

**WARSAW UNIVERSITY  
OF TECHNOLOGY**

Faculty of Electronics and  
Information Technology



# PH.D. THESIS

Igor Margasiński, M.Sc. Eng.

**Anonymous Transport in  
Peer-to-Peer  
Overlay Networks**

**WARSAW  
2008**

# WARSAW UNIVERSITY OF TECHNOLOGY

**Faculty of Electronics and  
Information Technology**

## **Ph.D. THESIS**

Igor Margasiński, M.Sc. Eng.

**Anonymous Transport in Peer-to-Peer Overlay Networks**

Supervisor  
Professor Michał Pióro, Ph.D., D.Sc.

Warsaw, 2008



*You can be only a person for the world,  
but for a person you are the world.*

Gabriel García Márquez



# Abstract

Broadband access to the Internet has given a way for explosion of large scale distributed overlay networks. Peer-to-peer (P2P) overlay networks have grown to be one of the leading Internet applications and constitute a significant part of the Internet traffic. From its inception, development of P2P communications has gone hand in hand with the demand for privacy. A distributed architecture of P2P overlays provides an extremely promising environment for implementation of effective privacy mechanisms. The decentralization, as an immanent feature of the peer-to-peer communication, eliminates single points of potential observation and overcomes the need for trusted third parties. Nevertheless, we observe that highly secured P2P overlays of today struggle to provide satisfactory traffic performance.

The main contribution of this thesis is an original proposal, design and analysis of an anonymous peer-to-peer system called P2PRIV (peer-to-peer direct and anonymous distribution overlay). The basic novel features of P2PRIV are: a peer-to-peer parallel content transport architecture and separation of the anonymization process from the transport function. We have found that these features allow considerable savings of service time while preserving a high degree of anonymity. In the thesis we evaluate anonymity measures of P2PRIV (using a revised information entropy measurement model) as well as its traffic measures (including service time and network dynamics), and compare anonymity and traffic performance of P2PRIV with a well known system called CROWDS.

The entropy measurement model for the anonymity analysis has been revised with respect to the P2P environment usability and the practical possibilities of a P2P adversary. In addition to adaptive attack scenarios, widely described in the state of the art, we also considered more realistic static attacks. Static attacks demonstrated that boundless extension of forwarding path lengths can degrade anonymity. Apart from the passive observation, we also designed extremely invasive active attacks adjusted to the proposed P2P architecture.

We designed a traffic performance measurement model for the evaluation of the mean download time and system dynamics caused by an unpredictable users' migration and traffic bursts after a publication of a new popular content. We verified the model and observed that

empirical analysis provided by simulations of the CROWDS system have been analogous to theoretical values.

We found that P2PRIV effectively protects user privacy by assuring a high degree of anonymity. The proposed system significantly decreases the download time as compared with traditional cascade schemes, and achieves close to optimum results for low to medium loaded networks. Taking into account network dynamics, we found that the proposed system is more flexible and reacts faster to dynamically changing conditions, such as peers/content migration and traffic bursts introduced by a new data publication. The new system scales well and proves highly flexible in large networks.

Patent for P2PRIV is currently pending. Patent application (“A Method for an Anonymous and Direct Data Sending in Telecommunications Network”, #P384095, inventor: Igor Margasiński, assignee: Warsaw University of Technology, submitted on December 2007) successfully passed the preliminary evaluation of the Polish Patent Office. Expert studies of PPO confirmed the novelty of P2PRIV concept.

# Streszczenie

Popularyzacja szerokopasmowego dostępu do Internetu ostatecznie przełamała praktyczne ograniczenia hamujące rozwój rozproszonych sieci nakładkowych. Nakładkowe sieci peer-to-peer (P2P) wiodą dziś prym wśród aplikacji sieciowych, a wolumen ruchu P2P stanowi znaczną część ruchu w sieci Internet. Od samego początku, rozwój P2P szedł ramię w ramię z potrzebą ochrony prywatności. Właśnie to środowisko jest szczególnie obiecujące w realizacji skutecznych mechanizmów prywatności, a zerwanie z paradygmatem centralizacji usług pozwala na eliminację zaufanej trzeciej strony – powiernika prywatności. Jednak obecne rozwiązania, cechujące się wysokim poziomem bezpieczeństwa, wciąż nie zapewniają satysfakcjonującej wydajności.

Głównym przedmiotem rozprawy jest propozycja, zaprojektowanie i analiza nowego systemu anonimowego peer-to-peer, o nazwie P2PRIV (peer-to-peer direct and anonymous distribution overlay). Zasadnicze nowatorskie cechy P2PRIV to: równoległa architektura transportu danych użytkowych, oraz oddzielenie procesu anonimizacji od transportu tych danych. Zaproponowane rozwiązanie pozwala na znaczne obniżenie czasu obsługi zgłoszeń oraz zachowanie wysokiego poziomu anonimowości. W rozprawie dokonano oceny anonimowości systemu P2PRIV (z zastosowaniem zweryfikowanego modelu systemów anonimowych, opartego na pojęciu entropii informacji) oraz oceny wydajności systemu (w tym czasu obsługi i odporności na zmiany struktury sieci). Efektywność i wydajność badanego rozwiązania zostały porównane ze znanym i reprezentatywnym systemem CROWDS.

Znany ze „stanu sztuki” model systemów anonimowych został zweryfikowany w zakresie zasadności założeń i dostosowany do specyficznego środowiska nakładkowych sieci rozproszonych. Poza najczęściej rozważanym w literaturze scenariuszem ataku adaptacyjnego, przeanalizowano bardziej realistyczny wariant ataków statycznych. Włączenie analizy ataków statycznych pokazało, że nadmierne wydłużanie ścieżek anonimizujących może obniżać anonimowość systemu. Dodatkowo, poza obserwacją pasywną, w pracy rozważono scenariusze ataków aktywnych oraz opracowano szczególnie

inwazyjny typ tego ataku. Scenariusz ten został bezpośrednio dopasowany do charakterystycznych podatności proponowanego rozwiązania.

Do modelowania ruchu systemów anonimowych zaprojektowano i zrealizowano środowisko symulacyjne. Zastosowana metoda pozwala na badanie złożonych warunków pracy sieci P2P, takich jak reakcja na publikację nowego, popularnego zasobu informacyjnego oraz migracja węzłów sieci. Poprawność modelu została zweryfikowana poprzez porównanie wyników symulacyjnych z wartościami analitycznymi dla warunków brzegowych, dla systemu CROWDS.

Zaobserwowano, że system P2PRIV skutecznie chroni prywatność użytkowników poprzez zapewnianie wysokiego poziomu anonimowości. Zaproponowany system pozwala na znaczne podwyższenie szybkości anonimowego transportu nakładkowego w porównaniu do tradycyjnych systemów kaskadowych i osiąga wyniki bliskie optymalnych dla nisko i średnio obciążonych sieci. System P2PRIV jest bardziej odporny na zmiany warunków pracy sieci, takie jak migracja węzłów oraz wzrost natężenia ruchu na skutek publikacji nowego, popularnego zasobu informacyjnego. Nowy system cechuje się dobrą skalowalnością i odpornością na zmiany struktury sieci nakładkowych dużej skali.

Przedstawiony w rozprawie system został zgłoszony jako wynalazek do Urzędu Patentowego RP. Zgłoszenie („Sposób anonimowego i bezpośredniego przesyłu danych w sieci telekomunikacyjnej”, zarejestrowane 20 grudnia 2007, pod nr P384095, twórca: Igor Margasiński, na rzecz: Politechniki Warszawskiej) uzyskało pozytywną ocenę wstępną. W sprawozdaniu UP RP o stanie techniki wskazano, że żadne spośród branych pod uwagę rozwiązań nie podważa nowości i poziomu wynalazczego zgłoszenia.

# Contents

<b>1</b>	<b>Introduction .....</b>	<b>15</b>
1.1	Motivation.....	15
1.2	Contribution.....	16
1.3	Theses .....	17
1.4	Organization.....	18
<b>2</b>	<b>Definitions and assumptions.....</b>	<b>19</b>
2.1	Anonymity and privacy.....	19
2.2	Peer-to-peer overlay networks .....	20
2.2.1	Hybrid P2P.....	22
2.2.2	Pure P2P.....	24
2.3	Summary.....	26
<b>3</b>	<b>Techniques for network anonymity .....</b>	<b>28</b>
3.1	Third-party proxies .....	29
3.2	Mix-nets .....	34
3.3	Blending into the crowd.....	35
3.4	Anonymous P2P overlays .....	37
3.5	Summary.....	38
<b>4</b>	<b>Anonymity measures .....</b>	<b>39</b>
4.1	Entropy.....	39
4.2	Effective anonymity set size .....	40
4.3	Degree of anonymity.....	41
4.4	Attack model.....	42
4.4.1	Attack example: CROWDS .....	42
4.4.2	Static attacks.....	43
4.5	Summary.....	48
<b>5</b>	<b>Traffic performance measures .....</b>	<b>50</b>
5.1	Empirical model of P2P traffic .....	50
5.2	Download time.....	52
5.3	Dynamics .....	54
5.4	Summary.....	56
<b>6</b>	<b>Peer-to-Peer direct and anonymous distribution overlay (P2PRIV system).....</b>	<b>57</b>
6.1	Overview of P2PRIV .....	57
6.2	Design of P2PRIV peer.....	59
<b>7</b>	<b>Anonymity analysis for P2PRIV .....</b>	<b>67</b>
7.1	Quantitative analysis of P2PRIV secure working point.....	67
7.1.1	Passive-static attacks .....	67
7.1.2	Passive-adaptive attacks.....	70
7.1.3	Active attacks.....	73
7.1.4	Summary .....	78
7.2	Qualitative analysis.....	80

*Contents*

7.2.1	Passive-static attacks .....	80
7.2.2	Passive-adaptive attacks .....	81
7.2.3	Active attacks .....	83
7.2.4	Summary .....	85
<b>8</b>	<b>Traffic performance analysis for P2PRIV .....</b>	<b>88</b>
8.1	Download time .....	89
8.2	Dynamics .....	90
8.3	Scalability analysis for P2PRIV .....	90
8.4	Summary .....	96
<b>9</b>	<b>Conclusions .....</b>	<b>98</b>
9.1	Summary of the contribution .....	98
9.2	Future work .....	100
9.2.1	Traffic performance modeling for anonymous system .....	100
9.2.2	DHT suit .....	101
9.2.3	Anonymous content publication .....	101
9.2.4	Applications .....	102
	<b>Bibliography .....</b>	<b>103</b>
	<b>Index of Acronyms .....</b>	<b>111</b>

# List of Figures

FIGURE 2-1 CLASSIFICATION OF PEER-TO-PEER OVERLAY NETWORKS.....	22
FIGURE 2-2 EXAMPLE OPERATION OF HYBRID PEER-TO-PEER OVERLAY.....	23
FIGURE 2-3 EXAMPLE OPERATION OF PURE PEER-TO-PEER OVERLAY.....	24
FIGURE 3-1 NETWORK RANDOM WALK ALGORITHM.....	36
FIGURE 3-2 EXAMPLE OPERATION OF ANONYMOUS PEER-TO-PEER OVERLAY.....	38
FIGURE 4-1 RANGE OF OBSERVATION FOR STATIC AND THE ADAPTIVE ATTACKS.....	44
FIGURE 4-2 ENTROPY OF CROWDS, STATIC AND ADAPTIVE ATTACKS, $N = 50$ .....	46
FIGURE 4-3 ENTROPY OF CROWDS, STATIC AND ADAPTIVE ATTACKS, $N = 1000$ .....	46
FIGURE 4-4 ENTROPY OF CROWDS, STATIC AND ADAPTIVE ATTACKS, $N = 50$ .....	47
FIGURE 4-5 ENTROPY OF CROWDS, STATIC AND ADAPTIVE ATTACKS, $N = 1000$ .....	47
FIGURE 5-1 ARCHITECTURE OF P2P TRAFFIC SIMULATOR.....	51
FIGURE 5-2 MEAN DOWNLOAD TIME FOR CROWDS RANDOM WALK IN A PERIOD OF TWO DAYS AFTER START OF NETWORK OPERATION, $N = 100$ , MAXIMUM REQUEST ARRIVAL RATE.....	53
FIGURE 5-3 MEAN DOWNLOAD TIME FOR CROWDS RANDOM WALK IN A PERIOD OF TWO DAYS AFTER START OF NETWORK OPERATION, $N = 100$ , LOW REQUEST ARRIVAL RATE.....	53
FIGURE 5-4 REACTION OF CROWDS RANDOM WALK TO THE NEW CONTENT PUBLICATION, $N = 100$ , LOW REQUEST ARRIVAL RATE, $D = 20\%$ .....	55
FIGURE 5-5 REACTION OF CROWDS RANDOM WALK TO THE NEW CONTENT PUBLICATION, $N = 100$ , LOW REQUEST ARRIVAL RATE, $D = 30\%$ .....	55
FIGURE 6-1 P2PRIV ARCHITECTURE.....	57
FIGURE 6-2 P2PRIV CONNECTION AND CONTENT EXCHANGE.....	58
FIGURE 6-3 STATE DIAGRAM OF P2PRIV PEER.....	60
FIGURE 6-4 PSEUDO-CODE DESCRIPTION OF A P2PRIV SINGLE NODE OPERATION.....	61
FIGURE 6-5 PSEUDO-CODE DESCRIPTION OF THE CLONE AND DOWNLOAD SUBROUTINE.....	62
FIGURE 6-6 PSEUDO-CODE DESCRIPTION OF THE FIND AND REPLY SUBROUTINE.....	63
FIGURE 6-7 PSEUDO-CODE DESCRIPTION OF THE UPLOAD SUBROUTINE.....	64
FIGURE 6-8 OPERATION OF P2PRIV CLONE AND DOWNLOAD.....	66
FIGURE 7-1 ENTROPY OF P2PRIV, PASSIVE-STATIC ATTACKS, $N = 50$ .....	69
FIGURE 7-2 ENTROPY OF P2PRIV, PASSIVE-STATIC ATTACKS, $N = 1000$ .....	70
FIGURE 7-3 ENTROPY OF P2PRIV, PASSIVE-ADAPTIVE ATTACKS, $N = 50$ .....	72
FIGURE 7-4 ENTROPY OF P2PRIV, PASSIVE-ADAPTIVE ATTACKS, $N = 1000$ .....	72
FIGURE 7-5 ENTROPY OF P2PRIV, ACTIVE-STATIC ATTACKS, $N = 50$ .....	76

*List of Figures*

FIGURE 7-6 ENTROPY OF P2PRIV, ACTIVE-STATIC ATTACKS, $N = 1000$ .....	76
FIGURE 7-7 ENTROPY OF P2PRIV, ACTIVE-ADAPTIVE ATTACKS, $N = 50$ .....	77
FIGURE 7-8 ENTROPY OF P2PRIV, ACTIVE-ADAPTIVE ATTACKS, $N = 1000$ .....	77
FIGURE 7-9 DEGREE OF ANONYMITY FOR P2PRIV AND CROWDS, PASSIVE-STATIC ATTACKS, $N = 100$ .....	81
FIGURE 7-10 DEGREE OF ANONYMITY FOR P2PRIV AND CROWDS, PASSIVE-ADAPTIVE ATTACKS, $N = 100$ .....	82
FIGURE 7-11 DEGREE OF ANONYMITY FOR P2PRIV, CC INTERCEPTION ACTIVE-STATIC AND ACTIVE-ADAPTIVE ATTACKS, $N = 10$ .....	84
FIGURE 7-12 DEGREE OF ANONYMITY FOR P2PRIV, CC INTERCEPTION ACTIVE-STATIC AND ACTIVE-ADAPTIVE ATTACKS, $N = 100$ .....	85
FIGURE 8-1 NUMBER OF NODES AND LINKS FOR CASCADE (A) AND PARALLEL (B) ANONYMIZATION.....	88
FIGURE 8-2 MEAN DOWNLOAD TIME FOR P2PRIV AND CROWDS RANDOM WALK, $N = 100$ .....	89
FIGURE 8-3 REACTION OF P2PRIV (LOWER GRAPHS) AND CROWDS RANDOM WALK (UPPER GRAPHS) TO THE NEW CONTENT PUBLICATION, $N = 100$ , $\lambda = \lambda_{\text{MAX}} = 33,3^{-1}$ [MIN <sup>-1</sup> ].....	91
FIGURE 8-4 REACTION OF P2PRIV (LOWER GRAPHS) AND CROWDS RANDOM WALK (UPPER GRAPHS) TO THE NEW CONTENT PUBLICATION, $N = 100$ , $\lambda = 240^{-1}$ [MIN <sup>-1</sup> ]. .....	92
FIGURE 8-5 MEAN DOWNLOAD TIME FOR DIFFERENT NETWORK SIZES ( $N$ ) OF P2PRIV AND CROWDS RANDOM WALK. ....	93
FIGURE 8-6 REACTION OF P2PRIV (LOWER GRAPHS) AND CROWDS RANDOM WALK (UPPER GRAPHS) TO THE NEW CONTENT PUBLICATION, $N = 1000$ , $\lambda = \lambda_{\text{MAX}} = 33,3^{-1}$ [MIN <sup>-1</sup> ].....	94
FIGURE 8-7 REACTION OF P2PRIV (LOWER GRAPHS) AND CROWDS RANDOM WALK (UPPER GRAPHS) TO THE NEW CONTENT PUBLICATION, $N = 1000$ , $\lambda = 240^{-1}$ [MIN <sup>-1</sup> ] .....	95

# List of Tables

TABLE 7-1 EFFECTIVE ANONYMITY SET SIZE FOR P2PRIV .....	79
TABLE 7-2 DEGREE OF ANONYMITY FOR CROWDS AND P2PRIV .....	86
TABLE 8-1 MEAN DOWNLOAD TIME FOR CROWDS AND P2PRIV .....	96

# Acknowledgments

This thesis would not have been possible without the help of many people I would like to acknowledge here.

First, I want to thank my research advisor, Prof. Michał Pióro, for his support, reviews, and invaluable suggestions. I greatly appreciate his commitment to keeping me motivated throughout the process and allowing my ideas to come forward. It was a great honor to work with him.

Next, I would like to thank Prof. Andrzej Jajszczyk and Prof. Zbigniew Kotulski for their reviews and suggestions to improve this dissertation.

My special thanks goes to my colleagues and officemates from the Department of Computer Networks and Switching. I have been very fortunate to have worked with people who have always been willing to help. Thank you for all your great advice, support and friendship.

I also want to thank my family for their love and understanding and continuing support. My special thanks goes to my sister for her help in dealing with the linguistics mysteries of my dissertation.

# Chapter 1

## Introduction

### 1.1 Motivation

Telecommunications increasingly impact our daily lives. Every day activities, such as learning, shopping, meeting people, or managing finances involve network techniques. Today's economy is based directly on information. These changes, often viewed as a new age of civilization, demonstrate that information, including personally attributable information, quickly becomes a key commodity.

Unfortunately, the abuse of networks user privacy, in its sociological, financial, and technical aspects, is growing with the same pace [73]. Obtaining and using of personal information, mostly to gain a commercial advantage, constitute a major and global attack on network users. Privacy intrusion is intentional, because personal information is a "hot" commodity. Network privacy abuse has many faces. Apart from a simple theft (such as stealing credit card numbers), there are many more subtle attacks launched, in particular – consisting of compiling user profiles (personally attributable information on user's habits, preferences, interests, and even geographical location [20], [24], [65]). Irrespective of the kind and immediate outcome of the attack, the long-term results are user discomfort, ranging from mild irritation up to the total denial of services use, and commercial marginalization of the Internet as a whole (this seems a rather farfetched claim, but in our opinion it is more likely than one is ready to admit) [98]. These serious consequences call for effective countermeasures.

The success of the Internet network is based on its openness. The distributed model of the network provides means of communications robust against geographical locations, financial and formal boundaries. The effortless exchange of information on the Internet acts as a catalyst in the development of the information society. Still, when it comes to the application layer communication, a centralized architecture is often retained. The World Wide Web

## *Introduction*

service, as a core Internet application, has preserved among its users the perception of the largest wide area network as a client-server environment, where designated network nodes control delivery of content ([7], [67]). The extraordinary growth of the Internet made it an influential media. However, a reduction of its independence is the price. In this field, new and non-profit services arise. The broadband Internet access opened the door for an explosion of large scale, distributed applications. A confirmation of this direction is a migration of the WWW, which in the existing formula seems to be exhausted. We can observe decentralization of the WWW and development of new Web services compiled from many independent sources – a Web 2.0 phenomenon [74]. However, the real lifeblood of the changes are new peer-to-peer overlays. Today's volume of the P2P traffic consumes about 80% of the overall Internet communications. This environment is extremely promising for an implementation of effective privacy mechanisms, because getting rid of the centralization paradigm allows for elimination of a trusted third party – privacy trustee. The known anonymity techniques introduce significant traffic overheads. Hence, when we consider large size content transport, typical for P2P overlays [40], increased delays become onerous issue.

### 1.2 Contribution

In this work we present a novel peer-to-peer system. The system introduces a parallel content transport architecture, separating the anonymization process from the transport function. We applied an information entropy measurement model to analyze anonymity measures of the new system and to compare the system with a classical architecture. The model has been revised with respect to the P2P environment usability and the practical possibilities of a P2P adversary. We considered adaptive attack scenarios, that are widely described in the state of the art, and extended our analysis to more realistic static attacks. Apart from the passive observation, we also designed extremely invasive active attacks, which still cannot be detected quickly in the proposed, parallel architecture. To evaluate the practical usefulness of the solution, we analyzed the system's traffic performance measures including a mean download time and the system dynamics caused by unpredictable users' migration and traffic bursts after a publication of a new, popular content. Similarly to the security evaluation, we compared the traffic performance measures with the analogous

measures obtained from the well known cascade system. Finally, we analyzed scalability of both architectures, as this vital feature of P2P allows for spontaneous growth of distributed overlay networks.

### 1.3 Theses

Our first claim is that we can design such a peer-to-peer overlay system that separates the anonymization process and the content transport function, and retains an acceptable anonymity level. In particular, we claim that the well known network anonymity techniques can be used for anonymization of a specific management communications adjusted to provide further anonymous and direct parallel transport of the shared information content. To prove this assertion we use the recent analytical methods based on information entropy and revise them towards modeling of the peer-to-peer overlay networks anonymity. To describe the anonymity of the solution we quantify the level of the anonymity of the new system and also compare the effectiveness of the system with the classical solution.

Secondly, we show that the proposed solution allows for both a significant reduction of the content's download time and for an improvement of the overlay network dynamics. To prove this claim, we have designed the simulation environment of P2P traffic to model the complicated conditions of dynamic P2P users and content migration. The simulation model has been verified by comparing the simulated measures with the analytical measures that we managed to obtain for representative boundary conditions. To evaluate the traffic performance of the new anonymous system, we have replicated both the new system and the well known representative of a traditional cascade system, and compared the results with analytically calculated optimum values.

Our third challenge is to make sure that the proposed solution scales well. We show that the large size of the network does not significantly reduce the system performance. In order to accomplish this task, we have carried out an analysis analogous to the one mentioned above, but with altered number of network nodes, and we also sought to achieve a general dependency of a network size and the system performance.

## 1.4 Organization

The thesis is organized as follows. Chapter 2 provides an introduction to electronic privacy issues and contains definitions of anonymity and peer-to-peer overlays. Chapter 3 discusses related work and describes techniques for network anonymity. Chapter 4 explains methodology for anonymity measurement. Chapter 5 describes and verifies simulation methods for traffic performance analysis of P2P overlays. Chapter 6 contains a description of the novel P2P anonymous system and its anonymization idea. Analytical evaluation of the anonymity is presented in Chapter 7 while traffic performance simulations are presented in Chapter 8. Conclusions of our contributions and discussion of a future work are presented in Chapter 9.

# Chapter 2

## Definitions and assumptions

### 2.1 Anonymity and privacy

In the common meaning, the term anonymity describes the state of being unnamed, unidentified, or generally speaking unrecognized. Popular expressions, like “anonymous letter”, “undisclosed location”, “anonymous principal”, or “traveling incognito” correspond to individuals’ desire to retain their anonymity while engaging in certain activities. In the field of telecommunications, anonymity has a similar meaning. Anonymity, as a one of the information hiding techniques, represents a key method for assuring electronic privacy. In contrast to other network security techniques, anonymity is not aimed at protection or hiding of some communications or network resources, but rather at hiding their association with particular actors, for example network users. According to the role of the actor in the telecommunication network, there can be distinguished three types of anonymity:

- *sender anonymity* – hiding information about the initiator of a communication;
- *receiver anonymity* – hiding information about the destination point of a communication;
- *unlinkability* – hiding information about association between the initiator and a destination within a communication.

Moreover, anonymity as a technical term was influentially defined by Pfitzmann and Köhntopp [76]:

*Anonymity is the state of being not identifiable within a set of subjects, the anonymity set.*

The definition used widely in the state of the art is simple and natural. The surrounding set of subjects, potentially associated with a specific communications, provides its anonymity and is

### *Definitions and assumptions*

analogous to the effect of “blending into the crowd” of others subjects. This denotation of anonymity will be used in the rest of the thesis. Certainly, the definition is too general for a quantification of anonymity. The range of the “anonymity set” alone speaks strongly in favor of anonymity; however we should also consider anonymity’s other practical aspects. The task of providing numerous anonymity sets can be easily (simply) accomplished by involving other users in the activity of “feigning” and hiding a particular user. However, in this way, the traffic performance is usually reduced and download time elongated.

## 2.2 Peer-to-peer overlay networks

Broadband access to the Internet has opened a door to an explosion of large scale distributed overlay networks. Peer-to-peer (P2P) overlays reached the status of one of the leading Internet applications and constitute a significant part of the Internet traffic. From its inception, development of P2P communications has gone hand in hand with the demand for anonymity. Still, we observe that present-day highly secured P2P overlays struggle to provide good traffic performance.

P2P architecture – which tempts with its scalability, reliability, low cost of implementation, and obviously with its freedom of information exchange – is also not a subject of surveillance of any single unit [61]. Therefore, a fear of a lack of security emerges, since roles of an administration support and its locations are difficult to define. The entities responsible for providing security seem uncertain, because of the wide dispersion of these responsibility roles. When we consider pure model of P2P, all of its users are entrusted with network administration duties. However, it is unreasonable to require high networking skills from each user. Therefore, as we consider P2P overlays, it is important to be vigilant about the design of its protocols, which should be hold up to peers compromising. Collaboration of particular nodes should not compromise other network users. Nevertheless, P2P applications are sometimes perceived as an “underground” of the Internet and they often balance on the edge of legality. In corporate networks, this independent nature of peer-to-peer overlays is sometimes even recognized as a “sabotage” activity. This fear has a reasonable base, because the P2P development has often been based on amateur projects introducing various

### *Definitions and assumptions*

vulnerabilities and hidden functionality. Notice that P2P overlays actually build an individual network architecture overlaid on the TCP/IP network and duplicate networking functions, for example routing or transport. This “virtual” network provides numerous possibilities of its misuse against its users and their privacy.

