Computer Science and Artificial Intelligence Laboratory

Technical Report

MIT-CSAIL-TR-2004-051
AIM-2004-018

July 25, 2004

Distribution Volume Tracking on Privacy-Enhanced Wireless Grid

Ozlem Uzuner

massachusetts institute of technology, cambridge, ma 02139 usa — www.csail.mit.edu

Abstract

In this paper, we discuss a wireless grid in which users are highly mobile, and form ad-hoc and sometimes short-lived connections with other devices. As they roam through networks, the users may choose to employ privacy-enhancing technologies to address their privacy needs and benefit from the computational power of the grid for a variety of tasks, including sharing content. The high rate of mobility of the users on the wireless grid, when combined with privacy-enhancing mechanisms and ad-hoc connections, makes it difficult to conclusively link devices and/or individuals with network activities and to hold them liable for particular downloads. Protecting intellectual property in this scenario requires a solution that can work in absence of knowledge about behavior of particular individuals.

Building on previous work, we argue for a solution that ensures proper compensation to content owners without inhibiting use and dissemination of works. Our proposal is based on digital tracking for measuring distribution volume of content and compensation of authors based on this accounting information. The emphasis is on obtaining good estimates of rate of popularity of works, without keeping track of activities of individuals or devices.

The contribution of this paper is a revenue protection mechanism, Distribution Volume Tracking, that does not invade the privacy of users in the wireless grid and works even in the presence of privacy-enhancing technologies they may employ.

1 INTRODUCTION

1.1 What is the Wireless Grid?

A computer grid is a collection of distributed resources shared among a group of users. Wireless grids [1] support a diverse set of devices that communicate with each other using a diverse set of protocols. Some of these protocols, like Bluetooth, are subscriptionless, and would allow users to connect to the grid via machines in their proximity. In this setting, the computational capacity of these small devices is no longer a constraint on the activities they can perform because as long as they have access to the grid, these devices can issue requests to the computationally more able devices on the grid and receive the results of their requests.

Given these parameters, imagine a scenario where user A wants to download a particular song to his MP3 player. His MP3 player is Bluetooth enabled. It detects and connects to the laptop of user B who is in its proximity. The laptop happens to be on the wireless grid. The request for download is sent from the MP3 player to the laptop. The network that the laptop is on finds the requested song and transfers it to the MP3 player. The MP3 player can now terminate the Bluetooth connection.

This scenario highlights couple of important characteristics of the wireless grid and its users. Namely, that:

1. Users are highly mobile and their connections to the grid can be ad-hoc and short-lived.
2. There is a wide variety of devices and protocols used for communication on the wireless grid.
3. Users can use subscription-based as well as subscriptionless services to connect to networks and tracking the behavior of individuals can be very difficult.

In addition, in this paper, we assume that:

1. Users can have subscriptions with multiple service providers.
2. Service providers have wireless grid infrastructure, which identifies the machines, learns from them about their computational and bandwidth resources as well as their willingness to participate in the grid, and considers this information during resource allocation.
3. Service providers do not share infrastructure.
4. Users can move through the networks of service providers to which they subscribe.
5. As the users move through the networks, they are identified, authenticated, and authorized automatically by the service providers.
6. Service providers participate in the protection of privacy and anonymity of their users and therefore allow them to authenticate with the servers while masking their private information, such as MAC addresses, IP address or, location if they choose to do so.

1.2 Intellectual Property in Wireless Grids

The dynamic, mobile and transient nature of the users of the wireless grid and the ability to share computational resources for ad-hoc activities creates interesting challenges for intellectual property protection on the wireless grid.

In addition to the high rate of mobility and the ability of users to form subscriptionless short-lived ad-hoc connections, a range of privacy-enhancing technologies used in mobile networking, such as those that allow users to make anonymous connections [2], enable users to be private, at least to some extent, with respect to their personal information. The combination of these characteristics make it difficult to link people and their mobile devices with patterns in network activities.

