

# Rumo à Caracterização de Disseminação Ilegal de Filmes em Redes BitTorrent

Adler Hoff Schmidt, Marinho Pilla Barcellos, Luciano Paschoal Gasparly

<sup>1</sup>Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)  
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

{ahschmidt,marinho,paschoal}@inf.ufrgs.br

**Abstract.** *Content sharing through BitTorrent (BT) networks accounts nowadays for a considerable fraction of the Internet traffic. Recent monitoring reports revealed that the contents being shared are mostly illegal and that movie is the most popular media type. Research efforts carried out to understand content production and sharing dynamics in BT networks do not provide precise information in respect to the behavior behind illegal film dissemination, being this the main objective and contribution of this paper. To perform such analysis, we monitored during 30 days all film torrent files published on the main BT public community. Furthermore, we joined the respective swarms, without downloading content, in order to obtain additional information regarding illegal sharing. As result, we present, characterize and discuss who produces and who publishes torrents of copyright-infringing files, what is produced and who acts as first provider of the contents.*

**Resumo.** *O compartilhamento de conteúdo por meio de redes BitTorrent (BT) é atualmente um dos principais responsáveis pelo volume de dados na Internet. Relatórios de monitoração recentes constataram que os conteúdos sendo compartilhados são, em ampla maioria, ilegais e que filme é o tipo de mídia mais comum. Esforços de pesquisa realizados para entender a dinâmica de produção e de compartilhamento de conteúdo em redes BT não oferecem informações precisas sobre o comportamento por trás da disseminação ilegal de filmes, sendo esse o principal objetivo e contribuição deste artigo. Para realizar tal análise, monitorou-se todos os arquivos torrent de filmes publicados na principal comunidade pública de BT durante 30 dias e ingressou-se nos enxames, sem compartilhar conteúdo, a fim de obter informações adicionais acerca de compartilhamento. Como resultado, apresenta-se, caracteriza-se e discute-se quem produz e quem publica torrents de cópias ilícitas, o que é produzido e quem atua como primeiro provedor dos conteúdos.*

## 1. Introdução

Redes BitTorrent (BT) são atualmente a principal opção para usuários compartilharem conteúdo através da Internet [Schulze and Mochalski 2009]. Segundo estudo apresentado pela Envisional [Envisional 2011], aproximadamente dois terços dos 2,72 milhões de *torrents* administrados pelo principal rastreador BT são de conteúdos ilícitos, algo que reforça a noção intuitiva de BT ser largamente utilizado para compartilhar arquivos que infringem direitos autorais. O mesmo estudo aponta que 35,2% desses conteúdos ilícitos são cópias ilegais de filmes.

Apesar de existirem algumas publicações caracterizando compartilhamento de conteúdo em redes BT [Zhang et al. 2010b, Zhang et al. 2010a, Le Blond et al. 2010, Cuevas et al. 2010], nenhuma concentrou-se em observar aspectos específicos do processo de disseminação de conteúdos ilícitos, e muito menos de cópias ilegais de filmes (como, por exemplo, usuários responsáveis pela criação das cópias, tecnologias de digitalização utilizadas, etc). Pouco se sabe, por exemplo, sobre quem são os usuários responsáveis por criar cópias ilegais, quais são os processos de digitalização empregados, quem publica *torrents* desses conteúdos, e quem fomenta, nos estágios iniciais, os enxames formados em torno de cópias ilegais.

Algumas razões que justificam a importância de entender o compartilhamento ilegal de filmes em redes BT são discutidas a seguir. Primeiro, esse tipo de conteúdo é o principal responsável pelo volume de tráfego dessas redes. Segundo, caracterizar fidedignamente a atividade dos disseminadores desses conteúdos é base para formular mecanismos de combate a esse comportamento indesejado. Terceiro, responsáveis por conteúdos (no caso deste artigo, filmes) protegidos por direitos autorais podem amparar-se em conhecimento acerca do comportamento de usuários mal intencionados e criar estratégias para minimizar proliferação indevida de cópias ilegais.

Diante do problema e da motivação em abordá-lo, neste artigo apresenta-se resultados de um estudo experimental sistemático realizado para caracterizar disseminação ilegal de filmes em redes BT. Procura-se desvendar quem produz e quem publica cópias ilícitas, o que é produzido e quem atua como primeiro provedor. Além disso, estabelece-se relações entre agentes envolvidos e realiza-se exercício visando observar dinâmicas existentes (e não facilmente perceptíveis) no processo de disseminação ilegal de filmes. Para realizar o estudo, estendeu-se e utilizou-se a arquitetura de monitoração TorrentU [Mansilha et al. 2011], desenvolvida pelo grupo. Em um mês de monitoração, obteve-se 11.959 *torrents*, 1.985 nomes de usuários da comunidade, 94 rastreadores e 76.219 endereços IP únicos. Observou-se, ainda, atividades realizadas por 342 grupos de digitalização.

O restante do artigo está organizado como segue. A seção 2 apresenta conceitos acerca de compartilhamento ilícito de filmes e discute trabalhos relacionados. A seção 3 discorre sobre a arquitetura de monitoração e as decisões tomadas quanto à sua instanciação. A seção 4 relata e discute os resultados obtidos. A seção 5 encerra o artigo com considerações finais e perspectivas para trabalhos futuros.

## 2. Fundamentos e Trabalhos Relacionados

Esta seção está organizada em duas partes. Primeiro, revisita práticas adotadas por grupos de digitalização e características dos processos utilizados para digitalizar conteúdo. Na sequência, descreve e discute os trabalhos relacionados de maior relevância.

