International Journal of Information and Computation Technology.
ISSN 0974-2239 Volume 3, Number 8 (2013), pp. 841-844

http://www. irphouse.com /ijict.htm

Content Pollution in P2P System

Simmi Chawla

RDIAS, New Delhi, (Affiliated to GGSIPU, Delhi)

Abstract

One way to combat P2P file sharing of copyrighted content is to deposit into the file sharing systems large volumes of polluted files. Without taking sides in the file sharing debate, in this paper we undertake a measurement study of the nature and magnitude of pollution in the Fast Track P2P network, currently the most popular P2P file sharing system. We develop a crawling platform which crawls the majority of the Fast Track Network's 20,000+ super nodes in less than 60 minutes. From the raw data gathered by the crawler for popular audio content, we obtain statistics on the number of unique versions and copies available in a 24-hour period. We develop an automated procedure to detect whether a given version is polluted or not, and we show that the probabilities of false positives and negatives of the detection procedure are very small. We use the data from the crawler and our pollution detection algorithm to determine the fraction of versions and fraction of copies that are polluted for several recent and old songs. We observe that pollution is pervasive for recent popular songs. We also identify and describe a number of anti-pollution mechanisms.

Keywords: Network Measurements, P2P networks.

1. Introduction:

One sabotage technique that is particularly prevalent today is that of pollution. Here, a “Pollution Company “first tampers with copyrighted content with the intention of rendering the content unusable. It then deposits the tampered content in large volumes in theP2P network. Unable to distinguish polluted files from unpolluted files, unsuspecting users download the files into their own file-sharing folders, from which other users download the polluted files. In this manner, the polluted copies of a given

song spread through the file sharing system, and the number of polluted copies can eventually exceed the number of clean copies of a given song. The goal of the pollution company is to trick users into frequently downloading polluted copies; users may then become frustrated and abandon P2P file sharing.

In this paper I undertake a detailed measurement study of the nature and magnitude of pollution in KaZaA, currently the most popular P2P file sharing system. I emphasize that the purpose of this paper is not to take sides on the P2P file-sharing debate nor to condone nor to condemn pollution. The goal instead is to understand P2P pollution, how pervasive it is currently in P2P networks, how quickly it spreads, and to identify measures for countering P2P pollution attacks. We will see that pollution is indeed pervasive, with more than 50% of the copies of many popular recent songs being polluted in KaZaA today. Because P2P file sharing is having a major impact on Internet traffic and usage, it is important to gain deep insights into P2P pollution, which is now a central part of the P2P landscape.

2. Classification of P2P Pollution:

Content Pollution: This is currently the more common form of pollution. The polluting party targets a particular digital recording (e.g., song or video). It then manufactures decoys for the recording by modifying it in one or more ways, including replacing all or part of the content with white noise, cutting the duration, shuffling blocks of bytes within the digital recording, inserting warnings of the illegality of file sharing in the recording, and inserting advertisements. We observed that today a popular pollution technique is to insert tens of seconds of undecodable white noise into the middle of the song.

Metadata Pollution: The other strategy is to not tamper with the digital recordings themselves but instead tamper with metadata. This often involves taking an older recording, whose copyright has expired, and changing its song title, album title, and artist name to that of the targeted recently released recording. Thus, when a user requests the target recording, the user will mistakenly obtain a different recording.

3. Anti-pollution Mechanisms

Given that pollution in P2P file sharing systems is pervasive, it is natural to consider what can be done to defend against the pollution attack. In this section we describe a number of potential anti-pollution mechanisms. We classify the mechanisms into two categories:

Detection without downloading: After receiving search results, the mechanism attempts to determine whether the files in the results are polluted without actually downloading any portion of the files.

Detection with downloading: For this class, the mechanism detects whether a file is polluted by first downloading portions (or all) of the file. Clearly, from the perspectives of the user and of network traffic, the first class of mechanisms is

preferable, as resources are not wasted downloading high-bit-rate Polluted multimedia files (or portions thereof).

4. Conclusion

We developed the KaZaA Crawling Platform to obtain measurement data for this study. This crawler is of independent interest. Developing the crawler was challenging since KaZaA uses a proprietary protocol with most of its signaling messages being encrypted. Also, a farm of server nodes, each running a large number of threads, was necessary to crawl the 20,000+KaZaA super nodes in an acceptable amount of time. In future work we will further exploit the crawler to gain insight into IP and geographic information on the sources of content.

References

[1] F. Oberholzer, K. Strumpf, "P2P's Impact on Recorded Music Sales," Second Workshop on Economics of Peer-to-Peer Systems, Cambridge, Massachusetts, June 2004

[2] J. Kurose, K.W. Ross, "Computer Networking: A Top-Down Approach Featuring the Internet," Addison-Wesley, 2005.

[3] K.P. Gummadi, R.J. Dunn, S. Saroiu, S.D. Gribble, H.M. Levy and J. Zahorjan, "Measurement, Modeling, and Analysis of a Peer-to-Peer File-Sharing Workload," Proceedings of the 19th ACM Symposium on Operating Systems Principles (SOSP-19), October 2003.

[4] J. Liang, R. Kumar and K.W. Ross, "Understanding KaZaA,"submitted, 2004.