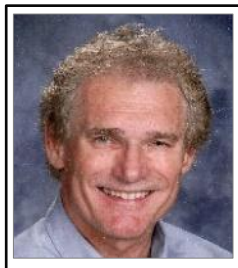
Software and Methods Supporting Legal Proceedings

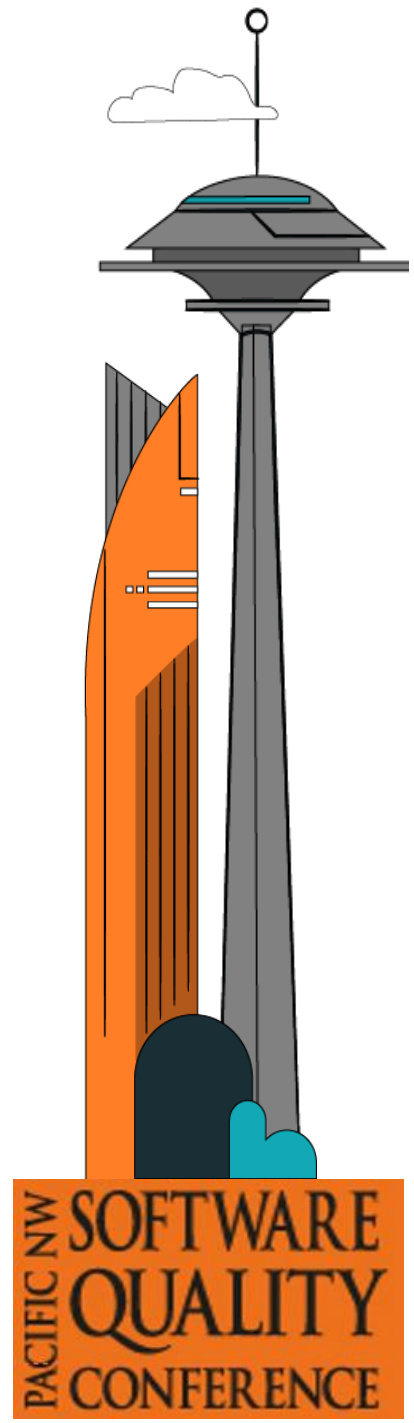
Home-Brewed and Speculative or Verifiably Reliable and Definitive?

Pacific Northwest Software Quality Conference
Portland, Oregon, October 14-16, 2024



Kal Toth

Kalman C. Toth, **Ph.D, P.Eng.**
kalmanctoth@gmail.com
https://www.sovereignimage.com
https://www.linkedin.com/in/kaltoth/



Presentation Slides

PNSQC 2024 “at-a-glance” and “technical papers”
https://www.pnsqc.org/kal_toth_2024_2.php
http://www.sovereignimage.com/PNSQC2024/

About

Preamble

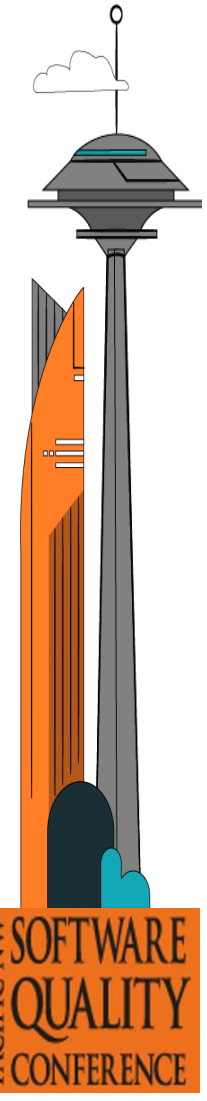
- ❑ Media companies have launched tens of thousands of copyright infringement cases
- ❑ Plaintiffs claim copyrighted movies have been illegally shared using BitTorrent
- ❑ I've collected facts and data while supporting at least 12 such cases

Foundational Issues

- ❑ Monitoring BitTorrent tools are software-based
- ❑ Software and methods controlling data collection reliable enough?
- ❑ Proprietary, "forensic", "reliable", "accurate", but not commercially available.

Paper Addresses

- ❑ How BitTorrent technology works
- ❑ How BitTorrent monitoring tools work
- ❑ Critical deficiencies in monitoring software and methods
- ❑ Changes mitigating certain of the critical deficiencies
- ❑ Limitations not readily resolved.



Introduction

BitTorrent Technology

- Used for large-scale legal sharing of software, movies, games, etc.
- Also used for illegal content sharing - copyrighted movies.