The grid, on the other hand, complicates the digital copyright problem by allowing individuals to use devices of others for their downloads and applications. This implies that the identity of the machine that holds the content may not be relevant to identifying the person responsible for the download. For example, imagine a scenario where a user with a wireless Palm PC wants to download and play a song. While he has the processor cycles and bandwidth, he lacks the disk space for this download. He requests a download to a machine in his proximity, from which he can listen to the song through streaming media. The presence of a downloaded song on a machine does not, therefore, mean that the owner of that machine has downloaded it.

The wireless grid puts these two challenges together. In these networks, it is difficult to relate observed network activity to the source and destination of distributed content or to link downloaded works to particular individuals. The question we try to address in this paper is: In an environment where people cannot be uniquely linked to devices or activities, how do we protect intellectual property?

In the following sections, we briefly present the digital copyright problem and outline what we want to achieve with a solution to this problem. Next, we review some proposed solutions from the literature, and discuss their shortcomings and strengths in the wireless grid. We continue with a proposal for author compensation through Distribution Volume Tracking (DVT), a form of digital tracking which is based on fingerprints of works. In section three, we discuss how digital tracking can work in the wireless grid, especially in the presence of increasingly more complex privacy-enhancing mechanisms. In section four, we describe the role of the ISPs for distribution volume tracking. We conclude with proposals from literature on payment mechanisms that can work with DVT and a discussion of open questions.

The main contributions of this paper include:

1. A proposal to use fingerprint-based distribution volume tracking for revenue protection in the wireless grid and in the presence of privacy-enhancing mechanisms that protect the privacy and identity of users in a variety of ways,
2. Demonstration, with examples, of how DVT can be deployed in the wireless grid and an analysis of the

required supporting technical and policy infrastructure for using DVT both as an independent volume tracker to obtain estimates of popularity of works, and as an independent verifier that can be used to double-check the rate of distribution reported by other sources.

2 DIGITAL COPYRIGHT PROBLEM: A BATTLE OF WILLS, A BATTLE OF INCENTIVES

The digital copyright problem revolves around two major groups of stakeholders: content owners and content users.

*Content owner*s are threatened by the ease of unauthorized replication and distribution of digital copies of their works and claim that they need strict control over their works, such as through the digital rights management (DRM) systems and supporting legislation, to protect their incentives to produce. The more control they have over their works, the content owners claim, the more willing they will be to share their works with the public.

Unfortunately, in addition to protecting the intellectual property of content owners by preventing unauthorized uses, the technological protection mechanisms such as DRMs also limit authorized uses, such as fair use, and invade consumer privacy while doing so [3]. This is against the interests of the second major stakeholder group, namely the *content users*.

In reaction to use of DRMs without regard to erosion of their fair use privileges, the invasion of their privacy and attacks on their anonymity, the content users have been bypassing the DRMs and circulating DRM-free versions of works. As a result, the content owners have been strengthening DRMs. And, both parties are stuck in a futile battle of wills. This battle affects the welfare of the whole society as it affects the wealth of informative and artistic works available to the community.

Once DRM-free versions of works have been created, limiting access to intellectual property becomes impossible. Therefore, in the wireless grid where the modes of distribution can be as diverse as the number of available communication protocols, and users can share all kinds of resources including hardware dedicated to storing content, we suggest addressing the digital copyright problem by focusing on revenue protection for content owners, and fair use, privacy and anonymity protection for content users.

2.1 Distribution Volume Tracking and Alternatives

In this section, we look at some representative solutions to the digital copyright problem and discuss whether they would be applicable to the wireless grid.

DRMs, i.e., technical protection mechanisms that encapsulate works and control their use, and technologies that use encrypted content and trusted computational devices [4] have so far been the solution insisted upon by the content owners. However, these mechanisms prevent some legally authorized uses of works along with unauthorized ones, instigate the users to bypass them, and thus fail to protect revenues. Therefore, we believe they will also fail to address the digital copyright problem in the wireless grid.

