A TAXONOMY FOR AND ANALYSIS OF ANONYMOUS COMMUNICATIONS NETWORKS

DISSERTATION

Douglas Kelly, GG-14

AFIT/DCS/ENG/09-08

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

The views expressed in this dissertation are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

A TAXONOMY FOR AND ANALYSIS OF ANONYMOUS COMMUNICATIONS NETWORKS

DISSERTATION

Presented to the Faculty

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

Douglas J. Kelly, BS, MS, MBA

March 2009

AFIT/DCS/ENG/09-08

A TAXONOMY FOR AND ANALYSIS OF ANONYMOUS COMMUNICATIONS NETWORKS

DISSERTATION

Douglas J. Kelly, BS, MS, MBA

Approved:

//SIGNED//	16 Mar 09
Dr. Richard A. Raines (Chairman)	Date
//SIGNED//	16 Mar 09
Dr. Barry E. Mullins (Member)	Date
//SIGNED//	16 Mar 09
Dr. Rusty O. Baldwin (Member)	Date
//SIGNED//	16 Mar 09
Dr. Michael R. Grimaila (Member)	Date

Accepted:

//SIGNED//	18 Mar 09
Dr. M. U. Thomas Dean, Graduate School of Engineering and Management	Date

Abstract

Any entity operating in cyberspace is susceptible to debilitating attacks. With cyber attacks intended to gather intelligence and disrupt communications rapidly replacing the threat of conventional and nuclear attacks, a new age of warfare is at hand. In 2003, the United States acknowledged that the speed and anonymity of cyber attacks makes distinguishing among the actions of terrorists, criminals, and nation states difficult. Even President Obama's Cybersecurity Chief-elect recognizes the challenge of increasingly sophisticated cyber attacks. Now through April 2009, the White House is reviewing federal cyber initiatives to protect US citizen privacy rights. Indeed, the rising quantity and ubiquity of new surveillance technologies in cyberspace enables instant, undetectable, and unsolicited information collection about entities. Hence, anonymity and privacy are becoming increasingly important issues. Anonymization enables entities to protect their data and systems from a diverse set of cyber attacks and preserves privacy.

This research provides a systematic analysis of anonymity degradation, preservation and elimination in cyberspace to enhance the security of information assets. This includes discovery/obfuscation of identities and actions of/from potential adversaries. First, novel taxonomies are developed for classifying and comparing well-established anonymous networking protocols. These expand the classical definition of anonymity and capture the peer-to-peer and mobile ad hoc anonymous protocol family relationships. Second, a unique synthesis of state-of-the-art anonymity metrics is provided. This significantly aids an entity's ability to reliably measure changing anonymity levels; thereby, increasing their ability to defend against cyber attacks. Finally, a novel epistemic-based mathematical model is created to characterize how an adversary reasons with knowledge to degrade anonymity. This offers multiple anonymity property representations and well-defined logical proofs to ensure the accuracy and correctness of current and future anonymous network protocol design.

Acknowledgments

Special thanks go to my advisor and committee members who were extremely supportive during some very difficult times while pursuing this degree. Succinctly, my committee members and agency executive support enabled me to grow professionally and contribute eight academic papers in my field.

Most PhD candidates deservedly praise their wives and/or significant others for their tremendous love, support, and understanding while advancing their career. I am able to attest to the significant advantage such support provides for I had no such personal support to draw strength from. Nonetheless, I acknowledge God for giving me the fortitude to continue to make progress in my studies and ultimately become successful during this challenging but rewarding academic experience.

Table of Contents

List of Figures

List of Tables

List of Acronyms

AA	Anonymous Agent
ACS	Anonymous Communications System
AC	Adversary Capability
ACM	Association for Computing Machinery
AFIT	Air Force Institute of Technology
AP	Anonymity Property
AODV	Ad hoc On-demand Distant Vector routing protocol
AS	Anonymity Set
AZ	Anonymity Zone
BNF	Backus Normal Form
BSS	Basic Service Set
CA	Communications Anonymity
CCS	Calculus of Communicating Systems
CIA	Central Intelligence Agency
CSP	Communicating Sequential Processes
CT	Cubic Taxonomy
CUV	Conditionally Universally Verifiable
CWSN	Clustered Wireless Sensor Network
D-AZ	Destination Anonymity Zone
DIA	Defense Intelligence Agency
DoD	Department of Defense
DoS	Denial of Service
ESS	Extended Service Set
FBI	Federal Bureau of Investigation
GA	Group Anonymity
GCA	Group Communications Anonymity
GPS	Navistar Global Positioning System
HTTP	Hyper Text Transport Protocol
IBSS	Independent Basic Service Set
IEEE	Institute for Electrical and Electronics Engineers

IOI	Items of Interest
IP	Internet Protocol
IRC	Internet Relay Chat
LA	Location Anonymity
LBS	Location-Based Services
MA	Mutual Anonymity
MAC	Media Access Control
MANET	Mobile Ad hoc Network
MO	Middle Outsider
MV	Mix Verifiable
NA	Non-anonymous Agent
NASA	National Aeronautics and Space Administration
NGA	National Geospatial Agency
NSA	National Security Agency
NT	Network Type
PALM	Possibilisitic Anonymity Logical Model
PAN	Personal Area Network
P2P	Peer-to-Peer
PGP	Pretty Good Privacy
QI	Quasi-Identifier
QoS	Quality-of-Service
RA	Receiver Anonymity
RFC	Request For Comments
RREQ	Route Request
RREP	Route Reply
RSA	Rivest, Shamir, Adleman encryption algorithm
RT	Research Table
SA	Sender Anonymity
SSH	Secure Shell
SMTP	Simple Mail Transfer Protocol
SV	Sender Verifiable

TCP	Transport Control Protocol
TLS	Transport Layer Security
UDP	User Datagram Protocol
UV	Universally Verifiable
VoIP	Voice over IP

A TAXONOMY FOR AND ANALYSIS OF ANONYMOUS COMMUNICATIONS NETWORKS

I. Introduction

This chapter introduces the current and historical issues related to anonymity in cyberspace. In Section 1.0, a brief history of anonymity is outlined. The problems and available solutions for anonymous communications are described in Section 1.1. The research objectives, in Section 1.2, are provided. The subsequent assumptions/limitations and implications of this research are given in Sections 1.3 and 1.4, respectively. Lastly, Section 1.5 summarizes this chapter.

1.0 Background

Anonymity derives from the Greek word ανωνυμία (*anonumos*), meaning nameless, and is the state of being unknown or unacknowledged. Thus, anonymity connotes an inability to link a name to a specific set of actions. Also, the term cyberspace, from the Greek work Κυβερνήτης, describes anything associated with computers, information technology, the Internet and the diverse Internet culture. In societies throughout history, anonymity has always been a pervasive, dichotomous issue. For instance, millionaires differ on the value of anonymity in philanthropic giving [Sch94] and the sociological debate about anonymity [Hum98, Mar99] is not new. Some believe anonymity is essential in protecting privacy and freedom of expression while others believe anonymity is superfluous and only encourages the propagation of dubious dogma as well as abusive, illegal activity.

In the boundless digital world and global society of the Internet, recently dubbed cyberspace, anonymity is also an increasingly important issue [AbF01, Nis97, Nis98, Nis99, Rig95, Wal01, Woo06]. The Internet was first and foremost designed to share information, not protect user privacy. During the 1970s, when military and academic research organizations were the primary users, this was acceptable as the nascent Internet was a relatively anonymous network anyway. With the rapid growth of the Internet as a means of communication and information dissemination, concerns about Internet privacy and security are escalating. In 1980's, Chaum began work on untraceable e-mail [Cha81]. Technology emerged to protect user privacy on very sensitive, controversial newsgroups, such as Dave Mack's for alt.sex.bondage [Rig95] and the anonymous dining cryptographer problem [Cha88]. Then in 1992, Cyberpunk [Pas00] introduced anonymous e-mail. In 1997, nine privacy experts recognized as a major concern the pursuit of perfect identity with biometrics and DNA and converting anonymous transactions to identifiable ones [Ven97]. Furthermore, the increase of new surveillance technologies such as computer matching and profiling, video cameras, and electronic location monitoring enable information collection without an individual's explicit knowledge or consent provides future research issues [Mar01]. The Internet has become an amazingly powerful surveillance tool: anyone has the capability to spy on anyone else [DiP04]. Today, in an effort to prevent cyberstalking, posting annoying Web messages or sending anonymous e-mails has been deemed a federal crime in the United States resulting in stiff fines and two years in prison [Mcc06].

1.1 Problem Statement

Any entity operating in cyberspace is susceptible to debilitating cyber attacks. As part of the National Strategy to Secure Cyberspace in 2003, the United States acknowledged that the speed and anonymity of cyber attacks makes distinguishing among the actions of terrorists, criminals, and nation states difficult [BuG03]. With the ability to gather intelligence and disrupt communications in cyberspace rapidly replacing the threat of conventional and nuclear warfare, a new age of warfare is upon us. President's Obama's Cybersecurity Chief nominee is reviewing federal cyber initiatives and recognizes the challenge of the increasing sophistication of cyber attacks. Now through end-of-April 2009, the National and Homeland Security Councils are conducting a review of federal cyber initiatives' to stop and deter cyber attacks and protect the privacy rights of our US citizens. As millions of individuals and organizations become subject to more and more online monitoring, cataloging, and recording, the economic and security risks as well as potential threats from adversaries becomes greater and greater. Indeed, today's Internet is an incredibly effective, uncontrolled weapon for eavesdropping and spying. Therefore, anonymity and privacy are increasingly important issues. Web-browsing, message-sending, and file-sharing are three key activities where individuals and organizations may prefer a certain degree of anonymity in ubiquitous distributed environments [GuF04]. For a typical Internet user, anonymity means using all available Internet services while keeping an identity or Internet Protocol (IP) address hidden from an adversary. Pure anonymity prevents the adversary from discovering a user's true IP address. Pseudo-anonymity hides the IP address from adversaries but securely stores the IP address to make the user reachable by non-adversarial users.

A number of Anonymous Communications Systems (ACS) have been developed to achieve anonymity including Crowds [RmR98], Herbivore [GoR02], Mixminion [DaR03], Tor [DiM04], and WonGoo [LuF05]. These technologies offer varying degrees of anonymity to protect the user's identity and provide privacy over a communications system. The effectiveness of anonymous protocols depends heavily on a number of factors including: the number of anonymous users; how messages are routed; adversary knowledge and ability; and other environmental factors for both the Internet [GuF02, Kes01] and mobile ad hoc networks [KoL07, LiK05]. The ability to comparatively and quantitatively analyze these anonymity protocols and anonymity services to better understand how anonymity is lost, maintained or improved during an attack is an area of open research. Furthermore, developing novel conceptual and mathematical frameworks for specifying, designing and verifying anonymity properties and protocols is an area ripe for adding to the body of knowledge.

1.2 Research Objectives

The primary research objectives are to develop a novel taxonomy, appropriate anonymity metrics, and a mathematical model to systematically analyze the anonymity properties of anonymous communications networks. Three distinct sub-objectives are to be realized. First, a creative conceptual taxonomy for analyzing anonymity in communications networks is developed. Extensive survey paper(s) on burgeoning anonymity issues such as location anonymity in mobile ad-hoc networks and multicast or group anonymity are examples of literature contributions. Second, to fully comprehend the nontrivial aspects of defining, measuring and preserving anonymity in a variety of

situations, a number of anonymity metrics and their advantages and disadvantages are analyzed. Finally, a modified formal mathematical framework for verifying anonymity properties and reasoning about the enhancement, preservation, degradation and elimination of anonymity in communications networks is explored. The results are significant and motivate even more anonymity research in application domains such as Voice over Internet Protocol (VoIP), video teleconferencing, and mobile ad-hoc networks (MANETs).

1.3 Assumptions/Limitations

The research assumptions vary for each sub-objective. Without loss of generality, for the anonymous taxonomy, a clear distinction between wired and wireless anonymous networks is assumed even though the Internet is becoming an increasingly heterogeneous networked environment. This is justified because the requirements for providing anonymity in highly mobile and wireless networks is unique enough to warrant such a separation as the literature clearly indicates in the next chapter. One key limitation is the difference between link, network and application layer anonymity is not specifically modelled; however, this would make an excellent extension to this research. Also, only three key categorizations are highlighted in the taxonomy. Whereas other categorizations such as verifiability type, anonymization technique, or application domain may be equally valid choices, the three selected complement and even extend the current, albeit limited, taxonomy research. However, unlike other taxonomies or proposed protocols, no adversary assumptions are made. The adversary capabilities are included as part of the taxonomy. For the anonymity metrics, each makes their own assumptions about the

underlying anonymous protocol and/or anonymization technique/algorithm. This is why a single anonymity metric is not applicable to all situations; hence, the need for more appropriate, robust metrics. Finally, the key assumptions of adversarial logical omniscience and no temporal and dynamic capability are made in the formal model. Some of these assumptions can be relaxed if the theorem-proving or model checking software used to solve NP-hard problems is available to facilitate and expedite anonymity-based deductive proofs or satisfiable decision procedures; however, no such software was used. These limitations are discussed more in later chapters; but, again, removing such assumptions is highly encouraged as an extension of this research.

1.4 Implications

This research produced an innovative taxonomy, anonymity metric comparison, and intuitive rigorous formal model to systematically define, quantify, and analyze how anonymity is degraded, preserved or enhanced in existing and proposed wired and wireless anonymous communications networks. These synergistic results accentuate the significance and subtlety of anonymity and contribute to future anonymous protocol design and development across one or more application domains.

1.5 Summary

This chapter introduced anonymity, provided a brief motivation for the necessity of the research, delineated the research objectives as well as assumptions, limitations and implications, and the positive impact this research will have on future anonymity research.

Chapter 2 reviews the pertinent prevailing literature on anonymity history, anonymity nomenclature, wired and wireless anonymous networking protocols, anonymity quantification and anonymity formalization. Chapter 3 provides a discussion on this anonymity research and methodology. Chapter 4 provides analysis and results of the anonymous network taxonomy research. A synthesis of existing and proposed anonymity metrics is examined in Chapter 5. The analysis and results of the formal adversary anonymity reasoning model in Chapter 6 is described. Chapter 7 summarizes the contributions of this research and recommends future research to extend the results presented herein.

II. Literature Review

2.0 Chapter Overview

This chapter provides an extensive literature review covering the state-of-the-art concepts in anonymous communications systems. The background of Section 2.1 offers definitions for and historical accounts of privacy, identity, anonymity, pseudonymity, and reputation. The advantages and disadvantages of anonymity and an example reputation system are described. The anonymity properties, the adversary, the attacks, and mix technology are examined in the nomenclature Section 2.2. In Section 2.3, the explanation of extant and prospective wired and wireless anonymous networking protocols is given. Thereafter, ten different ways to quantify anonymity are discussed in Section 2.4. Section 2.5 introduces the basic concepts in formally analyzing anonymous systems. Thereafter, epistemic-based formal methods are explored in Section 2.6. The well established theoretical approach of using process calculi to model systems in computer science is discussed in Section 2.7. The functional framework of Section 2.8 is covered and Section 2.9 concludes this chapter.

2.1 Background

This section covers the history of and introduces the terminology of privacy, identity, anonymity, pseudonymity and reputation. The advantages and disadvantages of anonymity and the eBay reputation system are also highlighted.

2.1.1 Privacy.

The desire for privacy motivates much of the research into anonymity systems. Even Aristotle in 384 to 327 B.C. had a keen interest in privacy when he differentiated between two spheres of life: public (*polis*, city) and private (*oikos*, home). Today, the derived English words *politics* and *economics* still embody the same spirit of separation [WrS05]. However, Aristotle's interest in privacy was neither the first nor last.

With the adoption of the Justices of the Peace Act in 1391 under the reign of Edward III, privacy has been a key part of British law [Mic61]. The act outlawed peeping Toms and eavesdroppers who invade the privacy of others [Ano06]. Nonetheless, privacy as an individual right has only begun to be widely acknowledged in the past 150 years [WrS05].

United States Supreme Court Justice Louis Brandeis and lawyer Samuel Warren proposed that the right to privacy [WsB90] as a natural extension of the individual right to liberty. Liberty as a right had initially been understood with respect to preventing physical assault, but as newer business models and media coverage started to significantly affect society, intrusion into private lives for public consumption has became of concern to many. The ideal of liberty was extended to include unfair intervention into aspects of a person's life that might be embarrassing or dangerous if publicized. They sought "a general right to privacy for thoughts, emotions and sensations" but lost their first major courtroom case by a four-to-three decision at the New York Court of Appeals in Roberson v. Rochester Folding Box Co. in 1902 [PaO02, Unk12]. In reference to earlier work by a Michigan Supreme Court Justice, privacy was

defined as “the right to be let alone” [Cra76]. This concept is still fundamental to almost all definitions of personal privacy.

Serious interest in privacy, however, appears to have begun only in the second half of the twentieth century [WrS05]. The modern concept of privacy at an international level is found in the 1948 United Nations Universal Declaration of Human Rights, which protects territorial and communications privacy in its twelfth article [Com05, Uni97]. Similarly, article 17 of the International Covenant on Civil and Political Rights recognize privacy as a basic human right [Ano06]. Both the European Union [Ano06] and the United States Department of Commerce [Uni04] employ measures to protect privacy, however these rights are still emerging and in a state of flux.

Not everyone supports the notion of individual privacy protection. Privacy from a purely economic basis [Pos81] holds that personal information should be kept private only if the economic value to society of such information is decreased by it becoming public knowledge. Furthermore, the only personal value in concealing private information is in deceiving or manipulating others for personal gain, and therefore is not of economic use to society as a whole. This view proposes corporate privacy as having value, but asserts that personal privacy is not beneficial to a nation’s economy and so should not be protected in law. This view of privacy is not widely accepted; however, and many modern world societies have enacted laws that protect individual privacy to varying degrees.

2.1.2 Identity.

Many anonymity-related concepts obfuscate information relating to a user's (or agent's) identity. Identity takes several forms, but the archetypical example is the *name* [WrS05]. The name of an individual is intended to be a unique identifier within some group so that individual can be distinguished from others in that group. When discussing the anonymity properties of a user, the existence of a unique identity is implicit.

However, a distinction must be made between a user's representation in a system and their real identity. Multiple users may collaborate to form a single online identity or a single user may have multiple representations online. The full implications of this are not entirely understood, as the simplifying assumption that a single user is linked to a single representation is almost universally made in anonymity research [WrS05]. Although this seems logical, there are many other interpretations of what an identity or "name" is including an Internet Protocol (IP) address (either IPv4 or IPv6), Media Access Control (MAC) address, geographical location, or e-mail address.

2.1.3 Anonymity.

Anonymity is a fundamental identity hiding property and totally removes identifying information about the user. Even so, identifying information may be added into a data channel within an anonymous system as needed. As such, anonymity provides the choice to limit identity hiding as much or as little as desired by explicitly revealing identifying information as necessary [WrS05].

Total anonymity is the focal point for identity hiding research. Additionally, anonymous systems are typically based on a small number of approaches with Chaum's

mix [Cha81] being the most prevalent. Most active research topics on anonymity are variations of these basic ideas. Figure 1 shows the yearly anonymity publications in IEEE Xplore [IEE09] and the Freehaven bibliography [Fre09], an authoritative source of select anonymity publications from 1980 to the present.

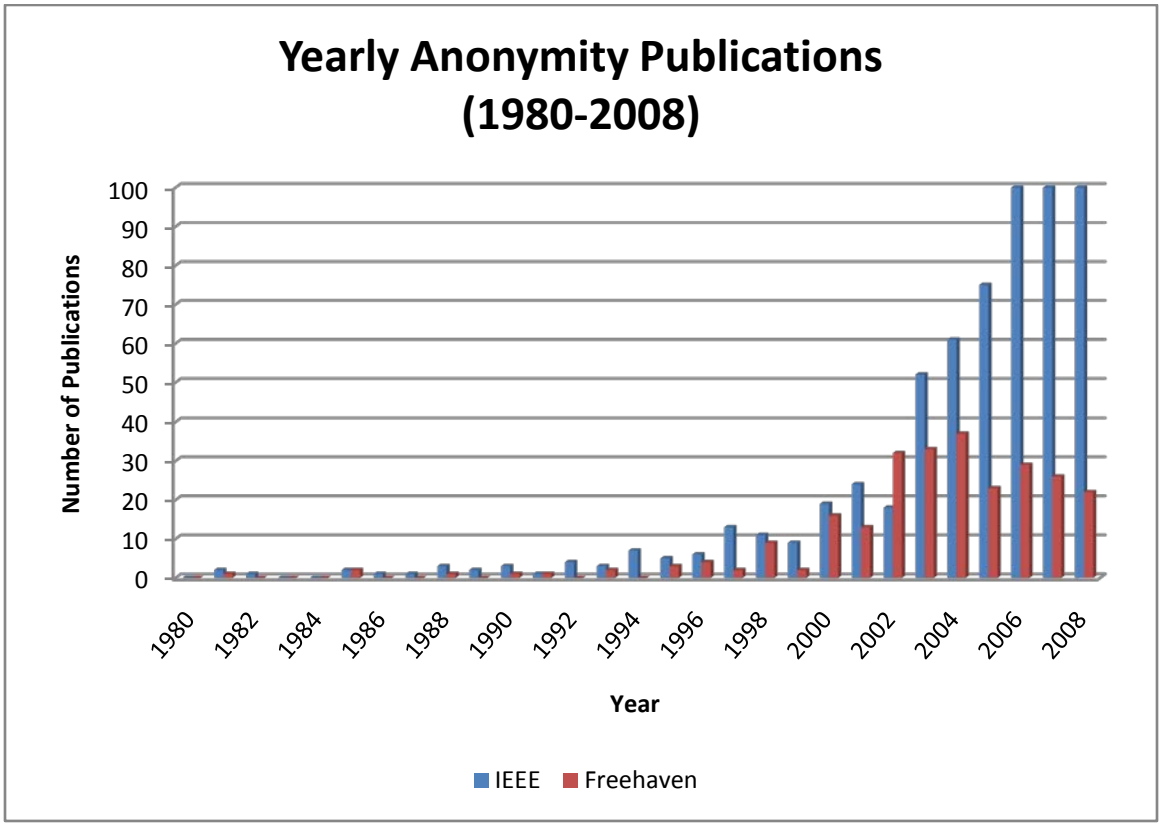


Figure 1: Yearly Anonymity Publications

Although not an exhaustive list, the trend is quite clear. The field of anonymous system technologies started receiving attention from the large research community around the year 2000 and interest in anonymous system research is growing.

Despite the focus on anonymous systems, total anonymity is a two-edged sword [WrS05]. For publishing, mailing lists, and web surfing applications, anonymity can be

highly desirable. However, for other systems, no possibility of tracking identities is detrimental [WrS05]. Sometimes identity needs to be tracked over the course of an extended transaction, but not between transactions. For this reason *pseudonymous communication*, which provides a certain amount of information associated with an identity, is required for a number of practical identity hiding systems [WrS05]. The advantages and disadvantages of anonymity in general are discussed next.

2.1.3.1 Advantages.

Any society has a natural inclination towards conservatism, including the global society of the Internet. So anonymity is often seen as a counter-balance to such conservatism. Anonymity inherently offers the advantages of promoting freedom of expression and protecting user privacy.

The Internet allows any user to instantly reach and possibly influence millions of others. In essence, Internet technology offers users a fast, inexpensive way to publish anything, anywhere, anytime. There are many long-standing precedents for anonymity in publishing. For example, the Founding Fathers of the United States anonymously advocated the adoption of the Constitution by publishing the *Federalist Papers* under the pseudonym Publius [Luc06]. Prior to the American Revolution, many resorted to secret publication to avert English prosecution [GoW98].

More recently, the United States Supreme Court favored protection for anonymous publication of political speech. As Justice Stevens wrote:

> "*Under our Constitution, anonymous pamphleteering is not a pernicious, fraudulent practice, but an honorable tradition of advocacy and of dissent. Anonymity is a shield from the tyranny of the majority* [GoW98]."

Most newspapers allow anonymously signed letters and credit articles to the "AP Newswire" [Rig95]. Additionally, in academic environments, anonymous peer reviews of proposals and articles are expected and common. Thus, anonymous publication is a time-honored tradition. This makes anonymous speech an integral part of free speech, and free speech an essential part of any healthy democratic society.

Anonymity is also important for protecting user privacy in sensitive online forums involving sexual abuse, sexual conduct, religious beliefs, cultural issues, racial issues, harassment, and whistle blowing [Rig95]. Anonymity gives users a non-attributable channel to vent their benign or divisive opinions without fear of eventual identification and retribution. Thus, anonymity circumvents the majority from controlling the actions of the minority. Some prefer to be anonymous to ensure their views are evaluated on merit, not authorship name or association. Without anonymity, user actions or opinions may result in censorship, physical injury, social inequity, financial loss or legal action. Protecting users from such risks means preserving their privacy and circumventing social inequities in the global Internet society. This is a justifiable cause for the introduction and preservation of anonymity on the Internet.

Given the historical precedents of anonymity and growing demand for anonymous technologies, anonymity on the Internet is here to stay. Anonymity offers the advantages of promoting freedom of speech and protecting user privacy on the global society of the Internet. Nevertheless, anonymity does have disadvantages.

2.1.3.2 Disdvantages.

Abuse and illegal activity are the most obvious drawbacks to anonymity. Governments, businesses and other organizations fear an inability to control abusive and illegal activity on the Internet. A libel suit was brought against online service Prodigy for anonymous postings. Although it ended with a temporary victory for Prodigy [Ano04], other site operators dread being held accountable for such nefarious activity and have developed a strong aversion to anonymity.

The concern about excessive abuse has merit. As mentioned in the previous section, the ability for any user to instantaneously publish printed information to millions of users around the world is a powerful one. People of all cultures, races and nations tend to more quickly and readily confer credence to the written word as opposed to the spoken word. As Walter Mossberg in the Wall Street Journal wrote, operating "… under the cloak of anonymity … makes it easier to spread wild conspiracy theories, smear people, conduct financial scams, or victimize others sexually" [Ano04]. Thus, online anonymity abuse can profoundly and adversely affect others. Fortunately, the majority of abuses can be attributed to new anonymous users and this type of abuse eventually diminishes [Rig95]. Even so, some abuse is instigated by disreputable individuals who are lured by the ability to effortlessly carry out certain actions with impunity. These actions include kidnapping, terrorism, harassment, threats, hate-speech, financial scams, and disclosure of trade secrets, personal information or intellectual property [Rig95]. For example, hiding behind anonymity to espouse nationally, ethnically, racially, or religiously hateful views is unacceptable and harmful to society. Some feel dealing directly with these societal issues is preferable to allowing concealment behind anonymous services. Yet for

centuries, societies have had similar issues. Offensive and inappropriate e-mails on the Internet may best be dealt with in the same manner as the real-world society – ignoring them [Rig95]. However, former U.S. President George Bush recently made posting annoying Web messages and sending anonymous e-mails a federal crime [Mcc06] based on existing telephone harassment law Title 47 [Uni05]. Illegal activity is not so simply dismissed.

Controlling illegal activity is virtually impossible on the Internet since anonymity ensures the identity of the perpetrator cannot be discovered or linked to specific actions. The topic of child pornography is often cited to vividly highlight the disadvantages of anonymous services. Two Texas men were indicted for using the online pseudonyms "Poo Bear" and "Wild One" to lure two young boys and commit sexual acts [Rig95]. The number of criminals using Internet anonymity services to participate in illegal activity is increasing and has motivated lawmakers to limit the use of anonymity. Recently lawmakers barred 29,000 known sex offenders from using MySpace to anonymously solicit minors [Lem07]. Hence, using anonymity services makes committing crimes such as this easier. On the other hand, law enforcement agencies encourage citizens to use anonymous e-mail to report crimes [Ale07, Ano07g, Jor07, Rob07]. Businesses that rely on trade secrets and/or intellectual property to maintain competitive advantage fear anonymity services will undermine existing laws to protect this information.

Given the disadvantages of excessive abuse and illegal activity, it is no wonder many organizations are dissuaded from fully embracing anonymity. They do not want to be

held responsible for acts of terrorism or kidnapping due to anonymous messages passing through their system. Hence, anonymity's disadvantages are not trivial.

2.1.4 Pseudonymity.

One simple form of hiding identity is to use a pseudonym. Pseudonymity stems from Greek (*pseudos*, false) and refers to the adoption of a false name. This is also commonly known as an allonym (*allos*, other), *nom de plume* (pen name) or *nom de guerre* (name of war), after the traditional pre-computer use of pseudonyms as a method by which authors could publish politically inconvenient material without the threat of retaliation [WrS05].

Pseudonymity, in terms of usable online systems, associates a user with at least one semi-persistent identifier. The normal purpose is to allow types of transactions, relying on user history and behavior that are not possible in a totally anonymous system. This is of particular use in systems that rely on networks of trust between users, and thus cannot rely on a one-time session identifier approach.

Pseudonymity can be achieved using an anonymous infrastructure with suitable user information and history stored with the explicitly transmitted data. If the communication infrastructure is inherently anonymous then pseudonymity is an easier proposition as data can be released as desired without fear of extra information leakage from the system. Care must be taken that the interaction between deliberately released data and other data within the system does not interact reveal more than is intended.

Pseudonymity may therefore be seen as a problem that exists at a 'higher' level than anonymity. An anonymous channel may have some form of persistent user identification

that is kept secret between the sender and receiver. Pseudonymity typically entails a combination of other security properties such as secrecy, anonymity and authentication.

2.1.5 Reputation.

Reputation and trust are closely linked properties, particularly within the context of anonymity systems [WrS05]. Reputation allows a user to make an informed decision about whether or not to trust another user. This is important in commercial systems where users are required to invest real economic interests in other users of a system. The potential risks of such a system are high, especially in cases where there are no legal restrictions on the parties involved in a transaction. In these cases, which are common on the Internet which allows commerce between countries with differing legal systems, reputation is critical to users and legitimate businesses alike. Anonymity systems rely on distributed networks of untrusted users. Reputation algorithms provide a degree of assurance that network users will behave as advertised. Similarly, for pseudonymous online systems, reputation enforces "good" behavior between users. As such, in many of the practical applications of anonymity and pseudonymity, reputation is the key to a usable system.

2.1.5.1 eBay.

The most well-known reputation-based system is the seller rating on eBay [WrS05]. eBay is a popular online auction site that manages the buying and selling of a large quantity of items all over the world. Ebay emulates a global auction where buyers bid against each over a fixed period of time. The item is sold to the highest bidder. When the transaction ends, both the buyer and the seller are encouraged to provide a positive,

neutral or negative rating and text-based feedback about the behavior of the other party in the transaction. When considering an item, potential buyers may examine the ratings of a seller and decide whether to trust the seller and make the purchase. The greater the number of positive feedback reports a seller indicates a higher level of trustworthiness.

A seller wants to protect their reputation to attract more business in the future. As such, the seller is unlikely to perform any action that could damage their reputation. This approach towards trust management in commerce systems has been the subject of some study [Del05, JuF05, JuF06, JuF07, Li06, LiX06, MiR06, YaI04, YaI05, ZaM99]. Even before the invention of eBay and similar systems, reputation as a method of enforcing positive behavior in markets had been well-known and received much attention.

2.2 Nomenclature

This section reviews terminology and concepts of anonymity systems. These include the anonymity properties, adversary, attacks and mix. The more abstract term “agent” is often used instead of the simpler term “user” throughout.

2.2.1 Fundamental Anonymity Properties.

The fundamental anonymity properties covered in the academic literature include sender, receiver, communications and location anonymity. For completeness, the unobservability property is also discussed.

Sender anonymity prevents a particular message from being linked to a particular sender identity. Figure 2 depicts *sender anonymity* in an anonymous system. A message Bob receives is not linkable to Alice or any other sender in the

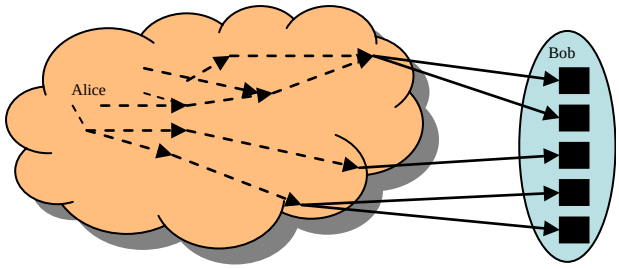


Figure 2: Sender Anonymity [Ser05]

anonymous cloud. Furthermore, no message to Bob or any other receiver is linkable to Alice. Thus, sender identity is hidden. The DC-net [Cha88] mechanism achieves *sender anonymity*.

Receiver anonymity prevents a particular message from being linked to a particular receiver identity. *Receiver anonymity* is shown in Figure 3. A message Alice sends is

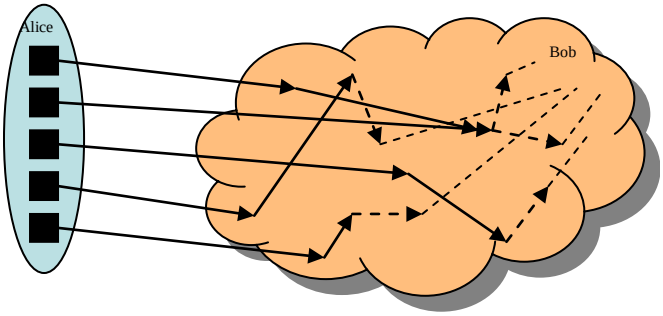


Figure 3: Receiver Anonymity [Ser05]

not linkable to Bob or any other receiver in the anonymous cloud. Furthermore, no message from Alice or any other sender is linkable to Bob. Thus, receiver identity is hidden. Broadcast [PaM86, Wai90] and private information retrieval [CoB95] are two mechanisms that achieve *receiver anonymity*.

Communication anonymity means a particular message cannot be linked to any sender-receiver pair and no message is linkable to a particular sender-receiver pair. Figure 4 shows *communication* (a.k.a. *relationship*) *anonymity* where a message is not

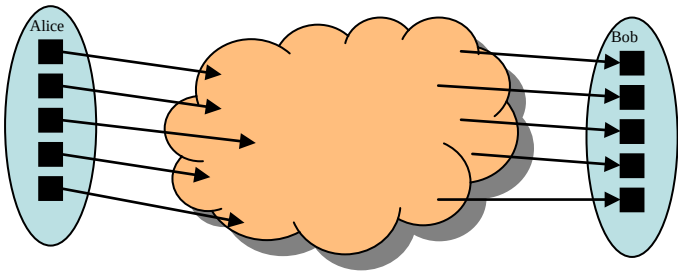


Figure 4: Communication Anonymity [Ser05]

linkable to the Alice-Bob pair or any other pair. Furthermore, no message from the Alice-Bob pair or any other sender-receiver pair is linkable for others. Thus, sender-receiver pair relationships are hidden. The MIX-net [Cha81] mechanism achieves *communication anonymity*. *Communication anonymity* is a weaker property than either of *sender* and *receiver* anonymity. This means although the sender and receiver cannot be linked, it may be clear the pair are participating in some communication [WaN07].

Location anonymity means a particular message is not linkable to any sender or receiver location, motion, route or topology information. An adversary has access to routing information on nodes or in packets but is unable to discover location, link information of a node, true routing path or tree information.

Unobservability means the adversary is unable to observe items of interest (IOI) as opposed to agent identities or relations. Unobservability can be achieved in one of two ways. First, if an adversary is unable to observe any message or IOI from any agent whether the IOI exists or not. Second, is if the anonymity of the other agent(s) related to an IOI is identical to other agent(s) related to that IOI. For instance, all agents simultaneously send the same size message across the network. The relationship of *unobservability* to anonymity is [PfK00]

$$\textit{Unobservability} \Rightarrow \text{Anonymity} \tag{1}$$

$$\text{Anonymity} + \text{Dummy Traffic} \Rightarrow \textit{Unobservability}. \tag{2}$$

Unobservability implies anonymity by keeping messages indistinguishable as well as identities anonymous as indicated by equation (1); however, anonymity does not imply *unobservability*. Looking at (2), anonymity plus dummy (indistinguishable) traffic implies *unobservability*.

Unobservability may be divided into *sender unobservability*, *receiver unobservability* and *communication unobservability* [PfK00]. *Sender unobservability* means it is undetectable whether any sender within the unobservability set sends. For example, in Figure 2 if Alice or any other sender transmits a message, the adversary is unable to either observe any or distinguish among the sender messages. Thus, sender messages are hidden. *Receiver unobservability* means it is undetectable whether any receiver within the unobservability set receives. For example, in Figure 3 if Bob or any other receiver gets a message, the adversary is unable to either observe any or distinguish among the receiver messages. Thus, receiver messages are hidden. *Communication unobservability* means it is not detectable whether anything is sent out of a set of could-be senders to a set of could-be receivers. For example, in Figure 4 any message sent by Alice or any other sender and received by Bob or any other receiver is undetectable. Thus, sender-receiver pair messages are hidden. It is not detectable whether within the communication unobservability set of all possible sender-recipient(s)-pairs a message is exchanged in any relationship. The larger the unobservability set, the stronger the unobservability.

2.2.2 The Adversary.

An adversary is an agent whose aim is to degrade or eliminate anonymity. The objective of an adversary is to link sender and receiver, identify the sender or receiver for a particular message, or trace a sender forward/receiver back to messages or disrupt the system.

A *global* adversary is omnipresent and has full access to the entire network of nodes and links. A *local* adversary has limited omnipresence and has full access to only a portion of the network nodes and links. This corresponds to the adversary possessing complete or restricted information or knowledge about the system. It may also refer to the veracity of this information. The adversary may either know things to be true or only believe things to be true.

A *passive/external* adversary is an outsider that can only observe messages traversing the network and is typically invisible. This adversary can only compromise communication channels between nodes. In other words, it is a non-empty set of agents, part of the surrounding of the anonymous system and capable of compromising links. An *active/internal* adversary is a visible insider and may alter messages traversing the network. This adversary controls nodes in the network. In other words, this describes a non-empty set of agents which are part of the anonymous system and capable of participating in normal communications and controlling at least some nodes.

Typically, the adversary is *dynamic* and collects information about the path selection algorithm, its parameters and as much information as possible about network activities from compromised nodes and links. The adversary uses all available facts to infer who sent or received which messages in a computationally bounded or even unbounded

manner. The adversary may behave deterministically with a scheduled plan of attack, probabilistically depending on the relative frequency of sequences of observed actions or events, or non-deterministically (unpredictably).

A combination of adversarial types constitutes the threat model. A strong threat model is a well-funded adversary who may compromise both nodes (*internal*) and links (*external*), observe all network traffic (*passive, global*), alter traffic (*active*) and operates mixes (*dynamic*) [DaR03]. Although this may appear to be a rather excessive assumption, any anonymous system that withstands strong adversarial attacks provides very strong security. However, in practice such threat models may lead to unrealistic designs. Therefore, available adversarial resources are considered carefully and countermeasures tailored accordingly to the anticipated threat level. In brief, anonymous communications systems are designed with an assumed adversary threat model in mind.

2.2.3 The Attacks.

Whatever the threat model, the goal of an attack is to link sender and receiver, identify the sender or receiver for a particular message, or trace a sender forward/receiver back to messages. The attacks and defenses for a passive and active adversary are provided in Table 1.

The goal of passive traffic analysis attack is to observe all traffic. A defense is to obscure traffic patterns by adding noise traffic, obfuscating timing or having same size messages. The purpose of a timing attack is to link incoming and outgoing message based on route time traversals. Synchronizing batching increases the anonymity set and is a good defense; however, it results in greater network load and less operator flexibility.

Table 1: Attacks and Defenses (Passive Adversary)

Attack	Goal	Defense
Traffic Analysis	Observe traffic	Obscure traffic patterns
Timing	Examine route time traversals	Synchronous batching
Content	Extract identifying information	Encryption
Counting	Long or short term communication	Obscure traffic patterns
Intersection	Correlate active times	Spread message out over time

based on route time traversals. Synchronizing batching increases the anonymity set and is a good defense; however, it results in greater network load and less operator flexibility. Extracting data or location identifying information is the aim of the content attack. Employing encryption to not reveal identifying information is a common defense. The counting attack scheme counts long or short term communications to reveal identifying information. Similar to traffic analysis defense, obscuring traffic patterns can thwart this attack. Lastly, the intersection attack targets networks without dummy messages to produce constant message stream and correlate the times sender and receiver are active. A defense spreads messages out over time to increase the set of possible senders. Active adversary attacks and subsequent defenses are shown in Table 2.

Table 2: Attacks and Defenses (Active Adversary)

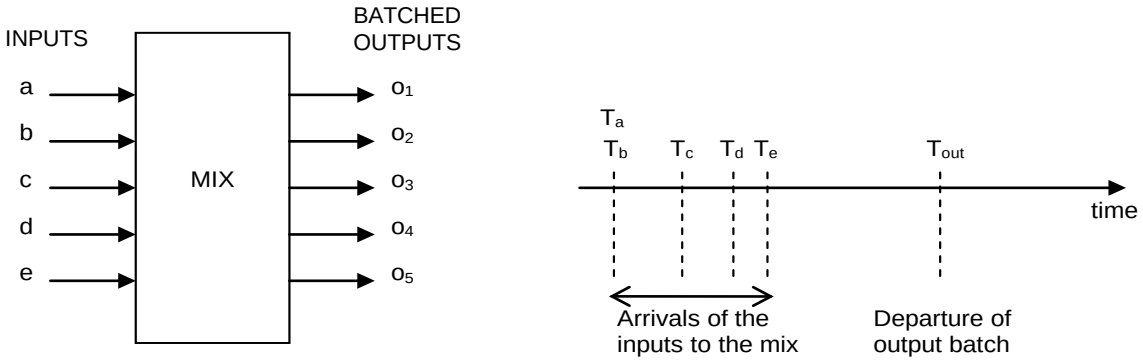
Attack	Goal	Defense
Traffic Analysis	Corrupt/delay traffic Partition traffic	Impose transmission deadline Little defense
Blending/n-1	Isolate target message	"heartbeat" messages
Denial of Service (DoS)	Deny use Degrade performance/anonymity	Digital currency (puzzles)
Tagging	Modify messages	Integrity checks
Colluding	Multiple-mix compromise	Drop messages
Sybil	Add mixes to control paths	None
Compulsion	Force mix to reveal decrypt keys	Forward secure
Reputation	Deny access, Cease existence	Digital currency (puzzles)
Replay	Re-use valid messages	Use nonces or timestamps

The goal of active traffic analysis attack is to corrupt or delay traffic and establish many attacker controlled routers. There are few effective defenses as these attacks are difficult to accomplish. Imposing transmission deadlines at each hop may partly mitigate the delay traffic. Isolating the target message is the reason behind a blending attack. Not relying on batches and sending "heartbeat" messages instead is a defense. Heartbeat messages are sent through the network back to the originating sender. If all heartbeat messages are not received, an *n-1* attack is occurring and the sender may either cease operations or inject dummy traffic to improve the anonymity of valid messages. The Denial of Service (DoS) attack objective is to force a large number of cryptographic operations or deplete bandwidth to deny use or degrade performance/anonymity. A defense using digital currency to make clients pay for router services can be effective. A hard to perform but easy to verify client puzzle, such as use of a client puzzle in Tor, demonstrate its effectiveness [Fra06]. A tagging attack modifies messages. Performing integrity checks on messages counters this attack. The target of a colluding attack is to get multiple mixes to work together to compromise mixes. Dropping messages if an unplanned path is taken ensures agents cannot traverse adversary-controlled paths. The Sybil attack adds mixes and controls message paths. It is believed that no defense exists for this type of attack. The compulsion attack forces a mix to provide decryption keys. Ensuring forwarding nodes are anonymous also or forward secure is a good defense. Denying access to the network or making an anonymous service unpopular are two goals of a reputation attack. Defending against this is similar to DoS attacks: use digital currency to deny or slow access. The replay attack goal is to reuse or alter prior authentic messages later to masquerade as a valid user. A simple way to thwart a replay attack is

using one-time-only nonces in messages so subsequent similar messages are ignored. Another is embedding time-stamps in messages for synchronized systems. Unfortunately, one good defense, injecting unique sender and receiver identities into messages, runs counter to the purpose of providing anonymity. The mix technology is described next.

2.2.4 The Mix.

A mix is the most extensively researched and implemented anonymous technology. The original mix was designed to make e-mails untraceable [Cha81]. Other applications of a mix include secure electronic voting, anonymous telecommunications, and anonymous Internet communications. Subsequent mix variations protect against or avoid specific attacks and/or seek to boost performance in specific application domains. A representative mix is shown in Figure 5.



(a) Changes appearance and flow of inputs and batches outputs.

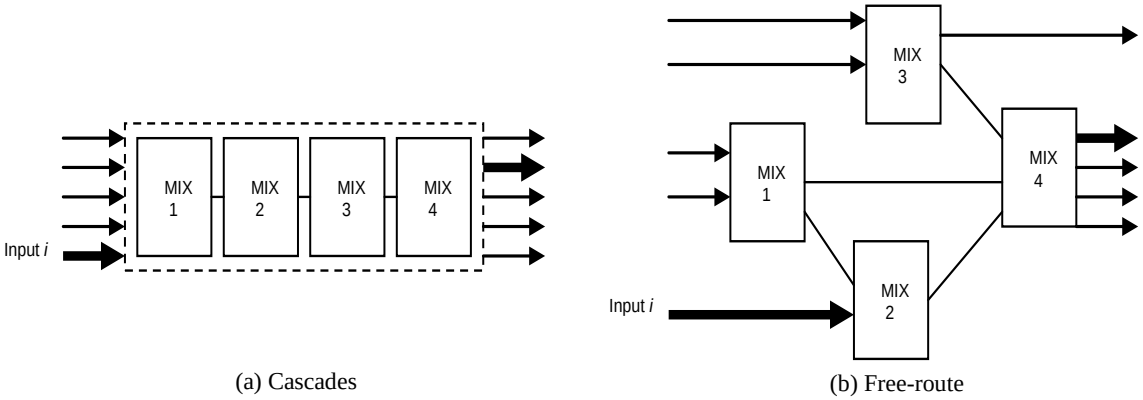
(b) Different input arrival times at mix, one output departure time.

Figure 5: A Mix [SaP06]

Figure 5(a) shows the major mix component. A mix accepts input messages on links *a*, *b*, *c*, *d*, and *e* and generates uncorrelated, batched output messages to links o_1, o_2, o_3, o_4,

and o_5 by altering the flow and appearance of each message. For alter flow, the message is delayed and/or reordered. For appearance, the message is re-encrypted and/or padded. The mix decrypts the encrypted input messages and removes all sender information such as timing information from the headers. For instance in Figure 5(b), different input arrival times T_a=T_b, T_c, T_d, and T_e are simultaneously output at time T_{out}. This provides unlinkability and defends against traffic analysis attacks. Once a specific condition is achieved, the mix forwards a mixed batch of output messages to receivers or another mix.

Multiple mixes are connected together to form a mix topology and are called mixnets. The two main topologies are illustrated in Figure 6.



(a) Cascades (b) Free-route

Figure 6: Mix Topologies [SaP06]

Cascades consist of a fixed number of sequential mixes a message traverses in the anonymous network. In Figure 6(a), mix one transforms the inputs and concurrently transfers outputs to mix two. Mix two repeats the transformation and forwards to mix three. This continues until mix four outputs the untraceable inputs. All inputs traverse a single path. Alternatively, free-route networks consist of a variable number of mixes a message traverses in the anonymous network. In Figure 6(b), mix two accepts an input and forwards it to mix four; however, not all inputs follow the same path. While the

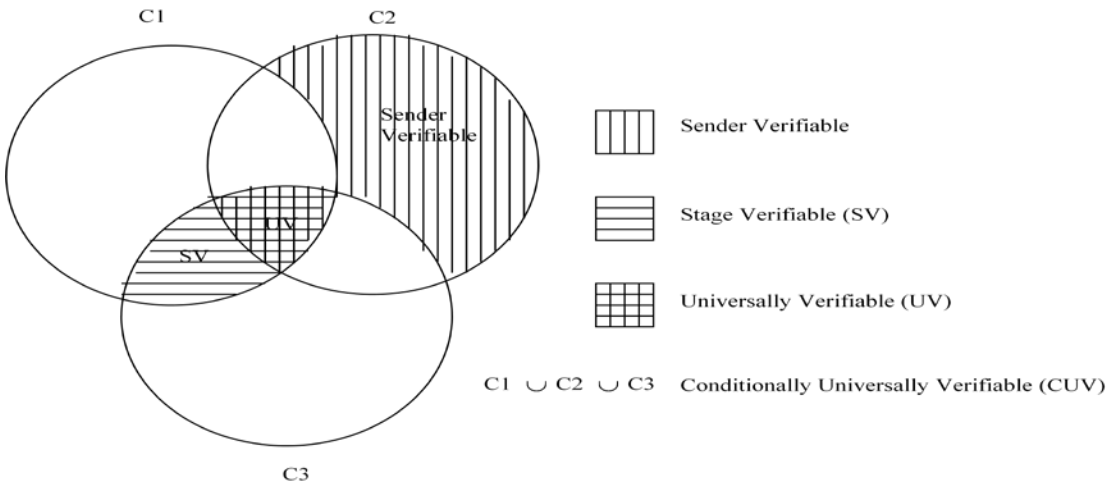
cascade topology provides overall better security properties compared to the free-routing topology in mixnets, under certain conditions, the free-routing topology provides more robust anonymity [BeP01].

Verifiability, a common robustness technique in cascade topologies to protect against integrity attacks, checks the correctness of each mixnet output. The following correctness criteria [SaP06] determines whether input messages are

*C*1) Transformed as expected.
*C*2) Uncorrupted.
*C*3) Equal in number (no added/deleted messages).

