A Concept of an Anonymous Direct P2P Distribution Overlay System

Igor Margasiński, Michał Pióro
Institute of Telecommunications, Warsaw University of Technology
{I.Margasinski, M.Pioro}@tele.pw.edu.pl

Abstract

The paper introduces a peer-to-peer system called P2PRIV (peer-to-peer direct and anonymous distribution overlay). Basic novel features of P2PRIV are: (i) a peer-to-peer parallel content exchange architecture, and (ii) separation of the anonymization process from the transport function. These features allow a considerable saving of service time while preserving high degree of anonymity. In the paper we evaluate anonymity measures of P2PRIV (using a normalized 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.

1. Introduction

The broadband Internet access has given a way for explosion of large scale distributed overlay networks. Peer-to-Peer (P2P) overlays have grown to one of the leading Internet applications and constitute a significant part of the Internet traffic. From its inception, development of P2P communications goes hand in hand with demand for anonymity. Still, we observe that highly secured P2P overlays of today struggle to provide good traffic performance. A novel P2P system proposed in this paper introduces a parallel content exchange architecture separating the anonymization process from the transport function. The anonymity and traffic performance measures of the new system are presented and compared with analogous measures obtained for a classical architecture.

1.1. Related work

The first concept of network anonymity was introduced in the seminal paper of Chaum [6]. 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 exchanged 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 [6].

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) ([4], [8], [13]). Dummies enhance anonymity and allow particular Mixes to flush faster. However additional and "empty" traffic finally delays delivery of parallel user data.

The concept of Mix-net has been used in a wide range of applications such as E-mail [9], Web browsing [3], ISDN [18]. Other solutions [5] seem to play a less important role or (as CROWDS [20]) utilize an 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

1550-445X/08 $25.00 © 2008 IEEE
DOI 10.1109/AINA.2008.117

IEEE computer society

anonymous messages without usage of mixing or public key encryption techniques. P2P overlays as well as other applications usually adapt Mix-net to assure anonymity. Many variants of Mix-net such as Free Haven [10], Tarzan [11] and MorphMix [19] were introduced for P2P. Other system, as Freenet [7] and GNUNet [2], use heuristics with encryption to achieve anonymity. It should be noted that the basic common mechanism used to achieve P2P anonymization is traffic forwarding by a set of middleman nodes.

1.2. Outline of the paper

The rest of the paper is organized as follows. Section 2 describes the system and its anonymization idea. Analytical evaluation of anonymity is presented in Section 3 while traffic performance simulations are described in Section 4. Conclusions and future work are presented in Section 5.

2. P2PRIV overview

P2PRIV 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 exchange 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 Mix-net. A motivation behind the parallel architecture with direct content transport is decrease of the service time while preserving high degree of anonymity.

P2PRIV peers are symmetric and do not include any privileged nodes (supernodes). The P2PRIV specific connection and content exchange architecture utilizes a classical concept of chaining with encryption anonymization, for example Mix-net. It also uses structured lookup system, which can be based on DHT algorithms (distributed hash table, [1]). The cascade anonymization mechanism assures anonymity of an entire P2PRIV control messages exchange, including the communications of distributed content location process. We can distinguish two steps of P2PRIV operation:

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 requestor 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 requestor sends a token with a file id to a randomly chosen peer. Then, the selected peer flips an asymmetric coin to decide whether to forward the 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 control messages of constant length, generated by cloning, can be effectively exchanged by the Mix cascades [12].

Step 2: *Data connection* – transport of the requested content. After a random interval of time and based on the content id received earlier, the copies of the content are directly downloaded by selected (cloned) peers from nodes which store data. One of these nodes is the initial requestor. Files can be looked up by the DHT algorithm. Like the cloning exchange, lookup 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 a content automatically increases the number of its copies. Notice, that in our solution the anonymization process is separated from the content transport, in contrast to classical schemes.

3. Anonymity

Below we will analyze the degree of anonymity of P2PRIV using entropy measurement model ([14], [21]). The CROWDS system, which combines anonymity and performance with simplicity and reputability, will be used as a reference. The anonymity analysis will cover receiver anonymity referred to as an unlinkability of the requested content and the requestor. The anonymous publication process, which can be performed in many different ways using the state of the art methods, is not considered in the proposed scheme and omitted in the analysis. We will not analyze sender anonymity as well. However, when we assume that P2PRIV utilizes the DHT storage, peers of P2PRIV are involved in the process of storing and sending data independently from decisions of they users.

