Anonymous Communication Networks

Protecting Privacy on the Web

Kun Peng
Contents

1 Anonymity in Network Communication ................................. 1
   1.1 Right to Be Anonymous ........................................ 2
   1.2 When We Need to Be Anonymous ............................. 8
   1.3 The Current Situation and Where We Start .................. 10

2 Mix Networks .................................................................... 15
   2.1 Definition of Mix Network ...................................... 16
   2.2 Classification of Mix Networks ................................. 17
      2.2.1 Decryption Chain or Re-encryption ..................... 17
      2.2.2 General or Separate Verification ....................... 19
      2.2.3 Tag Attached to Input .................................... 21
      2.2.4 Summary .................................................... 22
   2.3 Efficient and Secure GMN: a Simple and Efficient Solution ... 22
      2.3.1 The GMN in [92] .......................................... 24
      2.3.2 Analysis and Summary .................................... 26
   2.4 Efficient and Secure SMN: the Most Important Mix Network .. 27
      2.4.1 SMN Employing Multiplicative Homomorphic Encryption Algorithm ......................................................... 32
      2.4.2 SMN Employing Additive Homomorphic Encryption Algorithm ......................................................... 35
   2.5 Grouped Shuffling: a Trade-off to Improve Efficiency of Mix Networks ......................................................... 38
      2.5.1 Group Shuffling in [93] .................................... 39
      2.5.2 Another Grouped Shuffling Protocol to Support Efficient SMN ......................................................... 44
   2.6 Survey and Analysis of SMN ...................................... 49
      2.6.1 The Assumptions Needed in the Mix Networks in [51] and [52] ......................................................... 50
      2.6.2 Security Concerns in [90] and [102] ..................... 53
      2.6.3 Failure in Functionality of the Mix Network by Wikstrom ......................................................... 55
   2.7 Efficiency of SMN: Claim and Reality .......................... 57
      2.7.1 Unfair Usage of Short Exponents ......................... 58
2.7.2 Other Unfair Factors in Efficiency Claim and Comparison 59
2.7.3 Re-evaluating Efficiency 59
2.8 Summary 61

3 Application of Mix Network to E-Voting: a Case Study 63
3.1 Mix Network for E-Voting 63
  3.1.1 The Recent Mix Network Schemes 64
  3.1.2 The New Mix Network 66
  3.1.3 Security and Efficiency Analysis 72
  3.1.4 Extension and Application 76
3.2 Mix Network-Based E-Voting 78
3.3 Security Concerns and Suggested Solution in Practice 81
  3.3.1 Various Relation Attacks and the Existing Countermeasures 83
  3.3.2 Main Idea of the New Countermeasure 88
  3.3.3 A Prototype and Its Drawbacks 91
  3.3.4 Optimization and Security Analysis 93
3.4 Off-Line Pre-Computation in Mix Networks 99
  3.4.1 Security Model of Shuffles 102
  3.4.2 The Basic Design 103
  3.4.3 Off-Line Pre-Computation and Batch Verification 106
  3.4.4 The Final Shuffling Protocol: Modifying Encryption and Further Improving Efficiency 108
  3.4.5 Comparison and Conclusion 111

4 Onion Routing 115
4.1 The Basic Idea 115
4.2 Formal Definition of Security 117
4.3 Second Generation: Tor 124
  4.3.1 Design of Tor 127
  4.3.2 Application of Symmetric Cipher and Diffie-Hellman Key Exchange in Tor: Efficiency Improvement 133

5 Optimisation and Practical Application of Onion Routing 137
5.1 Verifiable TOR: a Verifiable Application of Onion Routing 137
  5.1.1 Preliminaries 138
  5.1.2 Advanced TOR with a Verification Mechanism 139
5.2 Efficiency Improvement Using Diffie-Hellman Chain 142
  5.2.1 A Simple Optimisation of Tor and its Drawback: Simpler but Still Direct Application of Diffie-Hellman Key Exchange 143
  5.2.2 A New and More Advanced Technique: Diffie-Hellman Chain 144
  5.2.3 Analysis and Comparison 147
## Contents

5.2.4 Further Optimisation: Small Compressed Onion Routing ........................................ 148
5.3 A Practical Application: Paid and Anonymous Use of Cloud Softwares ........................................ 151
  5.3.1 How to Obtain Anonymous Usage Permit of a Cloud Software: Anonymous Token ...................... 153
  5.3.2 Two-Way Onion Routing to Support Anonymous Usage of Cloud Software .......................................... 155
  5.3.3 Efficiency Optimisation: Two-Way Onion Routing with Compact Diffie-Hellman Handshakes ...................... 158
  5.3.4 Security Analysis .................................................. 162