### 2.1. Conceitos Associados ao Compartilhamento Ilícito de Filmes

Grupos de digitalização são os responsáveis pela criação, através de meio ilícitos, de cópias de filmes [Wikipedia 2011a]. Eles podem ser compostos por um ou mais membros e recebem crédito pela sua atividade agregando a sua identificação ao nome dos *torrents* por eles criados. Via de regra, consumidores experientes não reconhecem um *torrent* de filme como confiável (e evitam usá-lo) caso ele não identifique o grupo de digitalização

responsável. Logo, essa “etiqueta” é obedecida tanto por produtores quanto por consumidores. Ela provê uma maneira de dar notoriedade aqueles que estão realizando essa atividade ilegal, ao mesmo tempo em que os grupos buscam, para preservar sua reputação, assegurar o “casamento” correto entre nome dos *torrents* e conteúdos, bem como a correta classificação da qualidade desses *torrents*.

A decisão de usuários em realizar (ou não) *download* de determinadas cópias é influenciada, também, pela identificação, nos *torrents*, dos processos de digitalização empregados [Wikipedia 2011b]. A tabela 1 lista oito processos amplamente utilizados. Cada um deles é caracterizado por uma sigla, por uma fonte, *i.e.*, a mídia a partir da qual a cópia ilícita é gerada, e por uma expectativa de tempo, após a primeira estréia oficial do filme, para se encontrar uma cópia autêntica gerada usando o processo em questão. Os processos aparecem na tabela em ordem crescente de qualidade resultante esperada para as cópias digitalizadas.

**Tabela 1. Processos de digitalização**

Sigla	Fonte	Lançamento
CAM	Gravado no cinema	1 Semana (S)
TS	Gravado no cinema com fonte exclusiva de áudio	1 S
TC	Material sendo projetado no cinema	1 S
PPVRip	Exibição para clientes de hotéis	8 S
SCR	Cópia distribuída a críticos e usuários especiais	Imprevisível
DVDScr	DVD distribuído para usuários especiais	8-10 S
R5	DVD não editado, lançado somente na região 5	4-8 S
DVDRip*	DVD acessível ao público	10-14 S

\* Digitalizações a partir de fontes de maior qualidade, como Blu-ray, foram consideradas DVDRip.

## 2.2. Trabalhos Relacionados

Nos últimos anos a comunidade de pesquisa em redes par-a-par produziu alguns trabalhos ligados à monitoração de redes BT. Nesta seção descreve-se e discute-se os mais relacionados ao presente artigo. Por questão de escopo, são organizados em três grupos: infraestruturas de monitoração, caracterizações gerais do “universo” BT e caracterizações detalhadas de produção e consumo em redes BT.

Bauer *et al.* [Bauer et al. 2009] propuseram uma infraestrutura de monitoração que realiza medições ativas. A monitoração consiste em contatar rastreadores, obter endereços IP e contatar *hosts*, para confirmá-los como participantes “válidos” de enxames BT. Jünemann *et al.* [Junemann et al. 2010] desenvolveram uma ferramenta para monitorar *Distributed Hash Tables* (DHT) associadas a enxames BT. A ferramenta divide-se em três módulos. O primeiro permite coletar dados da rede par-a-par, como a quantidade de pares, endereços IP, portas utilizadas e países de origens, ao percorrer a DHT. O segundo analisa os dados e gera gráficos de acordo com métricas definidas pelos usuários. O terceiro busca, nos resultados do segundo módulo, valores que excedam limiares estipulados pelo usuário, gerando avisos. Ainda no campo de infraestruturas de monitoração, Chow *et al.* [Chow et al. 2007] apresentaram BTM: um sistema para auxiliar a detecção de pirataria que lança mão de monitoração automática de enxames BT. Ele é organizado em módulos responsáveis, respectivamente, pela procura de *torrents* na rede e pela análise

dos mesmos. O discernimento entre quais dos materiais monitorados violam direitos autorais, e quais não, é completamente realizado pelo usuário através das regras que podem ser definidas para processamento dos dados coletados.

No que se refere a caracterizações gerais do “universo” BitTorrent, Zhang *et al.* [Zhang et al. 2010b] analisaram *torrents* de cinco comunidades públicas. A descoberta de pares deu-se através de comunicação com rastreadores ou consulta a DHTs. Os autores apresentam, entre outros aspectos, quais são as principais comunidades de BT, os graus de participação de cada publicador de *torrents*, as cargas e localizações dos principais rastreadores, a distribuição geográfica dos pares e as implementações de clientes BT mais utilizadas. Seguindo uma metodologia similar à desse trabalho, Zhang *et al.* [Zhang et al. 2010a] realizaram uma investigação sobre *darknets* em BT, *i.e.*, comunidades privadas. Entre os resultados apresentados, os autores comparam características de enxames impulsionados por *darknets* com de enxames “oriundos” de comunidades públicas. Como observação geral desses dois estudos, ressalta-se o interesse em “fotografar” momentos do ciclo de vida de enxames BT na tentativa de quantificá-los e de abstrair modelos. Não fez parte de seu escopo, contudo, analisar dinâmica e caracterizar padrões de disseminação de conteúdo ilícito.