To achieve practical results it is important to assume realistic capabilities of an adversary corresponding to the specific environment of the attack. P2PRIV is 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 user and can utilize the system's provided information at 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, we will analyze local attacks where the adversary can control only a part of the system. We will study an 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 headquarter for a summarizing analysis.

3.1. Passive-static attacks

Static attackers cannot predict which nodes will be randomly selected to form the CC for a particular request anonymization. The adversary can distinguish two sets of peers $\{S_1, S_2\}$ among all N nodes and assign their members probabilities of being the requestor $\{p_1, p_2\}$. S_1 consists of peers which communicate directly with collaborating nodes C during a transport of the requested data, and S_2 are remaining suspected nodes. Then the average CC length is

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

The adversary concludes that the requestor is not among collaborating peers. A number of honest nodes from the cascade are

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

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's consider the most pessimistic scenario, where all of clones download the content from different peers. Then the average number of honest nodes from the CC which communicate with collaborating nodes equals

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

Each of them can be the content requestor with probability

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

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

$$S_2 = N - C - S_1, \tag{5}$$

Each of the S_2 members can be the requestor with probability

$$p_2 = \frac{1 - S_1 p_1}{S_2}. \tag{6}$$

According to the information theory [22], entropy of P2PRIV for this scenario will be

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

Using the entropy measurement model ([14], [21]) the degree of the anonymity (normalized entropy) provided by P2PRIV is

$$d_{PS} = \begin{cases} 0 & p_1 \geq 1 \vee p_2 \geq 1 \\ \dfrac{\frac{C}{N}\log_2\left(p_1^{-1}\right)}{\log_2(N-C)} & p_2 \leq 0 \\ \dfrac{\left(1-\frac{C}{N}\right)\log_2\left(p_2^{-1}\right)}{\log_2(N-C)} & p_1 \leq 0 \\ \dfrac{\frac{C}{N}\log_2\left(p_1^{-1}\right)+\left(1-\frac{C}{N}\right)\log_2\left(p_2^{-1}\right)}{\log_2(N-C)} & p_1 \in (0,1) \wedge p_2 \in (0,1), \end{cases} \tag{8}$$

where

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

The degree of the anonymity describes the uncertainty of the adversary in finding the requestor and takes values from [0,1]. Figure 1 shows the degree of anonymity d as a function of the parameter C.

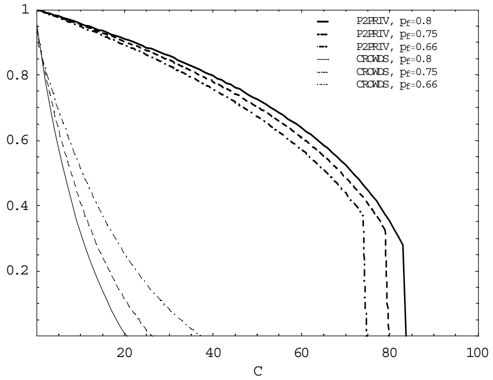


Figure 1. Degree of anonymity for P2PRIV and CROWDS, passive-static attacks, *N* = 100.

3.2. 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 number of honest nodes communicating with collaborating nodes will be

$$S_{1PA} = \frac{C}{N}(n-1)+1, \tag{9}$$

$$S_{1PA} = \frac{(N-C)\big((p_f-2)C+(p_f-1)N\big)}{(p_f-1)N^2}. \tag{10}$$

Analogously to (5) the number of remaining nodes is

$$S_{2PA} = N - C - S_{1PA}, \tag{11}$$

with assigned probability of being the requestor

$$p_{2PA} = \frac{1-S_{1PA}p_1}{S_{2PA}}. \tag{12}$$

Entropy of P2PRIV in this attack scenario is

$$H_{PA} = -S_{1PA}p_1\log_2(p_1) - S_{2PA}p_{2PA}\log_2(p_{2PA}) \tag{13}$$

and the degree of anonymity equals

$$d_{PA} = \begin{cases} 0 & p_1 \geq 1 \vee p_{2PA} \geq 1 \\ \dfrac{\vartheta\log_2\left(p_1^{-1}\right)}{\log_2(N-C)} & p_{2PA} \leq 0 \\ \dfrac{\varsigma\log_2\left(p_{2PA}^{-1}\right)}{\log_2(N-C)} & p_1 \leq 0 \\ \dfrac{\vartheta\log_2\left(p_1^{-1}\right)+\varsigma\log_2\left(p_{2PA}^{-1}\right)}{\log_2(N-C)} & p_1 \in (0,1) \wedge p_{2PA} \in (0,1), \end{cases} \tag{14}$$

where

$$p_{2PA} = \frac{(p_f-1)N\big(N+(p_f-2)C\big)}{(p_f-2)(C-N)\big(C(p_f-2)+N(N-p_fN+p_f-1)\big)},$$

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

Figure 2 shows results for adaptive observation (14).