6 Practical Systems to Achieve Anonymity: How to Use Them ........................................ 165
  6.1 Installation and Usage of Tor .................................................. 166
    6.1.1 Download and Installation .................................................. 166
    6.1.2 Configuration .................................................. 168
  6.2 Installation and Usage of I2P .................................................. 169
    6.2.1 Download and Installation .................................................. 169
    6.2.2 Configuration .................................................. 173
  6.3 Installation and Usage of JAP/JonDo .................................................. 174
    6.3.1 Download and Installation .................................................. 175
    6.3.2 Configuration .................................................. 176
  6.4 Installation and Usage of QuickSilver ........................................ 178
    6.4.1 Download and Installation .................................................. 178
    6.4.2 Configuration .................................................. 180

References .................................................. 183

Index .................................................. 191
Chapter 1

Anonymity in Network Communication

Anonymity is a fundamental right of a democratic society. In most democratic nations, anonymity is very important and the laws and government regulations have been set up to protect information privacy in every aspect of society including the fast-developing computer and communication network. Users’ anonymity is one of the most important requirements to information privacy. According to the Privacy Act Regulations [3], anonymity is an important property in applications of network communication. Due to wide application of IT and network techniques, user anonymity in network applications has become a critical challenge of security of our society, especially in areas like finance and terrorism and crime prevention. According to the Australian Federal Privacy Commissioner, the number of complaints and enquiries about anonymity remains high [1] and IT and Internet issues have become very important [2] in recent years. There are various motivations for this. Many participants of network applications like e-finance, e-commerce and e-health want to conceal their activity and identity so that their personal privacy can be protected. In some other applications like e-voting the participants do not want their identities to be linked to their activities. In all these applications, anonymity has to be implemented although their requirements for anonymity may be different. A complicating factor is that computers and computer networks make it quite easy to maintain and distribute digital information. Millions of bits can be transferred without error, stored, and analyzed in seconds. People are aware of the potential danger of this information processing power and are afraid of losing control of their personal data in both their private and business lives. Yet, their behavior is often inconsistent with that fear.

Without a satisfactory solution to anonymity of network communication, many network applications cannot gain customers’ confidence and thus cannot replace the traditional counterparts. For example, coins and notes in
traditional cash are indistinguishable and do not reveal consumers’ identities while each e-cash coin is a distinguishable string and may be linked to a consumer. So a customer used to traditional cash and traditional shopping manner may worry about his anonymity and be reluctant to use e-cash and on-line shopping. In another example, in traditional paper-based election systems a scrambled ballot box can guarantee anonymity of the voters. If no electronic scrambling is provided, the voters will be concerned about their anonymity and be unwilling to vote, which is especially serious in a nation with compulsory elections like Australia. In e-health systems, medical information must be transferred, shared and viewed while the patients’ anonymity must be protected. When anonymity is implemented, a good balance must be kept with other security and practical properties like integrity, efficiency and information recovery (in abnormal situations). For example, digital signature is a normal tool to guarantee information integrity, but often violates users’ anonymity. In some network applications, special digital signature techniques like group signatures and ring signatures are employed to achieve integrity and anonymity simultaneously. However, group signatures and ring signatures schemes are usually too inefficient for practical systems. So integrity, anonymity and efficiency are contradictory in some applications. Another contradiction is between personal privacy and national security. Individuals want to reserve anonymity in communication while the government may want to monitor suspicious communication. With a democratic social system facing a serious terrorism threat, a government must achieve a good balance between anonymity and national security. After September 11, the American government began to limit usage of encryption in communication. However, this is not necessarily a good solution. A more flexible method is to design recoverable anonymity such that messages through private communication can be recovered with a court order.

1.1 Right to Be Anonymous

Privacy and anonymity are essential parts of today’s society. In some environments, such as medical treatments and banking, the privacy and anonymity of individuals are protected by law. Other aspects of modern life may not have such clear protections. Today, people are increasingly using computer networks to accomplish activities that were always assumed to be private or anonymous in the non-virtual world. The current protections provided by law and law enforcement are slowly being driven to adapt to this new life in the virtual world.