Defective BitTorrent Monitoring Tools Harm Wrongfully Accused

- Litigation costs; much stress; lost privacy and jobs; damaged relationships.

Common Software-based BitTorrent Monitoring Tools

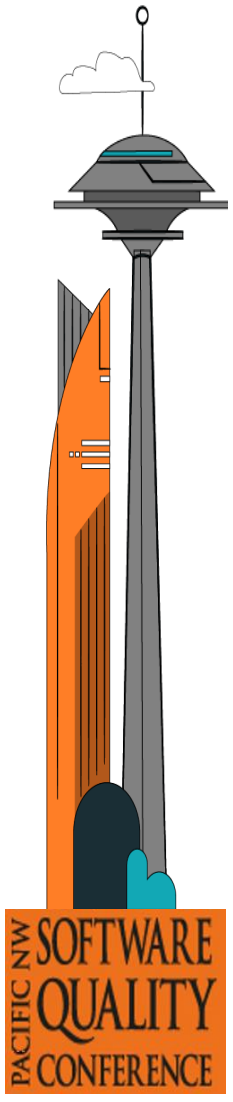
- NARS, MaverickMonitor, and VXN Scan

Copyright Infringement in BitTorrent (BT) Context

- Obtaining complete or substantial playable copies of copyrighted movies
- Distributing complete or substantial playable copies of copyrighted movies.

Admissibility of Technical Evidence – "Daubert" Principles

- Sufficiency of facts and data; data collection empirically tested and peer reviewed; published (e.g. conferences); accepted by scientific community.



BitTorrent Building Blocks

BT Client SW – SW on owners' devices for fragmenting and sharing movies in pieces

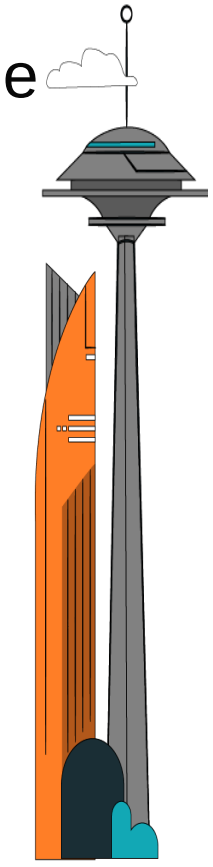
- ❑ Implement P2P protocol for exchanging movies in "pieces" between IP addresses
- ❑ Popular Clients: BitComet, BitTorrent, Deluge, qTorrent, rTorrent, Vuze, uTorrent (69%).

Swarm – represents IP addresses of BitTorrent clients sharing pieces of a given movie

- ❑ **Seeders**: hold all pieces of a movie and upload pieces to participating peers
- ❑ **Peers**: initially hold no pieces; download pieces from seeders and peers in swarm.

.Torrent File – initial seeder creates a .torrent file for sharing pieces of the movie

- ❑ Path/file holding the movie, file size, piece size, # of pieces, piece-hashes
- ❑ Hashing .torrent file metadata yields infohash identifying swarm sharing movie
- ❑ Initial seeder posts .torrent files on torrent sites for peers to locate swarm.



BitTorrent Tracker(s) – .torrent files "announce" URLs of trackers

- ❑ Trackers maintain IP address directories for each swarm identified by infohash
- ❑ Seeders and peers provide their IPs to trackers for updating swarm directories.

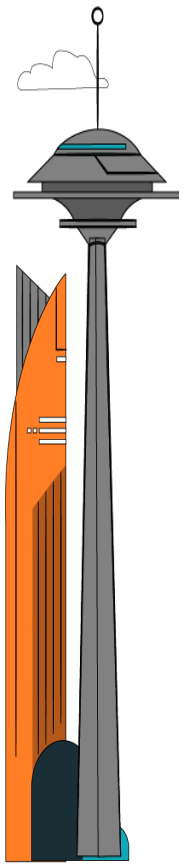
BitTorrent Sharing, Completeness, Playability

Sharing Constraints

- ❑ BT clients do not authenticate users – do not identify who they are
- ❑ Seeders/peers can join/quit at any time – pieces not always available.
- ❑ Clients can throttle the uploading of pieces into BitTorrent
- ❑ Peers who download pieces without uploading are called “leechers”
- ❑ Clients can “choke” leeching peers who refuse to upload pieces to them

BitTorrent Protocol - handshake, have, interested, piece data, … (wireshark)

- ❑ Peers receive pieces from seeders/peers randomly over the movie timeline
- ❑ Timeline gaps gradually filled in until all pieces received/verified or peer quits

