited to justify similar conditions in a precedent-setting Australian file-sharing suit (*Dallas Buyers Club LLC v iiNet Ltd.*, 2015).

While some of the most exploitative opportunities for trolling have been foreclosed by the above judgments, a few copyright surveillance companies have found ways to secure compliance from ISPs to identify alleged infringers without proceeding through the courts. The most notable of these has been Rightscorp, which pursues settlements for just $20, albeit on a mass scale (Mullin, 2014). In Canada, CEG-TEK has pursued somewhat larger settlements by taking advantage of the country's new copyright enforcement regime, which requires ISPs to forward notices from copyright owners to subscribers (Roberts, 2015). Just as some courts have come to oppose copyright trolling, new business models based on copyright surveillance and identification are being developed to fit changing legal environments.

8. Pervasive Surveillance, Contingent Enforcement

Widespread and vigorous copyright enforcement can be justified by the harm that infringement causes: reducing profits for artists and creative industries, thereby limiting incentives for the production of new creative works. While it is demonstrably false that every infringing act results in harm (particularly as some un-

authorized uses are actively encouraged by copyright owners, see Lee, 2008), it is undeniable that forms of infringement such as file-sharing have, to some extent, been "revenue-depleting" for certain industries (Bridy, 2011, p. 711). If the aim of copyright is to lessen such harm by reducing infringement, then pervasive surveillance and severe enforcement might be legitimate approaches. If there are reasons to believe that a heavy-handed approach to copyright enforcement is counter-productive to this aim (see Bachmann & Jaishankar, 2011; Harris, 2012), then as with any form of illegality, we might debate which kinds of infringement are best addressed through which enforcement measures, or to what extent the law is being "overenforced" (Balganesh, 2013b).

It is not my objective in this paper to determine the appropriate balance between the rights of copyright owners and users, or how best to combat infringement. Instead, my interest is in the rise of the copyright surveillance industry and its consequences. Digital media and networks have made it easier than ever for individuals to copy and distribute copyrighted content, but monitoring and enforcement technologies now also have a global reach. Copyright owners previously had limited insight into how their works were used and distributed (particularly for non-commercial purposes) and little ability to control such behavior. But Content ID can scan hundreds of years of video fed into YouTube each day (Google, 2014), and similar systems are being adopted by a growing number of media-sharing platforms. With limited resources, millions of IP addresses connected through BitTorrent can be monitored (Zhang et al., 2011), and the resurgence of mass litigation against file-sharers has seen hundreds of thousands of these IP addresses brought before courts for identification. These developments have dramatically extended areas of contact between individuals and copyright owners.

A number of authors have raised the fear that copyright enforcement systems were transforming our networked society into a dystopia of total surveillance and "perfect regulation" (Lessig, 2006, pp. xiii-xv) or "perfect [law] enforcement" (Mulligan, 2008). However, while copyright enforcement systems are now widely deployed, they form an uneven regulatory patchwork that is far from perfect in its discriminations. In the cases examined above, haphazard contingencies determine whether or not enforcement measures come into effect. Content ID does not enforce all copyright equally (enforcement depends on the rightsholder), and BitTorrent trolls can choose among legal jurisdictions and ISPs when seeking court orders. Many forms of copyright surveillance are not tied to any enforcement actions at all, and copyright owners frequently tolerate unauthorized use of their works for promotional purposes (Lee, 2008).

Perfect enforcement is therefore an impossible and undesirable goal, even for many rightsholders. Instead, we see pervasive copyright surveillance and uneven, contingent enforcement. As a consequence, individuals are left uncertain about which actions will be tolerated and which will be pursued as instances of infringement (Katyal, 2009, p. 418), or what uses of copyrighted content qualify as "fair" (Katyal, 2009, pp. 411-413; Lee, 2008). False positives also occur regularly as automated systems misidentify content, or copyright owners assert illegitimate claims (Depoorter & Walker, 2013). Individuals wishing to contest these claims can be left in the position of the Lansdowne Library video producers, unsure of how or where to appeal a judgment. Those who are misidentified as infringers by copyright trolls are left weighing the price of a settlement against the costs of demonstrating their innocence in court (Balganesh, 2013a). Copyright enforcement might be inconsistent and uncertain, but those caught in its net experience significant harms.