Three basic classes (also called “generations”) of P2P can be pointed out. Notice that this general classification corresponds also to an evolution of P2P privacy:

- I. *hybrid P2P* – includes features of a client-server network architecture and relies on the central node or nodes; an observation of the central node enables a global observation of the overlay’s users;
- II. *pure P2P* – is composed of nodes with equal functionality; each node enables an observation of an local environment of an overlay (usually a group of neighbor nodes’ users);
- III. *anonymous P2P* – is dedicated to assuring anonymity; in the ideal anonymous P2P overlay an observation of an overlay node or a group of nodes and its communication links does not allow for tracking any users of the overlay.

Figure 2-1 illustrates general possibilities of privacy abuse in each class of P2P. In the first class of P2P overlays a central node exists, which is directly exposed to its activity/traffic observation and thereby it is also vulnerable to tracking of all connected nodes. After spectacular copyrights infringement trials, one of the most famous services of this type (*Napster* [96]) was closed. However, appetites for independent exchange of information were whetted. A natural direction of P2P overlays development was a migration from a simple hybrid model to a purely distributed architecture, where all peers have equal possibilities (for example *Gnutella* [42], [48]). In this scenario there is no single node exposed to tracking. However, it is still possible to track particular users of peer-to-peer overlay system. The next class of P2P was intended to also eliminate this weakness. These solutions are usually based on an assignment of a role of proxy to peers. Random proxy peers build anonymity of a sender and a receiver by forwarding traffic for others (for example in *Freenet* system [17]).

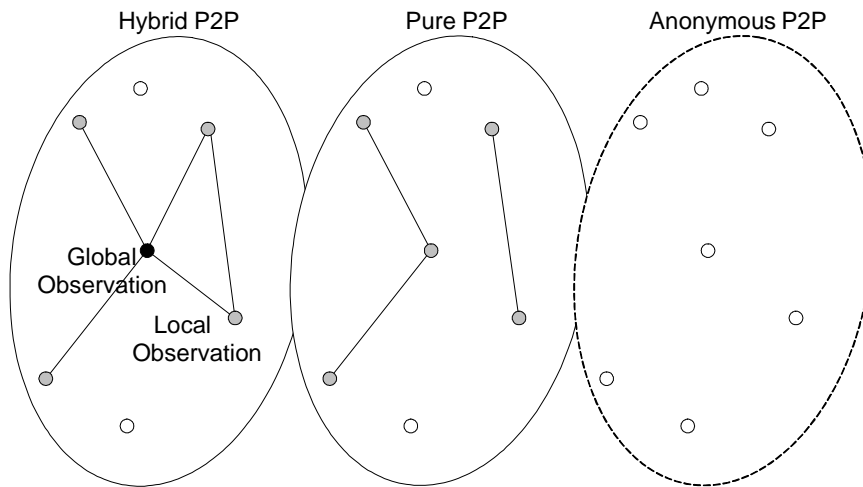


Figure 2-1 Classification of peer-to-peer overlay networks.

### **2.2.1 Hybrid P2P**

The file sharing boom (including an exchange of large size multimedia content) was initiated by hybrid systems. These solutions were based on the simple assumption: the client-server model is used if it is easier to benefit from the client-server model (usually in the process of content look-up), and if the presence of the server is not longer necessary, the peer-to-peer communication occurs (mostly in the final phase of content transport).

Let us follow an operation of the hybrid peer-to-peer variant using an example of the influential Napster system. In this class of overlays all nodes connect to a central node where they can access information about other nodes and their content resources. The basic role of the central node (server) is the maintenance and the distribution of a valid index of overlay nodes. Additionally, the central node represents a so called “chat room” allowing other users to communicate with each other. The system is based on a client-server protocol designed for the TCP transport. Users of the system, by means of client-side applications, connect to the server and register by specifying their personally attributable information. The administrator of the central node can exclude (“ban”) selected users. Clients inform the server about their content resources and also request from the server analogous information about other clients. The server responds with the address of the node which stores the copy of the requested file

### *Definitions and assumptions*

(originally MPEG-1 Audio Layer 3 format). After that, the initiator connects directly to the designated network node. Nodes notify the central node about their status and their link capacities. This gives the server simple load balancing capabilities. Additionally, overlay nodes declare a throughput of their links. This information allows the server to designate the most appropriate peer for an access to requested content.

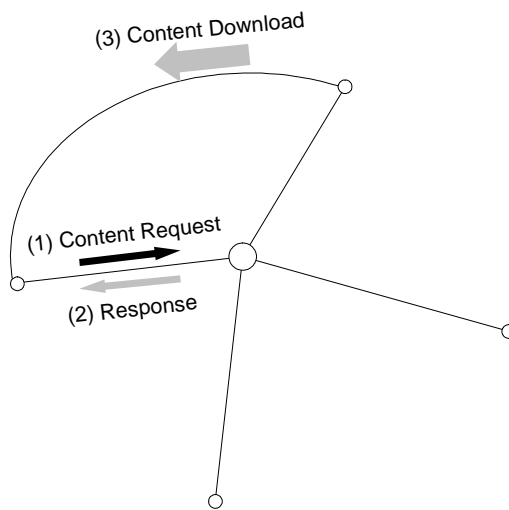


Figure 2-2 Example operation of hybrid peer-to-peer overlay.

The presented scheme is uncomplicated and also effective. This effectiveness, which is based on a combination of both client-server and peer-to-peer models, had played a key role in popularization of P2P overlay networks. However, the described solution had quickly proved not to be practical, because of its privacy deficiencies. The information about activities of all parties involved in P2P communications is concentrated in the central point of the hybrid peer-to-peer overlay – the server. This variant, in spite of its general assignment to the group of peer-to-peer applications, relies on a server (or servers) and operates under its absolute control, yet, with all of its consequences – including its inability to survive.

### 2.2.2 Pure P2P

An operation of this class of peer-to-peer overlay networks can be observed based on an example of “opened” and well known protocol called Gnutella. There is no designated central node in this system and all peers have the same functionality. Additionally, the network does not contain any single repository with index of network nodes or their resources. Before a new participant can connect, the user of this new node has to know at least one address of other peers. Then it can send a request message for a particular content, which will be broadcasted to successive nodes. The reply will come only from peers, which store the requested file. Finally, the sender and the receiver connect directly to transport the requested content. The Gnutella protocol is based on five types of messages to achieve these goals:

- “PING” and “PONG” – nodes look-up,
- “QUERY” and “QUERY HITS” – content look-up,
- “PUSH” – content download (firewalls bypass).

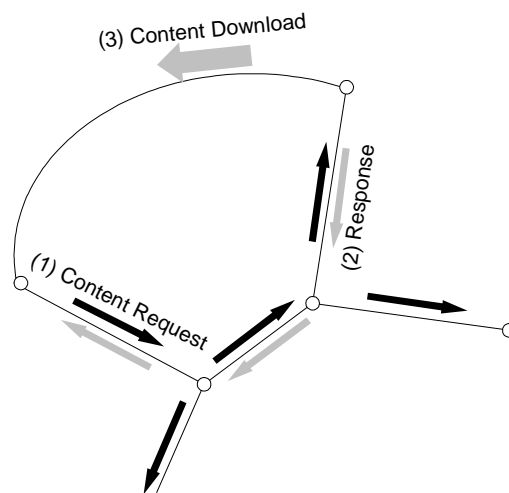


Figure 2-3 Example operation of pure peer-to-peer overlay.

All Gnutella messages contain “TTL” (*Time to Live*) field that, together with a message identifier, permits for the elimination of routing loops.

### *Definitions and assumptions*

Figure 2-3 shows the procedure of a content look-up (using the QUERY and the QUERY HIT messages). The look-up for the nodes (PING and PONG messages) operates analogously except for the last step, which obviously is not required. This step – content download – is realized over HTTP (HyperText Transfer Protocol) ([8], [39]).

The overlay systems with pure peer-to-peer architecture significantly hinder the global observation of its users. Additionally, we can obtain a more reliable and scalable solution, because of its robustness against the single point of failure problem. However, simple protocols like Gnutella suffer from scalability problems, denial of service vulnerabilities, and also low efficiency caused by traffic overheads of PING and QUERY messages broadcasts. Indeed, many more practical solutions were proposed. Still, they usually adapt some hybrid features (for example *eDonkey* [51], [104], *FastTrack/Kazaa* [59] systems). These overlay networks do not have pure P2P architecture, because in their hierarchical structures some designated nodes (called “super nodes”) fulfill additional tasks.

The described systems belong to the so called “unstructured” overlay networks and represent a core of popular implementations. Unstructured solutions have an ad-hoc nature rather than a structured architecture based on a mathematical basis. All connections and relations between network nodes are directly determined by successive users’ actions. Current research offers a lot of promise in the area of structured solutions. A lot of structured peer-to-peer networks like *CAN* [81], *Chord* [94], *Pastry* [87], *Tapestry* [105], and *Koorde* [53] are based on exact basis derived from a graph theory, e.g., PRR-trees, de Bruijn, Butterfly, and Hypercube graphs. The common feature of these solutions is usage of *DHT* ([79], [54], [60], [50], [4], [93]) – *Distributed Hash Table* algorithms.

DHT allows for fast look-up of content resources dispersed among pure P2P nodes. In the DHT structure each content resource is identified by a unique numeric key. This identifier (called a “DHT key”) determines a storage location of the resource. Therefore, network nodes have no direct influence on the content their store, as the DHT algorithm decides about the location of each stored file in the overlay network. The DHT can be treated as a black box allowing just for one operation – finding a content resource by giving a DHT key. DHT APIs give usually a `lookup(key)` function which returns a network address of the node currently storing the file. DHT keys are calculated using a one way hash function, such as SHA-1 [34]

### *Definitions and assumptions*

or MD5 [86] algorithms. The hash function takes a file as an input and returns a key allowing for an identification of this file. Two basic tasks of DHT operation can be distinguished:

- allocation of content resources among pure P2P nodes,
- acquisition of information about a content location without any central repository.

The performance of the first task is closely related to the execution of the second task. The specific allocation of content resources should allow for its fast look-up. Today's DHT algorithms usually guarantee look-up of path lengths less than or equal to  $O(\log N)$  hops. In accordance to the applied structure of the graph and appropriate routing algorithm, after  $O(\log N)$  or less hops the request message should reach the location of the requested file where  $N$  is the total number of network nodes. Additionally, the first task should provide load-balancing into the process of mapping file keys to network nodes.

In spite of the impact of inclusion of hybrid P2P features on privacy degradation, even purely distributed solutions cannot effectively protect the privacy of single users, because it is possible to observe selected areas of the overlay by taking control of a group of overlay nodes.

### 2.3 Summary

Nowadays peer-to-peer is a phenomenon that cannot be ignored. The existence of private and uncensored communications is no longer an option but a must. We can observe a high demand for solutions which can assure a private existence in a surrounding digital world. Telecommunication networks, like never before, allow for a fast and ubiquitous exchange of information. However, like never before, a great number of data, logs and activity traces are gathered and processed. One should believe that a free sharing of information is possible in the modern, telecommunications supported world. The main area opened for a free electronic dialogue is the realm of distributed, peer-to-peer overlay networks. However, from the beginning of P2P development, its spontaneous and "backstage" nature was also responsible for its potential threads and vulnerabilities. After a few years of evolution, P2P solutions are

### *Definitions and assumptions*

widely recognized as a mean of communication with a well organized, distributed and scalable architecture. Still, distributed topology cannot automatically assure a high level of privacy. Private peer-to-peer communications require an implementation of effective techniques for network anonymity. The third class of P2P overlay networks – anonymous P2P – is described in Section 3.4.

# Chapter 3

## Techniques for network anonymity

Partial solutions to the problem of electronic privacy violations represent a number of viewpoints, with complementary (rather than orthogonal and disjoint) classes of mechanisms in each of the viewpoints ([14], [69]). For the purpose of further discussion we will stick to two basic viewpoints:

- protection policy viewpoint:
  - negotiation of the declared level of privacy protection, based on a certain level of mutual trust;
  - privacy protection based on the assumption of untrustworthy and hostile environment;
- architectural viewpoint:
  - client-side and network-based solutions.

The solutions of a privacy negotiation like the Platform for Privacy Preferences Protocol (P3P) ([19], [80]) are not a privacy protection mechanisms *per se* and cannot be considered as privacy enhancing technologies (PET), but rather as tools allowing a user to be informed of potential privacy risks, and to decide if his/her interest in contacting a given service is worth taking the risk [64]. P3P introduces a uniform and machine-readable format for privacy policies and for user's private data collected by service providers. A user's browser can read this information and automatically decide, e.g., whether to send user's id information or to allow Cookies ([57], [62]). P3P is a mechanism that does not guarantee in any way the advertised privacy policy. However, as a standardized mechanism, it may be attractive to service providers in the long term, as it can increase mutual trust. P3P can help in achieving harmony between companies' economical needs for information (which is

required to provide services) and customers' rights to privacy and control over their personal information.

Client-side utilities can also only play a secondary role. The functions of protection mechanisms running on a client machine are: monitoring and control of all connections to and from a user's computer (i.e., a personal firewall), system "cleaning", blocking intrusive traffic and detecting "trojan horses". A growing number of such applications combine many individual functions. However, they are not able to, for example, hide (conceal) data that may identify a host, like an IP address.

### 3.1 Third-party proxies

A Proxy acts as a middleman in the process of communication. Adding a proxy between a sender and a receiver makes the hiding of all kinds of information about both parties from each other possible. Each party can only access information about a proxy. In addition, a secure connection can be established between a user agent and a proxy, for example by using the SSL/TLS protocol (Secure Socket Layer, Transport Layer Security) [25]. If such a connection is used, other parties, such as ISP (Internet Service Provider), a LAN (Local Area Network) administrator, or random eavesdroppers cannot access the transferred information. Another advantage of a proxy is the ease of control and filtration of the transferred content.

Using a proxy to anonymize Internet services is a widely known and popular technique, due to its high efficiency of hiding users' identity data, easy access to the service, no additional requirements imposed upon users, simple architecture, insignificant delays, relatively low cost of system implementation, and no need for the modification of existing nodes and protocols. However, the use of third party proxy servers has also serious disadvantages. Proxy servers have access to information about the activity of users. Anonymity service providers induce the belief that this data is not collected, used or shared. However, if the anonymity service provider for any reason breaks this promise, the user will be exposed to an even greater risk than in a case of traditional usage of the Internet, because information collected by different opponents is much more difficult to combine and profile.

Furthermore, proxy servers do not protect against traffic analysis attacks. An eavesdropper (i.e. a third party) can observe the volume of transmitted data and correlate inputs and outputs (proxy server requests).

The single proxy is a trusted third party. This immanent feature of third-party proxy systems forces to trust the anonymity service provider. In general this limitation is difficult to eliminate. However, specifically for the WWW system, a method that overcomes the above shortcoming was proposed in ([66], [68]). VAST system (Versatile Anonymous System for Web Users), described there, is a “zero-trust” method, which assumes a hostile and untrustworthy environment. It is thus complementary to negotiation-based approaches such as P3P. VAST is also an active method – in some aspects its operation resembles an attack (classified in [2] as subterfuge or deflection) upon a party trying to profile a user. This feature, by some considered as “ethically dubious”, is the price to pay for the effectiveness of the method. We claim that this approach is justified by the volume of privacy violations and harm inflicted by them on the users. Other known privacy protection systems are based on hiding (e.g., substituting, encrypting) or masking (e.g., dispersing) the user traffic. When information is eventually revealed (e.g., decrypted) or unmasked, the protection is gone. The idea of VAST is to introduce a dummy traffic that could have originated from the user, but actually does not. Therefore, the observed traffic, regardless of whether eavesdropped, inferred from traffic analysis or just recorded by a network node, cannot be reliably attributed to a given user, and thus is useless for the attacker.

VAST uses one middleman node – a kind of a proxy server. To achieve anonymity with regards to this server, and also to render useless traffic analysis attacks, a specific kind of a dummy traffic generation mechanism is placed between a distant proxy and a local agent. More Web pages than actually requested by a user are transferred from a proxy to a client. Information about which content is actually the object of interest to the user rests with the user himself. An agent (a JAVA applet) cooperates with the user’s browser. While the user reads the page contents, the agent simulates Web activity of this user, by requesting random Websites from the proxy. This idea originates from the observation of a typical Web navigation behavior. A user does not request Websites at all times. Requests are sent in various time intervals, after the user has read the contents. VAST, unlike other existing

solutions, does not conduct any additional major activity while the transaction is taking place, but it rather utilizes free time in between, inherent in Web browsing.

VAST consists of two key elements: an agent (a Java applet) placed in the Web browser environment, and a proxy placed between the agent and a destination Web server.

The primary functions of the VAST agent that communicates with the proxy server by means of the secure SSL/TLS protocol include: (a) simulation of user Web activity; (b) generation of URL (Uniform Resource Locator) addresses as a background for addresses requested by the user; (c) receiving configuration parameters from the user and transmitting them to the proxy; (d) requesting pages selected by the user and pages selected by the dummy traffic generator; (e) receiving resources from the proxy; (f) dividing resources into a group of pages chosen by the user and dummy pages; (g) presentation of pages chosen by the user (skipping the dummy pages); (h) analysis of the level of user anonymity, calculated as a proportion of resources downloaded by the user to resources downloaded by the dummy traffic generator (which serves as a simulator of user activity); (i) presentation of actual anonymity level to the user.

The VAST proxy server is very similar to popular anonymous proxy systems. The main difference is the absence of a user interface. This function was moved to the VAST agent. The primary functions of the VAST proxy server are: hiding all user-identifiable data (including IP addresses) from a destination Web server, encrypting all data transmitted between the VAST agent and the VAST proxy (including the URL addresses of resources), optionally – encrypting all communication between the VAST proxy and a destination Web server, blocking cookies, scripts, programs and Java applets sent from destination Web servers.

For the purposes of the description of the VAST system operation the following two terms are introduced:

- *Web transaction* – a series of HTTP client requests and corresponding server responses, which represent a single Web page;
- *subject session* – a collection of Web transactions generated by a user, where all transactions can be connected with each other by links contained in the Web pages.

### *Techniques for network anonymity*

We presume that a potential adversary, who has access to the transmitted data, is able to extract individual transactions and sessions from observed communication. The dummy traffic generation corresponds to the establishment of additional sessions. Transactions which belong to these sessions typically take place while the user is reading the content of pages already received. The Agent also generates dummy traffic requests assigned to the original user session. This makes any reliable distinction of a “true” user session impossible. Specific properties of sessions generated by a human – identifiable semantic relations between transactions – are then lost. When the user starts a new session, the agent also restarts dummy sessions. An eavesdropper (who knows the algorithm of agent applet, which is open source), can not distinguish if a particular request comes from the user or from the simulator. The anonymity service provider – the strongest possible attacker – is only able to separate particular sessions. The provider may assume that one of these sessions is of interest to the user, but the provider does not know which one it is. The provider also does not know which requests from a particular session come from the user. The user can configure the number of dummy sessions. The user should configure the system before the first use – it is necessary to input a list of search engines preferred by the user. The Agent will then employ these engines to generate dummy traffic. The Agent will use a dictionary of queries downloaded from the VAST proxy server. It is important for the dictionary to contain a large number of queries. If the user enters a query not contained in the dictionary, the VAST Agent will issue a warning. In this case the VAST service provider may be able to infer that the query was not generated by the simulator. A request for a page via a search engine marks the beginning of a new session. The same rules apply to the beginning of dummy sessions. At first, the user requests are not submitted immediately. The choice, as to which transaction is executed first, is made randomly. In subsequent transactions, user requests have priority over simulator requests. However, if their frequency is higher than the frequency of dummy transactions, the user gets an appropriate warning.

The VAST system is designed to provide high performance, as perceived by the user – similar to the performance of traditional Web browsing. Additional anonymity focused operations occur when the user is reading the page contents. A question remains: how does the dummy traffic actually delay browsing? Sometimes the user just glances at the page. How long will he be forced to wait for the next page? VAST may block (disable) data,

### *Techniques for network anonymity*

including banner ads, which originate from third party servers (i.e. AdServers). Statistical analysis conducted by the VAST system authors showed that the size of ads published on most popular Web pages and Web portals often exceeds 50% of the total page size. The number of requests necessary to download a single page is often a multiple of requests to destination servers. In the VAST system, requests to third party servers may be replaced by dummy requests. Users often allow download of numerous advertising elements. Therefore, it is a valid claim that the replacement of ads with dummy traffic, which provides privacy protection, would be also acceptable (and welcome).

VAST evolved from popular single proxy systems. One of the novel ideas of VAST – the use of Web search engine resources to generate dummy traffic between the local agent and the distant proxy – may also be seen as its weakness. For users, who pay for the amount of downloaded data, it means higher costs. We should stress that the system can block advertisement elements originating from third party servers. This, in turn, means that the graphic files from third parties are “traded” for dummy traffic. As usual, there is a price to pay for anonymity. To preserve full security, the user cannot start navigating from direct URLs, but only from queries input into popular search engines. The requested phrases should be included in the VAST dictionary, which can be quite vast indeed. This means that, in some cases, the user would have to take a moment to think how to change his/her request to find what he is really looking for. Anonymity from the perspective of the VAST service provider is accomplished through masking. The provider may only presume, with certain probability chosen by the user that particular requests come from the user. The volume of dummy traffic should be maintained on a certain level, in order to perform an effective masking of the user. VAST is a technique which overcomes the most serious weakness of anonymous proxies – a service provider’s access to a private user’s data. However, this advantage is reached only in a specific environment – the Web system. VAST is an example of taking advantage of a specific environment for which it can be dedicated. Still, single proxies cannot assure privacy and anonymize a transport of general information content since in this case they pose a serious thread themselves.

## 3.2 Mix-nets

A major step toward the improvement the security level provided by proxy-based systems has been the replacement of a single proxy with a network of many intermediate nodes, where packets are routed at random through this network, and each node mixes and encrypts packets with a different public key. The identity of both a sender and a receiver is never disclosed to any single proxy in a network, and due to specific routing of encrypted packets an attack based on traffic analysis is unlikely to succeed.

This concept of network anonymization was introduced in Chaum's seminal paper [16]. The Mix-net system proposed there has become a foundation of modern anonymity solutions. Mix-net is an anonymous network composed of nodes called Mixes that forward anonymous messages. The strength of the solution consists in: (i) a specific operation of nodes which "mixes" forwarded messages, and (ii) an asymmetric encryption of messages transported between them. The purpose of such mixing is to hide the correlation between received and forwarded messages. In general, received data units are padded to a constant size length, encrypted, delayed for a batch aggregation and then sent (flushed) in a random order. Anonymous messages are sent usually via a chain of Mixes to eliminate presence of a trusted party and also to omit single point of failure imposed by a single Mix. In Mix-net, each message is encrypted recursively with public keys of Mixes from a forwarding path. Finally, a message for each successive forward has different bit representation and place in a flow of other Mix-net messages. This makes Mix-net communications practically untraceable and secured against eavesdropping [16]. In the original Mix-net, the route through a cascade of Mixes was fixed. Further improvements allowed for a random path selection in the so called free-route networks. Hybrid models with restricted number of connections and path selection narrowed to restricted-routes were also proposed in [22]. Still, both fixed cascades and fully interconnected Mix networks with random routes have assorted constraints and the advantage between them depends primarily on their application and a scale of the network ([11], [13], [18], [47]). Along with Mix interconnection issues, Mix-nets evolution proceeds also in the field of a single node design [89]. The original Chaumian Mixes flush all received messages. The so called Pool Mixes assure higher anonymity by keeping some number of messages in a pool buffer while flushing the rest.

The pool size can be static or dynamic. Besides pool selection algorithm a designation of conditions required for Mix flushing is also pronounced. Two major conditions of flushing are: a time based and a threshold based. The timed Mixes flush periodically while in threshold Mixes the number of received messages triggers the Mix flushing. A generalized framework for expressing batching strategies of mixes was presented in [31]. Recent research deals with Continuous Mixes that delay each message individually based on probabilistic distribution [55]. Message splitting towards sending one content via different routes was proposed in [90].

The development of mixing methods is primarily aimed at resolving a tradeoff between delay time imposed by batching and reordering of messages and the anonymity level. Regardless of potential discontinuities in incoming traffic, Mixes have to wait for sufficient number of messages to achieve untraceable mixing. One solution to this problem is generation of fake messages (dummy traffic) ([10], [21], [30]). Dummies enhance anonymity and allow particular Mixes to flush faster. However, additional and “empty” traffic finally delays delivery of parallel user data.

### 3.3 Blending into the crowd

The concept of Mix-net has been used in a wide range of applications such as e-mail (in *Babel* [47] or in the *Type I cyberpunk anonymous remailers* [45] with *Pretty Good Privacy – PGP* ([41], [106]) encryption progressed in *Mixmaster* [71] and *Mixminion* [23]), Web browsing [9], ISDN ([78], [52]) and general IP traffic anonymization (*PipeNet* [102] and *Onion Routing* ([46], [82], [83], [95], [56]) with circuit-based routing, followed by *TOR* ([70], [72])). Other solutions ([1], [15], [97]) seem to play a less important role or (as introduced in CROWDS system [85]) utilize a simple idea of “blending into a crowd” by traffic forwarding via a group of nodes before its delivery.

Anonymity of CROWDS is based on a random walk algorithm. A random set of CROWDS nodes forward anonymous messages without usage of mixing or public key encryption techniques (Figure 3-1).

*Techniques for network anonymity*

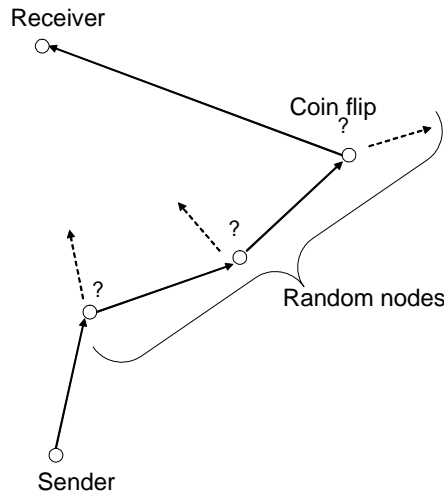


Figure 3-1 Network random walk algorithm.

CROWDS member, who wishes to send an anonymous message, selects a random node of the CROWDS network (the so called “Jondo”) and then sends the node the message. The message contains a destination address, but the source address is neglected. Then, the selected node flips an asymmetric coin to decide whether to forward the message to the next random node or send it directly do the destination. All random nodes repeat the same activity and finally the message is sent to its destination (based on the included address). The decision whether to forward the message to a next proxy node or to the message’s receiver is random. However, commonly a selection of a proxy is more probable than directing the message to the destination. This probabilistic forwarding assures anonymity, because any node of the network can ascertain the message’s origin. The coin asymmetry is described by a probability  $p_f$ . The proxy node forwards the message to the next random proxy node with the probability  $p_f$  and sends it to the destination with a probability  $1 - p_f$ . Then the mean forwarding path length of network random walk equals

$$P = \sum_{i=2}^{\infty} i p_f^{i-2} (1 - p_f) = \frac{p_f - 2}{p_f - 1} . \quad (3-1)$$

Authors of the CROWDS system suggest  $p_f = 0.75$  configuration, which corresponds to a mean length of random forwarding path equaled to 5. Basically, thanks to the network random walk mechanism (described above), messages can be sent anonymously. This is

possible, because proxy nodes cannot determine either whether the message was received from its initiator or whether the message's direct sender is also a proxy. The CROWDS system can be treated as a simplification of Mix-net anonymization with excluded mixing and asymmetric cryptography mechanisms.