Passando ao último grupo de trabalhos analisados, Blond *et al.* [Le Blond et al. 2010] monitoraram por 103 dias as três comunidades de BT mais populares, traçando perfis dos provedores de conteúdo e dos consumidores mais participativos. Conseguiram identificar 70% dos provedores, listar os principais conteúdos sendo compartilhados e caracterizar os participantes mais ativos (pares presentes em vários enxames). Cuevas *et al.* [Cuevas et al. 2010] investigaram os fatores socioeconômicos de redes BT, ressaltando os incentivos que os provedores de conteúdos têm para realizar essa atividade. Três grupos de publicadores foram definidos: os motivados por incentivos financeiros, os responsáveis por material falso e os altruístas. Com um mês de medições, esses grupos foram caracterizados por *Internet Service Providers* (ISPs) aos quais estão associados, tipos de conteúdos disponibilizados, incentivos para as suas atividades e renda monetária especulada. Os trabalhos de Blond *et al.* e Cuevas *et al.* são os que mais se assemelham ao apresentado neste artigo, em especial no nível detalhado de monitoração e nas técnicas empregadas. Destaca-se, entretanto, que o escopo desses trabalhos não foi o de analisar disseminação ilegal de filmes nem tampouco de conteúdo ilícito em geral. Logo, aspectos que desempenham papel importante na compreensão de esquemas ilegais de distribuição foram deixados de lado. Além disso, os trabalhos parecem apresentar limitações técnicas importantes. A título de exemplo, incluem nas estatísticas resultados associados a *torrents* com pouco tempo de vida nas comunidades (forte indicador de conteúdo falso), comprometendo análises realizadas e conclusões obtidas.

Nesta subseção revisou-se alguns dos trabalhos mais relevantes e correlatos a este artigo. Observa-se um esforço da comunidade de pesquisa em redes par-a-par em criar ferramental de monitoração e conduzir caracterizações. Nenhum dos trabalhos, porém, preocupou-se em investigar como redes BT vêm sendo usadas para disseminação ilegal de filmes e de outros conteúdos ilícitos. Acredita-se ser este um tópico de grande relevância, em especial para que se possa, com conhecimento de dinâmicas até agora obscuras, subsidiar a proposição de estratégias e mecanismos eficazes que propiciem a proteção de

conteúdo protegido por direito autoral e, até mesmo, contribuir para ampliação do uso de redes BT em cenários mais sensíveis. Até onde sabemos, este é o primeiro trabalho que procura mapear, de forma sistemática, processo de disseminação de conteúdo ilegal em redes BT. As próximas seções detalham a arquitetura de monitoração empregada, aspectos de sua instanciação e os principais resultados obtidos.

### 3. Infraestrutura de Monitoração Utilizada

Esta seção apresenta a infraestrutura de monitoração utilizada. A subseção 3.1 apresenta a arquitetura de monitoração empregada, denominada TorrentU, e extensões implementadas para permitir a caracterização almejada. Em seguida, a subseção 3.2 detalha como a arquitetura foi instanciada.

#### 3.1. Arquitetura de Monitoração TorrentU e Extensões

TorrentU [Mansilha et al. 2011] é uma arquitetura flexível projetada e desenvolvida para permitir a monitoração de redes BitTorrent. Como a figura 1 ilustra, a arquitetura segue a abordagem clássica gerente/agente e, portanto, possui basicamente dois componentes: observador e telescópios. Observador é o componente que faz o papel de *front-end*, isto é, gerente, permitindo que o operador configure o sistema e observe os dados coletados em tempo real (assim como o histórico dos dados). Telescópios, por sua vez, atuam como agentes, sendo os componentes responsáveis pela monitoração do universo BitTorrent e pelo retorno de resultados de acordo com as requisições enviadas pelo Observador. Telescópios são subdivididos em três partes, denominadas “lentes”, sendo cada uma responsável por monitorar um grupo diferente de elementos do universo: comunidades, rastreadores e pares. Essa modularização permite que as lentes existentes possam ser substituídas, assim como novas possam ser facilmente incorporadas na arquitetura (sem modificação de seus componentes essenciais).

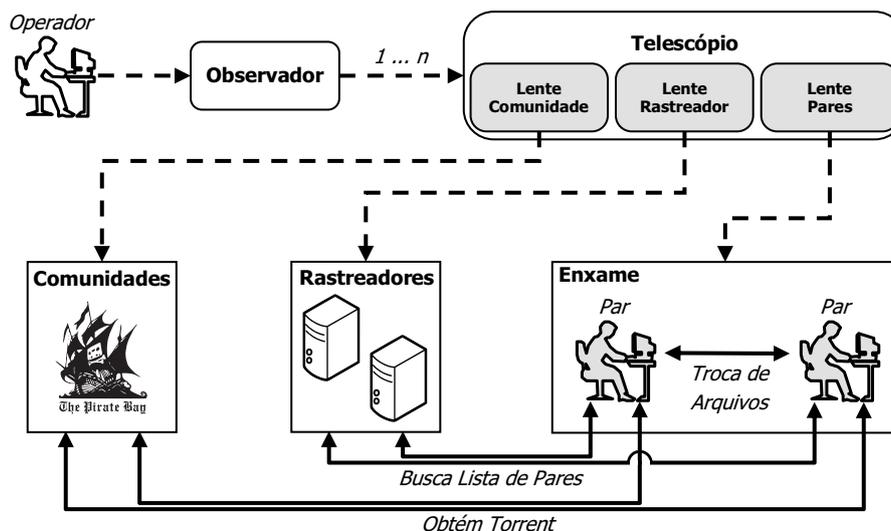


Figura 1. Arquitetura TorrentU

Lançando mão da flexibilidade oferecida por TorrentU, algumas funcionalidades, originalmente não contempladas pela arquitetura, foram implementadas e integradas. Entre as extensões, destacam-se duas. A primeira, criada para permitir a identificação dos

primeiros pares semeadores de enxames, consiste em lente que captura *torrents* logo que publicados em comunidades. A segunda extensão, também materializada por meio de nova lente, realiza monitoração contínua das páginas de *torrents*, armazenando tempo de vida dos enxames, números de semeadores e sugadores, e testemunhos postados. O objetivo, nesse caso, é a produção de “fotografias” de enxames ao longo do tempo.