The verifiability mechanism must satisfy all three criteria. This region is indicated by $C1 \cap C2 \cap C3$ and the classification of cascade mixnets is shown in Figure 7. Sender verifiable (SV), mix verifiable (MV), universally verifiable (UV), and conditionally universally verifiable (CUV) are the classifications.

The sender verifiable (SV) mechanism detects corrupt output messages and the mixnet only satisfies the horizontally hashed C2 area as shown in Figure 7. The mix



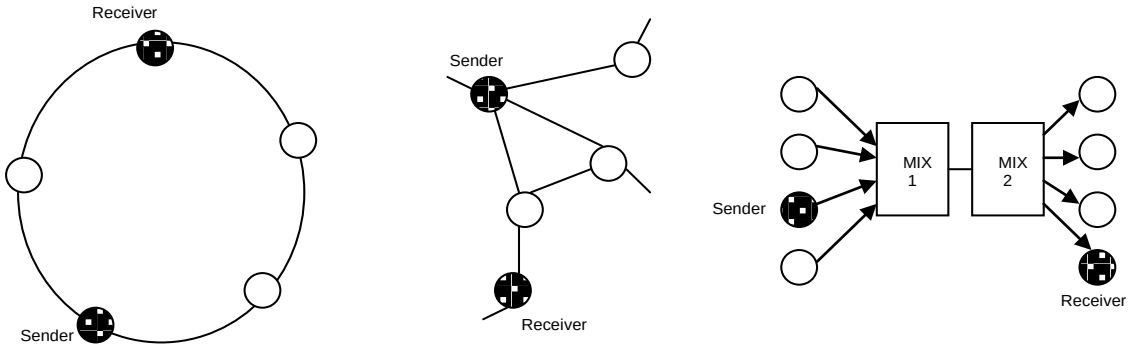
C1 = Transformed as expected.
C2 = Uncorrupted.
C3 = No added/deleted input messages.

Figure 7: Verifiable Cascade Mixnets Classification based on Satisfied Correctness Criteria [SaP06]

verifiable (MV) mechanism has each mix verify its own batch output but does not require SV.

Together the mixes execute supplementary subprotocols to ensure output batch correctness. The mixnet satisfies $C1 \cap C3$ but not necessarily C2. In a universally verifiable (UV) mixnet, even if all mixes are corrupted, an incorrect output batch is not possible and satisfies all three criteria or the $C1 \cap C2 \cap C3$ region. Each mix must prove an output uniquely corresponds to an input without revealing such a relationship. Conditionally universally verifiable (CUV) provide probabilistic guarantees on output batches but not necessarily on all batch outputs. Hence, a CUV mixnet satisfies one, two, or three criteria or the $C1 \cup C2 \cup C3$ region.

Several variations on mix methods [Cha88, ChK03, DiM04, Jon04, LeS02, ReR98, ShL00] and other peer-to-peer approaches [BoW05, ChW06, GoR03, HaL05, Kon05, LiX06, LuF04, ReP02, RsZ04, XiX03, XiX03a, ZhH04] have been proposed as solutions to provide anonymity in communication networks. In Figure 8, the three main anonymity solutions are shown.



(a) Peer-to-Peer with Sender and Receiver Anonymity

(b) Peer-to-Peer with only Sender Anonymity

(c) Mixnet with Unlinkability

Figure 8: Anonymity Solutions [SaP06]

In Figure 8(a) and Figure 8(b), the sender has two or more connected peers. If the adversary is unable to eavesdrop on all of its connections and the peers are not compromised, the sender's communications are untraceable [PaM86]. Hence, sender and communication anonymity can be achieved. However, the sender may be identifiable and traceable to the mix input in Figure 8(c). Hence, only communication anonymity may be achieved. The Figure 8(a) solution is effective for broadcast communications and providing sender and receiver anonymity [Cha88, PaM86, Wai90]. Figure 8(b) solution is effective for low latency communications [ReP02, ReR98]. However, both peer-to-peer solutions are susceptible to single node disruptions and a powerful adversary may degrade or eliminate anonymity. Also, peer-to-peer solutions are not necessarily robust, efficient, or scalable for secure applications. The mixnet solution provides better anonymity and is more robust, efficient, and scalable for secure applications [ReP03].

The different approaches to anonymity and classification of mixnets based on verifiability are shown in Figure 9. The root of the tree anonymity is broken out as peer-to-peer or a mixnet. The peer-to-peer subtree was already discussed using Figure 8. The mixnet topology expands to cascade and free-routes as covered above using Figure 6 and Figure 7. The free-route is either synchronous or asynchronous. The asynchronous subtree branches to remailers and low latency onion routing. Both are reviewed in more detail in the next section. The cascade subtree subdivides mixes by cryptographic function of decryption, hybrids, and reencryption. Decryption mixnets [Cha81] require the sender to encrypt the message with the keys of each intermediate mix, called a onion, and may use the RSA [RiS78] or

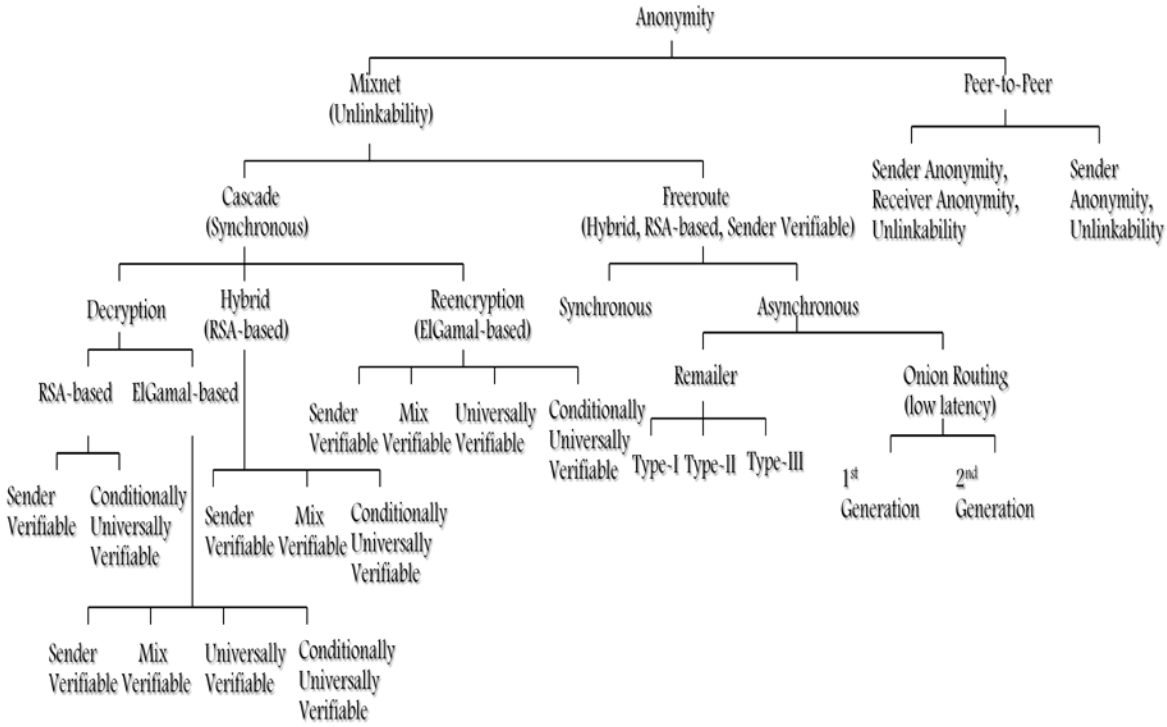


Figure 9: Overall Classification on Anonymity and Mixnets [SaP06]

ElGamal public key cryptosystem. As the number of intermediate mixes or onion size increases, public key operations become expensive.

A more efficient variant of the decryption mixnet is the hybrid mixnet [GoR96, JaJ01, Mol03]. It uses symmetric as well as public key operations to achieve efficiency and is RSA-based. However, RSA-based decrypt and hybrid mixnets have weaknesses: a sender traceable onion, a sender must encrypt for each intermediate mix, the sender onion size decreases as it traverses the network, and a fixed decryption sequence. The ElGamal-based reencryption mixnet overcomes these weaknesses. The leaves of the cascade subtree are identified with appropriate classifications as explained above using Figure 7. The anonymous communication networks are reviewed next.

2.3 Anonymous Networks

Anonymous networks may be divided into wired and wireless protocols. They typically vary in routing scheme, transmission medium, topology, and protocol implementation which affect the adversarial threat. Hence, providing anonymity in each network requires a different approach particularly when mobility is involved.

Wired or fixed anonymous networks have been thoroughly studied [Cha81, Cha88, PaM86, PfP91, RaS93, ReS98, RmR98]. These networks consist of a set of uncompromised nodes with distinctive identities called the anonymity set. The items of interest are predominantly network transmissions. Many anonymous schemes assume the network topology is fixed, while others assume the entire topology is known *a priori* [KoH05]. These assumptions do not hold in mobile wireless networks.

Wireless anonymous mobile networks research [BeS03, DeH04, GrG03] examines protecting privacy or location information in stationary sensor networks but does not consider mobility's impact on anonymity. Other research [AtH99, HqW04, SaM95] focuses only on protecting anonymity for mobile users in last-hop wireless networks which degenerates to analyzing wired network anonymity.

In both *Wired* and *Wireless* networks, the network routing scheme is a major factor affecting anonymity [LhM04]. Four generic network routing schemes are shown in Figure 10. There is a single sender node on the far left. The nodes incident or near the lines on the right are the receiver node(s).

(a) Unicast (b) Multicast (c) Broadcast (d) Anycast

Figure 10: Network Routing Schemes [Wik07]

In Figure 10(a) unicast, a one-to-one relationship exists between sender and receiver. A single receiver is uniquely identified. In Figure 10(b) multicast and Figure 10(c) broadcast, a one-to-many relationship exits exists between sender and receivers. Each uniquely identified receiver gets all information from the sender. In Figure 10(d) anycast, a one-to-many relationship also exists between sender and receivers. However, only one uniquely identified receiver gets the information from any given sender at any given time. Anycast is used for connectionless or User Datagram Protocol (UDP) based protocols.

For *Wired* networks, practically all in-depth research on anonymity assumes a unicast routing strategy. Exceptions include the Dining Cryptographers Network (DC-Net) [Cha88], P^5 [ShB02], Hordes [LeS02], MAM [XiL06], and Cashmere [ZhZ05]. For *Wireless* networks, a mobile wireless node typically broadcasts to neighboring nodes.

2.3.1 Wired Networks.

This section introduces the myriad of implemented or proposed wired anonymous networks. Each protocol is summarized and major advantages and/or disadvantages highlighted.

2.3.1.1 Anonymizer.

Anonymizer [Boy97] is a Hyper-Text Transport Protocol (HTTP) proxy that filters

out identifying headers and sender addresses from the Web browser [GuF04]. This is a fast way for users to surf anonymously without revealing their identity to Web servers and provides sender anonymity. The mix topology and path consist of a single node, the Anonymizer-Server. The strengths are low-latency, easy implementation, and increase of anonymity set compared to non-anonymous systems. However, security is weak since no chaining, encryption, log safeguarding, or forward secrecy is offered. Furthermore, with only one node, a DoS attack is easy and an adversary monitoring requests can quickly link sender and receiver.

2.3.1.2 Java Anon Proxy.

Java Anon Proxy (JAP) or WebMIX [Egg05] is a working anonymous web surfing network over the Internet. A single address is shared by many users so sender and communication anonymity are protected from both the adversary and receiver (website). The client interacts with cascade mixes and uses a predetermined sequence of mixes (i.e., a fixed path). Users connect with encryption through intermediary mixes to the web server. Its strength is users may choose between different mix cascades and multiple users traversing the same mix increases the anonymity set and mix dummy traffic inhibits traffic analysis.

2.3.1.3 PipeNet.

PipeNet [Dai98] is a simple theoretical model for web surfing over the Internet. It is a low-latency Internet Protocol (IP)-level cousin of a Type II remailer network such as Mixmaster, with extra dummy traffic to defend against timing attacks. All users send a legitimate or dummy message each time unit to the identical cascade mix using virtual

link encryption. The cascade consists of a sequence of pre-established (fixed) 3 to 4 node path. The strengths of strong anonymity and traffic analysis protection are offset by the weaknesses of impracticality, DoS vulnerability, and inefficiency. The model is idealistic not practical. Although a very influential early anonymous communication network proposal, PipeNet has not been designed much less implemented and is not a serious candidate for practical development. The DoS vulnerability stems from a malicious user's ability to not send a message thereby bringing the entire system down. The efficiency problem is due to the constant-bandwidth long-lived encrypted links incurring serious performance costs to provide security against a strong adversarial model of pervasive eavesdropping on the network.

2.3.1.4 Onion Routing (Tor).

Onion Routing [DiM04, ReS98] is a mature research anonymous communications network for interactive anonymous Internet traffic such as the Web, Internet Relay Chat (IRC) and Secure Shell (SSH). Onion Routing establishes circuits with layered asymmetric keys (hence, the onion nomenclature) and hides the sender and receiver address. It is implemented at the application or Transmission Control Protocol (TCP) layer and offers sender, receiver and communication anonymity. Onion Routing relies on Transport Layer Security (TLS) to provide forward secrecy and dummy messages. The first generation (type I) mix topology is cascade mixes called Onion Routers with a fixed five (5) onion router path selection strategy. The second generation (type II) mix topology is free-route with variable, random hop and cyclic path selection of up to 50 onion routers [GuF04]. Each mix station is independent and randomly chooses the next

mix in the path. The strengths are application independent connections and wrapped encryption to established circuits which is an excellent deterrent against traffic analysis attacks. The main weakness is no attempt is made to protect against a global, active adversary. Hence it is vulnerable to attackers who can control (or monitor) many diverse portions of the network simultaneously.

2.3.1.5 Freedom Network.

Freedom Network [GoS99] provides an anonymous Internet connection that is similar to Onion Routing; however, it is implemented at the IP layer rather than the application level. It provides sender anonymity for Web browsing but may also be used for IRC, SSH, Telnet and E-mail. The topology is a cascade mix, random path length and acyclic. The sender may randomly choose the no cycle path, but the path length is fixed at three intermediate nodes [GuF04]. The strengths are efficiency and reasonably secure against DoS attacks. The weaknesses are application-dependence and vulnerability to generic traffic attacks.

2.3.1.6 Cyberpunk (Type I remailer).

Cypherpunk [Pas00] is a type I remailer using layered asymmetric encryption for messages with a proper Pretty Good Privacy (PGP) key but does not mix messages. It provides communication anonymity only. The path is a sequence of remailers. The strength is strong anonymity. First, no pseudonyms are supported; no secret identity table is maintained, and no mail logs are kept to identify users. This diminishes the risk of "after-the-fact" tracing. Second, remailers accept encrypted e-mail, decrypt it, and remail the resulting message. This prevents an eavesdropping adversary from linking

incoming and outgoing messages. Third, remailers use chaining to achieve more robust security. Chaining sends a message through several anonymous remailers. The weakness is messages are not mixed and when message size gets smaller a link between sender and receiver is possible if the adversary monitors requests.

2.3.1.7 Mixmaster (Type II remailer).

Mixmaster [Cot01] is a type II remailer enhances protection against eavesdropping attacks and uses Simple Mail Transfer Protocol (SMTP) by adding sender anonymity. The path is still a fixed sequence of remailers. Strengths are the use of message padding and mixing to reduce the vulnerability to content or timing attacks. Another is the use of unique identifier and timestamps to mitigate replay attacks. Weaknesses are messages are unicast only and no reply message capability exists.

2.3.1.8 Mixminion (Type III remailer).

Mixminion [DaR03] is a type III remailer that improves upon Mixmaster. The added improvements include replies, integrated directory servers, dummy traffic, forward anonymity, replay prevention using key rotation, and exit policies. For instance, Mixminion batches message-based free-route mixes with secure single-use reply blocks. It also uses Transport Layer Security (TLS) over TCP and adds receiver anonymity. The path is a free-route mix. A strength is replies are allowed. Another is mix nodes cannot distinguish forward messages from reply messages, so forward and reply messages share the same anonymity set which provides forward anonymity. Other strengths are it runs in a real-world Internet environment, requires minimal node synchronization, and defends against known anonymity-breaking attacks such as replay attacks.

2.3.1.9 DC-Net.

The Dining Cryptographer Network (DC-Net [Cha88]) is the 1st P2P approach to achieve perfect sender and receiver anonymity and allows a single sender to broadcast a message to multiple receivers [Cha88, Wai90]. It is the only known non-rerouting protocol. A strength is perfect sender anonymity. The receiver gets the message under certain circumstances (odd parity) that prevents anyone but the sender from knowing who sent the message. The strength over rerouting protocols is lower overhead due to shorter delays and no re-routing traffic [GuF04]. A weakness is due to the broadcast medium, only a single sender may send a message. Another weakness is sharing secret coin flips with other parties requires significant coordination and synchronization between nodes which is inefficient on larger scales. In fact, it requires $O(n^3)$ protocol messages per anonymous message in a network of n agents [WaN07]. This makes DC-Net impractical and un-scalable.

2.3.1.10 Herbivore.

Herbivore [GoR03] is used for anonymous Web surfing and other Internet applications. It addresses the practical issues DC-Net does not like who sends when and the joining and leaving of a network by dividing the communication of the shared secret into three steps [Jon04]. It uses a star-topology instead of broadcasting to reduce the communication requirements to preserve anonymity. The strengths are a more efficient and scalable design. A weakness is network nodes may crash and depart the network at any time and degrade anonymity to a small degree.

2.3.1.11 Crowds.

Crowds [ReR98] is for anonymous Web surfing and extends the Anonymizer protocol. A sequence of mixes (jondos) with random hop selection per hop with cycles replaces the single node point of failure. This achieves sender anonymity. As long as the sender does not reveal identifying information in the request [Jon04], communication anonymity is also achieved. The strengths are users blend into a crowd and the unicast probabilistic routing. However, since the last jondo contacts the end server directly [Jon04], no receiver anonymity is achieved. This is a weakness.

2.3.1.12 Hordes.

Hordes [LeS02] improves Crowds. Jondos are User Datagram Protocol (UDP) proxies instead of HTTP proxies. Also, a multicast instead of reverse path return is used, thus sender anonymity is achieved. As with Crowds, if the sender does not reveal identifying information to the receiver, communication anonymity is achieved. However, receiver anonymity is not achieved as the last jondo still contacts the receiver directly. The multicast return and UDP proxies achieve the strength of low-latency. Similar to Crowds, Hordes allows cycles on the forwarding path.

2.3.1.13 P^5.

P^5 [ShB02] is for anonymous Internet applications. Users are placed into anonymity groups and messages are broadcast in a hierarchical tree structure. Using broadcast ensures receiver anonymity. To achieve sender and communication anonymity, nodes send uniformly distributed constant noise [Jon04] to ensure the impossibility of distinguishing between noise and real traffic. This makes for an efficient and scalable

system. A weakness is P^5 requires the most bits to send one anonymous bit compared to the other protocols. However, P^5 message dropping algorithm mitigates this somewhat [Jon04] by allowing bandwidth or processing constrained nodes to drop packets in a uniform or non-uniform manner as necessary; thereby, reducing the number of bits traversing the network.

2.3.1.14 Tarzan.

Tarzan [FrM02] is a peer-to-peer anonymous IP network overlay that uses layered encryption and multi-hop routing. The sender pre-selects the relay node path, creates static tunnels through these nodes, and generates dummy traffic to provide anonymity. It achieves sender, receiver and communication anonymity for Web surfing and has the strength of using less processor intensive symmetric keys. A tunnel failure incurs both significant computation overhead and delay [ZhZ05].

2.3.1.15 WonGoo.

WonGoo [LuF04] is based on mix and Crowds and is a scalable P2P system for low-latency anonymous communication resistant to both eavesdropping and traffic analysis. Layered encryption and random forwarding result in strong anonymity and high efficiency. A detailed comparison of WonGoo, Crowds and mix in [LuF05] shows its efficiency and anonymity.

2.3.1.16 Cashmere.

Cashmere [ZhZ05] is a resilient anonymous layer built on a structured P2P overlay. Instead of relaying traffic through fragile single-node Chaum-mixes to achieve

anonymity, Cashmere relays traffic through more robust relay groups of mix nodes thereby lowering the chance of a path failure and increasing the success of end-to-end message delivery. When an agent of the relay group receives a message, it anycasts the message to the next relay group as well as broadcasts the decrypted contents to all relay group agents. Cashmere provides sender and communication anonymity and can be extended to provide receiver anonymity. However, issues of key management and key revocation still must be resolved.

2.3.1.17 MAM.

MAM [XiL06] is a self-organizing and distributed mutual anonymous multicast and unicast protocol for applications such as video conferencing, distance learning and software updates. It is designed for high mutual anonymity degree, efficient message delivery, distributed and dynamic behavior and self-optimization [XiL06]. Two challenges are managing group agent memberships and group keys. MAM works best with smaller networks as the protocol is sub-optimal if the vast majority or all agents in the network want to hide their identity.

2.3.2 Wireless Networks.

The dynamic topology of wireless networks due to mobility, routes failures, and nodes entering/leaving makes proactively maintaining topology knowledge very costly and divulges private node knowledge to adversaries. The wireless IEEE 802.11 standard specifies particular topologies supporting transparent to allow node mobility to higher protocol layers [IEE99]. These topologies include Basic Service Set (BSS) networks, Extended Service Set (ESS) networks and Independent Basic Service Set (IBSS)

networks. Figure 11 illustrates the two basic networks.

A BSS network has mobile nodes within the same area which communicate via a single access point. Each mobile node transmits all frames to the access point, who forwards them within the same area or over the backbone distribution system. An ESS comprises one or more BSS networks where each access point acts as an Ethernet bridge and communicates over the distribution system. These topologies can achieve the same anonymity as *Wired* anonymous networks.

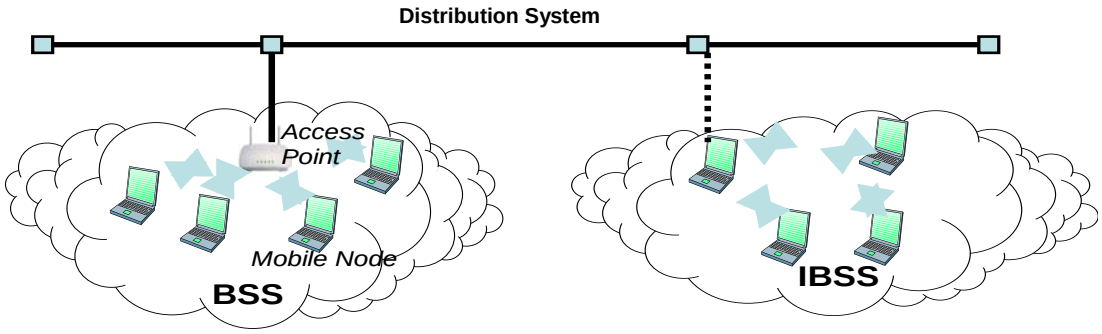


Figure 11: BSS and IBSS Networks

In contrast, IBSS or ad-hoc network nodes within the same area communicate directly with each other. The dotted line indicates one or more nodes might still have access to the backbone distribution system. This requires a different approach to achieve anonymity. Ad hoc networks self-organize, deploy quickly and lack infrastructure. Nodes may be highly mobile or stationary and have a wide range of capabilities [KoV98]. A few researchers have offered anonymous solutions for Mobile IPv6 [HaJ01, HqW04] and personal areas networks (PANs) [HqW04, Sch02]. Numerous protocols address the routing problem this poses. Each protocol is summarized and major advantages and/or disadvantages highlighted.

2.3.2.1 SDAR.

SDAR [BoE04] is a non-source-based routing, proactive neighbor detect, Mix-net onion, and path hijacking resistant protocol for MANETs deployed in hostile environments. Sender nodes initiate path establishment by broadcasting a path discovery message with specific trust requirements to neighboring nodes to ensure only trustworthy nodes construct routing paths to preserve node anonymity. It uses a public key cryptography trapdoor. However, it has a trapdoor, scalability and security issue [SoK05]. The long private decryption key results in very high computational complexity when the number of route request (RREQ) packets gets large for forwarding nodes. The long private key results in high computational complexity when forwarding nodes create encrypted signature routing messages during path discovery. Finally, part of the routing message may be deleted and modified by a forwarding node or adversary.

2.3.2.2 AnonDSR.

AnonDSR [SoK05] is a purely on-demand, MIX-net onion, no neighbor exposure, and crypto-protected receiver protocol [KoH07] for MANETs. It is composed of the security parameter route establishment, anonymous source-receiver route discovery, and anonymous cryptographic onion data transfer protocols. In route establishment, an adversary performing an active modification or reply attack or executing the passive eavesdropping attack cannot succeed. In route discovery, an adversary cannot modify the public key, trapdoor or onion and a replay attack is detectable. In data transfer, the onion protects all data communications. As path length increase, AnonDSR scales better than SDAR especially for anonymous route establishment.

2.3.2.3 MASK.

MASK [ZhL06] is a proactive neighbor detect, virtual circuit data delivery, no neighbor exposure, and broken destination (receiver) anonymity protocol [KoH07]. It is capable of MAC-layer and network-layer communications and offers sender, receiver, location and communication anonymity under a passive adversary model for large-scale theater-wide communications (multiple MANETs) or small-scale tactical communications in Urban Terrain Military Operations. It establishes source-destination virtual circuits and uses dynamic pseudonyms for path presentations [ZaW05]. It is resistant to message coding, flow recognition, replay and timing attacks, and offers high routing efficiency compared to classical AODV [PeB03]. Unlike ANODR, MASK is not sensitive to node mobility and allows anonymous MAC-layer communications. Two weaknesses are the final destination is contained within every RREQ message plaintext thereby breaking destination anonymity and reliance on a tight synchronization of neighbor keys and pseudonyms [SeP06].

2.3.2.4 ARM.

ARM [SeP06] is an anonymous on-demand routing protocol for MANETs that is secure against two assumed adversaries: cooperating nodes inside the network and an external, global, passive adversary that monitors all network traffic. It offers sender, receiver, and communication anonymity in both static and dynamic networks. It assumes every node has a permanent identity known by other nodes, source and destination share a secret key and pseudonym, every node establishes a broadcast key with its 1-hop neighborhood, and symmetric wireless links. Both random padding and time-to-live

values are applied to RREQ and RREP messages. The main advantages are higher efficiency than ASR, ANODR and SDAR, improved receiver anonymity over SDAR and MASK, and preserved communication anonymity against a powerful adversary unlike ANDOR, ASR, SDAR and MASK [SeP06].

2.3.2.5 ODAR.

ODAR [SyC06] uses Bloom filters for storage-, processing- and communication-efficiency, is based on asymmetric cryptosystems, and provides sender, receiver, communications and location anonymity. A Bloom filter is a space-efficient probabilistic bit-vector data structure for storing the elements of a set, and testing whether or not any given element is a member [Blo70]. A key management mechanism for distributing keys during source route construction provides strong end-to-end communication anonymity.

2.3.2.6 AMUR.

Anonymous MUlticast Routing Protocol (AMUR) [BaL07] uses Bloom filters and Diffie-Hellman key exchange protocols to provide efficient anonymity in ad hoc network environments. It is an extension of the unicast routing approach in ODAR to a multicast environment and augments the trapdoor approaches used in SDAR, AnonDSR and SDDR. The filters encode a source multicast tree in every multicast packet to provide an efficient means to preserve sender, receiver, and communication anonymity. However, the protocol offers no protection against a globally omniscient and active adversary and subsequent insertion and denial of service attacks.

2.3.2.7 HANOR.

HANOR [LiH06] is based on a hierarchical MANET architecture with multi-hop clustering, called groups, found in some military communication networks. It leverages the inherited group management security features to reduce the prohibitive computation and communication limitations of flat routing schemes such as AnonDSR, ASR, MASK and SDAR in larger-scale MANETs while preserving anonymity and providing additional intra-group and inter-group communication anonymity. However, the protocol was designed assuming a local, passive, and solitary adversary threat model instead of a much stronger global, active and multiple adversarial threat model.

2.3.2.8 ANODR.

ANODR [KoH03] is based on a "*broadcast with trapdoor information*" concept to achieve an untraceable and intrusion tolerant protocol for MANETs deployed in a hostile environment. It is an on-demand, first contact flood, virtual circuit data delivery, no neighbor exposure, and crypto-protected receiver anonymity protocol [KoH07]. It uses a route pseudonym approach and a symmetric key boomerang type onion, a layered cryptographic structure on which appending and peeling off are performed by the same forwarding nodes. It prevents strong adversaries from tracing a packet flow back to its source or destination (communication anonymity) and ensures that adversaries are unable to identify local message-forwarding nodes (location privacy). However, it has a trapdoor and anonymity issue [SoK05]. First, each forwarding node must impractically try all known shared secret keys. Second, how to establish shared session keys during the RREQ and route reply (RREP) phases is unspecified.

2.3.2.9 SDDR.

SDDR [ElK03] is based on a distributed route construction algorithm used for establishing anonymous routing paths in ad hoc networks such as wireless battlefield, on-the-fly conference, or emergency/rescue environments. The goal is to allow intermediate nodes to build paths without putting the communicating nodes anonymity at risk. SDDR does not require a global view of the network topology, is resilient against path hijacking, and provides protection against replay and modification attacks. Its limitations are an inability to change routes if under attack, constrained path lengths and non-minimal node computation power and storage requirements. Hence, it is very vulnerable to DoS attacks. It also ignores sender and receiver anonymity and does not provide strong location privacy [RaM06].

2.3.2.10 ASR.

ASR [ZhW04] is based on asymmetric cryptosystems and is designed to ensure the security of discovered routes and preserve sender, receiver, communications and location anonymity against known passive and active attacks. Unlike SDDR, it offers forwarding node, strong location, and communication anonymity. Unlike ANDOR, it offers sender, receiver, and strong location anonymity. However, it has the disadvantages of large computational latency, key size, and power consumption and an inability to dynamically repair failed routes. For instance, every forwarding node must generate a fresh public/secret key pair for every RREQ message it forwards and decrypt each RREP with every private key in its routing table [SeP06].

2.3.2.11 ZAP.

ZAP [WuB05] is a zone-based anonymous protocol designed to achieve destination *k*-anonymity in positioning routing algorithms. In this group-based approach, it uses wireless broadcast to give "false" positions near the destination and is based on a "crowd" of nodes so that anonymity depends on crowd size. It assumes uniformly distributed nodes, robust flooding, always-available GPS and public keys, symmetric radio channels, equal probability of being a source or destination node, and a global, passive, adaptive adversary. *k*-anonymity is preserved by initially choosing a large fixed zone or dynamically maintaining a *k*-sized zone based on node density and mobility.

2.3.2.12 AODPR.

AODPR [RaM06] uses a dynamic handshake mechanism to achieve sender, receiver, communication, and location anonymity for an ad hoc network of any node density. It uses a Virtual Home Region (VHR [Wux05])-based DIstributed Secure POsition SERvice (DISPOSER[Wux04]) where nodes stay in one VHR to obtain and report position information. A node varies density by being connected to neighbors in all four directions (*quad placement*), in a line of intermediate nodes (*line placement*), or to just one neighbor node (*least placement*). The source estimates the minimum number of hops to the destination and forwarding nodes also calculate distance to the destination. It computes a time variant temporary identifier from node time and position to circumvent a traffic analysis attack, thwart a wormhole attack, and protect against a DoS attack.

2.3.2.13 AO2P.

AO2P [WxB05] is based on asymmetric cryptosystems, uses a receiver contention scheme for route discovery (an anycast approach), uses pseudonyms and temporary MAC addresses for data delivery, and is designed for high density networks. It offers sender, forwarding node, communications and location anonymity but not receiver anonymity. A modified protocol R-AO2P [WuB05a] does improve receiver anonymity. However, Ao2P also has the disadvantage of large computational latency, key size, and power consumption. Hence, it may not scale well for larger networks.

2.3.2.14 SAS.

SAS [MiX06] is a simple and efficient scheme for establishing anonymity during node discovery and routing in clustered wireless sensor networks (CWSN). Neighboring nodes share pairwise symmetric keys and are assigned non-contiguous, uniformly distributed dynamic pseudonyms. This guarantees complete anonymity even in the presence of malicious and colluding neighboring nodes. It assumes the algorithm HEED [YoF04] is used to form clusters and that sensor network nodes are static thereafter. Therefore, the true dynamic nature of ad hoc networks is not captured.

2.3.2.15 ASC.

ASC [KaM07] is connection-oriented, based on a symmetric cryptosystem, and uses path and link encryption, and virtual circuit identifiers. It does not rely on any trusted agent or centralized mechanism and preserves sender, receiver, communications and location anonymity for video and audio streaming applications in MANETs. Compared to ANODR and AO2P, it may be the first anonymous routing protocol fast enough to

route real-time traffic while preserving anonymity and uses an adaptive transmission power scheme to improve network security and performance.

2.3.2.16 ASRPAKE.

Anonymous Secure Routing Protocol with Authenticated Key Exchange (ASRPAKE) [XiR07] is a proposed elliptic curve cryptosystem-based ring signature scheme designed to achieve anonymous authentication key agreement in MANETs. ASRPAKE augments the other MANET-based anonymous protocols of AnonDSR, MASK, SDAR, and ASR. As long as the entire routing path is not compromised, it offers end-to-end anonymity from the original sender to the intended receiver. Also, its embedded suite of authenticated key exchange mechanism ensures the security of the shared session key between sender and receiver. Quantifying anonymity is discussed next.

2.4 Quantifying Anonymity

To achieve anonymity, actions should be separated from the agents who perform them for some adversary. Anonymity in general as well as the anonymity of each particular agent or message is context dependent on the number of agents or messages, time frame, attributes, etc. A good deal of research has investigated different ways to measure anonymity. Typical analytical approaches to describe anonymous systems use simple quantifications and basic probabilistic models. Other approaches, covered in the following sections, produce formal frameworks for the more general description of anonymous systems. These formal approaches provide inspiration to search for future measures and methods for analyzing anonymous systems. A variety of practical anonymity metrics include, but are not limited to, anonymity set size, individual

anonymity degree, entropy anonymity, effective anonymity set size, normalized entropy anonymity degree, negligibility-based identity-free anonymity, localized real-time anonymity, combinatorial anonymity degree, evidence theory anonymity, *k*-anonymity and multicast anonymity.

2.4.1 Anonymity Set Size.

Anonymity set size is a traditional way to measure anonymity in an ACS. For example, the message sender is embedded in an *anonymity set* [Cha88, KeE98] of other honest, uncompromised senders. The cardinality of this anonymity set provides a numerical measure of Sender Anonymity. This metric has been used to evaluate the design of the Stop-n-Go MIXes [KeE98].

Informally, if the adversary knows the number of potential agents N prior to an attack and has compromised a number of agents C during the attack, then the *anonymity set size* $n = N - C$ quantifies the level of anonymity achieved after the attack. Formally, an equivalent derived definition is below.

(Derived) Definition 1 [KeE98] Assume an adversary threat model E, set of all agents A where $|A| < \infty$, anonymity set $AS \subset A$, message M, and agent $i \in A$. Let O denote the role (either a sender or receiver) of agent i. Further assume a priori anonymity set $AS' \subset A$ where $N = |AS'|$ and comprised set of agents $I \subset AS'$ where $C = |I|$ and $1 \leq C \leq N\text{-}1$. If the a priori probability $Q > 0$ that agent i has role O with respect to M with compromised agents I, then $i \in AS' - I$ with posterior probability $P \neq 0$. Any method to provide anonymity has an *anonymity set size* $n = N - C$.

The adversary's chances of identifying the agent's *i* role *O* increases (decreases) as the anonymity set size *n* decreases (increases). The set of possible agents depends on the knowledge of the adversary. Thus, anonymity is relative with respect to the adversary.

In open environments, the anonymity set of a receiver changes over time. Since the intersection of two different anonymity sets is likely to be smaller than either of the anonymity sets, different intersections of anonymity sets could be used to gain information about a specific agent or group of agents. Effectively, this leads to an anonymity set whose size shrinks as the adversary observes additional acts of communication by the same agent. The worst case is when an adversary reduces the anonymity set to size one or $n = N - C = n - (n - 1) = 1$. If the probability distribution of an agent performing an action is not uniform, then the anonymity set size may be a poor measure of anonymity in any real anonymous system. An individual anonymity degree metric is examined next.

2.4.2 Individual Anonymity Degree.

From the perspective of the adversary, the *anonymity degree* for each agent *i* in anonymity set *AS* at any point in time can be characterized by the scale in Figure 12.

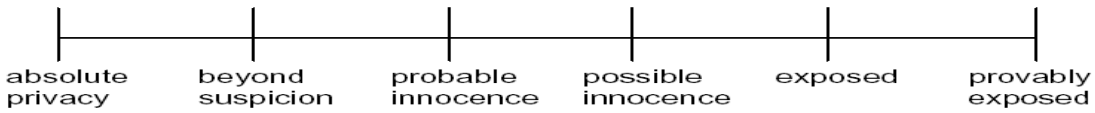


Figure 12: Individual Anonymity Degree Scale [ReR06]

The anonymity degrees range from absolute to none from left to right. The scale qualitatively describes anonymity degree and was first introduced in the design of Crowds [ReR98].

Consider an adversary trying to determine who the sender of a message is. On the far left, *absolute privacy* means no agent ever sends any message in the ACS. *Beyond suspicion* means agent *i* is no more likely to have sent the message than anyone else. This is the highest achievable level of anonymity for any set of agents and is also known as *total anonymity* or *strongly probabilistic*. *Probable Innocence* means agent *i* is no more likely to have sent the message than not sent the message. *Possible Innocence* means there is a non-trivial chance that an agent other than *i* sent the message. *Exposed* means there is a non-trivial chance that agent *i* is the sender of the message. *Provably Exposed* means agent *i* is the sender of the message. This means the adversary is absolutely certain who the sender of the message is and no anonymity exists. The next information theoretic entropy anonymity measure looks at the average uncertainty across the entire anonymity set.

2.4.3 Entropy Anonymity.

To overcome the limitations of the anonymity set metric, other researchers independently proposed information theoretic anonymity degree [DiC02, SeD02] based on information entropy [Sha48] that quantifies the level of uncertainty inherent in a set of data. The information-theoretic metrics of *entropy*, *conditional entropy*, *channel capacity*, and *effective anonymity set size* measures how random the probability distribution is and considers the global anonymity of the communication system.

Intuitively, each can be used as a measure to describe the average degree of anonymity or uncertainty of a system against a specific attack. The formal entropy definition based on [Kon05] follows,

Definition 2 [Kon05] For an event space *AS*, let X_{AS} be a discrete random variable with probability distribution $\Pr_i = \Pr[X_{AS} = i]$ where $j \in AS$. If the event space *AS* denotes an anonymity set, then X_{AS} represents the identity (similar to assigning an anonymity degree probability for each identified agent *i* as covered in the previous section). However, if the event space *AS* denotes the set of all items of interest or IOI (i.e., sender, receiver and messages), then X_{AS} represents the end-to-end routing path (being eavesdropped) between any sender and any receiver. The adversary's a priori knowledge is measured by $H(X_{AS})$ or *entropy*

$$H(X_{AS}) = -\sum_{i \in X_{AS}} \Pr(i) * \log_2 \Pr(i). \tag{3}$$

where *AS* is the anonymity set. The adversary's posteriori knowledge is measured by the *conditional entropy*

$$H(X_{AS} \mid C) = -\sum_{i \in X_{AS}, j \in C} \Pr(i, j) * \log_2 \Pr(i \mid j) \tag{4}$$

where *C* is the set of intercepted IOI (messages) or compromised IOI (agents), $\Pr(i,j)$ is the joint probability of agent *i* and intercepted IOI *j* and $\Pr(i|j)$ is the conditional probability that agent *i* is identified given the intercepted IOI *j*, where

$$\Pr(i \mid j) = \Pr(i \mid j) / \sum_{i \in X_{AS}, j \in C} \Pr(i, j). \tag{5}$$

In terms of anonymous communications, the *entropy* $H(X_{AS})$ of X_{AS} is the amount of uncertainty about the anonymous events, before executing the protocol. The *conditional entropy* $H(X_{AS}|\ C)$ gives the uncertainty of the adversary about the anonymous events after performing the observation [ChP07]. The higher the entropies are, the more uncertain the adversary is about the outcome. The communication *channel capacity* [ChP07] gives the maximum channel rate information is transmitted and measures anonymity loss or $\max_{Pr}[H(X_{AS}) - H(X_{AS}|\ C)]$.

Consider an entropy example. Let N be the number of agents and C be the number of compromised agents. Combining the previous *anonymity set size* definition, $n = N - C$, with the *entropy* anonymity, $H(X_{AS})$, the maximum entropy anonymity measure, H_{max}, at any point in time is $H_{max} = \log_2(N - C) = \log_2(n)$. Thus, *entropy* is also called *effective anonymity set size* [Dia05c]. For the Crowds protocol, *effective anonymity set size* is a function of N, C, and forwarding probability p_f and is

$$H(X_{AS}) = \frac{N - p_f(N - C - 1)}{N} \log_2 \left[\frac{N}{N - p_f(N - C - 1)} \right] + p_f \frac{N - C - 1}{N} \log_2 \left[\frac{N}{p_f} \right]. \qquad (6)$$

2.4.3.1 Effective Anonymity Set Size.

The *effective anonymity set size* metric measures the degree of success for an adversary on mixes and must be computed for each individual message going through the mix [Dia05c]. The anonymity provided by a mix can be determined for incoming messages (sender anonymity) or outgoing messages (receiver anonymity). For sender anonymity, the entropy of the probability distribution relating outgoing messages with all possible inputs is computed. For receiver anonymity, the entropy of the probability

distribution relating the chosen input with all possible outputs is computed. For both, the anonymity measure applies equally to each output/input message for a given period of time called a round. The actual anonymity metric computation depends on the type of mix the messages go through and if any dummy traffic is generated by the mix. If dummy traffic is generated, it matters if the dummy messages are inserted with the output messages or in the pool of input messages within the mix.

Let r be a round, a_r be the number of input messages, n_r be the number of messages in the pool, s_r be the total number of sent/output messages, and $P(n_r)$ be the probability a message leaves as a function of the number n_r of messages in the pool. Also, let $\Pr(I_{i,k})$ be the probability an output message matches the input message k of round i and $\Pr(O_{r,q})$ be the probability an input message matches the output message q of round r.

First, assume no dummy traffic. The sender anonymity H_S and receiver anonymity H_R metrics for a deterministic and binomial mix are shown in Table 3.

Table 3: Sender and Receiver Anonymity Metrics without Dummy Traffic [Dia05c]

Mix Type	Sender (H_s)	Receiver (H_R)
	$H_S = -\sum_{i=0}^{r} a_i \bullet \Pr(I_{i,k}) \log_2(\Pr(I_{i.k}))$	$H_R = -\sum_{r=0}^{\infty} s_r \bullet \Pr(O_{r,q}) \log_2(\Pr(O_{r,q}))$
Deterministic Mix	$\Pr(I_{i,k}) = \frac{1}{n_r} \prod_{j=i}^{r-1} (1 - P(n_j))$	$\Pr(O_{r,q}) = \frac{P(n_r)}{s_r} \prod_{j=i}^{r-1} (1 - P(n_j))$
Binomial Mix	$\Pr(I_{i,k}) = \frac{1}{n_r} \prod_{j=i}^{r-1} (1 - \frac{s_j}{n_j})$	$\Pr(O_{r,q}) = \frac{1}{n_r} \prod_{j=i}^{r-1} (1 - \frac{s_j}{n_j})$

Sender anonymity H_S is computed using the number of input messages a_r and the probability distribution of the output message matching the input messages $\Pr(I_{i,k})$ in the familiar entropy formula. The message probability distribution $\Pr(I_{i,k})$ depends on the

mix type. For a deterministic mix, the probability is the product of the probability the output message matches an input message from the current round $1/n_r$ and the probability the output message matches an input message still in the pool from a previous round $\prod_{j=i}^{r-1}(1-P(n_j))$. For a binomial mix, the probability is the product of the probability the output message matches an input message from the current round $1/n_r$ and the probability the output message matches an input message not previously sent out $\prod_{j=i}^{r-1}(1-\frac{s_j}{n_j})$ where $\frac{s_j}{n_j}$ is the percent of sent messages to total messages in the mix from prior rounds.

Receiver anonymity H_R is computed using the number of sent messages s_r and the probability distribution of the input message matching the output messages $\Pr(O_{r,q})$ in the familiar entropy formula. Theoretically, the adversary has to wait forever to compute receiver anonymity for any particular input message; however, practically, the adversary estimates receiver anonymity after waiting only a few rounds after the input message arrived at the mix. The message probability distribution $\Pr(O_{r,q})$ depends on the mix type. For a deterministic mix, the probability is the product of the probability the input message matches an output message $\frac{P(n_j)}{s_j}$ which only makes sense if a message has been output in the current round or $s_j > 0$ and the probability the input message matches an output message from a previous round $\prod_{j=i}^{r-1}(1-P(n_j))$. For a binomial mix, the computation is the same as sender anonymity.

Next, assume dummy traffic is generated. The sender anonymity H_{DS} metrics for

Table 4: Sender Anonymity with Dummy Traffic [Dia05c]

Sender Anonymity with dummy traffic (H_{DS})		
$H_{DS} = -p_d \log_2(p_d) - (1-p_d)\log_2(1-p_d) + (1-p_d)H_S$		H_s = sender anonymity w / o dummy traffic p_d = probability target message is a dummy
Output Insertion	$p_d = d_k / s_k$	d_k = dummy messages inserted in round k s_k = total messages sent at round k
Pool Insertion	$p_d = D_r / n_r$ $D_r = d_r + \sum d_i \prod_{j=i}^{r-1} (1 - \frac{s_j}{n_j})$	p_d = probability target message is dummy D_r = avg number of dummy messages in pool at round r n_j = number of messages in pool at round j

output and pool insertion are shown in Table 4.

The sender anonymity metric H_{DS} is a function of the probability the output message is a dummy message p_d and sender anonymity without dummy traffic H_S. The probability the output message is a dummy message depends on where the dummy message is inserted. If the mix inserts the dummy messages on the output, then this probability is simply the ratio of inserted dummy messages d_k to total messages sent s_k in round k or $p_d = d_k/s_k$. If the mix inserts the messages in the pool, this probability is the ratio of average dummy messages inserted in the pool D_r to messages in the pool n_r in round r or $p_d = D_r/n_r$. Of course, D_r is the number of dummy messages inserted this round d_r and in previous rounds $\sum_{i=1}^{r-1} d_i \prod_{j=i}^{r-1} (1 - \frac{s_j}{n_j})$.

2.4.4 Normalized Entropy Anonymity Degree.

Normalized entropy anonymity degree d is a relative entropy measure and is

$$d = H(X_{AS} \mid C) / H(X_{AS}) \tag{7}$$

An anonymous communication scheme has either perfect or preserved anonymity when $d = 1$. Perfect anonymity holds if $H(X_{AS})$ is the maximum entropy measure where $H_{max} = -\log_2 \Pr(i)$ where $\Pr(i)=1/n$, $n=|AS|, \forall\, i \in AS$ or when all agents in the anonymity set are *Beyond Suspicion*. Otherwise, anonymity is preserved if $H(X_{AS}) < H_{Max}$ or when agents have a non-uniform probability distribution. Any anonymity change may be measured by computing d after an attack and elapsed amount of time. Preserving anonymity is the "holy grail" of anonymous systems. Realistically, however, anonymity tends to degrade over time at a rate related to the increase of adversary knowledge. Hence, anonymity degree is characteristically bounded between [0,1].

Assume an adversary intercepts C during the attack and gains additional knowledge. This knowledge is reflected by the adversary adjusting the probability distribution for the receiver anonymity set. For instance, removing r agents from anonymity set AS such that $\Pr(r)=0, \forall\, r \in AS$ and/or changing k agent probabilities such that $\Pr(k) \neq \Pr(i)$, $i=k$, $k \in AS$. This decreases the adversary's uncertainty or $H(X_{AS}|C) < H(X_{AS})$. In the best case, the adversary may only be able to reassign a uniform probability distribution across the reduced sized anonymity set size $n - r$ such that $\Pr(i')_{i'}=1/(n-r), \forall\, i' \in AS,\ i'=1\ldots n-r$. Obviously, the closer d is to one, the less the system is compromised and the closer d is to zero, the more the system is compromised. Hence, ACS's may be quantitatively compared based on how much or how quickly anonymity is degraded. This entropy measure is not always practical so a negligibility-based anonymity measure [KoH07] is next.

2.4.5 Negligibility-based Identity-free Anonymity.

The negligibility-based anonymity probability metric assumes the adversary is a polynomial time algorithm (i.e., has limited resources) in terms of the number of participating nodes *N* in the anonymous network. Due to *identity-free* routing, the adversary cannot identify any mobile node's routing identity (e.g., IP address, MAC address). The goal is to achieve a *negligible* (indistinguishable) difference between true-randomness and pseudo-randomness, which is asymptotically less than the reciprocal of any polynomial of input *x* where *x* is the number of nodes, not cryptographic key length. The formal definition is

Definition 3 (*Negligible* [KoH07]). A function $\mu : \mathbb{N} \to \mathbb{R}$ is negligible if, for every positive integer c and all sufficiently large *x*'s (i.e., $\exists\ N_c\ \forall x > N_c), \mu(x) < \frac{1}{x^c}$.

It shall be shown that the probability of no anonymity is negligible (e.g., decreasing exponentially toward 0) when the number of mobile network nodes *N* increases linearly. A *venue* is the smallest area the adversary is able to pinpoint the mobile agent in radius *R* without differentiating two or more identity-free agents in a venue $A = \pi R^2$ as shown in Figure 13.

