Real-World Sybil Attacks in BitTorrent Mainline DHT

Liang Wang † Jussi Kangasharju †‡

†Department of Computer Science, University of Helsinki, Finland
‡Helsinki Institute for Information Technology, University of Helsinki, Finland
Contents

Background
Measurement Platform
Types of Attacks
Honeypot Design
Real-World Attacks
Threats & Solutions
Conclusion
BitTorrent MLDHT

BitTorrent is the most influential P2P software

BitTorrent adopted DHT to support trackless mode
  More reliable and robust to single point of failure
  Avoid legal issues and becoming the lawyer’s target

Two incompatible versions
  Mainline DHT (MLDHT)
  Vuze DHT
  Both are Kademlia-based DHT
BitTorrent MLDHT

Building block of modern P2P software

More and more P2P softwares support MLDHT, over 16,000,000 users

Forms a peculiar ecosystem

Different implementation has slightly different interpretation of the standard protocol

Only minimum functionalities of Kademlia

MLDHT is simple but weak
Measurement Platform

Two dedicated machines monitor the system
Prevent the “sample gap” due to soft- and hardware failure
Two crawlers with different crawling policy

Continuously take “snapshot” of MLDHT
15 mins interval, started since 2010-12-07
Over 67,000 samples are collected
Measurement Platform

A full-fledged monitoring system
Five main components
Automatically collect and analyze the sample
Together with several honeypots to catch specific attacks

MLDHT Monitor System
- Crawler
- Analyzer
- Maintainer
- Injector
- Visualizer
Two Basic Attacks
- Sybil Attack

First introduced by Douceur in “The Sybil Attack” in 2002

Inject multiple fake identities into the system

Use them as a starting point to perform further attacks

Can also be considered as routing table attack
Two Basic Attacks

Horizontal Attack
- Spread sybils widely across the system
- Pollute as many routing table as possible

Vertical Attack
- Isolate an node by filling routing table with malicious nodes
- Target can be content (more general than Eclipse Attack)

Hybrid Attack
- Combined both Horizontal and Vertical attacks
BitTorrent MLDHT
- Four Control Messages

PING
– Probe a node’s availability.

FIND_NODE
– Given a target ID, find the K closest neighbors of the ID.

GET_PEERS
– given an infohash, get the initial peer set.

ANNOUNCE_PEER
– a peer announces it belongs to a swarm.
Normal Operation

1. Announce Peer (29,59)
2. Get Peers (82,59)
3. Peer 65
4. Peer 95

BT Protocol
Two Basic Attacks
- Horizontal Attack

Peer 11
Peer 29
Peer 33
Peer 36
Peer 43
Peer 51
Peer 57
Peer 65
Peer 71
Peer 78
Peer 82
Peer 95

Find Node (82, 57)
Find Node (29, 78)
Find Node (43, 65)

I am 57
I am 65
I am 78

Peer 71
Peer 65
Peer 57
Peer 51
Peer 43
Peer 36
Peer 33
Peer 29
Peer 11
Two Basic Attacks
- Horizontal Attack

Horizontal attack spreads sybils widely across the system.

The aim is to pollute as many routing tables as possible.

The number of sybils in one routing table is not the concern.

A successful horizontal attack can let the attacker sniff most of the control messages and therefore hijack the system.

In MLDHT, this attack can be carried out with very limited physical resources.
Two Basic Attacks
- Vertical Attack

1. Announce Peer (29,59)
2. Get Peers (82,59)
3. Peer 78
4. Sybil

Peer 71
Peer 78
Peer 95
Peer 82
Peer 57
Peer 51
Peer 65
Peer 36
Peer 33
Peer 43
Peer 29
Peer 11
Peer 57
Peer 51
Two Basic Attacks
- Vertical Attack

Vertical attack attempts to insert as many sybils as possible in one specific routing table.

Similar to eclipse attack, but broader in the sense it can target content.