O algoritmo 1 apresenta uma visão geral do procedimento de monitoração executado. Como pode-se perceber pela descrição das funções, os *torrents* recentemente publicados são capturados. Para cada *torrent*, o(s) respectivo(s) rastreador(es) é/são caracterizado(s) e mensagem(ens) para obtenção de lista(s) de pares participantes, enviada(s). Caso obtenha-se essa(s) lista(s), um processo de caracterização dos primeiros pares participantes do enxame é realizado e, ao finalizá-lo, a lente responsável pela captura de fotografias da comunidade é iniciada para o *torrent* em questão. São cinco parâmetros que determinam o comportamento desse algoritmo. *Tempo* determina quanto durará a campanha de monitoração. *Rodadas* indica o número de tentativas a serem realizadas para contatar rastreadores. *Quantidade* representa o tamanho da lista de pares requisitada aos rastreadores. *Limiar* determina em quais enxames será feita troca de mensagens *bit-field*. *Periodicidade* consiste no intervalo de tempo a ser respeitado para produzir cada fotografia de um dado enxame. *Intervalo* representa o tempo de espera entre cada rodada de execução do algoritmo.

**input:** *tempo, tentativas, quantidade, limiar, periodicidade e intervalo*

```
for  $i \leftarrow 0$  to tempo do
  torrents[]  $\leftarrow$  CapturarTorrentsRecentes();
  for  $j \leftarrow 0$  to torrents.size() do
    torrent  $\leftarrow$  torrents[ $j$ ];
    DownloadTorrent(torrent);
    LerArquivo(torrent);
    CaracterizarRastreadores(torrent);
    peerList  $\leftarrow$  ObterListaPares(torrent, tentativas,
    quantidade);
    CaracterizarPares(peerList);
    if peerList.size() < limiar then
      | TrocarBitfields(torrent);
    end
    IniciarCapturaFotografias(torrent, periodicidade);
  end
  Esperar(intervalo);
end
```

**Algoritmo 1:** Visão geral do procedimento de monitoração

Ressalta-se que a ênfase deste artigo reside nos resultados da caracterização de disseminação ilegal de filmes e não na descrição da arquitetura de monitoração. Ao leitor interessado em detalhes acerca do funcionamento da arquitetura sugere-se consulta a artigo anterior [Mansilha et al. 2011] produzido pelo nosso grupo de pesquisa.

### 3.2. Instanciação da Arquitetura

Telescópios foram instanciados em três nodos do PlanetLab [PlanetLab 2011] e em um servidor privado. O objetivo dessa redundância foi, basicamente, tolerar falhas e evitar descontinuidade do processo de monitoração. Já o componente Observador foi instanciado em uma única estação. Entre as comunidades BitTorrent existentes, optou-se por monitorar o PirateBay [Piratebay 2011]. Tal deve-se ao fato de ser a comunidade aberta mais popular, disponibilizar somente *torrents* publicados em seus servidores, manter registro de usuários responsáveis pela publicação de cada *torrent*, e prover classificação de cada usuário baseada em sua reputação.

O processo de monitoração foi instanciado utilizando-se as seguintes configurações (podem ser entendidas como parâmetros recém detalhados do algoritmo 1): 2 tentativas para obtenção de lista de pares com cada rastreador, 50 pares por lista, limiar definindo máximo de 10 pares com os quais serão trocadas as mensagens *bitfield*, periodicidade de 8 horas entre cada fotografia da comunidade e intervalos de 2 minutos entre rodadas de monitoração. A monitoração foi realizada por período de um mês (05/2011 à 06/2011), produzindo dados brutos que somaram 11.959 *otorrents*, 94 rastreadores e 187.140 endereços IP.

## 4. Resultados

A base de 11.959 *torrents* precisou ser submetida a um processo de filtragem, para que exames indesejados fossem retirados e, assim, não influenciassem a análise. Inicialmente removeu-se todos os *torrents* cujos rastreadores não puderam ser contatados devido a inconsistências nas URLs informadas.. Nesse grupo enquadraram-se 4.181 *torrents*. Em um segundo momento, retirou-se aqueles *torrents* que tiveram menos de 8 horas de vida na comunidade, levando à glosagem de mais 4.791 *torrents*. Exames com dados inconsistentes ou removidos tão precocemente da comunidade representam, muito provavelmente, conteúdos falsos. A não remoção desses exames pode exercer forte influência na análise dos resultados. Apesar da importância do processo de filtragem, até onde sabemos em nenhum dos trabalhos relacionados houve tal preocupação.

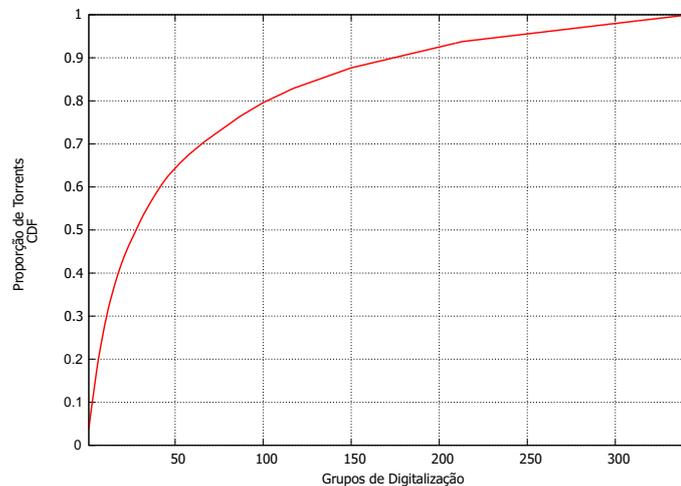
Após as duas filtrações, 2.987 *torrents* remanesceram e foram analisados. Os resultados são apresentados a seguir. Inicialmente caracteriza-se produtores e publicadores de conteúdos ilícitos, investigando seus graus de atividade e possíveis relações entre esses agentes. Na sequência, relata-se os processos de digitalização mais empregados e detalha-se a dinâmica, ao longo do tempo, de lançamento de *torrents* (dado um conjunto conhecido de conteúdos)  $x$  processos utilizados. Por fim, analisa-se os primeiros semeadores, procurando relações entre eles, os criadores de cópias digitais e os publicadores de *torrents*.