This paper's selection of cases has been used to make three broad points. First, many kinds of mass copyright surveillance can be carried out with limited resources, and there is sufficient demand for these services to support a small, dedicated industry. Surveillance companies like MediaDefender demonstrated the vast reach of their algorithmic methods in the early 2000s, and web giants like YouTube can scale their monitoring capabilities to match the vast volumes of content passing through their servers. The second point is that these algorithmic judgments are inherently imperfect and unevenly applied, contributing to deep uncertainties in copyright enforcement. The harm caused when automated methods misidentify or overreach against infringement can be significant, as in the Lansdowne Library video, and those affected may have limited recourse. As a third point, it is important to recognize the systematic harms that result from expansive copyright enforcement regimes (Cohen, 2012, ch. 4), even when these operate within the law. The phenomenon of copyright trolling combines mass copyright surveillance and mass litigation, extracting settlements out of as many people as possible, but not submitting evidence to the scrutiny of a trial. Such efforts disrupt copyright's traditional "equilibrium" of under-enforcement (Balganesh, 2013b) by pursuing non-commercial cases of infringement which were largely outside the scope of enforcement before internet technologies facilitated both widespread sharing and mass surveillance. Therefore, while uses of internet technologies have harmed the traditional business models of some copyright owners, the systematic harms enabled by the business of copyright surveillance and enforcement also need to be acknowledged.

9. Conclusion

The commercialization of internet activity since the

1990s has entailed treating some digital flows as intellectual property, the idea being that much of the content circulating through the internet has an "owner" with exclusive rights to its distribution. The copyright surveillance and enforcement industries serve their clients by identifying copyrighted works and controlling their distribution, as well as identifying individuals allegedly engaged in infringement. Algorithmic tools are used to solve these problems at scale, so that even small monitoring firms can have massive reach. There is sufficient demand for these services to ensure that copyright surveillance and enforcement systems will continue to be developed and refined, particularly as private enforcement regimes like Content ID are adopted across a growing number of online platforms.

While copyright enforcement is driven by private actors, it depends on state authorities and credible threats of legal action to compel compliance and assistance from third parties. The legal foundation of these efforts makes them vulnerable to changing judicial attitudes, as well as political reforms to copyright regimes. Copyright trolling in particular has faced a growing backlash in recent years, with some courts limiting or imposing supervision over such operations. But innovative new ways have been developed to generate revenue from systematic copyright claims. These kinds of exploitative enforcement will remain a danger as long as digital flows can easily be attributed to individuals, and copyright regimes impose large monetary penalties for commonplace behavior.

Not all of the concerns discussed in this paper are specific to intellectual property rights. Any attempt to screen the vast volumes of data in circulation for illegality will require the use of algorithms to make legal distinctions, or ways to hold individuals accountable for digital flows. But the reason why copyright has been such a powerful driver of internet governance debates and policies is the large financial interest that media industries have in controlling the distribution of creative works. This same interest now supports a copyright surveillance industry, which in turn enables widespread copyright enforcement, along with enforcement's systemic harms. It can be argued that these harms are the price of preserving copyright in the era of digital networks, but there are now many examples of how such an approach can go too far, and reasons to wonder whether this is an appropriate justification. It is only by closely attending to the effects of copyright regimes and the practices they support that we can make an informed judgment on the best way forward. This requires not just the active interest of scholars, but also that copyright regimes (especially algorithmic and extra-judicial regimes) be open to scrutiny.

Acknowledgments

This paper benefited from the guidance of Kevin Haggerty and the opportunity to present at the 2014 Surveillance & Society conference.

Conflict of Interests

The author declares no conflict of interests

References

Anderson, N. (2007). Peer-to-peer poisoners: A tour of MediaDefender. *Ars Technica*. Retrieved from http://arstechnica.com/tech-policy/2007/03/mediadefender

Anderson, N. (2010). US anti-P2P law firms sue more in 2010 than RIAA ever did. *Ars Technica*. Retrieved from http://arstechnica.com/tech-policy/news/2010/10/us-anti-p2p-law-firms-sue-more-in-2010-than-riaa-ever-did.ars

Bachmann, M., & Jaishankar, K. (2011). Suing the genie back in the bottle: The failed RIAA strategy to deter P2P network users. In *Cyber criminology: Exploring internet crimes and criminal behavior* (pp. 155-172). Boca Raton: CRC Press.

Balganesh, S. (2013a). Copyright infringement markets. *Columbia Law Review*, *113*, 2277-2332.

Balganesh, S. (2013b). The uneasy case against copyright trolls. *Southern California Law Review*, *86*, 723-781.

Bridy, A. (2011). Is online copyright enforcement scalable? *Vanderbilt Journal of Entertainment & Technology Law, 13*(4), 695-737.

Chothia, T., Cova, M., Novakovic, C., & Toro, C. G. (2013). The unbearable lightness of monitoring: Direct monitoring in BitTorrent. In A. D. Keromytis & R. D. Pietro (Eds.), *Security and privacy in communication networks* (Vol. 106, pp. 185-202). Heidelberg: Springer.