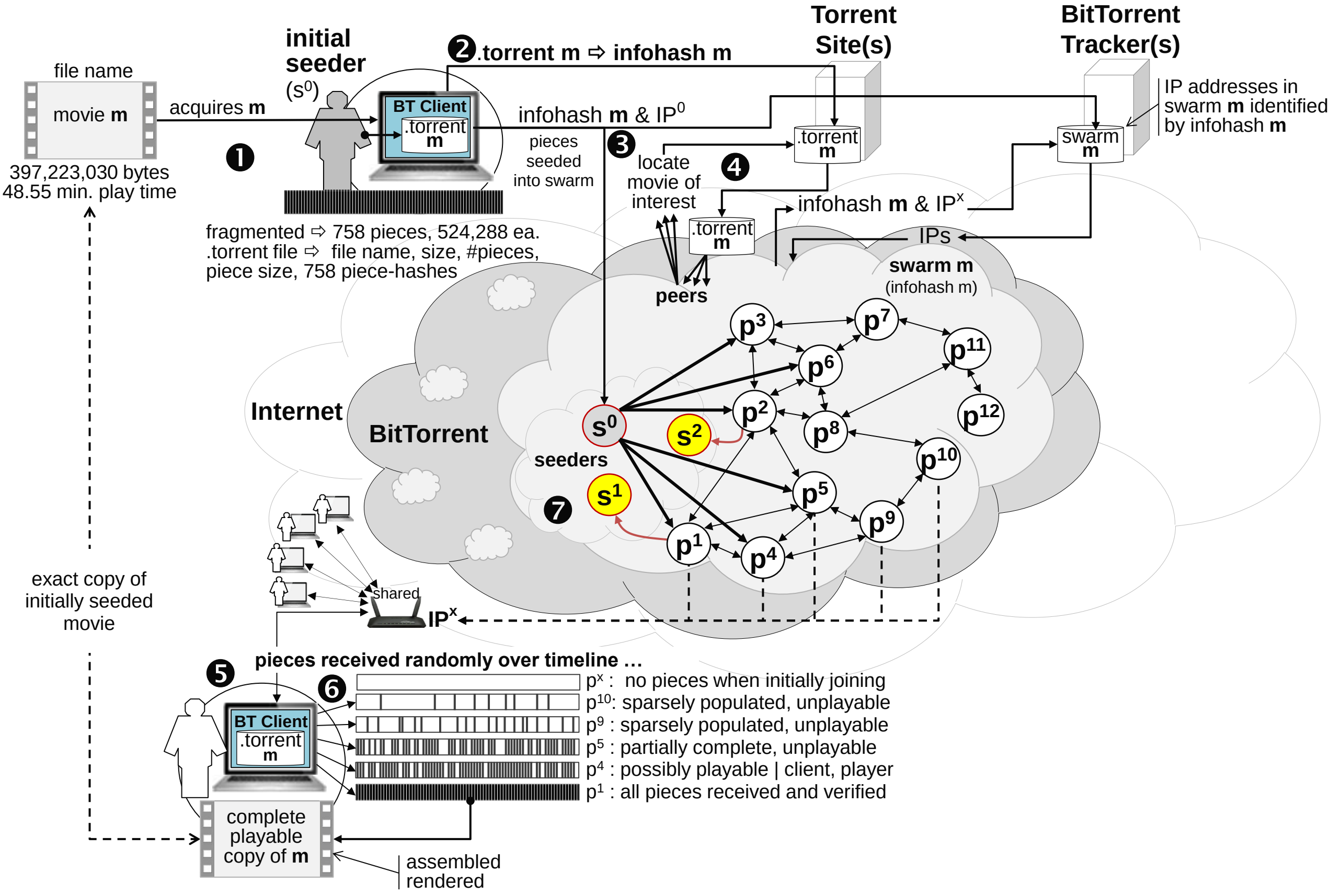


Completeness and Playability

- ❑ Obtained pieces are assembled and rendered into an exact playable copy
- ❑ Peers obtaining all pieces from a swarm can become seeders or leave.

1. BitTorrent does not guarantee peers will obtain complete playable copies of movies.
2. No guarantee .torrent metadata and file names characterize the asserted content.
3. Virtual Private Networks (VPNs) can be used to hide IP addresses used.