### 4.1. Produtores e Publicadores de Conteúdo Ilícito

Conforme apresentado na subseção 2.1, grupos de digitalização são os principais responsáveis pela criação de cópias ilícitas de filmes sendo compartilhadas nas redes BT. Neste artigo eles são referidos como *produtores*. Dos 2.987 *torrents* analisados, 2.066 (69,16%) identificam o produtor responsável por cada cópia. Esses 2.066 *torrents* foram criados por 342 produtores distintos. A tabela 2 enumera, em ordem decrescente, os 10 principais produtores. Ao lado, na figura 2, ilustra-se CDF (*Cumulative Density*

*Function*) representando no eixo horizontal os produtores, ordenados por quantidade de conteúdo criado, e no eixo vertical, a proporção acumulada de *torrents* criados. Como é possível observar, poucos produtores são responsáveis por grande parcela do conteúdo criado; praticamente 80% das cópias foram criadas por 100 produtores (29,23% dos 342).

Grupo	Torrents	
	#	%
Waf	78	4,02
Mr_Keff	69	3,56
Tnt Village	62	3,20
DutchTeamRls	61	3,14
Dmt	61	3,14
Imagine	59	3,04
Lkrg	51	2,63
Miguel	46	2,37
Martin	46	2,37
Nlt	44	2,27



**Figura 2. Contribuição cumulativa dos produtores**

Com o arquivo digital criado, o próximo passo para disseminá-lo é a publicação de *torrent* na comunidade. Os responsáveis por essa etapa são os publicadores, que, no escopo deste artigo, correspondem a usuários cadastrados no PirateBay realizando *upload* de *torrents*. Os 2.987 *torrents* foram publicados por um total de 517 usuários distintos. A tabela 3 apresenta os usuários mais ativos junto com a quantidade e proporção de *torrents* publicados. Essa tabela apresenta, além do nome do usuário, a sua categoria, que representa um “termômetro” da sua reputação na comunidade. Existem quatro categorias de usuários: VIP, confiável, ajudante e regular (estado inicial de qualquer usuário). A figura 3 apresenta uma CDF com os usuários em ordem decrescente de *torrents* publicados no eixo horizontal e a proporção de *torrents* no vertical. Ao observar esse gráfico, pode-se, novamente, perceber como poucos usuários são responsáveis pela maioria do conteúdo publicado. Em números, tem-se que 100 usuários (19,34% dos 517) publicaram quase 75% do conteúdo. Além disso, destaca-se que 25,5% dos 517 usuários eram de tipos especiais (não regulares) e foram eles os responsáveis pela publicação de 59,9% dos *torrents*.

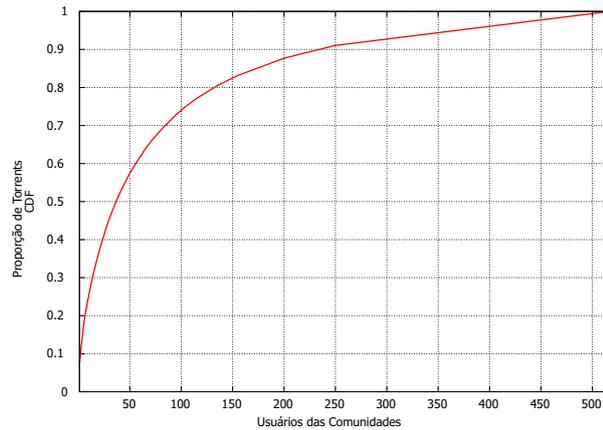
Após análise isolada da atividade de produtores e consumidores, procurou-se evidências quanto à existência de relação na ação de ambos agentes. Dois casos típicos foram observados: um publicador disponibilizando todos os materiais de um produtor e um grupo de publicadores trabalhando para um único produtor. Exemplificando o primeiro caso tem-se o usuário “sadbawang”, que publicou 77 dos 78 *torrents* do grupo “Waf”. Para ilustrar o segundo caso, tem-se que os publicadores “.BONE.” e “froggie100” foram responsáveis pela maioria dos conteúdos criados pelo grupo “Imagine”.

#### 4.2. Processos de Digitalização Empregados

Como já apontado na subseção 2.1, cópias digitais de um mesmo filme são diferenciadas pelas suas qualidades, que, por sua vez, são resultantes do processo de digitalização uti-

**Tabela 3. Ranqueamento**

Usuário	Tipo	Torrents	
		#	%
.BONE.	VIP	211	7,06
HDVideos	Regular	91	3,04
sadbawang	VIP	77	2,57
Sir_TankaLot	Confiável	73	2,44
MeMar	VIP	67	2,24
l.diliberto	VIP	57	1,90
SaM	VIP	47	1,57
martin_edguy	Confiável	46	1,54
miguel1983	VIP	46	1,54
virana	Confiável	41	1,37

**Figura 3. Contribuição cumulativa dos publicadores**

lizado. Dos 2.987 *torrents* analisados, 2.344 (78,47%) identificavam o processo utilizado para criação de cada mídia. A tabela 4 apresenta os processos, os seus graus de ocorrência e os três grupos de digitalização que mais criaram mídias empregando cada processo.