### 3.4 Anonymous P2P overlays

Let us observe the operation of anonymous peer-to-peer overlays using the example of popular *Freenet* system. The Freenet's objective is to provide anonymity of a content receiver and a content sender. This is accomplished by mutual relaying of requests by particular network nodes using heuristic methods to finally deliver the message to its destination node. Additionally, communications between peers are encrypted. In Freenet, correspondingly to Gnutella, the TTL field protects against excessive storms of broadcast messages.

It is noteworthy that Freenet users do not have the ability to decide what content is stored on their computers. This feature (similar to DHT P2P overlays) accomplishes sender anonymity since the protocol is responsible for a content association to a peer. Additionally, each copy of a file is stored by several peers and indexed with a content hash function. A specific look-up service has been designed to correlate file names with proper hash keys. Each Freenet peer maintains a routing table with routes to several neighbors, and with hash keys, assigned to nodes which probably may store the content. At the start of a file download an originating peer sends a content request to its neighbor with the most similar key. The request is then forwarded in this manner until it reaches the node with a copy of the requested file. A proxy node, which detects a routing loop, redirects the request to a successive node with the most similar hash key. After the successive look-up, routing tables are being updated. The size of routing table entries is limited and if it exceeds the limit the less popular routes are deleted. The requested content is downloaded via the same route and all proxy nodes cache the file.

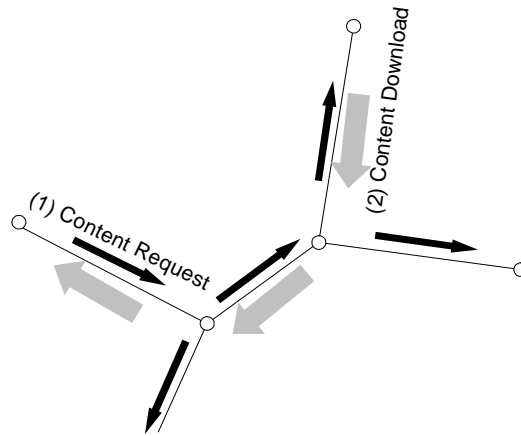


Figure 3-2 Example operation of anonymous peer-to-peer overlay.

### 3.5 Summary

The simple scheme of heuristics and encryption adaptation (used also in *Gnutnet* [6], [58]) to pure P2P architecture to assure anonymity seems to be effective. However, content relaying by middleman nodes – usually personal or SOHO computers – is not in favor of traffic performance of such systems. To improve traffic performance of Freenet, the new routing protocol was designed (the so called “next generation routing algorithm”). Although a significant download time reduction was achieved, it was at the cost of the system’s anonymity. Moreover, even without routing modifications, Freenet anonymity is seriously questioned. Based on simulation results from [12], Freenet system does not guarantee a proper level of anonymity for each network node in every case, and at random circumstances selected nodes are exposed.

Other P2P overlays, as well as various other applications, usually adapt the general idea of Mix-net anonymization to assure anonymity. Many variants of Mix-net such as Free Haven [26], Tarzan [27] and MorphMix ([84], [99]) were introduced for P2P. It should be noted that the basic common mechanism used to achieve P2P anonymization is traffic forwarding by a set of middleman nodes.

# Chapter 4

## Anonymity measures

In the previous chapters we introduced the definitions of anonymity and peer-to-peer overlay networks and described techniques for assuring network anonymity and their application in peer-to-peer overlays. The goal of this chapter is the presentation of methods for quantifications of anonymity and designation of anonymity metrics to facilitate an evaluation of efficiency of anonymity systems, including anonymous P2P overlays.

The term “anonymity set size” introduced in the definition of anonymity (compare with Section 2.1) is not sufficient for an accurate evaluation of this information hiding technique, mainly because the homogeneity of the anonymity set also impacts the level of anonymity. For example, let us consider an anonymous system in which an adversary can distinguish a set of 100 users, any one of which could be the sender of a traced message. Thus, a defined system assures high anonymity, because the anonymity set is numerous. However, it is possible for one of the users to be observed as the sender with a much higher probability than the other 99 users. In this case, the size of the anonymity set is of a minor importance and its extensions will not significantly improve anonymity. Additionally, the size of the anonymity set corresponds not only to the efficiency of an anonymization technique, but also to the size of the communication network in which it is applied. We need to provide a measurement methodology able to reflect the homogeneity of the anonymity set, its impact on anonymity, and also to enable an evaluation of anonymization efficiency independently from scale of considered systems.

### 4.1 Entropy

In the year 2002, two papers authored by Diaz *et al.* [32] and Serjantov *et al.* [88] simultaneously and independently introduced a new methodology for anonymity

measurement based on Shannon's information theory. The information entropy proposed by Shannon [91] describes the uncertainty associated with a random variable. It can be applied to anonymity quantification by assignment of probability of being an initiator of a specified action in the system to its particular users. Certainly, the sum of all these probabilities should equal 1. Then, based on the information provided by a system (shown by the system to an adversary) it is possible to measure the uncertainty of finding a real initiator.

Let  $X$  be a discrete random variable, and

$$p_i = \Pr(X = i) \quad (4-1)$$

where  $i$  corresponds to the number of a particular subject/node/user of the analyzed system. For each user from the set of all system users  $N$ , the adversary can assign probability of being the initiator  $p_i$ . Then the entropy  $H$  will be described by

$$H = -\sum_{i=1}^N p_i \log_2(p_i) . \quad (4-2)$$

In Equation (4-2), a base-2 logarithm was used. Therefore the unit of the expressed entropy is a *bit*. This measure discloses a number of bits required for an adversary to explicitly point out the initiator. The adaptation of information theory seems to be a proper method for anonymity quantification as the uncertainty of the observer increases proportionally with  $H$ . The anonymity corresponds intuitively to a blending into the crowd of other similar subjects. Entropy measurement, as a quantification of a disorder of a structure or a system, gives a description of this phenomenon and an analytical instrument to measure how particular subjects of the system are distinguishable among the whole population of subjects.

## 4.2 Effective anonymity set size

Serjantov *et al.* [88] used the entropy measure to calculate the effective anonymity set size  $S_A$  of anonymous systems. Let  $H$  be the entropy of the system, then

$$S_A = 2^H . \quad (4-3)$$

This quantification allows stressing anonymity of a particular system as an equivalent of the perfect system with  $2^H$  users. For example, if an analyzed system has 100 users and its effective anonymity set size equals 50, it means that in this case the system represents anonymity of a perfect system with 50 users. It can be illustrated as a perfectly homogenous crowd which surrounds the undistinguishable initiator.

### 4.3 Degree of anonymity

To enable the comparison between heterogeneous anonymity systems, a normalization of entropy was proposed by Diaz *et al.* [32]. The normalized entropy has been defined as entropy  $H$  divided by the maximum entropy of the system. Let  $H_{\max}$  be the maximum entropy for a current number of system users. Entropy reaches the maximum value when all possible users are equiprobable

$$H_{\max} = \log_2(N) , \quad (4-4)$$

where  $N$  is the number of users. Then the normalized entropy – degree of anonymity – can be described by

$$d = \frac{H}{H_{\max}} = - \frac{\sum_{i=1}^N p_i \log_2(p_i)}{\log_2(N)} , \quad (4-5)$$

where

$N$  – anonymity set size (the number of suspected subjects, for example users or network peers),

$p_i$  – probability assigned to  $i$  subject of anonymity set, describing how likely this subject is perceived by the adversary as the initiator.

This metric (4-5) describes the uncertainty of the system observer (the adversary) in finding the initiator of a specific action (for example sending a request for a specific content) and takes values from [0,1].

#### 4.4 Attack model

To achieve practical results it is important to assume realistic capabilities of an adversary corresponding to the specific environment of the attack. P2P overlays are primarily dedicated to public WANs, especially the Internet. We assume that users do not establish private groups and that no additional trust or access control mechanisms are provided. Any person can become the system's user and can utilize the system providing information in his own way. Initially, we assume that the adversary obeys the protocol and conducts a passive observation. Next we will consider active attacks enabling attackers to change protocol operation to disclose more information. Keeping in mind the large scale of public overlays and the pure P2P architecture, we will analyze local attacks where the adversary can control only a part of the system. We will study the impact of the number of collaborating nodes  $C$  on the degree of anonymity, and end with a look at a global attack. We assume that collaborating nodes can collect all information the system is "leaking" and send this statistics via an independent channel to the adversary headquarters for a summarizing analysis.

##### 4.4.1 Attack example: CROWDS

The CROWDS system (described in Section 3.3) can be used as an example of an attacked system, because it combines anonymity and performance with simplicity and reputability ([92], [103]). We will apply the entropy measurement model to quantify anonymity of the CROWDS system. The adversary, who foists colluding nodes to the network, can assign probabilities of being the initiator to particular network nodes. Based on [85] and [32] we can assign the following probability to a predecessor of the first colluding node from the forwarding path

$$p_{c+1} = 1 - p_f \frac{N - C - 1}{N} . \quad (4-6)$$

The rest of nodes will have assigned equal probabilities since the adversary has no additional information about them. All colluding nodes should not be considered,

$$p_i = \frac{p_f}{N} . \quad (4-7)$$

According to (4-2), the entropy of the system will be described by

### Anonymity measures

$$H_{paCROWDS} = \frac{N - p_f(N - C - 1)}{N} \log_2 \left( \frac{N}{N - p_f(N - C - 1)} \right) + \frac{p_f}{N} (N - C - 1) \log_2 \left( \frac{N}{p_f} \right). \quad (4-8)$$

The CROWDS maximum entropy is reached when all honest nodes are equiprobably recognized by the adversary as the initiator

$$H_{\max CROWDS} = \log_2(N - C), \quad (4-9)$$

then the normalized entropy equals

$$d_{paCROWDS} = \frac{H_{paCROWDS}}{H_{\max CROWDS}} \quad (4-10)$$

$$d_{paCROWDS} = \frac{(N - p_f(N - C - 1)) \log_2 \left( \frac{N}{N - p_f(N - C - 1)} \right) + p_f(N - C - 1) \log_2 \left( \frac{N}{p_f} \right)}{N \log_2(N - C)}. \quad (4-11)$$

#### 4.4.2 Static attacks

In the model proposed by [32] and [88] it is assumed that the adversary has yet colluding nodes among network nodes, which actively anonymize specified request (for example nodes from the forwarding random walk path of CROWDS system). Practically, the scenario may be different, and what is more, the probability that the adversary can find this group of nodes (referred to as an “active set”) also determines the quality of the system anonymization.

The scenario described above (Section 4.4.1) should be called an adaptive attack, because it is assumed that the adversary is capable of adopting an area of its observation to the scope of activity of system users. It is important to also consider a more general case, where the adversary cannot be certain of a successive collaboration of proper active set. This uncertainty should be quantified as well. For example, in the CROWDS system an increase of  $p_f$  parameter can easily increase the system active sets. However, if the active set is too numerous, even for large networks, collaboration of active nodes is simple – highly probable. Notice that when the adversary has no collaborating nodes among the active set then all the nodes of the system are equalprobable and the uncertainty of the adversary is maximized. The need for considering the impact of the observer uncertainty in finding proper active nodes was noticed in [32] and a weight mean formula was proposed to

compute final  $d$ . This seems to be an intuitive attitude. However, the same results can be achieved by using a conditional entropy formula. In [12] a conditional entropy was proposed to describe the generalized scenario of the observation. The conditional entropy describes the entropy of a random variable  $X$  under condition of elimination of the entropy of other random variable  $Y$ . The conditional entropy expresses then the uncertainty associated with one aspect when the other aspect is certain.

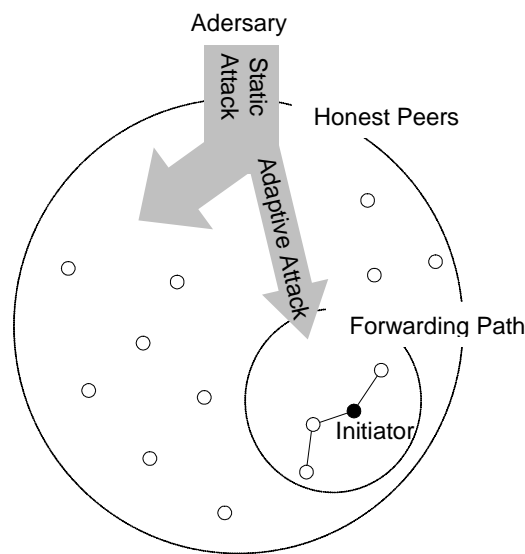


Figure 4-1 Range of observation for static and the adaptive attacks.

This attack will be referred to as a static attack, as in this scenario the adversary “injects” colluding nodes in a static manner and cannot dynamically predict (adapt, like in previous scenario referred to as adaptive attack) which random nodes will actively anonymize the specified request. Let us return to the CROWDS example. A probability that none of the collaborating nodes can become a member of the random walk forwarding path is

$$p_r = \frac{N-C}{N} (1-p_f) \sum_{i=0}^{\infty} \left( \frac{N-C}{N} p_f \right)^i = 1 - \frac{C}{N - p_f(N-C)}, \quad (4-12)$$

then entropy for passive-static attacks equals

*Anonymity measures*

$$H_{psCROWDS} = -\frac{C}{N - p_f(N - C)} \frac{N - p_f(N - C - 1)}{N} \log_2 \left( \frac{N - p_f(N - C - 1)}{N} \right) + \left( 1 - \frac{C}{N - p_f(N - C)} \right) p_f \frac{N - C - 1}{N} \log_2 \left( \frac{p_f}{N} \left( 1 - \frac{C}{N - p_f(N - C)} \right) \right), \quad (4-13)$$

and the normalized entropy is

$$d_{psCROWDS} = \frac{H_{psCROWDS}}{H_{\max CROWDS}} \quad (4-14)$$

$$d_{psCROWDS} = -\frac{C}{N - p_f(N - C)} \frac{N - p_f(N - C - 1)}{N \log_2(N - C)} \log_2 \left( \frac{N - p_f(N - C - 1)}{N} \right) + \left( 1 - \frac{C}{N - p_f(N - C)} \right) p_f \frac{N - C - 1}{N \log_2(N - C)} \log_2 \left( \frac{p_f}{N} \left( 1 - \frac{C}{N - p_f(N - C)} \right) \right). \quad (4-15)$$

We will analyze how  $p_f$  configuration impacts the entropy of the CROWDS system for both attack scenarios. It is important to remember that  $p_f$  value directly affects the active set size – the number of network nodes actively involved in the anonymization process (compare with Section 3.3). Figure 4-2 and Figure 4-3 show the entropy of CROWDS in the full spectrum of available  $p_f$  configuration. We use maximum entropy  $H_{\max CROWDS}$  (4-9) as a reference. We analyzed three variants of network collaboration level:

- $C = 10\%$  – scenario usually considered in the state of the art;
- $C = 5\%$  – more realistic collaboration level for large and public access overlays (where it is more difficult for the adversary to prevail the level of honest nodes); and
- $C = 20\%$  – scenario for small overlays (among a small population of honest nodes it is easier for the adversary to introduce the significant range of colluding nodes).

Anonymity measures

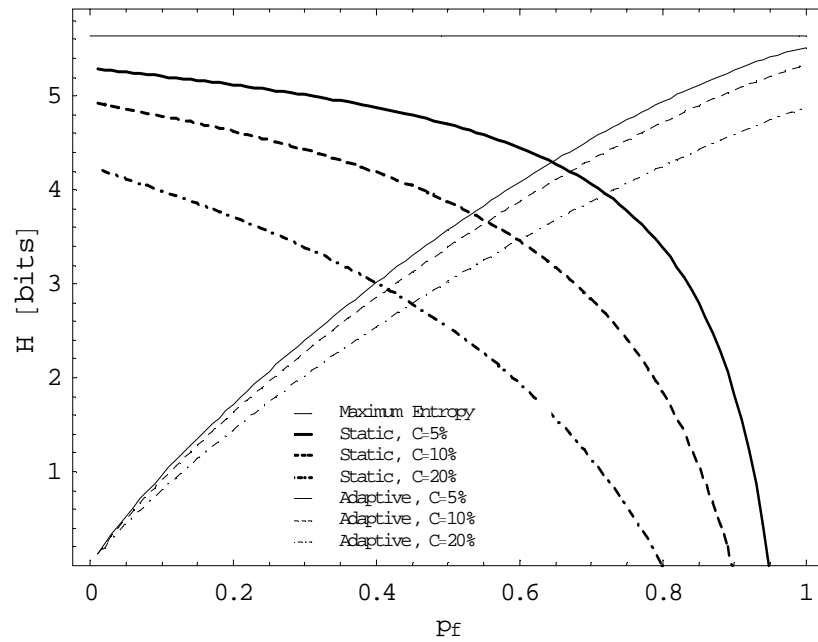


Figure 4-2 Entropy of CROWDS, static and adaptive attacks,  $N = 50$ .

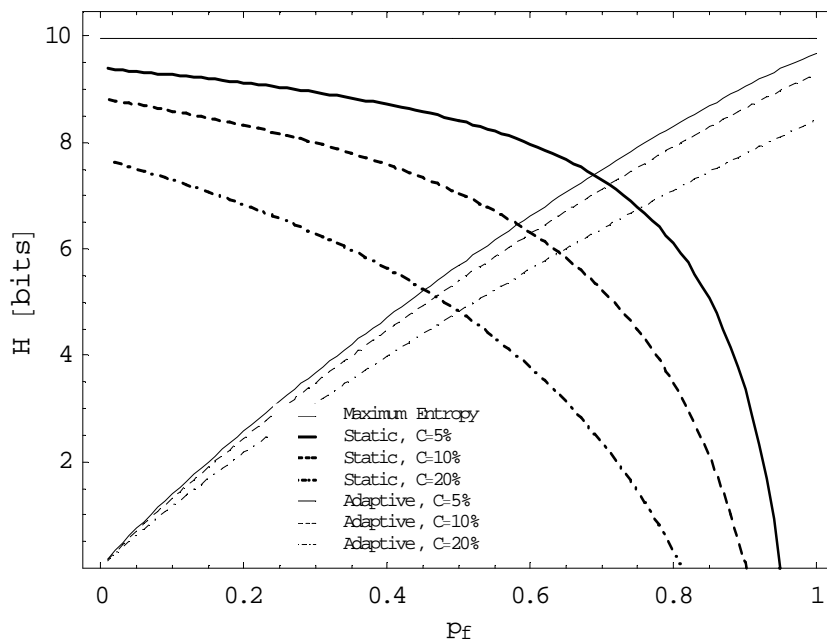


Figure 4-3 Entropy of CROWDS, static and adaptive attacks,  $N = 1000$ .

Figure 4-4 and Figure 4-5 show the entropy of CROWDS as a function of the number of colluding nodes  $C$  for both adaptive and static attacks.

Anonymity measures

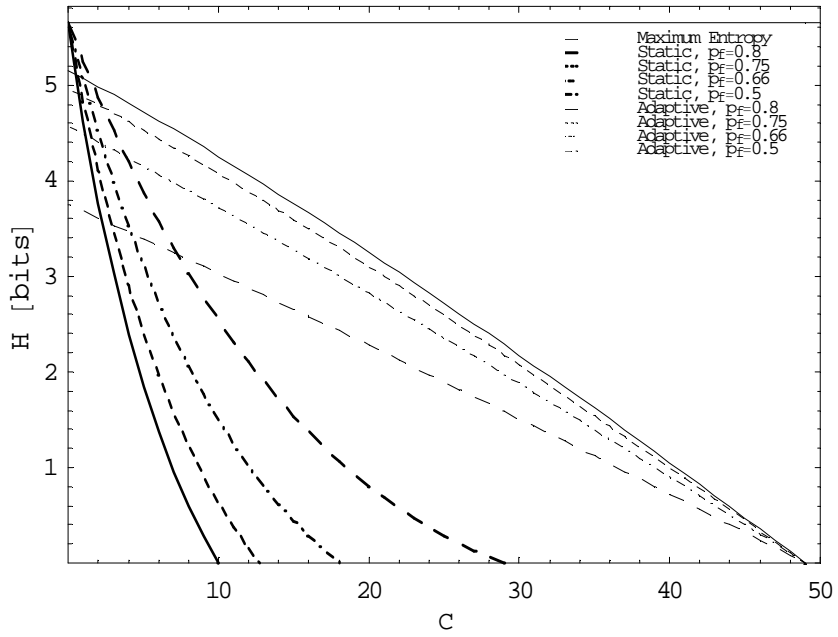


Figure 4-4 Entropy of CROWDS, static and adaptive attacks,  $N = 50$ .

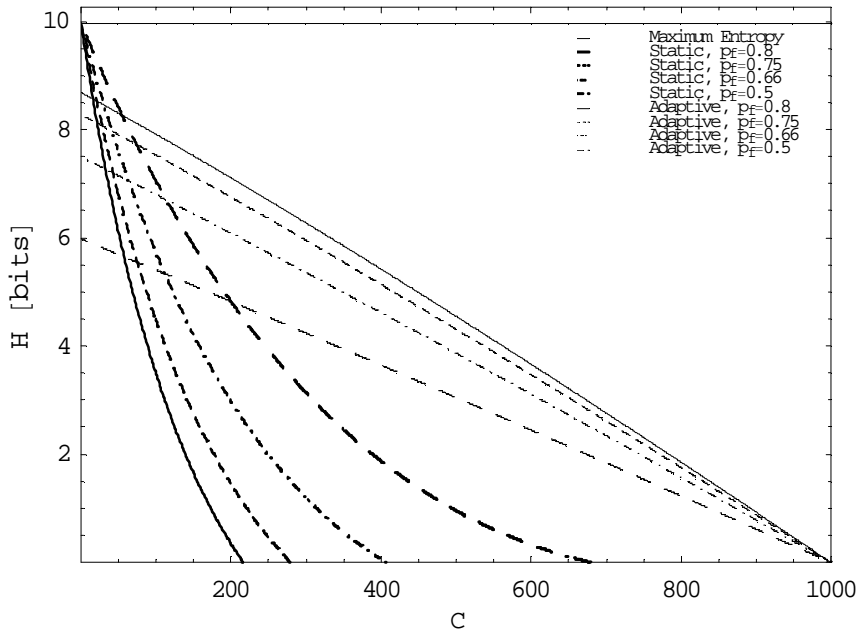


Figure 4-5 Entropy of CROWDS, static and adaptive attacks,  $N = 1000$ .

## 4.5 Summary

The results for the two considered attack scenarios are substantially opposite. In the adaptive scenario low entropy, close to zero, is obtained for low  $p_f$  values, and high, close to maximum, entropy is achieved for large  $p_f$ . In the static scenario, the dependency is quite different and the best results are achieved for the lowest  $p_f$  values. As  $p_f$  grows, the entropy grows logarithmically smaller. In a small network this decrease (static attack) of entropy is slightly faster in contrast to the adaptive scenario where, in the small network, the decrease of the entropy is slower than for large overlays. This analysis shows that a set of nodes actively involved in the anonymization process should not be too numerous. Longer cascades not only impose larger traffic overheads, but can also make it easier for the adversary to become a member of the active set. Especially in small networks the security of particular systems can be effectively compromised. In small networks, nodes from a forwarding path constitute a significant part of all network nodes. It should be reminded that the analyzed CROWDS system does not include mixing or asymmetric encryptions techniques for traffic protection.

The results show that the secure configuration of forwarding path lengths should be adjusted to the size of the CROWDS overlay network. Taking into account large overlays (with 1000 nodes and 5% collaborators) the  $p_f$  configuration of CROWDS should be from the range of [0.5,0.8]. Lower values than 0.5 expose the originator of a particular request against the adaptive adversary. Values higher than 0.8 compromise him to the static attacker. Using Formula 3-1 we can emphasize that acceptable CROWDS mean path lengths in large overlays are:

- minimum:

$$P_{\min CROWDS} = 3 (p_f = 0.5); \quad (4-16)$$

- maximum:

$$P_{\max CROWDS} = 6 (p_f = 0.8). \quad (4-17)$$

Forwarding paths shorter than  $P_{\min CROWDS}$  cannot provide sufficient “crowd” of nodes, which actively anonymize the initiator. If the adversary is yet among this set of nodes there

### *Anonymity measures*

should be additional 2 other honest nodes. On the other hand, the forwarding paths longer than 6 nodes ( $P_{\max CROWDS}$ ) becomes too easy to enter, because the size of the “crowd” provided by nodes passively anonymizing the active set becomes insufficient. Our analysis confirmed that the optimum configuration for a realistic level of CROWDS collaboration is about  $p_f = 0.75$  (recommended by the system authors).

As one can expect, the entropy largely depends on the number of colluding nodes. What is more, we can observe a significant impact of static observation on the anonymity of the CROWDS system. CROWDS entropy is significantly lower for static attacks than for adaptive scenarios.

Still, both static and adaptive variants of attacks are vital to the anonymity analysis as they correspond to different aspects of the system’s anonymity. The static scenario shows more realistic capabilities of the adversary and constitutes a critical point of view on the expansion of the system active sets. However, a more pessimistic attack – adaptive observation – is possible. Even though this scenario happens comparatively rarely, it is important to analyze its consequences.

# Chapter 5

## Traffic performance measures

Certainly, techniques of information hiding, such as anonymity, require traffic overheads and can therefore potentially degrade the network traffic performance. Hence, usefulness of a particular anonymity solution depends not only on its security level, but also on necessary traffic overheads. For P2P overlays a basic performance factor is a mean download time (DT) quoted as a mean time required for a content transport after a submission of the request by the user. As we consider public access peer-to-peer overlays it is important to take into account dynamically changing conditions of a network structure and of a location of shared content. Unpredictable users' migration and traffic bursts after a new publication of a popular content characterize the open P2P environment. Besides the mean download time and scalability, we will analyze dynamics of the system in reaction to a new content publication. The scenario covers DT impacts of the request arrival rate and content migration.

### 5.1 Empirical model of P2P traffic

For the purpose of the complicated dynamic conditions analysis, we have created a peer-to-peer traffic simulation environment. Each peer of the simulated network retrieves the same algorithm suitable to a simulated protocol (for example Jondo of CROWDS). The simulator traces tasks for each symmetric peer independently. The peers can collect a specified content and can randomly leave the overlay – depriving other users of the content copies. Figure 5-1 shows an outline of the simulator architecture.

The simulator has been implemented in ECMAScript/JavaScript ([35], [36]) language (ISO/IEC 16262) and requires only a Web browser to operate. The simulator contains about 500 lines of code. We have dedicated a Web server to provide flexible management for

### Traffic performance measures

various configurations of simulated overlays. This pervasive platform allowed us to employ many hosts for simulation computations.