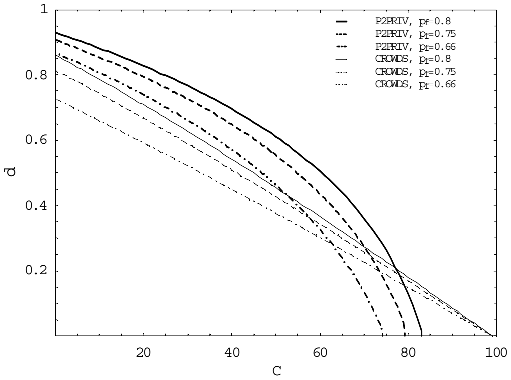


Figure 2. Degree of anonymity for P2PRIV and CROWDS, passive-adaptive attacks, *N* = 100.

3.3. Active attacks

Now we analyze static and adaptive behaviors of active adversary. In traditional anonymous chaining systems active attacks must be more subtle and complex ([15], [16], [17]) than they could be in a 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 parallel architecture, it is important to take into account extreme scenario of cloning interception.

The length of CC before first collaborating node interception P_A can be calculated

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

Among P_A the number of nodes communicating directly with collaborating nodes is

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

hence members of

$$S_{1AS} = \frac{C}{N}n_{AS},\ S_{1AA} = \frac{C}{N}(n_A-1)+1 \tag{17}$$

will have assigned probabilities

$$p_{1A} = \left(P_A - \frac{C}{N}P_A\right)^{-1}. \tag{18}$$

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

$$S_{2AS} = N-C-S_{1AS},\ S_{2AA} = N-C-S_{1AA} \tag{19}$$

will be considered as the requestor with probabilities

$$p_{2AS} = \frac{1-S_{1AS}p_{1A}}{S_{2AS}},\ p_{2AA} = \frac{1-S_{1AA}p_{1A}}{S_{2AA}}. \tag{20}$$

Then degrees of anonymity of P2PRIV for active attacks are

$$d_{AS} = \begin{cases} 0 & p_{1A} \geq 1 \vee p_{2AS} \geq 1 \\ \dfrac{\frac{C}{N}\log_2\left(p_{1A}^{-1}\right)}{\log_2(N-C)} & p_{2AS} \leq 0 \\ \dfrac{\left(1-\frac{C}{N}\right)\log_2\left(p_{2AS}^{-1}\right)}{\log_2(N-C)} & p_{1A} \leq 0 \\ \dfrac{\frac{C}{N}\log_2\left(p_{1A}^{-1}\right)+\left(1-\frac{C}{N}\right)\log_2\left(p_{2AS}^{-1}\right)}{\log_2(N-C)} & p_{1A} \in (0,1) \wedge p_{2AS} \in (0,1), \end{cases} \tag{21}$$

$$d_{AA} = \begin{cases} 0 & p_{1A} \geq 1 \vee p_{2AA} \geq 1 \\ \dfrac{\psi\log_2\left(p_{1A}^{-1}\right)}{\log_2(N-C)} & p_{2AA} \leq 0 \\ \dfrac{\zeta\log_2\left(p_{2AA}^{-1}\right)}{\log_2(N-C)} & p_{1A} \leq 0 \\ \dfrac{\psi\log_2\left(p_{1A}^{-1}\right)+\zeta\log_2\left(p_{2AA}^{-1}\right)}{\log_2(N-C)} & p_{1A} \in (0,1) \wedge p_{2AA} \in (0,1), \end{cases}$$

where

$$p_{1A}=\frac{N^3(p_f(C-N)+N)}{(N-C)\left(p_f{}^2(C-N)^3+N^3+p_fN(2C^2-3NC+N^2)\right)},$$

$$p_{2AS}=\frac{N^3\left(N+p_f(C-N)\right)}{p_f{}^2C(C-N)^3+N^3(C-N^2)+p_fN\left(2C^3-3C^2N-(N-1)CN^2+N^4\right)},$$

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

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

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

Figure 3 shows the degree of anonymity for P2PRIV active attacks.

