'Business Usage of P2P File Sharing' Demonstration

Authors:

James Walkerdine, Computing, InfoLab 21, Lancaster University, Lancaster, LA1 4WA, UK.
walkerdi@comp.lancs.ac.uk

Danny Hughes, Research and Development, Isis Forensics, PO Box 793, Lancaster, LA1 9ED, UK.
danny@comp.lancs.ac.uk

Kevin Lee, School of Computer Science, University of Manchester, Manchester, M13 9PL, UK.
klee@cs.man.ac.uk

Demonstration:

This demonstration will allow users to explore how two large-scale P2P file sharing systems (Gnutella and Bittorrent) are accessed through business networks. This will be done through three separate demonstration media; (i) a real-time visualization of business P2P use, (ii.) an interactive graphical exploration of the impact of viral media on existing trace data and (iii) on-demand statistics for delegates own institutions.

1. **Real-time visualization of business P2P use:** Using Isis Forensics monitoring software, a continuous and real-time visualization of the characteristics of business P2P use will be shown. This will include the overall proportion of peers on business networks and a breakdown of business use by both country and business size.

2. **Exploring the effect of Viral Media:** As described in the following abstract, we have observed that the participation of big business on P2P networks is closely related to the release of viral media. Using a GUI-based interactive tool, delegates will be able to explore the effect of viral media events from our past traces and observe how viral media spreads across the networks of big businesses, government and even military institutions.

3. **On Demand statistics for delegates own institution:** throughout the period in which the demonstration runs, delegates will be able to register to receive a detailed statistical break-down of their own institution's participation in P2P file sharing (anonymized data only).

Space Required:

The demonstration requires a table at least 1.5M * 0.75M, preferably with sufficient space behind it to place a projector screen (if one is available). All other equipment will be brought by the demonstrators. The demonstration will require an ADSL or better Internet connection.

Time Required:

The time period required to present this demonstration is flexible. The real-time visualization is automated and could potentially execute throughout the conference period. Similarly, on-demand statistics of delegate's institutions can be provided 'out of band'. Furthermore, the interactive exploration of the effect of viral media can either execute in an automated fashion throughout the conference, or delegates can be guided through their interaction with this and the other props, as part of a structured demonstration session.

The Effect of Viral Media on Business Usage of P2P

James Walkerdine
Computing, InfoLab 21,
Lancaster University,
Lancaster, LA1 4WA, UK.
+44 (0) 1524 510352
walkerdi@comp.lancs.ac.uk

Danny Hughes
Research and Development,
Isis Forensics, PO Box 793
Lancaster, LA1 9ED, UK.
+44 (0) 7092 891873
danny@comp.lancs.ac.uk

Kevin Lee
School of Computer Science,
University of Manchester,
Manchester, M13 9PL, UK.
+44 (0) 161 2756132
klee@cs.man.ac.uk

Abstract

P2P file-sharing poses a number of serious problems for business network administrators including: unpredictable network usage, increased vulnerability to security threats and the danger of legal action. Due to these problems, most businesses attempt to restrict the use of P2P. This paper and the associated demonstration explore the nature of business participation in P2P and particularly highlight the effect of viral media on the participation of large businesses in P2P networks.

1. Introduction

Today, peer-to-peer (P2P) file sharing systems generate the majority of Internet traffic [1]. Unfortunately, these systems cause a number of serious problems for business network administrators:

1. *Unpredictable Network Usage*: This is particularly problematic due to the large size of P2P files and their highly concentrated and dynamic popularity.

2. *Security Threats*: Malicious files are widespread on P2P file sharing systems.

3. *Danger of Legal Action*: P2P file sharing supports a variety of illegal activities. Furthermore, in many localities, access providers may be held responsible for the behaviour of their users.

Despite these problems, Identifying and blocking P2P traffic is difficult due to the large number of rapidly evolving P2P protocols [2]. This paper firstly considers the scale and characteristics of participation in P2P file-sharing by users on business networks. Secondly, we explore the significant effects of *viral media* on the penetration of networks belonging to large businesses.

2. Business Participation in P2P

Isis Forensics was founded to address the P2P-related security concerns of businesses by providing monitoring, training and consultancy services. As part of this, Isis routinely monitors a number of P2P file sharing networks. Specifically, this paper focuses upon a two-week case-study of the Gnutella [3] network, performed from the 30th of March to the 12th of April using a *passive application-level* tracing methodology [1] and a single super-node connection to the Gnutella network.

Using this trace, we firstly explore the scale of business participation by analyzing PONG *connection messages* logged during our trace. Each unique IP address was resolved and categorized as belonging to either a business or non-business network. We found that business networks are responsible for a small, but significant proportion of PONG connection traffic - on average, 3.7% of the whole.

We also measured the rate of free riding [4] throughout our trace by analyzing the number of PONG messages which report zero shared files. We found a significant level of free riding on both business and non-business networks. Our trace revealed an average free-riding level of over 96% for 'general' users and a level of over 99% on business networks. This higher rate of free-riding is most likely due to prohibitions against P2P file sharing in the work-place (observation of these peers accepting incoming connections shows that Gnutella is not blocked on their host network).