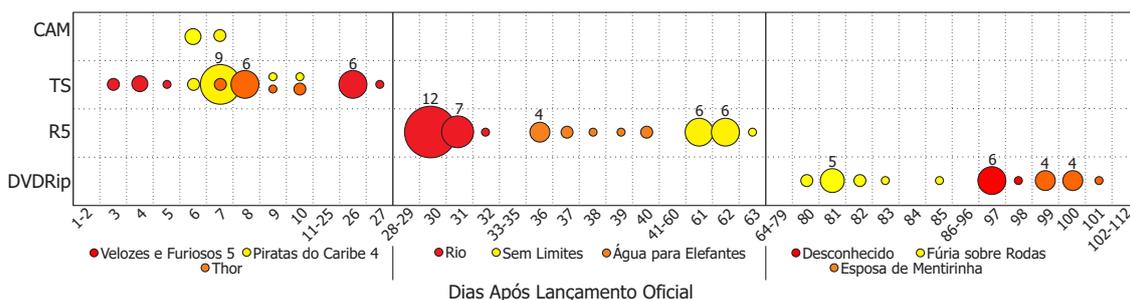
**Tabela 4. Principais processos**

Processo	Torrents		Principais Grupos
	#	%	
DVDRip	1825	77,85	Waf, Mr_Keff, Tnt Village
TS	144	6,14	Imagine, Dtrg, Cm8
R5	132	5,63	Imagine, Dmt, Vision
DVDScr	109	4,65	Ddr, Mtr, Xtreme
CAM	68	2,90	Lkrg, Imagine, Team Tnt
PPVRip	27	1,15	Iflix, Dmt, Flaw13ss
SCR	22	0,93	Kickass, Scr0n, 7speed
TC	16	0,68	Mtr, Team Tc, Rko

Analisando os resultados, observa-se que “DVDRip” representa o processo de digitalização mais comum, sendo o empregado por 77,85% dos *torrents* analisados que identificaram o processo. Tal predominância deve-se a duas razões principais. Primeiro, o conjunto de filmes que esses *torrents* podem estar representando é muito maior. Qualquer filme com 16 semanas transcorridas do seu lançamento oficial pode ser encontrado nesse formato e o resultado desse processo são mídias com a qualidade máxima, resultando no desinteresse pelas criadas por outros processos. Segundo, digitalizar um filme por meio desse processo é trivial se comparado com os outros; qualquer usuário que possuir um DVD original pode fazê-lo em seu computador pessoal. Outro grupo de processos que se destaca é o formado por “CAM”, “TS”, “DVDScr” e “R5”. O alto grau de ocorrência, em comparação aos outros processos que não “DVDRip”, deve-se a uma questão de custo/benefício entre: a dificuldade de obter-se a fonte para digitalização, o tempo necessário após o lançamento oficial do filme e a qualidade final da mídia. A título de exemplo tem-se que, apesar da qualidade resultante do processo “TC” ser a melhor entre a estréia do filme e 4-8 semanas transcorridas, “CAM” e “TS”, por utilizarem

fonte facilmente acessíveis, são processos mais empregados (0,68% x 9,04%). Vale observar, também, como produtores “menos ativos” destacam-se pelas suas especializações em processos que necessitam de fontes de difícil acesso. O grupo “Imagine”, por exemplo, apesar de ser somente o sexto colocado da tabela 2, apresenta-se como o principal produtor de “TS” e “R5”. Logo, as atividades desse grupo tornam-se tão importantes quanto, se não mais, as dos primeiros colocados da tabela 2.

Passando-se, agora, a caracterizar a dinâmica de processos de digitalização e de *torrents* disponibilizados em portais BT em paralelo ao ciclo de vida de filmes lançados no cinema, a figura 4 sintetiza resultado de observação realizada. Antes de apresentar discussão, contudo, é necessário informar que o período mínimo de monitoração para ser possível capturar todas as digitalizações realizadas sob um mesmo filme é de 16 semanas. Como não dispôs-se desta janela de tempo, trabalhou-se com 9 filmes, todos presentes no *dataset* coletado ao longo de 30 dias, cujo lançamento tivesse ocorrido há 0-4, 5-8 e há mais de 8 semanas (de acordo com o informado na IMDb [IMDB 2011]). No gráfico, o eixo horizontal representa os dias transcorridos e o eixo vertical, os processos de digitalização. Para apresentar uma “fotografia” mais fidedigna, todos *torrents* que não tiveram um tempo mínimo de vida de uma semana na comunidade e que não haviam sido publicados por usuários renomados foram desconsiderados.



fenômeno é observado quando uma fonte de melhor qualidade é encontrada para realizar o processo, acarretando em melhor qualidade final da mídia gerada.

### 4.3. Provedores de Conteúdo Ilícito

Ainda como parte da caracterização de disseminação ilegal de filmes em redes BT, interessou-se em determinar os pares (usuários) que fomentam enxames, na condição de semeadores, em seus instantes iniciais. Para identificá-los, foi necessário que a arquitetura de monitoração estivesse devidamente configurada para, assim que *torrents* fossem publicados no portal, pudessem ter seus respectivos rastreadores contatados e listas de pares participantes do início do enxame, obtidos. Quanto menor o intervalo decorrido entre a publicação de um *torrent* e o início da monitoração do enxame, maior a probabilidade de encontrar no enxame apenas o(s) primeiro(s) semeador(es). No contexto da investigação conduzida (lançando mão da arquitetura TorrentU e, em última análise, do procedimento ilustrado pelo algoritmo 1), esse intervalo girou em torno de 4 minutos.