Computer and telecommunication technologies have generated great concerns regarding the protection of privacy and anonymity on global networks. Many scholars, policymakers and “netizens” have discussed appropriate methods to protect privacy in electronic transactions and to ensure protection of personal identity on networks. International organizations, such as the
Organization for Economic Cooperation and Development (OECD), have also been active in providing relevant principles. As a result, various laws and government policies, industry self-regulation, technological solutions and private contract-based approaches have been suggested as appropriate methods for privacy protection on the Internet, and respective responsibilities among governments, businesses, users and international organizations have been the focus of recent debates.

Anonymity as a mechanism for negotiating social relationships, especially between individuals and external power institutions, has been developed and defined in various ways. The traditional concept of anonymity is often against public and private institutions with ability and power. The main focus was the restriction of access to an individual, which was expected to mitigate the power imbalance between these institutions and individuals. With the development of new technologies that broaden and diversify the collectors and users of personal information, there have been attempts to negotiate the individual’s relationship with the external environment by providing the individual with control over information about himself or herself. But in the interactive network, users’ anonymity cannot be dealt with effectively by current legal and technological measures. In such a situation, what would constitute a factor that could ensure an individual’s self-autonomy and self-governance in relationship with external forces? The greatest difficulty for individuals who become the objects of surveillance in the current technological environment is that users’ identities have become increasingly exposed, while the subjects of surveillance and their activities have become less identifiable. Therefore, the major impetus for the power imbalance between the subjects and objects of surveillance in the network is their differences in identifiability. The right not to be identified should be the most important component of privacy on the Internet and that, by not being bluntly identified, individuals can protect themselves from the potential risk and threat of not easily identifiable entities of surveillance and their activities. This right not to be identified seems to reflect accurately the sentiment of individual users in the networked environment. The sentiment seems to be a more complex or confusing one than wanting to be left unidentified, which may be expressed as “please do contact me and give me benefits, but I still do not want to fully give up my control (but I do not know how to have that control)”. Thus, the issue here is less one of restricting immediate access to the individuals than of the welcoming permission of the access combined with a hope for the restriction of some unknown, unidentifiable, future use of their information and identity. As such, with the new characteristics of interactive network media, the elements that are needed to ensure self-autonomy and self-governance also change. The right to be left unidentified was critical in maintaining one’s autonomy, dignity and self-governance in the context of powerful governments and new communications media such as print. But the networked environment and subsequent sociocultural changes have influenced the relative importance between being left unidentified and actively seeking control. Therefore, the
concept of anonymity changes dramatically from maintaining passive liberty and freedom from an external institution’s interference to allowing limited access without identification of the users when desired.

A network user is often faced with a choice between giving up the benefits and services of the site and providing at least some kind of information, because the use of many sites or services is not allowed unless various personal information is given. Choosing between providing personal information and giving up the information and services that an individual wants from the network is particularly difficult in the current technological environment because, in many cases, it is not known what will happen to the personal information once it is out on the network. Individual users, governments, data-using industries and all other potential users of information are uncertain about whether and how personal information will be used in the future. So even if people do voluntarily provide personal information and make conscious decisions, this decision making process tends to be based on incomplete information and uncertainty. Incomplete information is the reason why network users hesitate to provide personal information, but at the same time do not want to give up their activities on the network because of that unknown, potential risk. When people are required to provide or reveal their information, they hesitate, but if the calculated benefits are greater than potential risks, they give in. Many people later have regrets, especially when their mailboxes are flooded with unsolicited emails or they receive notice of a summons based on some of their activities on the network. Now they realize that not having anonymity on the network is a risky business, and they may begin to conduct self-censorship. Network anonymity becomes an essential part of maintaining the autonomy of network activities. Anonymity has long been discussed as an important element comprising privacy, but network anonymity in particular has become a widely discussed topic recently. The emergence of anonymous remailers that conceal a sender’s identity and location has sparked so much controversy that sometimes laws are enacted to prohibit anonymous messages. Many argue against anonymous communication on the network, focusing on extreme cases such as anonymous threats and libellous messages. However, Froomkin [38] conducted an analysis of the costs and benefits of anonymity, acknowledging that anonymity has both valuable and harmful consequences. The most often cited cost of allowing anonymous communication is the difficulty in detecting illegal and immoral activities [38]. One of the primary reasons for monitoring Internet users is to prohibit infringement of intellectual property rights or to prevent cybercrime. Even some software products contain secret links to servers that allow them to pass along a variety of information regarding when consumers use the software, where, how long and the like, often without their knowledge. It is also possible that the manufacturer could command and control the software or even remotely disable the software if it seems to be working on the “wrong machine”. In such an environment, network anonymity is completely compromised in the name of the often cited justification of piracy control. It has been argued that in cyberspace, the right
to read anonymously that is protected in real space would be totally lost by copyright management systems and fee-based approaches to online activities.