The bounded network has a spatial agent distribution expressed as the probability density function $\rho = f_{XY}(x, y)$. The probability a given agent is located in a subarea A_1 of the system area *A* or Pr[*node in* A_1] is computed by integrating ρ over this subarea A_1. The metric is extendable to *k* agents and the venue area may be any bounded shape.

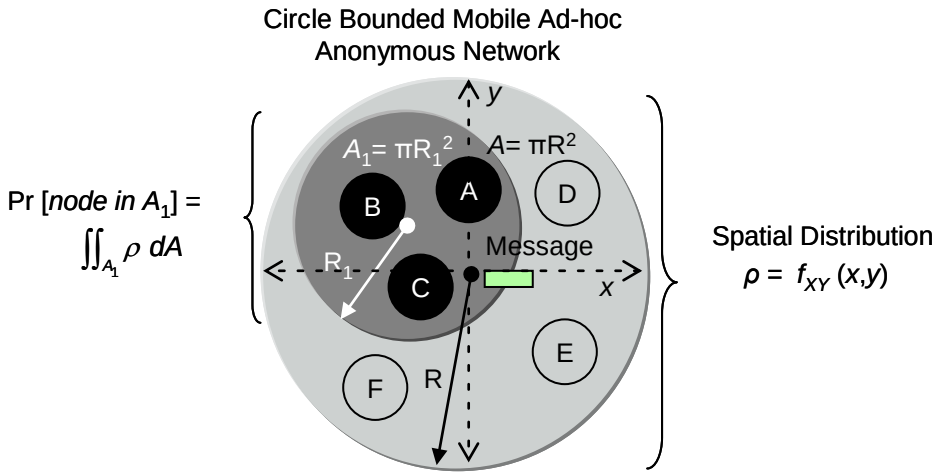


Figure 13: Negligibility-based Anonymity Metric (Pr[*node in* A_1]) given agent Spatial Distribution (ρ)

The probability a given node is located in a subarea A_1 of the system area A is computed by integrating ρ over this subarea

$$\Pr[node\ in\ A_1] = \iint_{A_1} f_{XY}(x,y)dA \tag{8}$$

which is universally applicable to any mobility pattern.

An example random waypoint mobility model is $\rho = f_{XY}(x,y) \approx \frac{36}{a^6}(x^2 - \frac{a^2}{4})(y^2 - \frac{a^2}{4})$ [BeR03]. With N agents, $\rho_N = \sum_{i=1}^{N} \rho_i$, where ρ_i is agent i's probability density function and $\rho_N = N\rho$ if roaming agents are independently and identically distributed. Let x be a random variable of the number of nodes in the area, then the probability of exactly k nodes in area A_1 is

$$\Pr[x = k] = \iint_{A_1} (\frac{\rho_N^k}{k!} e^{-\rho_N})\ dA. \tag{9}$$

The probability a venue is empty is

$$P_{empty} = \Pr[x = 0] = \iint_{\pi R_1^2} e^{-N\rho}\ dA = O(e^{-N\rho}) \tag{10}$$

since $e^{-N\rho}$ remains an exponential in differential and integral calculus. Thus, as the number of nodes *N* increases linearly, P_{empty} approaches zero.

If all nodes are moving, the adversary needs at least one empty venue to trace the identity-free node *v*. The probability the adversary traces node *v* along a sequence of *m* empty venues is

$$P_{trace\_motion} = (P_{empty})^m = O(e^{-N\rho m}) \tag{11}$$

This is the *negligible-based, identity-free anonymity* metric with respect to network size *N*. Localized anonymity for real-time systems is explored next.

2.4.6 Localized Real-time Anonymity.

To measure local anonymity in a non-adaptive, real-time system, *source-hiding* and *destination-hiding* properties in a formal PROB-channel model are analyzed and quantified [TgH04]. Assume a system has *senders* (s) transmitting encrypted *sent messages* (α) to the anonymous system. After transforming and delaying the sent messages, the *delivered messages* (β) reach the *receivers* (r). The passive adversary attempts to break sender anonymity by computing $P(\beta, s)$ and receiver anonymity by computing $P(\alpha, r)$, respectively. A system is *source-hiding* with parameter θ if the adversary cannot assign a sender to any delivered message with a probability greater than θ, i.e., if

$$\forall_\beta \forall_s (P(\beta, s) \le \theta). \tag{12}$$

This is also called *source* or *sender anonymity* [PfK00].

Similarly, a system is *destination-hiding* with parameter Ω if the adversary cannot assign a receiver to any sent message with a probability greater than Ω, i.e., if

$$\forall_{\alpha}\forall_{r}(P(\alpha,r)\leq\Omega). \tag{13}$$

This is also called *destination, receiver* or *recipient anonymity* [PfK00].

However, it is essential to give a theoretically based but also practically usable objective numerical measure for local anonymity. An analysis of the previous global entropy-based metrics [TgH04a] on the anonymous message transmission, continuous time PROB-channel model [TgH04] reveals shortcomings like an anonymous system appears near-optimal yet the adversary still is able to guess the sender of some messages with high probability. Also, the exponential computational complexity of the adversary globally tracking and assigning sender probabilities is impractical. Thus, an argument is made to use the maximum probability that an attacker can assign to a sender or receiver with respect to a particular message as a measure. This amounts to the sender specifying a Quality-of-Service (QoS) *threshold* for anonymity services depending on underlying frequency parameters ($\tau_{\min}$ and $\tau_{\max}$) and channel delay characteristics ($f(\delta)$). Such a measure may be of more interest to individual users of the system to better capture the local aspects of anonymity.

For instance, if no sender sends more than one message within a minimum time interval $\tau_{\min}$ and all senders send at least one message in a maximum time interval $\tau_{\max}$, then a practical upper limit $\hat{P}(\beta,s)$ and guaranteed localized *source-hiding* measure is

$$\hat{P}(\beta,s)=\frac{\sum_{i=1}^{\Delta_{\min}}\max_{(i-1)\tau_{\min}\leq q\leq i-\tau_{\min}} f(\delta)}{|S|\cdot\sum_{i=1}^{\Delta_{\max}}\min_{(i-1)\tau_{\max}\leq q\leq i-\tau_{\max}} f(\delta)} \tag{14}$$

where $f(\delta)$ is a message and time-invariant channel delay distribution function, $\Delta_{max} = \left\lfloor \frac{\delta_{max} - \delta_{min}}{\tau_{max}} \right\rfloor$ and $\Delta_{min} = \left\lfloor \frac{\delta_{max} - \delta_{min}}{\tau_{min}} \right\rfloor$, δ_{max} and δ_{min} are predefined, message and time-invariant maximal and minimal channel delays, and S is the set of senders.

Simplifying this equation demonstrates how this *localized sender anonymity* measure reduces to an optimal global anonymity measure of *perfect sender anonymity*. Assuming the channel delay distribution function $f(\delta)$ is uniform ($f(\delta) = f_{max} = \frac{1}{\delta_{max} - \delta_{min}}$) and MIN/MAX properties hold ($\tau_{min} \leq \tau_{max} \leq [\delta_{max} - \delta_{min}]$), the upper limit $\hat{P}(\beta, s)$ and guaranteed localized *source-hiding* measure becomes

$$\hat{P}(\beta, s) = \frac{\Delta_{min}}{|S| \cdot \Delta_{max}} \approx \frac{\tau_{max}}{|S| \cdot \tau_{min}}. \tag{15}$$

Furthermore, if each sender sends messages with the same periodicity ($\tau_{min} = \tau_{max}$), perfect anonymity is achieved as the adversary ascribes a uniform probability distribution for all senders S

$$\hat{P}(\beta, s) \approx \frac{1}{|S|}. \tag{16}$$

Hence, specifying the message sender frequency with the parameters τ_{min} and τ_{max} allows three different ways to measure localized sender anonymity including

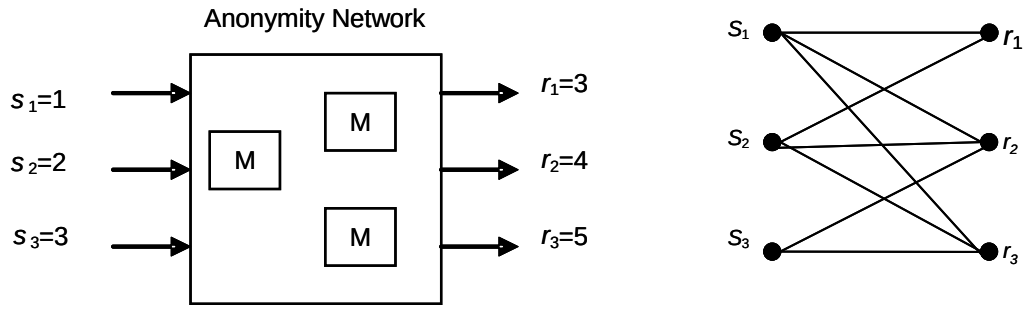
1) Message sender frequency is constrained ($\tau_{min} \leq \tau_{max}$),
2) Uniform distributed channel delay ($f(\delta) = \frac{1}{\delta_{max} - \delta_{min}}$) and MIN/MAX property hold ($\tau_{max} \leq [\delta_{max} - \delta_{min}]$)
3) Message sender frequency is periodic ($\tau_{min} = \tau_{max}$)

Unfortunately, a similar *destination-hiding* measure is not realizable due to the limitations of the PROB-channel model [TgH04, TgH04a]. Specifying the message frequency of receivers would either require the difficult task of coordinating the senders in a distributed environment or injecting dummy traffic on the channel, implying an active adversary. A combinatorial anonymity degree follows.

2.4.7 Combinatorial Anonymity Degree.

The anonymity set size, effective anonymity set size, entropy anonymity, and normalized entropy anonymity measures primarily determine the anonymity degree from the perspective of a single agent or message. The *combinatorial anonymity degree* [EdS07] is a combination of the individual agent anonymity levels and is a complementary system-wide measure based on the permanent of a matrix. The measure reveals the whole communication pattern between senders and receivers in a minimally (∇_{min}) and maximally (∇_{max}) delay-bounded real-time anonymous mix network.

Given a set of n senders ($s_i \in S$) and n receivers ($r_j \in R$) of an anonymous network and a set of possible mappings between the inputs and outputs (E), a bipartite graph $G = (S, R, E)$ represents the anonymous mix network. The timestamps of the entering and exiting messages are the only observable information. If $n = 3$, then $S = \{s_1, s_2, s_3\}$ and $R = \{r_1, r_2, r_3\}$ and an example anonymous three mix network and bipartite graph is shown in Figure 14.



(a) Observed Entry and Exit Times (b) Corresponding Bipartite Graph

Figure 14: Sample Mix Network and Graph ($\nabla_{min} = 1$, $\nabla_{max} = 4$)

In Figure 14(a), if $\nabla_{min} \leq r_j - s_i \leq \nabla_{max}$, then the input s_i maps to output r_j and is an edge in graph G or $(s_i, r_j) \in E$. For example, if $\nabla_{min} = 1$ and $\nabla_{max} = 4$, then $r_1 - s_1 = 3 - 1 = 2$ and $\nabla_{min} \leq 2 \leq \nabla_{max}$ so $(s_1, r_1) \in E$ but $r_1 - s_3 = 3 - 3 = 0$ and $0 < \nabla_{min}$ so $(s_3, r_1) \notin E$. In Figure 14(b), the corresponding bipartite graph is shown.

From these observed input-output timestamp correlations, the global adversary forms probability distributions on links and constructs a special doubly-stochastic n x n matrix P. An anonymous mix network is shown in Figure 15.

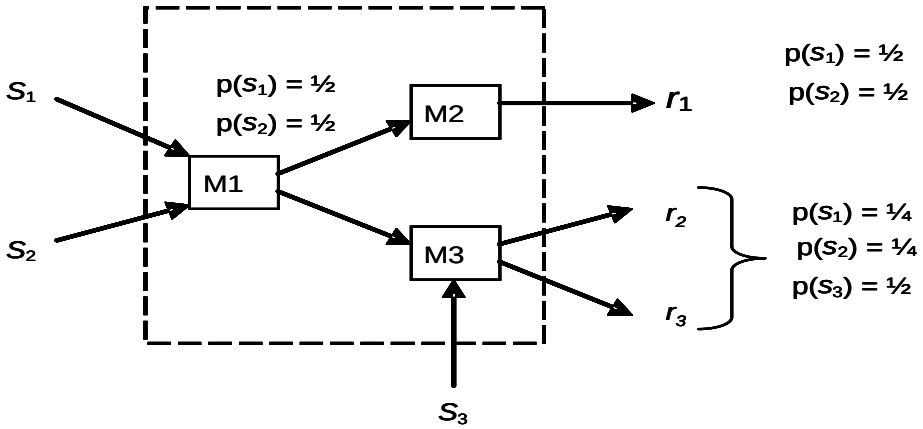


Figure 15: Example Mix Network with Probabilities

Three messages enter and exit the system, and each message entering a mix is equally likely to follow any outgoing link. The probabilities represent the likelihood of messages being on a particular link. The resulting matrix *P* is in Figure 16.

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}$$

Figure 16: Corresponding Doubly-Stochastic Matrix

The permanent of the matrix *per*(*P*) is computed as follows:

$$per(P) = \sum_{\pi} \prod_{i=1}^{n} P(i, \pi(i)) \tag{17}$$

where $\pi(i)$ is the a priori probability *per*(*P*) and is bounded by the inequality $n!/n^n \leq per(P) \leq 1$ via the proven *Van der Waerden conjecture* [Egr81, Fal81]. Referencing the doubly-stochastic *n* x *n* matrix example in Figure 16 where *n* = 3, *per*(*P*) = (½)(½)(0) + 2(¼)(¼)(½) = $\frac{1}{4}$ and the a priori lower bound is $n!/n^n = 3!/3^3 = 6/27 = \frac{2}{9}$.

The *combinatorial anonymity degree d*(*P*) represents the system-wide strength of the anonymous network and is

$$\mathrm{d}(P) = \begin{cases} 0 & n=1 \\ \dfrac{\log(\mathrm{per}(P))}{\log(\frac{n!}{n^n})} & n>1 \end{cases}. \tag{18}$$

Clearly, with only one sender and receiver (n=1), no anonymity exists ($d(P)$=0). With more than one sender and receiver (n>1), anonymity degree is quantified as the ratio of the log of the matrix permanent over the log of the lower bound of the a priori probability. When $per(P) = n!/n^n$ or the matrix permanent equals the lower bound, perfect anonymity is achieved ($d(P)$=1) otherwise a lower level of anonymity is achieved ($d(P)$<1). Continuing with the example mix network where $n = 3$, $d(P) = \log(per(P))/\log(n!/n^n) = \log(\frac{1}{4})/\log(\frac{2}{9}) = 0.92$. Hence, the system-wide *combinatorial anonymity degree* is strong but not perfect. Another anonymity measure is based on evidence theory.

2.4.8 Evidence Theory Anonymity.

The evidence theory based approach measures communication anonymity in wireless mobile ad-hoc networks [Dij06]. Evidence theory represents the belief-based epistemic knowledge of the adversary. Evidence is measured by the number of detected packets within a given time period. Basic probability assignments for all packet delivery paths are assigned and evidence theory quantifies anonymity in the number of *bits*. This approach is more general and practical than the entropy based metrics where the probability assignments are predefined [Dij06].

A captured packet is evidence that proves communication between two or more mobile nodes. The quantity of evidence, $w(V)$, for two communicating mobile nodes is

$$w(V) = \min_{U \subseteq V}\{w(U)\}, \quad |V| \geq 2 \tag{19}$$

where X is the set of all mobile nodes within the system, $\mathscr{P}(X)$ is the power set of X, $V \in \mathscr{P}(X)$ is the packet-sequenced ordered set and $U \subseteq V$. The normalized value $m(V)$ is the ratio for an acting communications relation defined in $\mathscr{P}(X)$ for each $V \in \mathscr{P}(X)$ or

$$m(V) = w(V) / \Sigma_{U \in \mathscr{P}(X)} w(U). \tag{20}$$

From evidence theory (a.k.a., *Dempster-Shafer theory* [Sen02, Sha76]), the *basic probability assignment* function is $m : \mathscr{P}(X) \rightarrow [0,1]$ such that $m(\varnothing) = 0$ and $\Sigma_{V \in \mathscr{P}(X)} m(V) = 1$. Every set $V \in \mathscr{P}(X)$ for which $m(V) \neq 0$ is a *focal element*. A *focal element* is a sender and receiver pair $v \in V$ the adversary believes is communicating indicated by an assignment of a non-zero probability measure m. $\langle \mathscr{F}, m \rangle$ is the set of all focal elements induced by m called a *body of evidence*. Given this assignment, the upper and lower bounds of the anonymity measure are defined. The lower bound *belief measure* is a function Bel: $\mathscr{P}(X) \rightarrow [0,1]$ and combined with a basic probability assignment m is $Bel(V) = \Sigma_{U|U \subseteq V} m(U)$. The upper bound *plausibility measure* is $Pl(V) = \Sigma_{U|U \cap V > 0} m(U)$ and $Pl(V) \geq Bel(V)$.

To measure uncertainty, the entropy-like measures $E(m) = \Sigma_{V \in \mathscr{F}} m(V) \log_2 Pl(V)$ and $C(m) = \Sigma_{V \in \mathscr{F}} m(V) \log_2 Bel(V)$ based on the plausibility [Hoh82] and belief [Yag83] are proposed. Because too many irrelevant sets are considered, $E(m)$ is not a satisfactory upper bound anonymity measure in wireless environments. Hence, the *discord* function $D(m)$ is the generalized anonymity measure in number of *bits* [Dij06]

$$D(m) = -\sum_{V \in F} m(V) \log_2 (1 - \sum_{U \in F} m(U) \frac{|U - V|}{|U|}). \quad (21)$$

The $\sum_{U \in F} m(U) \frac{|U - V|}{|U|}$ term factors out any irrelevant or conflicting evidence. $D(m)$ is a weighted version of belief measure $C(m)$ where $E(m) \leq D(m) \leq C(m)$ holds. $D(m)$ measures average anonymity for any given communication scenarios without probability pre-assignment to each individual node.

A wireless ad-hoc networking system with seven nodes, $X = \{A, B, C, D, E, F, G\}$, and eleven possible communicating pairs is shown in Figure 17.

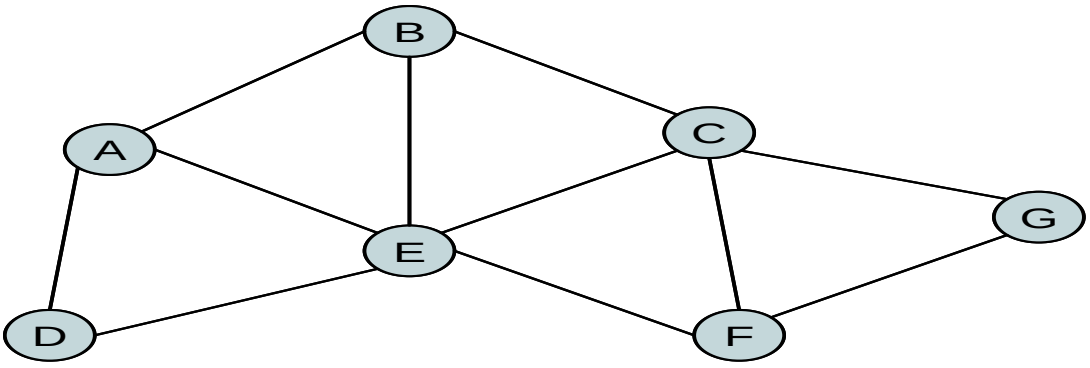


Figure 17: Example 7 Node MANET [Dij06]

A sophisticated adversary knows the exact location of each mobile node and can detect the transmitted packet source within the communication range of each mobile node. So the adversary partitions the MANET into multiple hexagon zones with at most one node per zone as shown in Figure 18.

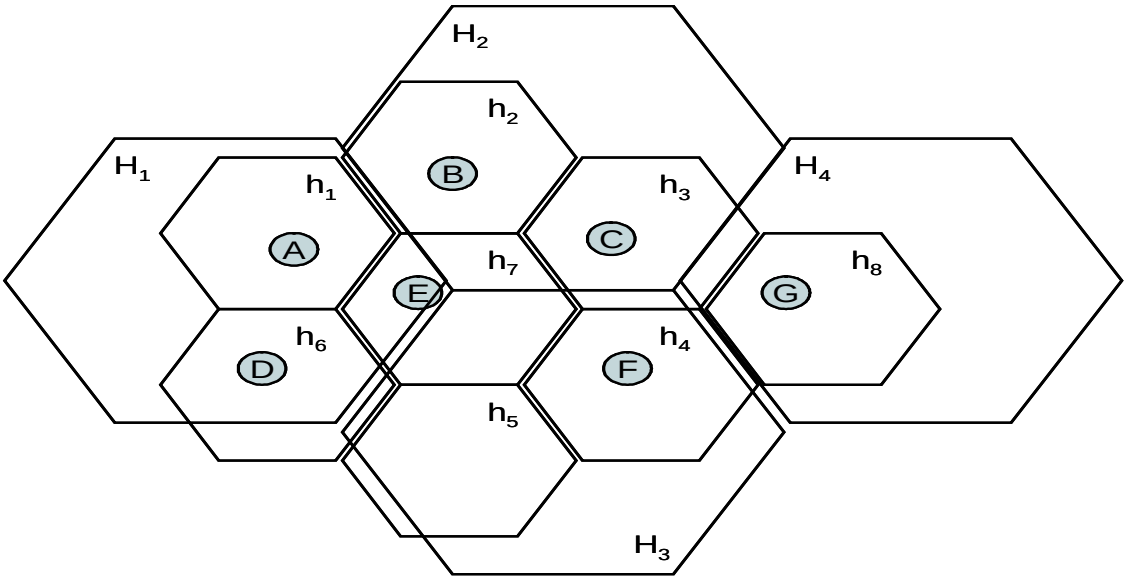


Figure 18: Communication Area Partitions

The adversary is able to monitor packets to/from these zones $h_1 - h_8$ and learn the topology in Figure 17. For instance, with a time period Δt, the adversary detects exactly one sent packet from the hexagon zones h_1, h_2, and h_4 corresponding to nodes *A*, *B*, and *F*, respectively. The adversary computes $w(V)$, $m(V)$, $Bel(V)$, and $Pl(V)$ where $V \in \mathscr{P}(X)$ as shown in Table 5.

Table 5: Body of Evidence

#	$\mathscr{F}$	$w(V)$	$m(V)$	$Bel(V)$	$Pl(V)$
1	<*A*, *B*>	1	1/11	1/11	8/11
2	<*A*, *D*>	1	1/11	1/11	6/11
3	<*A*, *E*>	1	1/11	1/11	8/11
4	<*B*, *A*>	1	1/11	1/11	8/11
5	<*B*, *C*>	1	1/11	1/11	7/11
6	<*B*, *E*>	1	1/11	1/11	8/11
7	<*F*, *E*>	1	1/11	1/11	6/11
8	<*F*, *C*>	1	1/11	1/11	5/11
9	<*F*, *G*>	1	1/11	1/11	3/11
10	<*A*, *B*, *C*>	1	1/11	1/11	9/11
11	<*A*, *B*, *F*>	1	1/11	1/11	9/11
	$\sum$	11	1		

The *w*-values in lines 1-9 are derived directly from observing the wireless system; the *w*-values in line 10 and line 11 are derived by applying (19) and using (20), each *focal element* such as <*A, B*> and <*A, B, E*> has a non-zero *m*-value of $1/11^{th}$. Based on the *lower bound E*(*m*), *upper bound C*(*m*), and *discord D*(*m*) equations above, the adversary computes the anonymity measures *E*(*m*) = 0.76 *bits*, *C*(*m*) = 1.74 *bits*, and *D*(*m*) = 3.17 *bits*. The maximum entropy is $\log_2 | X | = \log_2 11 = 3.46$ *bits*. Therefore, the anonymity measure of the mobile ad-hoc network ranges from 0.76 to 3.17 *bits* and is, on average, 1.74 *bits* within the time period Δt.

2.4.9 *k*-Anonymity.

In general, *k-anonymity* is a privacy preservation method to ensure an adversary is unable to distinguish an identity/item of interest among at least *k*-1 other identities/items of interest and is a NP-hard problem [AgF05, MaW04]. Many research efforts have proposed approaches to achieve *k-anonymity* and preserve data privacy [AgF05, KiG06, LeD06, MaW04, MwX06, NeC06, SaS98, Swe02] or location privacy [GeL04, GeL05, GeL07, GhK06, KaG06, Liu07, WuB05]. Some research efforts recommend multi-dimensional anonymization measures of *l-diversity* [MaG06], *m-invariant* [Liu07] and *t-closeness* [LiL07] which go beyond the typical *k-anonymity* approaches [Iye02] to improve data or location privacy under specific adversary attacks. This section describes three measures: *data privacy k-anonymity*, *destination k-anonymity zone*, and *personalized location k-anonymity*.

2.4.9.1 Data Privacy *k*-Anonymity.

The first *k-anonymity* measure [Swe02] addresses data privacy protection of releasable person-specific table-based information to third party organizations. The assumption is that the data holder can accurately identify *quasi-identifiers* [Dal86], namely a set of private data attributes that also appear in external information. These quasi-identifiers include explicit identifiers such as name, address, and phone number, as well as attributes such as birth date and gender which may uniquely identify an individual. The goal is to limit an adversary's ability to link released person-specific data to other information. This formal definition of *k-anonymity* follows.

Definition 3 (*k-Anonymity* [Swe02]). Let $RT(A_1,...,A_n)$ be a releasable table *RT* with attributes $\{A_1,...,A_n\}$ and QI_{RT} be the associated quasi-identifier set $\{A_i,...,A_j\} \subseteq \{A_1,...,A_n\}$. The releasable table *RT* satisfies *k-anonymity* if and only if each sequence of values in $RT[QI_{RT}]$ appears with at least *k* occurrences in $RT[QI_{RT}]$. An example of an *RT* table that adheres to *k-anonymity* is in Table 6.

Table 6: *k-anonymity* example, where *k*=2 and QI= {*Race, Birth, Gender, Zip*} [Swe02]

Tuple	Race	Birth	Gender	Zip	Problem
t1	Black	1965	M	0214*	Short breath
t2	Black	1965	M	0214*	Chest pain
t3	Black	1965	F	0213*	Hypertension
t4	Black	1965	F	0213*	Hypertension
t5	Black	1964	F	0213*	Obesity
t6	Black	1964	F	0213*	Chest pain
t7	White	1964	M	0213*	Chest pain
t8	White	1964	M	0213*	Obesity
t9	White	1964	M	0213*	Short breath
t10	White	1967	M	0213*	Chest pain
t11	White	1967	M	0213*	Chest pain

The quasi-identifier is $QI_{RT} = \{Race, Birth, Gender, Zip\}$ and k=2. For each tuple, the values that make up the quasi-identifier appear at least twice in *RT*. In other words, each sequence of values in $RT[QI_{RT}]$ has at least *2* occurrences of those values in $RT[QI_{RT}]$. Specifically, $t1[QI_{RT}] = t2[QI_{RT}]$, $t3[QI_{RT}] = t4[QI_{RT}]$, $t5[QI_{RT}] = t6[QI_{RT}]$, $t7[QI_{RT}] = t8[QI_{RT}] = t9[QI_{RT}]$, and $t10[QI_{RT}] = t11[QI_{RT}]$. So data privacy is preserved.

2.4.9.2 Destination k-Anonymity Zone.

This zone-based *k-anonymity* measure [XiB05] addresses destination location privacy protection in positioning routing protocols in mobile ad-hoc networks. The assumptions are uniformly distributed nodes, high node density, globally available position information and public keys, and symmetric communication channels. Also, the adversary is assumed to trace node behavior and obtain location updates but is unable to identify the sender or location position requesting nodes. An anonymity zone is generated for each destination called the D-AZ as shown in Figure 19.

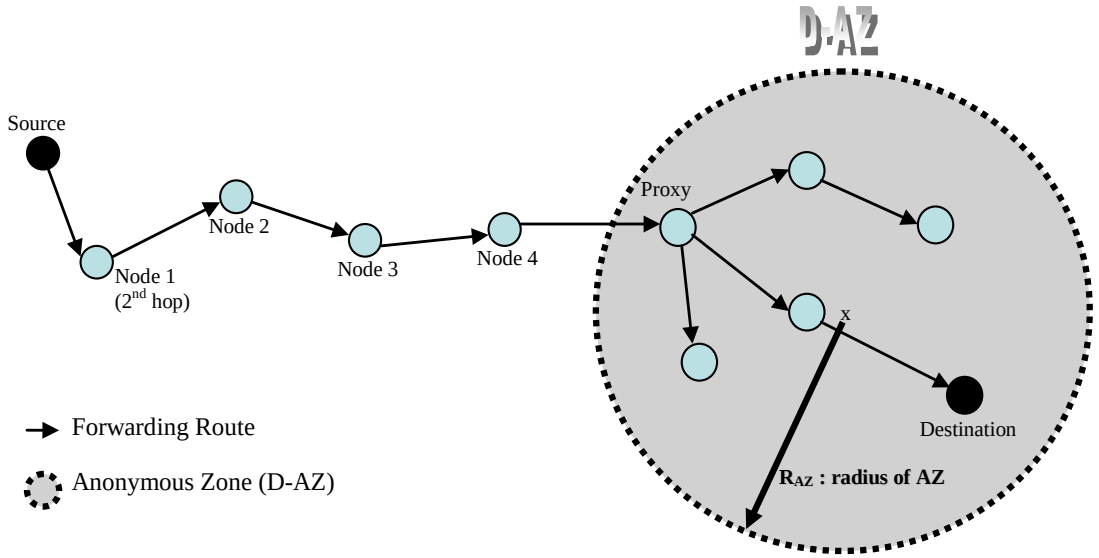
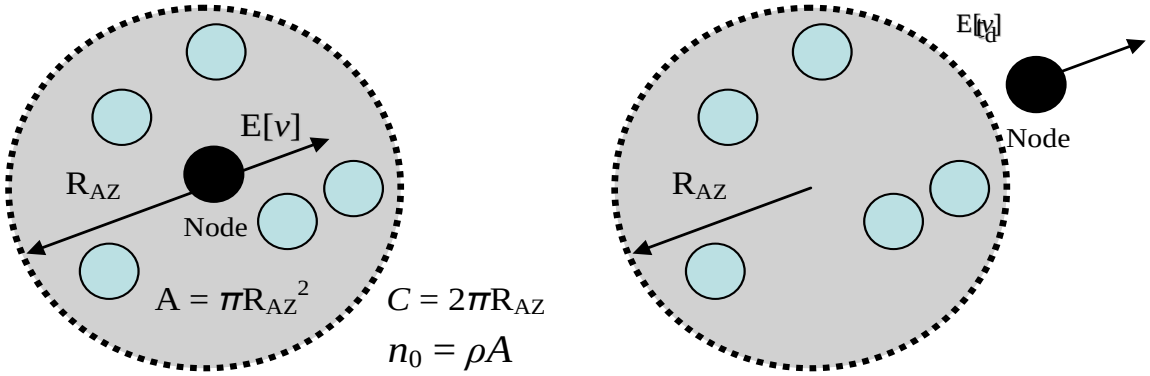


Figure 19: *k-anonymity* Based Private Positioning Routing [XiB05]

The source node generates the D-AZ by specifying the adversary observable center *x* and radius R_{AZ} information in the route request (RREQ) packet. The RREQ also carries destination challenge information to keep the destination private. The problem is node mobility degrades destination anonymity, especially with an intersection attack [XiB05]. A fixed D-AZ and adaptive D-AZ are approaches to preserving location privacy and achieving destination *k-anonymity*. For the fixed D-AZ, the source node originally uses a large-sized D-AZ ($n_0 >> k$) where n_0 is the initial number of nodes in the zone and as time passes and nodes move out the source aims to keep *k* or more nodes in the zone. A fixed D-AZ scenario is depicted in Figure 20.



(*a*) Node inside D-AZ (*b*) Node outside D-AZ

Figure 20: Fixed D-AZ *k-anonymity*

In Figure 20(*a*), the destination zone has radius R_{AZ} in meters (*m*), area *A* in square meters (m^2), circumference *C* in meters (*m*), node density ρ in nodes per square kilometer ($nodes/km^2$) and initial nodes n_0. Assuming $R_{AZ} = 200\ m$ and density $\rho = 50\ nodes/km^2$,

then $n_0 = \rho A = \rho\pi(R_{AZ})^2 = (50\pi\ nodes/km^2)((200m)\ (\frac{1}{1000}\ km/m))^2 = (50\pi\ nodes/km^2)(\frac{1}{25} km^2) = 2\pi\ nodes = 6\ nodes$. As indicated, $n_0 = 6\ nodes$ is the initial number of nodes in the D-AZ. In Figure 20(*b*), after a period of time t_d in seconds (*sec*) a node exits the D-AZ with constant velocity $E[v]$ in meters per second (*m*/*sec*). The probability of preserving destination *k-anonymity* is

$$P\{n \geq K-1\} = p(1-\sum_{i=1}^{K-1} P\{n=i\}) \tag{22}$$

where p is the probability the destination node stays in D-AZ and $P\{n=i\}$ is the probability that i nodes (k-1 other nodes) stay in the D-AZ. Assume 2-anonymity, $k = 2$, is the goal, then $i = k\text{-}1 = 1$ so $P\{n \geq 1\} = p(1-P\{n=1\})$. $P\{n=i\}$ is further defined as

$$P\{n=i\} = \frac{(n_0-1)!}{i!(n_0-i-1)!} p^i (1-p)^{n_0-i-1} \tag{23}$$

where i is the number of nodes in the D-AZ, n_0 is the initial number of nodes, and p is the probability a node stays in the D-AZ. In the example, $n_0 = 6$ and $i = 1$ so (23) is $P\{n=1\} = \frac{(6-1)!}{1!(6-1-1)!} p^1(1-p)^{6-1-1} = \frac{5!}{4!} p(1-p)^4 = 5p(1-p)^4$. Substituting $P\{n=1\}$ into (22) yields $P\{n \geq 1\} = p(1-5p(1-p)^4)$. Now p is further defined as

$$p = p\{t_d > t_1\} = \int_{t_1}^{\infty} f_{t_d}(t_d)dt_d = e^{-t_1/\bar{t}_d} \tag{24}$$

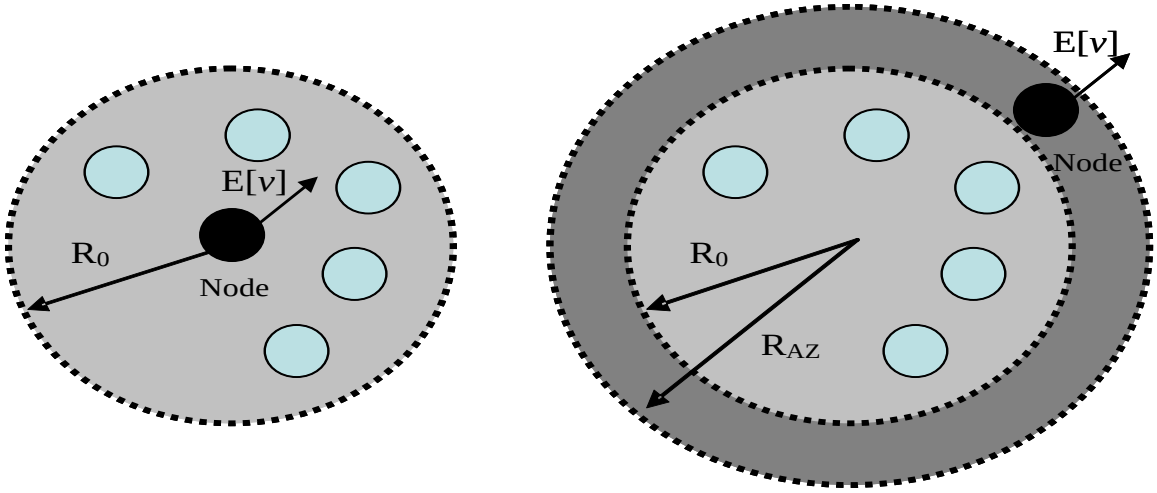
where t_d is the time the destination node stays in the D-AZ in seconds (*sec*), $f_{t_d}(t_d)$ is the probability density function of the destination staying in the D-AZ (exponential in this case), $\bar{t}_d$ is the mean node time in the D-AZ in seconds (*sec*), and *p* is the probability the destination node stays in the D-AZ beyond time t_1. Finally, $\bar{t}_d$ is

$$\bar{t}_d = \pi A / E[v]C = \pi R_{AZ} / 2E[v] \tag{25}$$

where *A* is the area of D-AZ, *C* is the circumference of D-AZ, R_{AZ} is the zone radius, and $E[v]$ is the node velocity. Assuming mobile nodes move at a velocity of $E[v]$ = 1 *m/sec* and the same radius R_{AZ} = 200 *m* as before, then (25) simplifies to $\bar{t}_d = 200\pi/2$ *sec* = 100π *sec*. Plugging into (24) yields $p = e^{-t_1/\bar{t}_d} = e^{-t_1/100\pi}$. After waiting t_1 = 60 *sec*, the probability the destination node stays in the D-AZ is $p = e^{-60/100\pi} = e^{-3/5\pi} = 0.826$. Since *p* = 0.826, the probability of preserving destination *2-anonymity* after one minute using the Fixed D-AZ method is $P\{n \geq 1\} = p(1-5p(1-p)^4) = (0.826)(1-5(0.826)(1-0.826)^4)$ = 0.823. After waiting t_1 = 300 *sec*, the probability the destination node stays in the D-AZ is $p = e^{-300/100\pi} = e^{-3/\pi} = 0.385$. The probability of preserving destination *2-anonymity* after only five minutes drops to $P\{n \geq 1\} = (0.385)(1-5(0.385)(1-0.385)^4)$ = 0.279. Since anonymity degrades rapidly after only a few minutes, an adaptive D-AZ approach is considered.

For adaptive D-AZ, the source determines the size of D-AZ (=*k* nodes) based on node density and as time passes expands D-AZ based on mobility to encompass nodes moving outside the D-AZ. An adaptive D-AZ scenario is depicted in Figure 21. In Figure 21(*a*),

the destination zone has initial radius R_0 and *k-anonymity* where $k = 6$. In Figure 21(*b*), after a period of time a node exits the D-AZ and the radius R_{AZ} is updated to



(*a*) Node inside D-AZ (*b*) Node outside D-AZ

Figure 21: Adaptive D-AZ *k-anonymity*

ensure *k-anonymity* after time t_1. Preserving *k-anonymity* requires the source to linearly expand the radius as

$$R_{AZ}(t_1) = c(t_1 + t_0) - R_0 \tag{26}$$

where $R_0 = \sqrt{\frac{k}{\pi\rho}}$ is the initial radius, $t_0 = -\bar{t}_d \ln(P_k)/k$ is the time when achieving *k-anonymity* is low (defined as $P_k(t) \leq$ threshold probability p_0), t_1 is the time when the radius is expanded, c is the constant R_0/t_0, and $R_{AZ}(t_1)$ is the expanded radius at time t_1. Again, ρ is node density and $\bar{t}_d$ is the mean node time in the D-AZ. Additionally, $P_k(t) = e^{-kt/\bar{t}_d}$ is the probability that k nodes are in the D-AZ after time t. If the goal is again $k = 2$ with node density $\rho = 50$ *nodes*/*km*^{2}, then the initial radius $R_0 = \sqrt{\frac{2\ km^2}{50\pi}} * 1000\ m/km =$

$1000\sqrt{\frac{1}{25\pi}}m = 113m$. Also, if the mean time in the zone is $\bar{t}_d = 100\pi\, sec$ and the threshold probability is $p_0 = 0.8$, then $t_0 = -\bar{t}_d \ln(p_0)/k = -100\pi \ln(0.8)/2 = -50\pi \ln(0.8) = 35\ sec$. Thus, the initial radius of 113 meters must be expanded after 35 seconds. With $R_0 = 113m$ and $t_0 = 35\ sec$, $c = R_0/t_0 = (113/35)m/sec = 3.23\ m/sec$. Finally, $R_{AZ}(t_1) = c(t_1 + t_0) - R_0 = 3.23m/sec(t_1 + 35sec) - 113m = (3.23(t_1 + 35) - 113)m$. In other words, at time t_1 the radius is linearly expanded to $R_{AZ}(t_1)$ to preserve *2-anonymity*.

2.4.9.3 Personalized Location *k*-Anonymity.

The third *k*-anonymity model protects against various privacy threats through sharing location information. When requesting *k*-anonymity, each mobile agent specifies an acceptable minimum *k*-anonymity level and maximum temporal and spatial resolution. A scaleable and efficient CliqueCloak algorithm, which perturbs location information in messages, provides high quality *personalized location k-anonymity* for forwarding agents. An agent is location *k*-anonymous if and only if the location information sent from a mobile agent is indistinguishable from the location information of *k*-1 other agents. The location-based service (LBS) system consists of anonymity servers, mobile agents, a wireless network, and LBS servers. The two location *k*-anonymity techniques are spatial expansion and temporal cloaking.

Let *S* be the set of received messages from the mobile agents. Each received message $m_s \in S$ has a unique identifier u_{id} and a three dimensional *spatio-temporal point* of timestamp *t* and coordinates (*x*, *y*). Let *T* be the set of anonymized messages and

$m_t = R(m_s) \in T$ be the anonymized version of message m_s. The function R: $S \rightarrow T$ is bijective. If $m_t = R(m_s)$, then the message identifiers are the same $m_t.u_{id} = m_s.u_{id}$. If $R(m_s) = \varnothing$, message m_s is not anonymized. The *spatio-temporal cloaking box* of anonymized message m_t is denoted as $B_{cl}(m_t)$.

Let $M = \{m_{s_1}, m_{s_2}, ..., m_{s_l}\}$ be a set of messages in S. The formal definition of *location k-anonymity* states that for a message $m_s \in S$ and its anonymized message $m_t \in T$, the following conditions must hold

Definition 4 (*Location k-anonymity* [GeL04, GeL05])

$$\exists T' \subset T, \text{s.t. } m_t \in T', |T'| \geq m_s.k,$$
$$\forall_{\{m_{t_i}, m_{t_j}\} \subset T'}, m_{t_i}.u_{id} \neq m_{t_j}.u_{id} \text{ and}$$
$$\forall_{m_{t_i} \in T'}, B_{cl}(m_{t_i}) = B_{cl}(m_t).$$

This *location k-anonymity* means for each anonymized message $m_t = R(m_s)$ there exist at least $m_s.k - 1$ other anonymized messages ($\exists T' \subset T, \text{s.t. } m_t \in T', |T'| \geq m_s.k$) from different nodes ($\forall_{\{m_{t_i}, m_{t_j}\} \subset T'}, m_{t_i}.u_{id} \neq m_{t_j}.u_{id},$) within the same spatio-temporal cloak box ($\forall_{m_{t_i} \in T'}, B_{cl}(m_{t_i}) = B_{cl}(m_t)$). These conditions form a constraint graph G_m.

The challenge is to find a set of messages $m_t = R(m_s) \in T'$ within a minimal spatio-temporal cloaking box to satisfy the above definition. Another challenge is given the message $m_s \in S$, finding the set M containing m_s and the k-1 group of messages that can

be anonymized with m_s. The Clique-Cloak local *k*-anonymity search algorithm in Figure 22 solves the latter problem.

The first parameter of the LOCAL-k_SEARCH procedure is the agent's desired minimum *k*-anonymity level, the second parameter is the received message m_{s_c}, and the third is the constrained subgraph $G_m^{'}$. This algorithm detects a suitable clique in the

$LOCAL-k\_SEARCH(k, m_{s_c}, G_m^{'})$
(1) $U \leftarrow \{m_s \mid m_s \in nbr(m_{s_c}, G_m^{'})$ and $m_s.\ k \leq k\}$
(2) **if** $|U| < k-1$
(3) **return** $\varnothing$
(4) $l \leftarrow 0$
(5) **while** $\neq |U|$
(6) $l \leftarrow |U|$
(7) **foreach** $m_s \in U$
(8) **if** $(|nbr(m_s, G_m^{'}) \cap U| < k-2)$
(9) $U \leftarrow U \setminus \{m_s\}$
(10) Find any subset $M \subset U$, s.t. $|M| = k-1$ and $M \cup \{m_{s_c}\}$ forms a clique
(11) **return** M

Figure 22: ClickCloak Local-k Search Algorithm [GeL07]

subgraph $G_m^{'}$, which contains m_{s_c} and its neighbors in graph G_m, denoted as $nbr(m_{s_c}, G_m^{'})$. The goal is to find a *k*-sized clique that satisfies the *location k-anonymity* definition. In line 1, before searching, a set *U* of cliques is constructed. In lines 2-3, if no *k*-sized cliques are found, the algorithm exits. In lines 4-9, the set *U* is filtered until no more modifications are required. Each message $m_s \in U$ is verified to have at least *k*-2 neighbors in line 8. If not, m_s is removed in line 9. In lines 10-11, the subset of *k*-1 cliques are returned.

Two metrics measure anonymity effectiveness: *anonymization success rate* and *relative anonymity level* [GeL05]. The success rate is the rate at which anonymized messages meet the anonymization constraint or

$$Anonymization\ Success\ Rate = \frac{|\{m_t \mid m_t = R(m_s), m_t \in T, m_s \in S'\}|}{100^{-1}|S'|}, S' \subset S. \qquad (27)$$

The number of anonymized messages is in the numerator and the number of received messages in the denominator. A higher percentage is preferable. The *relative anonymity* is the amount of anonymous messages in the cloak box normalized by the required message level ($\frac{1}{|T'|}$) or

$$Relative\ Anonymity = \frac{1}{|T'|}\sum_{m_t = R(m_s) \in T'} \frac{|\{m \mid m \in T \wedge B_{cl}(m_t) = B_{cl}(m)\}|}{m_s.k}, T' \subset T. \qquad (28)$$

This measure does not go below 1.

In summary, the first *k-anonymity* preserves data privacy and both zone-based destination *k-anonymity* and *personalized k-anonymity* preserve location privacy in mobile ad-hoc networks.

2.4.10 Multicast Anonymity.

Multicast services are required by various applications such as video teleconferencing, Internet-based education, NASA TV, and software updates. Anonymity degree metrics in unicast communications are not directly applicable in multicast environments [XiL06]. The fundamental difference is the multicast group.

This one-to-many relationship may be represented as a tree structure between senders and receivers. The typical unicast one-to-one relationship is simply a single path in this tree structure. In [XiL06], a k-ary incomplete tree structure with L+1 layers and a Layer 0 root node is assumed. The three types of nodes in a multicast network are anonymous agents (AA), non-anonymous agents (NA) and middle outsiders (MO). Only AA nodes require their identities to be hidden from all agent/non-agent nodes. MO nodes only provide packet forwarding services.

The metric used to analyze sender anonymity degree in this multicast environment is the probability the identity of the AA node is revealed or P_{reveal}. If the AA node identity is broken, P_{reveal} = 1; otherwise, the probability is computed according to a weight. The weight for each node is the probability the adversary believes the node's parent or one of the children is an AA node. Assuming the adversary randomly chooses nodes to compromise, the probability of each node in the tree being broken is

$$P_{broken} = N / \sum_{i=0}^{L} \sum_{j=1}^{k} q_{i,j} \tag{29}$$

where $q_{i,j}$ is a value given to each node in the tree, L is tree depth, k is tree degree, and N is the number of nodes the adversary already broke. If the root node is broken, the adversary already has all the necessary information. Otherwise, the probability P_{attack} that the real root or sender will be identified and subject to attack next time is computed. The overall probability that the root identity is revealed P_{reveal} is

$$P_{reveal} = P_{broken} + (1 - P_{broken})P_{attack}, and \quad (30)$$

$$P_{attack} = \sum_{j=1}^{k} (w_{1,j} / \sum_{i=0}^{L} \sum_{j=1}^{k} w_{i,j}) \quad (31)$$

where $w_{i,j}$ is the weight given to the broken tree head node or the j^{th} node in the i^{th} layer. The weight formula is not shown but is correlated to adversary ability (P_{broken}) and sender multicast tree structure (k, L).

Receiver anonymity degree in this multicast environment is the probability the identity of the AA node as a receiver is revealed P'_{reveal}. Again, the probability P'_{attack} that the real receiver will be identified and subject to attack is

$$P'_{reveal} = (1 - (1 - P_{broken}))^2 + (1 - P_{broken})^2 P'_{attack}, \ and \quad (32)$$

$$P'_{attack} = \sum_{j=1}^{k} (w_{u-1,[t/k]} / \sum_{u=1}^{L} \sum_{t=1}^{k} w_{u,t}) / k \ , \ u > 1 \quad (33)$$

where $w_{u,t}$ is the weight of the AA node and $w_{u\text{-}1,[t/k]}$ is the weight of its parent node. The weight formula is not shown but is correlated to adversary ability (P_{broken}) and receiver multicast tree structure (k, L).

The two probabilistic formulas of sender anonymity degree, P_{reveal}, and receiver anonymity degree, P'_{reveal}, for multicast communications are defined above. These anonymity degree formulas depend on adversary ability (P_{broken}), tree degree (k), and tree depth (L). Overall, anonymity degree improves when P_{broken} decreases, k increases, and L increases [XiL06].

For example, assume agent A multicasts a message to receiver E. The adversary constructs a binary, incomplete tree (L=2, k=2) and computes P^{S}_{reveal} and P^{R}_{reveal} as illustrated in Figure 23(a), (b) and (c), respectively.