Por meio da metodologia mencionada, identificou-se os primeiros semeadores de 692 (23,16%) dos 2.987 *torrents* analisados. Todos aqueles em que não foi possível identificar o(s) primeiro(s) semeador(es) eram enxames que: estavam vazios, existiam previamente à publicação do seu *torrent* no PirateBay ou cujo(s) rastreador(es) não foi possível contatar por erro na tentativa de comunicação. Passando à análise dos resultados, os 692 *torrents* foram semeados por um total de 775 pares; para alguns enxames observou-se, já no seu início, mais do que um semeador. Os 775 semeadores identificados estão associados a 318 endereços IP únicos. A figura 5 ilustra o grau de participação de cada IP.

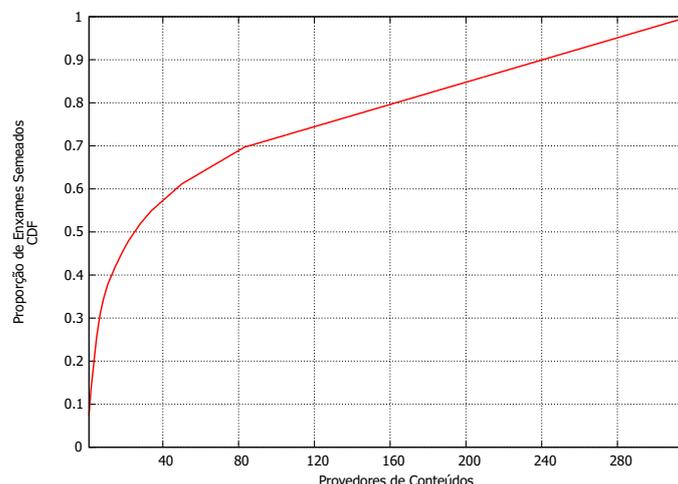


Figura 5. Contribuição cumulativa dos *provedores*

Dois aspectos destacam-se a partir da análise da figura 5. Primeiro, 25 endereços IP (7,86% dos 318) participaram como semeadores de cerca da metade dos enxames. Tal é um indicador de que usuários especializados podem estar utilizando *seedboxes* [Wikipedia 2011c] para disseminar seus conteúdos. Segundo, todos os semeadores a partir do 82º serviram exclusivamente a um enxame, caracterizando a participação de usuários “domésticos” no fomento de parcela significativa de enxames BT.

A tabela 5 apresenta a procedência dos principais semeadores, destacando país de origem, *Internet Service Provider* (ISP) de cada IP e quantidade de enxames semeados. Em contraste, apresenta-se na tabela 6 as principais localizações dos semeadores, ignorando a quantidade de enxames que cada um serviu. Como pode-se observar, a França destaca-se por 9,03% dos semeadores estarem localizados nesse país, curiosamente sendo todos servidos pelo ISP “Ovh”.

País	ISP	# Enxames
NZ	Obtrix	57
FR	Ovh	45
FR	Ovh	32
PL	Mokadi	32
GB	Ovh	31
FR	Ovh	24
NZ	Obtrix	21
FI	Lsinki	15
FR	Ovh	14
NL	Upc	12

**Tabela 5. Principais semeadores**

País	# IPs
IN	41
US	33
SE	33
FR	28
NL	26
JP	24
GB	14
PK	11
DE	10
AU	7

**Tabela 6. Distribuição semeadores**

Os provedores de conteúdo são os terceiros e últimos agentes responsáveis pelo processo de disseminação de cópias ilícitas de filmes através de BT. Ao observá-los, constatou-se que existem relações de dependência, ou subordinação, entre provedores (primeiros semeadores), produtores (grupos de digitalização) e publicadores (usuários da comunidade). Três casos típicos foram encontrados e são discutidos a seguir. O primeiro caso são provedores e publicadores dependentes dos produtores. Um exemplo são os grupos “Dmt”, “Mr\_keff” e “Miguel”, que tiveram cerca de 90% dos seus *torrents* publicados e semeados pelo mesmo usuário. O segundo caso consiste na observação de que provedores podem estar subordinados a produtores. Exemplos desse caso são os grupos “Kickass”, “Ddr” e “Extratorrentrg” que, apesar de terem seus *torrents* publicados por um grupo heterogêneo de usuários, sempre são semeados pelo mesmo provedor. O terceiro e último caso está relacionado com a possibilidade de provedores serem dependentes de publicadores. Como exemplo, tem-se os publicadores “Theroach”, “Riff” e “Safcuk009”, que disponibilizaram *torrents* de mídias criadas por grupos variados, porém sempre semeados pelos mesmos provedores.

## 5. Conclusões e Trabalhos Futuros

O compartilhamento de conteúdo por meio do protocolo BitTorrent é um dos principais responsáveis pelo atual volume de tráfego da Internet. Sabe-se que a maior parte desse volume é constituída pelo compartilhamento de conteúdos ilícitos. Sabe-se, também, que filme é o principal tipo de mídia sendo compartilhado ilegalmente. Apesar da reconhecida importância do tema, nenhum estudo procurou observar e mapear a dinâmica de disseminação dessa natureza de conteúdo. Para suprir essa lacuna, realizou-se a extensão de uma arquitetura de monitoração, que foi instanciada para observar o “universo” BT por 30 dias. A grande massa de dados obtida foi, então, organizada e cuidadosamente

analisada. Até onde sabemos, este é o primeiro estudo científico que busca caracterizar disseminação ilegal de filmes em redes BitTorrent.