What is often missing in the data about network anonymity is its advantages, as Froomkin \[38\] suggests. Important for considerations of free speech and democracy, anonymity may be the only way for ordinary individuals to protect themselves from governments’ and private corporations’ active use and profiling of their personal information in the networked environment \[38\]. Under the current conditions, the only way for network users to be ensured a minimum amount of privacy on the internet would be to conceal their identity or give a false one. Concealing an individual’s information has been discussed, as well as practised and allowed, relatively more than disguising his or her identity. But if individuals do not want to identify themselves on the network while still being free to pursue all their activities, a logically possible solution is to provide false information. Concealing one’s identity on the internet can be achieved by providing incorrect registration information or using a false identity. Many people do conceal their identity and many commercial books have been published on how to conceal identity on the internet. To have a right to protect themselves from revealing their personal information, consumers should be assisted by the right to lie. But will this right to lie be acknowledged by society or the law? Opinions vary, but those with authority and commercial power tend to say no at present. A number of statutory and regulatory restrictions on anonymous or pseudonymous communication in the United States exist and the constitutional protection of anonymous or pseudonymous communication is not clear, especially in non-political speech \[39\]. The difficulty and costs of concealing and disguising one’s identity on the network also surpass the legal realm. ISPs and websites require correct registration information and if registrants or users are found to have provided a false identity or information, they are at a clear legal, social and cultural disadvantage when damages or disputes occur. Because of these disadvantages resulting from providing false personal information, a right to conceal and disguise one’s identity is not yet widely recognized in a practical sense on the network. This may pose a serious risk to individual privacy because it might be the most effective and sometimes the only practically available way of ensuring privacy and anonymity on the network.

The ways in which people interact with the information environment are changing. The development of telecommunications technology and its convergence with computer technology has generated dramatic changes in the ways in which information can be collected and managed. Understanding the changes in the information environment and what it means to people’s lives and experiences is critical in formulating a new conception of privacy as a changed social condition and proceeding with policymaking endeavors. In the current technological and regulatory scheme, individual privacy in terms of anonymity is less protected on the Internet than in real space. For example, in real space, people usually have a right not to be listed in the telephone directory or to read without always revealing their identity. But on the net-
Anonymous Communication Networks: Protecting Privacy on the Web

work, anonymity cannot be ensured without a practical option to disguise user identity and conceal personal information. In addition, given the voluntary nature of revealing and providing personal information, to apply traditional policy measures that attempt to “provide” privacy “for” individuals by limiting identifiable data users’ information activities to the network environment is ineffective and futile. Users’ self-solutions, enabled by their right to secrecy and deception on the network, are a logical and practical approach to ensure the least amount of network anonymity and privacy needed for personal autonomy. The most pertinent method to achieve this purpose seems to be providing individual network users with some right to engage in Internet activities without being visibly identified and allowing for an active search for network anonymity both legally and technically. Therefore, policy measures for network privacy should focus on ensuring individual users’ searches for anonymity by recognizing the right to be silent about their identities and the right to disguise their identities rather than providing restrictions on easily identifiable external forces and institutions.

It is not difficult to imagine that the solution suggested here would face challenge and opposition. However, many of the concerns are based on social and cultural reflections rather than on legal or logical foundations. Our society seems to favour disclosure over secrecy and speech over silence. The sentiment that concealing data is a bad thing and more than less information is better seems to be a deeply-held social value. Therefore, privacy as a tool to give the right to prevent personal information-sharing may have a natural and inherent disadvantage that can be compared to the obvious distinction between “sunlight” and “shadow”. A close review of society’s customs and practices suggests that secrets and lies are essential elements of society’s function. Every society tolerates, and even respects, some forms of untruth. People tell lies about themselves and their motives and actions, a reality reflected in many cases of secrecy exercised in government and the news media. Wikileaks lists the occasions of government secrecy, including military relations, diplomacy, juvenile proceedings and the identity of information sources, as examples of such sanctioned secrecy. There are many circumstances in which attaining knowledge is considered undesirable and these circumstances are supported by various justifications such as national interest and protection from foreign entities. Ironically, in the case of undercover operations, deception is used legitimately to uncover other deception and secrets. Therefore, deception can be good or bad, much as we distinguish between white and black lies, or between small and big lies. The basis for such distinctions, in the case of government secrets and undercover operations, is that when the benefits outweigh potential harm and risk, secrecy and deception seem to be justified. Also, good purpose and good intention seem to be other justifications for this kind of deception. Thus, the clearly negative connotations associated with deception, disguise, dishonesty and concealment may have a rather relative meaning when applied to specific circumstances. Concealing and deceiving identity to gain anonymity on the current network environment is a tool for protection from unknown harm or
Anonymity in Network Communication