How BitTorrent Shares a Movie



BitTorrent Monitoring and Infringement Reporting

How BitTorrent Swarms are Monitored

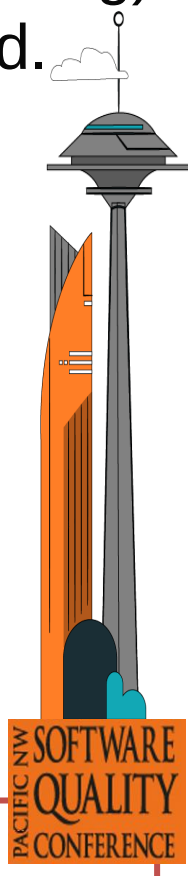
- ❑ **Targeting:** Swarms suspected of sharing movies examined to decide targeting
- ❑ **Monitoring:** Pieces downloaded from targeted swarms without uploading (leeching)
- ❑ **Protocol Capture (PCAPs):** record date/time, IP addresses, pieces exchanged.

How Swarms and IP Addresses are Targeted

1) .torrent files searched to identify swarms suspected of infringing a movie
2) Pieces collected from IP addresses in swarm until a full copy obtained
3) If full copy matches copyrighted movie, swarm and IPs in swarm targeted.

Critical Issues

- ❑ **Targeting:** PCAPs are not provided (not recorded or not available)
- ❑ **Monitoring:** PCAPs reveal only 2-3 pieces obtained from an IP address.



Critical Issues

a. Swarm Targeting: full copy from multiple IPs in a swarm compared/matched to copyrighted movies.

PCAPs are not recorded when collecting full copies and matching to copyrighted movies.

b. Monitoring: IP addresses are reported as infringing despite collecting only 2-3 pieces from the IP.

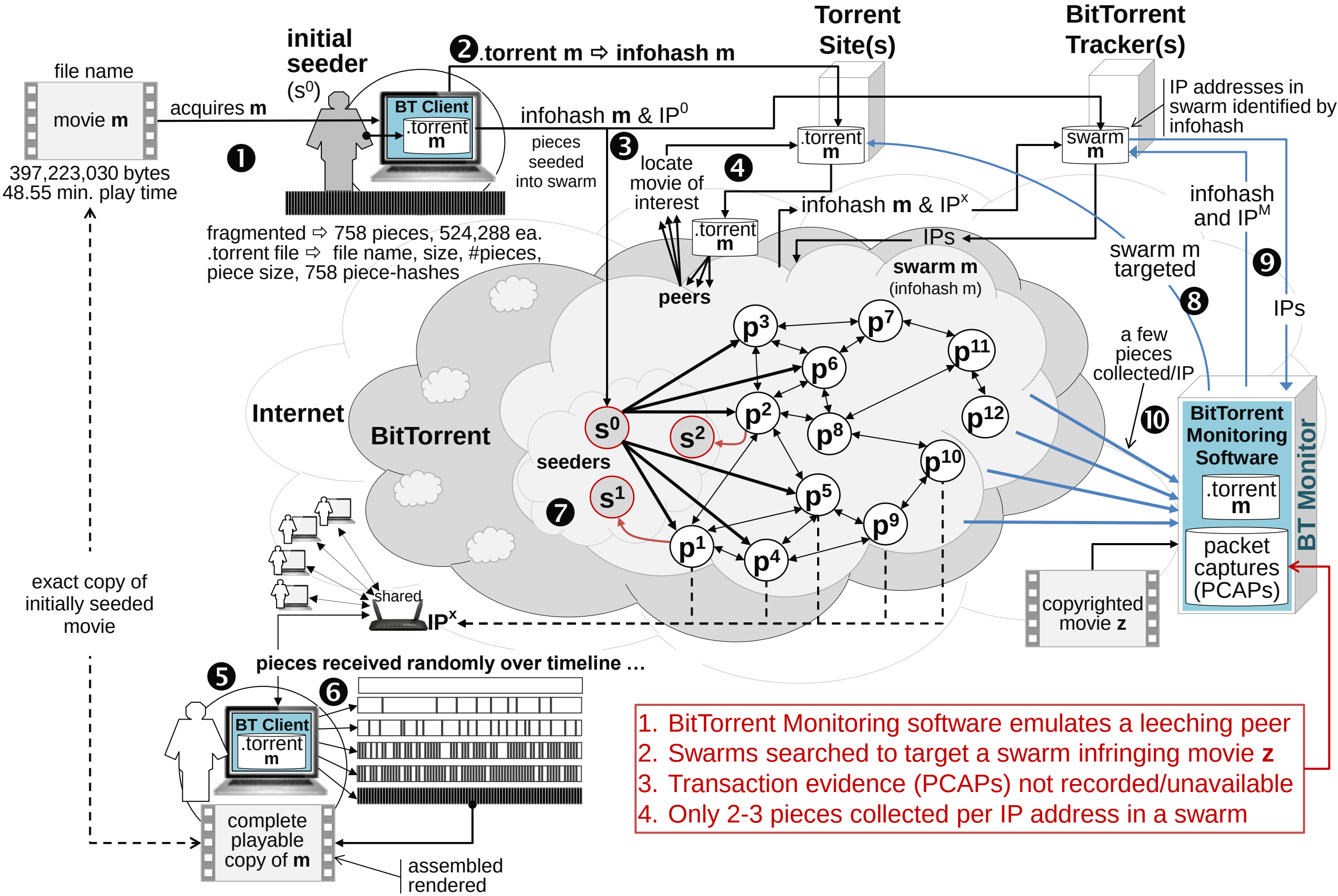
0.3% of a movie is not playable* this does not represent substantial copying / infringement.