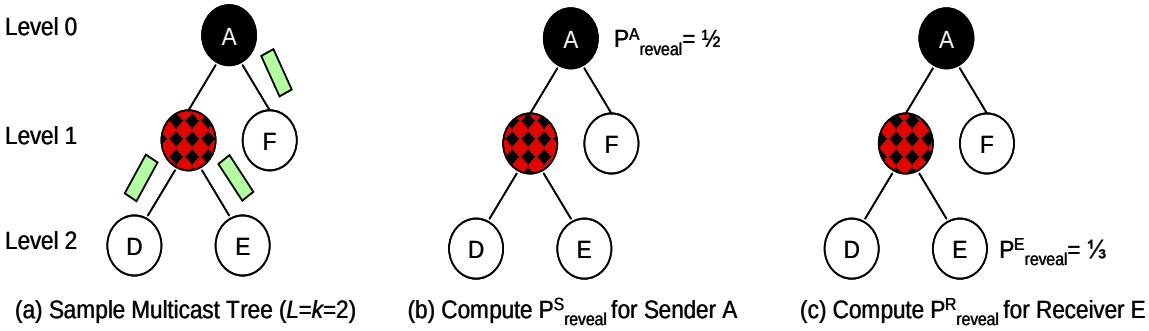


Figure 23: Example of Adversary Multicast Tree and Anonymity Degree Computations (L=k=2).

Assume C = 1 and $q_{i,j}$ = 1 so P_{broken} = 1/5. Also, assume P^{S}_{attack} = 3/8 and P^{R}_{attack} = 13/27. Sender anonymity is P^{A}_{reveal}= 1/5 + (4/5)(3/8) = 2/10 + 3/10 = ½. Receiver anonymity is P^{E}_{reveal}= $(1-(1-1/5))^2$ + $(1 – 1/5)^2$*(13/27) = 1/25 + (16/25)(13/27) = ⅓.

2.5 Formalizing Anonymity

Formal methods provide a rigorous approach to defining and modeling security concepts and aid in the analysis, design and evaluation of secure systems. Using mathematical notation to describe a system, these methods increase reliability and verifiability in software from the requirements phase onwards. Several formal methods for analyzing anonymity have been developed in the literature. These characteristically fall under approaches based on epistemic logic [EiO07, GaH05, HaO03, SyG95, SyS99], process-calculi [AdD03, BhP05, DeP06, HuS04, RyS01, ScS96], functional views [HaO03, HuS04], or automata [KaM06]. Conceptually, these formal approaches use an

adversary-defender modeling (ADM [Mer06]) process to model anonymous protocols as shown in Figure 24. This simply entails a refinement from a general to application specific system model.

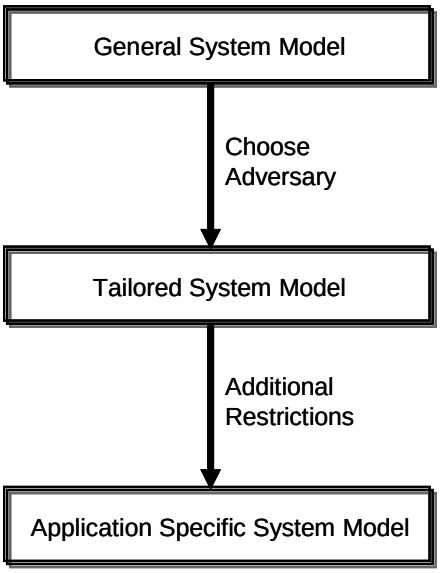


Figure 24: Universal Adversary-Defender Modeling Process [Mer06]

Starting with a general system model defined in the formal method of choice, an adversary is selected. Since anonymous communications take place with a specific adversary in mind, this is an essential first step. As mentioned earlier, the adversary may be weak to strong and have varying anonymity levels which results in a tailored system model. Next, additional environmental and agent restrictions are assumed. Environmental factors may be globally/neighborly available location information, uniform/non-uniform and dense/sparse node densities, noiseless/noisy communication channels, or delay sensitive/insensitive traffic. Agent choice and behavior may be probabilistic/unpredictable, adaptive/non-adaptive, or finite/infinite when sending, receiver or forwarding anonymous messages. These extra limitations produce an

application specific system model for analyzing comparable anonymous communication systems. Then an explicit anonymity property can be verified to be preserved or degraded for the particular application specific model.

For instance, one study formally and quantitatively analyzes sender anonymity in a message-based anonymous communications system under various routing strategies [GuF04]. The general system model is a collaborating set of *n* agents $A = \{a_i : 0 \le i < n\}$ to achieve anonymity as shown in Figure 25.

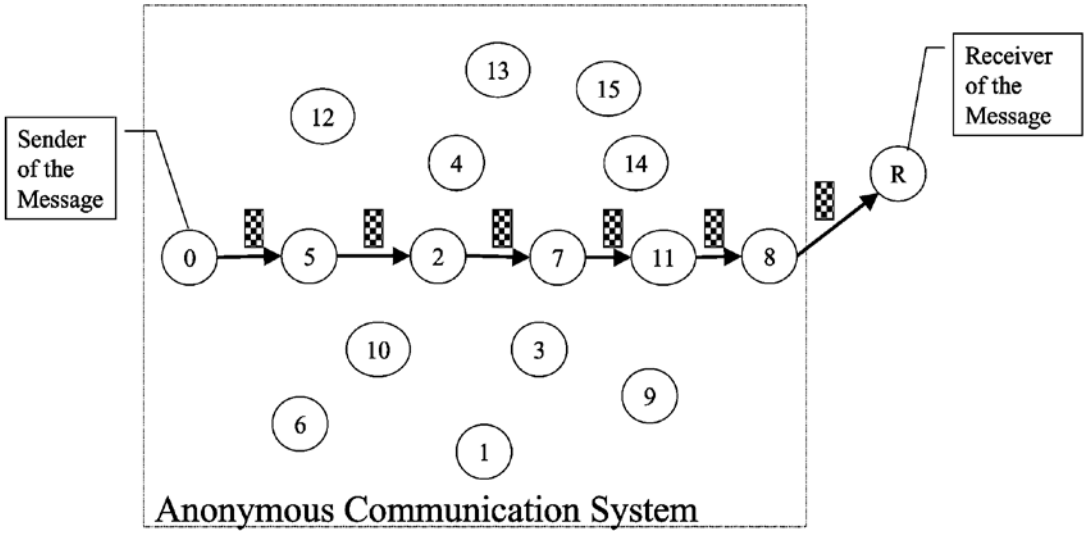


Figure 25: General System Model [GuF04]

The sender sends a message to the receiver through the anonymous communication system consisting of sixteen agents and to preserve its identity. A passive adversary threat model with a fixed number of compromised nodes is chosen. Figure 26 displays the tailored system model with this threat model in mind.

The adversary has already compromised six agents 1, 5, 7, 8, 10, and 15 as well as the receiver *R* and collects information from these agents. The system anonymity metric is the adversary's probability of identifying the message sender.

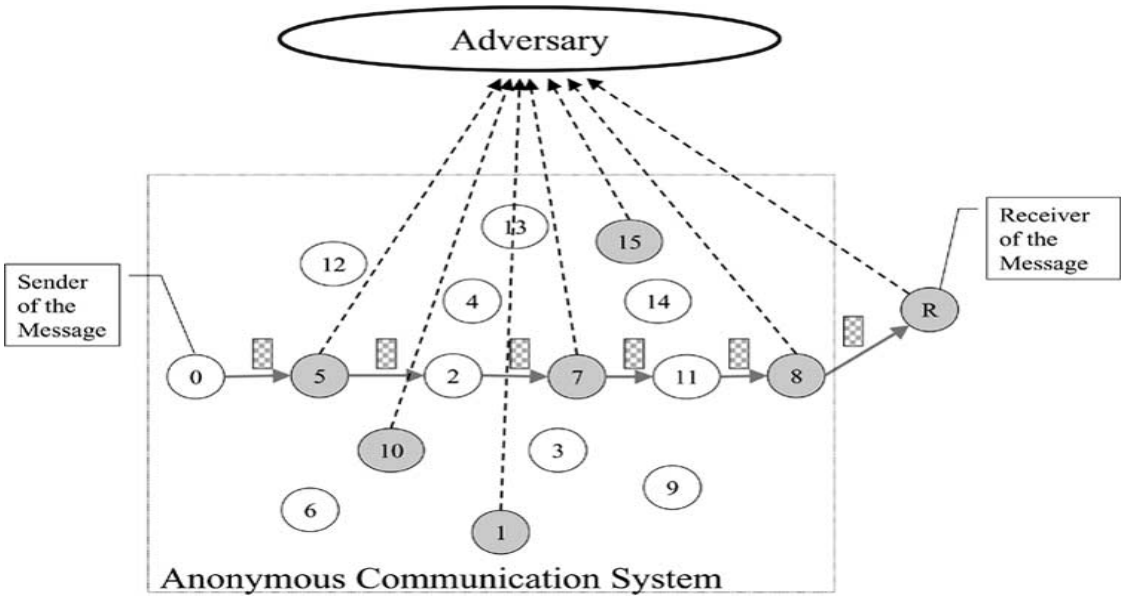


Figure 26: Tailored System Model [GuF04]

The adversary's behavior is framed algorithmically in four steps as indicated in Figure 27. Every message the receiver receives affords the adversary an opportunity to collect key information (Steps 1 and 2), eliminate possible sender nodes (Step 3), and

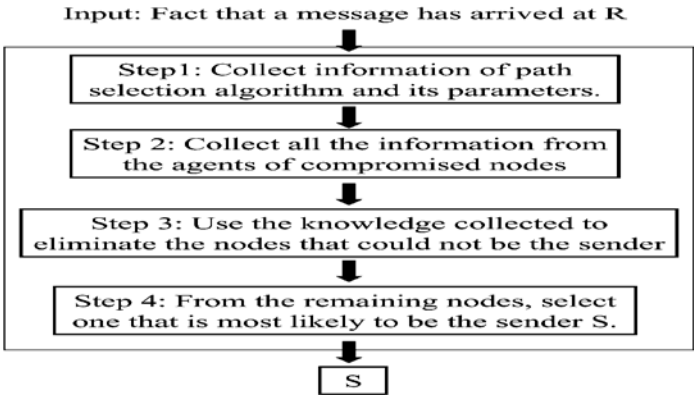


Figure 27: Algorithmic Adversary Framework [GuF04]

update probabilities of the remaining nodes (Step 4). Additional restrictions include agents using a cascade or free-route topology, a probabilistic geometric or uniform variable path length, and cyclic or acyclic path type which defines the agent's behavior. The sender has no knowledge of compromised agents while the adversary has full knowledge of path selection algorithm, and the adversary collects all information from compromised agents to reveal sender identity and correlate received messages. Depending on the agent selections, several application specific system models of a message-based system may be defined either graphically or algorithmically. These are the internal mechanisms of the agents and the adversary and are not shown. Hence, the universal adversary-defender model applies to this as well as other studies. The rest of this section reviews the use of approaches in security, with a focus on applications for the design or description of anonymity systems.

2.5.1 Conceptual Framework.

Before meticulously exploring anonymity mathematical frameworks, it is useful to first cover a more holistic and intuitive anonymity framework or taxonomy. Such a conceptual approach complements the formal framework by accentuating the significance and subtlety of anonymity, acting as a state-of-the-art model for theoretical theorem-proving and model checking and empirical statistical investigations into anonymity, and contributing to future anonymous protocol design and development across one or more application domains. Unfortunately, there is a dearth of literature for such intuitive taxonomies. Three conceptual frameworks for anonymity are known to have been

developed: one for group support systems (GSS) [VaD92], another for collaborative peer groups [SuP03], and another for connection anonymity [TiO05].

2.5.1.1 Group Support System Framework.

Anonymity is important in group support systems because it offers a low-threat communicative environment to reduce evaluation apprehension, encourage open and honest contributions without the fear of direct reprisals, and depersonalize contributions to allow valuing based on merit not authorship for both individuals and groups [VaD92]. The group support conceptual framework is displayed in Figure 28. The four main parts include the anonymity factors, the anonymity types, individual anonymity and group process/outcome. The arrows represent the influence each left part has on the subsequent right part and indicate a natural flow from the anonymity factors to the eventual group outcome.

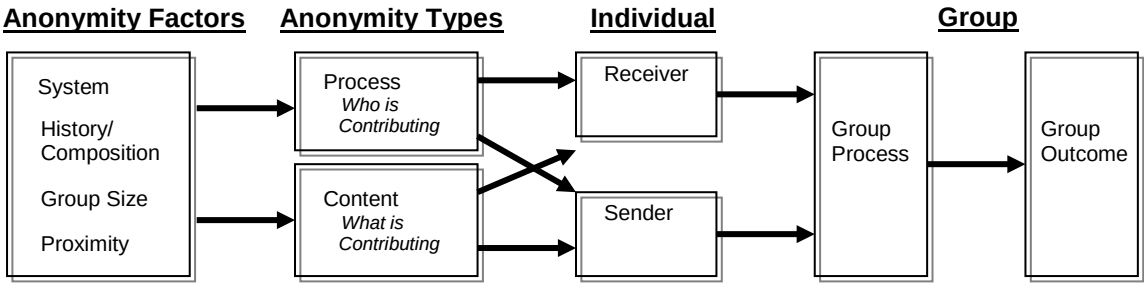


Figure 28: Conceptual Framework for the Study of GSS Anonymity [VaD92]

The anonymity factors are system characteristics, group history and composition, group size, and group agent proximity. Each factor results in either process and/or content anonymity types. Process anonymity is the ability of a group agent to know who the contributing agents are. Content anonymity is the ability of a group agent to know

what information was contributed by which group agent. Both determine the level of individual receiver and sender anonymity preserved. The subsequent perceived or known degree of anonymity, not simply the presence or lack thereof, has direct implications on the group process that either negatively or positively affects group outcome. For example, a system which only allows instantaneous concurrent contributions of a group size of four individuals in close proximity (residing in the same room) would have a lower degree of anonymity than a system which allows delayed contributions of a group size of ten individuals in disperse proximity (sitting at their own computers in different rooms).

2.5.1.2 Collaborative Peer Group Framework.

A lower level collaborative peer group conceptual framework, the Janus architecture [SuP03], was also proposed. This P2P network is a middleware architecture and software toolkit to facilitate the development and deployment of applications where self-organizing peers aggregate in a controlled manner and new types of communication primitives achieve collective goals. Janus peer groups do not possess identities. Each peer holds a template that defines group specific capabilities and other information. A new peer, such as Node 1 in Figure 29, scans to discover peer groups with matching templates. If no match is found, Node 1 becomes a group of one like Peer 7. If a match is found with, say, Peer 3 and/or Peer 6, a communication channel is open and Node 1 joins the group. As Figure 29 shows, these actions may merge previously disjoint peer groups; or upon leaving, split groups. Each peer maintains a table of its neighbors, called a *local view*, and a *next neighbor* table as revealed in Figure 30.

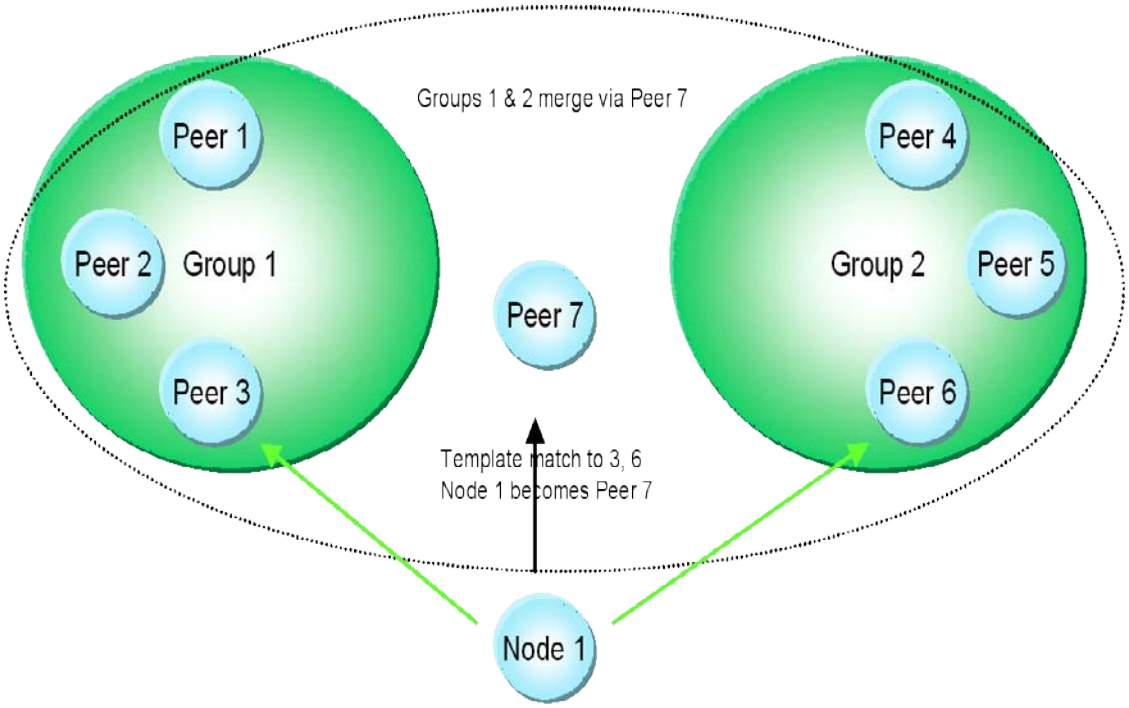


Figure 29: Formation of Janus Groups [SuP03]

Node(i)	*Local view (v)*	*Next neighbor (v')*
1	2,3,4	[1,5],[1,6,7],[1]
2	1,5	[2,3,4],[2]
3	1,6,7	[2,3,4],[3],[3]
4	1	[2,3,4]
5	2	[1,5]
6	3	[1,6,7]
7	3	[1,6,7]

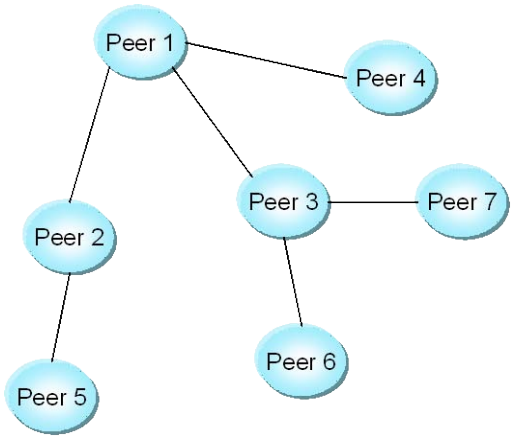


Figure 30: Peer Neighbor Information Tables [SuP03]

For instance, the local view of Peer 1 includes the peer neighbors 2, 3, and 4 and the next hop neighbors [1,5], [1,6,7], and [1], respectively. The multicast primitive transmits messages to a group of at least *k* identity-less peers and the message is either delivered or an error returned. The proximal cast primitive allows a subset of groups to disseminate messages to neighbors collectively. The collect cast primitive enables subset of groups

to gather messages from neighbors collectively. Stable peer groups are easy to handle, but dynamic peer groups may cause more errors if peers suddenly enter or leave groups. Thus, this model works well for hundreds to thousands of nodes of small degree only or low density networks.

2.5.1.3 Connection Anonymity Framework.

Anonymity is important for protecting the communications channel between sender and receiver. With the evolution of anonymous technologies from simple proxies to complex systems, a more structured meta-level approach to designing and comparing current anonymity strategies and techniques is desirable. A connection anonymity conceptual framework is depicted in Figure 31.

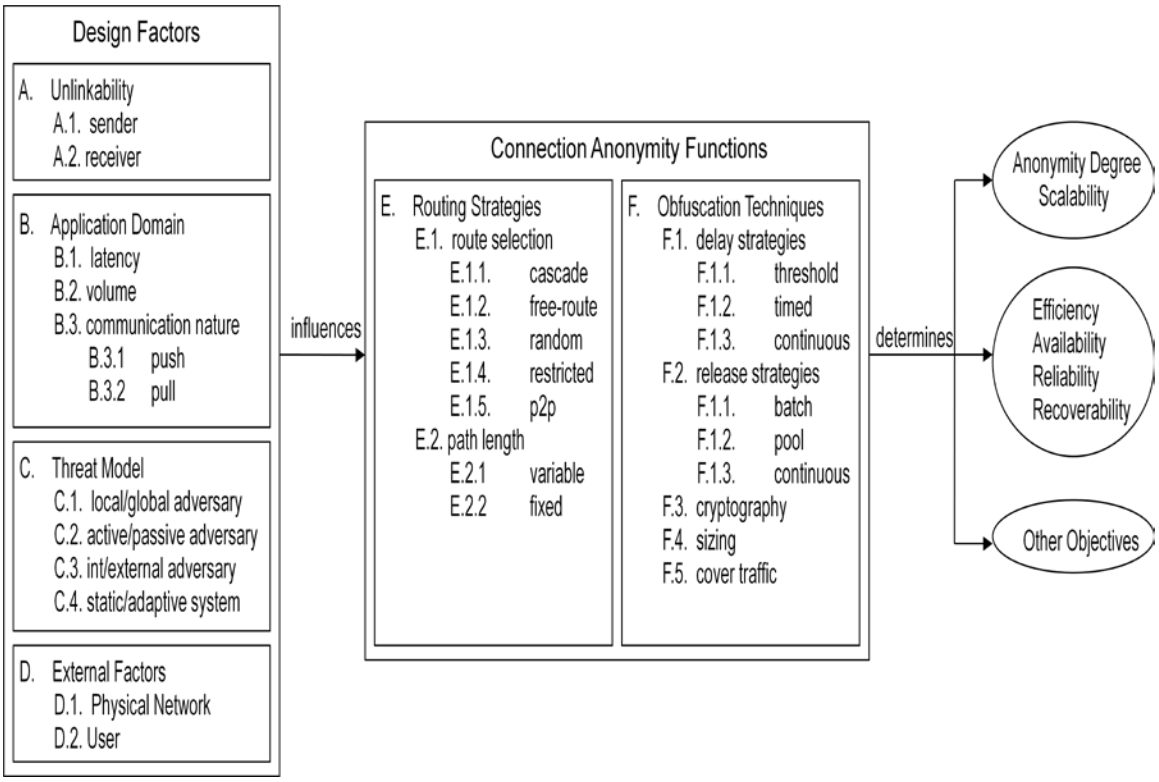


Figure 31: A Conceptual Framework for Connection Anonymity [TiO05]

The three main components are the design factors, the connection anonymity functions, and the objectives. To easily identify individual framework items, each specific item is numbered. The design factors are heuristic measures useful in the design and evaluation of connection anonymity services. The four design factors consist of unlinkability, the application domain, the threat model and the external factors. Listed under each are its sub-components. Unlinkability (A.) means two or more items of interest such as agents, messages, events or actions are no more or no less related afterwards than they were before given a priori knowledge. Unlinkability consists of sender (A.1) and receiver (A.2) anonymity. The application domains (B.) include store-and-forward applications such as e-mail, interactive applications such Internet Relay Chat, and real-time applications such as Voice-over-IP (VoIP) or video conferencing. Each has distinct latency (B.1) and volume (B.2) requirements. Each application technology may be classified as push (B.3.1) or pull (B.3.2). The threat model (C.) highlights adversary capabilities of an individual, large corporation or national entity with legal powers. The adversary may be local-global (C.1), active-passive (C.2) and/or internal-external (C.3). Adversaries are usually adaptive, but the system itself may be either static or adaptive (C.4) when recovering from an attack. Since attacks are design or implementation specific and directly affect anonymity degree, attack techniques are excluded from this abstract framework. The two external factors (D.) physical network (D.1) and the user (D.2) indirectly affect anonymity degree and technology effectiveness. Each design factor influences connection anonymity functions.

The fundamental functions of connection anonymity are routing strategies (E.) and obfuscation techniques (F.). For routing strategies, the route selections (E.1) are either

cascades (E.1.1) which chains multiple mixes together, free-route (E.1.2) which permits the sender to choose the route, random (E.1.3) which enables plausible deniability, restricted (E.1.4) which combines cascades and free-route or structured peer-to-peer (E.1.5) which boost scalability and resiliency. Path lengths (E.2) are fixed (E.2.1) in cascade and free-routes and variable (E.2.2.) in random and P2P routing. For obfuscation techniques, the delay strategies (F.1) are threshold (F.1.1) mixes which collect a fixed number of messages, timed (F.1.2) mixes which flush messages periodically, and continuous (F.1.3) mixes which do not batch messages. Release strategies (F.2) include batch (F.2.1) where all messages are simultaneously released, pool (F.2.2) mixes which flush a random number of messages, and continuous (F.2.3) mixes that cyclically delay messages. The remaining obfuscation techniques include cryptographic (F.3) and sizing (F.4) transformations to circumvent certain attacks and resource-intensive cover traffic (F.5) to enhance anonymity.

The anonymity functions determine the overall objectives of the anonymity system. The objectives are anonymity degree which quantifies the level of anonymity, scalability which defines allowable system sizes, efficiency which emphasizes acceptable anonymity levels, availability, reliability and recoverability.

2.5.1.4 Summary.

These three meta-level frameworks for group support systems, collaborative groups and connection anonymity delineate the factors and issues in their respective areas. They are useful abstract formalisms for classifying and clarifying a variety of different

approaches to anonymous technologies and may eventually lead to a more comprehensive and discerning taxonomy and formal framework for anonymity.

2.5.2 Probabilistic and Nondeterministic Systems.

Anonymity may be formally modeled in probabilistic or nondeterministic systems. Most research focuses on individual agent anonymity, not group anonymity. The anonymous communications protocols such as DC-net, Crowds and Onion Routing use random mechanisms that may be described probabilistically. Agent or adversary choice and behaviors may be probabilistic or nondeterministic. The formal frameworks typically employed to model anonymity are process calculi, epistemic logic, and functional views and are described later in this chapter. Hence, a formal method's approach to anonymity may be purely nondeterministic, purely probabilistic or both probabilistic and nondeterministic.

A purely nondeterministic (a.k.a. possibilistic) approach to anonymity has been studied [RyS01, ScS96]. For nondeterministic anonymity, the actions of a system *S* are anonymous (*A*), known (*B*), or hidden (*C*) to the adversary. The anonymous set of abstract actions $A = \{a.i \mid i \in I\}$ indicates that action *a* may be performed by identifiable agent *i* in the anonymity set of identities *I*. For instance, the process calculi may model anonymity as a non-unique observable trace in a purely nondeterministic manner. A limitation of this approach is the inability to differentiate between fair and unfair coins. However, fairness is essential to ensure anonymity and the ability to only express possible/impossible nature of a trace and not the probability of a trace is insufficient for some application domains.

A purely probabilistic approach factors out all nondeterministic influences and focuses either on agent probability or observable effects on agent probability [Pal05]. If agent probability is the focus, then anonymity may be defined as strong probabilistic anonymity, beyond suspicion, probable innocence, possible innocence, or probabilistic α-anonymity. If observable effects on agent probabilities are the focus, then conditional probabilistic anonymity is used as the definition of anonymity where probabilities are dynamically updated. In one purely probabilistic approach [HaO03], the agents are probabilistic with possibly unknown probabilities. Anonymity is proven to hold for any agent probability distribution. The formal method is epistemic logic but an equivalent function view approach is suggested.

A combined probabilistic and nondeterministic approach [BhP05, Pal05] is the most general. The agents are nondeterministic (unpredictable) and the anonymity internal system mechanism (protocol) is probabilistic (coin toss). The protocol is proven to not leak probability information to the adversary. The formal method is typically process algebra. For instance, the notion of anonymity may be observables for processes in probabilistic π–calculus with probabilistic automata semantics [BhP05]. Perfect anonymity means no information is deduced from observables about the possible agent. The probabilistic automata model of computation is chosen since nondeterministic agent behavior does not equate to unknown agent probabilities. However, repeated experiments on random mechanisms allow the adversary to infer probability between agents and observables [BhP05].

2.5.3 Group Principals.

In this section, a group principal (agent) approach [SyS99] to formally reason about anonymity systems based on epistemic logic is described. This approach focuses on *group* anonymity instead of individual agent anonymity. This shift from individual agents to groups of agents is appropriate for modeling anonymity systems, which intrinsically rely on the interaction of groups of agents to preserve anonymity.

The logic defines four *group principals* [SyS99] to express group-based knowledge. These principals (agents) are the collective group (*G), and-group (&G), or-group ($\oplus G$) and threshold-group (n - G). The collective group is knowledge gained from combining individual agent knowledge in group G. The and-group is knowledge known by every agent in the group G. The or-group is knowledge known by at least one agent of group G. The threshold-group is collective knowledge of any subgroup of G with cardinality of n. Alternatively, an n-threshold group is an or-group of collective groups, each with cardinality of at least n.

Each agent in the set $P = \{P_1, P_2, ..., P_n\}$ of principals uses a local clock to track the observed time-order of events. In the model agents have a *history* of performed actions, *log* of time-stamped actions, a set of predefined or deduced environmental *facts*, and a set of *recent* actions performed by the agent. Each agent has a unique local state s_i represented by <*state_id*, *history*, *log*, *facts*, *recent*> where *state_id* is the sequence of previous states.

This framework models send and receive actions that are performed within a run of the system and are entered into or purged from the log of any agent that observes the action. For the formal language, if P_i and P_j are agents, and M is a message, then

send(M, P_i, P_j) and receive(M, P_j, P_i) are the primary actions and P_i **said** M, P_i **received** M, P_i **said to** P_j M, and P_j **received from** P_i M are the corresponding logical formulas. If φ is any formula, $\Box_{Pi}\,\varphi$ means agent P_i knows φ and $\Diamond_{Pi}\,\varphi$ means agent P_i possibly knows φ. A set of axioms based on group principals allows agents to gain knowledge from the system as each action is performed. The use of deduction rules expresses the knowledge that a particular agent may gain, and thus the potential of an adversary compromising the anonymity of an agent in a group in the system.

Let A be the adversary, P be the agent or group to remain anonymous and $\varphi(P)$ be the fact to hide from the adversary. Seven anonymity definitions, logical expressions and meanings are shown in Table 7. These anonymity definitions are purely nondeterministic (possibilistic). The *unknown* definition is impossible since the logic and language ensures that every agent is always a suspect. The *(≥ N)-anonymizable* definition says if agent P is suspect, then at least N-1 other agents are also suspect. If the adversary is

Table 7: Group Principals Anonymity Definitions [SyS99]

Definition	Formula	Meaning
Unknown	$\neg(\Diamond_A\,\varphi(P))$	Adversary does not know that P possibly performed action.
(≥N)-anonymizable	$\Diamond_A\,\varphi(P) \Rightarrow (\Diamond_A\,\varphi(P_1) \wedge \ldots \wedge \Diamond_A\,\varphi(P_{n-1}))$	If P is a suspect, then at least N-1 other agents are suspect.
Possible Anonymity	$\Diamond_A\,\varphi(P) \wedge \Diamond_A \neg\,\varphi(P)$	Adversary has no knowledge about P's actions.
(≤ N)-suspected	$\Box_A(\varphi(P) \vee \varphi(P_1) \vee \ldots \vee \varphi(P_{n-1}))$	Adversary suspects N or fewer agents including P.
(≥ N)-anonymous	$\Diamond_A\,\varphi(P) \wedge \Diamond_A\,\varphi(P_1) \wedge \ldots \wedge \Diamond_A\,\varphi(P_{n-1})$	Adversary suspects N or more agents including P.
(≤M)-suspected ⇒ (≥N)-anonymous	$\Box_A(\varphi(P) \vee \varphi(P_1) \vee \ldots \vee \varphi(P_{m-1})) \Rightarrow (\Diamond_A\,\varphi(P) \wedge \Diamond_A\,\varphi(P_1) \wedge \ldots \wedge \Diamond_A\,\varphi(P_{n-1}))$	Adversary suspects N to M agents, $N \le M$
Exposed	$\Box_A\,\varphi(P)$	Adversary knows P performed action.

unable to rule out the possibility or impossibility of agent *P* performing the action, then no knowledge about agent *P* exists and *P* has *possible anonymity*. With *N* or fewer suspects, the *(≤ N)-suspected* definition is equivalent to *up-to* |*I*| *anonymity* [HaO03]. With *N* or more suspects, the *(≥ N)-anonymous* definition is equivalent to *k-anonymity* where *N=k*. The definition *(≤ M)-suspected* ⇒ *(≥ N)-anonymous* bounds the adversary to suspecting from *M* to *N* agents. Finally, when the adversary knows who performed the action, agent *P* is *exposed*. Another framework based on knowledge-based logic and deductive reasoning is discussed next.

2.5.4 Multi-agent Systems.

In this section, a *multi-agent system* [HaO02, HaO03, Wei99] framework is reviewed. This framework mathematically represents an anonymous system based on epistemic logic. This approach is compatible with many other standard approaches for representing and reasoning about systems and is rich enough to accommodate a variety of system representations [HaO02, HaO03]. However, first the concept of the abstract agent architecture is explored as shown in Figure 32.

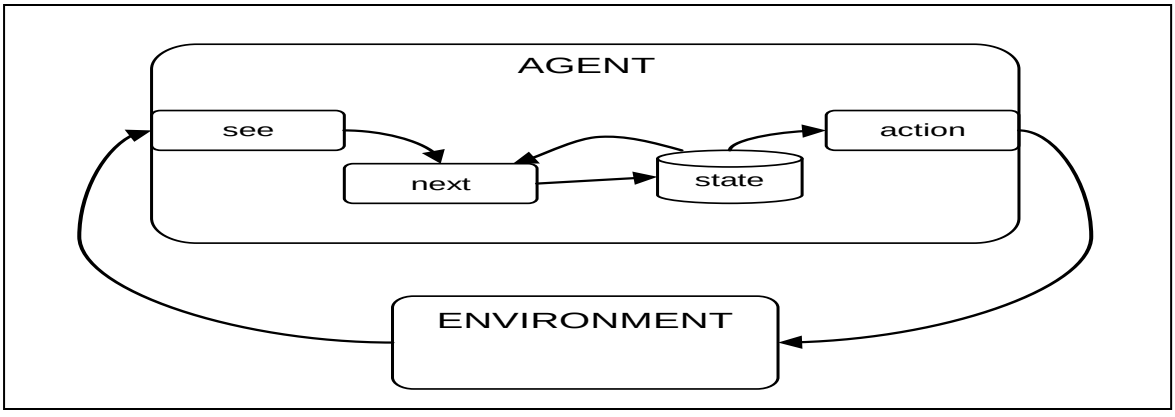


Figure 32: Abstract Agent Architecture [Wei99].

An abstract view of agents assumes that the agent's environment may be represented as a set $S = \{s_1, s_2, \ldots\}$ of *environmental states*. The environment is in one of these states s_i at any given instant. The agent has a set $I = \{i_1, i_2, \ldots\}$ of *internal states* as well as a set $P = \{p_1, p_2, \ldots\}$ of *precepts* which are the agent's interpretation of each environmental input. The agent may perform the set $A = \{a_1, a_2, \ldots\}$ of *actions*.

The agent has three decision functions: *see*, *next*, and *action*. The perception function *see* captures the agent's ability to observe its environment; the function *next* updates the internal state based on its own perceptions; and the action-selection function *action* selects the appropriate action and performs the action in the environment. Each function maps the appropriate input(s) to a corresponding output. The *see* function maps environmental states to precepts or *see*: $S \rightarrow P$. The *next* function maps an internal state and precept to an internal state or *next*: $I \times P \rightarrow I$. The *action* function maps internal states to actions or *action*: $I \rightarrow A$.

This abstract agent architecture reveals the properties of state-based agents and models an agent's abstract functions but fails to explain what the agent's state might be or examine how the *see*, *next* and *action* functions are decided. A concrete epistemic based agent architecture is proposed [HaO03] where anonymity is expressed and agent decisions are realized through logical deduction.

A *multi-agent system* consists of *n* agents, each of which is in some *local state* at a given point in time. An agent's local state encapsulates all the information to which the agent has access. The local state of an agent might include initial information regarding keys, the messages sent and received, and a timestamp. The framework makes no assumptions about the precise nature of the local state; hence, high-level anonymity

properties do not depend on the local agent states. This is a major disadvantage if an adversary with limited view of the system, i.e., a local adversary, needs to be modeled. The entire system may be in some *global state*, a tuple consisting of the environmental state and the local state of each agent. Thus, a global state has the form $(s_e, i_1, \ldots, i_n)$ where s_e is the environment state and i_j is agent i's state, for $j = 1 \ldots n$.

This approach is based on a *run*. A *run* is a function that maps time to global states. Intuitively, a *run* is a complete description of what happens over time in one possible execution of the system. The *run* is analogous to the concept of *traces* used in the CSP process calculus. A *point* is a pair (r,m) consisting of a run r and a time m where both r, $m \in$ Integers. Logical deductions concerning the properties of agents are made based on these *points*. At a point (r,m), the system is in global state $r(m)$. If $r(m) = (s_e, i_1, \ldots, i_n)$, then $r_i(m)$ is user i's local state at the point (r,m).

An important advantage of the framework is that it is easy to formally define what an agent knows at a point in a system. Formally, a *system* consists of a set of runs or executions. Let $P(R)$ denote the points in system R. Given a system R, $K_i(r,m)$ is the set of points in $P(R)$ that i thinks are possible at (r,m), i.e.,

$$K_i(r,m) = \{(r',m') \in P(R): r'_i(m') = r_i(m)\}. \tag{34}$$

Agent i knows a nontrivial fact φ at a point (r,m) if φ is true at all points in $K_i(r,m)$. To be more precise, truth values must be assigned to basic formulas in a system. Assume a set Φ of primitive propositions describes basic facts about the system. In the context of anonymous protocols, a fact, φ, may be "*Alice* sent the message *M* to *Bob*". An

interpreted system Γ consists of a pair (R,π) where R is a system and π is an *interpretation*, which assigns to each primitive proposition in Φ a truth value at each point (r,m). Thus, for every primitive proposition $p \in \Phi$ and point (r,m) in R, $(\pi(r,m))(p) \in \{\textbf{true}, \textbf{false}\}$.

Now, a formula or fact φ (or ψ) is **true** at a point (r,m) in an interpreted system Γ, written $(\Gamma,r,m) \models \varphi$ (or ψ) where $\models$ is logical entailment [Sik94], by induction using the following formulas

$$(\Gamma,r,m) \models p \textbf{ iff } (\pi(r,m))(p) = \textbf{true} \tag{35}$$

$$(\Gamma,r,m) \models \neg\varphi \textbf{ iff } (\Gamma,r,m) \not\models \varphi \tag{36}$$

$$(\Gamma,r,m) \models \varphi \wedge \psi \textbf{ iff } (\Gamma,r,m) \models \varphi \textbf{ and } (\Gamma,r,m) \models \psi \tag{37}$$

$$(\Gamma,r,m) \models K_i\varphi \textbf{ iff } (\Gamma,r',m') \models \varphi \textbf{ for all } (r',m') \in K_i(r,m) \tag{38}$$

The formula $K_i\varphi$ in (41) means "agent *i* knows fact φ". Conversely, the formula $\neg K_i\varphi$ means "agent *i* does not know fact φ". Formal logic is reviewed next.

2.6 Logics

Formal logics are used as a mathematical model to internally specify a language of reasoning or action and externally design metalanguages to specify, design, and verify certain behavioral properties in a dynamic environment. The three aspects of any logic are *well-formed formulas*, *proof-theory*, and *model-theory* [Wei99]. *Well-formed formulas* (*wffs*) are assertions made in the formal language of the underlying logic. *Proof-theory* is the axioms and inference rules and state entailment [Sik94] relationships among

wffs. *Model-theory* interprets the formal meaning of the *wffs*. The syntax is the language and proof-theory and semantics is the model-theory [Wei99]. Formal methods make extensive use of propositional, modal, deonetic, dynamic, and temporal logics. Propositional logic represents factual information, modal logic represents other meanings of formulas, deonetic logic specifies what ought to be or one is obligated to do, dynamic logic is modal logic of action, and temporal logic is the logic of time [Wei99].

Propositions are proved using inference rules from facts known to be true and basic axioms are assumed to be true. The underlying rules differ between the various formal logics and express notions of belief, knowledge, uncertainty, or even ignorance, within specific domains.

The application of formal logics to the analysis of anonymous protocols is an important way to verify anonymous systems and their anonymity properties [AdD03, GaH05, HaO03, HuS04, SyG95, SyS99]. Logics can detect various protocol problems and are reasonably easy to use. However, logics are a high level abstraction for a system, and do not prevent lower-level protocol implementation flaws to pass undetected [Ker07]. The following is a review of two more prevalent modal logics in security proofs: epistemic and temporal logics.

2.6.1 Modal Logics.

Modal logics consider questions of necessity and possibility. This family of logics is concerned with qualifiers that concern the state, or *modality*, of propositions based on sets of defining axioms. The basic syntactic elements, or “modalities”, are the two

statements that represent possibility ◊ (diamond) and necessity □ (box) operators of proposition p:

◊p: it is possible that p
□p: it is necessary that p

However, each may be expressed in terms of the other using negation:

◊$p \equiv \neg\Box\neg p$, "it is possible that p" ≡ "it is not necessary that not p"
□$p \equiv \neg\Diamond\neg p$, "it is necessary that p" ≡ "it is not possible that not p"

Many forms of modal logic rely on different sets of axioms. The most common axiom set is modal logic *S5* [Lew18]:

1. $\Box(p \rightarrow q) \rightarrow (\Box p \rightarrow \Box q)$
2. $\Box p \rightarrow p$
3. $\Diamond p \rightarrow \Diamond\Box p$

The first axiom expresses the distribution property of the necessitation operator □ over the implication operator → statement with two propositions p, q. Specifically, if it is necessary that p implies q then if it is necessary that p then it is also necessary that q. The second axiom defines a reflexive relation property (called *T* for truth) that if p is necessary then p is true. The third axiom describes a Euclidean relation property (called *5*) that if it is possible that p, then it is necessary that it is possible that p. These *S5* axioms allow a wide range of expressive power, and provide a basis for more advanced forms of modal logic based on *equivalence relations* [Wik07a]. Numerous other sets of axioms also exist.

Interestingly, the *possible worlds* concept is sometimes erroneously compared with the *many-worlds* [Ano02, EiR85] interpretation of quantum mechanics. The *many-*

worlds concept provides an interpretation of nondeterministic processes (such as measurement) without positing the so-called collapse of the wave function [EiR85] which introduces a quantum superposition of a possibly infinite number of identical "parallel universes", all of which actually exist, while the *possible worlds* concept provides an interpretation (in the sense of a formal semantics) for modal claims. These concepts differ in two main aspects. First, the states of quantum-theoretical *many-worlds* are mechanically entangled [EiR85] while entanglement for *possible words* is meaningless. Second, quantum-theoretical *many-worlds* are all physically possible while possible worlds are logically but not necessarily physically possible.

Anonymous systems and properties may be expressed using the modal logic syntax and semantics mentioned above. Modal concepts may prove useful in constructing a meaningful definition of anonymity for more advanced models. The anonymity-relevant epistemic and temporal logics are reviewed next.

2.6.2 Epistemic Logic.

Epistemic logics are concerned with propositions of knowledge, uncertainty, and ignorance. Seminal work on epistemic logic [BrA06, EiO07, GaH05, HaO03, SyG95, SyS99] abounds. Knowledge refers to an agent's justified beliefs based of observed facts. In contrast, *doxastic* logics [GrT96] are concerned with agent beliefs only and are based on lower levels of justification. Logics of knowledge add operators to express the knowledge held by a particular agent. $KT45^n$ [HuR04] is an epistemic logic.

2.6.3 KT45^n Logic.

Modal logic systems are fragments of classical logics, which strike a balance between expressive power (of first order predicate logic or other formalisms) and computational simplicity (of prepositional logic) [BlV06]. The normal modal logic system KT45^n has many modes of knowledge including K_i for each agent $i \in A$ where $A = \{1, 2, ..., n\}$ and E_G for everyone, C_G for common, and D_G for distributed knowledge of a group of agents $G \subseteq A$. In KT45^n, the *K* emphasizes knowledge (or lack thereof) of *n* logically omniscient agents. The *T* for truth, *4* for positive introspection and *5* for negative introspection imply reflexive, transitive, and Euclidean (i.e., equivalence relation) semantic properties, respectively [HuR04]. Intuitively, KT45^n means *n* agents know things (*K*), only know true things (*T*), know what they know (*4*), and know what they do not know (*5*). The syntax, inference rules, and semantics are briefly described next.

2.6.3.1 KT45^n Syntax.

A KT45^n formula ϕ is defined by the Backus normal form (BNF) grammar [HuR04]

$$\phi ::= \bot \mid T \mid p \mid \neg\phi \mid \phi \wedge \phi \mid \phi \vee \phi \mid \phi \rightarrow \phi \mid \phi \leftrightarrow \phi \mid K_i\phi \mid E_G\phi \mid C_G\phi \mid D_G\phi \quad (39)$$

where *p* is any atomic formula and $i \in A = \{1, 2, ..., n\}$ and $G \subseteq A$. The grammar in (39) specifies exactly the formulas ϕ of KT45^n modal logic, given a set of atomic formulas *p*. The formula ϕ syntax consists of false ($\bot$), true ($\top$), *p*, five propositional operators ($\neg, \wedge, \vee, \rightarrow, \leftrightarrow$) and four knowledge modalities (K_i, E_G, C_G, D_G). $K_i\phi$ means "agent *i*

knows ϕ". $E_G\phi$ means "everyone in group G knows ϕ" or $\mathrm{E}_G\phi \equiv \bigwedge_{i \in G \subseteq A} \mathrm{K}_i\phi$.; however, not everyone may know that everyone knows. Thus, the state of everyone knowledge may increase until it is common knowledge. $C_G\phi$ means "ϕ is common knowledge among G" or $C_G\phi \equiv E_G\phi \wedge E_G E_G\phi \wedge E_G E_G E_G\phi \wedge \ldots$ Hence, C_G denotes an infinite conjunction of increasing knowledge [HuR04]. $D_G\phi$ means "knowledge of ϕ is distributed among G" although no one in G may know ϕ. The various KT45^n rules are covered next.

2.6.3.2 KT45^n Rules.

The KT45^n propositional inference rules are enumerated in Table 8. These inference rules are used to prove the validity of anonymity formulas. The KT45^n introduction and elimination inference rules for the varying degrees of knowledge are enumerated in Table 9. The closed consequence rules are the "Modus Ponens" equivalents in KT45^n. Substitution rules allow knowledge to traverse the various levels from an individual agent to common knowledge. The introspection and truth knowledge rules for K_j, C_G and D_G are the formal representations of the "4", "5" and "T" properties in KT45^n. The "4" rules are *K*4, *C*4 and *D*4. The "5" rules are *K*5, *C*5, and *D*5. The "T" rules are *KT*, *CT* and *DT*. The K_i dashed boxes mean the formulas are known to agent *i*. The E_G boxes mean the formulas are known to everyone in group *G*. The C_G boxes mean the formulas are common knowledge to those in group *G*. The D_G boxes mean the formulas are distributed, albeit not necessarily known to those in group *G*.

Table 8: KT45^n Propositional Rules [Hal05, HuR04]

Op	Name	Introduction	Elimination
$\bot$	False	n/a	$\dfrac{\bot}{\phi}\,\bot\,\mathrm{e}$
$\neg$	Negation	$\dfrac{\begin{array}{|c}\phi\\\vdots\\\bot\end{array}}{\neg\phi}\,\neg\,\mathrm{i}$ $\qquad$ $\dfrac{\begin{array}{|c}\neg\phi\\\vdots\\\bot\end{array}}{\phi}\,\mathrm{PBC}$	$\dfrac{\phi\ \ \neg\phi}{\bot}\,\neg\,\mathrm{e}$
$\neg\,\neg$	Double Negation	$\dfrac{\phi}{\neg\,\neg\phi}\,\neg\,\neg\,\mathrm{i}$	$\dfrac{\neg\,\neg\phi}{\phi}\,\neg\,\neg\,\mathrm{e}$
$\wedge$	Conjunction	$\dfrac{\phi\ \ \psi}{\phi\wedge\psi}\,\wedge\,\mathrm{i}$	$\dfrac{\phi\wedge\psi}{\phi}\,\wedge\,\mathrm{e}_1$ $\qquad$ $\dfrac{\phi\wedge\psi}{\psi}\,\wedge\,\mathrm{e}_2$
$\vee$	Disjunction	$\dfrac{\phi}{\phi\vee\psi}\,\vee\,\mathrm{i}_1$ $\qquad$ $\dfrac{\psi}{\phi\vee\psi}\,\vee\,\mathrm{i}_2$	$\dfrac{\phi\vee\psi\quad\begin{array}{|c}\phi\\\vdots\\\chi\end{array}\quad\begin{array}{|c}\psi\\\vdots\\\chi\end{array}}{\chi}\,\vee\,\mathrm{e}$
$\rightarrow$	Material Implication	$\dfrac{\begin{array}{|c}\phi\\\vdots\\\psi\end{array}}{\phi\rightarrow\psi}\,\rightarrow\,\mathrm{i}$	$\dfrac{\phi\ \ \phi\rightarrow\psi}{\psi}\,\rightarrow\,\mathrm{e}$ $\qquad$ $\dfrac{\neg\psi\ \ \phi\rightarrow\psi}{\neg\phi}\,\mathrm{MT}$
$\leftrightarrow$	Equivalence	$\dfrac{\phi\rightarrow\psi\ \ \ \psi\rightarrow\phi}{\phi\leftrightarrow\psi}\,\leftrightarrow\,\mathrm{i}_1$ $\dfrac{\phi\rightarrow\psi\ \ \ \psi\rightarrow\phi}{\psi\leftrightarrow\phi}\,\leftrightarrow\,\mathrm{i}_2$	$\dfrac{\phi\leftrightarrow\psi}{\phi\rightarrow\psi}\,\leftrightarrow\,\mathrm{e}_1$ $\dfrac{\phi\leftrightarrow\psi}{\psi\rightarrow\phi}\,\leftrightarrow\,\mathrm{e}_2$

Table 9: KT45^n Modal Knowledge Rules [Hal05, HuR04]

Op	Name	Introduction	Elimination
K_i	Agent Knowledge	$\dfrac{\begin{array}{|c}\hline K_i \\ \vdots \\ \phi \\ \hline\end{array}}{K_i\phi}\ K_i i$	$\dfrac{K_i\phi}{\begin{array}{|c}\hline K_i \\ \vdots \\ \phi \\ \vdots \\ \hline\end{array}}\ K_i e$
E_G	Everyone Knowledge	$\dfrac{\begin{array}{|c}\hline E_G \\ \vdots \\ \phi \\ \hline\end{array}}{E_G\phi}\ E_G i$	$\dfrac{E_G\phi}{\begin{array}{|c}\hline E_G \\ \vdots \\ \phi \\ \vdots \\ \hline\end{array}}\ E_G e$
C_G	Common Knowledge	$\dfrac{\begin{array}{|c}\hline C_G \\ \vdots \\ \phi \\ \hline\end{array}}{C_G\phi}\ C_G i$	$\dfrac{C_G\phi}{\begin{array}{|c}\hline C_G \\ \vdots \\ \phi \\ \vdots \\ \hline\end{array}}\ C_G e$
D_G	Distributed Knowledge	$\dfrac{\begin{array}{|c}\hline D_G \\ \vdots \\ \phi \\ \hline\end{array}}{D_G\phi}\ D_G i$	$\dfrac{D_G\phi}{\begin{array}{|c}\hline D_G \\ \vdots \\ \phi \\ \vdots \\ \hline\end{array}}\ D_G e$

The ordinary formula ϕ cannot be brought into such dashed boxes, because the mere truth of ϕ does not mean that agent *i* or group *G* knows it [HuR04]. Additional KT45^n knowledge rules are enumerated in Table 10.