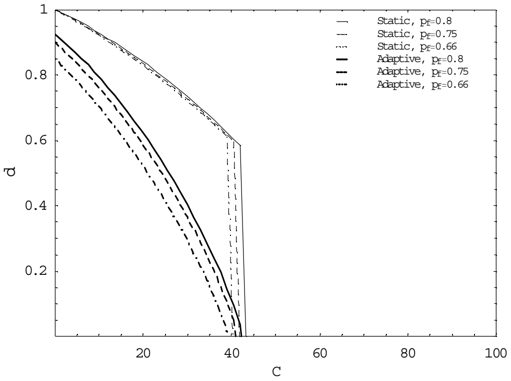


Figure 3. Degree of anonymity for P2PRIV, CC interception active attacks, *N* = 100.

Notice that we cannot directly compare preceding results with CROWDS because of different attack possibilities for cascade and parallel systems. As in previous scenarios, the results show good resistance of P2PRIV for a realistic percentage of colluding nodes.

3.4. Summary of the results

The minimum degree of anonymity depends on system usage and particular users requirements. However as in [14], we restrict acceptable normalized entropy to $d \approx 0.8$. Taking into account adaptive attacks, the degree of anonymity of CROWDS ($p_f = 0.75$ recommended by CROWDS authors) falls below this level for $C = 5\%$ while P2PRIV still retains the proper anonymity level even for pessimistic active-adaptive attacks (see Table 1). 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 a particular system. We should remind that CROWDS does not include mixing and asymmetric encryptions techniques. Notice that P2PRIV proved high robustness under the same conditions.

Table 1. Degree of anonymity for CROWDS and P2PRIV, *C* = 5%.

Attack	CROWDS			P2PRIV		
	p_f=0.66	p_f=0.75	p_f=0.8	p_f=0.66	p_f=0.75	p_f=0.8
Passive-Static	0.69	0.63	0.58	0.97	0.98	0.98
Passive-Adapt.	0.69	0.78	0.82	0.84	0.88	0.91
Active-Adapt.	-	-	-	0.79	0.84	0.87

Using the anonymity measurement model ([14], [21]) 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 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 is proper selection of cascade length (p_f value configuration). We have observed that p_f should not be lower than 0.66. The comparison between CROWDS and 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).

4. Traffic performance

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 distribution overlays a basic performance factor is a mean download time (DT) quoted as a mean time required for a content delivery after a submission of the request by the user. Open P2P environment also requires consideration of unpredictable users migration and traffic bursts after a new publication of a popular content. Besides the mean download time and scalability, we will analyze dynamics of the system in reaction to a new content publication. The scenario will cover request arrival rate and content migration impacts. For the purpose of the complicated dynamic conditions analysis we have created a peer-to-peer simulation environment. The simulator traces tasks of symmetric peers independently. We have used Poisson distribution to model an arrival process. As a reference we have simulated the CROWDS random walk algorithm. CROWDS admits simplifications of the anonymous cascade schemes for better traffic performance. Nodes of CROWDS (called *jondos*) do not mix or delay forwarding content and also do not use asymmetric cryptography. CROWDS system was originally dedicated for Web browsing thus we included the content caching functionality for each forwarding node. Based on the results showed in Table 1 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 for both systems. The average number of peers in a cascade is 4 for $p_f = 0.66$ and reaches value 5 for $p_f = 0.75$ (1). These values correspond also to comparative traffic volume for P2PRIV and CROWDS. Notice that P2PRIV includes one more link for the same cascade length because of its parallel architecture. As an additional reference we will use minimal download rate μ_{Min} marked for simplification as FTP. We assume rather typical for today P2P overlays values of link throughputs and file size. Let average link throughput between peers be B = 512kbps and average file size V = 32MB, then

$$\mu_{Min} = \frac{B}{V} = 0{,}002\left[s^{-1}\right]. \qquad (22)$$

4.1. Download time

Figure 4 shows the mean values and 95% confidence intervals of DT for CROWDS and P2PRIV systems as the function of parameter λ^{-1}. To analyze systems mean download time we have computed six simulation series (with 30 realizations each) starting from the maximum request arrival rate per each node

$$\lambda_{Max} = \frac{\mu_{Min}}{P} = 0{,}0005\left[s^{-1}\right]. \qquad (23)$$

We have found that P2PRIV DT is close to FTP for low amount of requests. Diagram shows the superiority of P2PRIV regardless of the request arrival rate. The content delivery has been at least four times faster than for CROWDS. The observed increase of DT for the close-to-maximum request arrival rate is lower for P2PRIV.