Watermarks, i.e., imperceptible identifying information embedded in the works, form the corner stone of another proposed solution. These mechanisms mark each work with a unique identifier so that each digital copy can be uniquely recognized. The unique identifiers can encode any information the content owner likes; including the identity of the content user who purchased the work [5]. This, unfortunately, means that, in case of theft of content from an authorized content user, for example, watermarks end up implicating innocent people for unauthorized distribution of content.

One of the revolutionary solutions to the digital copyright problem came from Gerovac and Solomon who proposed replacing copy protection with revenue protection, based on digital tracking that uses headers and descriptors to identify video content [6]. Our solution builds on this proposal. However, instead of using fixed headers, we propose using fingerprints of works based on their unique content and expression [7]. These fingerprints will be generated, using the computational resources of the grid, every time works are transferred within a network.

This kind of fingerprint-based digital tracking has several advantages. One major advantage is that the fingerprint of a work can be recomputed before every transfer. This makes the fingerprint robust to foul play; and making the fingerprint of a work unrecognizable requires making the work itself unrecognizable by making substantial changes to it. Second advantage is that the fingerprint matches between works do not have to be exact. Therefore, by requiring further examination of works that are similar beyond a threshold, we can capture works that are derived from each other, sometimes without proper attribution.

We have so far implemented a prototype digital fingerprinting system that can accurately identify books and their translators [7]. A challenge for full-scale deployment of this system is the performance bottleneck. The grid can address this problem, by delegating the fingerprinting task to computationally more powerful machines, so that works can be fingerprinted for their identity during transfer.

We intend to use our digital fingerprinting mechanisms for tracking the distribution volume of works, without tracking information about individual user behavior. We argue that for proper compensation of the content owners, all we need is a good estimate of the popularity of works as determined by the number of downloads. In the non-digital world, this is equivalent to the number of copies distributed/ sold of hard copy of a work.

The rate of distribution/ popularity of works can help proper compensation of content owners in several ways. For example, the rate of popularity of works can be used to guide the negotiations between content owners and licensees over the terms of compulsory licenses and extended compulsory licenses covering digital downloads [8]. Alternatively, agencies such as ASCAP, that sell compulsory licenses on behalf of content owners, can use the popularity information when compensating artists.

In short, Distribution Volume Tracking (DVT) using digital fingerprints:

1. Can Provide a robust way of accounting for popularity of works and enables the content owners to be compensated adequately,
2. Can Alleviate the privacy, anonymity and fair use concerns of users, does not unjustly link users to unauthorized behavior, minimizing the incentives of the content users to bypass it,
3. And, is not affected by the privacy-enhancing mechanisms that can be employed by users.

3 Distribution Volume Tracking in Presence of Privacy-enhancing Mechanisms in the Wireless Grid

DVT is a promising solution to the digital copyright problem that does not track individual user behavior or try to link users with network activities. This nature of DVT addresses the characteristics and needs of the privacy-enhanced networks including the wireless grid for intellectual property protection.

In this section, we show how DVT can be used in the wireless grid, in absence of encryption by users, by analyzing the following scenarios:

1. A peer-to-peer (P2P) network on the wireless grid with no dynamic rerouting of transfers.
2. A wireless grid which has to dynamically reroute transfers to handle the high rate of mobility of its users and a P2P network on this grid in which the P2P software deliberately reroutes transfers through a random number of nodes or generates phony download requests, to disguise the identity of the recipient and the identity of the work requested

In both of these scenarios, we assume that the wireless grid has a resource allocator that knows the machines that are present on the wireless grid at any time and can efficiently control the allocation of resources.

The resource allocator handles the fingerprinting task just like any other computationally intensive task; once it receives a request to fingerprint a work, it delegates this task the computationally powerful machines available on the grid.

3.1 Tracking in Peer-to-Peer Networks in Wireless Grid

The file transfers on the wireless grid are likely to be P2P. In other words, transfer requests will be between nodes A and B, each of which are end-users. Tracking distribution volume in these networks can be done at the ISPs servers. ISP can do this by issuing a fingerprinting request to the resource alloctor and thus keeping track of the distribution volume as works go through the servers.