invasion and for the maintenance of individual control over private space. This kind of “defensive lying” has clear benefits and a positive purpose and often is permitted in society. It is argued here that, similarly, activities with the purpose of gaining anonymity on the network deserve social permission.

Furthermore, transparency and certainty are always sought by those who want order and discipline in a society. There is a close connection between deceit and power, as deception is often used by people to seek power or avoid its exertion on them. Lies may function to protect the weak from the strong. In that sense, the right to conceal and disguise one’s identity on the network is a very important tool to influence the distribution of social and political power. In the interactive network, anonymity is the most important element for individuals to have a mechanism of counterbalance against powerful invaders and unknown risks. Network anonymity works as a critical device for mitigating power relations. Therefore, the right to actively seek one’s anonymity so as not to be easily identified, rather than the right to be left alone, deserves to be the centre of the privacy concept on the network. The question that confronts us is this: are we ready and willing to allow a right to lie for the sake of a right to privacy? In this new networked environment, perhaps we cannot have one without the other. Are we willing to sacrifice transparency and bureaucratic efficiency for the sake of network privacy? In the networked environment, we cannot have both. Thus, it would be futile to discuss technical and social methods to achieve privacy when we have not decided whether we are ready to change our fundamental moral concept to achieve it. It is one thing to claim that privacy is an important right, but it is quite another to actually sacrifice other important values for it. Only after we answer these questions can we approach the greater issue of privacy and self-autonomy, which has significant financial, social and human consequences for the future.

In summary, worrying about their privacy and anonymity and taking measures to protect them is a right of the users of network communication although it is sometimes negotiable. At the 2000 RSA Security Conference, privacy law expert Stewart Baker observed that there are four basic rules of privacy and anonymity opinion and behavior today:

• Each individual firmly believes that he or she has a right to complete anonymity in all situations.

• Each individual also firmly believes that other people do not have that right. Messages or letters from anonymous callers are always seen as suspicious, if not outright threatening. Individuals are also concerned that while they would never do so, others might employ anonymity to commit crimes.

• If an individual chooses to give up some private personal information, that information cannot be recovered.

• Most individuals choose to give up private information in exchange for trivial things, such as access to a web site or to register a new purchase.
While disheartening, these attitudes do not constitute a reason to abandon privacy and anonymity research. Rather, it is due to the lack of good alternatives that these rules have evolved.

1.2 When We Need to Be Anonymous

Usage of Tor, one of the most famous anonymous networks, is investigated and analysed in [23]. The investigation provides a deep analysis of the Tor network in the wild, by setting several exit nodes and distributing them worldwide. Taking special cautionary measures to comply with the legal and ethical aspects of users’ privacy, it performed an analysis of the application usage of the Tor network through a deep packet inspection (as opposite to a simple port-based classification), and show that most of the traffic exchanged through Tor is undesirable BitTorrent traffic. The study also observed an important fraction of “unknown” traffic. It reveals that the vast majority of this traffic is actually encrypted BitTorrent traffic. The analysis shows then that the BitTorrent traffic on top of Tor accounts for much more traffic size than commonly believed. It also studied the HTTP and BitTorrent usage over Tor and compared Tor user behaviors to those of typical Internet users.

As explained in [23] and will be detailed later, Tor is a circuit-based low-latency anonymous communication service. Its main design goals are to prevent attackers from linking communication partners and from linking multiple communications to or from a single user. Tor relies on a distributed overlay network and onion routing to anonymize TCP-based applications like web browsing, secure shell, or peer-to-peer communications. When a client wants to communicate with a server via Tor, he selects \( n \) nodes of the Tor system and builds a circuit using those selected nodes. Messages are then encrypted \( n \) times using the following onion encryption scheme: messages are first encrypted with the key shared with the last node (called the exit node of the circuit) and from node \( n-1 \) to node \( 1 \). As a result of this onion routing, each intermediate node only knows its predecessor and successor, but no other nodes of the circuit. In addition, the onion encryption ensures that only the last node is able to recover the original message. A Tor client typically uses multiple simultaneous circuits. As a result, all the streams of a user are multiplexed over these circuits. For example, a BitTorrent user can use one of the circuits for his connections to the tracker and other circuits for his connections to the peers. Finally, some ISP may block access to Tor network by filtering the IP addresses of Tor nodes. To circumvent this censorship, the Tor project has created the so-called bridges. These are new types of Tor routers that are not listed in the main Tor directory, and hence cannot be blocked. Tor restricts access to this list and gives a small subset (3 bridges IP addresses) per unique requester IP for a fixed period of time.