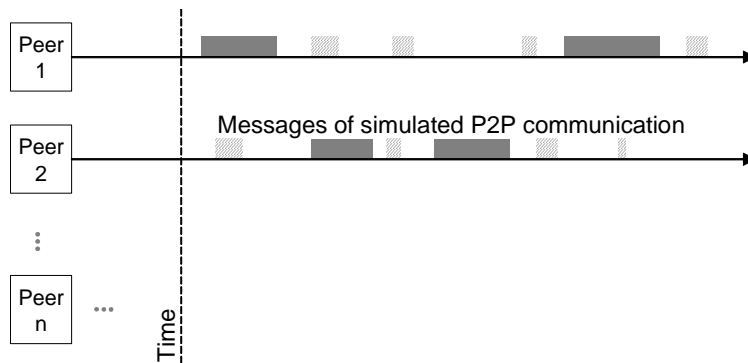


Figure 5-1 Architecture of P2P traffic simulator.

The simulator allows the analysis of the following parameters of an overlay traffic:

- mean request arrival rate ( $\lambda$ ) – intensity of requests for a content;
- mean download time between two neighbor peers ( $\mu^{-1}$ ) – resultant of a link throughput and an amount of sent data;
- mean request arrival rate for a specified content ( $\lambda_{Dc}$ ) – intensity of requests for a specified (for example newly published) content;
- mean migration rate ( $\lambda_{Dm}$ ) – intensity of users' leaving and arriving in the overlay network (“churn”).

We use Poisson distribution to model a request arrival process. We simulated the CROWDS random walk algorithm to verify the simulation model. In further analysis, like in the anonymity evaluation, we will also use the CROWDS system as a reference. It is a comparative system, which fairly represents cascade systems as it admits simplifications of the anonymous cascade schemes for better traffic performance. Nodes of CROWDS do not mix or delay forwarding content and also do not use asymmetric cryptography (compare

with Section 3.3). These simplifications give the CROWDS better traffic performance results. CROWDS system was originally dedicated for Web browsing, thus we included the content caching functionality for each forwarding node.

## 5.2 Download time

We assume rather typical, for today P2P overlays, values of link throughput and shared file size. Let average link throughput between peers be  $B = 512$  kb/s and average file size of the shared content  $V = 32$  MB. To analyze systems mean download time we have computed series of simulation with 30 realizations each starting from the maximum request arrival rate per each node

$$\lambda_{\max} = \frac{\mu_{\min}}{P} = 0.0005[s^{-1}]. \quad (5-1)$$

where  $P$  is a mean random walk path length. From Equation 3-1  $P$  equals

$$P = \frac{p_f - 2}{p_f - 1}, \quad (5-2)$$

and a  $\mu_{\min}^{-1}$  denotes download time between two directly connected nodes (referred in the rest of the work as FTP for simplification),

$$\mu_{\min} = \frac{B}{V} = 0.002[s^{-1}]. \quad (5-3)$$

Figure 5-2 and Figure 5-3 show 95% confidence intervals and 25% to 75% quantiles (marked as boxes) surrounding the mean values of DT for the CROWDS system as the function of parameter  $\lambda^{-1}$ . In the analysis we have assumed CROWDS configuration which is accurate with recommendation of the CROWDS authors and with our analysis:  $p_f = 0.75$ . It means that the mean number of overlay nodes forwarding a single request equals 5 (3-1).

*Traffic performance measures*

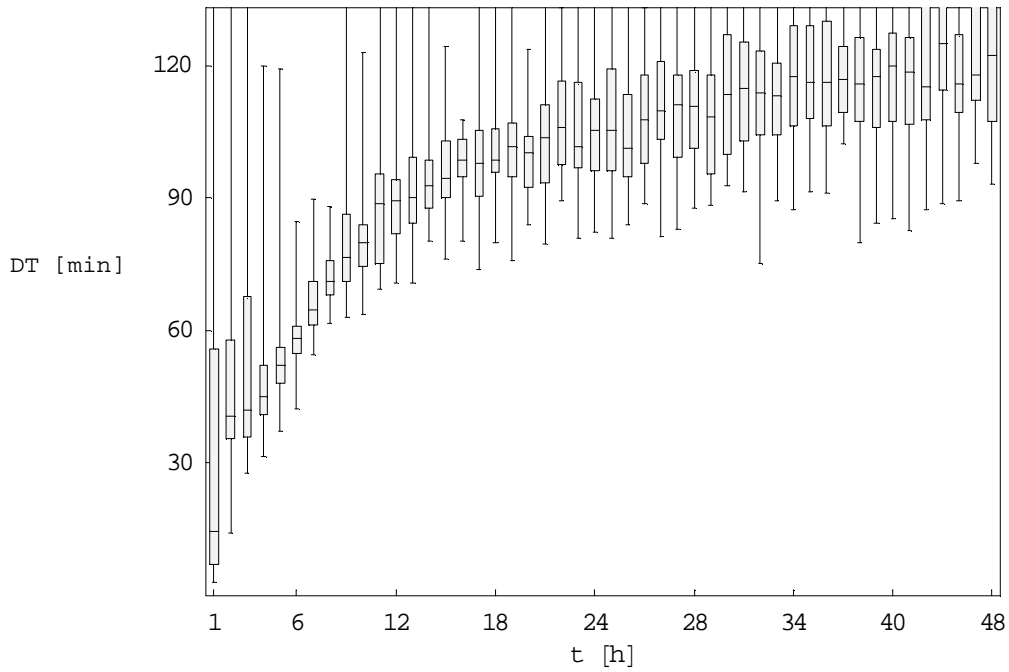


Figure 5-2 Mean download time for CROWDS random walk in a period of two days after start of network operation,  $N = 100$ , maximum request arrival rate.

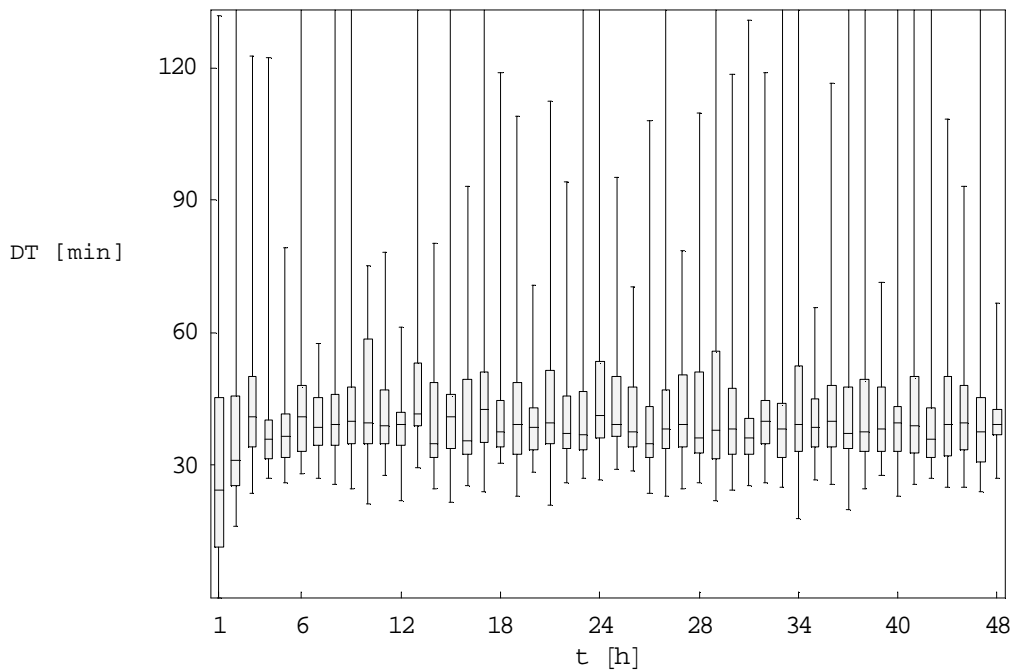


Figure 5-3 Mean download time for CROWDS random walk in a period of two days after start of network operation,  $N = 100$ , low request arrival rate.

### Traffic performance measures

Our first remark is that the simulated CROWDS random walk overlay works on the brink of stability for a analytical maximum request arrival rate. Simulations results for a lower request arrival rate showed the stable operation of the overlay.

Secondly, for a low arrival rate we can observe download time much above half an hour. The mean DT of a single file, measured in a period of two days of the system operation, equals 38.82 minutes. The analytically obtained mean time of a file transfer between neighbor nodes equals  $\mu_{\min}^{-1} = 8\frac{1}{3}$  minutes. In the CROWDS random walk each file is sent through a cascade of nodes. In the configuration of  $p_f = 0.75$  the mean number of links in the cascade between source and destination peers equals 4. The simulation results show that DT for the CROWDS random walk is about 4.6 times longer than a FTP DT for a single link.

### 5.3 Dynamics

Next we will apply the simulation model to consider the mean DT characteristics under dynamically changing network traffic conditions. We will analyze system behavior starting from a new file publication. We analyze the scenario where the new and popular content is just shared by one of the overlay nodes. Additionally, we take into account the common practice of some users to connect to the overlay only for the purpose of a particular content download and to leave the network just after its successful delivery. Let  $D$  be the part of all requests which corresponds to the new file. We will take into account “selfish” users’ behavior where simultaneously  $D$  percent of copies leaves the overlay network for each request. In this manner we have joined two parameters of the simulation which describe the popularity of a selected content and the migration of overlay users ( $\lambda_{Dc}$  and  $\lambda_{Dm}$ ).

Figure 5-4 and Figure 5-5 show 95% confidence intervals and 25% to 75% quantiles (marked as boxes) surrounding the mean values of DT. We simulated the overlay under dynamically changing traffic conditions:  $D = 10\%$ ,  $D = 20\%$ , and  $D = 30\%$ .

*Traffic performance measures*

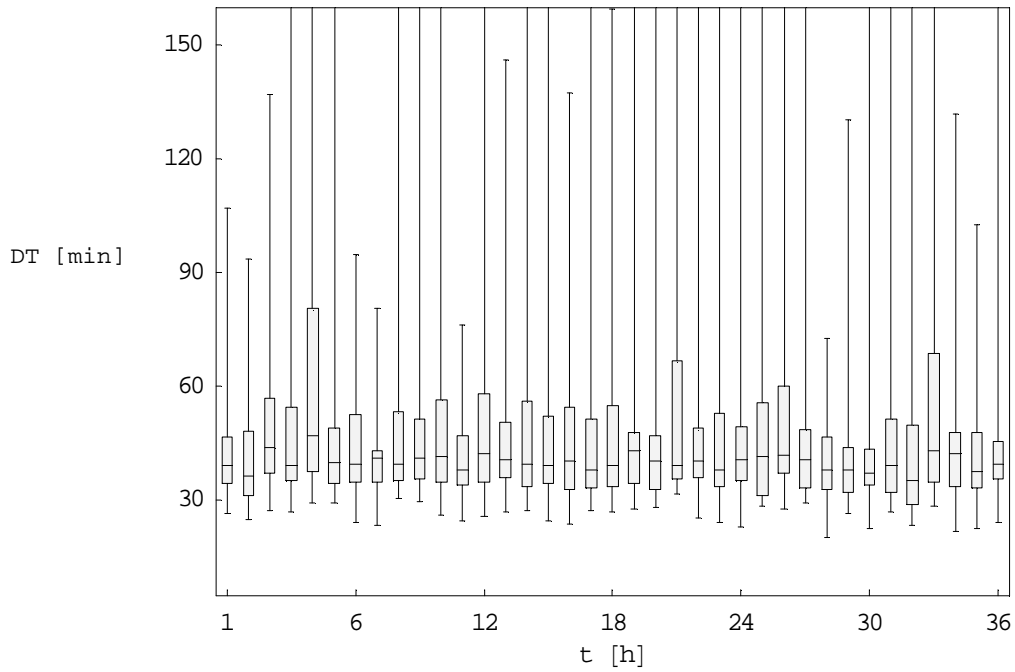


Figure 5-4 Reaction of CROWDS random walk to the new content publication,  $N = 100$ , low request arrival rate,  $D = 20\%$ .

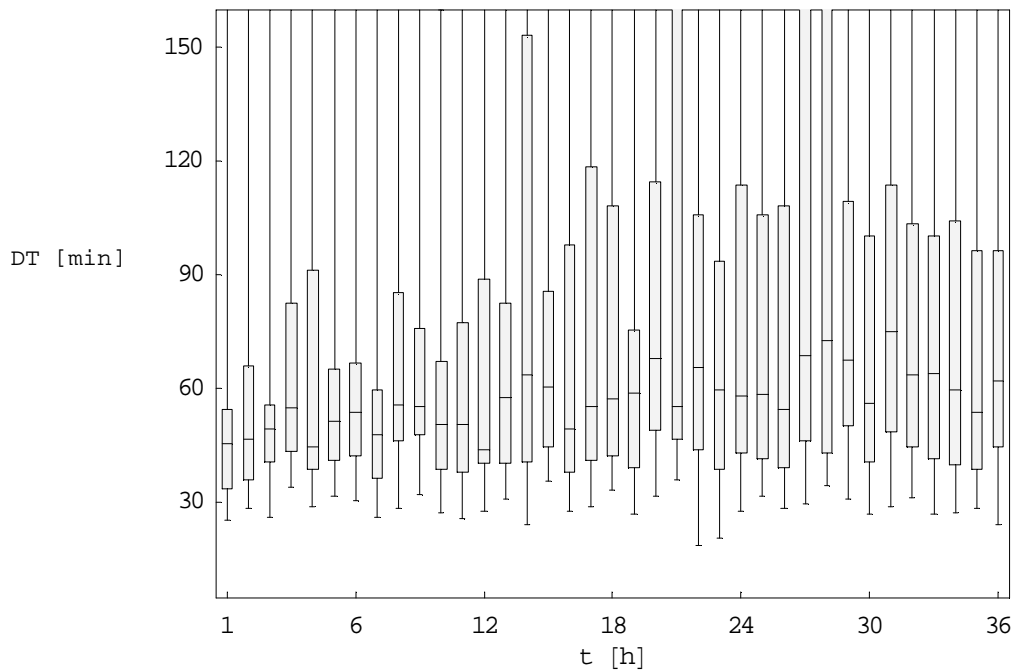


Figure 5-5 Reaction of CROWDS random walk to the new content publication,  $N = 100$ , low request arrival rate,  $D = 30\%$ .

### *Traffic performance measures*

The first presented results (Figure 5-4) were obtained for dynamics  $D = 20\%$ . We can observe that mean download time is slightly longer after a new content publications and returns to an initial level after about 5 hours. The further increase of dynamics to the value of  $D = 30\%$  (presented on Figure 5-5) caused an instability of the system – DT increases in the analyzed period of 36 hours. However, this scenario is highly pessimistic as about  $\frac{1}{3}$  of all user requests in the overlay network are directed towards the same, new content.

#### 5.4 Summary

We have observed that the empirical analysis of the network random walk algorithm, provided by our simulation environment, is analogous to theoretical values, under boundary conditions. The analytical calculations included: available capacity (traffic intensity of the maximum request arrival rate) and the mean download time for a low-loaded network (low request arrival rate).

For the analytically obtained maximum request arrival rate, we have observed that the simulated system works on the brink of stability. For low traffic intensity the mean download time provided by the simulated system is slightly higher than the analytically obtained results that do not deal with delays introduced by network nodes.

The observation of system's behavior under dynamically changing conditions shows that stable operation of network random walk from CROWDS can be retained for dynamics lower or equal to 20%.

# Chapter 6

## Peer-to-Peer direct and anonymous distribution overlay (P2PRIV system)

### 6.1 Overview of P2PRIV

Peer-to-Peer Direct and Anonymous Distribution Overlay (P2PRIV) [63] is an application layer solution for P2P overlay networks assuring a sender's and a receiver's anonymity. The basic novel idea of the solution consists in parallel content transport instead of widespread cascade transmission between chaining nodes. Certainly, anonymity assured by P2PRIV imposes traffic overheads, as in any other system of this type, for example CROWDS. A motivation behind the parallel architecture with direct content transport is the decrease of the service time while preserving a high degree of anonymity.

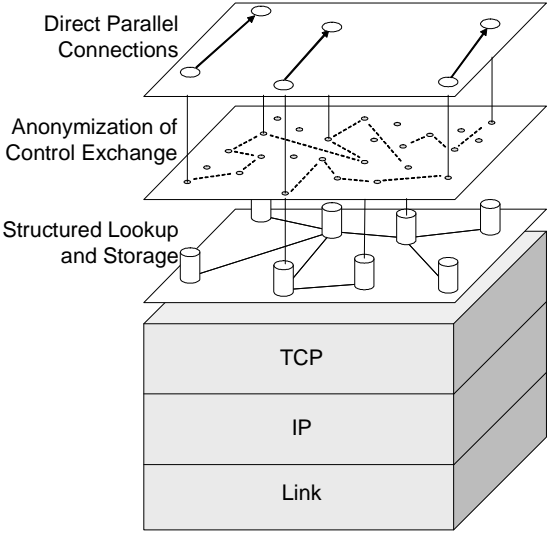


Figure 6-1 P2PRIV architecture.

The architecture of P2PRIV system is presented in Figure 6-1. P2PRIV peers are symmetric and do not include any privileged nodes (supernodes). The P2PRIV utilizes a classical concept of chaining with encryption anonymization, for example Mix-net. It also uses a structured look-up system, which can be based on DHT algorithms (distributed hash table). The classical cascade anonymization mechanism assures anonymity of an entire P2PRIV management communications, including the distributed content look-up process.

We can distinguish two steps of P2PRIV operation (Figure 6-2).

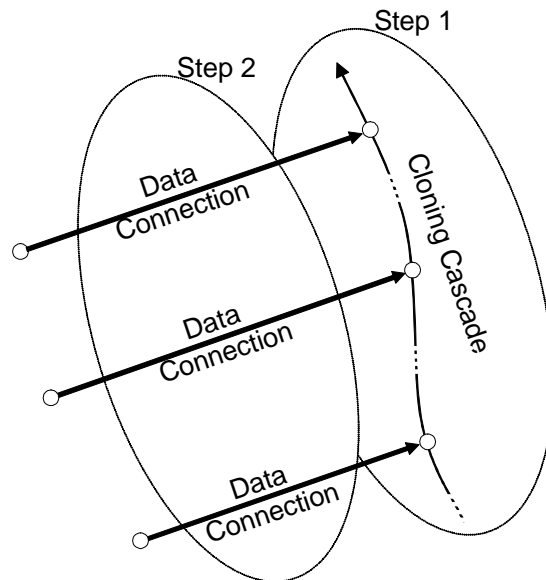


Figure 6-2 P2PRIV parallel transport of content.

**Step 1: Cloning** – exchange of management messages applied for anonymous random selection of a subset of peers referred to as a *cloning cascade* (CC). Each such CC contains the initiator and its clones; each peer can be potentially selected for such a clone. This step is similar to the network random walk mechanism of CROWDS. The initiator sends a file id to a randomly chosen peer. Then, the selected peer flips an asymmetric coin to decide whether to forward such a token (with probability  $p_f$ ) to the next random peer. This communication may be additionally secured and anonymized by Mix-net mechanisms, as

numerous but short management messages of constant length generated by cloning (tokens) can be effectively exchanged by the Mix cascades [28].

**Step 2: Data Connection** – transport of the requested content. After a random interval of time and based on the content id received earlier, copies of the content are directly downloaded by selected (cloned) peers from nodes which store data. One of these nodes is the initiator. Files can be looked-up by the DHT algorithm. Like the cloning communications, look-up messages can be secured by Mix-net mechanisms. The resulting data redundancy (the file is downloaded and stored by each clone) improves content accessibility, because the popularity of content automatically increases the number of its copies. Notice, that in our solution the anonymization process is separated from the content transport process, in contrast to classical schemes.

## 6.2 Design of P2PRIV peer

This section contains a detailed description of P2PRIV system with an explanation of its communications and operation from the point of view of a peer. Notice that the P2PRIV is a pure P2P overlay, hence all P2PRIV peers are symmetric and represent the same functionality (compare with Section 2.2.2). Figure 6-3 shows a finite state machine diagram of the P2PRIV communications.

The basic states of P2PRIV peer are: “LISTEN”, “CLONED” and “DATA CONNECT”. LISTEN represents an idle operation of a peer waiting for requests (marked on the diagram as “recv:”) from other peers or from the user (marked with “user:” symbol). CLONED state is triggered when the peer joins a Cloning Cascade (compare with Step 1 of P2PRIV operation). DATA CONNECT corresponds to the Data Connection step in which P2PRIV node downloads a specified content (Step 2 of P2PRIV operation). The diagram contains also: “COIN FLIP” and “LOOKUP” states. COIN FLIP applies to a decision process based on a binary random selection and LOOKUP initiates a DHT lookup procedure.

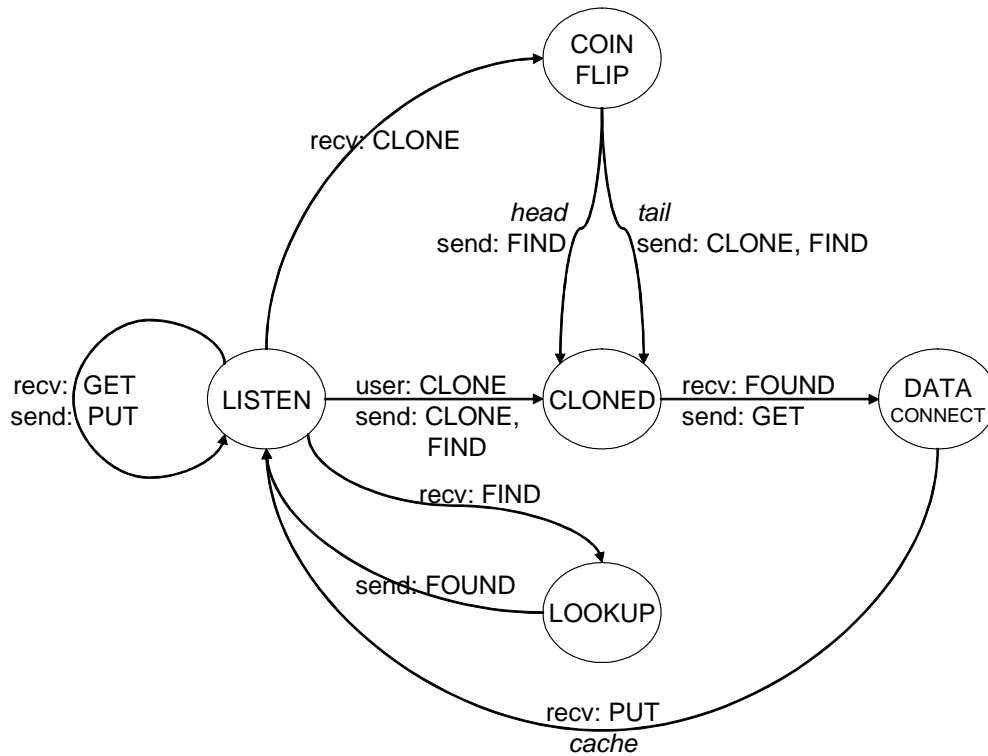


Figure 6-3 State diagram of P2PRIV peer.

The P2PRIV communication is based on five messages: “CLONE”, “FIND”, “FOUND”, “GET” and “PUT”. The following pseudo-code includes a description of their detail roles (Figure 6-4, Figure 6-5, Figure 6-6, and Figure 6-7).

```
(1)  while ∞ do
(2)      User_Request:=false
(3)      if Mix_Receive ({"CLONE", File_Id}, Irrelevant_Return_Addr)
(4)      or
(5)      User_Request := User_Api_Read ({"CLONE", File_Id})
(6)      then Clone_and_Download (File_Id, User_Request)
(7)
(8)      if Mix_Receive ({"FIND", File_Id}, Some_Proxy_Addr)
(9)      then Find_And_Reply (File_Id, Some_Proxy_Addr)
(10)
(11)     if Tcp_Receive ({"GET", File_Id}, Clone_or_Initiator_Addr)
(12)     then Upload (File_Id, Clone_or_Initiator_Addr)
```

Figure 6-4 Pseudo-code description of a P2PRIV single node operation.

The pseudo-code of a single node of P2PRIV overlay is organized as a one infinite loop with three main sections. The first section (starting from the Line 3, Figure 6-4) refers to a CC joining process and a further content download. When a peer receives a CLONE message containing a file id it will check whether the message comes from other peer or from a local user. Then it starts a “Clone and Download” subroutine presented in Figure 6-5. The next section (starting from the Line 8, Figure 6-4) refers to a reaction of a peer to a content look-up request. If a FIND message is received from other P2PRIV node, then a “Find and Reply” subroutine will begin (see Figure 6-6), where a P2PRIV node looks for a specified content on behalf of other anonymous P2PRIV node. The last section (starting from Line 11, Figure 6-4) refers to reaction to an upload request from other node. In this case the node was pointed out by DHT interface as a storage of the specified file. If it occurs and a GET message is received then an “Upload” subroutines will begin (Figure 6-7).

```
(1) subroutine Clone_and_Download (Download_File_Id, Is_Request_from_User)
(2)   CC_Forward:=false
(3)   if (Is_Request_from_User)
(4)     CC_Forward:=true
(5)   else if (Coin_Flip (pf))
(6)     CC_Forward:=true
(7)   if (CC_Forward)
(8)     Mix_Send ({"CLONE", Download_File_Id}, First_Random_Addr)
(9)
(10)    Mix_Send ({"FIND", Download_File_Id}, Second_Random_Addr)
(11)
(12)    while not
(13)      Mix_Receive
(14)        ({"FOUND", Download_File_Id, "@", File_Owner_Addr},
(15)          Second_Random_Addr)
(16)    do Wait
(17)    after (randomTime) Tcp_Send ({"GET", Download_File_Id},
(18)      File_Owner_Addr)
(19)
(20)    while not Tcp_Receive ({"PUT", File}, File_Owner_Addr)
(21)    do Wait
(22)    Cache (File)
(23)    if Is_Request_From_User
(24)      Alert ("Requested file has been anonymously downloaded!")
```

Figure 6-5 Pseudo-code description of the Clone and Download subroutine.

The Clone and Download subroutine describes a process of a CC joining and fulfilling tasks of a clone: a location of a requested content and its download. If the request is originated anonymously (via the Mix-net) from other node, the P2PRIV peer will flip an asymmetric coin to decide whether to forward the message anonymously to the next random peer – a next CC member (Line 5, Figure 6-5). Otherwise, if the request originates from a user, the forwarding process will surely proceed (Line 8, Figure 6-5).

*Peer-to-Peer direct and anonymous distribution overlay (P2PRIV system)*

Next, the P2PRIV node sends a FIND message to a randomly chosen peer via means of a Mix-net. The selected node will perform a DHT look-up procedure and reply with an address of the peer which stores the file. This reply is sending on a Chaumian untraceable return address allowing a preservation of anonymity of this request.

After receiving the reply message (FOUND) with an address of a storage and a file id (Line 13, Figure 6-5), the P2PRIV peer sends a randomly delayed request for the file (GET) directly to the storage node and waits for PUT reply message (Line 17, Figure 6-5). This request can be performed directly as the destination node cannot detect whether this request originates from the real initiation or from one of clones. Finally, the requested file is downloaded and cached (Lines 20-22, Figure 6-5). If the P2PRIV node is the real initiator, then the user will be informed about the successful download (Line 24, Figure 6-5).

```
(1) subroutine Find_and_Reply (Lookup_File_Id, Reply_Dest_Addr)
(2)   if File_Owner_Addr:=Lookup (Key (Lookup_File_Id))
(3)     Mix_Send
(4)       ({"FOUND", Lookup_File_Id, "@", File_Owner_Addr},
(5)         reply_Dest_Addr)
```

Figure 6-6 Pseudo-code description of the Find and Reply subroutine.