Table 10: KT45^n Derived Rules [Hal05, HuR04]

Op	Name	Derived
K_i	Closed Consequence	$\frac{K_i\phi \wedge K_i(\phi \rightarrow \psi)}{K_i\psi} K \rightarrow E$
	Substitution	$\frac{K_i\phi \ \ for\ \ each\ \ i \in G}{E_G\phi} KE \qquad \frac{E_G\phi \ \ i \in G}{K_i\phi} EKi$
	Introspection	$\frac{K_i\phi}{K_iK_i\phi} K4 \qquad \frac{\neg K_i\phi}{K_i\neg K_i\phi} K5$
	Truth	$\frac{K_i\phi}{\phi} KT$
E_G	Closed Consequence	$\frac{E_G\phi \wedge E_G(\phi \rightarrow \psi)}{E_G\psi} E \rightarrow E$
C_G	Closed Consequence	$\frac{C_G\phi \wedge C_G(\phi \rightarrow \psi)}{C_G\psi} C \rightarrow E$
	Substitution	$\frac{C_G\phi}{E_G \ldots E_G\phi} CE \qquad \frac{C_G\phi \ \ i_j \in G}{K_{i1} \ldots K_{ik}\phi} CK$
	Introspection	$\frac{C_G\phi}{C_GC_G\phi} C4 \qquad \frac{\neg C_G\phi}{C_G\neg C_G\phi} C5$
	Truth	$\frac{C_G\phi}{\phi} CT, G \neq \varnothing$
D_G	Closed Consequence	$\frac{D_G\phi \wedge D_G(\phi \rightarrow \psi)}{D_G\psi} D \rightarrow E$
	Introspection	$\frac{D_G\phi}{D_GD_G\phi} D4 \qquad \frac{\neg D_G\phi}{D_G\neg D_G\phi} D5$
	Truth	$\frac{D_G\phi}{\phi} DT$

2.6.3.3 KT45^n Semantics.

Epistemic logics consider the semantic *possible worlds* that can be constructed from the knowledge held within the system. Thus, if an agent knows a fact *p*, it will not consider those worlds in which ¬*p* is true. In expressing adversary models and agent behavior, knowledge that can be deduced by an agent from observed facts is of great importance to the anonymity the system provides. From an anonymity perspective, the objective is to avoid revealing facts that would decrease the number in valid possible worlds.

A model $\mathfrak{M} = (W, (R_i)_{i \in A}, L)$ of the multi-modal logic KT45^n with the set *A* of *n* agents is specified by three things [HuR04]:

1. Set of possible worlds *W*;
2. Accessibility relations R_i for each $i \in A$;
3. Labeling function L: $W \rightarrow P(\text{Atoms})$.

KT45^n uses relational structures called Kripke models whose elements are thought of variously as being possible worlds, moments of time, evidential situations, or states of a computer [Gol05]. Kripke semantics focus on intuitive graphs and address the key ideas of time flow (discrete integer), computations state transitions (accessibility relations) and possible world networks (worlds labeled with atomic propositions).

2.6.4 Logical Posibilistic Anonymity.

Logical possibilistic (a.k.a. purely nondeterministic) anonymity delineates what the adversary knows is possible or impossible in an anonymous system. Table 11 lists four definitions of *minimal anonymity*, *total anonymity*, *up-to anonymity* and *k-anonymity* [HaO03]. The formula $\delta_{i,a}$ means "agent *i* performed action *a*". I_A is the anonymity set.

Table 11: Possibilistic Anonymity Formulas [HaO03]

DEFINITION	FORMULA	ADVERSARY j KNOWLEDGE
Minimal Anonymity	$\neg K_j \delta_{i,a}$	Action Hidden
Total Anonymity	$\bigwedge_{i' \neq j} P_j \delta_{i',a}$	Anybody Perform Action
Up to $\lvert I_A \rvert$ *Anonymity*	$\bigwedge_{i' \in I_A} P_j \delta_{i',a}$	Up to $\lvert I_A \rvert$ Agents Perform Action
k-Anonymity	$\bigvee_{\{\lvert I_A \rvert \geq k\}} \bigwedge_{i' \in I_A} P_j \delta_{i',a}$	$\geq k$ Agents Perform Action

Minimal anonymity means the adversary does not know that an agent performed an action. More precisely, the formula means adversary *j* does not know, represented by the negated modal unary operator $\neg K_j$, that agent *i* performed action *a*, represented by the atomic formula $\delta_{i,a}$.

Total anonymity means the adversary believes the action could have been performed by anybody in the system except the adversary. $P_j \delta_{i,a}$ is an abbreviation for $\neg K_j \neg \delta_{i,a}$ meaning adversary *j* does not know that agent *i* did not perform action *a*. Thus, the adversary *j* thinks it possible, P_j, that any agent $i' \in A\text{-}\{j\}$ denoted as $\bigwedge_{i' \neq j}$ could have performed *a* or $\delta_{i',a}$ where *A* is the set of agents in the system.

Up to anonymity means the adversary believes the anonymous action could have been performed by up to $\lvert I_A \rvert$ agents in the system. More precisely, adversary *j* believes it is

possible, P_j, that any anonymous agent $i' \in I_A$ performed action a, $\delta_{i',a}$.

K-anonymity means the adversary believes the anonymous action may have been performed by at least k agents in the system. More precisely, the formula means adversary j believes it is possible, P_j, that any anonymous agent $i' \in I_A$ could have performed action a and the size of all possible anonymity sets is at least k denoted by $\bigvee_{\{|I_A| \geq k\}}$. In [HaO03], this was denoted as $\bigvee_{\{|I_A| = k\}}$, but this only means equal to k, so $\bigvee_{\{|I_A| \geq k\}}$ is used herein instead.

These represent varying degrees of anonymity with respect to the adversary j. These logical possibilistic formulae mean the adversary only believes it is probable that a certain number of agents could have performed the anonymous action.

The subset of germane grammar is

$$\phi ::= p \mid \neg\phi \mid \phi \wedge \phi \mid \phi \vee \phi \mid K_j \phi \mid P_j \phi \tag{64}$$

Hence, the anonymity definitions contain formula p, two binary operators and three unary operators ($\neg$, K_j, P_j). The negation ($\neg$), conjunction ($\wedge$) and disjunction ($\vee$) operators correspond to their typical meanings in propositional calculus. The K_j operator corresponds to the modal box operator ($\Box$) and non-variable predicate calculus universal quantifier ($\forall$). The K_j operator distributes over $\wedge$, not $\vee$. The $P_j \phi$ is short for $\neg K_j \neg \phi$ and means "adversary j thinks ϕ is possible"; however, exactly how possible is unspecified and not quantified. The P_j operator corresponds to the diamond operator ($\Diamond$)

and non-variable predicate calculus existential quantifier ($\exists$). The P_j operator distributes over $\vee$, not $\wedge$.

2.6.5 Logical Probabilistic Anonymity.

Logical probabilistic anonymity extends the possibilistic definition to quantify to what degree the adversary knows an anonymous action is possible in the system. Table 12 lists the four definitions of *α-anonymous*, *strongly probabilistic anonymous*, *weakly probabilistic anonymous*, and *conditionally anonymous* [HaO03]. These definitions are of the form $\Pr_j(\varphi) \leq \alpha$ where $\Pr_j$ is an adversary assigned *posterior* probability, φ is any fact, and $\alpha \leq 1$. The formula $\theta_{i,a}$ means "agent *i* performed action *a*" with the added implication that if the action was not performed then the adversary does not know about it (e.g., $\neg\theta_{i,a} \rightarrow \neg K_j[\theta_{i,a}]$); hence, the adversary is unable to assign probabilities to unperformed actions.

Table 12: Probabilistic Anonymity Formulas [HaO03]

Definition	Formula	Action Probability
α-anonymous	$\Pr_j(\theta_{i,a}) < \alpha$	Less than some probability threshold $\alpha \leq 1$.
Strongly probabilistically anonymous	$\Pr_j(\theta_{i,a}) = \Pr_j(\theta_{i',a})$	Uniformly distributed (*totally anonymous*).
Weakly probabilistically anonymous	$\Pr_j(\theta_{i,a}) \leq \Pr_j(\theta_{i',a})$	Non-uniformly distributed (*beyond suspicion*).
Conditionally anonymous	$\Pr_j(\theta_{i,a}) = \beta$	Unchanged after action (*a priori = a posterior*)

Let $\beta = \mu(e_r(\theta_{i,a}) \mid e_r(\varphi))$ where $e_r(\varphi)$ means action φ has occurred, $e_r(\theta_{i,a}) \mid e_r(\varphi)$ means action $\theta_{i,a}$ occurs after action φ, and $\mu(e_r(\theta_{i,a}) \mid e_r(\varphi))$ means assigning a probability that agent *i* performed action *a* given the prior action φ. Hence, β is an *a priori* probability of $\theta_{i,a}$.

α-anonymous means the adversary's assigned *posterior* probability, $\Pr_j(\theta_{i,a})$, must be less than one or some probability threshold α. *Strongly probabilistically anonymous* means the adversary is only able to assign a uniform distribution to the anonymity set of agents so agent *i*'s action has *total anonymity*. More specifically, the *posterior* probability of agent *i* performing action *a*, $\Pr_j(\theta_{i,a})$, is equal to the probability of any other anonymous agent *i*' performing the same action *a*, $\Pr_j(\theta_{i',a})$.

Weakly probabilistically anonymous means the adversary is able to assign a non-uniform distribution to the anonymity set of agents yet agent *i* is *beyond suspicion* or *possible innocent*. More specifically, the *posterior* probability of agent *i* performing action *a*, $\Pr_j(\theta_{i,a})$, is less than or equal to the probability of any other anonymous agent *i*' performing the same action *a*, $\Pr_j(\theta_{i',a})$.

Conditionally anonymous means the adversary *posterior* probability, $\Pr_j(\theta_{i,a})$, is the same as the *a priori* probability, β. Hence, the adversary is unable to learn anything new given $\theta_{i,a}$. This is equivalent to preserving anonymity or when normalized entropy anonymity degree is one (d=1).

2.6.6 Temporal Logics.

Temporal logics add time to propositions which allows logics to express not only the truth of propositions, but also when the truth holds. This greatly enhances the expressive power of logic but at the cost of added complexity [WrS05]. Modal temporal logics may be able to express additional properties in anonymous systems. For example, it may be desirable to prove that a certain fact concerning an agent is true at a particular moment in time, such as having a certain pseudonymous identity performing an action. However, it

may be undesirable for an adversary to known this information for extended periods of time and discover the real identity. Temporal logics allow propositions that are true at certain times, but not at others. For example, one approach [Men05] views time as a sequence of events and defines four operators, two weak and two strong [WrS05] or alternatively two about the past and two about the future. Let θ be an arbitrary event and define two operators:

· Past Operators

$P\,\theta$: θ has at some time been true.
$H\,\theta$: θ has always been true.

· Future Operators

$F\,\theta$: θ will at some time be true.
$G\,\theta$: θ will always be true.

Similar to KT45^n, the duality of operators hold so $P\,\theta = \neg H \neg\,\theta$ is "θ has at some time been true" = "it is not always the case that θ has not been true". Also, $F\,\theta = \neg G \neg\,\theta$.

Modal temporal logics are the most common [ChH04, Gol05, HuD01, Hui04, KoS04, MoS06, OrL06, SuK04]. The KARO logic [HuD01] offers ways to do automate reasoning about agent-based systems using an expressive combination of modal logics. One method uses branching-time temporal logic [JiK05] and a KT45^n-like logic with a clausal resolution calculus. The Typed Modal Logic (TML) combined with a temporal logic [OrL06] offers ways to model and reason about evolving trust and beliefs for multi-agent systems. Spatial Propositional Neighborhood Logic [MoS06] is a semi-decidable, modal logic for spatial reasoning that can be polynomially reduced to a decidable temporal logic based on time intervals preserving, at least, valid formulas. Another new

modal logic [ChH04] for the π-calculus, an extension of the modal μ-calculus with Boolean expressions over names, is introduced as an appropriate temporal logic for the π-calculus to perform model checking.

However, there has been little research into using temporal logics to express anonymity, or even security properties. This may be due to the complexity of temporal logics, combined with the ability to abstract away the temporal element of protocols [WrS05]. Few existing protocols use explicit timing information, relying instead on single-use values, *cryptographic nonce* [And01], which indicates an event took place without any reference to the time domain. The alternative framework of process calculi is examined next.

2.7 Process Calculi

Process calculi provide a mathematical notation for describing communicating processes. Computers are viewed as communicating agents in larger networks. Since anonymous systems are concerned with communication between agents, process calculi is an excellent way to express anonymity.

2.7.1 Communications Sequential Processes (CSP).

Communicating Sequential Processes (CSP [BrH84, Hoa04]) is a formal language for describing patterns of interaction in concurrent systems and is a member of the family of mathematical theories of concurrency. CSP was initially introduced in 1978 but has evolved substantially to include real-time [ReR88], probabilistic [SeM96] and larger scale system expansions [Cre01]. CSP has the basic constructs of a typical programming

language such as choice operators and logical expressions. The core concept is a process as a mathematical abstraction of the interactions between a system and its environment.

2.7.1.1 System Model.

A system is modeled in terms of events it can perform and is composed of a number of processes. Processes are defined in terms of a sequence of possible events using the prefix operator ($\rightarrow$). For example, $x \rightarrow y \rightarrow P$ means performing event x then event y acts like process P. Intuitively, $LIGHT = on \rightarrow off \rightarrow LIGHT$ means turning *on* then *off* acts like process *LIGHT*. This is pictorially represented in Figure 33.

The circles represent states of the process, and the arrows represent transitions between states. The top circle is the starting state. Each down arrow is labeled by the event which occurs on making that transition. Arrows leading from the same node must have unique labels. The unlabeled arrow from the bottom to the top circle is an immediate and imperceptible transition, making the process unbounded [Hoa04]. Hence, process LIGHT may turn *on* then *off* again continuously. A *traces*(*P*) is a finite sequence of events that *P* may perform. For instance, an empty trace $\langle\rangle$ or three-event trace $\langle on, off, on\rangle$ are two instantiations of *traces*(*LIGHT)*.

A process *P* is refined by a process *Q*, denoted as $P \sqsubseteq Q$, if $traces(Q) \subseteq traces(P)$. Two processes are equal $P = Q$ if each refines the other, namely $P \sqsubseteq Q$ and $Q \sqsubseteq P$. The definition of anonymity requires processes to be equal in this manner. An automated

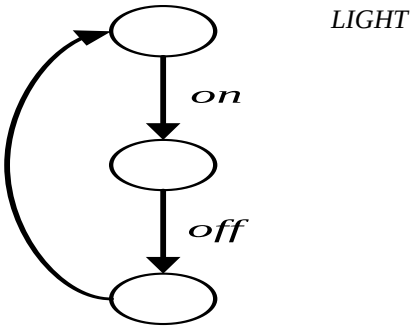


Figure 33: Unbounded Process *LIGHT* = *on* → *off* → *LIGHT*

model-checking tool is used to check for such equality. For instance, let two concurrent processes (agents) be defined as $P = x \rightarrow P$ and $Q = (x \rightarrow Q \mid y \rightarrow Q)$ where x and y are events of sending messages and | is a choice operator. Hence, P may only send message x but Q may send both x and y messages. The processes P and Q are depicted in Figure 34.

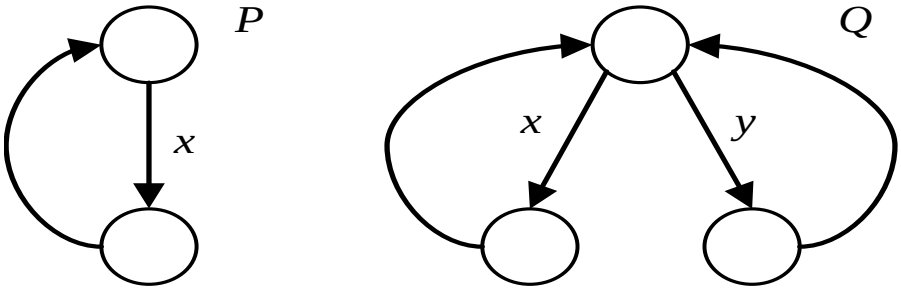


Figure 34: Two Processes (agents) *P* and *Q*

If Q decides to send one x message, then $traces(Q) = \langle x \rangle$. However, if P sends one x message, then $traces(P) = \langle x \rangle$. Since $traces(Q) \subseteq traces(P)$, then $P \sqsubseteq Q$. Also, $Q \sqsubseteq P$ so $P = Q$ and the processes are equal. In other words, if the adversary observes a single message x, then the traces are indistinguishable and sender anonymity is preserved. However, if Q decides to send any y messages, then the traces are distinguishable and no

anonymity exists. Given the sequential execution of the two processes *P* and *Q*, the following operations may be performed.

· Basic Operations

$P(n)$	: Process *P* parameterized with value *n*.
$?x{:}\mathrm{E} \rightarrow P(x)$	: Perform any event $x \in P$, then behave like $P(x)$.
$P \,\square\, Q$	: Deterministically choose between the initial events of *P* and *Q*, and then behave accordingly.
$b\&P$	: If (boolean) *b* then enable *P* else *STOP*.

• Parallel Composition

$P \| Q$	: *P* and *Q* require full synchronization of events.
$P \|\|_x Q$	: *P* and *Q* require full synchronization of set of *X* events.
$P \| \| \| Q$	: *P* and *Q* without synchronization.
$P \backslash Q$	: Hide set *Q* events from adversary.
Pa/b	: Rename all variables *a* in *P* to *b*.

• Primitive Processes

STOP	: Deadlocked process.
SKIP	: Successfully terminating process.

CSP focuses on the simplest form of sets of observations of process traces, *traces*(*P*), and process equality, $P \sqsubseteq Q$ and $Q \sqsubseteq P$. Other more complex observations such as *failures*, *divergences*, and *refusals* contain additional information about system state and enhance the ability to reason about a process.

2.7.1.2 Applications.

CSP has been applied in industry as a practical tool for specifying and verifying concurrent aspects of a variety of different systems including the T9000 Transputer

[Bar95] and a secure ecommerce system [HaC02]. Anonymity has also been formalized in CSP [ScS96].

The model draws an analogy between existing features of CSP and aspects of anonymity. For example, hiding CSP events from the view of other processes models the anonymous sending of a message. Parallel execution of processes models an anonymity set of processes that could have performed an action. The anonymity property is the existence of indistinguishable *traces*, a sequence of actions observable to the adversary, for any sender. By assuming a reliable broadcast channel and a passive adversary and analyzing the trace observations, process equivalence or, synonymously, sender anonymity is proven for the three-agent dining cryptographer network [Cha88]. The model is highly specialized and only has the broadest applicability to other anonymity systems.

Nonetheless, this is one of the few examples of a formal methods proof of anonymity and provides inspiration for further work into proving anonymity properties with process calculi. Adding the probabilistic aspect [ScS96] is essential to successfully modeling real anonymity-providing services.

2.7.2 π-Calculus.

The $\pi-$calculus is a derivative of Calculus of Communicating Systems (CCS [Mil89]). CCS and CSP describe communicating processes and offer the same level of expressive power. However, the $\pi-$calculus extends the basic capabilities of CCS to include *mobility*: agents can form new and destroy old links with other agents. An agent may therefore begin in one area of a system and, in the course of execution, relocate to an

entirely new portion of a system. Processes send and receive messages along defined channels and these messages may include the name of a channel. This powerful addition allows the dynamic creation of new topologies in the system. The basic structure of the calculus is presented below.

2.7.2.1 Syntax.

The fundamental structure of $\pi-$calculus enumerates over a set of names and includes a prefix and process syntax. Let ***N*** be a countable set of *names*, *x*, *y*, …. The set of prefixes, α, β, … syntax is

$$\textit{Prefixes} \quad \alpha ::= x(y) \mid \bar{x}\, y \mid \tau. \tag{40}$$

The prefixes are basic process actions of *input*, *output*, and *silent*, respectively or

1) $x(y)$ is the *input* of the name *y* from channel *x*;
2) $\bar{x}\, y$ is the *output* of the name *y* on channel *x*;
3) τ is any *silent* action.

The set of $\pi-$calculus processes syntax is

$$\textit{Processes} \quad P ::= \sum_i \alpha_i.P_i \mid \nu x P \mid \quad P \mid P \quad \mid !P \mid [x = y]P \mid [x \neq y]P. \tag{41}$$

The processes are *guarded choice*, *restriction*, *composition*, *replication*, and *if-then-else*, respectively or

1) $\sum_i \alpha_i.P_i\, x$ is *guarded choice* or execution of an action

   where **0**=*inaction*, $\alpha.P$ =*unary sum*, $P+Q$=*binary sum*;
2) $\nu x P$ is *restriction*;
3) $P \mid P$ is *composition*;
4) $!P$ is *replication*;

5) $[x = y]P \,|\, [x \neq y]Q$ is **if** x=y **then** P **else** Q where $P \neq Q$.

2.7.2.2 Semantics.

Operational semantics is specified via a transition system labeled by *Actions* $\mu,\mu', \ldots$ given by the grammar

$$\textit{Actions} \quad \mu ::= xy \,|\, \bar{x}\, y \,|\, \bar{x}(y) \,|\, \tau. \tag{42}$$

The actions are *input prefix* (xy), *free name output* ($\bar{x}\, y$), *bound output* ($\bar{x}(y)$), and *silent* (τ). The bound name of an action μ, $bn(\mu)$, is defined as $bn(xy) = bn(\bar{x}\, y) = bn(\tau) = \varnothing$; $bn(\bar{x}(y)) = \{y\}$. Names may be passed along channels. Processes have the ability to run both sequentially and in parallel. Replication can be expressed and the scope of names may be restricted to processes using the ν operator.

2.7.2.3 Variants and Applications.

The $\pi-$calculus has spawned variants designed for the analysis of various interacting systems. One variant is spi-calculus [AbG97] which adds cryptographic primitives. Another is an extension of the modal μ-calculus [Alb02] with Boolean expressions over names, and primitives for name input and output as an appropriate temporal logic for the $\pi-$calculus [ChH04]. Other variants include Update Calculus [PaV97], Probabilistic Asynchronous $\pi-$calculus [HeP00], and $\pi_{prob}-$calculus [ChP05]. The latter $\pi_{prob}-$calculus is able to analyze probabilistic security protocols

involving probabilistic choice in applications such as sending certified e-mail and protecting the anonymity of communicating agents. Recently, pattern-matching spi-calculus [HaJ06] has been introduced to provide a framework, methods and tools, to rigorously analyze security protocols. Proving security protocols using the $\pi-$calculus and its variants uses *observational equivalences* between processes by comparing protocol models and abstract specification of security properties specifications. Using the calculus, equivalence is established between the model of the protocol and the abstract properties.

Executable languages based on $\pi-$calculus have been developed such as an executable specification for asynchronous $\pi-$calculus [ThS05]. The existence of languages in which $\pi-$calculus models can more easily be expressed would increase the utility of the calculus. In [BhP05], the Dining Cryptographer anonymous system is modeled and the probabilistic extension π_p- calculus is proposed. The flexibility offered by the calculus is ideal for representing many of the network topologies used in modern anonymity systems. The existing body of knowledge of $\pi-$calculus security proofs provides a source of techniques that may be fruitful in proving anonymity properties.

2.7.3 Comparison.

In theoretical computer science, CSP and $\pi-$calculus are the most common formal methods in security research. Other existing process calculi include the International Organization for Standards (ISO) Language of Temporal Ordering Specification (LOTOS [EiV89]) for formal descriptions of systems, Algebra of Communicating Processes with Abstraction (ACP [BeK85]) for asynchronous process cooperation via synchronous

communication and many additional $\pi-$calculus variants. CSP and $\pi-$calculus differ in three important ways: semantics, maturity, and mobility.

First, both deal with the rigorous mathematical study of the semantics of programming languages and models of computation [WiK07c]; however, each uses a different, albeit possibly relatable [ZhH06] semantic approach. CSP uses denotational semantics [Bou89, ScS71] whereas $\pi-$calculus uses algebraic semantics [GoT77, ZhN05]. Denotational semantics loosely deals with compilation and translates each language phrase into a mathematical formalism rather than another computer language. The computer program is interpreted as a function that maps inputs to outputs. Algebraic semantics is a form of axiomatic semantics [Hoa69] based on mathematical logic to prove the correctness of computer programs. Each language phrase is interpreted as a description of the relevant logical axioms or algebraic forms. In both, semantically demonstrating description equivalences between systems is the method for proving anonymous communications.

Second, $\pi-$calculus is a less mature language and formalism than CSP. CSP is supported by mature proof tools such as logically embedded Higher Order Logic for Z specifications (HOL-Z) and special purpose Failure-Divergence Refinement (FDR [FDR97]) model-checker. Ways to transform the CSP abstract language into executable forms have been proposed [Gar03, Pel05, Ste03]. The ability to efficiently execute abstract models and proofs is of immense practical value in addition to theoretical value for experimenting in real-world environments. There have also been efforts to produce an executable form of $\pi-$calculus such as Nomadic Pict [UnS01], but these are not as well developed as in CSP.

Lastly, $\pi-$calculus, unlike CSP, is able to explicitly model mobility. Channel names passed in data messages enable non-static links between agents in the system. The ability to create and destroy links models of dynamic interactions between anonymous agents in mobile ad hoc networks. Both CSP and the $\pi-$calculus can be extended to express cryptographic operations, asymmetric communications and probabilistic protocol behaviors; however, only $\pi-$calculus is able to express mobility. This is a key advantage even with CSP's extensive mature tool support.

2.8 Function Views

Function views and *opaqueness* are other defined and succinct ways to formally express anonymity. The main advantage of these are restrictions can be placed on relationships between agents and actions. This functional relationship expression allows a local adversary to be modeled by limiting the adversary view of such relationships. Defining a function from a set of actions to a set of agents who performs those actions and by specifying the opaqueness of the function to the adversary, anonymity may be represented.

2.8.1 Function Knowledge.

An adversary's uncertainty associated with a given function is modeled using function knowledge. The aspects of knowledge about a function are its graph *f*, image *im f* and kernel *ker f*. The graph *f* is the set of ordered pairs $(x, f(x))$, for all x in domain X. The *im f* is the function value at x, namely $f(x)$ or y. The *ker f* is a binary equivalence relation of the function domain X, is a subset of the Cartesian product $X \times X$, and is symbolically defined as $ker\ f := \{(x, x') \mid f(x) = f(x')\}$ where $x, x' \in X$. The *function view*

is a mathematical abstraction of partial knowledge of a function, namely a nondeterministic approximation of graph *f*, a subset of *im f*, and a *ker f* equivalence relation. Functional knowledge of function f: $X \rightarrow Y$ is represented by the triple $N = (F,I,K)$, where domain X is a set of actions, codomain Y is a set of agents, $F \subseteq X \times Y$ maps actions to agents, $I \subseteq Y$ is the anonymity set, and $K \sim X$ is an equivalence relation on the set of actions. Intuitively, (F,I,K) represents what the adversary may know about function *f*. Complete knowledge of function *f* is represented by (*f, im f, ker f*).

2.8.2 Opaqueness.

Anonymity is concerned with what an adversary does not know. *Opaqueness* formalizes this lack of functional knowledge. Given $N = (F,I,K)$, N is *k-value opaque* if $|F(x)| \geq k \ \forall x \in X$. In other words, each action x is at least *k-anonymous* to the adversary. Also, N is *Z-value opaque* if $Z \subseteq F(x) \ \forall x \in X$. In other words, for each action x no agent in Z may be ruled-out as having performed that action. Furthermore, N is *absolutely value opaque* if N is Y-value opaque. In other words, for each action x any agent $y \in Y$ could have performed it. Hence, opaqueness describes anonymity properties.

Z-value opaqueness is more precisely defined below. Intuitively, $f(x) = y$ if agent y has performed action x and $f(x)$ is undefined if no agent y has yet performed action x. If $f_{(r,m)}(x) = y$, agent y performed action x at point (r,m). Let Γ' be an interpreted system that satisfies $(\Gamma',r,m) \models f(x) = y$ whenever $f_{(r,m)}(x) = y$ [HaO03].

Definition 4 [HaO03]: In system Γ', f is *Z-value opaque* for adversary j at point (r,m) **iff**

$$(\Gamma',r,m) \models \bigwedge_{x \in X} \bigwedge_{z \in Z} P_j[f(x) = z].$$

The adversary *j* believes |*Z*| agents may have performed each action *x*. This function view opaqueness strongly resembles the previous definitions of anonymity. Hence, *function views* and *opaqueness* are other valid methods to express and quantify anonymity.

2.8.3 Modular Approach.

A modular approach [HuS04] uses partial knowledge about the function *f* to model and quantify anonymity using epistemic logic and process calculi. Epistemic logic models the system. The system is all possible states of a Kripke structure [Kri63]. This structure represents the adversary's view of the system and is a nondeterministic finite state machine with all states in the machine processing Boolean labels that express the evaluation of that state. The key aspect of this formalism is that *any* Kripke structure results in *function views* [HuS04]. Observational equivalences from process calculi express the observable differences between system configurations. As mentioned above, anonymity is defined in terms of *opaqueness*, the information an adversary may learn about a specific function within the *function view* framework. Higher levels of *opaqueness* conceal larger amounts of information in the function and equate to higher levels of uncertainty about which aspects of a system are linked.

One case study [HuS04] uses this framework to analyze an anonymity property of keeping communicating agent identities secret (sender/receiver anonymity) and a privacy property of keeping agent relationships secret (communication anonymity). Proving these properties hold is demonstrated but is not a trivial task.

This modular *function view* approach is an adaptable, intriguing approach to defining

and analyzing anonymity. A comparison between conventional and modular approaches is highlighted in Figure 35.

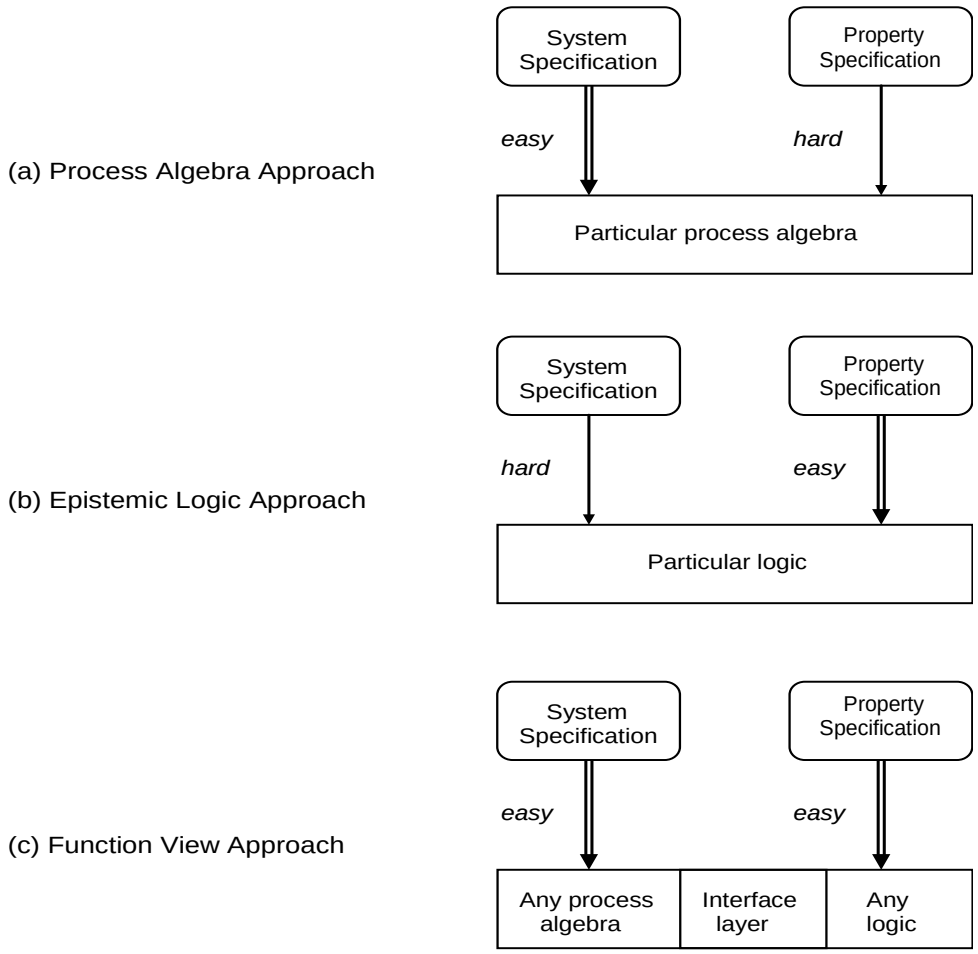


Figure 35: Modular Approach to Formalizing Information-Hiding Properties [HuS04]

For the process calculi approach in Figure 35(a), system specification is easy but property specification is hard. The particular process calculi may be CSP or $\pi-$ calculus. For the epistemic approach in Figure 35(b), system specification is hard but property specification is easy. The particular logic may be any modal logic such as $KT45^n$. For the modular function view approach in Figure 35(c), system and property specifications are easy. The interface layer allows any epistemic and process calculi to

be selected. This overall modular approach may provide keen insight into developing other frameworks for modeling, measuring, and analyzing anonymity.

2.9 Summary

This chapter provided a comprehensive coverage of state-of-the-art concepts in anonymous communications systems. The background section succinctly recounted the societal pursuit of personal privacy and describes identity, anonymity, pseudonymity, and reputation. The anonymity benefits of promoting freedom of expression and protecting user privacy and drawbacks of extreme abuse and illegal activity were discussed. The nomenclature section was a synthesis of the essential elements of anonymous systems and summarizes the anonymity properties, the adversary, the attacks, and mix technology. The three high-level anonymity properties of unidentifability, unlinkability, and unobservability were described. The three adversary capabilities that determine the threat model were mentioned. The goal of and defense for five active and nine passive attacks on anonymous systems were delineated. The anonymous communications networks described seventeen wired and sixteen wireless protocols designed for preserving anonymity. Over ten different ways to measure anonymity were illustrated in the quantifying anonymity section. The anonymity set size, individual anonymity degree scale, and information-theoretic entropy metrics are the classical approaches but negligibility-based, localized real-time, combinatorial, evidence-based, and multicast metrics have also been proposed. The remaining sections introduced formal methods for analyzing anonymity preservation in anonymous systems. The formalizing anonymity section explored three conceptual frameworks, the probabilistic versus nondeterministic

approaches to modeling anonymous system, the notion of group instead of individual anonymity, and multi-agent systems. Epistemic logic, such as $KT45^n$, and temporal logic were discussed in the logics section. The two most common process calculi, CSP and $\pi-$calculus, used in theoretical computer science for security research were described. Their semantic, maturity, and mobility differences were portrayed and some recent extensions are designated. Finally, a modular approach that combines both a process calculi anonymous system specification and epistemic logic anonymity property specification formal approach was explained in the function views section. This approach introduced function knowledge and opaqueness and requires the introduction of an interface between two different formal approaches.

III. Methodology

3.0 Chapter Overview

This chapter presents the methodology used in this research effort. The research is in three areas. First, an innovative anonymity network taxonomy is developed. Second, an evaluation and aggregation of emerging anonymity metrics is conducted. Lastly, a formal adversary anonymity reasoning framework is created. These three phases constitute three underdeveloped yet mutually complementary subtopics of open and relevant anonymity research. In Section 3.1, the motivation for exploring each of these phases is provided. Each research and development phase is elaborated on in Section 3.2. Section 3.3 concludes this chapter.

3.1 Motivation

This section further explains the reasons for pursuing these three areas of research. Figure 36 shows anonymity publications by topic from 1980 to 2008 from the authoritative bibliography source of Freehaven [Fre09]. The topics of "Anonymous Communications" and "Traffic Analysis" clearly lead the field of anonymity research with 101 and 66 papers, respectively. The anonymous communications topic is replete with theoretical and/or implemented wired and wireless anonymous protocols designed for particular applications such as e-mail, voice-over-IP, hostile military environments, video teleconferencing, and multicast services as described in Section 2.3. The traffic analysis topic contains papers that analyze various cyber attacks against these anonymous

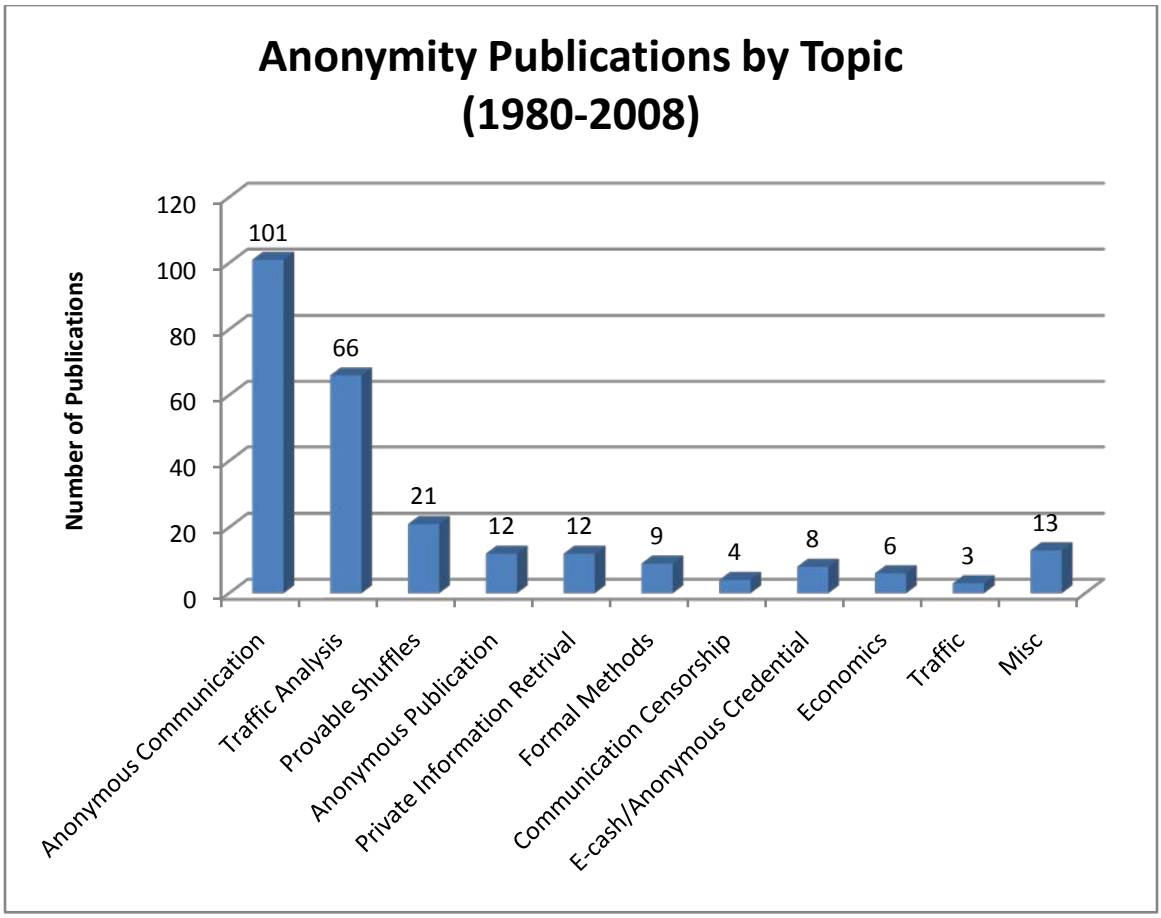


Figure 36: Freehaven's Anonymity Publications by Topic (1980-2008)

protocols. Unfortunately, these combined topics result in a large and diverse set of anonymity metrics to compare one anonymous protocol with another as discussed in Section 2.4. In contrast to these leading topics, the topic of "Formal Methods" has only nine published papers. A formal treatment entails building an appropriate mathematical model for representing anonymous protocols, and formulating, within that model, a definition of anonymity that captures the requirements of a particular application domain. Hence, research for this topic has been limited.

This motivated further investigation into research subtopics of anonymity taxonomy, metric synthesis, and epistemic-based formal methods. All known relevant anonymity publications by subtopic from 1980 to 2008 are displayed in Figure 37.

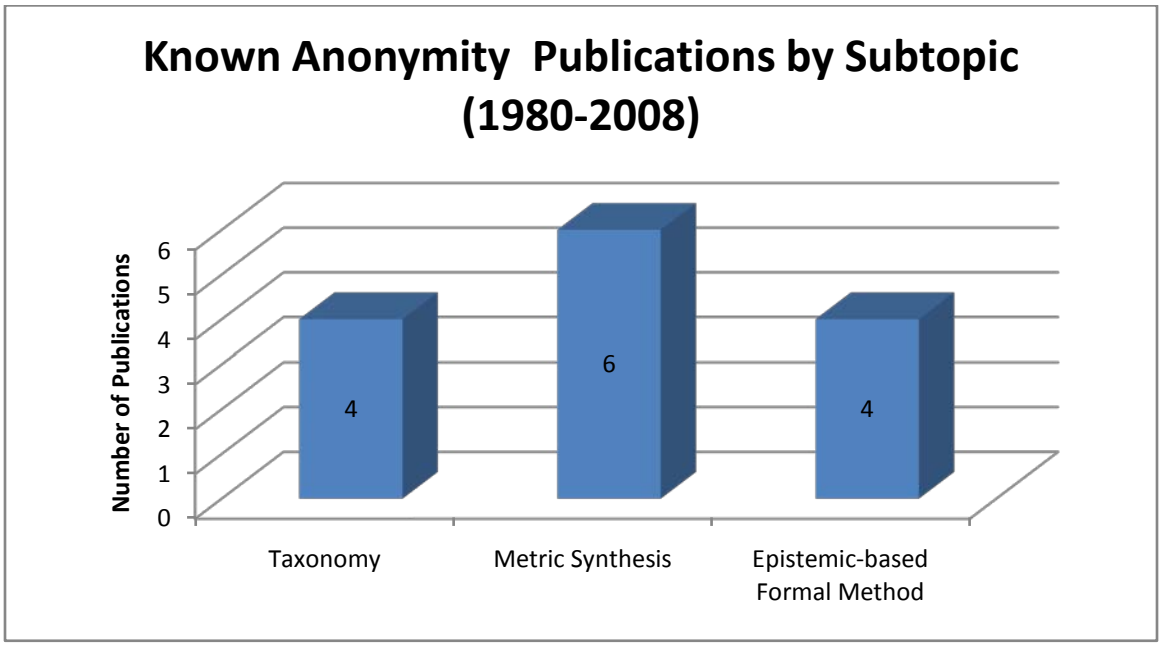


Figure 37: Anonymity Publications by Subtopic (1980-2008)

With only four or six papers published per subtopic over the last nearly three decades, these subtopics are prime areas for contributing to the field of anonymity research. Thus, this research extends the knowledge in the areas of anonymity taxonomy [Dia05c, DiP04, TiO05, VaD92], metrics synthesis [DcS02, Dij06, MuW08, NeM03, SeD02, TgH04a], and epistemic-based formal methods [GaH05, HaO03, HuS04, SyS99].

The anonymous network taxonomy examines a representative set of implemented or proposed wired and wireless anonymous protocols in the "Anonymous Communications" topic but, more importantly, classifies recent wireless anonymous networks. For these

anonymous protocols and "Traffic Analysis" performed, existing anonymity metrics are thoroughly examined. Finally, a logical formal model is created to model how an adversary reasons while attempting to degrade anonymity.

3.1.1 Develop Anonymous Network Taxonomy.

No taxonomy classifies anonymity in the diverse set of both wired and wireless anonymous communications networks. Current taxonomies are either for group support systems, low-density mobile ad hoc networks, fixed-connection-based networks, or cascade mixnets. Thus, an intuitive anonymous network taxonomy is developed to encapsulate and generalize the key ideas in state-of-the-art anonymous communications systems in order to categorize anonymous networking protocols, assumed adversary threat models, required anonymity properties, external environmental factors, and inherent interrelationships. This highlights the importance and intricacy of anonymity, serves as a modern model for theoretical and empirical investigations into anonymity, and fosters future anonymous protocol design and development across multiple application domains. Furthermore, it updates and merges key aspects of existing taxonomies with location anonymity and multicast or anycast group anonymity.

3.1.2 Evaluate Emerging Anonymity Metrics.

Anonymization enables organizations to protect their data and systems from a diverse set of cyber attacks and preserve privacy; however, recent research indicates that many anonymization techniques leak at least some information. Furthermore, there are confusing arrays of anonymity metrics and definitions for quantifying anonymity across a network. The ability to confidently measure this information leakage and changes in

anonymity levels across a network plays a crucial role in facilitating the free-flow of cross-organizational information sharing and promotes wider adoption of anonyimzation techniques. Although there are multiple methods of measuring analyzing anonymity, current research focuses on information theory, mobile ad hoc network, low-latency wired networks, or mixnet-specific metrics. In other words, there is no "one-stop-shop" research that comprehensively surveys this area for candidate measures; therefore, this research explores the state-of-the-art of anonymity metrics to provide a macro-level view of the systematic analysis of anonymity preservation, degradation, or elimination in cyberspace.

3.1.3 Create a Formal Model.

While the first phase offers a holistic approach to anonymity and the second phase thoroughly examines how anonymity has been, is and can be measured, the third phase creates a mathematical framework for anonymity. Rigorously demonstrating that a protocol meets expectations is an essential component of cryptographic protocol design. The same should hold for anonymous protocol design. The formal model should be rich enough to represent a large variety of real-life adversarial behaviors, and the definition should guarantee that the intuitive notion of anonymity is captured for any adversarial behavior under consideration. Thus, the goal is to expand upon existing epistemic-based formal anonymity methods and models. A possibilistic (i.e., non-deterministic) approach to anonymous system and several anonymity properties are specified. The primary step includes proving multiple anonymity definitions are satisfied given an epistemic syntactic specification and possible world's semantic interpretation. The contribution of this

research is the introduction of a formal adversary anonymity reasoning model to rigorously analyze how anonymity is preserved or degraded in an anonymous network.

3.2 Summary

This chapter presents the motivation and methodology for the development of an innovative taxonomy for the systematic analysis of anonymity properties and adversary knowledge in anonymous communications networks. First, with the aim to preserve privacy over a communications network, many anonymous protocols have been proposed along with many empirical investigations into specific adversary attacks over those networks but no known taxonomy addresses anonymity in the diverse set of both wired and wireless anonymous communications networks. Second, anonymization techniques still leak some information so an ability to confidently measure any changes in anonymity levels plays a crucial role in facilitating the free-flow of cross-organizational information sharing and promoting wider adoption of anonyimzation techniques. Third, many empirical investigations lack a rigorous approach to defining and modeling anonymity concepts to ensure information assurance as is customary when formally proving other security aspects of a system. An ability to comparatively and quantitatively analyze these anonymity protocols and anonymity services to better understand how anonymity is lost, maintained or improved during a cyber attack is an area of open research.

IV. Anonymous Network Taxonomy Analysis and Results

4.0 Chapter Overview

To preserve privacy over a communications network, numerous anonymous protocols have been proposed along with many empirical investigations into specific adversary attacks over those networks. However, there are no known taxonomies that address anonymity in the diverse set of both wired and wireless anonymous communications networks. This chapter describes a novel cubic taxonomy which explores the three key components of anonymity property, adversary capability, and network type. A two dimensional (2D) tree-based taxonomy is provided for over thirty anonymous protocols. This taxonomy expands the definition of anonymity and advances the state-of-the-art technological privacy-preserving mechanisms in anonymous networks against any adversary.