We also analyzed the effect of business size upon participation. Businesses were divided into two broad classes based upon their number of employees: 'small businesses', having fewer than 100 employees and 'large businesses'. We found that small businesses participate most heavily in Gnutella, while the participation of large businesses is consistently and significantly lower, as can be seen in figure 1 on the following page (except in the case of April 7th due to the effect of *viral media*, as explored in section 3). During our trace we observed an average of 5.3 small businesses per day compared to just 2.1 large businesses per day.

3. The Effect of Viral Media

One of the most notable effects observed in our trace was that of *viral media*. The term 'viral media' refers to electronic media which suddenly becomes very popular and is distributed between Internet users in an epidemic fashion. We find that, as one might expect, P2P file-sharing networks are a major vector for viral media. More interestingly, we find that the emergence of viral media has a dramatic effect upon the characteristics of businesses that participate in P2P file sharing.

As described in section 2, analysis of the P2P traffic originating from business networks shows that, in general, small businesses generate the majority of P2P traffic. However, the emergence of viral media has the capacity to dramatically alter this state of affairs. Figure 1 shows the number of unique business networks observed during each day of our trace. It can be seen that, in general, small businesses (i.e. those with fewer than 100 employees) participate more heavily in Gnutella; however, this trend reverses dramatically on April 7th, when the number of large business observed jumps to five times the observed average.
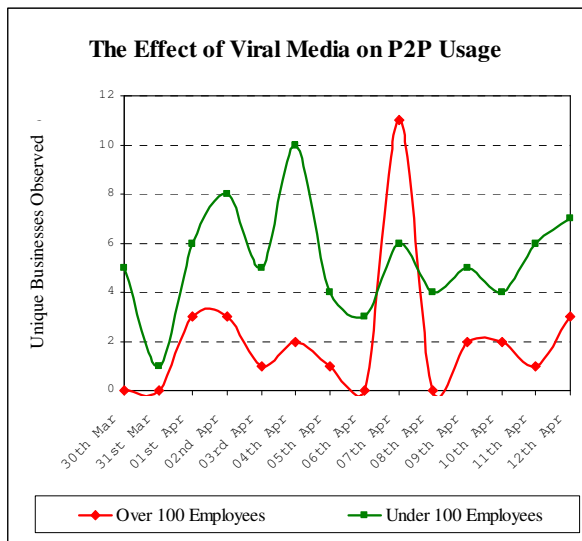

Figure 1 – The Effect of Viral Media on P2P Usage

We specifically analyzed QUERYHIT messages on April 6th, 7th and 8th. We found that, remarkably, *every* large business which appeared in our trace for the first time on April 7th was distributing search-responses for a single, *identical* file – a viral video (unfortunately its title is unsuitable for reproduction here). In keeping with the nature of viral media, the popularity of this file was short-lived and all of the business networks that we

observed for the first time distributing this file on April 7th did not re-appear during the remainder of our trace. The same was also true of the viral media file.

While this case-study is of relatively short duration, it demonstrates an oft-repeated pattern, wherein viral media emerges and causes a significant jump in the number of large businesses observed using P2P, followed by a rapid return to the norm as the viral media quickly loses popularity. These results support the assertion, made in section 2 that prohibitions against P2P in the workplace are partially responsible for the lower level of P2P participation through business networks (as these services are unlikely to be unblocked for short periods). It appears that the employees of large businesses choose not to participate until they are sufficiently temped by the release of viral media.

This effect, along with other trends in business P2P usage will be explored through the accompanying demonstration installation which will be available throughout the conference.

4. Summary

This abstract has briefly explored the participation of business networks in Gnutella using a two-week trace. Firstly, we show that business networks account for a small but significant proportion of P2P traffic. Secondly, we find that users participating through business networks demonstrate a consistently higher level of free riding. Thirdly, we find that small businesses are significantly more likely to participate in Gnutella than large businesses. Finally, we find that the participation of large businesses is highly dependent upon the release of viral media.

5. References

[1] "Monitoring Challenges and Approaches for P2P File Sharing Systems", Hughes D., Walkerdine J., Lee K., in the proceedings of the 1st International Conference on Internet Surveillance and Protection (ICISP'06), Cap Esterel, France, August 2006.

[2] "Accurate, Scalable Network-level Identification of P2P Traffic Using Application Signatures" Subhabrata S., Spatscheck O., Wang D., in the proceedings of the 13th International World Wide Web conference (WWW'04), New York, USA, May 2004.

[3] "The Gnutella Protocol Specification v0.6": available online at: http://rfcgnutella.sourceforge.net/src/rfc-0_6-draft.html.

[4] "Free Riding on Gnutella", Adar E., Huberman B., in First Monday, vol. 5, number 10, October 2000. http://www.firstmonday.dk/issues/issue5 10.