c. Leeching: BT monitoring tools download pieces from peers in swarms but do not upload pieces.

Mimicking leeching peers leaves monitoring tools vulnerable to “choking”.

* represents about 9 seconds of play time

BitTorrent Monitoring Scenario



1. BitTorrent Monitoring software emulates a leeching peer
2. Swarms searched to target a swarm infringing movie **z**
3. Transaction evidence (PCAPs) not recorded/unavailable
4. Only 2-3 pieces collected per IP address in a swarm

Speculative Identification and Infringement Reporting

movie **m**

initial seeder leaves swarm

initial seeder ❶

❹ upon completion peers can become seeders

seeders not reported

swarm m
(infohash m)

piece-availability:
✓# active peers
✓# active seeders

p^x

p^M

❷

IP^x

subscriber

private IPs

❸

BitTorrent Monitor

❺ swarm (m) targeted

.torrent m

PCAP data

IPs

recorded

❻ 2-3 pieces collected per IP

full copy (all pieces)

matched

copyrighted movie **z**

no PCAP files recorded validating origin of pieces comprising full copy hence no objective evidence swarm **m** used to infringe copyrighted movie **z**

Alternate Methodology
only report IPs yielding complete playable movies
rec. PCAPs to validate all pieces of movies obtained
match playable movies to copyrighted movies

❼ speculates targeted IPs will not experience failures
non-infringing IPs falsely reported as infringing

choking failures | monitoring failures | piece unavailability | abandoned downloads | client failures | space depletion

failure modes (unresolved)

Speculative Identification:

- users share IP addresses
- sharing not publicly visible
- IP addresses do not identify infringers (Cobbler [11])
- IP^x subscriber assumed to be responsible for infringement …
- Subpoena⇨ISP⇨ID subscriber

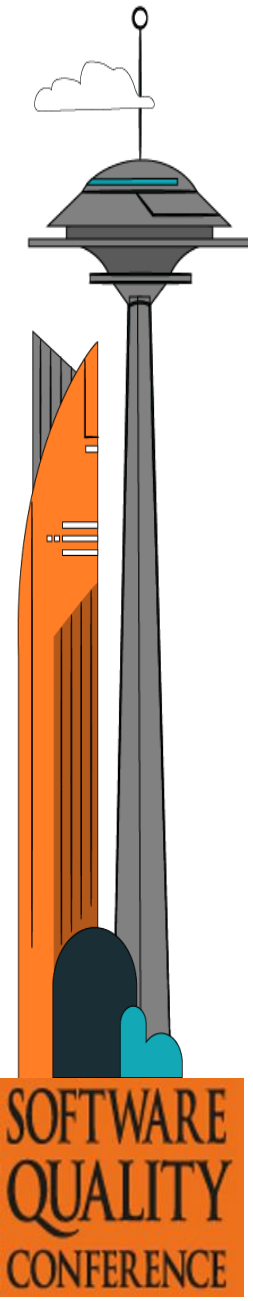
Speculative Reporting: IPs reported as infringing though 0.3% of the pieces collected (unplayable)

Closing Remarks

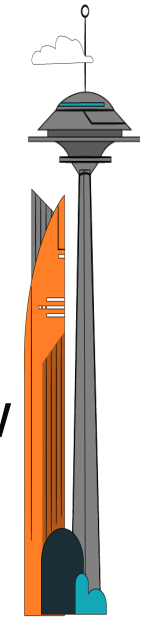
BT Monitoring Tools “AS IS”