The rest of the chapter is organized as follows. Section 4.1 defines the anonymity property component. The adversary capability component is delineated in Section 4.2. Section 4.3 details the network type component. Section 4.4 demonstrates the utility of CT by classifying anonymous networks in 3D cubic and 2D tree taxonomies. Section 4.5 concludes the chapter.

4.1 Anonymity Properties

Anonymity properties are generally classified into *unidentifiability*, *unlinkability*, and *unobservability*; however, only the former two are included in this taxonomy since the latter automatically implies anonymity as explained in Section 2.2.1. *Unidentifiability* means the adversary is unable to discern an agent's or group's identity, actions or other

items-of-interest (IOI) among a similar set of agents or groups. *Unlinkability* means the adversary is unable to relate agents, messages, actions or other IOI by observing the system. Moreover, an adversary's a priori and a posteriori knowledge are the same even after observing the IOI. The classical definition of anonymity is:

$$\text{Anonymity} = \textit{Unidentifiability} + \textit{Unlinkability}. \tag{43}$$

Each anonymity property may be defined by what information the anonymous system is designed to hide. Table 13 lists each property, its subcomponent type and hidden information. The next sections describe each property further.

Table 13: Anonymity Property

Property	Type	Hidden Information
Unidentifiability	Sender Anonymity (SA)	Message sender identity
	Receiver Anonymity (RA)	Message receiver identity
	Mutual Anonymity (MA)	Message identities from each other
	Group Anonymity (GA)	Message group identity
	Location Anonymity (LA)	Position, motion, link, or topology information
Unlinkability	Communication Anonymity (CA)	Sender-Receiver pair relationship from others
	Group Communication Anonymity (GCA)	Group-Group pair relationship from others

4.1.1 Unidentifiability

Unidentifiability is composed of *sender anonymity* (SA), *receiver anonymity* (RA), *mutual anonymity* (MA), *group anonymity* (GA), and *location anonymity* (LA) [PfK00]. SA prevents a particular message from being linked to a particular sender identity. RA prevents a particular message from being linked to a particular receiver identity. MA

hides the sender and receiver identities from each other. GA limits the adversary to linking a particular message to a group of agents. Agent identity is hidden among a group of indistinguishable agents. At a higher level of abstraction, group anonymity prevents a particular message from being linked to a particular group of agents. However, no known group anonymous services yet exist. The MAM aims to achieve both mutual and group anonymity. LA means a particular message is not linkable to any sender or receiver location, motion, route or topology information. The classic, current, and extended cubic unidentifiability property definitions are:

$$\textit{Classic Unidentifiability} = \text{SA} + \text{RA} \tag{44}$$

$$\textit{Current Unidentifiability} = \textit{Classic Unidentifiability} + \text{LA} \tag{45}$$

$$\text{Cubic Unidentifiability} = \textit{Current Unidentifiability} + \text{MA} + \text{GA} \tag{46}$$

4.1.2 Unlinkability

Unlinkability consists of *communication anonymity* (CA) and *group communication anonymity* (GCA). A particular message with CA cannot be linked to any sender-receiver pair and no message is linkable to a particular sender-receiver pair. CA is a weaker property than *sender* and *receiver* anonymity. GCA means a particular message cannot be linked to any sender group-receiver group pair and no message is linkable to a particular group sender-group receiver pair. All known anonymity research on the unlinkability property primarily deals with CA. The classic and extended cubic unlinkability property definitions are:

$$\textit{Classic Unlinkability} = \text{CA} \tag{47}$$

$$\textit{Cubic Unlinkability} = \textit{Classic Unlinkability} + \text{GCA} \tag{48}$$

Given these first two anonymity properties, the classic and expanded anonymity definitions are:

$$\begin{aligned}\text{Classic Anonymity} &= \textit{Classic Unidentifiability} + \textit{Classic Unlinkability} \\ &= \text{SA+ RA + CA}\end{aligned} \tag{49}$$

$$\begin{aligned}\text{Expanded Anonymity} &= \textit{Cubic Unidentifiability} + \textit{Cubic Unlinkability} - \text{Classic Anonymity} \\ &= \text{LA + MA + GA + GCA}\end{aligned} \tag{50}$$

Finally, the new cubic anonymity definition is:

$$\begin{aligned}\text{Cubic Anonymity} &= \textit{Cubic Unidentifiability} + \textit{Cubic Unlinkability} \\ &\quad \text{OR} \\ &= \text{Classic Anonymity + Expanded Anonymity}\end{aligned} \tag{51}$$

4.2 Adversary Capability

An adversary is an agent or set of agents whose aim is to degrade or eliminate anonymity. The adversary capabilities range from weak to strong and represent the assumed threat model. Table 14 lists capabilities, their type and a brief description. The next sections explain each capability further.

Table 14: Adversary Capability

Capability	Type	Description
Reachability	Global	Omnipresent
	Local	Limited omnipresent
Attackability	Passive/External	Compromise links
	Active/Internal	Compromise nodes
Adaptability	Static	*A priori* knowledge
	Dynamic	*Posterior* knowledge

4.2.1 Reachability.

Reachability is either *global* or *local*. A *global* adversary is omnipresent and has full access to the entire network of nodes and links. A *local* adversary has limited omnipresence and has full access to only a portion of the network nodes and links. This corresponds to the adversary possessing complete or restricted information or knowledge about the system. It may also refer to the veracity of this information. The adversary may either know things to be true or only believe things to be true.

4.2.2 Attackability.

Attackability is the combination of *passive/external* or *active/internal*. The objective of any attack is to link sender and receiver, identify the sender or receiver for a particular message, trace a sender forward/receiver back to messages or disrupt the system.

A *passive/external* adversary is an outsider that can only observe messages traversing the network and is typically invisible. This adversary can only compromise communication channels between nodes. In other words, it is a non-empty set of agents, part of the surrounding of the anonymous system and capable of compromising links.

An *active/internal* adversary is an insider and may alter messages traversing the network but is visible. This adversary controls nodes in the network. In other words, this describes a non-empty set of agents which are part of the anonymous system and capable of participating in normal communications and controlling at least some nodes.

4.2.3 Adaptability.

Adaptability describes whether the adversary or the anonymous system is *static* or *dynamic*. Typically, the adversary is *dynamic* and collects information about the path

selection algorithm, its parameters and as much information as possible about network activities from compromised nodes and links. The adversary uses all available facts to infer who sent or received which messages in a computationally bounded or even unbounded manner. The adversary may behave deterministically with a scheduled plan of attack, probabilistically depending on the relative frequency of sequences of observed actions or events, or non-deterministically (unpredictably). The adaptability of the anonymous system determines if or how much information is leaked to the adversary. A *static* system keeps adversary knowledge about the network and agent targets constant during and after an attack. The adversary retains only *a priori* knowledge. A *dynamic* system may attempt to counter an adversary's ongoing attack but may allow the adversary to learn additional information and update knowledge about the network and agent targets. So the adversary's *a posterior* knowledge may be greater than *a priori* knowledge. The network types are described next.

4.3 Network Types

Anonymous networks exist as either *wired* or *wireless*. Anonymous communications networks typically vary in routing scheme, transmission medium, topology, and protocol implementation which affect the adversarial threat. Hence, providing anonymity in each network requires a different approach particularly when mobility is involved. Table 15 outlines each type, its subtypes, related routing, and a brief description. Wired anonymous network classification is examined first, followed by wireless anonymous network classification.

Table 15: Network Types

Type	Sub-type	Routing	Description
Wired	Path Topology	Cascade	Fixed path length
		Free	Variable path length
		P2P	Dynamic path length
	Route Scheme	Unicast	One-to-one only
		Multicast	One-to-many
		Broadcast	One-to-all
		Anycast	One-to-one among possible many
	Path Type	Simple	No cycles
Wireless	Topology-based	Reactive	Identity-based, on-demand, high mobility
		Proactive	Identity-based, table-based, low mobility
		Hybrid	Combined reactive/proactive
	Position-based	Reactive	Identity-free, on-demand, high mobility
		Proactive	Identity-free, table-drives, low mobility
		Hybrid	Combined reactive/proactive

4.3.1 Wired.

Wired networks are decomposed into *path topology*, *route scheme*, and *path type* strategies. Each strategy assumes static *a priori* topology knowledge of the anonymous network for the duration of an adversary's attack.

The *Path Topology* routing approaches are *cascade* and *free* route for mixnets [SaP06] or distributed for P2P networks as mentioned in Chapter 2. In a *cascade* network, senders choose from a set of fixed paths through the anonymous network for message transfer. Cascades are unicast and may provide greater anonymity against an adversary who has compromised many nodes but are more vulnerable to blending attacks. Further, cascade networks have lower maximum anonymity [DaR03]. The anonymity set is limited to the number of messages the weakest node in the cascade can handle [DaR03]. In *free* route or P2P networks, senders may choose a route of variable length through the network for message transfer. In *free* route or peer-to-peer networks, senders choose a route of variable length *x* through the anonymous network to transfer

the message to the receiver. The path length L is a random variable conforming to a specific probability distribution. For instance, one strategy might use a geometric path-length distribution [GuF04]. Given the forwarding probability p_f, the randomly chosen path length is a nonnegative number conforming to the geometric distribution

$$P\{L = x\} = (1 - p_f)\, p_f{}^x, x \geq 0. \tag{52}$$

Another strategy uses a uniform path-length distribution [GuF04]. Given the lower bound a and upper bound b, the randomly chosen path length is a nonnegative number between a and b following a uniform distribution

$$P\{L = x\} = \frac{1}{b - a}, a \leq x \leq b \tag{53}$$

Free-route networks have higher maximum anonymity up to a certain path length [DaR03]. The anonymity set is larger because no single node acts as a bottleneck; hence, many nodes handle traffic in parallel as messages traverse the network [DaR03]. Once path length is determined, the path is chosen by randomly selecting intermediate nodes.

The *Route Scheme* is a major factor affecting anonymity. Practically all in-depth research on wired anonymity networks assumes a unicast routing strategy. Exceptions include the DC-Net, P^5, Hordes [LeS02], MAM, and Cashmere [ZhZ05].

Two *Path Type* approaches are *simple* and *complex* [GuF04]. In a *simple* path, no cycles are allowed. Intermediate nodes may only appear once on the path. In a *complex* path, cycles are allowed. In one strategy, the cycles may be disjoint. These cycles share no common nodes. Only intermediate nodes at the starting and ending point of a cycle can appear exactly twice on the path. In another, the cycles are arbitrary. The path begins and ends with the same node but intermediate nodes appear arbitrarily.

4.3.2 Wireless.

The *Wireless* Network Type is decomposed into *topology-based* and *position-based*. *Topology-based* protocols use information about links in the network to perform packet forwarding. *Position-based* routing protocols use geographical node position information to make routing decisions. A mobile wireless node typically broadcasts to neighboring nodes so no route scheme is strictly necessary when classifying anonymous wireless networks. Either routing protocol may be classified as *proactive*, *reactive*, or *hybrid*. *Proactive* protocols periodically exchange control messages to make routing adaptations in the network. The control messages may be sent locally to discover neighbor nodes or more distributed to obtain topology information from all network nodes. Either way, a route is known in advance. *Reactive* protocols do not discover routes in advance but rather attempt to find routes on-demand and routes request packet across the network prior to sending any data. *Hybrid* or "zone" protocols use a mix of both proactive and reactive routing techniques at the network node. No one routing protocol is universally applicable.

4.4 Anonymous Network Taxonomy Results

The cubic taxonomy (CT) can classify state-of-the-art anonymous network protocols. The utility of CT is demonstrated two ways. First, using the three-dimensional (3D) cubic taxonomy, a select few anonymous protocols are compared with all three components. Second, using a two-dimensional (2D) tree taxonomy, over thirty-three anonymous protocols are examined via the Anonymity Property and Network Type components only. It is believed this is the most comprehensive classification of wired

protocol family relationships and first known to capture wireless protocol family relationships. It is also the first graphical synthesized classification of both wired and wireless anonymous networks.

4.4.1 3D Cubic Taxonomy.

A novel 3D cubic taxonomy is developed to classify the desired anonymity properties, presumed adversary capabilities and selected network types inherent in an anonymous communications network. This top-level cubic taxonomy (CT) is shown in Figure 38.

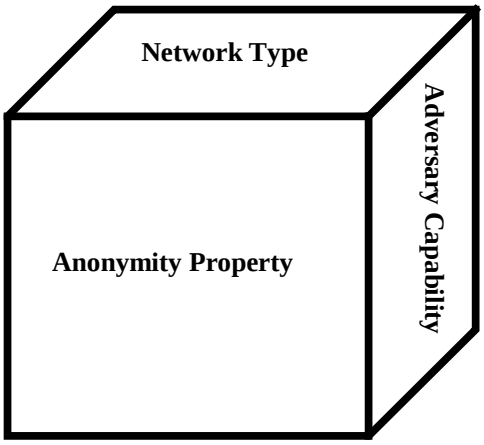


Figure 38: 3D Cubic Taxonomy (Top-Level)

The top-level contains three fundamental components: Anonymity Property, Adversary Capability, and Network Type. Anonymity Property addresses "What information must be hidden?" Hiding identity, relationship, location and/or other items of interest (IOI) from others in the anonymous network is typical. Adversary Capability addresses "From whom do we hide it?" and defines who the assumed adversary is and how strong the threat to the anonymous system is. Network Type addresses "How

hidden must it be?" by defining routing schemes, the transmission medium, network topology, and protocol interdependencies impact on anonymity. These three components are further decomposed as shown in Figure 39.

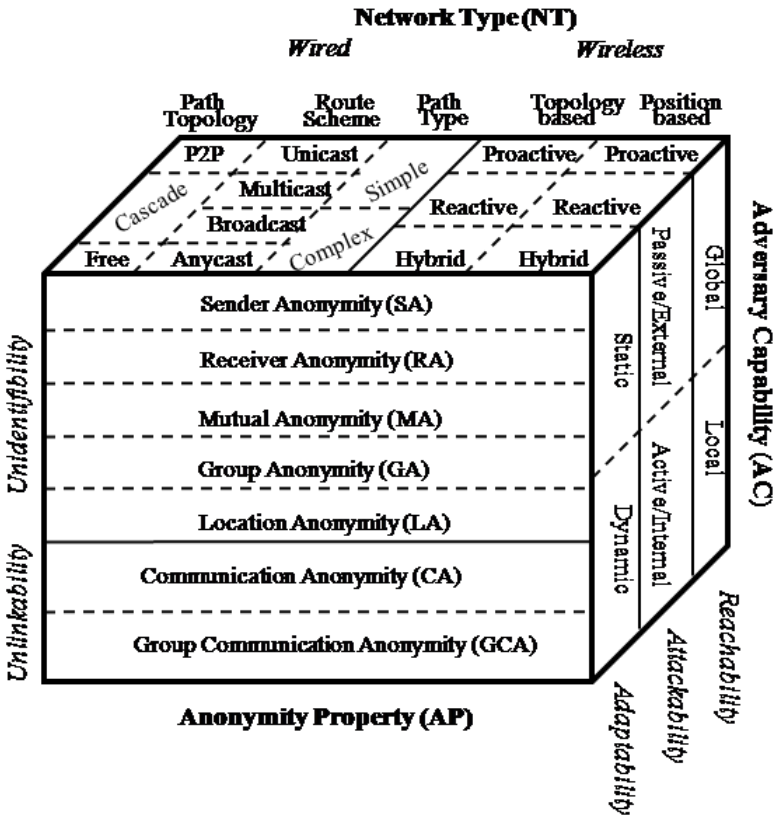


Figure 39: Cubic Taxonomy (CT) Components

At this mid-level, the Anonymity Property is broken down into the abstract *unidentifiability* and *unlinkability* terms. The Adversary Capabilities are broadly categorized as *reachability*, *attackability*, and *adaptability*. Finally, Network Type is either *wired* or *wireless*. These seven sub-components are further decomposed into their twenty-eight (28) "atomic" subcomponents.

The bottom-level consists of seven anonymity properties, six adversary capabilities and five network types decomposable into fifteen network sub-strategies. This is the first known 3D synthesized graphical classification of both wired and wireless anonymous networks.

The purpose of CT is to visually compare different anonymous network protocols and group them into identifiable protocol families. The taxonomy is used to classify a variety of wired and wireless anonymous networks. For instance, DC-Net, Crowds [DiM04, ReR98], and Tor [DiM04, Fra06] anonymous networks are compared in Figure 40.

For AP, each offers SA and RA against specific adversaries; in addition, Tor offers CA. For AC, DC-net assumes a strong passive global threat model whereas Crowds and Tor assume a weaker local adversary threat model. However, the latter two offer some degree of anonymity against an active, dynamic adversary who may control a limited number of collaborating jondos or compromised onion routers as well as selective passive traffic analysis attempts. For NT, all three are wired networks; however Tor employs a free route path topology whereas DC-Net and Crowds are P2P. DC-Net also uses a broadcast route scheme whereas Crowds and Tor use unicast and allow complex path types. Hence, formally analyzing similar anonymous protocols such as Crowds and Tor which offer anonymous web-surfing may prove to be an intriguing investigation. However, if two protocols are conceptually very different such as DC-Net and Tor, then any comparison would be difficult or simply invalid.

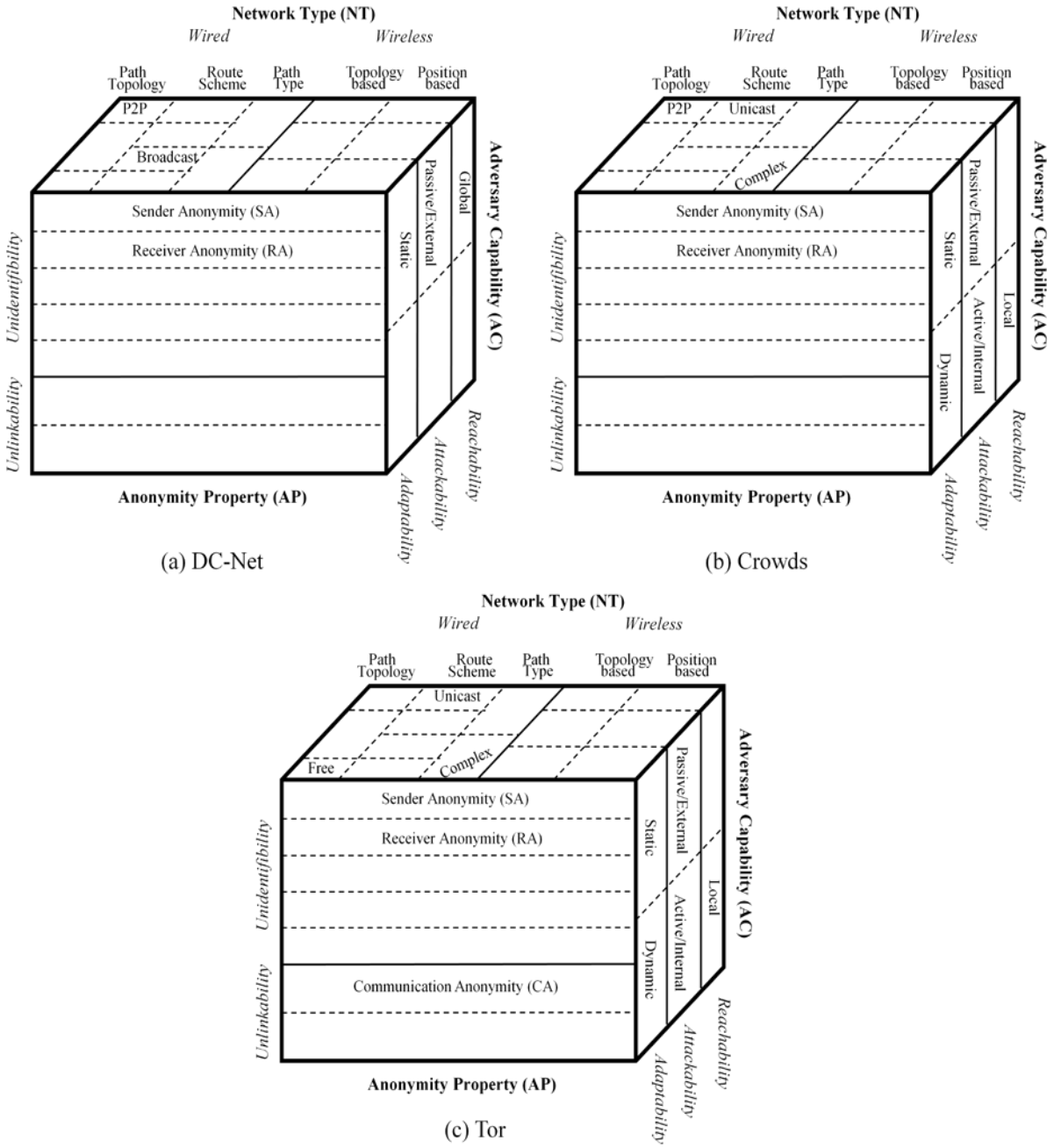


Figure 40: Cubic Taxonomy of Wired Anonymous Protocols

The Secure Distributed Anonymous Routing (SDAR) [BoE04] and Zone-based Anonymous Routing Protocol (ZAP) [WuB05] anonymous network protocols are compared in Figure 41.

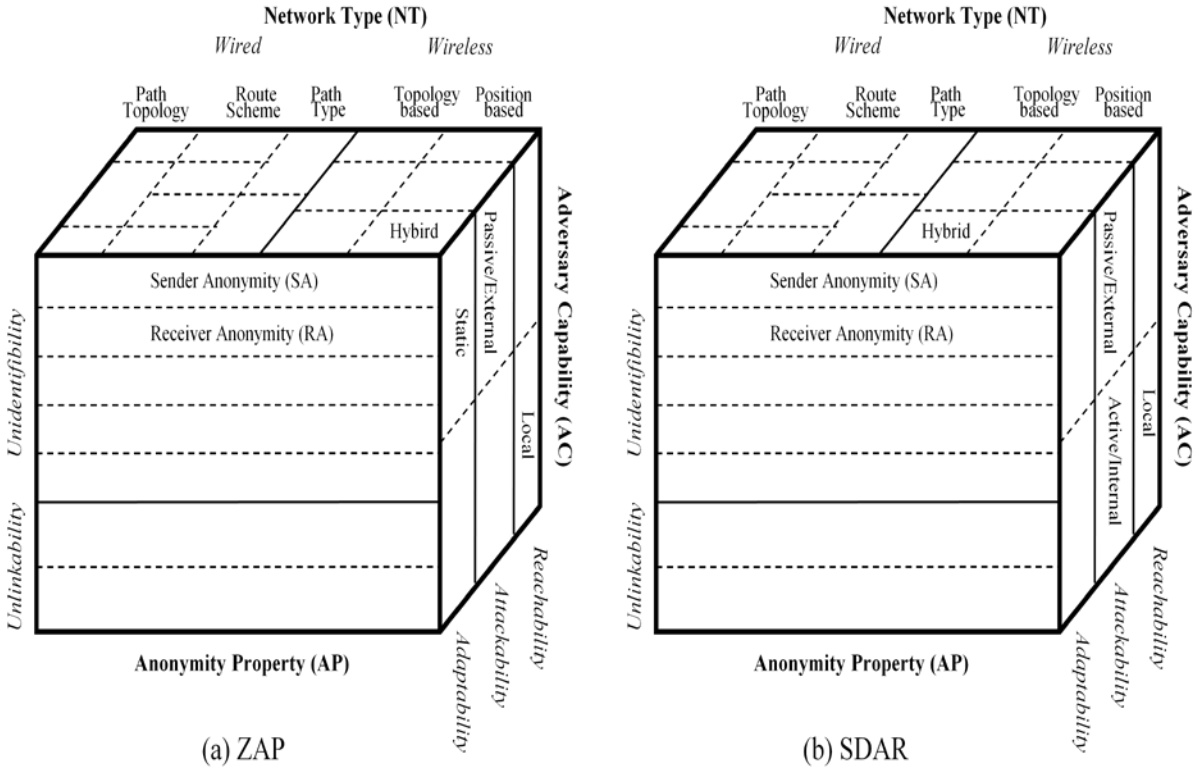


Figure 41: Cubic Taxonomy of Wireless Anonymous Protocols

In terms of NT, both are wireless networks; however ZAP is a hybrid, position-based protocol that uses destination flooding where as SDAR is a hybrid, topology-based protocol that uses multicast. In terms of AC, both assume a local, passive/external adversary; however, adaptability for ZAP may be static with a fixed receiver anonymous zone or dynamic with an adaptive receiver anonymous zone. Attackability may be active/internal for SDAR, but only passive/external for ZAP. In terms of AP, both offer SA and RA. Hence, formally representing these two protocols and/or quantitatively comparing their anonymity preservation and degradation may prove to be fruitful. In the end, a family of anonymous networking protocols may be more closely and rigorously analyzed.

4.4.2 2D Tree Taxonomy.

The 2D tree-based taxonomy is shown in Figure 42.

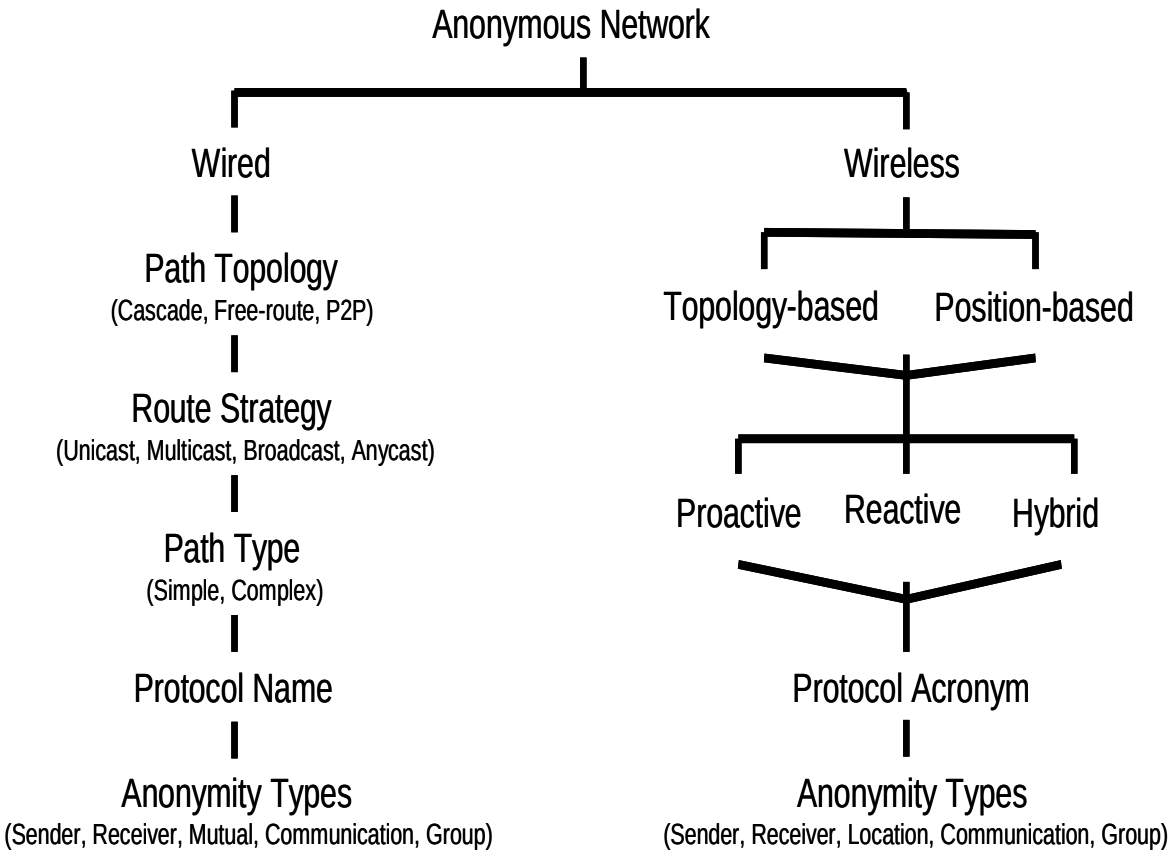


Figure 42: Tree Taxonomy with Anonymity Types

The internal tree structure from the Anonymous Network root node down to Protocol Name and Protocol Acronym nodes correspond to the Network Type classification displayed in Table 15. The leaf nodes represent the Anonymity Types specified in column 2 of Table 13. The overall classification of seventeen wired anonymous network protocols is shown in Figure 43.

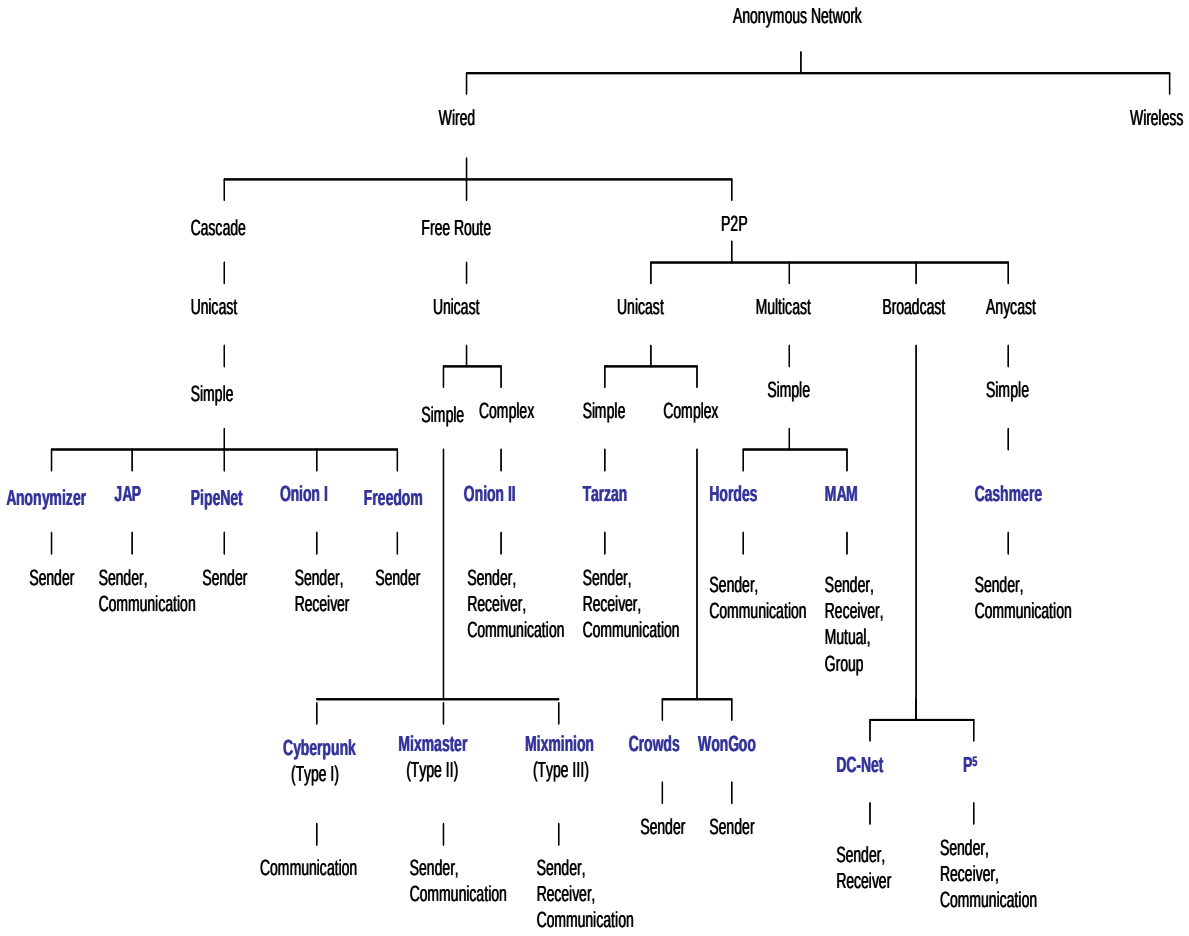


Figure 43: Classification of Wired Anonymous Networks

This taxonomy classifies classic and state-of-the-art wired anonymous networks. It adds path type and routing scheme classification and fills in the previously lacking P2P overall classification. Referring to the specific wired protocols as described in Section 2.3.1, Anonymizer, JAP, Onion-Routing I, PipeNet, and Freedom Network use *cascade* topologies. Onion-Routing II, Cyberpunk, Mixmaster, and Mixminion *free*-route topologies. Tarzan, Crowds, WonGoo, Hordes, MAM, DC-net, P^5, Herbivore, and Cashmere are P2P protocols. Herbivore uses a *broadcast* strategy whereas P^5 employs a *tree broadcast* strategy. Hordes and MAM use a *multicast* strategy. Only Cashmere uses

an *anycast* strategy. All but Onion Routing II, Crowds and WonGoo use a *simple* path type strategy. Crowds and WonGoo allow a *complex* arbitrary cycle path type. PipeNet, Freedom, Crowds, and WonGoo offer sender anonymity only. Onion Routing II, Mixminion, Tarzan, and P^5 offer classical anonymity of sender, receiver and communication anonymity. Herbivore does also if the receiver is inside the anonymous network. This 2D taxonomy is a valid classification of wired anonymous networks since Cyberpunk, Mixmaster, and Mixminion form a single protocol family under the Anonymity Network → Wired → FreeRoute → Unicast → Simple classification. This matches the recent and complementary Anonymity → Mixnet → Freeroute → Asynchronous → Remailer classification [SaP06]. However, this new taxonomy classifies more wired networks such as Cashmere, MAM and WonGoo and classifies P2P networks in addition to classical mixnets.

The overall classification of sixteen wireless anonymous network protocols is shown in Figure 44. This is the first known classification of wireless anonymous networks into protocol families. Referring to Section 2.3.2, AnonDSR, ARM, ODAR, HANOR, AMUR, ASRPAKE, SDAR and MASK are *topology-based* protocols. ANODR, SDDR, ASR, AODPR, AO2P, SAS, ASC, and ZAP are *position-based* protocols. SDAR, MASK and ZAP use the *hybrid* approach whereas the others use a *reactive* approach. All but SDAR, AnonDSR, ARM, HANOR, MASK, and ZAP offer location anonymity. ODAR, ASR, AODPR, MASK, and ASC claim to offer sender, receiver, communications and location anonymity. Only HANOR offers group anonymity.

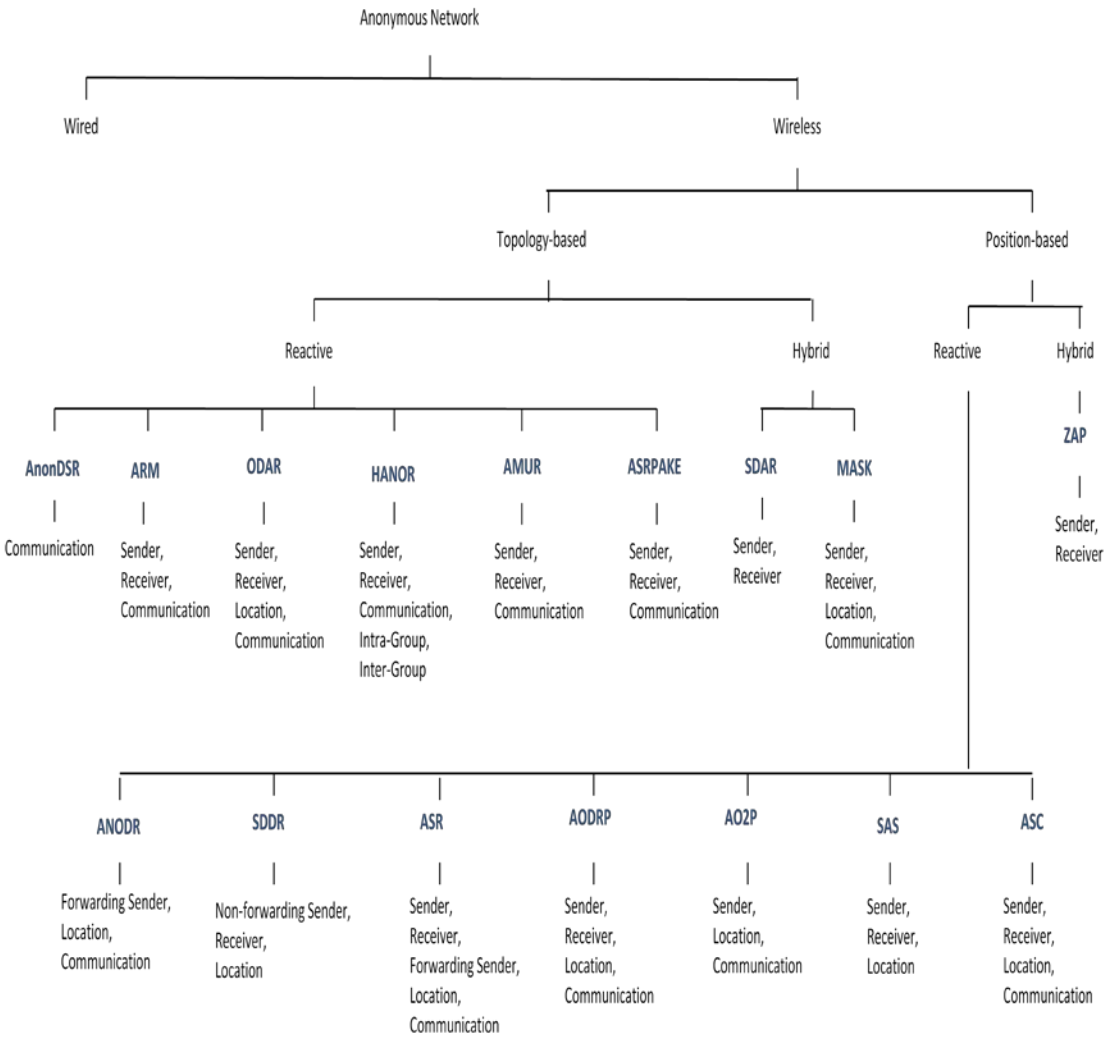


Figure 44: Classification of Wireless Anonymous Networks

The wireless protocol family classification offers a high-level view of the state-of-the-art wireless anonymous networks and corresponding anonymity properties.

4.5 Summary

This chapter describes an innovative CT to facilitate the systematic definition and comprehensive classification of anonymity of wired and wireless anonymous communications networks. The taxonomy considers seven desired anonymity properties,

six assumed adversary capabilities, and fifteen special network types. An expanded cubic anonymity definition is proposed and an assumed adversary capability is described. The wired and wireless network types are further refined. Finally, the cubic and tree-based taxonomies with state-of-the-art existing or proposed anonymous networks is given.

V. Anonymous Metrics Analysis and Results

5.0 Chapter Overview

This chapter presents the results of a synthesized quantified approach on measuring the changes in anonymity levels for a large variety of wired and wireless anonymous networks. This rest of this section is organized as follows. Section 5.1 describes the basic concepts in network and data anonymity. Four basic anonymity metrics used for data and/or network anonymity is covered in Section 5.2. Section 5.3 describes two database and one network data anonymity metric. In Section 5.4, three network-based metrics are explored. A qualitative comparison of all the metrics with respect to applicability, complexity, and generality is described in Section 5.5. Finally, Section 5.6 concludes the chapter and emphasizes the need for more anonymity metrics.

5.1 Anonymity Concepts

The anonymity metrics herein rely on probability and other theories. For clarity, pertinent concepts on network-based and data-based anonymity are reviewed and an intuitive example for each is provided. To ensure continuity with previous work, particular notation for each metric has been preserved whenever possible.

5.1.1 Network-based Metrics.

An example of message senders communicating with receivers over an anonymous network is shown in Figure 45. The set of senders is $S = \{A,B,C\}$ and set of receivers is $R = \{D,E,F\}$. More abstractly, either set may be the anonymity set (*AS*) [PfK00] and both

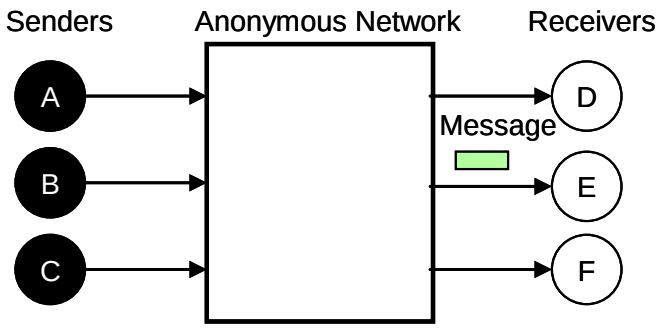


Figure 45: Anonymous Network Example

are sets of agents who perform some specific action. The type of underlying anonymous network often determines which metric is used. The anonymity properties measured in fixed networks include sender, receiver, and communication anonymity. Sender anonymity prevents a particular message from being linked to a particular sender identity. If the attacker believes the message sent to receiver E may be from any sender, then sender anonymity is preserved. Receiver anonymity prevents a particular message from being linked to a particular receiver identity. If the attacker knows that E received the sent message, then receiver anonymity is eliminated. Communication anonymity means a particular message cannot be linked to any sender-receiver pair and no message is linkable to a particular sender-receiver pair. If the attacker does not know the message sender but knows E received the message, the message sender-receiver relationship cannot be definitely established. However, communication anonymity is degraded since the attacker is able to exclude receivers D and F. In this case, the *AS* is the set of sender-receiver pairs ($AS=S \times R$). For mobile networks, the additional anonymity property of location anonymity is quantified to ensure sender, receiver, and communication anonymity. Location anonymity means a particular message is not linkable to any sender or receiver location, motion, route or topology information.

5.1.2 Data-based Metrics.

In privacy-preserving data publishing, sensitive attributes often lead to information leakage. Let table $T = \{t_1, t_2, \ldots t_n\}$ contain a subset $B = \{b_1, b_2, \ldots b_j\}$ of the set of all attributes $A = \{a_1, a_2, \ldots a_z\}$. The value of attribute a_i for tuple t is $t[a_i]$. Table 16 displays a sample network data table T that logs web search queries where $z = 7$, $j = 4$ and $B = \{IP\ Address, Date, Time, Query\}$. The set of sensitive attributes, S, are values that must be protected from an attacker. For instance, the *Query* attribute should be disassociated from the identifying *IP Address* attribute. The other set of attributes

Table 16: Original Network Data Table Example (T)

	IP Address	Date	Time	Query
1	96.234.69.21	2008-10-21	2345	Aids medicine
2	222.154.155.175	2008-10-21	2344	*m*-invariant
3	96.234.68.25	2008-10-20	2342	Cook book
4	96.234.69.21	2008-10-20	2341	Aids medicine
5	222.154.155.175	2008-10-15	2333	*l*-diversity
6	96.234.68.25	2008-10-13	2329	Cook book
7	96.234.68.25	2008-10-09	2327	*t*-closeness

are non-sensitive attributes, $NS = \{Date, Time\}$. A set of non-sensitive attributes that can be linked with external information to de-anonymize one or more agents in the table T constitute a quasi-identifier set such as $QI = \{IP\ Address\}$. Thus, an anonymizing algorithm sanitizes table T to an anonymized table T^* to prevent the attacker from discovering identifying information or relationships. A set of indistinguishable tuples with respect to specific identifying attributes is called an equivalence class, E, and corresponds with the anonymity set, AS, in the previous anonymous network example.

5.2 Basic Metrics

An anonymity metric quantifies how well the anonymization technique hides agent's identities or relationships against a specific attacker. Many of the metrics in the literature expand upon one or more of these four basic metrics.

5.2.1 Anonymity Set Size (ASS).

Anonymity set size (or analogously, *equivalence class set size* for data privacy) is a simple way to measure anonymity in an anonymized table or anonymous network. If the attacker knows the number of agents N prior to an attack (prior to release of the published network data and using background knowledge only) and compromises or eliminates C agents during the attack (after receiving the anonymized table T^*), the anonymity set size $n = N - C$ quantifies the level of anonymity achieved. Figure 46 depicts this metric in terms of sender anonymity.

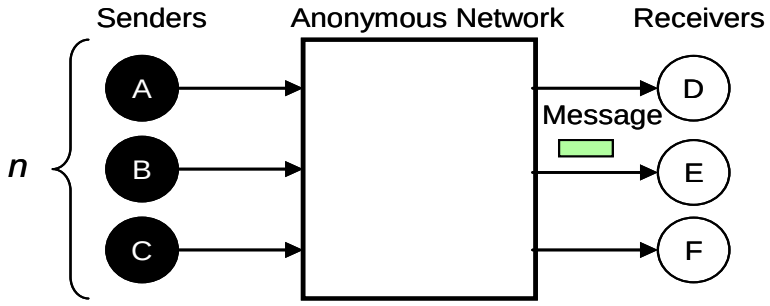


Figure 46: Anonymity Set Size Metric (n). $N = 6$, $C = 3$, $n = 3$.

The attacker's chances of identifying the agent's role of sender or receiver increases (decreases) as n decreases (increases). The attacker is often assumed to be able to distinguish between sender and receiver agents; thus, N may refer to the set of potential senders, receivers, or sender-receiver pairs, instead of the entire set of agents. This metric's levels of anonymity are in Table 17.

Table 17: Anonymity Set Size Levels

Level	Metric Value	
Preserved	$n = N$	$C = 0$
Degraded	$1 < n < N$	$1 \leq C < N - 1$
Eliminated	$n = 1$	$C = N - 1$

If no agents are compromised or eliminated ($C = 0$), then n is unchanged ($n = N$) and anonymity is preserved. If at least one agent is compromised or eliminated ($1 \leq C < N-1$), then *AS* decreases ($n < N$) and anonymity is degraded. The worst case is if $n = N - C = 1$ and anonymity is eliminated.

5.2.2 *k*-anonymity.

If only a minimal set size (k) is required, then the *k-anonymity* metric is used. *k-anonymity* refers to a minimum number of agents or agent pairs the attacker is required to keep in *AS* to preserve anonymity as illustrated in Figure 47. If the attacker believes at least two senders (e.g., A or B) sent the message, then 2-anonymity is achieved.

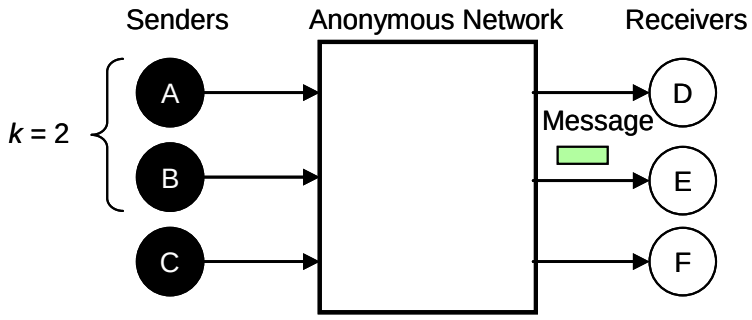


Figure 47: *k*-Anonymity Metric (*k*)

Analogously in the data publishing arena, *k-anonymity* [Swe02] for a table means that for each tuple there are at least k -1 other indistinguishable tuples with respect to a certain set of quasi-identifiers. The resulting generalized anonymity table T^* is in Table 18.

Table 18: Generalized 2-Anonymity Network Data Table (T^*)

	IP Address	Date	Time	Query
1	96.234.69.**	2008-10-2*	234*	Aids medicine
2	96.234.69.**	2008-10-2*	234*	Aids medicine
3	222.154.155.***	2008-10-**	23**	*m*-invariant
4	222.154.155.***	2008-10-**	23**	*l*-diversity
5	96.234.68.2*	2008-10-**	23**	Cook book
6	96.234.68.2*	2008-10-**	23**	Cook book
7	96.234.68.2*	2008-10-**	23**	*t*-closeness

This attempts to unlink agent identifying information between the released and external tables. If the attacker believes two or more agents could have made the query for each of the three equivalence classes, then 2-anonymity is achieved. In this example, three equivalence classes exist with at least two tuples per class. However, the equivalence class with generalized IP address 96.234.69.** has identical *Query* attribute values of "Aids medicine," thereby potentially leaking sensitive information. Hence, it lacks the diversity [MaG06] of the other two. The anonymity levels are indicated in Table 19.

Table 19: *k*-Anonymity Levels

Level	Metric Value
Preserved	$\geq k$
Degraded	$< k$
Eliminated	$k = 1$

If *AS* meets the minimum requirement ($\geq k$) for all messages or equivalence classes, then anonymity is preserved. If it is below the minimum ($< k$) for any given message or in any equivalence class, then anonymity is degraded. If the agent identity or relationship is identified ($k = 1$), then anonymity is eliminated.

5.2.3 Individual Anonymity Degree (IAD).

The *individual anonymity degree* for each agent *i* in *AS* at any point in time assigned by the attacker is characterized by the scale in Figure 48.

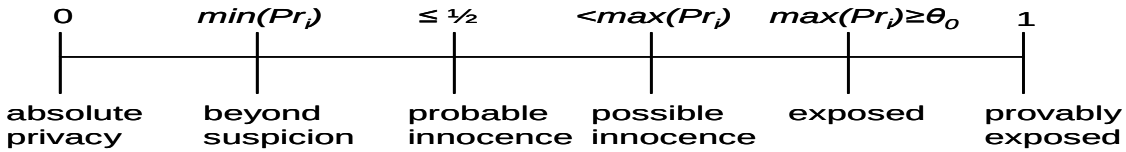


Figure 48: Individual Anonymity Degree Scale

The anonymity degrees range from absolute to none. The top half quantitatively expresses anonymity where $min(\Pr_j)$ is the minimum probability for all agents, $max(\Pr_j)$ is maximum probability for all agents, and θ_0 is some threshold probability. The bottom have qualitatively describes anonymity degree as mentioned in Section 2.4.2.

Consider sender anonymity where $AS = S$, $n = 3$, and $i \in AS$ as shown in Figure 49.