After launching an attack, the attacker waits for an interesting target ID, either node ID or content infohash. Then, the attacker inserts sybils close to the target to "isolate" it from the others.
Honeypot Design
- Three Types of Honeypots

Detector
- First round filtering, identifying suspicious nodes
- Work on the MLDHT protocol level

MLDHT honeypot
- Second round filtering, identify vertical and hybrid attacks
- Work on the MLDHT protocol level

BitTorrent honeypot
- Work on the BitTorrent protocol level
- Identify further malicious behaviors
Honeypot Design
- Three Types of Honeypots

Detector

First round filtering, identifying suspicious nodes

Use FIND_NODE to find “non-existent” ID in MLDHT
– If some node claims as the owner of the ID ---> Suspicious!

Use GET_PEERS to find “non-existent” infohash in MLDHT
– If someone comes to us with those IDs ---> Suspicious!

Try to catch horizontal attacks
– Because in reality, horizontal attack is the starting point of a successful large-scale attack
Honeypot Design
- MLDHT Honeypot

MLDHT honeypot
Detect vertical and hybrid attacks

Suppose our own ID is $x$, use GET_PEERS to find $x+1$
- Not such content with infohash $x+1$ at all in MLDHT
- Normal behavior is replying with $k$ nodes with closest ID
- However, lots of nodes came and used “scrape” to probe us
- Some immediately started vertical attacks by inserting sybils
- Some tried to set up BT level connection to get the metainfo
Honeypot Design
- BitTorrent Honeypot

BitTorrent honeypot
Work on the BitTorrent protocol level

Third round filtering, check if there are further actions

Most tried to get the metainfo for the infohash

But nobody really tried to retrieve any data
Real-World Attacks

Two main players (or attacker?)

Mr. ISP and Mr. 50

Suspicious behavior, but not really malicious (depends on the standards)

Quite amount of smaller players from various orgs
Real-World Attacks
- Mr. 50

Two (or more) virtual nodes from Amazon EC2 platform
large-scale horizontal attacks

The one with IP starting with 50 was the most active one

The app was optimized to perform horizontal attack
Not maintain states for the nodes
Only answer certain messages in order to stay in RT
Collect infohashes, later other nodes will get metainfo
Like a virus, even the honest node can help propagate
Real-World Attacks
- Mr. 50

Very active since the beginning of 2011
We can see the clear “test phase” and “deploy phase”
Real-World Attacks
- Mr. 50

Very active since the beginning of 2011
We can see the clear “test phase” and “deploy phase”
Real-World Attacks
- Mr. ISP

Several ISPs perform intensive “vertical attacks”

Very resourceful

Further investigation shows the main purpose is to localize BitTorrent traffic
Real-World Attacks
- Mr. ISP

Several ISPs perform intensive “vertical attacks”

Very resourceful

Further investigation shows the main purpose is to localize BitTorrent traffic
Real-World Attacks
- Mr. ISP

insider - connect to MLDHT with an IP in the ISP network
outsider - connect to MLDHT with an IP outside ISP
native node - node in the same ISP network
Real-World Attacks
- Mr. ISP

insider - connect to MLDHT with an IP in the ISP network
outsider - connect to MLDHT with an IP outside ISP
native node - node in the same ISP network
Threats & Analysis

Monitor a large fraction of user activities and traffic

Can disturb operation of system

Pollute or even effectively censor certain content

Hijack system with limited resource is possible

Violate user privacy
The trend of sybil attack is increasing

- Mr.50: 56%
- Mr.ISP: 27%
- Other: 14%
- Mr.184: 3%

# of sybils

Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
Possible Solutions

Root of the problem - Letting a node choose its own ID

Weakness has been well-understood over a decade, but the problem is still there

Various solutions, but no silver bullet
  - central, trusted authority to issue ID
  - charge money on ID
  - introduce social network into the system
Conclusion

Introduce two basic attacks in MLDHT

Hybrid attack can hijack whole system with limited physical resources

Extensive measurements to identify real-world attacks

Analyze potential damage and discuss possible solutions
Thanks!

Questions?