BT Monitoring Tools “GOAL”

Unresolved Limitations



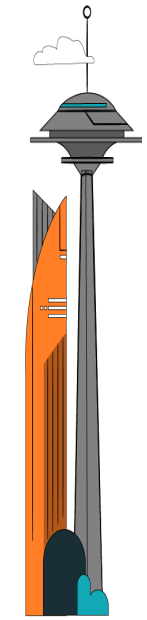
BT Monitoring AS IS: Speculative and Home-Brewed



Quality of Evidence

1. **Efficacy of Methodology**	**Speculative:** failure modes do **not** occur, collecting a few pieces is enough, and IP subscribers are responsible.
2. **Identifying Infringers Reliably & Accurately**	**Home-Brewed:** software process undefined, technical specifications do not exist, software not verifiably reliable.
3. **Adequacy of Testing & Quality Assurance**	**Testing Inadequate:** IP sharing not configured, not loaded, failure modes and feature-length movies not tested.
4. **Targeting Swarms and IP Addresses**	**No PCAP data:** no transaction data supporting assertion full copies obtained from swarms matched copyrighted movies.
5. **Critical Failure Mode Resolution**	**Design Flaws:** abandoned downloads, piece unavailability, space depletion, and choked monitoring not resolved.
6. **Infringement Evidence Sufficiency**	**PCAP Data:** confirms about 0.3% of the pieces of a movie are collected from IP addresses.
7. **Playability and Matching**	**Unplayable Content:** Such sparse content is not playable, hence no evidence identified IP addresses used to infringe.

BT Monitoring GOAL: Definitive and Verifiably Reliable



Quality of Evidence

1. **Efficacy of Methodology**	**Definitive:** collect all pieces from same IP in a swarm and match rendered (infringing) movie to copyrighted movie.
2. **Identifying Infringers Reliably & Accurately**	**SW Verifiably Reliable:** specify forensic purpose and tech specifications; produce peer-reviewable and testable software.
3. **Adequacy of Testing & Quality Assurance**	**Robust Testing:** test router/IP sharing, typical loads, failure modes, feature-length movies, repeatability and reproducibility.
4. **Targeting Swarms and IP Addresses**	**Validate Targeting:** provide PCAP evidence targeted swarms share all pieces of movie matching copyrighted movie.
5. **Critical Failure Mode Resolution**	**Collect All pieces from IP in swarm:** resolves abandonment, piece unavailability, space depletion, and choked monitoring.
6. **Sufficient Evidence IP Address Infringed**	**Validate Monitoring:** using PCAPs verify all pieces from each IP in targeted swarm matches piece-hashes in .torrent file.
7. **Playability and Matching**	**Infringement Verification:** verify all pieces obtained from an IP is playable and audio-visually matches a copyrighted movie.

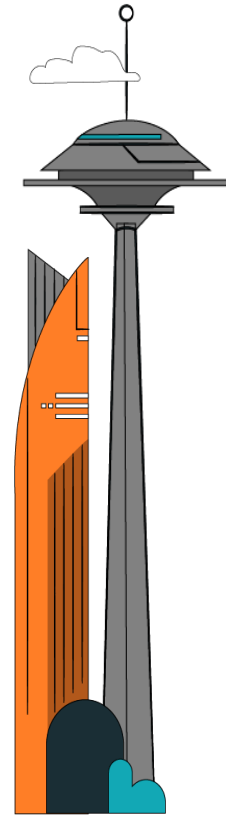
Unresolved Limitations

Shared IP Address Limitations

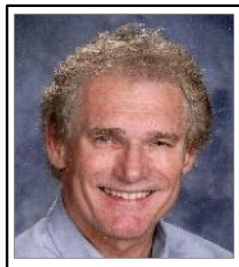
- ❑ shared routers hide private IP addresses from public IPs (e.g. monitoring tools)
- ❑ shared personal devices may also hide identities of persons using the device.

BitTorrent Choking Protocol

- ❑ monitoring tools mimic leeching peers typically choked by other peers
- ❑ tools may only be able to collect few pieces from a given peer’s IP address.



Thank you for your valuable time, questions and comments.



Kal Toth

Kalman C. Toth, Ph.D, P.Eng.
kalmanctoth@gmail.com
https://www.sovereignimage.com
https://www.linkedin.com/in/kaltoth/

Abstract, Bio, Presentation Slides

"conference at-a-glance" and "technical papers"

https://www.pnsqc.org/kal_toth_2024_2.php

http://www.sovereignimage.com/PNSQC2024/