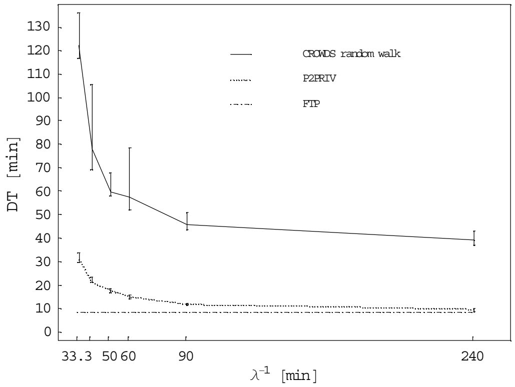
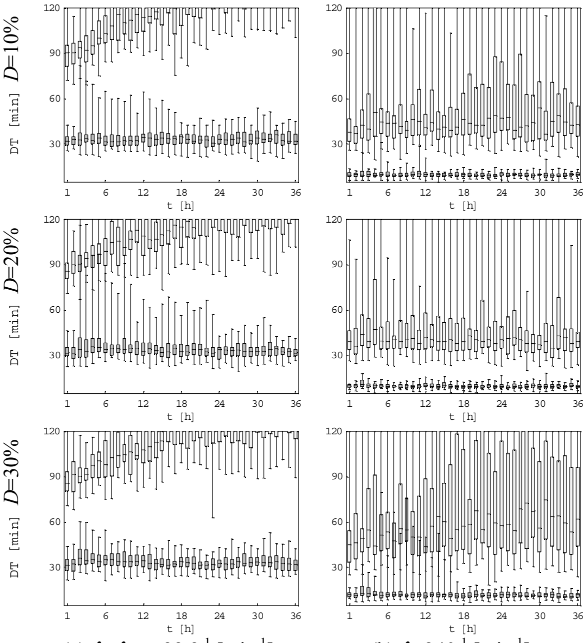


Figure 4. Mean download time for P2PRIV and CROWDS random walk, *N* = 100.

4.2. 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.



(a) $\lambda = \lambda_{Max} = 33{,}3^{-1}$ [min^{-1}] (b) $\lambda = 240^{-1}$ [min^{-1}]

Figure 5. P2PRIV (lower graphs) and CROWDS random walk (upper graphs) reaction to the new content publication, *N* = 100.

Figure 5 shows 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 architecture is more flexible and reacts faster to dynamically changing conditions.

We have found unstable operation of CROWDS random walk for rate λ_{Max} (Fig. 5a). This maximum request arrival rate does not cause instability of P2PRIV. Instead, we have observed permanently increased DT after a new file publication. Both systems have exhibited a stable operation for a low arrival rate and a moderate migration (Fig. 5b, $D = 10\%$ and $D = 20\%$). Performance characteristics of CROWDS under a low request arrival rate are similar to P2PRIV under rate λ_{Max}. For higher dynamics ($D = 30\%$), we have found instability in CROWDS even for a low request arrival rate. P2PRIV noticed only temporary (about 2 hours) peak under the same conditions.

4.3. Scalability

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 earlier traffic analysis of P2PRIV and CROWDS with altered size of simulated networks. Notice that throughout all traffic analysis we have excluded the files lookup problem. Process of finding data in distributed systems significantly impacts scalability, however this approach allows us to revise scalability of pure examined systems.

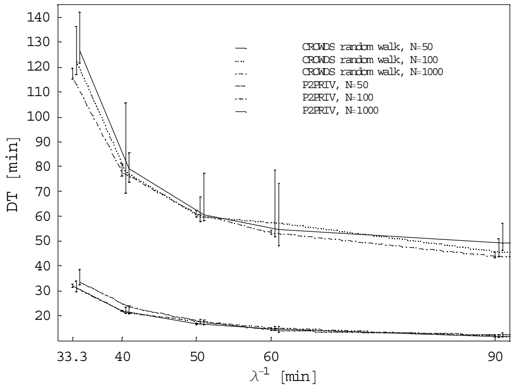
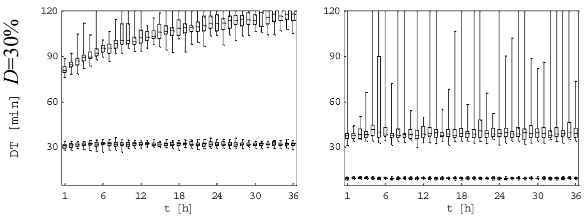


Figure 6. Mean download time for different network sizes (*N*) of P2PRIV and CROWDS random walk.