Tor has gained in popularity through the years, and its related traffic has certainly evolved. So it is interesting to analyse its traffic through deep packet
Anonymity in Network Communication

inspection, and not through a simple port-based classification. This provides more accurate classification of the traffic that is exchanged through the Tor network. In this way, a clear picture can be obtained about what applications are typically used on top of Tor. The statistics in [23] show that among all the Tor traffic, HTTP takes 34.3%, BitTorrent takes 25.3%, SSL takes 1%, other P2P/file-sharing than BitTorrent take 0.26%, insecure services like ftp, telnet and email take 1.3%, instant massaging takes 1.2%, other recognised protocols take 3.4%, and 29% packets are not recognised.

It is noticed in [23] that a significant part of the traffic is still unclassified. It represents more than 25% of the entire volume. This behavior suggests that such traffic likely belongs to any of the P2P protocols. To verify this, Chaabouni analyzed the distribution of destination ports for those unclassified connections. They observed that destination ports were uniformly distributed, which can lead to a belief that such traffic is BitTorrent traffic. In fact, to avoid port-based detection, BitTorrent clients choose a random port at installation time. This results in uniformly distributed ports. Although these proofs suggest BitTorrent to be responsible for this traffic, the analysis does not recognize it. This is most likely because this traffic is encrypted and thus unrecognizable. A step further is then to compute the entropy of sample data. The computed high entropy value confirmed that this data is either encrypted or compressed.

So most traffic in anonymous communication on Tor networks belongs to BitTorrent. A torrent is a set of peers sharing the same content. To join a torrent, a user sends an announce message to the tracker that maintains the list of all peers in that torrent. The announcement is an HTTP GET message containing the identifier of the requested torrent. Such identifier is known as the infohash of the torrent and is unique. Once the tracker receives the announce message for a specific torrent identified by the infohash, it selects a random subset of peers in that torrent and returns the endpoints (the IP and port of a peer) of those peers. Then, the user establishes a TCP connection and sends a handshake message to each peer. Finally, popular BitTorrent clients, e.g., µTorrent and Vuze, configure SOCKS proxies and give the option to use the proxies for connections to the tracker, to the peers, or both. Therefore, a BitTorrent client can use Tor, configuring the Tor interface as a SOCKS proxy, for communication to the tracker or the peers independently. The user can then decide to connect to the tracker via Tor, but have a direct connection to peers in order not to have performance penalty.

HTTP protocols take a large part of the anonymous traffic as well. The analysis in [23] shows that the HTTP protocol carries a wide spectrum of data going from simple text to rich media such as images and video. Furthermore, a large variety of applications are embedded into browsers to enrich the end user environment. Analyzing this data allows the readers to have a more comprehensive view of how the web is used on top of Tor. More precisely, among all the packets transported on Tor networks, 31.7% are for pictures, 27.9% are for text/html, 18% are for applications,
11.1% are for flashes, and 8.9% are for other services. The most significant content is, as expected, images and text/html. Surprisingly, applications (e.g., rar and zip) content represents a significant proportion of the observed traffic. In addition, it is noticed that 6% of the entire traffic is originating from Direct Download Link (DDL). This can be explained by the fact that some users may have switched from P2P networks known to be heavily monitored to DDL-based content, much harder to control. This behaviour switching has already been noticed in residential broadband Internet. On the other hand, flash and video usage representing 13.5% of the observed content, shows that the latency induced by the Tor relaying is not an actual brake for browsing Web 2.0.