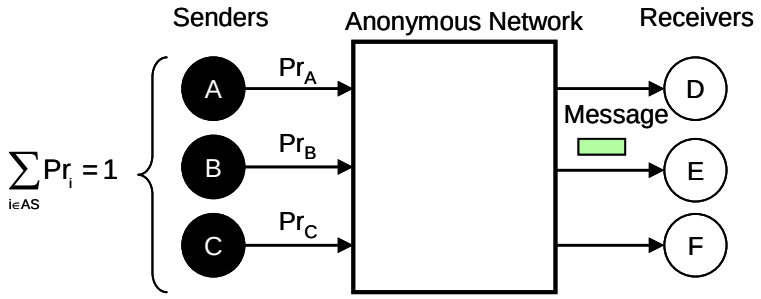


Figure 49: Individual Anonymity Degree Metric ($\sum_{i \in AS} \Pr_i = 1$)

For each agent *i*, the attacker assigns a probability $\Pr_i$ such that $\Pr_i \neq 0$. The probabilities determine where each agent falls on the anonymity degree scale.

On the far left of the scale, *absolute privacy* means agent *i* either never sends any messages or is not in *AS* so $Pr_j = 0$. The next four anonymity levels are depicted in Figure 50. The black arrows indicate which sending agents satisfy the corresponding anonymity definition.

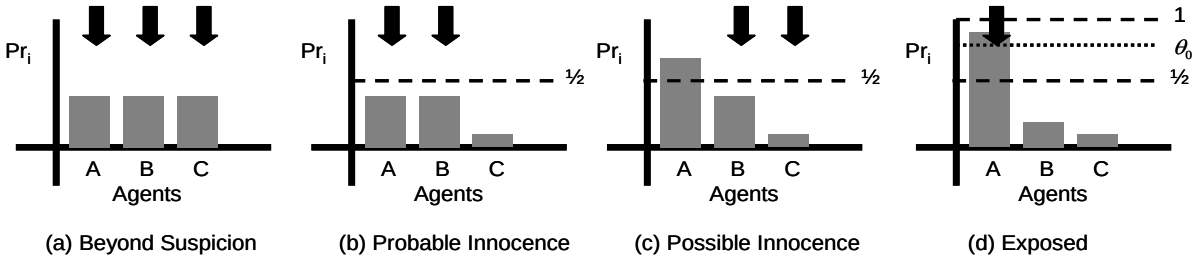


Figure 50: Individual Agent Anonymity Degrees

Beyond suspicion means agent *i* is no more likely to have sent the message than anyone else. In Figure 50(a), this is true of agents A, B, and C since $Pr_i = min(Pr_j) = ⅓, \forall j \in AS$. This is also known as *total, perfect,* or *strongly probabilistic anonymity*.

Probable Innocence means agent *i* is no more likely to have sent the message than not sent the message. In Figure 50(b), agent A and B are this since $Pr_A = Pr_B = 0.45$ but C is *beyond suspicion* since $Pr_C = min(Pr_j) = 0.10$.

Possible Innocence means there is a non-trivial chance that an agent other than *i* sent the message. In Figure 50(c), $Pr_A = max(Pr_j) > ½$ and $Pr_j < Pr_A$. Both agents B and C are *possible innocent*. By strict definition, agent C may also be considered *beyond suspicion*.

Exposed means there is a significant chance that agent *i* is the sender of the message or $Pr_i = max(Pr_j) \geq \theta_0, \forall j \in AS$. As Figure 50(d) shows, agent A is exposed.

Provably Exposed means the attacker knows agent *i* sent the message or $Pr_i = 1$ and $Pr_j = 0, \forall j \in AS, i \neq j$. This metric's anonymity levels are summarized in Table 20.

Table 20: Individual Anonymity Degree Levels

Level	Metric Name	Metric Value
Preserved	*Beyond Suspicion*	$\forall\, i,j\ (\Pr_i = \Pr_j),\ i \neq j$
Degraded	*Probable/Possible Innocence*	$\exists\, i,j\ ((\Pr_i > \Pr_j) \wedge (\Pr_i < \theta_0)),\ i \neq j$
Eliminated	*(Provably) Exposed*	$\exists\, i\ (\theta_0 \leq \Pr_i \leq 1)$

Anonymity is preserved if all agents have equal probability ($\forall\, i,j\ (\Pr_i = \Pr_j),\ i \neq j$) or are *Beyond Suspicion*. If agent probabilities differ ($\exists\, i,j\ ((\Pr_i > \Pr_j) \wedge (\Pr_i < \theta_0)),\ i \neq j$) or one or more agents are deemed innocent, then anonymity is degraded. If any agent ever becomes *Exposed* ($\exists\, i\ (\theta_0 \leq \Pr_i \leq 1)$), then anonymity is eliminated.

5.2.4 Entropy Anonymity Degree.

Entropy anonymity degree [DiC02, SeD02] quantifies the level of uncertainty inherent in a set of data in units of *bits*. The information-theoretic metric(s) measure how random the probability distribution is and considers the global anonymity of the system or table.

Entropy $H(X)$ involves an aggregation of the individual probabilities $\Pr_i$. The attacker's a priori knowledge is $H(X)$ as shown in Section 2.4.3, (1). The attacker's posterior knowledge is measured by the *conditional entropy* $H(X|C)$ as shown in Section 2.4.3, (2).

The higher the entropy, the more uncertain the attacker is about agent identity or relations. On an absolute scale, combining the *anonymity set size n* with *entropy* at any point in time yields the maximum entropy $H_{max} = \log_2(N - C) = \log_2(n)$. The lower bound of $H(X)$ is zero, but anonymity may be unacceptable at some minimum value $H_{min} > 0$. For example, if agent A is exposed ($\Pr_A = \theta_0$) and agents B and C are not ($\Pr_B = \Pr_C$

$= \frac{1-\theta_0}{2}$), then $H_{\min} = -(((1-\theta_0)\log_2 \frac{1-\theta_0}{2}) + (\theta_0 \log_2 \theta_0))$. On a relative scale, $H_0 = H(X)$ is any initial acceptable entropy value prior to a cyber attack ($H_0 \leq H_{\max}$) and $H_1 = H(X|C)$ is the entropy value after a cyber attack. Table 21 shows entropy anonymity levels.

Table 21: Entropy Anonymity Degree Levels

Level	Metric Value
Preserved	$H_0 \leq H_1 \leq H_{\max}$
Degraded	$H_{\min} < H_1 < H_0$
Eliminated	$0 \leq H_1 \leq H_{\min}$

Anonymity is preserved if the attacker's posterior knowledge falls within the acceptable range ($H_0 \leq H_1 \leq H_{\max}$). Anonymity is degraded if the attacker's posterior knowledge is lower than the a priori knowledge but above acceptable levels ($H_{\min} < H_1 < H_0$). Finally, anonymity is eliminated if H_1 falls below acceptable levels ($H_1 \leq H_{\min}$). An extension of entropy is called *normalized entropy anonymity degree* where $d = \frac{H_1}{H_0}$. The anonymity levels for d are shown in Table 22.

Table 22: Normalized Entropy Anonymity Degree Levels

Level	Metric Value
Preserved	$d \geq 1$
Degraded	$0 < d < 1$
Eliminated	$d \approx 0$

If $d \geq 1$, anonymity is preserved, otherwise anonymity is degraded. If $d \approx 0$, anonymity is eliminated.

5.3 Network-based Metrics

Network anonymity metrics measure the change in anonymity of communicating agent's identities or relationships against a specific attacker. Besides the more common anonymity set size and entropy network metrics, other specialized metrics are geared toward specific anonymous communications protocols. Three of these metrics are described next.

5.3.1 Combinatorial Anonymity Degree (CAD).

The *combinatorial anonymity degree* [EdS07] is a complementary system-wide measure based on the permanent of a matrix. The measure reveals the whole communication pattern between senders and receivers in a delay-bounded real-time anonymous mix network and measures communication anonymity shown in Figure 51.

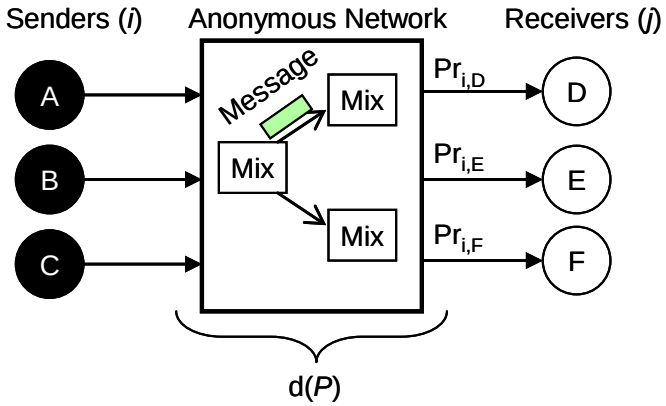


Figure 51: Combinatorial Anonymity Degree Metric (d(*P*))

Instead of assigning individual agent probabilities to sending or receiving agents, link probabilities $Pr_{i,j}$ where $i \in S$ and $j \in R$ are evaluated for the entire anonymous mix

network. The matrix *P* of link probabilities for the sample anonymous network is shown in Figure 52.

$$\boldsymbol{P} = \begin{pmatrix} \mathbf{Pr}_{\mathrm{A,D}} & \mathbf{Pr}_{\mathrm{A,E}} & \mathbf{Pr}_{\mathrm{A,F}} \\ \mathbf{Pr}_{\mathrm{B,D}} & \mathbf{Pr}_{\mathrm{B,E}} & \mathbf{Pr}_{\mathrm{B,F}} \\ \mathbf{Pr}_{\mathrm{C,D}} & \mathbf{Pr}_{\mathrm{C,E}} & \mathbf{Pr}_{\mathrm{C,F}} \end{pmatrix}$$

Figure 52: Attacker Constructed Doubly-Stochastic Matrix *P*

The permanent of the matrix *per*(*P*) is

$$per(P) = \sum_{\pi} \prod_{i=1}^{n} P(i, \pi(i)) \tag{54}$$

where $\pi(i)$ is the a priori probability and *per*(*P*) is bounded by the inequality $n!/n^n \leq per(P) \leq 1$ [Fal05]. The system-wide strength of the anonymous network is

$$\mathrm{d}(P) = \begin{cases} 0 & n=1 \\ \dfrac{\log(\mathrm{per}(P))}{\log(\frac{n!}{n^n})} & n>1 \end{cases}. \tag{55}$$

Thus, anonymity degree is the ratio of the log of the matrix permanent over the log of the lower bound of the a priori probability. The anonymity levels are displayed in Table 23.

Table 23: Combinatorial Anonymity Degree Levels

Level	Metric Value
Preserved	d(*P*) = 1
Degraded	0 < d(*P*) < 1
Eliminated	d(*P*) = 0

When $per(P) = n!/n^n$, perfect anonymity is achieved ($d(P) = 1$) otherwise a lower level of anonymity is achieved ($d(P) < 1$). With only one sender and receiver pair ($n = 1$) in *AS*, no anonymity exists ($d(P) = 0$).

5.3.2 Zone-based Receiver *k*-anonymity (ZRK).

The *zone-based receiver k-anonymity* metric [XiB05] addresses receiver location protection in positioning routing protocols. A sender generates an anonymity zone (AZ) with center x and radius R_{AZ} for each receiver as shown in Figure 53. The forwarding

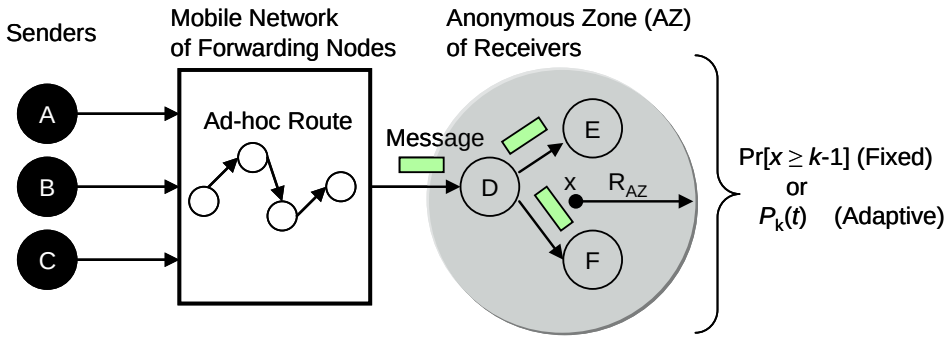


Figure 53: Zone-based Receiver k-Anonymity Metrics (Pr[$n \geq k$-1], $P_k(t)$)

agents in the network deliver the message to a proxy, in this case agent D, who broadcasts the message to all agents in the AZ.

Fixed and adaptive AZ solutions achieve receiver *k-anonymity*. For the fixed AZ, the sending agent uses an initial large-sized AZ ($n_0 >> k$) where n_0 is the initial number of agents in the zone. As time passes, agents move out of the zone and the sender wants to keep k or more agents in the zone. For the adaptive AZ, the sender determines the size of AZ (i.e., k nodes) based on agent density and expands AZ based on agent mobility. The fixed zone-based probability metric level uses a binomial formula to determine the probability of exactly i nodes in the AZ. The probability of preserving *receiver k-anonymity* is

$$P\{n \geq k-1\} = p(1-\sum_{i=1}^{k-1} P\{n = i\}) \tag{56}$$

where p is the probability the receiver agent stays in AZ and $P\{n = i\}$ is the probability that i agents (k-1 other agents) stay in the AZ. The adaptive zone-based probability metric has initial radius R_0 and updates the radius to R_{AZ} to ensure *k-anonymity* after time t_1. Preserving *k-anonymity* requires the sender to linearly expand the radius as

$$R_{AZ}(t_1) = c(t_1 + t_0) - R_0 \tag{57}$$

where $R_0 = \sqrt{\frac{k}{\pi\rho}}$ is the initial radius, $t_0 = -\bar{t}_d \ln(P_k)/k$ is the time when achieving *k-anonymity* is low ($P_k(t) \leq \mu$), t_1 is the time when the radius is expanded, c is the constant R_0/t_0, R_{AZ} (t_1) is the expanded radius at time t_1, ρ is agent density, and $\bar{t}_d$ is the mean agent time in the AZ. Additionally, $P_k(t)$ is the probability that k agents are in AZ after time t. Given pre-defined probability thresholds μ and μ_0, anonymity levels for these metrics are in Table 24.

Table 24: Zone-based Receiver *k*-Anonymity Levels

Level	Metric Value	
	Fixed	Adaptive
Preserved	Pr[$n \geq k$-1] > μ	if $P_k(t) > \mu$, keep R_{AZ}
Degraded	$\mu \geq$ Pr[$n \geq k$-1] > μ_0	if $P_k(t) \leq \mu$, expand R_{AZ}
Eliminated	$\mu_0 \geq$ Pr[$n \geq k$-1]	n/a

5.3.3 Evidence Theory Anonymity (ETA).

Evidence theory anonymity measures communication anonymity in wireless mobile ad-hoc networks. Evidence is measured by the number of detected packets within a given time period. Probability assignments for all packet delivery paths are generated dynamically and overall anonymity quantified in the number of *bits*. Figure 54 illustrates this metric.

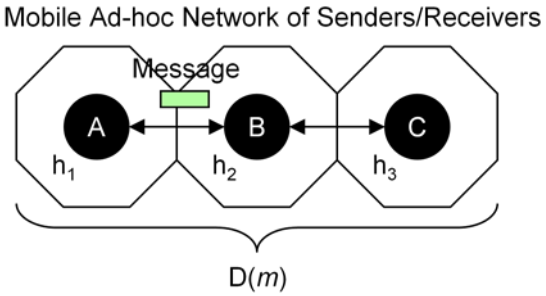


Figure 54: Evidence Theory Anonymity Metric (D(*m*))

The attacker can monitor packets to/from zones h_1, h_2 and h_3 and learn the network topology. For instance, with a time period Δt, the attacker detects exactly one sent packet from the hexagon zone h_2 corresponding to agent *B*. A captured packet is evidence that proves communication between two or more mobile nodes. The attacker computes $w(V)$, $m(V)$, $Bel(V)$, and $Pl(V)$ where U and V are ordered sets of agent communicating paths, $w(V)$ is the quantity of evidence for two communicating mobile agents, $m(V)$ is the probability of an acting communications relation, $Bel(V) = \Sigma_{U|U\subseteq V}\, m(U)$ is a belief measure, and $Pl(V) = \Sigma_{U|U\cap V>0}\, m(U)$ is a plausibility measure such that $Pl(V) \geq Bel(V)$.

To measure uncertainty, the entropy-like measures $E(m) = \Sigma_{V\in F}\ m(V) \log_2 Pl(V)$ and $C(m) = \Sigma_{V\in F}\ m(V) \log_2 Bel(V)$ are proposed where F is a focal element such that $m(V) > 0$. $E(m)$ is not a satisfactory upper bound anonymity measure since it includes irrelevant or conflicting evidence. Instead, the *discord* function $D(m)$ is used as a generalized anonymity measure [Dij06]

$$D(m) = -\sum_{V\in F} m(V)\log_2\left(1-\sum_{U\in F} m(U)\frac{|U-V|}{|U|}\right). \tag{58}$$

The $\sum_{U\in F} m(U)\frac{|U-V|}{|U|}$ term factors out any irrelevant or conflicting evidence. $D(m)$ is a weighted version of belief measure $C(m)$ where $E(m) \leq D(m) \leq C(m)$ holds and measures

average anonymity. Given pre-defined bit thresholds δ and δ_0, evidence theory anonymity levels are listed in Table 25.

Table 25: Evidence Theory Anonymity Metric Levels

Level	Metric Value
Preserved	$D(m) > \delta$
Degraded	$\delta \geq D(m) > \delta_0$
Eliminated	$\delta_0 \geq D(m)$

If $D(m)$ exceeds threshold δ, then communication anonymity is preserved. Anonymity is degraded if $D(m)$ is bounded between δ and δ_0. If it falls at or below δ_0, anonymity is eliminated.

5.4 Data-based Metrics

The data anonymity metrics provide privacy protection of releasable table-based information to third party organizations. The first two address database anonymity and third addresses network data anonymity. All three extend beyond *k*-anonymity and/or entropy anonymity degree.

5.4.1 *l*-diversity.

The *l*-diversity [MaG06] principle is an extension of entropy with the goal of resolving the attribute disclosure limitations of *k*-anonymity. Intuitively, for each equivalence class *E*, the sensitive attribute(s) must have *l* or more well-represented values. Table 26 illustrates a 2-diverse anonymized table. The {96.243.6*.**, 2008-10-**, 23**} equivalence class has diversity of three in the sensitive *Query* attribute with

"Aids medicine", "Cook book" and "*t*-Closeness". The {222.154.155.***, 2008-10-**, 23**} has two diversity with "*m*-invariant" and "*l*-diversity".

Table 26: 2-diverse Network Data Table *T**

	IP Address	Date	Time	Query
1	96.234.6*.**	2008-10-**	23**	Aids medicine
2	96.234.6*.**	2008-10-**	23**	Aids medicine
3	96.234.6*.**	2008-10-**	23**	Cook book
4	96.234.6*.**	2008-10-**	23**	Cook book
5	96.234.6*.**	2008-10-**	23**	*t*-closeness
6	222.154.155.***	2008-10-**	23**	*m*-invariant
7	222.154.155.***	2008-10-**	23**	*l*-diversity

The three metrics are Distinct *l*-diversity, Entropy *l*-diversity, and Recursive (*c*, *l*)-diversity as summarized in Table 27. Like *k*-anonymity, distinct *l*-diversity requires at least *l*-1 different sensitive attribute values in each *E*.

Table 27: *l*-Diversity Levels for Entire *T** Table

Level	Metric Value		
	Distinct	Entropy	Recursive
Preserved	$\geq l$	$H_{\min}(E) \geq \log_2 l$	$r_1 < c(r_l + \ldots + r_m)$
Degraded	$< l$	n/a	n/a
Eliminated	$= 1$	$H_{\min}(E) < \log_2 l$	$r_1 \geq c(r_l + \ldots + r_m)$

The entropy *l*-diversity metric $H(E)$ is:

$$H(E) = -\sum_{s \in S} p(E,s) * \log_2 p(E,s) \tag{59}$$

where *S* is the domain of the sensitive attribute and $p(E,s)$ is the percentage of tuples in *E* with sensitive value *s*. Let $H_{\min}(E)$ denote the minimum entropy for all *E*, this measures the diversity of the entire table *T**. If $H_{\min}(E) \geq \log_2 l$ then diversity is preserved, otherwise diversity is eliminated. However, entropy *l*-diversity is an inadequate measure if attribute values occur too frequently. As an alternative, Recursive (*c*, *l*)-diversity places an upper limit on the occurrences of the most frequent sensitive attributes value,

r_1, within each E. This limit is a c multiple of the sum of the less frequent values or $c(r_l + r_{l+1} + \ldots + r_m)$ where m is the number of values in E and r_i is the number of occurrences of the i^{th} value. For example, if $c = 1$ and $l = 2$ for the first equivalence class in Table 26, then m=3 since *Query* takes on three values. If the sensitive attribute value is "Aids medicine", then r_1=2 and the other occurrences are r_2=2 ("Cook book") and r_3=1 ("*t*-Closeness"); hence, (1,2)-diversity is preserved since $2 < 6$. However, if "Aids medicine" replaced the "Cook book" values, then m=2, r_1=4, and r_2=1. Since $4 < 1$ is false, (1,2)-diversity would not be preserved. The entire T^* table is recursive if each and every E is recursive.

5.4.2 *t*-Closeness.

To overcome attribute disclosure issues in *l*-diversity, the *t*-closeness data privacy metric [LiL07] takes into account the semantic relationships among the attributes values. In particular, it constrains the difference between sensitive attribute distributions in each E and entire table T^* to be no more the threshold t. This makes it more difficult for the attacker to gain knowledge from the released anonymized table T^*.

This measure is derived from the well-researched transportation problem of transforming one distribution to another with the least amount of total work. Given two discrete distributions $\mathbf{P} = \{p_1, p_2, \ldots, p_m\}$ and $\mathbf{Q} = \{q_1, q_2, \ldots, q_m\}$, the distance, D, between the distributions is:

$$\mathrm{D}[\mathbf{P}, \mathbf{Q}] = \sum_{i=1}^{m} \sum_{j=1}^{m} d_{ij} f_{ij} \tag{60}$$

where d_{ij} is the distance between p_i and q_j and f_{ij} is the minimal work flow of mass from p_i to q_j. The metric differs depending on whether the sensitive attribute is numerical or categorical. If numerical, then $r_i = p_i - q_i$ and distance metric is:

$$\mathrm{D}[\mathbf{P},\mathbf{Q}] = \frac{1}{m-1}\sum_{i=1}^{i=m} \left|\sum_{j=1}^{j=i} r_j\right|. \tag{61}$$

If categorical, the equal distance metric is:

$$\mathrm{D}[\mathbf{P},\mathbf{Q}] = -\sum_{p_i<q_i} (p_i - q_i). \tag{62}$$

Whichever metric is used, Table 28 shows *t*-closeness levels.

Table 28: *t*-Closeness Earth Mover's Distance (EMD) Levels

Level	Metric Value
Preserved	$0 \le \mathrm{D}[\mathbf{P},\mathbf{Q}] \le t$
Degraded	n/a
Eliminated	$t < \mathrm{D}[\mathbf{P},\mathbf{Q}] \le 1$

t-closeness is preserved if the attacker's posterior knowledge falls within the acceptable range ($\mathrm{D}[\mathbf{P},\mathbf{Q}] \le t$) and is eliminated if $\mathrm{D}[\mathbf{P},\mathbf{Q}]$ goes above *t*. The main advantage of *t*-closeness is, unlike *l*-diversity, it can measure anonymization techniques other than generalization and suppression. Another metric related to *t*-closeness but which further constraints the variability of the sensitive attribute values to be *m* or greater is *m*-invariance [XiT07]. It accounts for the anonymity of dynamic and re-releasable datasets as opposed to static, one-time releasable datasets.

5.4.3 L1 Similarity.

L1 Similarity [CoW08] quantifies anonymity by computing the difference between an anonymized object, *X*, and unanonymized object, *Y*. Both objects *X* and *Y* have

extractable distributional features. For example, object X may be a *k*-anonymous, *l*-diverse, or *t*-closeness network table *T** and object *Y* is the known universe of all network data tables. The attacker wants to compare feature distributions and reveal the identity of the anonymized object. This information theoretic metric, $sim(X,Y)$, is the maximum L1 distance minus the sum of the absolute differences or

$$sim(X,Y) = 2 - \sum_{z \in X \cup Y} | P(X = z) - P(Y = z) |. \tag{63}$$

The anonymity levels of the metrics are summarized in Table 29.

Table 29: L1 Similarity Levels

Level	Metric Value	Distributions
Preserved	$sim(X,Y) = 2$	Identical
Degraded	$sim_{\min} < sim(X,Y) < 2$	Different
Eliminated	$0 \le sim(X,Y) \le sim_{\min}$	Disjoint

Anonymity is preserved if the objects have identical distributions and the maximum value is obtained, $sim(X,Y) = 2$. Hence, the attacker is unable to gain additional knowledge from the released anonymized network data table. Anonymity is eliminated if the objects have nearly disjoint distributions and the attacker gains complete or substantial knowledge of identities and relationships beyond some acceptable threshold $sim_{\min}$. More realistically, the two distributions are likely to be different allowing the attacker to gain some additional knowledge. And this similarity metric quantifies exactly how similar or anonymous the network data table is and allows the comparison of various anonymization techniques on the same original network data table.

5.5 Metric Comparison

This section provides a high-level comparison of the metrics in terms of applicability, complexity, and generality. The definitions of each of the terms are reviewed and the metrics are evaluated.

The metric applicability may be data, network, or any. A data metric measures content privacy in one-time or repeated releasable datasets. The anonymization technique is usually an algorithmic sanitization of data through generalization and/or suppression. A network metric focuses on communications privacy over wired or wireless networks. Randomization is the most common technique employed to make traffic patterns more indistinguishable. Some metrics may apply to both data and communications privacy and use a variety of anonymity techniques. Table 30 lists the applicability definition.

Table 30: Applicability Definition

Value	Privacy Protected	Anonymity Technique
Data	Data Privacy for network/other domain releasble datasets	Generalization (Algorithmic Sanitization)
Network	Communciations Privacy over fixed or wireless networks	Randomization (Network Routing Perturbation)
Any	Data/Communications Privacy	Generalization/Supression/Randomization

The metric complexity may be low, medium, or high. If low, the metric is a simple integer value. If medium, individual or aggregated probabilities are computed. If high, one or more functions are computed to arrive at the anonymity measure. Table 31 lists the complexity definition.

Table 31: Complexity Definition

Value	Description
Low	An integer-valued metric
Medium	Involves assigning multiple probabilities and/or calculating an overall anonymity value
High	Requires computation of multiple functions

The metric generality is low, medium, or high. If low, the metric either is or has been efficiently applied to real data or network anonymity research. However, it may be protocol dependent and not be useful elsewhere. If high, it is abstract enough to be used across multiple domains. If medium, a trade-off between utility and mathematical rigor has been made. The generality definition is revealed in Table 32.

Table 32: Generality Definition

Value	Description
Low	Practical and efficient but limited to specific network logs or anonymous protocols
Medium	Balanced trade-off between practicality and mathematical rigor
High	Theoretically sound and useful for both data and communications privacy across multiple domains

A high-level qualitative assessment of the applicability, complexity, and generality of the anonymity metrics is in Table 33.

Table 33: Comparison of Anonymity Metrics

Metric	Applicability	Complexity	Generality
ASS	Any	Low	High
k-Anonymity	Any	Low	High
Entropy	Any	Medium	High
l-Diversity	Data	Medium	Medium
t-Closeness	Data	High	Medium
L1 Similarity	Data	Medium	High
IAG	Network	Medium	Medium
CAD	Network	High	Medium
ZRK	Network	High	Medium
ETA	Network	High	High

This table should spark much discussion among researchers and organizations interested in measuring anonymity levels in their own networks and protocols. Metrics with “any” applicability are anonymity set size, *k*-anonymity, and entropy. Only one

metric, L1 similarity, focused exclusively on network data applicability. With high generality, this may be a good candidate metric for further exploration and comparison of network data anonymization techniques. Interestingly, the metrics with a high computational complexity tend to also decrease in generality. What this may suggest is a more precise metric for each specific network data anonymization technique may be required. This underscores the fact that more network anonymity metrics are required.

5.6 Summary

This chapter comprehensively looks at ways to quantify anonymity. It conveyed, in a creative and consistent manner, state-of-the-art metrics to analyze the preservation, degradation, and elimination of anonymity relevant in discovering more network data anonymization specific metrics. First, the terminology and instructive examples were given for both data and network anonymity. Second, four common anonymity metrics of anonymity set size, *k*-anonymity, individual anonymity degree, and entropy anonymity were discussed. Third, the *l*-diversity, *t*-closeness, and L1 similarity data anonymization metrics were highlighted. It is believe that, the latter similarity metric is the only known network data specific measure. Fourth, the specialized network anonymity metrics of combinatorial degree, zone-based receiver *k*-anonymity, and evidence theory anonymity were covered. Last but not least, a macro-level comparison of the applicability, complexity, and generality of each metric was given. The most prevalent metrics used for both data and network anonymization techniques are low in complexity and high in generality. It may possible that multiple metrics are necessary for different network data anonymization techniques to give assurances of preserving privacy; thus, the search for

an elusive general, practical metric to compare various techniques continues. Nonetheless, knowing the available metrics and understanding the subtle changes in anonymity levels is essential for any organization determined to better defend against data and network attacks through cross-organizational network data sharing and message communications.

VI. Formal Anonymity Framework Analysis and Results

6.0 Chapter Overview

This chapter presents an innovative, intuitive Possibilistic Anonymity Logical Model (PALM) to rigorously reason about how an adversary can lower the information assurance of a system by degrading anonymity. The model is sufficiently expressive to allow a variety of anonymity definitions or anonymity properties to be expressed and proved for an anonymous network example.

The rest of the chapter is organized as follows. The proposed PALM model is explained in Section 6.1. Section 6.2 demonstrates the utility of the PALM model with a simple and expanded sender anonymity example. Model limitations are highlighted in Section 6.3. Finally, Section 6.4 concludes the chapter.

6.1 Created Mathematical Model

With the aim to preserve privacy over a communications network, a plethora of anonymous protocols have been proposed along with many empirical investigations into specific adversarial attacks over those networks. However, few formal methods have been developed and applied to anonymous systems with the goal of modeling how an adversary reasons about anonymity. Indeed, many analyses assume a passive, global adversary but fail to provide a rigorous approach for defining and modeling anonymity concepts to ensure information and data assurance as is customary when formally proving other security aspects of a system. Hence, this research proposes the Possibilistic

Anonymity Logical Model (PALM) for capturing the knowledge and reasoning ability of an adversary in an anonymous network.

Anonymous systems and properties may be expressed using the modal logic syntax and semantics as mentioned in Sections 2.5 through 2.8. For instance, if a passive, global adversary attempts to degrade anonymity in a multi-agent system, determining the possibility that a particular agent in a set of agents could have performed an action, such as sending a message, is of interest. The adversary wants to reduce the set of possible senders to the fewest number while the anonymous system wants to thwart the adversary from doing so. Modal concepts may prove useful in constructing a meaningful definition of anonymity for more advanced models.

6.1.1 PALM Model

The Possibilistic Anonymity Logical Model (PALM) is a formalism for capturing the knowledge and reasoning ability of an adversary in an anonymous network. PALM focuses on the four Halpern and O'Neill logical possibilistic anonymity definitions (*minimal, up to, total* and *k-anonymity*) in Table 11. Syntactically, PALM adds a unary possible operator, P_j, to $KT45^n$ modal logic and four new axiomatic anonymity formulas. Semantically, PALM assumes connectivity and best-case or worst-case Kripke possible world structures for a single adversary. The anonymity rules are shown in Table 34. In the first rule, the anonymity set is denoted as I_A. Also, *i* is any agent and *j* is the adversary. The last rule precludes the adversary from gaining knowledge directly from an honest agent. Subsequent models are listed in Table 35.

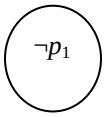
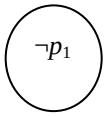
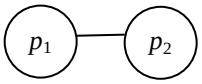
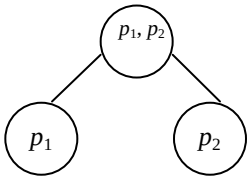
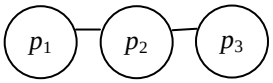
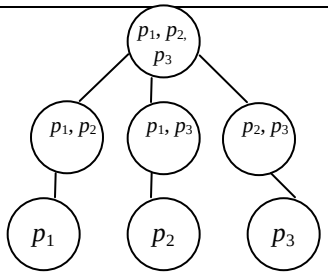
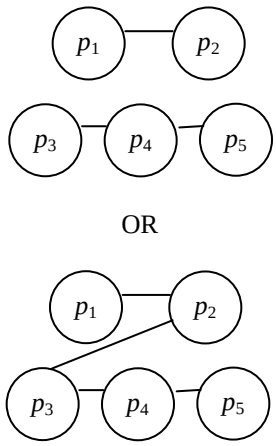
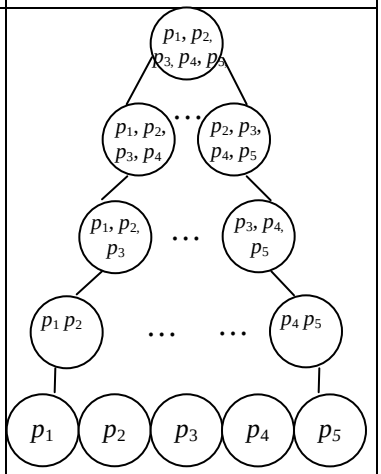
Table 34: Anonymity Rules

Formula	Meaning
$C_G(\bigvee_{\{\lvert I_A\rvert=k, i\in I_A, i\neq j\}} p_i)$	At least one agent sends a dummy message.
$\bigwedge_{i\neq j} C_G(\neg p_i \rightarrow P_j p_i)$	If an agent sends a real message, then the adversary thinks it is possible the agent sent a dummy message.
$\bigwedge_{i\neq j} C_G(p_i \rightarrow \neg K_j p_i)$	If an agent sends a dummy message, then the adversary does not know this.
$\forall i\ C_G(\neg K_i p_i \wedge P_i p_i)$	No agent knows their own sent message type (dummy or real).

6.2 Application of PALM Model

The utility of PALM is demonstrated using a five scenarios that formally (1) prove the validity of each possibilistic anonymity definition and (2) captures the adversary epistemic and nondeterministic reasoning ability about anonymity in multi-agent systems. To determine if these anonymity formulas are well-formed and able to be semantically captured, the $KT45^n$ modal logic system and rules are used in a simple and then expanded message-sender mystery example.

Table 35: PALM Anonymity Formulas and Semantic Models
(*n*=number of agents, *k*=anonymity set, *r*=number of real messages, and *d*=number of dummy messages)

Scenario Parameters	Anonymity Rules/ Formulas (Γ)	Best Case Model (*k* and *r* known)	Worst Case Model (*k* and $r \geq 1$ known)
I. No anonymity $n = 2$, $k = 1$, $r = 1$, $d = 0$	$C_G(p_1 \vee \neg p_1)$ $C_G(\neg p_1 \rightarrow K_j \neg p_1)$ $C_G(P_j p_1)$ $C_G(\neg K_1 p_1 \wedge P_1 p_1)$		
II. Minimal & III. Total $n = 3$, $k = 2$, $r \leq 1$, $1 \leq d \leq 2$	$C_G(p_1 \vee p_2)$ $C_G(p_1 \rightarrow \neg K_j p_1)$ $C_G(p_2 \rightarrow \neg K_j p_2)$ $C_G(\neg p_1 \rightarrow P_j p_1)$ $C_G(\neg p_2 \rightarrow P_j p_2)$ $C_G(\neg K_1 p_1 \wedge P_1 p_1)$ $C_G(\neg K_2 p_2 \wedge P_2 p_2)$		
IV. Up-to $\|I_A\|$ $n = 4$, $k = 3$, $r \leq 2$, $1 \leq d \leq 3$	$C_G(p_1 \vee p_2 \vee p_3)$ $C_G(p_i \rightarrow \neg K_j p_i)$ $i \neq j$ $C_G(\neg p_i \rightarrow P_j p_i)$ $i \neq j$ $C_G(\neg K_i p_i \wedge P_i p_i)$ $i \neq j$		
V. *k'*- to *k*-anonymity $n = 6$, $k = 5$, $r = 2$, $d = 3$, $k' = \lfloor k/r \rfloor$	$C_G(p_1 \vee p_2)$ $C_G(p_3 \vee p_4 \vee p_5)$ $C_G(p_i \rightarrow \neg K_j p_i)$ $i \neq j$ $C_G(\neg p_i \rightarrow P_j p_i)$ $i \neq j$ $C_G(\neg K_i p_i \wedge P_i p_i)$ $i \neq j$ OR $C_G(p_1 \vee p_2 \vee \ldots \vee p_5)$ $C_G(p_i \rightarrow \neg K_j p_i)$ $i \neq j$ $C_G(\neg p_i \rightarrow P_j p_i)$ $i \neq j$ $C_G(\neg K_i p_i \wedge P_i p_i)$ $i \neq j$		
Any $k = n\text{-}1$, $r \leq k\text{-}1$, $d = k - r$	$C_G(p_1 \vee \ldots \vee p_k)$ $C_G(p_i \rightarrow \neg K_j p_i)$ $i \neq j$ $C_G(\neg p_i \rightarrow P_j p_i)$ $i \neq j$ $C_G(\neg K_i p_i \wedge P_i p_i)$ $i \neq j$	k possible worlds	2^k-1 possible worlds

6.2.1 Simple Example.

This is a variation of the wise-men puzzle [HuR04]. There are two message sending agents on an anonymous network. The first is an honest agent. The second is an inquisitive adversary. The attack is an intersection attack of possibilities. There are two dummy messages and one real message. The real messages contain identifying information. The dummy messages obscure an agent's traffic sending patterns. Messages may be received in three different ways: DD, DR, and RD where D = dummy and R = real and the 1st letter is the message sent by the honest agent while the 2nd letter is the message sent by the adversary. RR is not possible since only one real message exists. The messages are randomly assigned to each agent but neither agent knows their own message type. Each sends their message to the other agent. The receiving agents know the received message type. Suppose the adversary asks the honest agent "Did you send a real message?" The honest agent truthfully says "I don't know". Now the adversary knows that he himself sent a dummy message. "I don't know" allows the adversary to rule out DR. If the honest agent received an R message from the adversary, he would have said "No" instead of "I don't know" since DR would have been the only way this could have been occurred. This leaves DD and RD; hence, the adversary knows he sent a dummy message. Thus, an adversary learns from and reasons with knowledge gained from the honest agent.

Formally, let $A = \{1, 2\}$ be the agents, group $G = A$, and agent $j = 2$ be the adversary. Let p_i mean "agent i sent dummy message D"; hence, $\neg p_i$ means "agent i sent real message R". The adversary knowledge and reasoning ability is expressed as logic formulas proceeded by the C_G operator. Thus, a single, global, and active adversary is

assumed. The first anonymity rule $C_G(p_1 \vee p_2)$ means at least one agent will send a dummy message; otherwise, no anonymity exists. The second set of rules $C_G(p_1 \rightarrow K_2 p_1)$ and $C_G(\neg p_1 \rightarrow K_2 \neg p_1)$ indicate the adversary knows the received message type. Analogously, the third set of rules $C_G(p_2 \rightarrow K_1 p_2)$ and $C_G(\neg p_2 \rightarrow K_1 \neg p_2)$ mean the honest agent also knows the received message type. Finally, the last rule $C_G(\neg K_1 p_1 \wedge \neg K_1 \neg p_1)$ represents the honest agents response of "I don't know" my sent message type.

Let $\Gamma = \{C_G(p_1 \vee p_2),\ C_G(p_1 \rightarrow K_2 p_1),\ C_G(\neg p_1 \rightarrow K_2 \neg p_1),\ C_G(p_2 \rightarrow K_1 p_2),\ C_G(\neg p_2 \rightarrow K_1 \neg p_2)\}$ be the initial common knowledge. Let $B = \{C_G(\neg K_1 p_1 \wedge \neg K_1 \neg p_1)\}$ be the additional knowledge the adversary learns from the honest agent. The next step is to prove adversary j knows about the dummy message or $K_j p_2 = K_2 p_2$. Thus, $\Gamma, B \vdash K_2 p_2$.

1	$C_G(p_1 \vee p_2)$	Premise (Γ)
2	$C_G(p_1 \rightarrow K_2 p_1)$	Premise (Γ)
3	$C_G(\neg p_1 \vee K_2 \neg p_1)$	Premise (Γ)
4	$C_G(p_2 \rightarrow K_1 p_2)$	Premise (Γ)
5	$C_G(\neg p_2 \vee K_1 \neg p_2)$	Premise (Γ)
6	$C_G(\neg K_1 p_1 \vee \neg K_1 \neg p_1)$	Premise (B)
7	C_G	
8	$\neg p_2$	Assume
9	$\neg p_2 \vee K_1 \neg p_2$	$C_G e$ 5
10	$K_1 \neg p_2$	$\rightarrow$ e 8,9 (Modus Ponens)
11	K_1	
12	$\neg p_2$	$K_1 e$ 10
13	$p_1 \vee p_2$	$C_G e$ 1
14	p_1	$\vee\ e_1$ 12,13
15	$K_1 p_1$	$K_2 i$ 11–14
16	$\neg K_1 p_1 \wedge \neg K_1 \neg p_1$	$C_G e$ 6
17	$\neg K_1 p_1$	$\wedge e_1$ 16
18	$\perp$	$\neg$e 15,17
19	$\neg\neg p_2$	$\neg i$ 8–18
20	p_2	$\neg\neg e$ 19
21	$C_G p_2$	$C_G i$ 7–20
22	$E_G p_2$	CE 21
23	$K_2 p_2$	EK_2 22

Hence, the adversary knows about the dummy message.

6.2.2 Expanded Example.

Assume there are *n* logically omniscient agents, *n*-1 honest agents and one inquisitive adversary, on an anonymous network. It is common knowledge that there are *k* sending agents where $1 \leq k \leq n$, zero or more real messages and at most *k* dummy messages. It is distributed knowledge that up to *r* real messages are assigned to the *k* agents where $r \leq k$-1. The messages are pseudo-randomly assigned one message per agent such that at most *r* real messages exist. Neither an agent nor the adversary can distinguish between a real or dummy message. Thus, *k* agents send messages, no more than *r* agents send a real message and $d = k - r$ agents send a dummy message over the anonymous network. Obviously, a larger *d* enhances sender anonymity. Each agent sends their respective message. However, the receiver agents do not know the received message type. The adversary must rely on the other agent's responses, if any, to gain more knowledge and degrade anonymity.

Under the current circumstances, if the adversary repeatedly asks the agents simultaneously 'Do you know if you sent a real message?', all *k* agents will repeatedly answer 'no'. The adversary may also ask "Did you send a message?" to determine anonymity set size *k*. In the best case, the adversary knows the number of possible real messages (i.e., *r* value(s)) and is able to reason with minimal knowledge (least possibilities). In the worst case, the adversary only knows that zero or more real messages are sent (i.e., $0 \leq r \leq k$-1) and may have to reason with maximum knowledge (most possibilities). In either case, the adversary builds a $KT45^n$ semantic PALM model and reasons about who sent real messages. Therefore, sender anonymity is subsequently investigated to validate the different degrees of *minimal*, *total*, *up to* $|I_A|$ and *k*-anonymity

formulas. The following five $KT45^n$ semantic model scenarios are used to prove the anonymity formulas:

Scenario I: No anonymity, Worst Case $(n = 2,\ k = 1,\ r = 1,\ d = 0)$

Scenario II: *Minimal*, Best Case $(n = 3,\ k = 2,\ r = 1,\ d = 1)$

Scenario III: *Total*, Worst Case $(n = 3,\ k = 2,\ r \le 1,\ 1 \le d \le 2)$

Scenario IV: *Up to* $|I_A|$, Worst Case $(n = 4,\ k = 3,\ r \le 2,\ 1 \le d \le 3)$

Scenario V: *k*-anonymity, Best Case $(n = 6,\ k = 5,\ r = 2,\ d = 3)$

A simplifying assumption is connectedness. Since the truth of modal properties at a world *x* in a Kripke model in $KT45^n$ depends only on worlds reachable from *x*, only connected graphs are considered to avoid concerns about definable properties of non-connected possible worlds [DaO05]. This corresponds to the Dolev-Yao model [DoY83] where all messages go through the adversary. In the best case(s), models are considered where only one or two binary equivalence relations exist for each of the *k* possible worlds *x*. In the worst case, models are considered where each 2^k-1 possible worlds *x* is reachable from the root and has one or more binary equivalence relation(s). The adversary's knowledge (K_j) and reasoning ability (equivalence relation, R_j) in the anonymous environment are the primary focus. In all models, the adversary *j* does not send any messages and only attempts to use logic to discover who sent real messages to identify sender(s) identity. Furthermore, p_i and $\neg p_i$ have the same meaning as the simple example. p_i means "agent *i* sent dummy message D". Table 35 summarizes the scenario

anonymity parameters, anonymity rules or formulas, and best and worst case semantic models.

In Scenario V, the best case model depends upon how well the adversary partitions the agents into anonymity sets or I_A's. Hence, the agents would enjoy *k*'- to *k*-anonymity where *k*' is the floor of the ratio of anonymous agents to real messages ($\lfloor k/r \rfloor$) if $r \neq 0$. This is only significant if the adversary is able to subdivide the anonymity set of agents into smaller anonymity sets based on previous knowledge or new knowledge gained from observing message traffic patterns and/or logical reasoning from the honest agent responses.

6.2.2.1 Scenario I: No Anonymity.

In this two agent (n = 2) scenario, only a single agent (k = 1) sends a single real message (r = 1) and no agents send dummy messages (d = 0); hence, no anonymity exists after the message is sent. However, before the agent sends the real message, as far as the adversary knows the agent may send a dummy or real message or $C_G(p_1 \vee \neg p_1)$ and also thinks it is possible for the agent to send a dummy message or $C_G(P_j p_1)$. Even in this simple model, the inability to distinguish between "before" and "after" is self-evident. But it is common knowledge the agent does not know the message type $C_G(\neg K_1 p_1 \wedge P_1 p_1)$. After the message is sent, the adversary asks the agent if a message was sent. The agent must say "Yes". Since no dummy message is sent (only a real one), it is now common knowledge or $C_G(\neg p_1)$. Of course, one could argue that it is always common knowledge since $C_G(\neg p_1 \rightarrow K_j \neg p_1)$.

Let $A = \{1, 2\}$ where $n = |A| = 2$ and adversary j and agent(s) $i \in A$, $G = A$ and $P(\text{Atoms}) = \{p_1\}$, then the formal KT45^2 model $\mathfrak{M} = (W, (R_i)_{i \in A}, L)$ is $W = \{x\}$; $R_j(x,x)$; $L(x) = \{p_1\}$. The graphical PALM model is shown in Figure 55 below.

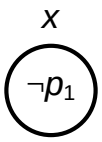


Figure 55: Scenario I PALM Model (KT45^n, n=2)

Only one possible world x exists. This world is where the agent sends a real message or $\neg p_1$. The model assumes a reflexive accessibility relation for the adversary or $R_j(x,x)$. These reflexive relations are assumed and not listed for the subsequent models. The varying degrees of knowledge are listed in Table 36.

Table 36: Scenario I Satisfied Formulas (ϕ) (Adversary Knowledge)

Op	x
p	$\neg p_1$
$\wedge$	$\neg p_1 \wedge \neg p_1$
$\vee$	$p_1 \vee \neg p_1$
$\neg$	n/a
$\rightarrow$	$p_1 \rightarrow p_1, \neg p_1 \rightarrow \neg p_1$
$\leftrightarrow$	$p_1 \leftrightarrow p_1, \neg p_1 \leftrightarrow \neg p_1$
K_j	$\neg p_1, \ p_1 \vee \neg p_1, p_1 \rightarrow p_1, \neg p_1 \rightarrow \neg p_1, p_1 \leftrightarrow p_1, \ p_1 \leftrightarrow p_1, \neg p_1 \leftrightarrow \neg p_1$
E_G	Same as K_j
C_G	Same as K_j, $\neg p_1 \rightarrow \ K_j \neg p_1, P_j p_1, \neg K_j p_1 \wedge P_1 p_1$
D_G	Same as K_j

The adversary's knowledge (K_j) and common knowledge (C_G) consist of the satisfied propositional formulas in world x for this model or $\mathfrak{M}, x \models \phi$. Using these satisfied

formulas, the anonymity formulas Γ and learned knowledge B_1, it is possible to validate the sequent $\Gamma, B_1 \vdash \phi$. First let $\phi = K_j \neg p_1$ then let $\phi = P_j p_1$.

Let $\Gamma = \{C_G(p_1 \vee \neg p_1), C_G(\neg p_1 \rightarrow K_j \neg p_1), C_G(P_j p_1)\}$ and $B_1 = \{C_G(\neg K_1 p_1 \vee P_1 p_1), C_G \neg p_1\}$.

Proof: $\Gamma, B_1 \vdash K_j \neg p_1$ is valid.