Figure 6 shows an impact of network size N on the download time for P2PRIV and CROWDS random walk (when looking at Figure 6 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).

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 with networks sizes enlarged to $N = 1000$.



(a) λ=λ_{Max}=33,3^{-1} [min^{-1}] (b) λ=240^{-1} [min^{-1}]

Figure 7. P2PRIV (lower graphs) and CROWDS random walk (upper graphs) reaction to the new content publication, *N* = 1000, *D* = 30%.

Figure 7 illustrates considerably higher systems stability of operation in comparison to results obtained for networks of hundred of nodes (Figure 5). Note that P2PRIV proves strong robustness (meaningless changes of DT) for high dynamics ($D = 30\%$).

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 a better stability than with $N = 100$, but still presents an unstable operation for request arrival rate λ_{Max}. What is more, P2PRIV for $N \approx 1000$ already tolerates high ($D = 30\%$) migration of peers and a content without introduction of any noticeable delays regardless of the request arrival rate.

5. Conclusions

The paper describes an original anonymization system dedicated for peer-to-peer distribution overlay networks, based on a specific connection and content exchange architecture. The proposed parallel content exchange concept enables direct and anonymous data transport between network nodes. We have analyzed the anonymity and the traffic performance provided by our system. We have found that P2PRIV effectively protects user privacy by assuring high degree of anonymity. For a realistic scope of collaboration, P2PRIV anonymity is close to maximum. Moreover, we have found that the proposed system significantly decreases the 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 on dynamically changing conditions such as peers/content migration and traffic bursts introduced by new data publication. P2PRIV scales well and proves high flexibility for large networks.

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 further analysis of impact imposed by anonymous techniques on network traffic performance. Moreover, we will study the distributed hash table interface adjusted to the considered parallel distribution, as well as practical implementation issues. We believe that P2PRIV can satisfy the requirement of private and low latency exchange of large content.

6. References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] D. Chaum, "The Dining Cryptographers Problem: Unconditional Sender and Recipient Untraceability," *Journal of Cryptology* 1/1 (1988). 65-75.

[6] D. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms," *Communications of the ACM*, 4(2), February 1981.

[7] 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.

[8] W. Dai, "Pipenet 1.1," Usenet post. Available: http://www.eskimo.com/~weidai/pipenet.txt, 1996.

[9] 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.

[10] 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.

[11] M. J. Freedman and R. Morris, "Tarzan: A peer-to-peer anonymizing network layer," in *9th ACM Conference on Computer and Communications Security*, Washington, DC, November 2002.

[12] C. Diaz, *Anonymity and Privacy in Electronic Services*, Ph.D. Thesis. Katholieke Univesiteit Leuven, 2005.

[13] C. Diaz, L. Sassaman, and E. Dewitt, "Comparison between two practical mix designs," in *Proceedings of* 9th *European Symposiumon Research in Computer Security* (ESORICS'04), pages 141–159. Springer-Verlag, LNCS 3193, 2004.

[14] 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.

[15] 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.

[16] B. Pfitzmann, "Breaking Efficient Anonymous Channel," in *Advances in Cryptology, Proceedings of EUROCRYPT 1994*, pages 332–340. Springer-Verlag, LNCS 950, 1994.

[17] B. Pfitzmann and A. Pfitzmann, "How to break the direct RSAimplementation of MIXes," in *Advances in Cryptology, Proceedings of EUROCRYPT 1989*, pages 373–381. Springer-Verlag, LNCS 434, 1990.

[18] 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.

[19] 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.

[20] M. K. Reiter, A. D. Rubin, "Crowds: Anonymity for web transactions," *ACM Transactions on Information and System Security*, 1(1), June 1998.

[21] A. Serjantov, 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.

[22] C. E. Shannon, *A Mathematical Theory Of Communication*, the Bell System Technical Journal, Vol. 27, pp. 379–423 and pp. 623–656, 1948.