As it was described earlier (compare with Clone and Download subroutine), the P2PRIV node can send the FIND message to the other P2PRIV node. The “Find and Reply” subroutine describes a reaction to this request. Here the P2PRIV node becomes an anonymizing proxy, which initiates the DHT look-up procedure for a specified file. Additionally, this proxy is reached by means of Mix-net. In the line 2 (Figure 6-6) a DHT look-up procedure is executed and a result is sent back to an initiator on its untraceable address provided by means of Mix-net.

```
(1) subroutine upload (upload_File_Id, upload_Dest_Addr)
(2)    Tcp_Send ({ "PUT", Upload_File_Id}, Upload_Dest_Addr)
```

Figure 6-7 Pseudo-code description of the Upload subroutine.

The last subroutine “Upload” (Figure 6-7) contains a short reaction to GET request (see Clone and Download subroutine where GET requests are initiated). In this case, the P2PRIV node replies with PUT message containing the requested content.

Other included subroutines refer to the following aspects of P2PRIV operations.

- TCP/IP communications is simplified by:

```
bool Tcp_Receive(data,      – Receiving data via a TCP/IP socket from a srcAddr
srcAddr)                address;
```

```
bool Tcp_Send(data,      – Sending data via a TCP/IP socket to a destAddr
destAddr)                address.
```

- Mix-net communications is simplified by:

```
bool Mix_Receive(data,    – Receiving data via a Mix-net from a Chaumian
untrcReturnAddr)        untraceable return address untrcReturnAddr;
```

```
bool Mix_Send(data,      – Sending data via a Mix-net to destAddr address.
destAddr)
```

- DHT look-up interface access is described by:

```
peerAddr Lookup(hash)    – Looking-up network node peerAddr which stores a
                        copy of a file described by DHT key hash;
```

*Peer-to-Peer direct and anonymous distribution overlay (P2PRIV system)*

`dhtKey key(fileId)` – Computing DHT key `dhtKey` from a file name `fileId`.

Certainly, Mix-net and DHT communications utilize TCP/IP stack, yet it is not exposed in the pseudo-code descriptions for a purpose of an algorithm explanation transparency.

- Other non-network subroutines are:

`bool User_Api_Read(data)` – Receiving `data` from a user interface;

`bool Coin_Flip(p)` – Binary random selection with a `p` probability;

`bool Alert(data)` – Sending `data` to a user interface;

`bool Cache(file)` – Save `file` to a local file system.

An example of a single peer Clone and Download operation is showed in Figure 6-8 marked with continuous lines. Additionally, the figure shows the Find and Reply and the Upload operation originated from other peer (marked with broken lines).

*Peer-to-Peer direct and anonymous distribution overlay (P2PRIV system)*

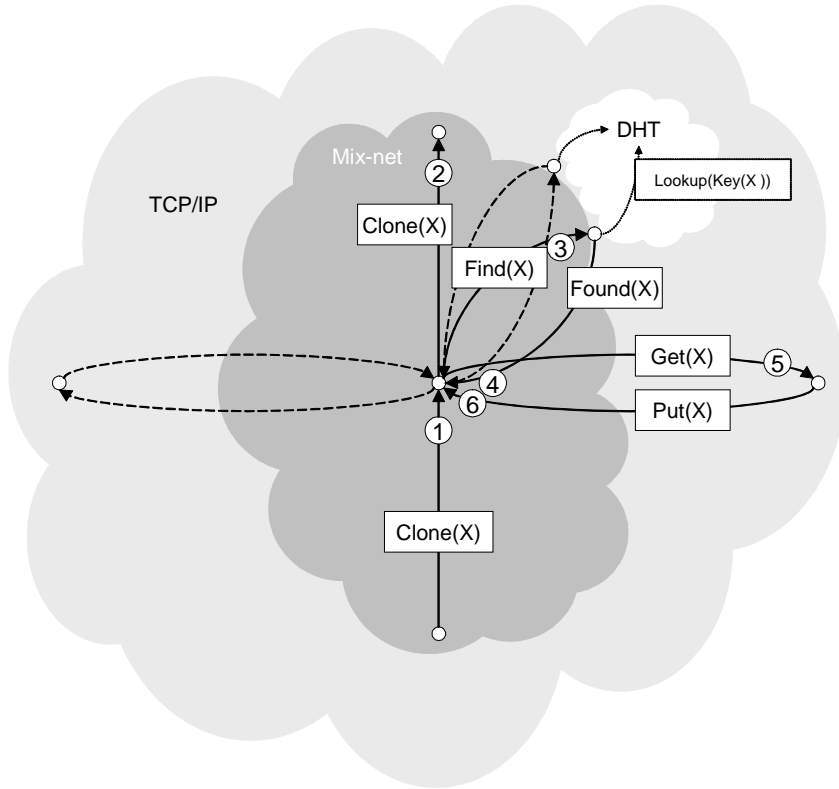


Figure 6-8 Simplified diagram of P2PRIV Clone and Download.

# Chapter 7

## Anonymity analysis for P2PRIV

The anonymity analysis will cover unlinkability of the request initiator and the requested content. We will evaluate how difficult for the adversary is to find a network node from which the request for a specific content was initiated. The goal of the P2PRIV system is an anonymous content transport and it does not include functions of a content publication. Since the anonymous publication process, that can be performed in many different ways using the state of the art methods (*Achord* [49], *Eternity* ([1], [5]), *TAZ* [44], *Free Heaven* [33], *Freenet* [17], *Infranet* ([37], [38]), *Publius* [101], and *Tangler* [100]), is not considered in the proposed scheme, we also omitted it in the analysis. We will not analyze sender anonymity as well. However, it should be remained that P2PRIV utilizes the DHT storage (compare with Section 2.2.2) and peers of P2PRIV are involved in the process of storing and sending data independently from decisions of the users.

### 7.1 Quantitative analysis of P2PRIV secure working point

The goal of this section is to describe in details the anonymity assured by P2PRIV. We will analyze the impact of  $p_f$  configuration on the system entropy. This analysis will cover small networks ( $N = 50$  nodes), where it seems to be more difficult to hide a real initiator and large networks ( $N = 1000$  nodes) – more realistic for open P2P overlays. Compliantly to earlier analysis (Chapter 4) we will focus on three variants of network collaboration levels  $C = 5\%$ ,  $C = 10\%$ , and  $C = 20\%$ .

#### 7.1.1 Passive-static attacks

Static attackers cannot predict which nodes will be randomly selected to form the CC for a particular request anonymization. However, the adversary can distinguish two sets of peers  $\{S_1, S_2\}$  among all  $N$  nodes and assign their members probabilities of being the

initiator  $\{p_1, p_2\}$ .  $S_1$  consists of peers which communicate directly with collaborating nodes  $C$  during the transport of the requested data, and  $S_2$  are remaining suspected nodes. The average CC length (as in Equation 3-1) is

$$P = \sum_{i=2}^{\infty} i p_f^{i-2} (1 - p_f) = \frac{p_f - 2}{p_f - 1} . \quad (7-1)$$

The adversary concludes that the initiator is not among collaborating peers. An average number of honest nodes from the cascade

$$n = P - \frac{C}{N} P = \frac{(p_f - 2)(N - C)}{(p_f - 1)N} . \quad (7-2)$$

The step 1 of P2PRIV operation is anonymized by Mix-net. However, in the step 2, the cloned peers communicate directly with other peers. Let us consider the most pessimistic scenario, where all of the clones download the content from different peers. Then the average number of honest nodes from the CC, that communicate with collaborating nodes, equals

$$S_1 = \frac{C}{N} n = \frac{C(p_f - 2)(N - C)}{(p_f - 1)N^2} . \quad (7-3)$$

Each of them can be the request initiator with probability

$$p_1 = \left( P - \frac{C}{N} P \right)^{-1} . \quad (7-4)$$

The attacker, who can observe  $S_1$  nodes involved in a transport of the requested data, should also consider that none of them is the initiator. He/she should also take into account the rest of the nodes

$$S_2 = N - C - S_1 , \quad (7-5)$$

Each of the  $S_2$  members can be the initiator with probability

$$p_2 = \frac{1 - S_1 p_1}{S_2} . \quad (7-6)$$

According to the information theory [91], information entropy of P2PRIV for this scenario will be

Anonymity analysis for P2PRIV

$$H_{psP2PRIV} = -\sum_{i=1}^N p_i \log_2(p_i) = -S_1 p_1 \log_2(p_1) - S_2 p_2 \log_2(p_2) \quad (7-7)$$

$$H_{psP2PRIV} = \begin{cases} 0 & p_1 = 1 \vee p_2 = 1 \\ \frac{C}{N} \log_2(p_1^{-1}) & p_2 = 0 \\ \left(1 - \frac{C}{N}\right) \log_2(p_2^{-1}) & p_1 = 0 \\ \frac{C}{N} \log_2(p_1^{-1}) + \left(1 - \frac{C}{N}\right) \log_2(p_2^{-1}) & p_1 \in (0,1) \wedge p_2 \in (0,1), \end{cases} \quad (7-8)$$

where

$$p_1 = \frac{(p_f - 1)N}{(p_f - 2)(N - C)}, \quad p_2 = \frac{(p_f - 1)N}{(p_f - 1)N^2 - (p_f - 2)C}.$$

Figure 7-1 and Figure 7-2 show the entropy  $H$  of small and large P2PRIV overlays as a function of the parameter  $p_f$ .

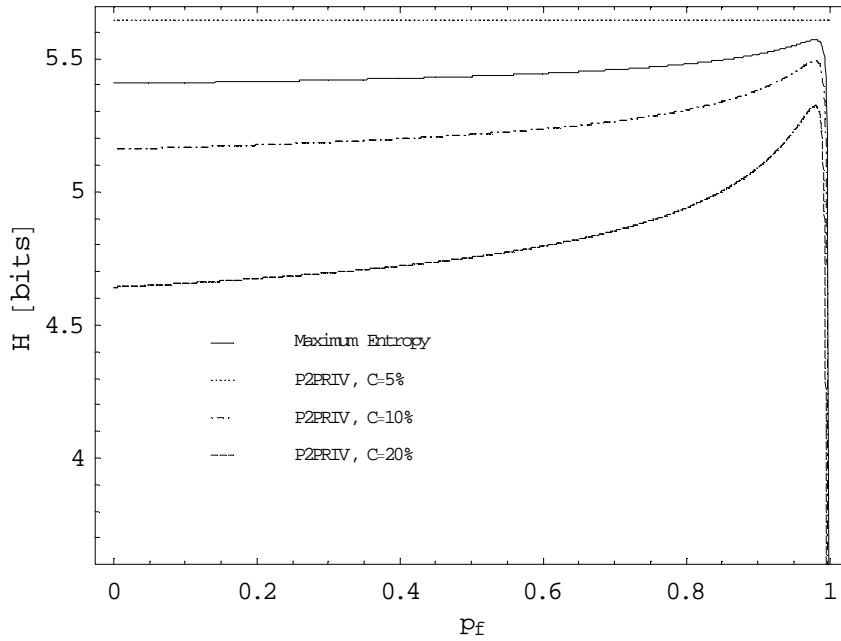


Figure 7-1 Entropy of P2PRIV, passive-static attacks,  $N = 50$ .

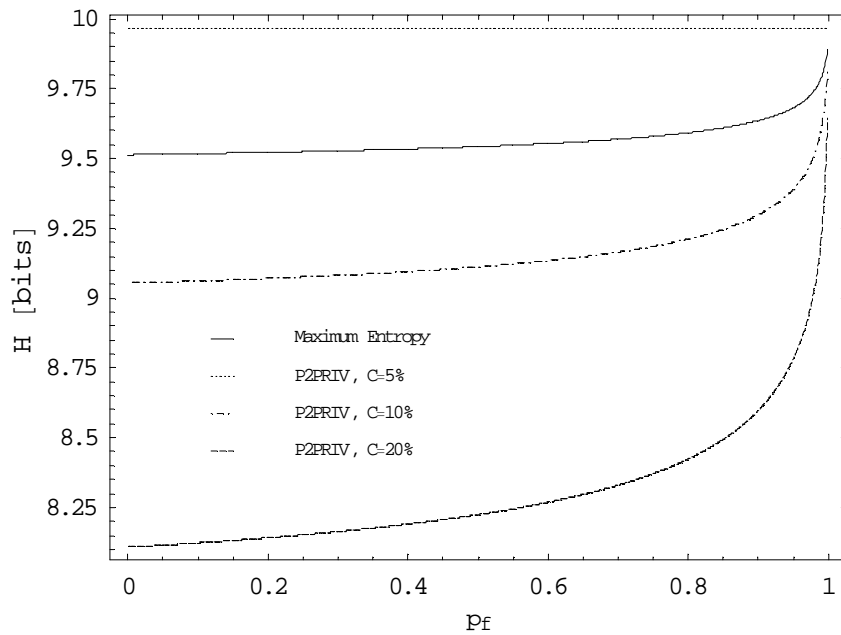


Figure 7-2 Entropy of P2PRIV, passive-static attacks,  $N = 1000$ .

In both cases  $p_f$  increment is in favor of the entropy of the P2PRIV system. However, in a small overlay network (Figure 7-1,  $N = 50$  nodes) high  $p_f$  values (close to 1) eventually degrade the entropy. For  $p_f = 0.9$  the cascade mean length reaches value of 6 nodes. This is a relatively large value for small networks. It is significant that in small networks it is easier for the adversary to become a member of CC and degrade P2PRIV entropy (from above 5 bits to even 0). Therefore, when we consider small overlays,  $p_f$  configuration should not be higher than about 0.9.

### 7.1.2 Passive-adaptive attacks

The previous passive-static attacks scenario corresponds to a realistic assumption that the adversary cannot predict which nodes will anonymize the request. The adaptive scenario is more pessimistic and considers implications of the presence of colluding nodes in the system area where active anonymization process occurs. If we assume that the collaborating peer belongs to a chosen CC then the average number of honest nodes communicating with collaborating nodes will be

Anonymity analysis for P2PRIV

$$S_{1pa} = \frac{C}{N}(n-1) + 1 \quad (7-9)$$

$$S_{1pa} = \frac{(N-C)((p_f-2)C + (p_f-1)N)}{(p_f-1)N^2} . \quad (7-10)$$

Analogously to (7-5) the average number of remaining nodes is

$$S_{2pa} = N - C - S_{1pa} , \quad (7-11)$$

with assigned probability of being the initiator

$$p_{2pa} = \frac{1 - S_{1pa} p_1}{S_{2pa}} . \quad (7-12)$$

Entropy of P2PRIV in this attack scenario is

$$H_{paP2PRIV} = -S_{1pa} p_1 \log_2(p_1) - S_{2pa} p_{2pa} \log_2(p_{2pa}) \quad (7-13)$$

$$H_{paP2PRIV} = \begin{cases} 0 & p_1 = 1 \vee p_{2pa} = 1 \\ \vartheta \log_2(p_1^{-1}) & p_{2pa} = 0 \\ \zeta \log_2(p_{2pa}^{-1}) & p_1 = 0 \\ \vartheta \log_2(p_1^{-1}) + \zeta \log_2(p_{2pa}^{-1}) & p_1 \in (0,1) \wedge p_{2pa} \in (0,1), \end{cases} \quad (7-14)$$

where

$$p_{2pa} = \frac{(p_f-1)N(N + (p_f-2)C)}{(p_f-2)(C-N)(C(p_f-2) + N(N - p_f N + p_f - 1))}$$

$$\vartheta = \frac{(p_f-2)C + (p_f-1)N}{(p_f-2)N}, \zeta = \frac{(p_f-2)C + N}{(p_f-2)N} .$$

Figure 7-3 and Figure 7-4 show the entropy  $H$  of small and large P2PRIV overlays as a function of the parameter  $p_f$ .

*Anonymity analysis for P2PRIV*

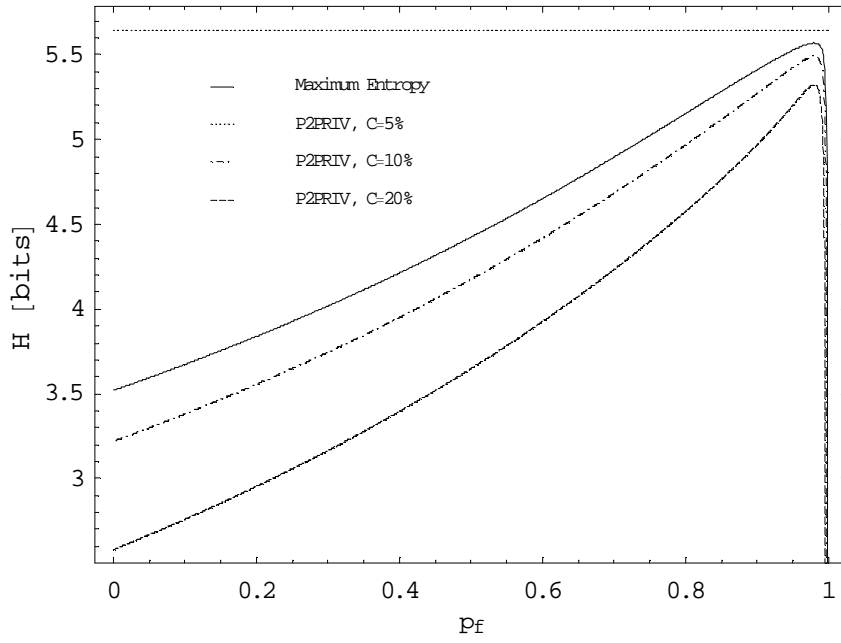


Figure 7-3 Entropy of P2PRIV, passive-adaptive attacks,  $N = 50$ .

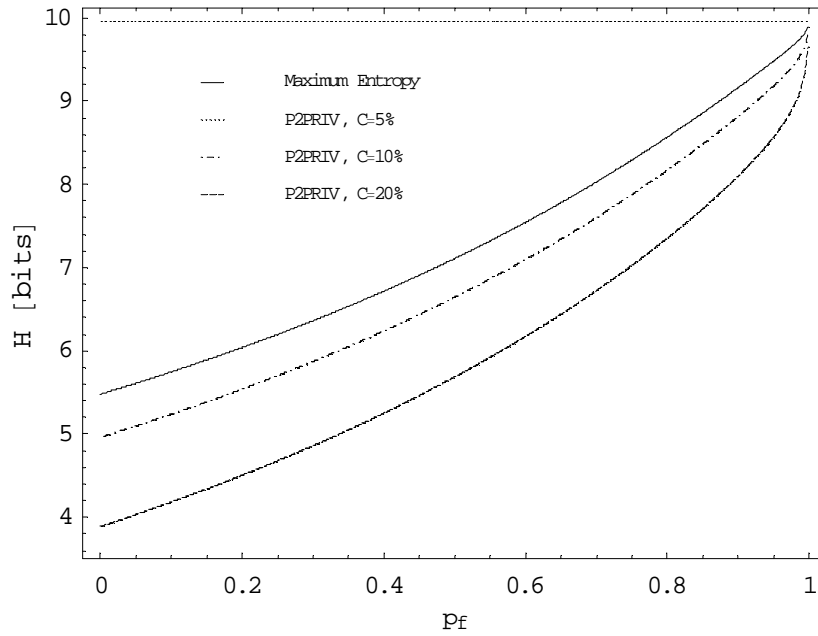


Figure 7-4 Entropy of P2PRIV, passive-adaptive attacks,  $N = 1000$ .

Similarly to the previous scenario of passive-static attacks we can observe that large values of  $p_f$  increase the entropy of P2PRIV. In small networks, values close to maximum finally negatively impact the system entropy. This similarity shows a P2PRIV robustness against the specific character of the static attack (compare with Section 4.4), because the results for adaptive attacks are practically the same as in the static scenario in the range of large  $p_f$  values. We can conclude that the P2PRIV system provides a high protection against a penetration of active sets.

The analysis also shows that the adaptive attacks reveal a higher importance of a proper  $p_f$  configuration. Contrary to the static attacks, low  $p_f$  value significantly impacts the system entropy and for large networks gives the adversary about 5 bits (from a total number of 10 bits) of information about the origin of a specified request.

### **7.1.3 Active attacks**

Now we analyze static and adaptive behaviors of an active adversary. In classical anonymous chaining systems active attacks must be more subtle and complex ([47], [75], [77]) than they could be in the proposed parallel architecture, since simple breaking of the chain transporting user data would be detected quickly. Breaking the cloning cascade in P2PRIV system does not affect user data delivery. Therefore, when we consider the proposed parallel transport architecture, it is important to take into account the extreme scenario of cloning interception.

The length of CC before the first collaborating node interception  $P_a$  is given by the following formula (7-15).

$$\begin{aligned}
 p(P_a = 1) &= \frac{C}{N} \\
 p(P_a = 2) &= \left(1 - \frac{C}{N}\right) p_f \frac{C}{N} + \left(1 - \frac{C}{N}\right) (1 - p_f) \\
 p(P_a = n) &= \left(1 - \frac{C}{N}\right)^{n-1} p_f^{n-1} \frac{C}{N} + \left(1 - \frac{C}{N}\right)^{n-1} p_f^{n-2} (1 - p_f) \\
 P_a &= \frac{C}{N} + \left(1 - \frac{C}{N}\right) p_f \frac{C}{N} + \left(1 - \frac{C}{N}\right) (1 - p_f) + \\
 &+ \sum_{i=3}^{\infty} i \left( \left(1 - \frac{C}{N}\right)^{i-1} p_f^{i-1} \frac{C}{N} + \left(1 - \frac{C}{N}\right)^{i-1} p_f^{i-2} (1 - p_f) \right) \\
 P_a &= \frac{N^3 + p_f^2 (C - N)^3 + p_f N (2C^2 - 3CN + N^2)}{(p_f (C - N) + N) N^2}.
 \end{aligned} \tag{7-15}$$

Among  $P_a$  the average number of nodes communicating directly with collaborating nodes is

$$n_a = \frac{(N - C) (N^3 + p_f^2 (C - N)^3 + p_f N (2C^2 - 3NC + N^2))}{(N + p_f (C - N)) N^3}, \tag{7-16}$$

hence members of

$$S_{1as} = \frac{C}{N} n_a, \quad S_{1aa} = \frac{C}{N} (n_a - 1) + 1 \tag{7-17}$$

will have assigned probabilities

$$p_{1a} = \left( P_a - \frac{C}{N} P_a \right)^{-1}. \tag{7-18}$$

And the rest of nodes respectively to the static/adaptive attack

$$S_{2as} = N - C - S_{1as}, \quad S_{2aa} = N - C - S_{1aa} \tag{7-19}$$

will be considered by the adversary as the initiator with probabilities

$$p_{2as} = \frac{1 - S_{1as} p_{1a}}{S_{2as}}, \quad p_{2aa} = \frac{1 - S_{1aa} p_{1a}}{S_{2aa}}. \tag{7-20}$$

Finally, the entropies of P2PRIV for active attacks are

$$\begin{aligned}
 H_{asP2PRIV} &= \begin{cases} 0 & p_{1a} = 1 \vee p_{2as} = 1 \\ \frac{C}{N} \log_2(p_{1a}^{-1}) & p_{2as} = 0 \\ \left(1 - \frac{C}{N}\right) \log_2(p_{2as}^{-1}) & p_{1a} = 0 \\ \frac{C}{N} \log_2(p_{1a}^{-1}) + \left(1 - \frac{C}{N}\right) \log_2(p_{2as}^{-1}) & p_{1a} \in (0,1) \wedge p_{2as} \in (0,1) \end{cases} & (7-21) \\
 H_{aaP2PRIV} &= \begin{cases} 0 & p_{1a} = 1 \vee p_{2aa} = 1 \\ \psi \log_2(p_{1a}^{-1}) & p_{2aa} = 0 \\ \zeta \log_2(p_{2aa}^{-1}) & p_{1a} = 0 \\ \psi \log_2(p_{1a}^{-1}) + \zeta \log_2(p_{2aa}^{-1}) & p_{1a} \in (0,1) \wedge p_{2aa} \in (0,1), \end{cases}
 \end{aligned}$$

where

$$\begin{aligned}
 p_{1a} &= \frac{N^3(p_f(C-N)+N)}{(N-C)(p_f^2(C-N)^3+N^3+p_fN(2C^2-3NC+N^2))} \\
 p_{2as} &= \frac{N^3(N+p_f(C-N))}{p_f^2C(C-N)^3+N^3(C-N^2)+p_fN(2C^3-3C^2N-(N-1)CN^2+N^4)} \\
 p_{2aa} &= \zeta \left[ N - C + \frac{C}{N} \left( \frac{1+(C-N)(N^3+p_fN(2C^2-3NC+N^2)+p_f^2(C-N)^3)}{N^3(N+p_f(C-N))} \right) \right]^{-1} \\
 \psi &= \frac{N^3(N+C)-p_fN(N^3-2C^3+3NC^2-2N^2C)+p_f^2C(C-N)^3}{(N^3+p_fN(2C^2-3CN+N^2)+p_f^2(C-N)^3)N} \\
 \zeta &= \frac{p_fN(2N^3-2C^3+5C^2N-5CN^2)-CN^3-p_f^2(C-N)^4}{(N^3+p_fN(2C^2-3CN+N^2)+p_f^2(C-N)^3)N}.
 \end{aligned}$$

Figure 7-5 and Figure 7-6 show the entropies for P2PRIV static attacks. Characteristics of adaptive scenarios are shown in Figure 7-7 and Figure 7-8.

Anonymity analysis for P2PRIV

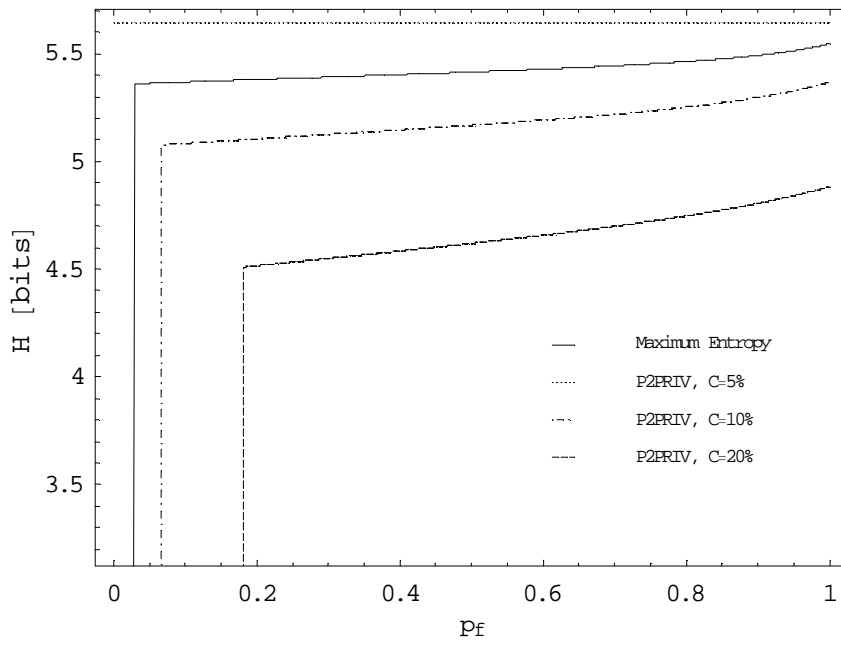


Figure 7-5 Entropy of P2PRIV, active-static attacks,  $N = 50$ .