Clarke, R. (1988). Information technology and dataveillance. *Communications of the ACM*, *31*(5), 498-512.

Cohen, J. E. (2012). *Configuring the networked self*. New Haven, CT: Yale University Press.

Curran, L. S. (2013). Copyright trolls, defining the line between legal ransom letters and defending digital rights: Turning piracy into a business model or protecting creative from internet lawlessness? *John Marshall Review of Intellectual Property Law*, *13*, 170-644.

Dallas Buyers Club LLC v iiNet Ltd. (2015). No. 317 (FCA 2015).

DeBriyn, J. (2012). Shedding light on copyright trolls: An analysis of mass copyright litigation in the age of statutory damages. *UCLA Entertainment Law Review*, *19*(1), 79-112.

Depoorter, B., & Walker, R. K. (2013). Copyright false positives. *Notre Dame Law Review*, *89*(1), 319-360.

Doan, A. (2000). NetPD wants to be web's police department. *Forbes*. Retrieved from http://www.forbes.com/2000/05/05/mu9.html

Elmer, G. (2004). *Profiling machines: Mapping the personal information economy*. Cambridge, MA: The MIT Press.

Ernesto. (2011). 200,000 BitTorrent users sued in the United States. *TorrentFreak*. Retrieved from https://torrentfreak.com/200000-bittorrent-users-sued-in-the-united-states-110808

Ernesto. (2012). Anti-pirates caught spying on thousands of torrents. *TorrentFreak*. Retrieved from http://torrentfreak.com/anti-pirates-caught-spying-on-thousands-of-torrents-120829

Farivar, C. (2012). Microsoft outsources copyright enforcement to small Redmond company. Retrieved from http://arstechnica.com/business/2012/05/microsoft-outsources-copyright-enforcement-to-small-redmond-company

Fernández-Díez, I. G. (2014). Comparative analysis on national approaches to the liability of Internet intermediaries for infringement of copyright and related rights. *WIPO*. Retrieved from http://www.wipo.int/export/sites/www/copyright/en/doc/liability_of_internet_intermediaries_garrote.pdf

Golden Eye (International) Ltd. & Anor v Telefonica UK Ltd. (2012). 723 (Ch) (EWHC 2012).

Google. (2014). How Google fights piracy. *Google*. Retrieved from https://drive.google.com/file/d/0BwxyRPFduTN2NmdYdGdJQnFTeTA

Google. (2015). Copyright removal requests. *Google*. Retrieved from https://www.google.com/transparencyreport/removals/copyright

Greenberg, B. A. (2014). Copyright trolls and presumptively fair uses. *University of Colorado Law Review*, *85*, 53-128.

Haggerty, K. D., & Ericson, R. V. (2000). The surveillant assemblage. *The British Journal of Sociology*, *51*(4), 605-622.

Harris, D. P. (2012). The new prohibition: A look at the copyright wars through the lens of alcohol prohibition. *University of Tennessee Law Review*, *80*, 101-211.

Haslach, M. R. (2013). Trouble for trolling: Courts reject copyright trolling tactics. *Washington Journal of Law, Technology & Arts*, *9*(2), 93-115.

houstonlawy3r. (2013, December 6). CLF 2013, a year in review. *Federal Computer Crimes*. Retrieved from http://cyberlawy3r.wordpress.com/2013/12/06/2013-a-year-in-review

Jurgenson, N. (2012). When atoms meet bits: Social media, the mobile web and augmented revolution. *Future Internet*, *4*(1), 83-91.

Katyal, S. (2003). The new surveillance. *Case Western Law Review*, *54*(2), 297-385.

Katyal, S. (2009). Filtering, piracy surveillance, and disobedience. *Columbia Journal of Law & the Arts*, *32*(4), 401-426.

King, D. (2007). Latest Content ID tool for YouTube. *Google Blog*. Retrieved from http://googleblog.blogspot.com/2007/10/latest-content-id-tool-for-youtube.html

La Rosa, D. (2014). YouTube pays out over $1 billion through Content ID. *Videoter*. Retrieved from http://videoter.com/youtube-content-id-pay-1-billion

Lee, E. (2008). Warming up to user-generated content. *University of Illinois Law Review*, *2008*(5), 1459-1548.

Lessig, L. (2001). *The future of ideas: The fate of the commons in a connected world*. New York: Random House.

Lessig, L. (2006). *Code: And other laws of cyberspace* (2nd ed.). New York: Basic Books.

Lobato, R., & Thomas, J. (2013). The business of anti-piracy: New zones of enterprise in the copyright wars. *International Journal of Communication*, *6*, 606-625.