An interesting question is which webs are most frequently visited through Tor. The analysis in [23] shows that 14.45% are search engines or portals, 11.50% are pornography webs, 11.45% are computers/Internet webs, 9.52% are social networking webs, 2.26% are blogs/web communication sites, 1.82% are streaming-media/MP3 webs, 1.66% are webs providing software downloading, 0.3% are hacking webs, 0.18% are political webs, 0.15% are illegal/questionable webs, and 0.06% are illegal-drug webs. Another interesting question is who uses Tor. The analysis in [23] shows that more than 70% of the clients were originating from only 10 countries. Germany and U.S represent more than a quarter of the clients. Such a high ratio may be explained by Internet demographics (especially the high Internet penetration in these countries) on one hand, and also by the increase and strengthening of anti-piracy and copyright laws during the past few years. The concentration of Tor clients among this small subset of countries and in particular, the absence of politically-sensitive countries among the top countries of the observed clients coupled with the announcements of the Tor project that bridges are still in their infancy and not yet often used by clients may be good indicators of the common usage of Tor. Eastern European nations (Poland, Romania and Russia) represent nearly 20% of the Tor clients and Chinese clients correspond to 5.8% of overall clients.

1.3 The Current Situation and Where We Start

In practice, achieving anonymity in network communication is not an easy task. According to [66], the challenge of private information may be simply stated: you can restrict it totally or you give it all:

- If privacy was guaranteed then every individual would have the capability to act anonymously in the virtual world. Therefore, the individuals could use the great power of the information processing systems for their own benefits. The danger here is that everyone, including criminals and terrorists, could use that power for their activities.
Anonymity in Network Communication

- If law enforcement agencies are given the capabilities to undo anonymity, then they can undo it for all people and have this great power to observe everyone in greatest detail.

- There are also “complete non-anonymity” proponents. In his “Transparent Society” [18], Brin imagines a world in which all aspects of everyone’s public life are subject to viewing by web cameras. He then proposed two variants: one with the government and law enforcement being the only ones who can view the images from the omnipresent cameras, and another where everyone, individuals as well as government, will have full, immediate access to any of the images via the Internet.

There may be no general solution to this all-or-nothing problem that satisfies everyone. Even so, we should begin to address this problem now. We suggest that initially focusing on application-specific anonymity techniques may help. For instance, not all people should have the same ability to use anonymity techniques, just as not all people have access to prescription drugs (restriction to some persons). Similarly, absolute anonymity should be guaranteed in electronic elections (restriction of application). Thus, we have the real challenge of how to achieve practical privacy through anonymity, i.e., while maintaining acceptable network performance. The rest of this chapter will address this technical question and will present the important results of the last twenty-plus years of research in this area.

The first attempt was made by Pfitzmann, whose definition and classification of privacy and anonymity from a technical angle in German were translated by Kesdogana and Palme [66] as follows.

- Anonymity is the state of being not identifiable within a set of subjects, known as the anonymity set. Anonymity in communications can be further distinguished as sender and recipient anonymity.

- Unobservability is the state of an item of interest being indistinguishable from any other item of interest.

- Unlinkability of two or more items or actions means that these items are no more and no less related than they were previously (attacker gains no information).

Unobservability can be reduced to a set of data items, senders or receivers. For example, a concrete requirement for messages is that each message cannot be linked to any potential sender or receiver from the set. At a higher level, relationship unobservability requires that it is not discernable whether anything is sent from a set of potential senders to a set of potential recipients. A definition of anonymity is incomplete if an attacker model (opponent model) is not specified. The attacker model describes the demands placed on the anonymity techniques and is also for the evaluation and comparison of proposed solutions. In general, a direct relation exists between the strength...
of the attacker model and the quality of the protection provided by a given solution. To guarantee formal and reliable anonymity, it is needed to assume that anonymity is to be provided in the presence of a powerful attacker. Thus, the capabilities of an attacker $A$ may vary. For the sake of simplicity, it is assumed here that the cryptography used is unbreakable. However, it is good to keep in mind that it is inappropriate to provide or demand more anonymity protection than the underlying cryptography can provide. Attackers may be classified as follows according to [66].

A1 Passive attacker. Attacker can observe all communication links.

A2 Passive attacker with sending capabilities. The A2 attacker is not much stronger than A1, yet the A2 attacker poses a bigger threat than A3 because it is by definition undetectable. Attacker may take part in the anonymity technique (i.e., attacker can send messages) if participation has not been explicitly forbidden for him.

A3 Active attacker. Attacker can control all communication links, switches, etc. and can attack all messages with delete, replay, and send, or delay actions.

By choosing the powerful attacker model it follows that a single transmission by a single person can be neither anonymous nor unobservable. The omnipresent attacker can observe the sender of a message (the sending act) and follow the message to the receiver, thereby detecting the communication relation without needing to read the content of the message. Hence, it is straightforward to notice that anonymity techniques require additional traffic, called cover traffic. Having the additional traffic, it is feasible to employ an embedding function for the subject traffic in order to confuse the adversary and conceal the particular sender, recipient, and their communication relationship. The following results are needed.