There are two problems with this approach to DVT. One problem is that we can potentially double-count the transfers, especially in cases where the transfer involves users of different ISPs. If ISPs only keep track of works distributed *from* nodes in their network, this problem can be addressed to some extent. The second problem is the computational load on the grid infrastructure of fingerprinting every single copy of a requested file. To make this computational load manageable, we can randomly sample the file transfers and get an idea about the rate of popularity of works instead of getting exact counts.

3.2 Tracking in Privacy-Enhanced Peer-to-Peer Networks in the Wireless Grid

In the wireless grid, dynamic rerouting of transfers is crucial for proper handling of mobile devices. This means that a transfer destined for a particular machine may have to be rerouted several times until it reaches its moving target. What is more, this rerouting has to be robust even when networks are complicated by privacy-enhancing mechanisms, such as common anonymous networking strategies that, re-route files through a random number of nodes to hide the real destination and the source [9].

This makes accurate accounting for downloads more complicated because each hop forced by either the peer-to-peer software or the router can look like an independent transfer. For example, for enhancing privacy, it is possible to route transfers through random nodes, all of which only know about the node before and the node after them. The transfer can continue through a random number of nodes after the ultimate goal node is reached, making it very difficult to identify the real recipient. The inability to calculate the exact number of transfers in this case also advocates for random sampling of the transfers, by ISPs, to estimate the *relative* popularity of shared works.

An alternative to distribution volume tracking at the ISP is distribution volume tracking at the peers. With this, we mean that the DVT software will be installed at the machines of individual peers and it will fingerprint documents before they are sent from each node. The collected accounting information can then be reported, in aggregate, to the responsible organization to be used for compensating the content owners.

A reputation-based mechanism would complement this proposal by giving end users incentives to install and run the fingerprinting software on their machines. For example, the users can be asked to correctly attribute the works they distribute, the accuracy of their attribution can be checked by the independent verifier – in this case the fingerprinting mechanism behind DVT, and the result can affect the user reputation. A better reputation can then be translated into better grid services and the competition for the resources on the grid can encourage more people to participate in this accounting system. Note that, this method does not require perfect participation, i.e., all users to participate. As long as an adequate sample of the population participates in this scheme, we will have enough information about the rate of popularity of content owners and their works.

In addition, estimating the relative popularity of works instead of keeping exact download counts addresses issues related to untrackable downloads, such as those made by subscriptionless wireless devices. Also, inaccuracies in counting due to unreliable network, dropped and reinitiated connections/ downloads are less of an issue as long as they are not due to attacks on the network and do not bias the accounting.

Short of being exact, DVT through digital fingerprinting provides one way of establishing the relative rate of popularity of works, can be used to back up any other hypotheses about

popularity of works and can adequately protect the revenues of content owners without inhibiting fair use, invading privacy or affecting anonymity of content users.

4 ROLE OF ISPS

4.1 Authentication and Security

In the scenarios we described, ISPs are responsible for the distribution volume tracking mechanism. This puts some additional burden on the ISPs. Operating under the assumption that the ISPs can negotiate with the content owners to be properly compensated for their efforts, we believe that the ISPs can also be responsible for collecting revenues from their subscribers and distribute this sum among the content owners, in proportion to the rate of distribution of works from and within their networks. This highlights the importance of user authentication for the ISPs.

4.2 Payment Mechanism

As mentioned earlier, the information about distribution volume of works collected by digital tracking systems can be used as a guide for compensating content owners. But, how is this information to be used? What kind of a payment mechanism would minimize the incentive to bypass the tracking and payment systems? For example, if each user is asked to pay per download, some people will have incentives to bypass the tracking systems and avoid payment[1]. However, if the users pay a flat fee regardless of the number of their downloads, then the tracking systems do not affect the amount spent by the users; they only affect how the amount collected through the flat fee is distributed among the content owners.