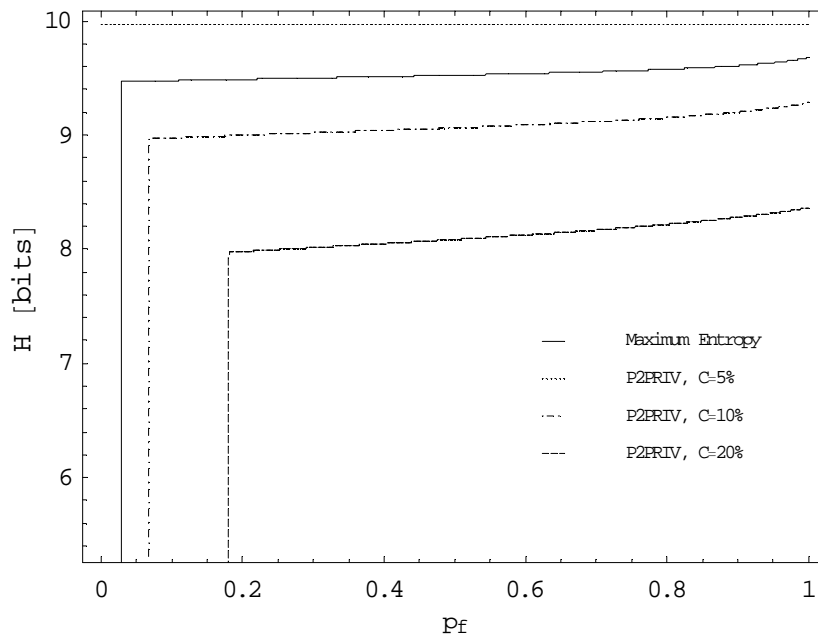


Figure 7-6 Entropy of P2PRIV, active-static attacks,  $N = 1000$ .

Anonymity analysis for P2PRIV

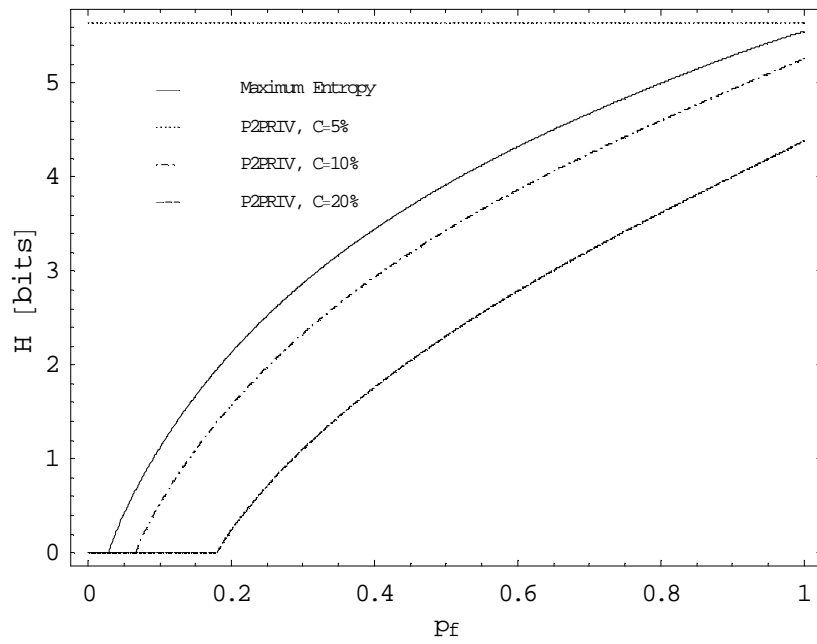


Figure 7-7 Entropy of P2PRIV, active-adaptive attacks,  $N = 50$ .

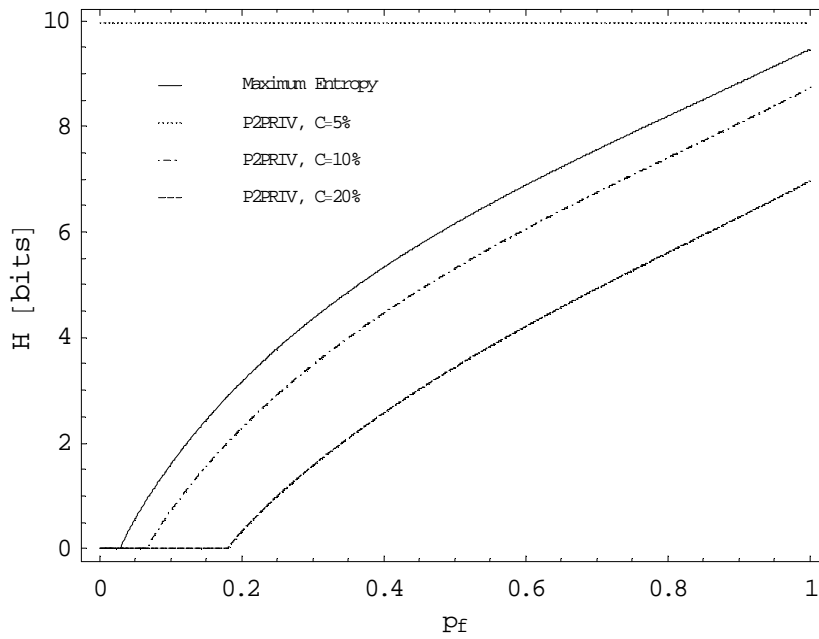


Figure 7-8 Entropy of P2PRIV, active-adaptive attacks,  $N = 1000$ .

Taking into account far reaching possibilities of active adversary, which is able to imperceptibly break the cloning cascade of P2PRIV, we have observed that P2PRIV can still assure proper level of anonymity. In configuration of high  $p_f$  values anonymity of P2PRIV is still close to maximum. However, these scenarios show that secure configuration of the system should cover  $p_f$  values higher than 0.5. For very low  $p_f$  values P2PRIV entropy against active-adaptive adversary reaches 0 values, which means that the initiator of the request becomes completely exposed.

#### **7.1.4 Summary**

Based on realistic assumptions of an adversary capabilities corresponding to the environment of public peer-to-peer overlays, we have observed that the P2PRIV system assures high entropy for its users. Passive attacks reveal that P2PRIV is robust against the specific character of a static attack and that the effective penetration of a proper cloning cascade is difficult to carry out for the adversary. The analysis of active attacks strengthened our pronouncement, obtained by the analysis of passive-adaptive attacks, that P2PRIV  $p_f$  configuration should not cover low values (below 0.5) as in this configuration entropy of the system is close to a minimum. Additionally, we have observed that in small overlay networks configuration of  $p_f$  close to 0.9 also negatively impacts the entropy. The system is dedicated to public and wide usage; however it should be also be able to work under temporal conditions of small number of users.

Table 7-1 contains a summary of the results obtained throughout analyzed scenarios. For the purpose of results' legibility, we have calculated the effective anonymity set size (compare with Section 4.2) of P2PRIV system for characteristic and permissible values of  $p_f$  parameter. Here, we can observe how numerous would be the perfectly homogeneous "crowd" of peers surrounding and hiding the initiator of the request for networks of 50 and 1000 nodes.

*Anonymity analysis for P2PRIV*

TABLE 7-1  
EFFECTIVE ANONYMITY SET SIZE FOR P2PRIV.

<i>C</i>	Attack	<i>N</i> = 50				<i>N</i> = 1000			
		<i>P<sub>f</sub></i>				<i>P<sub>f</sub></i>			
		0.5	0.66	0.75	0.8	0.5	0.66	0.75	0.8
5%	Passive-Static	43	44	44	45	750	760	770	770
	Passive-Adaptive	21	28	33	36	140	230	310	380
	Active-Static	43	43	44	44	740	750	760	760
	Active-Adaptive	15	23	29	32	71	160	240	290
10%	Passive-Static	37	38	39	40	550	570	580	590
	Passive-Adaptive	18	24	28	31	100	170	240	290
	Active-Static	36	37	38	38	540	550	560	570
	Active-Adaptive	11	17	22	24	40	89	140	170
20%	Passive-Static	27	28	30	31	300	320	330	340
	Passive-Adaptive	13	17	21	24	51	90	130	160
	Active-Static	25	26	26	27	270	280	290	300
	Active-Adaptive	5	8.3	11	12	11	25	39	49

## 7.2 Qualitative analysis

Below we will analyze the degree of anonymity (compare with Section 4.3) of P2PRIV using the entropy measurement model ([32], [88]) and compare these results with the degree of anonymity of the classical cascade architecture. The CROWDS system, which combines anonymity and performance with simplicity and reputability, will be used as the reference.

### 7.2.1 Passive-static attacks

The P2PRIV maximum entropy is reached when all honest nodes are equiprobably recognized by the adversary as the initiator

$$H_{\max P2PRIV} = \log_2(N - C) , \quad (7-22)$$

then based on Equation 4-5 and Equation 7-8 the degree of the anonymity (normalized entropy) provided by the P2PRIV system equals

$$d_{psP2PRIV} = \frac{H_{psP2PRIV}}{H_{\max P2PRIV}} , \quad (7-23)$$

after substitution

$$d_{psP2PRIV} = \begin{cases} 0 & p_1 = 1 \vee p_2 = 1 \\ \frac{\frac{C}{N} \log_2(p_1^{-1})}{\log_2(N - C)} & p_2 = 0 \\ \frac{\left(1 - \frac{C}{N}\right) \log_2(p_2^{-1})}{\log_2(N - C)} & p_1 = 0 \\ \frac{\frac{C}{N} \log_2(p_1^{-1}) + \left(1 - \frac{C}{N}\right) \log_2(p_2^{-1})}{\log_2(N - C)} & p_1 \in (0,1) \wedge p_2 \in (0,1), \end{cases} \quad (7-24)$$

where

$$p_1 = \frac{(p_f - 1)N}{(p_f - 2)(N - C)} , \quad p_2 = \frac{(p_f - 1)N}{(p_f - 1)N^2 - (p_f - 2)C} .$$

The degree of anonymity of the CROWDS system for passive-static attacks  $d_{psCROWDS}$  is given by Equation 4-15.

### Anonymity analysis for P2PRIV

Figure 7-9 shows the degrees of anonymity of P2PRIV and CROWDS systems as a function of the parameter  $C$  (the number of colluding nodes).

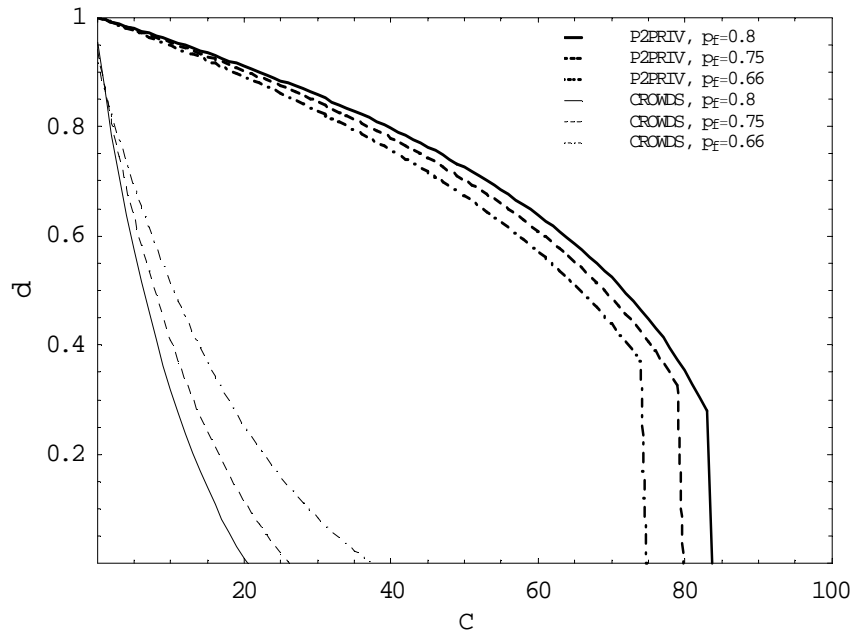


Figure 7-9 Degree of anonymity for P2PRIV and CROWDS, passive-static attacks,  $N = 100$ .

### 7.2.2 Passive-adaptive attacks

Using Equation 4-5 and Equation 7-14 the degree of anonymity of P2PRIV in this attack scenario equals

$$d_{paP2PRIV} = \frac{H_{paP2PRIV}}{H_{\max P2PRIV}} \quad (7-25)$$

Anonymity analysis for P2PRIV

$$d_{paP2PRIV} = \begin{cases} 0 & p_1 = 1 \vee p_{2pa} = 1 \\ \frac{\vartheta \log_2(p_1^{-1})}{\log_2(N-C)} & p_{2pa} = 0 \\ \frac{\zeta \log_2(p_{2pa}^{-1})}{\log_2(N-C)} & p_1 = 0 \\ \frac{\vartheta \log_2(p_1^{-1}) + \zeta \log_2(p_{2pa}^{-1})}{\log_2(N-C)} & p_1 \in (0,1) \wedge p_{2pa} \in (0,1), \end{cases} \quad (7-26)$$

where

$$p_{2pa} = \frac{(p_f - 1)N(N + (p_f - 2)C)}{(p_f - 2)(C - N)(C(p_f - 2) + N(N - p_f N + p_f - 1))}$$

$$\vartheta = \frac{(p_f - 2)C + (p_f - 1)N}{(p_f - 2)N}, \zeta = \frac{(p_f - 2)C + N}{(p_f - 2)N}.$$

In comparison, the degree of anonymity of the CROWDS system in this scenario  $d_{paCROWDS}$  is given by Equations 4-11. Figure 7-10 shows the results for adaptive observation.

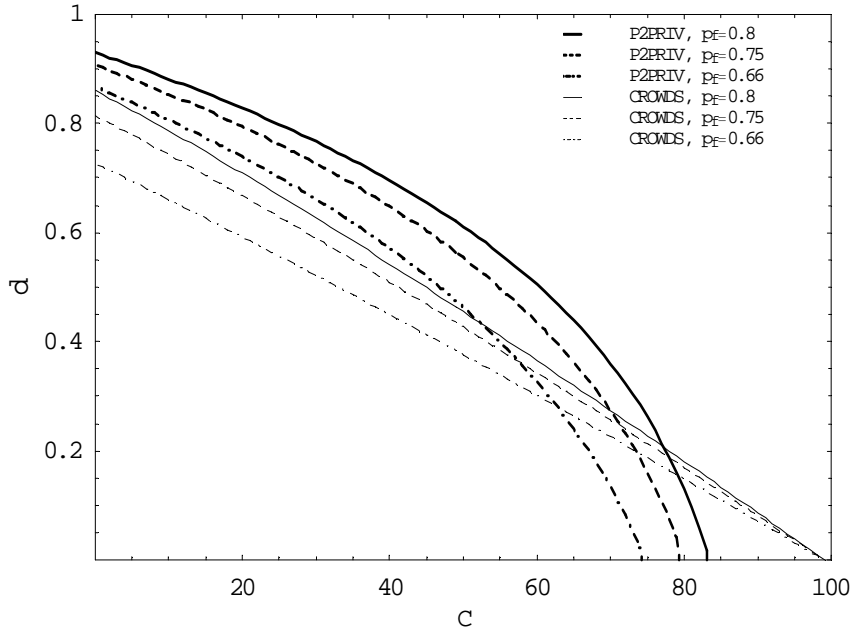


Figure 7-10 Degree of anonymity for P2PRIV and CROWDS, passive-adaptive attacks,  $N = 100$ .

### 7.2.3 Active attacks

Using Equation 7-21 and with the entropy measurement model proposed in ([32], [88]), the degrees of anonymity of P2PRIV for active attacks are

$$d_{asP2PRIV} = \frac{H_{asP2PRIV}}{H_{\max P2PRIV}} \text{ and } d_{aaP2PRIV} = \frac{H_{aaP2PRIV}}{H_{\max P2PRIV}}, \quad (7-27)$$

after substitution

$$d_{asP2PRIV} = \begin{cases} 0 & p_{1a} = 1 \vee p_{2as} = 1 \\ \frac{\frac{C}{N} \log_2(p_{1a}^{-1})}{\log_2(N-C)} & p_{2as} = 0 \\ \frac{\left(1 - \frac{C}{N}\right) \log_2(p_{2as}^{-1})}{\log_2(N-C)} & p_{1a} = 0 \\ \frac{\frac{C}{N} \log_2(p_{1a}^{-1}) + \left(1 - \frac{C}{N}\right) \log_2(p_{2as}^{-1})}{\log_2(N-C)} & p_{1a} \in (0,1) \wedge p_{2as} \in (0,1) \end{cases} \quad (7-28)$$

$$d_{aaP2PRIV} = \begin{cases} 0 & p_{1a} = 1 \vee p_{2aa} = 1 \\ \frac{\psi \log_2(p_{1a}^{-1})}{\log_2(N-C)} & p_{2aa} = 0 \\ \frac{\zeta \log_2(p_{2aa}^{-1})}{\log_2(N-C)} & p_{1a} = 0 \\ \frac{\psi \log_2(p_{1a}^{-1}) + \zeta \log_2(p_{2aa}^{-1})}{\log_2(N-C)} & p_{1a} \in (0,1) \wedge p_{2aa} \in (0,1), \end{cases}$$

where

$$p_{1a} = \frac{N^3(p_f(C-N) + N)}{(N-C)(p_f^2(C-N)^3 + N^3 + p_f N(2C^2 - 3NC + N^2))}$$

$$p_{2as} = \frac{N^3(N + p_f(C-N))}{p_f^2 C(C-N)^3 + N^3(C-N^2) + p_f N(2C^3 - 3C^2 N - (N-1)CN^2 + N^4)}$$

$$p_{2aa} = \zeta \left[ N - C + \frac{C}{N} \left( \frac{1 + (C-N)(N^3 + p_f N(2C^2 - 3NC + N^2) + p_f^2(C-N)^3)}{N^3(N + p_f(C-N))} \right) \right]^{-1}$$

$$\psi = \frac{N^3(N+C) - p_f N(N^3 - 2C^3 + 3NC^2 - 2N^2C) + p_f^2 C(C-N)^3}{(N^3 + p_f N(2C^2 - 3CN + N^2) + p_f^2(C-N)^3)N}$$

$$\zeta = \frac{p_f N(2N^3 - 2C^3 + 5C^2 N - 5CN^2) - CN^3 - p_f^2(C-N)^4}{(N^3 + p_f N(2C^2 - 3CN + N^2) + p_f^2(C-N)^3)N}$$

### Anonymity analysis for P2PRIV

Figure 7-11 and Figure 7-12 show the degree of anonymity for P2PRIV active attacks. Notice that we cannot directly compare preceding results with CROWDS because of different attack possibilities for cascade (Section 3.2) and parallel systems (Section 7.1.3). As in previous scenarios, the results show good resistance of P2PRIV for a realistic percentage of colluding nodes.

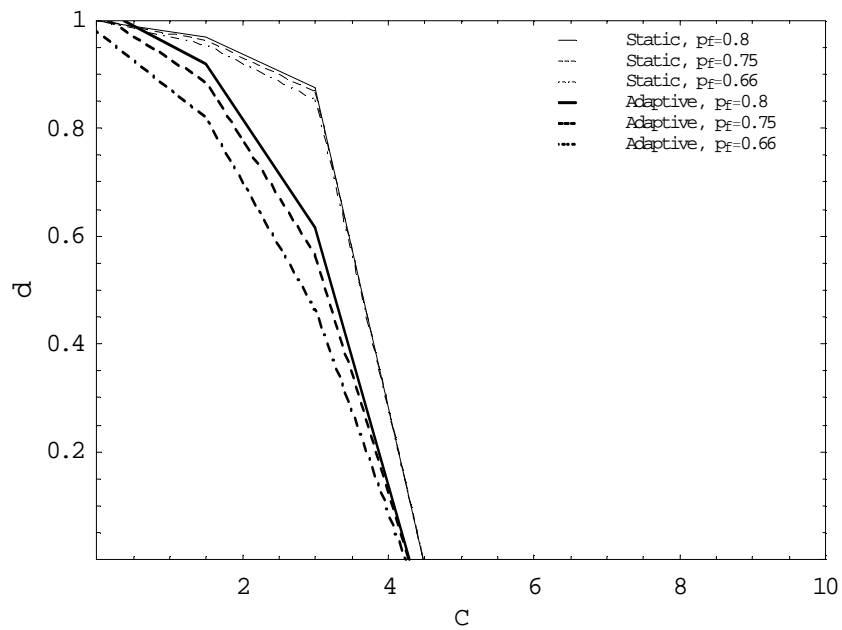


Figure 7-11 Degree of anonymity for P2PRIV, CC interception active-static and active-adaptive attacks,  $N = 10$ .

### Anonymity analysis for P2PRIV

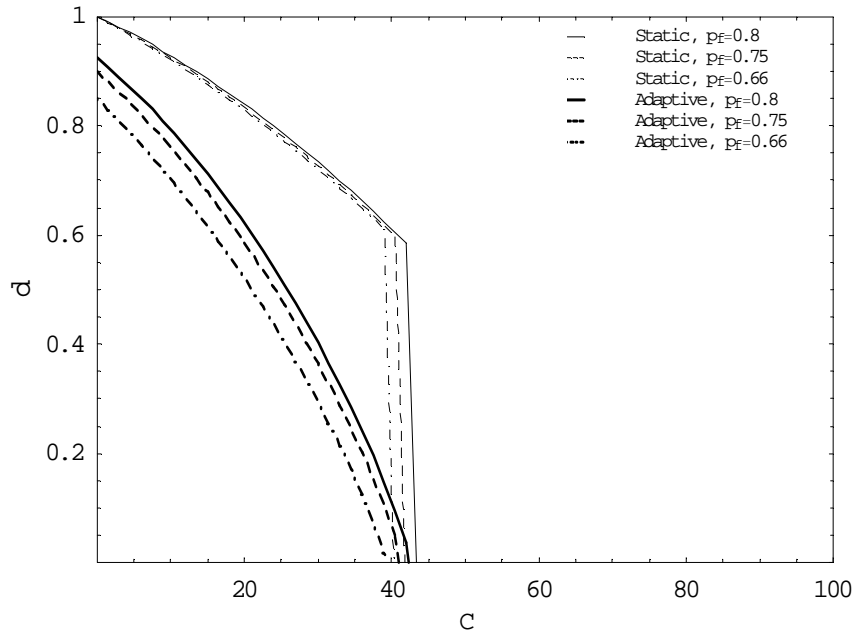


Figure 7-12 Degree of anonymity for P2PRIV, CC interception active-static and active-adaptive attacks,  $N = 100$ .

#### 7.2.4 Summary

The minimum degree of anonymity depends on the system usage and a particular user's requirements. However, as in [32], we restrict acceptable normalized entropy to  $d_{\min} = 0.8$ . Taking into account adaptive attacks, the degree of anonymity of CROWDS, with  $p_f = 0.75$  configuration recommended by CROWDS authors and acknowledged by our analysis, falls below  $d_{\min}$  level for  $C = 5\%$ , while P2PRIV still retains the proper anonymity level even for pessimistic active-adaptive attacks (see Table 7-2). The degree of anonymity offered by CROWDS is considerably low for static attacks. This scenario shows that a set of nodes actively involved in anonymization process should not be too numerous. Longer cascades impose not only larger traffic overheads, but also can make it easier for the adversary to become a member of this set and effectively compromise security of the CROWDS system. Notice that P2PRIV proved high robustness under the same conditions. The degree of anonymity against passive adversary of the P2PRIV system with configuration of  $p_f = 0.66$  is acceptable for the high collaboration level  $C = 10\%$ . Extremely pessimistic active-

Anonymity analysis for P2PRIV

adaptive attacks cannot degrade anonymity of P2PRIV with this configuration with less than 5% of active or adaptive collaborating nodes.

TABLE 7-2  
DEGREE OF ANONYMITY FOR CROWDS AND P2PRIV.

C	Attack	CROWDS				P2PRIV			
		$p_f$				$p_f$			
		0.5	0.66	0.75	0.8	0.5	0.66	0.75	0.8
5%	Passive-Static	0.73	0.69	0.63	0.58	0.97	0.97	0.98	0.98
	Passive-Adaptive	0.53	0.69	<u>0.78</u>	0.82	0.77	<u>0.84</u>	0.88	0.91
	Active-Static	-	-	-	-	0.97	0.97	0.98	0.98
	Active-Adaptive	-	-	-	-	0.68	0.79	0.84	0.87
10%	Passive-Static	0.60	0.51	0.41	0.32	0.94	0.95	0.95	0.96
	Passive-Adaptive	0.51	0.66	<u>0.74</u>	0.79	0.73	<u>0.81</u>	0.85	0.88
	Active-Static	-	-	-	-	0.94	0.94	0.95	0.95
	Active-Adaptive	-	-	-	-	0.60	0.72	0.78	0.81
20%	Passive-Static	0.40	0.25	0.11	0.0099	0.88	0.89	0.90	0.91
	Passive-Adaptive	0.45	0.59	0.67	0.71	0.65	0.74	0.79	0.83
	Active-Static	-	-	-	-	0.86	0.87	0.87	0.88
	Active-Adaptive	-	-	-	-	0.41	0.54	0.61	0.65

Using the anonymity measurement model ([32], [88]) with practical attacks approaches, we have found the proper P2PRIV degree of anonymity. Static-passive attacks have revealed resistance of P2PRIV higher than CROWDS in the entire scope of the collaboration. A significant impact on the P2PRIV anonymity was disclosed only after an injection of a large number of colluding nodes (above 25%). As expected, adaptive attacks have the largest impact. This less realistic scenario is a good reference, because of its pessimistic assumptions. Adaptive attacks show how important a proper selection of cascade length ( $p_f$  value configuration) is. The comparison between CROWDS and P2PRIV

*Anonymity analysis for P2PRIV*

passive-adaptive attacks showed that P2PRIV provides a higher degree of anonymity for realistic amount of collaborating peers – below 60%. The last analyzed scenario, active attacks, does not degrade P2PRIV protection meaningfully, despite of definitive invasion (breaking anonymization cascade by the first colluding node).

# Chapter 8

## Traffic performance analysis for P2PRIV

Based on the previous discussion (Section 7.2) and results showed in Table 7-2 we will simulate P2PRIV with  $p_f=0.66$  and CROWDS with  $p_f=0.75$  configuration, which corresponds to comparative degrees of anonymity for adaptive attacks (where CROWDS achieves its best results) for both systems. These values also correspond to a comparative traffic volume for P2PRIV and CROWDS, because the average number of peers in a cascade is 4 for  $p_f=0.66$  and reaches value 5 for  $p_f=0.75$ . Notice that P2PRIV includes one more link for the same cascade length because of its parallel transport architecture (Figure 8-1).