A partir dos dados obtidos foi possível identificar padrões de comportamento de disseminadores de filmes ilegais. No que remete aos produtores, descobriu-se que a maioria dos *torrents* possui identificação do grupo de digitalização responsável e que, na realidade, são poucos produtores criando a maioria dos conteúdos. Quanto aos publicadores, observou-se um comportamento similar ao dos produtores, no sentido de que poucos são responsáveis pela publicação de grande parte dos *torrents* de cópias ilícitas. Além disso, uma relação de subordinação foi observada entre produtores e publicadores. Ao analisar os processos de digitalização empregados, descobriu-se que certos produtores são especializados em certos processos e confirmou-se a evolução, ao longo do tempo, dos processos por meio dos quais os filmes ofertados são digitalizados. Por fim, abordou-se os responsáveis por inicialmente semearem os enxames desses conteúdos, identificando as relações de dependência existentes entre produtores, publicadores e semeadores.

Finalizada uma primeira iteração para caracterizar disseminação ilegal de filmes em redes BT, identifica-se um conjunto de oportunidades de investigações futuras. A primeira consiste em realizar monitoração mais longa, desejavelmente com um mínimo de 16 semanas, que permita acompanhar o “ciclo de vida” completo de filmes em redes BT, desde seu lançamento no cinema até o momento em que passa a ser distribuído em DVD. A segunda oportunidade de trabalho futuro consiste em observar outras comunidades públicas populares. A terceira, monitorar as *darknets*, comunidades privadas de *torrents*, que, provavelmente, representam os locais onde, em tese, enxames em torno de cópias ilegais de filmes aparecem primeiro. Por fim, uma quarta oportunidade é a observação de padrões e dinâmicas de consumo desses conteúdos. Por exemplo, é interessante procurar observar se há concentração de consumidores de filmes ilegais em determinadas regiões e se há comportamentos claros de migração de consumidores entre enxames.

## Referências

- Bauer, K., McCoy, D., Grunwald, D., and Sicker, D. (2009). Bitstalker: Accurately and efficiently monitoring bittorrent traffic. In *WIFS 2009, IEEE Workshop on Information Forensics and Security*, pages 181–185.
- Chow, K., Cheng, K., Man, L., Lai, P., Hui, L., Chong, C., Pun, K., Tsang, W., Chan, H., and Yiu, S. (2007). Btm - an automated rule-based bt monitoring system for piracy detection. In *ICIMP 2007, International Conference on Internet Monitoring and Protection*, pages 1–6.
- Cuevas, R., Kryczka, M., Cuevas, A., Kaune, S., Guerrero, C., and Rejaie, R. (2010). Is content publishing in bittorrent altruistic or profit-driven? In *Co-NEXT 2010, International Conference on Emerging Networking Experiments and Technologies*, pages 1–12.
- Envisional (2011). An estimate of infringing use of the internet. [http://documents.envisional.com/docs/Envisional-Internet\\_Usage-Jan2011.pdf](http://documents.envisional.com/docs/Envisional-Internet_Usage-Jan2011.pdf) [Acessado em 07/11].
- IMDB (2011). <http://www.imdb.com/> [Acessado em 07/11].

- Junemann, K., Andelfinger, P., Dinger, J., and Hartenstein, H. (2010). Bitmon: A tool for automated monitoring of the bittorrent dht. In *P2P 2010, IEEE International Conference on Peer-to-Peer Computing*, pages 1–2.
- Le Blond, S., Legout, A., Lefessant, F., Dabbous, W., and Kaafar, M. A. (2010). Spying the world from your laptop: identifying and profiling content providers and big downloaders in bittorrent. In *LEET 2010, USENIX Workshop on Large-Scale Exploits and Emergent Threats*, pages 4–4.
- Mansilha, R. B., Bays, L. R., Lehmann, M. B., Mezzomo, A., Gaspar, L. P., and Barcellos, M. P. (2011). Observing the bittorrent universe through telescopes. In *IM 2011, IFIP/IEEE International Symposium on Integrated Network Management*, pages 1–8.
- Piratebay (2011). <http://thepiratebay.org/> [Acessado em 07/11].
- PlanetLab (2011). <http://www.planet-lab.org/> [Acessado em 07/11].
- Schulze, H. and Mochalski, K. (2009). Internet study 2008-2009. <http://www.ipoque.com/sites/default/files/mediafiles/documents/internet-study-2008-2009.pdf> [Acessado em 07/11].
- Wikipedia (2011a). List of warez groups. [http://en.wikipedia.org/wiki/List\\_of\\_warez\\_groups](http://en.wikipedia.org/wiki/List_of_warez_groups) [Acessado em 07/11].
- Wikipedia (2011b). Pirated movie release types. [http://en.wikipedia.org/wiki/Pirated\\_movie\\_release\\_types](http://en.wikipedia.org/wiki/Pirated_movie_release_types) [Acessado em 07/11].
- Wikipedia (2011c). Seedbox. <http://en.wikipedia.org/wiki/Seedbox> [Acessado em 07/11].
- Zhang, C., Dhungel, P., Wu, D., Liu, Z., and Ross, K. (2010a). Bittorrent darknets. In *INFOCOM 2010, IEEE International Conference on Computer Communications*, pages 1460–1468.
- Zhang, C., Dhungel, P., Wu, D., and Ross, K. (2010b). Unraveling the bittorrent ecosystem. *IEEE Transactions on Parallel and Distributed Systems*, 22(7):1164–1177.