Netanel argues that the ISPs can be the middlemen that collect the flat-fees from their subscribers through "non-commercial use levy" and distribute the collected money to content owners, using the information gathered by the tracking mechanisms about distribution volume of different works [10].

The payment mechanisms that are based on flat fees are admittedly unfair to some users of the network, because people who download fewer works end up paying the bill for the users who download more works. However, as Odlyzko has shown, users prefer simpler methods of pricing and will be willing to overpay as long as pricing is simple [12].

Fisher suggests taxation of digital entertainment devices and services such as MP3 players and ISPs for collecting revenues. Such taxes can be used to expand the financial resources available for compensating the content owners and would work well with solutions based on estimates of popularity of works [13].

4.3 ISP Coordination and Competition

The wireless grid allows users to roam through networks, and authenticate with different network providers as they do so. Assuming that the ISPs do not share infrastructure, an open question is: Is it fair to charge users who use multiple ISPs multiple flat rate fees? Subscription to more networks does not imply more downloads. Can the ISPs coordinate to eliminate this over-charge?

5 OPEN QUESTIONS

Despite the challenges posed by the wireless grid environment, distribution volume tracking through digital fingerprinting of works can help reduce the deterioration of content owner's revenues. However, to complete this proposal, we also need to find solutions to several open questions:

1. Is it enough to install a tracking system at the service provider on the wireless grid setting? How else, and where else in the network can a digital tracking-based DVT system be implemented? We believe there are more strategies and scenarios that we can look at, and we will find more ways in which DVT can be effectively implemented in the wireless grid.
2. Is the assumption that users will not encrypt their transmissions realistic? And, is DVT completely ineffective if content is encrypted? The assumption may not be realistic, however, presence of encryption does not cripple DVT. As Schechter explains, on a peer-to-peer network, for example, interested parties can obtain the peer-to-peer client software and pose as a client to get access to unencrypted content [4]. We do not propose posing DVT as a client, but rather distributing it with the P2P software that people use. This would make sure that people on the network can participate in the distribution volume tracking, even in the presence of encryption. As mentioned earlier, a reputation-based mechanism would be useful for encouraging the users to, in fact, install and use this software.
3. The end users are inevitably the resource users and providers in a grid. This means that the resource allocator will be delegating the fingerprinting task to the end devices of the users. Can we trust the users to not attack the integrity of the software? McKnight et al. [14] discuss that there can be technical, social or legal consequences to this kind of behavior that can deter unacceptable behavior. A reputation-based mechanism can work in this case also. For example, the fingerprinting task can be delegated to multiple users and the results can be crosschecked. If a user provides inaccurate fingerprints, this hurts their reputation and the processes they ask to delegate to the grid can be given lower priority. A similar mechanism has been implemented in the P2P file sharing systems Gnutella and Kazaa[14].
4. Finally, can the artists manipulate the system to inflate the rate of popularity of their works? Unfortunately, the answer is yes. However, there are reputation-based solutions that can address this issue [15].

6 CONCLUSION

The high rate of mobility of the users on the wireless

[1] Anania and Solomon's work supports this argument. These authors suggest that flat-rate fees are indeed better than pricing mechanisms that are based on the download volume of individuals because "the per-bit pricing mechanisms encourage the sophisticated user to manipulate" the systems to avoid payment [11].

grid when combined with privacy-enhancing mechanisms that can hide location, device identity, user identity, or other private information, makes it difficult to conclusively link devices and/ or individuals with network activities and to hold them liable for particular downloads. Protecting intellectual property in this scenario requires a solution that can work in absence of knowledge about behavior of particular individuals.

Tracking individual behavior and preventing unauthorized uses has been a popular strategy for intellectual property protection. Problem with this approach is that it gives users incentives to break the technological protection systems. For the digital copyright problem to be solved with any technological solution, we need to eliminate, or at least minimize these incentives and make it not worthwhile to break the technological protection systems. Pricing works competitively, reducing the invasion of privacy, enabling fair use without extra charges, and in general finding a solution that is "fair" to all parties involved is the most effective way of achieving this goal.