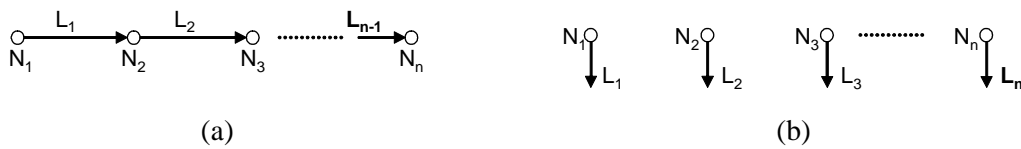


Figure 8-1 Number of nodes and links for cascade (a) and parallel (b) transport architectures.

As an additional reference we will use minimal download time  $\mu_{\min}^{-1}$  (FTP). Let the average link throughput between peers be  $B = 512$  kb/s and average file size  $V = 32$  MB, then

$$\mu_{\min} = \frac{B}{V} = 0.002 [s^{-1}] . \tag{8-1}$$

### 8.1 Download time

To analyze the systems' mean download time we have computed six simulation series (with 30 realizations each) starting from the maximum request arrival rate per each node

$$\lambda_{\max} = \frac{\mu_{\min}}{P} = 0.0005[s^{-1}] . \quad (8-2)$$

Figure 8-2 shows the mean values and 95% confidence intervals of DT for CROWDS and P2PRIV systems as the function of parameter  $\lambda^{-1}$ .

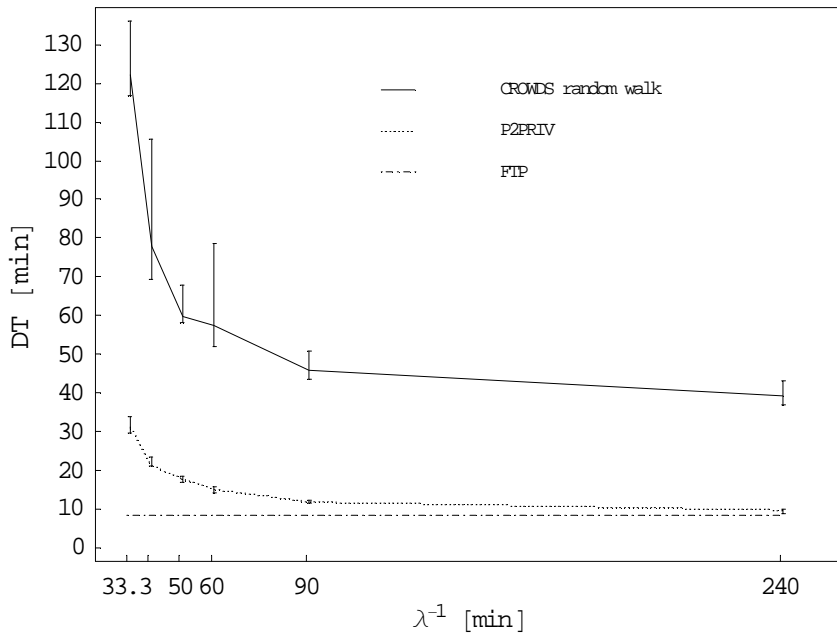


Figure 8-2 Mean download time for P2PRIV and CROWDS random walk,  $N = 100$ .

## 8.2 Dynamics

Below we consider the mean DT characteristics under dynamically changing network traffic conditions. We will analyze system behavior starting from a new file publication. Let  $D$  be the part of all requests which correspond to the new file. We will take into account the behavior of selfish users where simultaneously  $D$  percent of copies leaves the overlay network for each request.

Figure 8-3 and Figure 8-4 show 95% confidence intervals and 25% to 75% quantiles (marked as boxes) surrounding the mean values of DT for both architectures. The results indicate that the parallel transport architecture is more flexible and reacts faster to dynamically changing conditions.

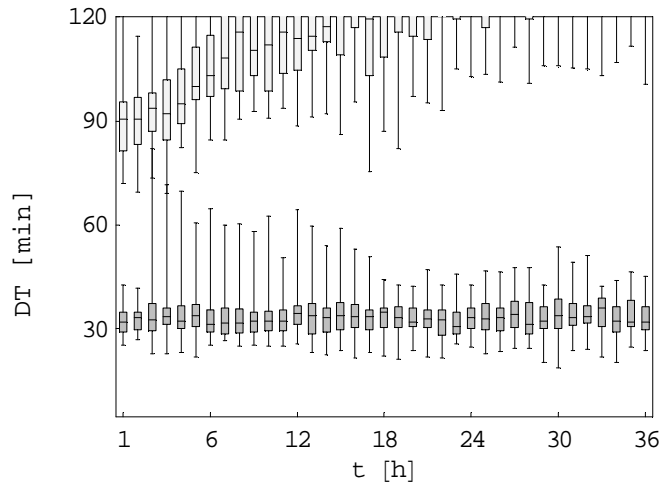
## 8.3 Scalability analysis for P2PRIV

A foreground advantage of P2P is its decentralized architecture and a necessary feature of any practical P2P design is scalability. This vital feature allows for a spontaneous growth of distributed overlays. We have repeated the earlier traffic analysis of P2PRIV and CROWDS with an altered size of simulated networks. Notice that throughout all traffic analysis we did not include the content look-up problem in simulations. The process of finding data in distributed systems mostly impacts scalability; however this approach allows us to revise scalability of pure examined anonymization systems.

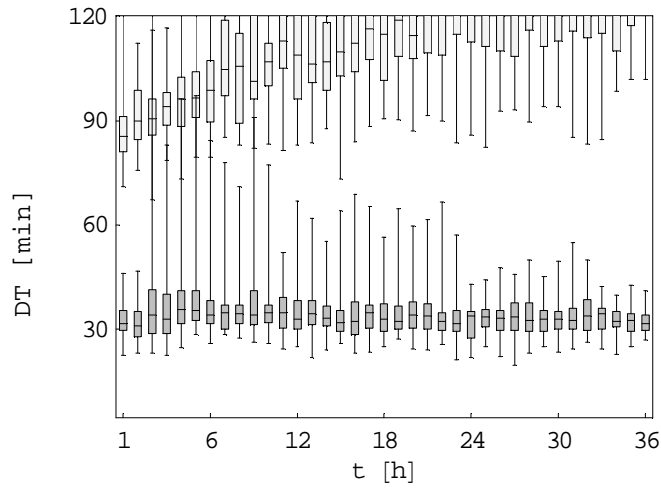
Figure 8-5 shows an impact of network size  $N$  on the download time for P2PRIV and CROWDS random walk (when looking at Figure 8-5 please note that the graphs for  $N = 50$  and  $N = 100$  are shifted to the right in order to avoid confidence intervals overlapping on the diagram).

*Traffic performance analysis for P2PRIV*

(a)  $D = 10\%$



(b)  $D = 20\%$



(c)  $D = 30\%$

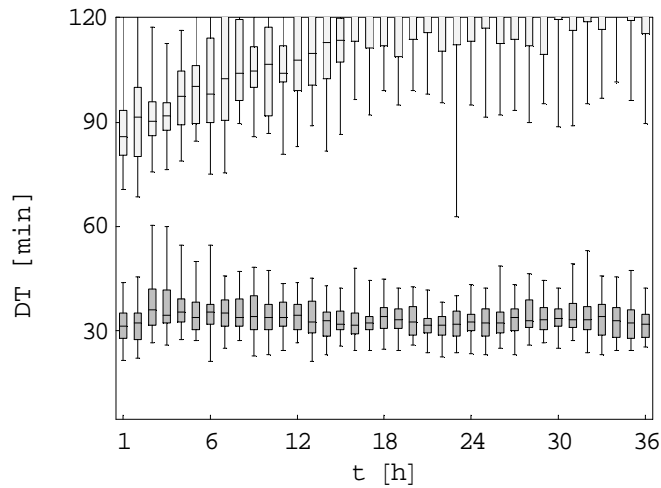
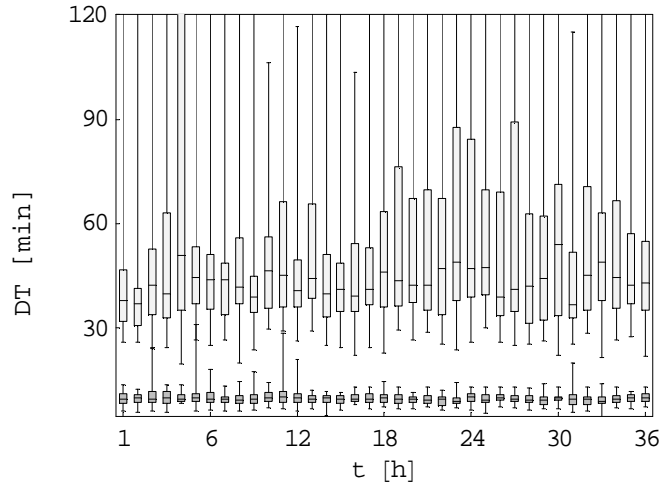


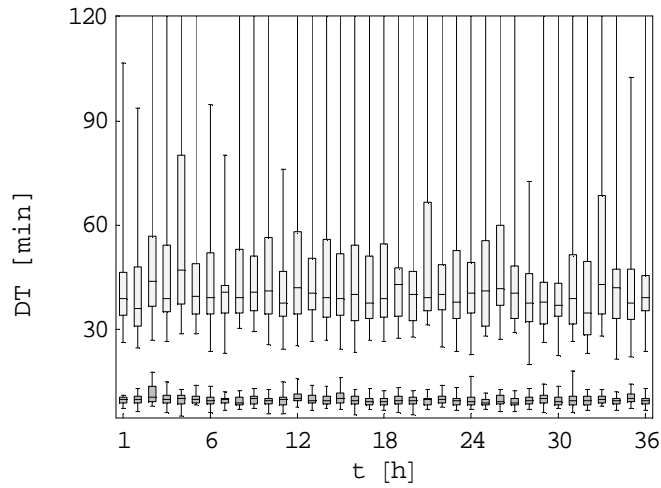
Figure 8-3 Reaction of P2PRIV (lower graphs) and CROWDS random walk (upper graphs) to the new content publication,  $N = 100$ ,  $\lambda = \lambda_{\max} = 33,3^{-1} [\text{min}^{-1}]$

Traffic performance analysis for P2PRIV

(a)  $D = 10\%$



(b)  $D = 20\%$



(c)  $D = 30\%$

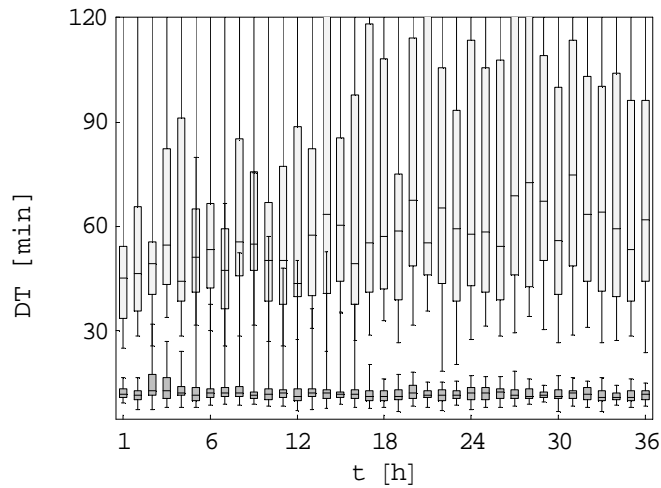


Figure 8-4 Reaction of P2PRIV (lower graphs) and CROWDS random walk (upper graphs) to the new content publication,  $N = 100$ ,  $\lambda = 240^{-1} [\text{min}^{-1}]$ .

*Traffic performance analysis for P2PRIV*

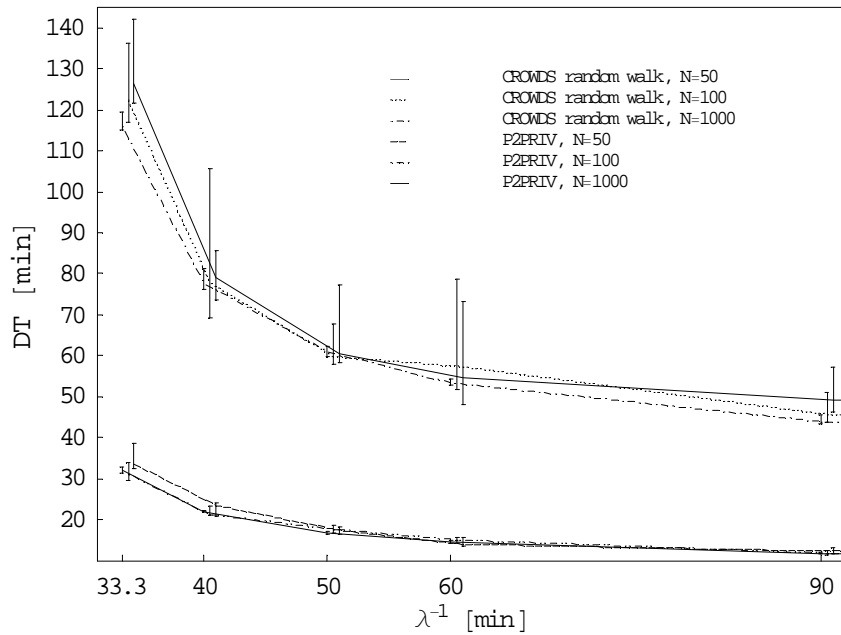


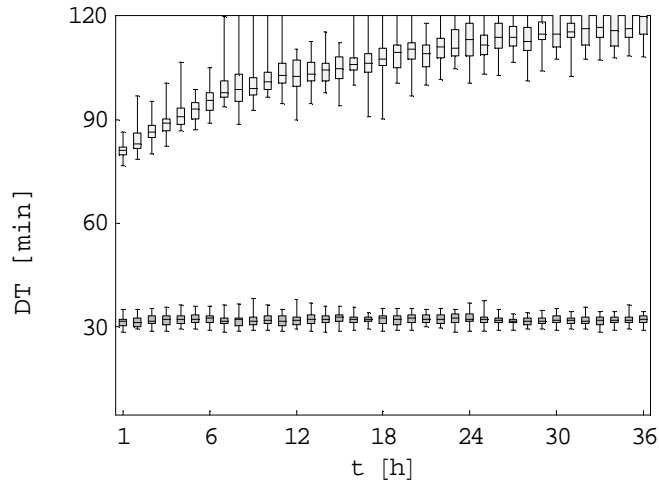
Figure 8-5 Mean download time for different network sizes ( $N$ ) of P2PRIV and CROWDS random walk.

In both systems the dependency of DT on the network size is negligible. For networks with thousand of peers we have observed insignificantly reduced DT in comparison with small networks (50, 100 nodes). Results bounded by narrow confidence intervals indicate that large networks are very reliable; however P2PRIV operation is more stable.

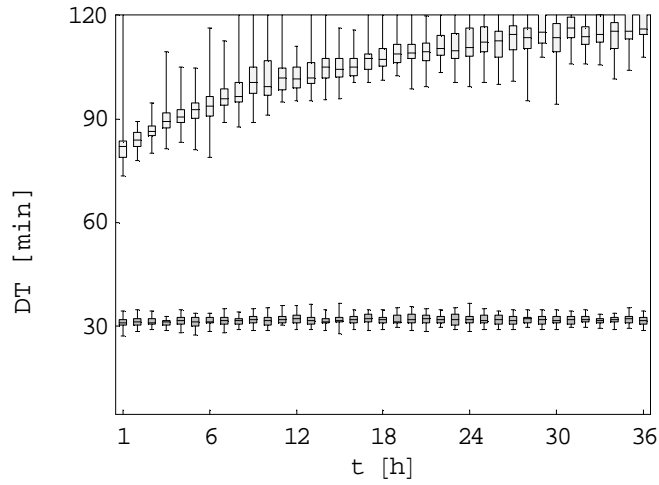
Below we present results of systems dynamics analogous to Chapter 5 and with networks size enlarged to  $N = 1000$ .

Traffic performance analysis for P2PRIV

(a)  $D = 10\%$



(b)  $D = 20\%$



(c)  $D = 30\%$

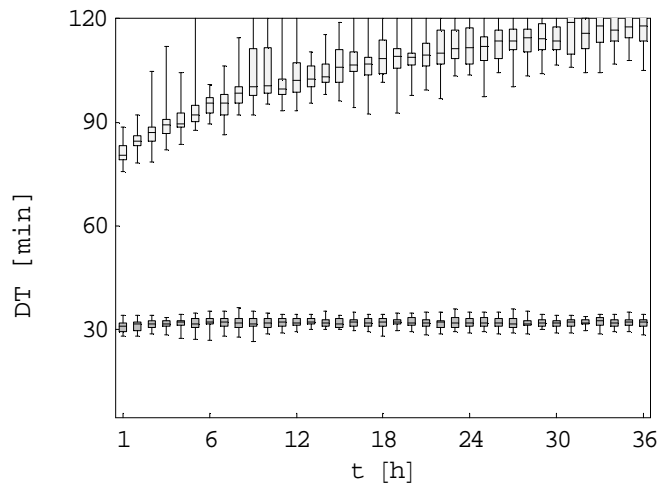
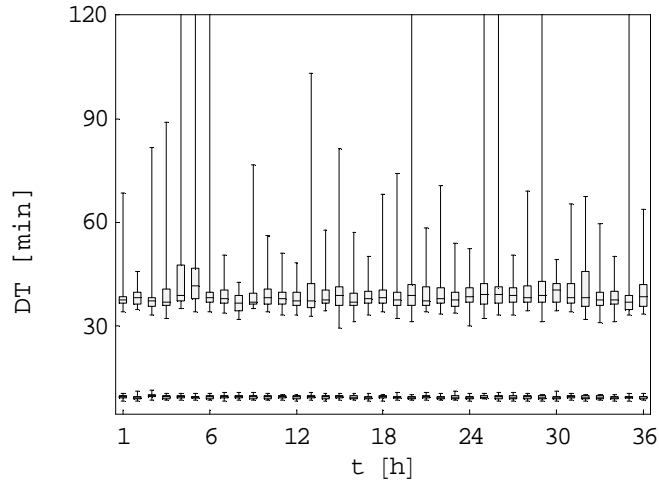


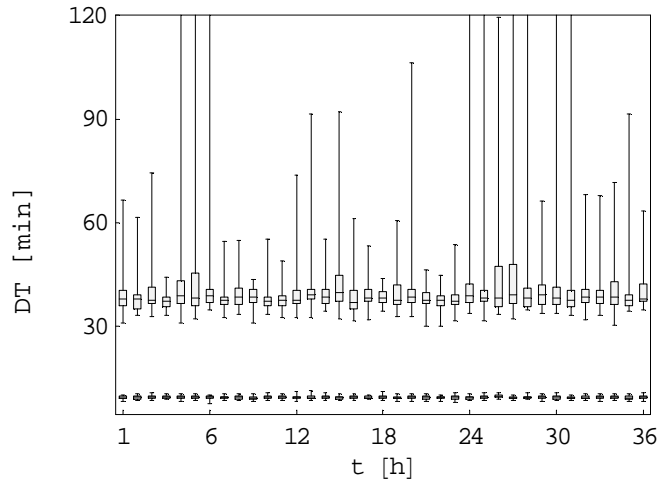
Figure 8-6 Reaction of P2PRIV (lower graphs) and CROWDS random walk (upper graphs) to the new content publication,  $N = 1000$ ,  $\lambda = \lambda_{\max} = 33,3^{-1} [\text{min}^{-1}]$

Traffic performance analysis for P2PRIV

(a)  $D=10\%$



(b)  $D=20\%$



(c)  $D=30\%$

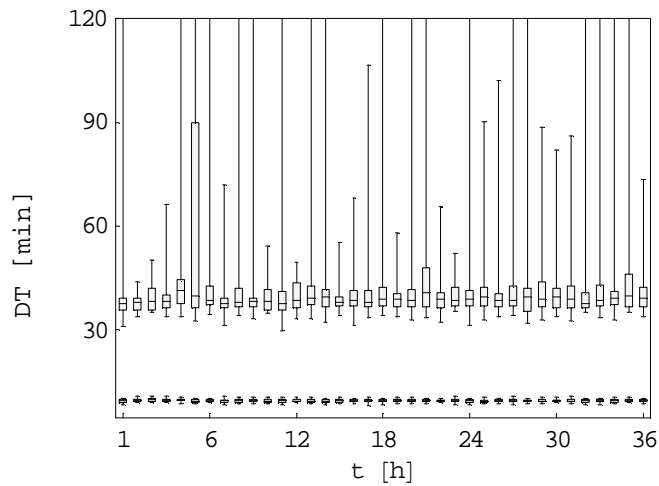


Figure 8-7 Reaction of P2PRIV (lower graphs) and CROWDS random walk (upper graphs) to the new content publication,  $N = 1000$ ,  $\lambda = 240^{-1} [\text{min}^{-1}]$

Figure 8-6 and Figure 8-7 illustrate considerably higher system's stability of operation in comparison to results obtained for networks of a hundred of nodes. Note that P2PRIV proves strong robustness (meaningless changes of DT) even for high dynamics ( $D = 30\%$ ).

#### 8.4 Summary

We have found that P2PRIV DT is close to FTP for low amount of requests. Presented diagrams show the superiority of P2PRIV regardless of the request arrival rate. The content transport has been at least four times faster than for CROWDS (Table 8-1). The observed increase of DT for the close-to-maximum request arrival rate is lower for P2PRIV.

TABLE 8-1  
MEAN DOWNLOAD TIME FOR CROWDS AND P2PRIV.

Requests Arrival Rate [ $s^{-1}$ ]	$2000^{-1}$	$2400^{-1}$	$3000^{-1}$	$3600^{-1}$	$5400^{-1}$	$14400^{-1}$
CROWDS [s]	6960.44	4656.96	3657.8	3201.63	2631.85	2329.83
P2PRIV [s]	1931	1319.87	1005.22	874.105	710.045	574.085

For dynamically changing conditions we have found unstable operation of CROWDS random walk for rate  $\lambda_{\max}$  (Figure 8-3). This maximum request arrival rate does not cause instability of P2PRIV. Instead, we have observed permanently increased DT after the new file publication. Both systems have exhibited a stable operation for a low arrival rate and a moderate migration (Figure 8-4:  $D = 10\%$  and  $D = 20\%$ ). Performance characteristics of CROWDS under a low request arrival rate are similar to P2PRIV under rate  $\lambda_{\max}$ . For higher dynamics ( $D = 30\%$ ), we have found instability in CROWDS even for the low request arrival rate. P2PRIV noticed only temporary (about 2 hours) DT peak under the same conditions.

*Traffic performance analysis for P2PRIV*

We have found that both P2PRIV and CROWDS random walk scale well. Moreover, a large scale of the network increases flexibility of both systems. CROWDS with  $N = 1000$  achieves better stability than with  $N = 100$ , but still presents an unstable operation for a request arrival rate  $\lambda_{\max}$ . What is more, P2PRIV for  $N = 1000$  already tolerates high ( $D = 30\%$ ) migration without introduction of any noticeable delays, regardless of the request arrival rate.

# Chapter 9

## Conclusions

### 9.1 Summary of the contribution

This work describes an original anonymization system dedicated for peer-to-peer overlay networks based on a specific parallel content transport architecture. The proposed concept enables direct and anonymous data transport between network nodes. We have analyzed both anonymity and traffic performance provided by our system.

For anonymity analysis we have applied an information entropy measurement model. The model (taken from [32], [88]) has been revised in order to achieve the P2P environment usability and to reflect the practical capabilities of a P2P adversary. We analyzed, the widely described in the state of the art literature, adaptive observation scenario, and also considered more realistic static attacks. Apart from the passive observation, we also adjusted extremely invasive active attacks, which still cannot be detected quickly in the new parallel architecture. Static attacks aware of boundless extension of forwarding path lengths. The long paths can impose not only larger traffic overheads but can also make it easier for the adversary to become a member of set of nodes actively involved in the anonymization process. For systems that do not protect the traffic by mixing methods the impact of a static attack is significantly higher. Both static and adaptive attacks are vital to the anonymity analysis, as they correspond to different aspects of system protection. The static scenario shows more realistic capabilities of the adversary and it exemplifies a critical point of view on the expansion of a forwarding paths. However, a more pessimistic attack – adaptive observation – is possible. Even though this scenario can take place comparatively rarely, its consequences should be considered. This scenario shows the effectiveness of the system's anonymity protection among network nodes actively involved in hiding the initiator.

Using the entropy measurement model we found that P2PRIV effectively protects user privacy by assuring a high degree of anonymity. For a realistic scope of collaboration,

## *Conclusions*

P2PRIV anonymity is close to maximum. Passive attacks reveal that P2PRIV is robust against the specific character of static attacks and provides a high protection against a penetration of active sets. Active attacks strengthened our remarks, obtained by analysis of passive-adaptive attacks, that P2PRIV should not operate with short management forwarding paths. Very long path lengths in small overlay networks also negatively impact the entropy. Comparisons of our solution with classical architecture showed that for static-passive attacks P2PRIV proved a higher robustness than CROWDS in the entire scope of collaboration. A significant impact on the P2PRIV anonymity was disclosed only after an injection of a large number of colluding nodes. Adaptive attacks had the largest impact. A comparison between CROWDS and P2PRIV passive-adaptive attacks showed that P2PRIV provides better anonymity for realistic amount of collaborating peers. The lastly analyzed scenario, active attacks, does not meaningfully degrade P2PRIV protection, despite of a definitive invasion (breaking management path by the first colluding node).

For purposes of the traffic performance analysis, we have created a simulation model of P2P traffic, allowing for an evaluation of the system mean download time and its dynamics. We have observed that the empirical analysis of a network random walk algorithm provided by our simulation environment has been analogous to results obtained analytically for representative boundary conditions. The system's behavior under dynamically changing conditions (caused by an unpredictable users migration and traffic bursts after a new publication of a popular content) shows that a stable operation of the network random walk from CROWDS can be retained for dynamics lower or equal to 20%.

Likewise, in the security evaluation we have compared the traffic performance measures of the CROWDS system with the proposed solution. We have found that the proposed system significantly decreases download time as compared with traditional cascade schemes, and achieves results close to optimum for low to medium loaded networks. Taking into account network dynamics, we have found that the proposed system is more flexible and reacts faster to dynamically changing conditions. We believe that the included analysis of the anonymous traffic performance is a vital contribution to the field of the anonymous systems modeling.

## *Conclusions*

Finally, we have analyzed scalability of both architectures. We have found that the new system scales well and proves a high flexibility for large networks.

To summarize: in this work, a novel peer-to-peer system has been proposed. It introduces a parallel content transport architecture separating the anonymization process from the transport function. We have proved that the separation of the anonymization process from the content transport function is possible with retention of an acceptable anonymity and the degree of anonymity comparable to classical solutions. We have showed that the proposed solution provides both a significant reduction of the content download time and an improvement in overlay network dynamics. The solution scales well and can operate vastly in wide area networks. We believe that P2PRIV can satisfy the requirement of private and low latency transport of large information content.

In December 2007, the P2PRIV concept was submitted to the Polish Patent Office (patent application #P384095) and successfully passed a preliminary evaluation. The P2PRIV has also received a positive reaction from the Technical Committee on Distributed Processing (TCDP) of IEEE Society: the paper “A Concept of an Anonymous Direct P2P Distribution Overlay System,” I. Margasiński, M. Pióro [63], has been accepted to the 22nd IEEE International Conference on Advanced Information Networking and Applications (AINA2008), Ginowan, Okinawa, Japan, as one of 145 from all 469 submitted papers. AINA is one of the most important international conferences supported by the TCDP.