Malibu Media LLC. (2014). *Plaintiff's status and informational report for its cases in the Northern District of Illinois* (Malibu Media LLC v. John Doe). United States District Court for the Northern District of Illinois. Retrieved from http://www.scribd.com/doc/216662394/Gov-uscourts-ilnd-292270-17-0

Mengers, P. (2012). Lansdowne teens' video turns into standoff with Sony, Michael Jackson estate. *Delaware County Daily Times*. Retrieved from http://www.delcotimes.com/general-news/20121126/lansdowne-teens-video-turns-into-standoff-with-sony-michael-jackson-estate-with-videos

Mengers, P. (2013). Chalk up another win for Lansdowne kids; audio restored to "Read It" video. *Delaware County News Network*. Retrieved from http://www.delconewsnetwork.com/articles/2013/10/10/news_of_delaware_county/news/doc52567162b99d9565138161.txt

Mennecke, T. (2005). ArtistDirect Purchases MediaDefender. *Slyck News*. Retrieved from http://www.slyck.com/story875_ArtistDirect_Purchases_MediaDefender

Monahan, T. (2009). Identity theft vulnerability: Neoliberal governance through crime construction. *Theoretical Criminology*, *13*(2), 155-176.

Mulligan, C. M. (2008). Perfect enforcement of law: When to limit and when to use technology. *Richmond Journal of Law & Technology*, *14*(4), 1-49.

Mullin, J. (2014). You could be liable for $150k in penalties—settle instead for $20 per song. *Ars Technica*. Retrieved from http://arstechnica.com/tech-policy/2014/06/meet-rightscorp-the-internets-new-for-profit-copyright-cop

New Media Rights. (2013). Teens make parody video, but Sony tells them to beat it… just beat it! *New Media Rights*. Retrieved from http://www.newmediarights.org/teens_make_parody_video_sony_tells_them_beat_it%E2%80%A6_just_beat_it

Polonsky, I. (2012). You can't go home again: The

righthaven cases and copyright trolling on the Internet. *Columbia Journal of Law & the Arts*, *36*(1), 71-101.

Poster, M. (2006). *Information please: Culture and politics in the age of digital machines*. Durham: Duke University Press Books.

Read It. (2012). Retrieved from https://www.youtube.com/watch?v=ZjwzcBTCxOo&

Ren, P. (2013). Fate of BitTorrent John Does: A civil procedure analysis of copyright litigation. *Hastings Law Journal*, *64*(4), 1343-1380.

Roberts, J. J. (2015). US firm runs mass copyright shakedown in Canada. *Gigaom*. Retrieved from https://gigaom.com/2015/03/06/us-firm-runs-mass-copyright-shakedown-in-canada

Roettgers, J. (2011). P2P Lawsuits gone wild. *Gigaom*. Retrieved from http://gigaom.com/2011/01/14/p2p-lawsuits-gone-wild

Roth, D. (2008). Hacking: The pirates can't be stopped. *Condé Nast Portfolio*. Retrieved from http://www.danielroth.net/archive/2008/01/hacking-the-pir.html

Sag, M. (2015). Copyright trolling, an empirical study. *Iowa Law Review*, *100,* 1105-1147.

Schwartz, M. (2012). Posting a parody video? Read this first. *Library Journal*. Retrieved from http://lj.libraryjournal.com/2012/11/copyright/posting-a-parody-video-read-this-first

Tarantino, B. (2012). Online infringement: Canadian "notice and notice" vs US "notice and takedown." *Entertainment & Media Law Signal*. Retrieved from http://www.entertainmentmedialawsignal.com/online-infringement-canadian-notice-and-notice-vs-us-notice-and-takedown

Voltage Pictures LLC v. John Doe and Jane Doe. (2014). No. T-2058-12 (FC 2014).

Wellman, B., & Haythornthwaite, C. (2002). Introduction. In *The Internet in everyday life* (pp. 3-41). Malden: Blackwell.

Zetter, K. (2007). Hackers smack anti-piracy firm MediaDefender again and again. *Wired*. Retrieved from http://www.wired.com/politics/security/news/2007/09/mediadefender

Zhang, C., Dhungel, P., Wu, D., & Ross, K. W. (2011). Unraveling the BitTorrent ecosystem. *IEEE Transactions on Parallel and Distributed System*, *22*(7), 1164-1177.

Zimmerman, D. L. (2014). Copyright and social media: A tale of legislative abdication. *Pace Law Review, 35*(1), 260-285.

About the Author

Mike Zajko

Mike Zajko is a PhD Candidate in the Department of Sociology at the University of Alberta, with an interest in communications and surveillance studies. He is currently studying the roles and responsibilities of internet intermediaries, including the responsibilities of internet service providers to facilitate surveillance, copyright enforcement, and cyber security.