We believe, distribution volume tracking through digital tracking of content allows content owners to be compensated adequately (proportional to the popularity of their works) and avoids the privacy and fair use concerns raised by alternatives to distribution volume tracking, such as the DRMs. The ability of DVT to provide unbiased accounting of popularity of works, the fact that it does not invade privacy, but in fact eliminates the need to invade it, makes DVT an attractive solution to the digital copyright problem especially on platforms where the users are increasingly more concerned with their privacy.

REFERENCES

[1] Lee W. McKnight, Diana Anius, and Ozlem Uzuner. *Internet Services. Quality of Service in Grids, Networks and Markets*, chapter "Virtual Markets in Wireless Grids: Peering Policy Prospects". MIT Press, forthcoming.

[2] Michael G. Reed, Paul F. Syverson, David M. Goldschlag. "Protocols using Anonymous Connections: Mobile Applications". *Security Protocols Workshop*, pp. 13-23, 1997. citeseer.ist.psu.edu/ article/ reed97protocols.html

[3] Jessica Litman. *Digital Copyright*. Prometheus Books 2001. ISBN 1-57392-889-5.

[4] Stuart E. Schechter, Rachel A. Greenstadt, and Michael D. Smith. "Trusted Computing, Peer-to-Peer Distribution, and the Economics of Pirated Entertainment". *The Second Workshop on Economics and Information Security*. 2003.

[5] Giovanni's Applied Digital Watermarking Schemes http:/ / www.bluespike.com/ giovanni-dws.html

[6] Branko Gerovac and Richard. J. Solomon. "Protect Revenues, Not Bits: Identify Your Intellectual Property," *Proceedings of Technological Strategies for Protecting Intellectual Property in the Networked Multimedia Environment*. Interactive Multimedia Association. http:/ / www.cni.org/ docs/ ima.ip-workshop/ Gerovac.Solomon.html, 1994.

[7] Özlem Uzuner, Randall Davis. "Content and Expression-Based Copy Recognition for Intellectual Property Protection". In the *Proceedings of the 3rd ACM Workshop on Digital Rights Management (DRM'03)*, 2003.

[8] R. Anthony Reese. "Copyright and Internet Music Transmissions: Existing Law, Major Controversies, Possible Solutions". *University of Miami Law Review*, Vol. 55, No. 2, Jan. 2001 http:/ / ssrn.com/ abstract=276333.

[9] Michael K. Reiter, Aviel D. Rubin. "Crowds: Anonymity for Web Transactions". *ACM Transactions on Information and System Security*, Vol. 1, No. 1, pp. 66-92, 1998.

[10] Neil W. Netanel. "Impose noncommercial Use Levy to Allow Free P2P File-Swapping and Remixing," *30th Annual Telecommunications Policy Research Conference*, 2002.

[11] Loretta Anania and Richard Jay Solomon. "Flat: The Minimalist B-ISDN Rate". Presented at *MIT Workshop on Internet Economics*, March 1995. http:/ / www.press.umich.edu/ jep/ works/ AnaniaFlat.html.

[12] Andrew Odlyzko. "Internet Pricing and the History of Communications". *Computer Networks* (Amsterdam, Netherlands: 1999), vol. 36, No. 5--6, pp. 493—517. 2001. http:/ / citeseer.ist.psu.edu/ odlyzko01internet.html.

[13] William Fisher. *Promises to Keep. Technology, Law, and the Future of Entertainment*. Forthcoming, Stanford University Press, 2004. http:/ / tfisher.org/ PTK.htm

[14] Lee W. McKnight, William Lehr and James Howison, "Coordinating User and Device Behavior in Wireless Grids". October 2003. http:/ / itc.mit.edu/ itel/ docs/ 2003/ wirelessgrids.pdf

[15] Roger Dingledine, Michael Freedman and David Molnar, "Accountability", *Peer-to-Peer: Harnessing the Power of Disruptive Technologies*, Chap.16, pp.271-- 340, O'REILLY Press, 2001.