## 9.2 Future work

### **9.2.1 Traffic performance modeling for anonymous system**

In our opinion the field of traffic performance modeling for anonymous systems is still in its infancy and most of the papers neglect this important issue. Consequently, our future work will include a more detailed analysis of the impact imposed by anonymous techniques on network traffic performance. An interesting goal is the consideration of overlay dynamics with simultaneous migration of many content resources. The traffic performance analysis, presented in this thesis, has covered the two following elements:

## *Conclusions*

- the evaluation of the download time in the stable operation of simulated overlays; and
- the measures of the download time after the publication of a new file – we have considered a series of copying and removing of the single content resource.

The future analysis should include modeling of anonymous traffic where many new files are copied and removed in the overlay network simultaneously. Additionally, this research can be based on a sociological analysis of users' habits, as this study can bring the simulation model closer to real networks.

### **9.2.2 DHT suit**

The peer-to-peer system proposed in this thesis requires a usage of a DHT look-up system. A future development of the system should cover a study of the distributed hash table interface adjusted to the considered concept of anonymous parallel transport. The future research can bring the new design of such a DHT algorithm, which corresponds to the specific transport imposed by the P2PRIV. An interesting goal is the optimization of DHT look-up and storage performance. In this adaptation, it should be considered that in the new overlay networks, based on the P2PRIV concept, it is known that one request is followed by a series of other requests for the same content.

As well as other practical implementation issues, the future research can include analysis of an integration of the P2PRIV system with an anonymous publication function.

### **9.2.3 Anonymous content publication**

For a pervasive use of the presented system it is necessary to include in it the anonymous publication function, which together with the described P2PRIV anonymous transport, can provide a comprehensive anonymous communications solution. The publication process can be treated as a separated function of the overlay, however, to achieve a coherent prototype, the publication scheme should be based on a concurrent with P2PRIV set of security primitives. The other motivating goal is selecting such a known

## *Conclusions*

anonymous publication system that can share as many security primitives embedded in the P2PRIV as possible.

### **9.2.4 Applications**

The proposed solution is dedicated to the transport of a large content. However, it is motivating to analyze its usefulness in a wider range of applications. The next vital question is if the new system can assure generic anonymous communications. In this case, it should be studied how robust P2PRIV is against long-term intersection attacks. Having in mind a single access to particular content by a single user, characteristic for a sharing of static, large content, we did not analyze this issue. However, in systems of general purpose, the problem of multiple accesses to the same resource should be considered. In peer-to-peer overlays it is difficult to ensure a long-term availability of individual peers, so the cascade is deemed to change over time (with the exception of the initiator). The generic purpose anonymous system, based on the P2PRIV architecture, should allow the initiator to maintain the composition of the cloning cascade for a long period of time.

# Bibliography

- [1] R. Anderson, “The eternity service,” in *Proceedings of 1<sup>st</sup> International Conference on the Theory and Applications of Cryptology (Pragocrypt’96)*, pages 242–252. Czech Technical University Publishing House, 1996.
- [2] S. Axelsson: *Intrusion Detection Systems: A Survey and Taxonomy*, Dept. of Computer Engineering, Chalmers Univ. of Technology, TR:99-15, March 2000.
- [3] Anonymizer.com
- [4] H. Balakrishnan, M. F. Kaashoek, D. Karger, R. Morris, and I. Stoica, “Looking Up Data in P2P Systems,” *Communications of the ACM*, Volume 46, Issue 2, Pages: 43 – 48, February 2003.
- [5] T. Benes, “The strong eternity service,” in *Proceedings of Information Hiding Workshop (IH’01)*, pages 215–229. Springer-Verlag, LNCS 2137, 2001.
- [6] K. Bennett and C. Grothoff, “GAP – practical anonymous networking,” in Roger Dingledine, editor, *Proceedings of The Privacy Enhancing Technologies Workshop*. Springer-Verlag, LNCS 2760, March 2003.
- [7] S. Berkovsky, N. Borisov, Y. Eytani, T. Kuflik, and F. Ricci, “Examining Users’ Attitude towards Privacy-Preserving Collaborative Filtering,” in *Proceedings of the Workshop on Data Mining for User Modeling (K-DUUM)*, in conjunction with the International Conference on User Modeling (UM), Corfu, Greece, June 2007.
- [8] T. Berners-Lee, R. Fielding, H. Frystyk: *Hypertext Transfer Protocol – HTTP/1.0*. RFC 1945, 1996.
- [9] O. Berthold, H. Federrath, and S. Köpsell, “Web MIXes: A System for Anonymous and Unobservable Internet access,” in *Proceedings of the Workshop on Design Issues in Anonymity and Unobservability*, pages 101–115, Berkeley, CA, USA, July 25–26 2000.
- [10] O. Berthold, H. Langos, “Dummy traffic against long term intersection attacks,” in *Designing Privacy Enhancing Technologies Proceedings of PET’02*, pages 110–128. Springer-Verlag, LNCS 2482, 2002.
- [11] O. Berthold, A. Pfitzmann, and R. Standtke: “The disadvantages of free mix routes and how to overcome them,” in *Proc. Workshop on Design Issues in Anonymity and Unobservability (25-26 July 2000)*, ICSI TR-00-011, pp. 27-42.
- [12] N. Borisov: *Anonymous Routing in Structured Peer-to-Peer Overlays*. PhD Thesis, UC Berkeley, 2005.
- [13] R. Böhme, G. Danezis, C. Diaz, S. Köpsell, and A. Pfitzmann, “Mix cascades vs. peer-to-peer: Is one concept superior?,” in *Designing Privacy Enhancing Technologies*, Proceedings of PET’04, pages 243–255. Springer-Verlag, LNCS 3424, 2005.

## Bibliography

- [14] K. Brzeziński, I. Margasiński, and K. Szczypiorski, “Prywatne wojny w sieci: poddaj się, okop, negocjuj lub stań do walki”, in *Proceedings of the Internet 2004*, Wrocław, December 2004. (in polish)
- [15] D. Chaum, “The Dining Cryptographers Problem: Unconditional Sender and Recipient Untraceability,” *Journal of Cryptology* 1/1 (1988). 65-75.
- [16] D. Chaum, “Untraceable electronic mail, return addresses, and digital pseudonyms,” *Communications of the ACM*, 4(2), February 1981.
- [17] I. Clarke, O. Sandberg, B. Wiley, and T. W. Hong, “Freenet: A distributed anonymous information storage and retrieval system,” in *Proceedings of Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability*, pages 46–66, July 2000.
- [18] L. Cottrell, “Mixmaster and remailer attacks”, 1994. Available: <http://web.inf.tu-dresden.de/~hf2/anon/mixmaster/remailer-essay.html>.
- [19] L. Cranor, M. Langheinrich, M. Marchiori, M. Presler-Marchall: *The Platform for Preferences 1.0 (P3P 1.0) Specification*, W3C Recommendation, April 2002. Recommendation, 16 April 2002.
- [20] D. Cvrcek, M. Kumpost, V. Matyas, and G. Danezis, “Study on the Price of Privacy,” in *Proceedings of Workshop on Privacy in the Electronic Society*, Alexandria, VA, USA, October 30, 2006.
- [21] W. Dai, “Pipenet 1.1,” Usenet post. Available: <http://www.eskimo.com/~weidai/pipenet.txt>, 1996.
- [22] G. Danezis: “Mix-networks with restricted routes,” in Roger Dingledine, editor, *Proceedings of Privacy Enhancing Technologies Workshop (PET 2003)*. SpringerVerlag, LNCS 2760, March 2003.
- [23] G. Danezis, R. Dingledine, and N. Mathewson, “Mixminion: Design of a Type III anonymous remailer protocol,” in *Proceedings of the IEEE Symposium on Security and Privacy*, May 2003.
- [24] G. Danezis, S. Lewis and R. Anderson, “How Much is Location Privacy Worth?” in *Proceedings of the Fourth Workshop on the Economics of Information Security (WEIS 2005)*. Harvard University, 2 – 3 June 2005.
- [25] T. Dierks, C. Allen: *The TLS-Protocol Version 1.0*, RFC 2246, 1999.
- [26] R. Dingledine, M. J. Freedman, and D. Molnar, “The free haven project: Distributed anonymous storage service,” in H. Federrath, editor, *Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability*. Springer-Verlag, LNCS 2009, July 2000.

## Bibliography

- [27] M. J. Freedman and R. Morris, "Tarzan: A peer-to-peer anonymizing network layer," in *9<sup>th</sup> ACM Conference on Computer and Communications Security*, Washington, DC, November 2002.
- [28] C. Diaz, *Anonymity and Privacy in Electronic Services*, Ph.D. Thesis. Katholieke Univesiteit Leuven, 2005.
- [29] C. Diaz, "Anonymity Metrics Revisited," in *Anonymous Communication and its Applications, number 05411 in Dagstuhl Seminar Proceedings*. Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, S. Dolev, R. Ostrovsky, and A. Pfitzmann (Eds), 6 pages, 2006.
- [30] C. Diaz, L. Sassaman, and E. Dewitt, "Comparison between two practical mix designs," in *Proceedings of 9<sup>th</sup> European Symposium on Research in Computer Security (ESORICS'04)*, pages 141–159. Springer-Verlag, LNCS 3193, 2004.
- [31] C. Diaz and A. Serjantov, "Generalising Mixes," in *Proceedings of Privacy Enhancing Technologies (PET'03)*. R. Dingledine (Ed.), Springer LNCS 2760, pp. 18-31, 2003.
- [32] C. Diaz, S. Seys, J. Claessens and B. Preneel, "Towards measuring anonymity," in *Roger Dingledine and Paul Syverson, editors, Proceedings of the Privacy Enhancing Technologies Workshop*. Springer-Verlag, LNCS 2482, April 2002.
- [33] R. Dingledine, M. Freedman, and D. Molnar, "The free haven project: Distributed anonymous storage service," in *Designing Privacy Enhancing Technologies*, pages 67–95. Springer-Verlag, LNCS 2009, 2000.
- [34] D. Eastlake, P. Jones, "Us secure hash algorithm 1 (sha1)," *Tech. Rep. 3174*, 2001.
- [35] ECMA-262 Standard ECMAScript Language Specification 3<sup>rd</sup> edition, December 1999.
- [36] ECMA-327 Standard ECMAScript 3<sup>rd</sup> Edition Compact Profile, June 2001.
- [37] N. Feamster, M. Balazinska, G. Harfst, H. Balakrishnan, and D. Karger, "Infranet: Circumventing web censorship and surveillance," Dan Boneh, editor, *USENIX Security Symposium*, pages 247-262, San Francisco, CA, 5-9 August 2002.
- [38] N. Feamster, M. Balazinska, W. Wang, H. Balakrishnan, and D. Karger, "Thwarting web censorship with untrusted messenger discovery," Roger Dingledine, editor, *Privacy Enhancing Technologies workshop (PET 2003)*, volume 2760 of LNCS, pages 125-140, Springer-Verlag, Dresden, Germany, March 2003.
- [39] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, and T. Berners-Lee: *Hypertext Transfer Protocol – HTTP/1.1*. RFC 2616, June 1991.
- [40] M. J. Freedman, E. Freudenthal, and D. Mazières, "Democratizing content publication with Coral," in *Proceedings of the 1<sup>st</sup> Symposium on Networked Systems Design and Implementation*, 2004.
- [41] S. Garfinkel and N. Makarevitch. *PGP: Pretty Good Privacy*. O'Reilly International Thomson, 1995.

## Bibliography

- [42] *The Gnutella Protocol Specification*. Available: <http://dss.clip2.com/GnutellaProtocol04.pdf>, 2001.
- [43] B. I. Goldberg and A. Shostack, "Freedom systems 2.1 security issues and analysis," *White paper, Zero Knowledge Systems, Inc.*, May 2001.
- [44] I. Goldberg and D. Wagner, "TAZ servers and the rewebber network: Enabling anonymous publishing on the world wide web," *First Monday*, 3(4), 1998.
- [45] I. Goldberg, D. Wagner and E. Brewer, "Privacy-enhancing technologies for the Internet," in *Proceedings of the 42<sup>nd</sup> IEEE Spring COMPCON*. IEEE Computer Society Press, February 1997.
- [46] D. Goldschlag, M. Reed, and P. Syverson, "Hiding Routing Information," in *Proceedings of Information Hiding (IH'96)*, pages 137–150. Springer-Verlag, LNCS 1174, 1996.
- [47] C. Gülcü and G. Tsudik, "Mixing E-mail with Babel," in *Proceedings of the Network and Distributed Security Symposium (NDSS'96)*, pages 2–16. IEEE, 1996.
- [48] I. Hall-Beyer. *Gnutella*. Available: <http://www.gnutella.com>.
- [49] S. Hazel and B. Wiley, "Achord: A variant of the chord lookup service for use in censorship resistant peer-to-peer publishing systems," in P. Druschel, M. F. Kaashoek, and A. I. T. Rowstron, editors, *Peer-to-Peer Systems, First International workshop, IPTPS 2002*, volume 2429 of LNCS, Cambridge, MA, USA, 7-8 March 2002. Springer-Verlag.
- [50] K. Hildrum, J. Kubiawicz, S. Rao, and B. Zhao, "Distributed Object Location in a Dynamic Network," in *Proceedings of 14<sup>th</sup> ACM Symposium on Parallel Algorithms and Architectures (SPAA)*, August 2002.
- [51] O. Heckmann, A. Bock, A. Mauthe, and R. Steinmetz, "The eDonkey File-Sharing Network," in *GI Jahrestagung (2)*, LNI, Vol. 51, pages 224-228, GI, 2004.
- [52] A. Jerichow, J. Müller, A. Pftzmann, B. Pftzmann, and M. Waidner, "Real-Time MIXes: A Bandwidth-Efficient Anonymity Protocol," in *IEEE Journal on Selected Areas in Communications*, 16(4), pages 495-509, 1998.
- [53] F. Kaashoek and D. Karger, "Koorde: A degree-optimal distributed hash Table," in *Proceedings of the 2<sup>nd</sup> International Workshop on Peer-to-peer Systems*, February 2003.
- [54] K. Karger, and M. Ruhl, "Finding nearest neighbors in growthrestricted metrics," in *Proceedings ACM Symposium on the Theory of Computing (May 2002)*, pages 741–750.
- [55] D. Kesdogan, J. Egner, and R. Büschkes: "Stop-and-go MIXes: Providing probabilistic anonymity in an open system," in *Proceedings of Information Hiding Workshop (IH 1998)*, pages 83–98. Springer-Verlag, LNCS 1525, 1998.
- [56] M. Klonowski, M. Kutylowski, and F. Zagorski, "Anonymous communication with on-line and off-line onion encoding," in *Proceedings of SOFSEM '05*, Volume 3381 of LNCS, pages 229-238, 2005.

## Bibliography

- [57] R. Kristol and L. Montulli: *HTTP State Management Mechanism*. RFC 2965, 2000.
- [58] D. Kügler, "An Analysis of GUNet and the Implications for Anonymous, Censorship-Resistant Networks," in Roger Dingledine, editor, *Privacy Enhancing Technologies workshop (PET 2003)*, volume 2760 of LNCS, pages 161+176, Dresden, Germany, March 2003. Springer-Verlag.
- [59] J. Liang, R. Kumar, and K. W. Ross, "The FastTrack overlay: A measurement study" *Computer Networks*, 50(6), pages 842-858, 2006.
- [60] D. Liben-Nowell, H. Balakrishnan, and D. Karger, "Analysis of the evolution of peer-to-peer systems," in *Proceedings in ACM Symposium on the Principles of Distributed Computing*. Monterey, CA (July 2002).
- [61] I. Margasiński, "Prywatność w systemach P2P," in Proceedings of the KSTiT 2006, Bydgoszcz, September 2006. (in Polish)
- [62] I. Margasiński, "Zapewnianie anonimowości przy przeglądaniu stron WWW," in *Proceedings of KST 2002*, Bydgoszcz, September 2002. (in Polish)
- [63] I. Margasiński and M. Pióro, "A Concept of an Anonymous Direct P2P Distribution Overlay System," in *Proceedings of the 22<sup>nd</sup> IEEE International Conference on Advanced Information Networking and Applications (AINA2008)* ISSN 1550-445X, ISBN 978-0-7695-3095-6, pp. 590-597, Ginowan, Okinawa, Japan, March 2008.
- [64] I. Margasiński and K. Szczypiorski, "Prywatność z protokołem P3P w transakcjach online," in *Proceedings of Enigma 2004*, Warsaw, May 2004. (in Polish)
- [65] I. Margasiński and K. Szczypiorski, "Prywatność w sieciach bezprzewodowych Wi-Fi, Bluetooth, ZigBee oraz RFID," in *Proceedings of Enigma 2005*, Warsaw, May 2005. (in polish)
- [66] I. Margasiński and K. Szczypiorski, "VAST: Versatile Anonymous System for Web Users," in *Proceedings of ACS-CISIM 2004*, J. Pejaś, A. Piegat (Eds) *Enhanced Methods in Computer Security, Biometric and Artificial Intelligence Systems*, Springer, ISBN 1-4020-7776-9, pp. 71-82, November 2004.
- [67] I. Margasiński and K. Szczypiorski, "Web Privacy: an Essential Part of Electronic Commerce," in Proceedings of the 3<sup>rd</sup> *International Interdisciplinary Conference on Electronic Commerce (ECOM-03)* ISBN 83-88617-75-3, pp. 65-72, Gdańsk, October 2003.
- [68] I. Margasiński and K. Szczypiorski, "Wszechstronna anonimowość klienta HTTP," in *Proceedings of the KST 2003*, Bydgoszcz, September 2003.
- [69] I. Margasiński, K. Szczypiorski, and K. M. Brzeziński, "Private Wars in Secret Life," in Proceedings of the *Next Generation Internet Networks (NGI 2005)*, ISBN: 0-7803-8900-X, Rome, 2005.
- [70] D. N. Mathewson and P. Syverson, "Tor: The second generation onion router," in *Proceedings of the 13<sup>th</sup> USENIX Security Symposium*, August 2004.

## Bibliography

- [71] U. Möller, L. Cottrell, P. Palfrader, and L. Sassaman, “Mixmaster Protocol — Version 2,” Draft, July 2003.
- [72] S. J. Murdoch, and G. Danezis, “Low-cost Traffic Analysis of Tor,” in *Proceedings of the 2005 IEEE Symposium on Security and Privacy*, May 8-11, 2005, Oakland, California, USA.
- [73] A. M. Odlyzko, “Privacy, economics, and price discrimination on the internet,” in N. Sadeh, editor, *International Conference on Electronic Commerce (ICEC 2003)*, pages 355-366. ACM Press, 2003.
- [74] T. O’Reilly: *What Is Web 2.0. Design Patterns and Business Models for the Next Generation of Software*. Available: <http://www.oreilly.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>
- [75] B. Pfitzmann, “Breaking Efficient Anonymous Channel,” in *Advances in Cryptology, Proceedings of EUROCRYPT 1994*, pages 332–340. Springer-Verlag, LNCS 950, 1994.
- [76] A. Pfitzmann and M. Köhntopp, “Anonymity, Unobservability and Pseudonymity – A Proposal for Terminology,” in *Hannes Federath (Ed.), Designing Privacy Enhancing Technologies*, Lecture Notes in Computer Science, LNCS 2009, pp. 1-9, Springer-Verlag, 2001.
- [77] B. Pfitzmann and A. Pfitzmann, “How to break the direct RSA implementation of MIXes,” in *Advances in Cryptology, Proceedings of EUROCRYPT 1989*, pages 373–381. Springer-Verlag, LNCS 434, 1990.
- [78] A. Pfitzmann, B. Pfitzmann and M. Waidner, “ISDN-mixes: Untraceable communication with very small bandwidth overhead,” in *Proceedings of the GI/ITG Conference on Communication in Distributed Systems*, pages 451–463, 1991.
- [79] C. Plaxton, R. Rajaraman, and A. Richa, “Accessing nearby copies of replicated objects in a distributed environment,” in *Proceedings of ACM Symposium on Parallel Algorithms and Architectures (SPAA)*, Newport, Rhode Island (June 1997).
- [80] M. Presler-Marshall: *The Platform for Privacy Preferences 1.0 Deployment Guide*, W3C Note, May 2001.
- [81] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Schenker, “A scalable content-addressable network,” in *Proceedings of ACM SIGCOMM*, pages 161–172, August 2001.
- [82] M. Reed, P. Syverson, and D. Goldschlag, “Anonymous Connections and Onion Routing,” in *IEEE Journal on Selected Areas in Communications*, 16(4):482–494, 1998.
- [83] M. Reed P. Syverson and D. Goldschlag, “Onion routing access configurations,” in *DARPA Information Survivability and Exposition (DISCEX 2000)*, IEEE CS Press, 2000.
- [84] M. Rennhard and B. Plattner, “Introducing MorphMix: Peer-to-Peer based Anonymous Internet Usage with Collusion Detection,” in *Proceedings of the Workshop on Privacy in the Electronic Society (WPES 2002)*, Washington, DC, USA, November 2002.
- [85] M. K. Reiter, A. D. Rubin, “Crowds: Anonymity for web transactions,” *ACM Transactions on Information and System Security*, 1(1), June 1998.

## Bibliography

- [86] R. Rivest, "The MD5 Message-Digest Algorithm," *RFC 1321*, MIT Laboratory for Computer Science and RSA Data Security, Inc., April 1992.
- [87] A. Rowstron, P. Druschel, "Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems," in *IFIP/ACM International Conference on Distributed Systems Platforms (Middleware)*, 2001.
- [88] A. Serjantov and G. Danezis, "Towards an information theoretic metric for anonymity," in *Roger Dingledine and Paul Syverson, editors, Proceedings of the Privacy Enhancing Technologies Workshop*, San Diego, CA, April 2002. Springer-Verlag, LNCS 2482.
- [89] A. Serjantov, R. Dingledine, and P. Syverson: "From a trickle to a flood: Active attacks on several mix types," in *Proceedings of Information Hiding Workshop (IH 2002)*. Springer-Verlag, LNCS 2578, 2002.
- [90] A. Serjantov and S. J. Murdoch, "Message Splitting Against the Partial Adversary," in *Proceedings of the Privacy Enhancing Technologies Workshop*. June 2005. LNCS 3856. Cavtat, Croatia.
- [91] C. E. Shannon, *A Mathematical Theory Of Communication*, the Bell System Technical Journal, Vol. 27, pp. 379–423 and pp. 623–656, 1948.
- [92] V. Shmatikov, "Probabilistic analysis of anonymity," in *Proceedings of the Computer Security Foundations workshop (CSFW-15 2002)*, pages 119-128, Cape Breton, Nova Scotia, Canada, 24-26 June 2002. IEEE Computer Society.
- [93] E. Sit and R. Morris, "Security considerations for peer-to-peer distributed hash tables," in *Proceedings of the 1<sup>st</sup> International Workshop on Peer-to-Peer Systems*, Cambridge, MA, March 2002.
- [94] I. Stoica, R. Morris, D. Karger, F. Kaashoek and H. Balakrishnan, "Chord: A Scalable Peer-To-Peer Lookup Service for Internet Applications," in *ACM SIGCOMM*, 2001.
- [95] P. Syverson, G. Tsudik, M. Reed, and C. Landwehr, "Towards an Analysis of Onion Routing Security," in *Designing Privacy Enhancing Technologies*, pages 96–114. Springer-Verlag, LNCS 2009, 2000.
- [96] *Testimony of Shawn Fanning*. Napster Inc. October 2000. Available: [http://judiciary.senate.gov/oldsite/1092000\\_sf.htm](http://judiciary.senate.gov/oldsite/1092000_sf.htm).
- [97] K. Szczypiorski, A. Zwierko, I. Margasiński, "MINX: Micropayments with Secure Network Exchange," in *Proceedings of the 3rd International Interdisciplinary Conference on Electronic Commerce (ECOM-03)*, ISBN 83-88617-75-3, pp. 65-72, Gdańsk, October 2003.
- [98] K. Szczypiorski, A. Zwierko, and I. Margasiński, "Micropayments with Privacy – a New Proposal for E-commerce," in *Proceedings of ACS-CISIM 2003*, K. Saeed, J. Pejaś (Eds) Information Processing and Security Systems, Springer, ISBN 0-387-25091-3, pp. 175-186, June 2004.

## Bibliography

- [99] P. Tabriz and N. Borisov, “Breaking the collusion detection mechanism of MorphMix,” in *Proceedings of the Privacy Enhancing Technologies Workshop (PET)*, June 2006.
- [100] M. Waldman and D. Mazières, “Tangler: a censorship-resistant publishing system based on document entanglements,” in *Proceedings of the 8<sup>th</sup> ACM Conference on Computer and Communications Security (CCS 2001)*, pages 126–135. ACM Press, 2001.
- [101] M. Waldman, A. Rubin, and L. Cranor, “Publius: A robust, tamper-evident, censorship-resistant and source-anonymous web publishing system,” in *Proceedings of the 9<sup>th</sup> USENIX Security Symposium*, pages 59–72, 2000.
- [102] W. Dai. *Pipenet 1.1*. Usenet post, 1996.
- [103] M. Wright, M. Adler, B. N. Levine, and C. Shields, “An analysis of the degradation of anonymous protocols,” in *Proceedings of the Network and Distributed Security Symposium (NDSS’02)*, San Diego, California, 6-8 February 2002.
- [104] J. Yang, H. Ma, W. Song, J. Cui, and C. Zhou, “Crawling the eDonkey Network,” in *Proceedings of GCC Workshops*, pages 133-136, IEEE Computer Society, 2006.
- [105] B. Y. Zhao, J. Kubiawicz, A. D. Joseph: *Tapestry: An infrastructure for fault-tolerant wide-area location and routing*, Technical Report UCB/CSD-01-1141, UC Berkeley, 2001.
- [106] P. Zimmermann, *The Official PGP User’s Guide*. MIT Press, 1995.

# Index of Acronyms

API	-	Application Programming Interface
CAN	-	Content-Addressable Network
CC	-	Cloning Cascade
DHT	-	Distributed Hash Table
DT	-	Download Time
HTTP	-	HyperText Transfer Protocol
IP	-	Internet Protocol
ISDN	-	Integrated Services Digital Network
ISP	-	Internet Service Provider
LAN	-	Local Area Network
MD5	-	Message-Digest algorithm 5
MPEG	-	Moving Picture Experts Group
P2P	-	Peer-to-Peer
P2PRIV	-	Peer-to-Peer diRect and anonymous dIstribution oVerlay
P3P	-	Platform for Privacy Preferences Protocol
PET	-	Privacy Enhancing Technologies
PGP	-	Pretty Good Privacy
PPO	-	Polish Patent Office
SHA	-	Secure Hash Algorithm
SOHO	-	Small Office/Home Office
SSL	-	Secure Socket Layer
TCDP	-	IEEE Technical Committee on Distributed Processing
TCP	-	Transmission Control Protocol
TLS	-	Transport Layer Security
TOR	-	The Onion Router
TTL	-	Time to Live
VAST	-	Versatile Anonymous System for Web Users
WAN	-	Wide Area Network
WWW	-	World Wide Web