- Group function (cover traffic). Since single transmissions are observable in the network, additional traffic is organized by the group function. It is essential that the attacker not be able to gain control over this additional traffic.

- Embedding function. The traffic generated by a particular user must be efficiently and untraceably embedded into the cover traffic.

If the attacker can control the cover traffic, all anonymity is lost. To avoid this and other attacks wherein the attacker exerts some control over the cover traffic, the CUVE requirements must be met as explained in [99].

- Completeness. All users can verify that their messages have been correctly sent, received, or transmitted.
• Un-reusability. Within any given session, 55A is the transmission of a packet using an embedding function and sent in the presence of cover traffic. No user can participate more than an allowed number of times.

• Verifiability. An adversary cannot change another user’s message without being discovered by the system.

• Eligibility. Only authorized users can participate in a given session.

The embedding function has to be applied in an environment. The only exception to this statement is in the case of recipient anonymity for broadcasted messages, where there is no room for any attackers. This simplifying assumption is needed in order to focus only on the anonymity techniques. This is similar to the case of encryption, wherein the application of the encryption algorithm has to be performed in a trusted environment. Otherwise, security cannot be guaranteed, since the attacker will have full knowledge of the process. However, unlike encryption, the users depend on other assumptions. In general, encryption is a unilateral function and anonymity a multilateral function, i.e., a user can encrypt his electronic diary on his machine but to act anonymously he always needs additional users (due to the requirement of cover traffic). Thus, while the security point of view recommends having the embedding function applied only within the trusted domains of the sender and receiver (the end-to-end solution), from a practical point of view it may be preferable for the group function to use central Trusted Third Parties (TTPs).

The desired goal of anonymity is only achievable if at least two honest participants work together. In general, it is always appropriate to assume that \( n > 1 \) users participate in the application of a given anonymity technique. A typical circumstance is where the network itself is unsecure and the trusted domains of the users are secure. TTPs sit between these two trusted end points and must fulfill some special trust requirements:

• A single point of outside trust should be avoided; the trust has to be distributed equally over all used \( N \) TTPs.

• TTPs should be as transparent to the user as possible, i.e., the correct functionality of the environment should be controlled by the user.

• TTPs should be independently designed and produced and have independent operators.

If \( (n - 1) \) of the users providing the cover traffic are dishonest, then obviously the technique cannot provide any protection. Many of the previous works in anonymity neglect to consider the corrupt user, and make the assumption that all participating users are honest. Certainly in open environments like the Internet, the attacker could be an alliance of \( (n - 1) \) dishonest persons. Unfortunately, there is no technical means to test the honesty of people and,
thus, no way of providing perfect protection. Note that using one-time pad cryptography can provide perfect security in the technical sense.

Anonymity protocols belong to the family of group communication protocols. Typically, networks (e.g., the Internet) are not designed to handle the huge amount of traffic produced by these protocols. In general, the situation is even worse for anonymity techniques since they depend upon dummy messages. To evaluate the performance of anonymity techniques, it may be necessary to abstract from a concrete network structure and consider it as a black box. In this network model, it is appropriate to assume that the cost of sending a message without an anonymity-providing technique is one and the transmission time is also one.

Since anonymity techniques need cover traffic in order to provide unobservable communication, it is important to maximize the number of real messages sent in a session. Consider a technique with a group function that handles \( n > 1 \) messages from \( n \) distinct users. Suppose \( k \) of the messages are real and \( m \) are dummy \((m = n - k)\). With anonymity, something more (bits or energy) has to be provided to meet the goal. So system effectiveness can be defined as \( k/n \). It is always less than 1, since there cannot be more real messages sent than all of the messages sent. If the message is sent via several TTPs, then these additional reroutings count as re-sending all of the messages again.

Clearly, it is desirable to have an anonymity technique with high system effectiveness approximately equal to one. Assuming that such a technique exists, it would mean that there exists \( n \) people who want to send \( n \) real messages. Since there are not always \( n \) people who want to send real messages within a given time period, the technique has to wait for enough real messages before beginning. The waiting time interval could be chosen with high probability that \( n \) people will want to send something. Thus, the people may have to wait until the specified time interval has elapsed. This time cost can be measured as time efficiency \( t \). Additionally, if the message is sent over several TTPs, that time has to be counted as well.