1	$C_G(p_1 \vee \neg p_1)$	Premise (Γ)
2	$C_G(\neg p_1 \rightarrow K_j \neg p_1)$	Premise (Γ)
3	$C_G(P_j p_1)$	Premise (Γ)
4	$C_G(\neg K_1 p_1 \vee P_1 p_1)$	Premise (B_1)
5	$C_G \neg p_1$	Premise (B_1)
6	C_G	
7	$\neg p_1$	C_Ge 5
8	$\neg p_1 \rightarrow K_j \neg p_1$	C_Ge 2
9	$K_j \neg p_1$	$\rightarrow$ e 7,8 (Modus Ponens)
10	$C_G(K_j \neg p_1)$	C_Gi 9
11	$E_G(K_j \neg p_1)$	*CE* 10
12	$K_j K_j \neg p_1$	EK_j 11
13	$K_j \neg p_1$	*KT* 12

Let $\Gamma = \{C_G(p_1 \vee \neg p_1), C_G(\neg p_1 \rightarrow K_j \neg p_1), C_G(P_j p_1)\}$ and $B_2 = \{C_G(\neg K_1 p_1 \vee P_1 p_1)\}$.

Proof: $\Gamma, B_2 \vdash P_j p_1$ is valid.

1	$C_G(p_1 \vee \neg p_1)$	Premise (Γ)
2	$C_G(\neg p_1 \rightarrow K_j \neg p_1)$	Premise (Γ)
3	$C_G(P_j p_1)$	Premise (Γ)
4	$C_G(\neg K_1 p_1 \vee P_1 p_1)$	Premise (B_2)
5	$E_G(P_j p_1)$	*CE* 3
6	$K_j P_j p_1$	EK_j 5
7	$P_j p_1$	*KT* 6

Therefore, both $K_j \neg p_1$ and $P_j p_1$ are valid formulas and no anonymity exists.

6.2.2.2 Scenario II: Minimal Anonymity.

In this scenario of three agents (n = 3), two agents (k = 2) send two messages, one real (r = 1) and one dummy (d = 1); hence, *minimal* anonymity exists for the agents. The adversary commonly knows at least one agent may send a dummy message or $C_G(p_1 \vee p_2)$. It is common knowledge the anonymity rules state if the first or second agent sends a real message, the adversary thinks it is possible it is a dummy message or $C_G(\neg p_1 \rightarrow P_j p_1)$ and $C_G(\neg p_2 \rightarrow P_j p_2)$, respectively. Also, if the agents send a dummy message, the adversary does not know this or $C_G(p_1 \rightarrow \neg K_j p_1)$ and $C_G(p_2 \rightarrow \neg K_j p_2)$. Neither agent knows their own message type either or $C_G(\neg K_1 p_1 \wedge P_1 p_1)$ and $C_G(\neg K_2 p_2 \wedge P_2 p_2)$. The agents make this common knowledge after the adversary asks "Do you know if you sent a real message?"

Let A = {1, 2, 3} where $n = |A| = 3$ and adversary j and agent(s) $i \in A$, $G = A$ and P(Atoms) = $\{p_1, p_2\}$, then the formal KT45^3 model $\mathfrak{M} = (W, (R_i)_{i \in A}, L)$ is $W = \{x, y\}$; $R_j(x,y)$; $L(x) = \{p_1\}$, $L(y) = \{p_2\}$. The graphical PALM model is shown in Figure 56 below.

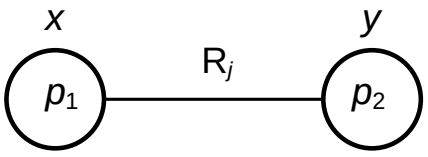


Figure 56: Scenario II PALM Model (KT45^n, n=3)

Two possible worlds exist x and y. A single reflexive, transitive, and symmetric accessibility binary relation for the adversary or $R_j(x,y)$ exists between the worlds. The varying degrees of knowledge are listed in Table 37. The adversary's knowledge (K_j)

and common knowledge (C_G) consist of the satisfied propositional formulas for each world x and y for this model or $\mathfrak{M}, x \models \phi$ and $\mathfrak{M}, y \models \phi$.

Table 37: Scenario II Satisfied Formulas (ϕ) (Adversary Knowledge)

Op	x	y
p	$p_1, \neg p_2$	$\neg p_1, p_2$
$\wedge$	$p_1 \wedge \neg p_2$	$\neg p_1 \wedge p_2$
$\vee$	$\neg p_1 \vee \neg p_2$ $p_1 \vee p_2$ $p_1 \vee \neg p_2$	$\neg p_1 \vee \neg p_2$ $p_1 \vee p_2$ $\neg p_1 \vee p_2$
$\neg$	$\neg(\neg p_1 \vee p_2)$ $\neg(p_1 \wedge p_2)$ $\neg(\neg p_1 \wedge \neg p_2)$ $\neg(\neg p_1 \wedge p_2)$	$\neg(p_1 \vee \neg p_2)$ $\neg(p_1 \wedge p_2)$ $\neg(\neg p_1 \wedge \neg p_2)$ $\neg(p_1 \wedge \neg p_2)$
$\rightarrow$	$\neg p_1 \rightarrow \neg p_2$ $p_1 \rightarrow \neg p_2$ $\neg p_1 \rightarrow p_2$ $p_2 \rightarrow p_1$ $\neg p_2 \rightarrow p_1$ $p_2 \rightarrow \neg p_1$	$p_1 \rightarrow p_2$ $\neg p_1 \rightarrow p_2$ $p_1 \rightarrow \neg p_2$ $\neg p_2 \rightarrow \neg p_1$ $\neg p_2 \rightarrow p_1$ $p_2 \rightarrow \neg p_1$
$\leftrightarrow$	$p_1 \leftrightarrow \neg p_2$ $\neg p_1 \leftrightarrow p_2$	$\neg p_1 \leftrightarrow p_2$ $p_1 \leftrightarrow \neg p_2$
K_j	$p_1 \vee p_2$ $\neg p_1 \vee \neg p_2$ $\neg(p_1 \wedge p_2)$ $\neg(\neg p_1 \wedge \neg p_2)$ $p_1 \rightarrow \neg p_2$ $\neg p_1 \rightarrow p_2$ $\neg p_2 \rightarrow p_1$ $p_2 \rightarrow \neg p_1$	$p_1 \vee p_2$ $\neg p_1 \vee \neg p_2$ $\neg(p_1 \wedge p_2)$ $\neg(\neg p_1 \wedge \neg p_2)$ $p_1 \rightarrow \neg p_2$ $\neg p_1 \rightarrow p_2$ $\neg p_2 \rightarrow p_1$ $p_2 \rightarrow \neg p_1$
E_G	$p_1 \vee p_2$ $\neg p_1 \vee \neg p_2$ $\neg(p_1 \wedge p_2)$ $\neg(\neg p_1 \wedge \neg p_2)$ $p_1 \rightarrow \neg p_2$ $\neg p_1 \rightarrow p_2$ $\neg p_2 \rightarrow p_1$ $p_2 \rightarrow \neg p_1$	$p_1 \vee p_2$ $\neg p_1 \vee \neg p_2$ $\neg(p_1 \wedge p_2)$ $\neg(\neg p_1 \wedge \neg p_2)$ $p_1 \rightarrow \neg p_2$ $\neg p_1 \rightarrow p_2$ $\neg p_2 \rightarrow p_1$ $p_2 \rightarrow \neg p_1$
C_G	$p_1 \vee p_2$ $\neg p_1 \vee \neg p_2$ $\neg p_1 \rightarrow P_j p_1$ $\neg(p_1 \wedge p_2)$ $\neg(\neg p_1 \wedge \neg p_2)$ $\neg p_2 \rightarrow P_j p_2$ $p_1 \rightarrow \neg p_2$ $\neg p_1 \rightarrow p_2$ $\neg K_1 p_1 \wedge P_1 p_1$ $\neg p_2 \rightarrow p_1$ $p_2 \rightarrow \neg p_1$ $\neg K_2 p_2 \wedge P_2 p_2$ $p_1 \rightarrow \neg K_j p_1$ $p_2 \rightarrow \neg K_j p_2$	
D_G	Same as K_j	Same as K_j

Using these satisfied formulas, the anonymity formulas Γ and learned knowledge *B*, it is possible to validate the sequent $\Gamma, B \vdash \phi$. First let $\phi = \neg K_j p_1$, then let $\phi = P_j p_1$. In the first proof, please note that for any model, formula ϕ is satisfiable **iff** its negation $\neg\phi$ is not valid [Gol05]. Let $\phi = \neg K_j p_1$, then $\neg K_j p_1$ is satisfiable **iff** $\neg(\neg K_j p_1) = K_j p_1$ is not valid in $\mathfrak{M}$.

Let $\Gamma = \{C_G(p_1 \vee p_2), C_G(p_1 \rightarrow \neg K_j p_1), C_G(p_2 \rightarrow \neg K_j p_2), C_G(\neg p_1 \rightarrow P_j p_1), C_G(\neg p_2 \rightarrow P_j p_2)\}$
and $B = \{C_G(\neg K_1 p_1 \vee P_1 p_1), C_G(\neg K_2 p_2 \vee P_2 p_2)\}$.

Proof 1: Γ, *B* |- $\neg K_j p_1$ is valid. *Minimal* formula valid for one agent.

1	$C_G(p_1 \vee p_2)$	Premise (Γ)
2	$C_G(p_1 \rightarrow \neg K_j p_1)$	Premise (Γ)
3	$C_G(p_2 \rightarrow \neg K_j p_2)$	Premise (Γ)
4	$C_G(\neg p_1 \vee P_j p_1)$	Premise (Γ)
5	$C_G(\neg p_2 \rightarrow P_j p_2)$	Premise (Γ)
6	$C_G(\neg K_1 p_1 \vee P_1 p_1)$	Premise (*B*)
7	$C_G(\neg K_2 p_2 \vee P_2 p_2)$	Premise (*B*)
8	C_G	
9	$K_j p_1$	Assume
10	K_j	
11	p_1	*KT* 9
12	$p_1 \rightarrow \neg K_j p_1$	$C_G e$ 2
13	$\neg K_j p_1$	$\rightarrow$ e 11,12 MP
14	$K_j \neg K_j p_1$	$K_i i$ 10–13
15	$\neg K_j p_1$	*KT* 14
16	$\perp$	$\neg\, e$ 9,15
17	$\neg K_j p_1$	$\neg\, i$ 9–16
18	$C_G(\neg K_j p_1)$	$C_G i$ 17
19	$E_G(\neg K_j p_1)$	*CE* 18
20	$K_j \neg K_j p_1$	EK_j 19
21	$\neg K_j p_1$	*KT* 20

Proof 2: $\Gamma, B \vdash P_j p_1$ is valid.

1	$C_G(p_1 \vee p_2)$	Premise (Γ)
2	$C_G(p_1 \rightarrow \neg K_j p_1)$	Premise (Γ)
3	$C_G(p_2 \rightarrow \neg K_j p_2)$	Premise (Γ)
4	$C_G(\neg p_1 \vee P_j p_1)$	Premise (Γ)
5	$C_G(\neg p_2 \rightarrow P_j p_2)$	Premise (Γ)
6	$C_G(\neg K_1 p_1 \vee P_1 p_1)$	Premise (B)
7	$C_G(\neg K_2 p_2 \vee P_2 p_2)$	Premise (B)
8	C_G	
9	$\neg P_j p_1$	Assume
10	$\neg\neg K_j \neg p_1$	Def. $P_j = \neg K_j \neg$ 9
11	$K_j p_1$	$\neg\neg$e 10
12	K_j	
13	$\neg p_1$	KT 11
14	$\neg p_1 \rightarrow P_j p_1$	$C_G e$ 5
15	$P_j p_1$	$\rightarrow$ e 13,14 MP
16	$K_j P_j p_1$	$K_i i$ 12–15
17	$P_j p_1$	KT 16
18	$\perp$	$\neg\, e$ 9,17
19	$\neg\neg P_j p_1$	$\neg\, i$ 8–18
20	$P_j p_1$	$\neg\neg\, e$ 19
21	$C_G(P_j p_1)$	$C_G i$ 20
22	$E_G(P_j p_1)$	CE 21
23	$K_j P_j p_1$	EK_j 22
24	$P_j p_1$	KT 23

6.2.2.3 Scenario III: Total Anonymity.

In this scenario of three agents (n = 3), two agents (k = 2) send two messages, one real (r = 1) and one dummy (d = 1) or no real (r = 0) and two dummy (d = 2); hence, *minimal* and *total* anonymity exists for the agents. The adversary commonly knows either agent may send a dummy message or $C_G(p_1 \vee p_2)$. It is common knowledge that the anonymity rules state if the first or second agent sends a real message, the adversary thinks it could be a dummy message or $C_G(\neg p_1 \rightarrow P_j p_1)$ and $C_G(\neg p_2 \rightarrow P_j p_2)$, respectively. It is common knowledge that if either sends a dummy message, the adversary does not

know this or $C_G(p_1 \rightarrow \neg K_j p_1)$ and $C_G(p_2 \rightarrow \neg K_j p_2)$, respectively. Neither agent knows their own message type either or $C_G(\neg K_1 p_1 \wedge P_1 p_1)$ and $C_G(\neg K_2 p_2 \wedge P_2 p_2)$. The agents make this common knowledge after the adversary asks "Do you know if you sent a real message?"

Let $A = \{1, 2, 3\}$ where $n = |A| = 3$ and adversary j and agent(s) $i \in A$, $G = A$ and $P(\text{Atoms}) = \{p_1, p_2\}$, then the formal KT45^3 model $\mathfrak{M} = (W, (R_i)_{i \in A}, L)$ is $W = \{x, y, z\}$; $R_j(x,y)$, $R_j(y,z)$; $L(x) = \{p_1\}$, $L(y) = \{p_1, p_2\}$ and $L(z) = \{p_2\}$. The graphical PALM model is shown in Figure 57 below.

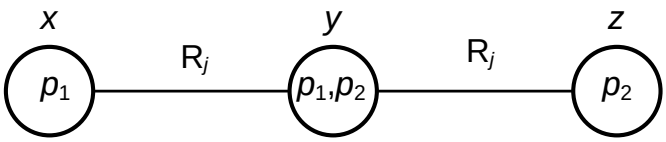


Figure 57: Scenario III PALM Model (KT45^n, n=3)

There are three possible worlds; x, y, and z. There are two reflexive, transitive, and symmetric accessibility binary relations for the adversary as well. The varying degrees of knowledge are in Table 38. The adversary's knowledge (K_j) and common knowledge (C_G) consist of the satisfied propositional formulas for each world x, y, and z for this model or $\mathfrak{M}, x \models \phi$, $\mathfrak{M}, y \models \phi$ and $\mathfrak{M}, z \models \phi$. Notice that the adversary knows fewer "things" or formulas (see K_j row, 2nd column) in world y compared to worlds x and z. What the adversary knows in y is constrained by the two relations to what is known in the other two worlds. Hence, a formula must be satisfied in all three worlds before the adversary may know it in world y. Also notice the reduction in common knowledge formulas compared to the previous model.

Table 38: Scenario III Satisfied Formulas (ϕ) (Adversary Knowledge)

Op	x	y	z
p	$p_1, \neg p_2$	p_1, p_2	$\neg p_1, p_2$
$\wedge$	$p_1 \wedge \neg p_2$	$p_1 \wedge p_2$	$\neg p_1 \wedge p_2$
$\vee$	$\neg p_1 \vee \neg p_2$ $p_1 \vee p_2$ $p_1 \vee \neg p_2$	$\neg p_1 \vee p_2$ $p_1 \vee p_2$ $p_1 \vee \neg p_2$	$\neg p_1 \vee \neg p_2$ $p_1 \vee p_2$ $\neg p_1 \vee p_2$
$\neg$	$\neg(\neg p_1 \vee p_2)$ $\neg(p_1 \wedge p_2)$ $\neg(\neg p_1 \wedge \neg p_2)$ $\neg(\neg p_1 \wedge p_2)$	$\neg(\neg p_1 \vee \neg p_2)$ $\neg(p_1 \wedge \neg p_2)$ $\neg(\neg p_1 \wedge p_2)$ $\neg(\neg p_1 \wedge \neg p_2)$	$\neg(p_1 \vee \neg p_2)$ $\neg(p_1 \wedge p_2)$ $\neg(\neg p_1 \wedge \neg p_2)$ $\neg(p_1 \wedge \neg p_2)$
$\rightarrow$	$\neg p_1 \rightarrow \neg p_2$ $p_1 \rightarrow \neg p_2$ $\neg p_1 \rightarrow p_2$ $p_2 \rightarrow p_1$ $\neg p_2 \rightarrow p_1$ $p_2 \rightarrow \neg p_1$	$p_1 \rightarrow p_2$ $\neg p_1 \rightarrow p_2$ $\neg p_1 \rightarrow \neg p_2$ $p_2 \rightarrow p_1$ $\neg p_2 \rightarrow p_1$ $\neg p_2 \rightarrow \neg p_1$	$p_1 \rightarrow p_2$ $\neg p_1 \rightarrow p_2$ $p_1 \rightarrow \neg p_2$ $\neg p_2 \rightarrow \neg p_1$ $p_2 \rightarrow \neg p_1$ $\neg p_2 \rightarrow p_1$
$\leftrightarrow$	$p_1 \leftrightarrow \neg p_2$ $\neg p_1 \leftrightarrow p_2$	$p_1 \leftrightarrow p_2$ $\neg p_1 \leftrightarrow \neg p_2$	$\neg p_1 \leftrightarrow p_2$ $p_1 \leftrightarrow \neg p_2$
K_j	p_1 $p_1 \vee p_2$ $p_1 \vee \neg p_2$ $\neg(\neg p_1 \wedge \neg p_2)$ $\neg(\neg p_1 \wedge p_2)$ $\neg p_1 \rightarrow \neg p_2$ $\neg p_1 \rightarrow p_2$ $\neg p_2 \rightarrow p_1$ $p_2 \rightarrow p_1$	$p_1 \vee p_2$ $\neg(\neg p_1 \vee \neg p_2)$ $\neg p_1 \rightarrow p_2$ $\neg p_2 \rightarrow p_1$	p_2 $p_1 \vee p_2$ $\neg p_1 \vee p_2$ $\neg(\neg p_1 \wedge \neg p_2)$ $\neg(p_1 \wedge \neg p_2)$ $p_1 \rightarrow p_2$ $\neg p_1 \rightarrow p_2$ $\neg p_2 \rightarrow \neg p_1$ $\neg p_2 \rightarrow p_1$
E_G	Same as K_j	Same as K_j	Same as K_j
C_G	$p_1 \vee p_2$ $\neg p_2 \rightarrow p_1$ $\neg p_1 \rightarrow P_j p_1$	$\neg p_1 \rightarrow p_2$ $p_1 \rightarrow \neg K_j p_1$ $\neg p_2 \rightarrow P_j p_2$ $\neg K_2 p_2 \wedge P_2 p_2$	$\neg(\neg p_1 \wedge \neg p_2)$ $p_2 \rightarrow \neg K_j p_2$ $\neg K_1 p_1 \wedge P_1 p_1$
D_G	Same as K_j	Same as K_j	Same as K_j

Using these satisfied formulas, the anonymity formulas Γ and learned knowledge B, it is possible to validate the sequent $\Gamma, B \vdash \phi$. The first proof of *Minimal* anonymity lets $\phi = \bigvee_{i \neq j} \neg K_j p_i$. The second proof of *Total* anonymity lets $\phi = \bigwedge_{i \neq j} P_j p_i$.

Let $\Gamma = \{C_G(p_1 \vee p_2), C_G(p_i \rightarrow \neg K_j p_i), C_G(\neg p_i \rightarrow P_j p_i)\}$ and $B = \{C_G(\neg K_i p_i \vee P_i p_i)\}$.

Proof: $\Gamma, B \vdash \bigvee_{i \neq j} \neg K_j p_i$ is valid. *Minimal* formula valid $\forall i$ agents, $i \neq j$.

1	$C_G(p_1 \vee p_2)$	Premise (Γ)
2	$C_G(p_i \rightarrow \neg K_j p_i)$	Premise (Γ)
3	$C_G(\neg p_i \rightarrow P_j p_i)$	Premise (Γ)
4	$C_G(\neg K_i p_i \vee P_i p_i)$	Premise (B)
5	C_G	
6	$p_1 \vee p_2$	Ce 1
7	p_1	Assume
8	$p_1 \rightarrow \neg K_j p_1$	C_Ge 2, $i = 1$
9	$\neg K_j p_1$	$\rightarrow$ e 7,8 MP
10	$\neg K_j p_1 \vee \neg K_j p_2$	$\vee$ i$_1$ 9
11	p_2	Assume
12	$p_2 \rightarrow \neg K_j p_2$	C_Ge 2, $i = 2$
13	$\neg K_j p_2$	$\rightarrow$ e 11,12 MP
14	$\neg K_j p_1 \vee \neg K_j p_2$	$\vee$ i$_2$ 13
15	$\neg K_j p_1 \vee \neg K_j p_2$	$\vee$ e 6,7-10,11-14
16	$\bigvee_{i \neq j} \neg K_j p_i$	Def. $\bigvee_{i \neq j} \neg K_j p_i \equiv \neg K_j p_1 \vee \neg K_j p_2$
17	$C_G(\bigvee_{i \neq j} \neg K_j p_i)$	C_Gi 16
18	$E_G(\bigvee_{i \neq j} \neg K_j p_i)$	CE 17
19	$K_j(\bigvee_{i \neq j} \neg K_j p_i)$	EK_j 18
20	$\bigvee_{i \neq j} \neg K_j p_i$	KT 19

Proof: $\Gamma, B \vdash \bigwedge_{i \neq j} P_j p_i$ is valid. *Total* formula valid $\forall i$ agents, $i \neq j$.

1	$C_G(p_1 \vee p_2)$	Premise (Γ)
2	$C_G(p_i \rightarrow \neg K_j p_i)$	Premise (Γ)
3	$C_G(\neg p_i \rightarrow P_j p_i)$	Premise (Γ)
4	$C_G(\neg K_j p_i \vee P_j p_i)$	Premise (B)
5	C_G	
6	$p_1 \vee p_2$	C_Ge 1
7	$\neg p_1 \wedge \neg p_2$	DeMorgans, 6
8	$\neg p_1$	$\wedge$ e$_1$ 7
9	$\neg p_2$	$\wedge$ e$_2$ 7
10	$\neg p_1 \rightarrow P_j p_1$	C_Ge 3, i = 1
11	$\neg p_2 \rightarrow P_j p_2$	C_Ge 3, i = 2
12	$P_j p_1$	$\rightarrow$ e 8,10 MP
13	$P_j p_2$	$\rightarrow$ e 9,11 MP
14	$P_j p_1 \wedge P_j p_2$	$\wedge$ i 12,13
15	$\bigwedge_{i \neq j} P_j p_i$	Def. $\bigwedge_{i \neq j} P_j p_i \equiv P_j p_1 \wedge P_j p_2$ 14
16	$C_G(\bigwedge_{i \neq j} P_j p_i)$	C_Gi 15
17	$E_G(\bigwedge_{i \neq j} P_j p_i)$	*CE* 16
18	$K_j(\bigwedge_{i \neq j} P_j p_i)$	EK_j 17
19	$\bigwedge_{i \neq j} P_j p_i$	*KT* 18

Therefore, both $\bigvee_{i \neq j} K_j p_i$ and $\bigwedge_{i \neq j} P_j p_i$ are valid formulas and the minimal and total anonymity properties hold.

6.2.2.4 Scenario IV: Up-to Anonymity.

In Scenario IV there are four agents (n = 4), three agents (k = 3) send three messages: no real (r = 0) and two dummy (d = 3), one real (r = 1) and two dummy (d = 2), or two real (r = 2) and one dummy (d = 1); hence, *minimal*, *total* and *up-to* $|I_A|$ anonymity exists for the agents depending on adversary knowledge. However, the adversary thinks the worst case is possible with up to three dummy messages sent ($1 \leq d \leq 3$). The adversary

commonly knows any agent may send a dummy message or $C_G(p_1 \vee p_2 \vee p_3)$. The other common knowledge is the same as before, except the formulas may be generally stated for any agent i as $C_G(p_i \rightarrow \neg K_j p_i)$, $C_G(\neg p_i \rightarrow P_j p_i)$ and $C_G(\neg K_i p_i \vee P_i p_i)$.

Let $A = \{1, 2, 3, 4\}$ where $n = |A| = 4$ and adversary j and agent(s) $i \in A$, $G = A$ and $\mathcal{P}(\text{Atoms}) = \{p_1, p_2, p_3\}$, then the formal KT45^4 model $\mathfrak{M} = (W, (R_i)_{i \in A}, L)$ is $W = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$; $R_j(x_1,x_2)$, $R_j(x_1,x_3)$, $R_j(x_1,x_4)$, $R_j(x_2,x_5)$, $R_j(x_3,x_6)$, $R_j(x_4,x_7)$; the labeling function L is monotonically increasing from leaf to root world or $L(x_1) = \{p_1, p_2, p_3\}$, $L(x_2) = \{p_1, p_2\}$, $L(x_3) = \{p_2, p_3\}$, $L(x_4) = \{p_1, p_3\}$, $L(x_5) = \{p_1\}$, $L(x_6) = \{p_2\}$, and $L(x_7) = \{p_3\}$. The PALM graphical worst case model is shown in Figure 58 below.

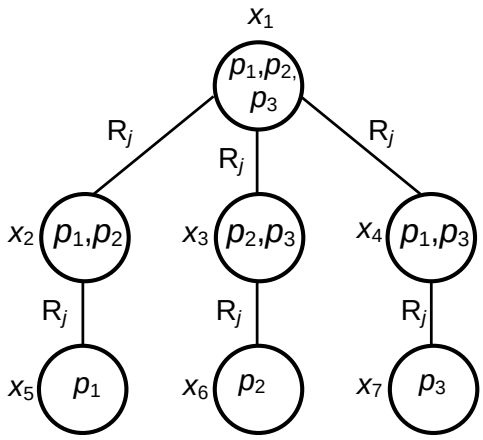


Figure 58: Scenario IV PALM Model (KT45^n, n=4)

This represents the adversary's *a priori* knowledge about the possible worlds assuming all k agents send messages (i.e., $I_A = A - \{j\}$, $k = |I_A| = 3$). However, assume after fewer than n-1 agents send messages; the adversary asks all agents simultaneously "Did you send a message?" Since the agents are honest, only $|I_A|$ say "Yes". The adversary now knows I_A and an updated model represents the adversary's *a posterior* knowledge (i.e., I_A

$\subset A$, $|I_A| < n$-1). Assume $I_A = \{1, 2\}$, the adversary would use the previous worst case PALM model where k=2, r=1, and d=1 as shown in Figure 59 below.

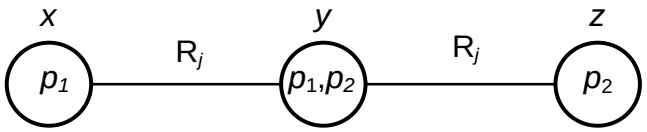


Figure 59: Scenario IV Improved PALM Model (KT45^n, n=3)

Clearly, proving the *up-to* $|I_A|$ anonymity formula $\bigwedge_{i' \in I_A} P_j p_i$ is identical to proving the *total* anonymity formula $\bigwedge_{i \neq j} P_j p_i$ in the previous example.

Let $\Gamma = \{C_G(p_1 \vee p_2), C_G(p_i \rightarrow \neg K_j p_i), C_G(\neg p_i \rightarrow P_j p_i)\}$ and $B = \{C_G(\neg K_i p_i \vee P_i p_i)\}$.

Proof: $\Gamma, B \vdash \bigwedge_{i' \in I_A} P_j p_{i'}$ is valid. *Up-to* $|I_A|$ formula valid $\forall i'$ agents, $i' \in I_A$.

1	$C_G(p_1 \vee p_2)$	Premise (Γ)
2	$C_G(p_i \rightarrow \neg K_j p_i)$	Premise (Γ)
3	$C_G(\neg p_i \rightarrow P_j p_i)$	Premise (Γ)
4	$C_G(\neg K_i p_i \vee P_i p_i)$	Premise (B)
5	C_G	
6	$p_1 \vee p_2$	C_Ge 1
7	$\neg p_1 \wedge \neg p_2$	DeMorgans, 6
8	$\neg p_1$	$\wedge$ e$_1$ 7
9	$\neg p_2$	$\wedge$ e$_2$ 7
10	$\neg p_1 \rightarrow P_j p_1$	C_Ge 3, i = 1
11	$\neg p_2 \rightarrow P_j p_2$	C_Ge 3, i = 2
12	$P_j p_1$	$\rightarrow$ e 8,10 MP
13	$P_j p_2$	$\rightarrow$ e 9,11 MP
14	$P_j p_1 \wedge P_j p_2$	$\wedge$i 12,13
15	$\bigwedge_{i' \in I_A} P_j p_i$	Def. $\bigwedge_{i' \in I_A} P_j p_i \equiv P_j p_1 \wedge P_j p_2$ 14
16	$C_G(\bigwedge_{i' \in I_A} P_j p_i)$	C_Gi 15
17	$E_G(\bigwedge_{i' \in I_A} P_j p_i)$	CE 16
18	$K_j(\bigwedge_{i' \in I_A} P_j p_i)$	EK_j 17
19	$\bigwedge_{i' \in I_A} P_j p_i$	KT 18

Therefore, $\bigwedge_{i' \in I_A} P_j p_i$ is a valid formula and the *up-to* $|I_A|$ anonymity property holds.

Scenario V is next.

6.2.2.5 Scenario V: k-Anonymity.

In this scenario of six agents ($n = 6$), five agents ($k = 5$) send five messages, two real ($r = 2$) and three dummy ($d = 3$); hence, *minimal*, *total*, *up-to* and *k-anonymity* exists for the agents depending on adversary knowledge. The adversary best case is possible with known two real messages. The adversary commonly knows any agent may send a dummy message or $C_G(\bigvee_{i \neq j} p_i)$. It is common knowledge that if an agent i sends a real message, the adversary thinks it could be a dummy message or $C_G(\neg p_i \rightarrow P_j p_i)$. It is common knowledge if agent i sends a dummy message, the adversary does not know this or $C_G(p_i \rightarrow \neg K_j p_i)$. No agent knows their own message type or $C_G(\neg K_i p_i \wedge P_i p_i)$. The agents make this common knowledge after the adversary asks "Do you know if you sent a real message?"

Let $A = \{1, 2, 3, 4, 5, 6\}$ where $n = |A| = 6$ and adversary j and agent(s) $i \in A$, $G = A$ and $\Phi(\text{Atoms}) = \{p_q: 1 \leq q \leq k\}$, the formal KT45^6 model $\mathfrak{M} = (W, (R_i)_{i \in A}, L)$ is $W = \{x_s: 1 \leq s \leq k\}$; $R_j(x_1,x_2)$, $R_j(x_2,x_3)$, $R_j(x_3,x_4)$, $R_j(x_4,x_5)$; and $L(x_i) = \{p_i\}$. A best case graphical PALM model assuming $|I_A| = k$ is shown in Figure 60.

This model represents the adversary's *a priori* knowledge about the possible worlds assuming all k agents may send a message (i.e., $I_A = A - \{j\}$, $k = |I_A| = 5$). Obviously, k-anonymity or 5-anonymity is achieved. However, after the messages are

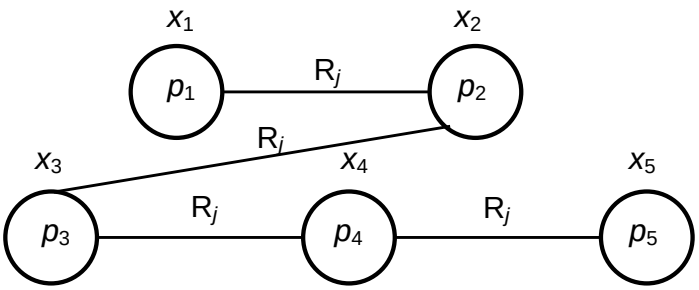


Figure 60: PALM Model (KT45^n, n=6)

sent, assume the adversary is able to distinguish between two separate "anonymity sets" I_{A1} and I_{A2} where $I_{A1} \cup I_{A2} \subseteq I_A$. In this example, since r = 2, the adversary knows one real message is sent per group. Assume I_{A1} = {1, 2} and I_{A2} = {3, 4, 5}, the adversary would use the PALM models where k=2, r=1, d=1 and k=3, r=1, d=2, respectively as shown in Figure 61.

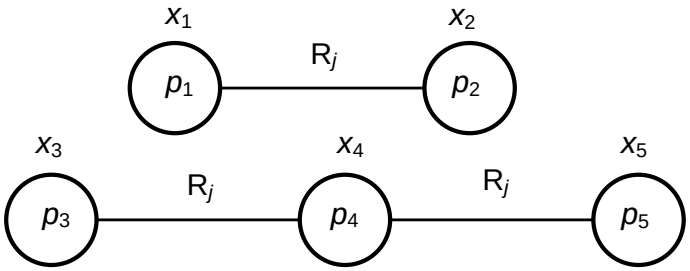


Figure 61: Improved PALM Model (KT45^n, n=6)

In effect, the adversary learned that $R_j(x_2,x_3)$ is not necessary. Obviously, at least k'-anonymity is achieved for all k agents. In this example, $k' = \lfloor k/r \rfloor = \lfloor 5/2 \rfloor = 2$ or 2-anonymity.

Clearly, the k-anonymity property holds but depends on how the adversary partitions the set of agents into anonymity sets. The majority of researchers assume the adversary pre-partitions the agents in the anonymous system before an attack.

The *k-anonymity* formula $\bigvee_{\{|I_A|\geq k\}} \bigwedge_{i'\in I_A} P_j p_{i'}$ may be rewritten as two *k'*-anonymity formulas $\bigwedge_{i'\in I_{A1}} P_j p_{i'} \vee \bigwedge_{i'\in I_{A2}} P_j p_{i'}$. Thus, proving the *k'-anonymity* formula is equivalent to proving a sequence of disjunctions of *up-to* $|I_{A1}|$ and *up-to* $|I_{A2}|$ formulas where $k' \leq |I_{A1}|, |I_{A2}| \leq k$.

Let $\Gamma = \{C_G(\bigwedge_{i'\in I_{A1}} P_j p_i), C_G(\bigwedge_{i'\in I_{A2}} P_j p_i)\}$.

Proof: Γ |- $\bigvee_{\{|I_A|\geq k'\}} \bigwedge_{i'\in I_A} P_j p_{i'}$ is valid. *k'-anonymity* formula valid $\forall i'$ agents, $i' \in I_A$.

1	$C_G(\bigwedge_{i'\in I_{A1}} P_j p_i)$	Premise (Γ)
2	$C_G(\bigwedge_{i'\in I_{A2}} P_j p_i)$	Premise (Γ)
3	C_G	
4	$P_j p_1 \wedge P_j p_2$	C_Ge 1, $i' \in I_{A1}$
5	$P_j p_3 \wedge P_j p_4 \wedge P_j p_5$	C_Ge 2, $i' \in I_{A2}$
6	$(P_j p_1 \wedge P_j p_2) \vee (P_j p_3 \wedge P_j p_4 \wedge P_j p_5)$	$\vee$i$_1$ 4,5
7	$\bigwedge_{i'\in I_{A1}} P_j p_i \vee \bigwedge_{i'\in I_{A2}} P_j p_i$	Def. $\bigwedge_{i'\in I_A} P_j p_{i'}$ 6
8	$\bigvee_{\{|I_A|\geq k'\}} \bigwedge_{i'\in I_A} P_j p_{i'}$	Def. $\bigvee_{\{|I_A|\geq k'\}}$ 7
9	$C_G(\bigvee_{\{|I_A|\geq k'\}} \bigwedge_{i'\in I_A} P_j p_{i'})$	C_Gi 8
10	$E_G(\bigvee_{\{|I_A|\geq k'\}} \bigwedge_{i'\in I_A} P_j p_{i'})$	*CE* 19
11	$K_j(\bigvee_{\{|I_A|\geq k'\}} \bigwedge_{i'\in I_A} P_j p_{i'})$	*EK$_j$* 10
12	$\bigvee_{\{|I_A|\geq k'\}} \bigwedge_{i'\in I_A} P_j p_{i'}$	*KT* 11

Therefore, $\bigvee_{\{|I_A|\geq k\}} \bigwedge_{i'\in I_A} P_j p_{i'}$ is a valid formula and only the *k'*-anonymity property holds.

6.3 Model Limitations

PALM models are easier to visualize, construct, and manipulate than operators on Boolean algebras inherent in process-calculi. However, it has the limitations of idealized knowledge, no temporal logic and no dynamic logic; hence, the need for alternative formalisms such as algebraic, neighborhood, and topological semantics. This is discussed in more detail below.

Humans and even computers lack the ability to "know all logically possible things" yet PALM assumes logical omniscience. An ability to reason with imperfect knowledge or only know a subset of all formulas is more realistic. Humans tend to believe things that are false and not believe things that are true. The logic of beliefs, desires, intentions or just plain common sense is not fully addressed in PALM. Also, PALM is unable to handle counterfactual conditions and non-monotonic reasoning (changing one's mind) as other formalisms do.

PALM does not include the concept of time. Time operators would allow a formula to be false now but true later or vice versa. In the no anonymity Scenario I model, issues about the lack of temporal logic in $KT45^n$ were evident. A combined time and knowledge logic may prove better than knowledge alone. However, some claim that the time dimension of analyzing security protocols only adds computational complexity and is easily abstracted away. Yet one formal approach uses Typed Model Logic plus [(OrL06)] to combine temporal and modal belief operators to specify, model, and reason about evolving theories of trust in agent based systems.

Finally, any dynamic change in the adversary's belief system is not captured. In Scenario V, the adversary first believed the worst case possible worlds existed but then

reasoned a better model existed. What caused this change? An ability to capture what actions took place to change the adversary's mind would prove most valuable. An action logic is simply not part of PALM – which can only reason after an action has taken place (e.g., message is sent) and assumes new knowledge is statically gained.

6.4 Summary

This chapter provides a rigorous, mathematical framework for modeling anonymous systems. The primary contribution of this chapter is formalizing how anonymity is preserved or degraded in an anonymous network based on adversary reasoning ability. The two primary knowledge operators K_j (agent) and C_G (common) and the epistemic and truth semantics made this possible. A simple anonymous network example, message-sender mystery, is discussed and proven with an expanded anonymous network example. Five scenarios are provided and the anonymity property formulas formally proved. Lastly, a few limitations of logical omniscience assumptions and lack of temporal and dynamic logic rules are highlighted.

VII. Conclusions and Recommendations

7.0 Chapter Overview

This chapter summarizes the dissertation research effort. The research conclusions are given in Section 7.2. Also, research contributions are delineated in Section 7.3. Lastly, Section 7.4 recommends future research to extend the research performed herein.

7.1 Research Conclusions

Historic to contemporary anonymity research issues have been surveyed. Over ten varying quantifications of anonymity are explained and the few conceptual and formal frameworks related to anonymity have been discussed. A methodology for the research was presented. The results include a novel cubic and tree-based taxonomy. In particular, seventeen wired and sixteen wireless anonymous communications protocols are explored and compared. In addition, a unique synthesis of anonymity metrics was identified. A formal epistemic logic framework was developed. Finally, the research proves that the $KT45^n$ logic is able to semantically represent possibilistic notions of anonymity but lacks action and temporal logics and bounded adversary aspects.

7.2 Research Contributions

Conceptual frameworks, metrics and formal models provide the ability to visualize anonymity protocols and anonymity services and better understand how anonymity is preserved, degraded or eliminated during a cyber attack in wired and wireless networks. The contribution of each of the three research areas is described next.

7.2.1 Anonymous Network Taxonomy

The contribution of the cubic/tree-based taxonomy (CT) is 3-fold. First, CT provides a definition of anonymity that extends the classical definition of anonymity to include four subtle yet important anonymity properties of mutual, group, group communication and location anonymity. Second, CT is the first known taxonomy to comprehensively cover both wired and wireless anonymous networks. CT complements previous wired anonymous network protocol family classifications and extends them with a novel peer-to-peer (P2P) anonymous network protocol family specification. CT is the only known taxonomy to capture the wireless anonymous protocol family relationships. Finally, the systematic classification and visually intuitive comparison of state-of-the-art wired and wireless anonymous protocols in this research is an innovative guide for future researchers' anonymity interests. The work in this area resulted in three fully referred conference papers [KeR08b, KeR09, KeR09a] and one soon-to-be published journal paper.

7.2.2 Anonymity Metrics

Knowing the available metrics and understanding the subtle changes in anonymity levels is essential for any organization determined to better defend against cyber attacks. This research gives researchers and organizations an ability to confidently measure information leakage given their specific anonymity requirements and application environment. The three accomplishments in this area include co-authoring a paper on analyzing client puzzles in Tor [Fra06], integrating data and network anonymity concepts

in a unique way [KeR08a], and exploring current metrics and issues in providing anonymity in mobile ad hoc networks [KeR08c].

7.2.3 Formal Adversary Anonymity Reasoning Model

One of the major benefits of formal methods is analytical techniques offer reasoning techniques that cover every possible state of a design, and well-defined proof techniques ensure the accuracy and correctness of a design. However, building a good mathematical model for representing anonymous protocols, and, even more so, formulating an appropriate definition of anonymity, is a non-trivial task. The model should be rich enough to represent a large variety of real-life adversarial behaviors, and the definition should guarantee the intuitive notion of anonymity is captured for any adversarial behavior under consideration. The formalization should be as clear and easy to work with as possible. This research took the first step towards building such an intuitive and mathematical model. This phase of the research resulted in a paper presented at the IEEE WIDA'08 conference [KeR08e].

7.2.4 Summary.

The contribution of this research to the field of computer science lies in its innovative development of a synergistic taxonomy, metrics, and formal model of anonymous networks. These contributions are summarized in Figure 62. In the taxonomy area, two complementary taxonomies were developed for classifying and comparing the myriad of wired and wireless anonymous protocols. Evolving issues in next generation mobile ad hoc anonymous wireless networks were highlighted. In terms of anonymity metrics, a client puzzle solution to mitigating DoS attacks on the Tor anonymous network was

analyzed. In addition, the seemingly disparate concepts of data and network based anonymity were merged to provide a common framework that researchers can use for future anonymity metric advances. A unique overview of state-of-the-art anonymity metrics was given. Finally, an epistemic-based model was created to model adversary reasoning ability.

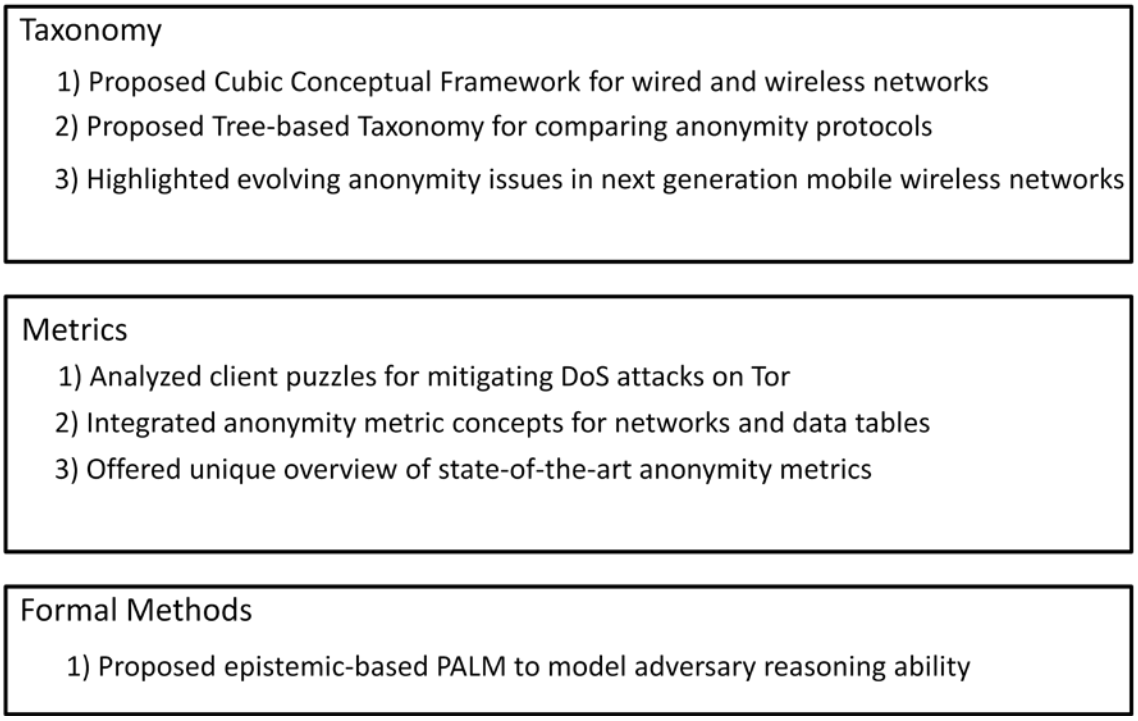


Figure 62: Summary of Contributions in Three Areas of Anonymous Networks

Figure 63 lists where each of the eight published research papers fall within each area. Four published papers are in the areas of anonymity network taxonomy. Three are in anonymity metric synthesis. One workshop paper falls in the area of epistemic-based formal methods.

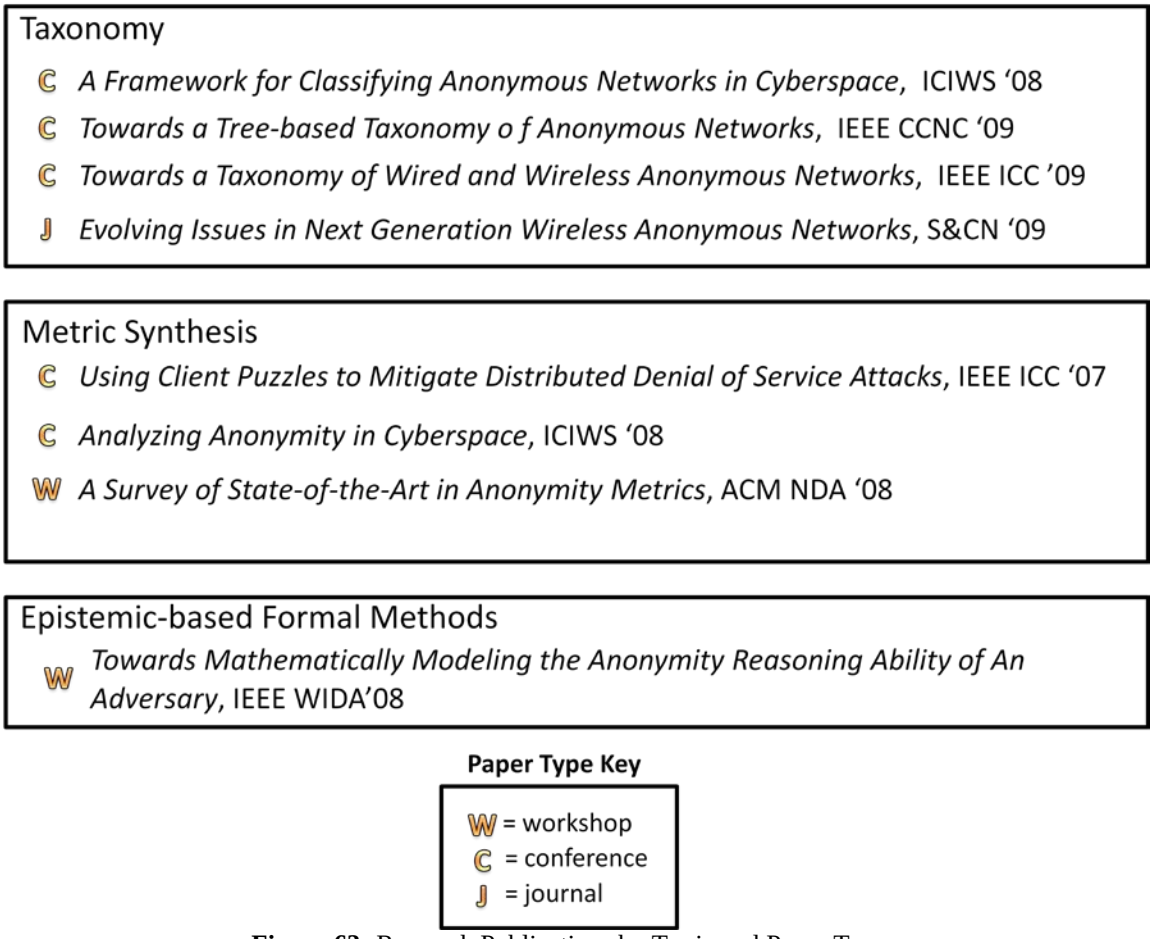
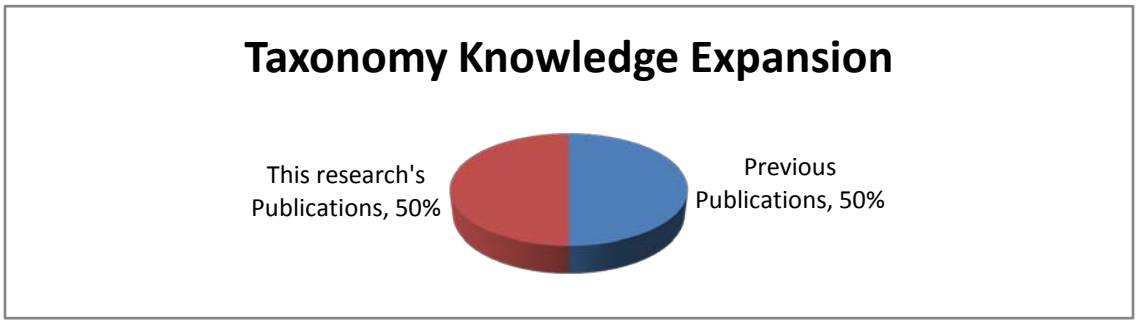


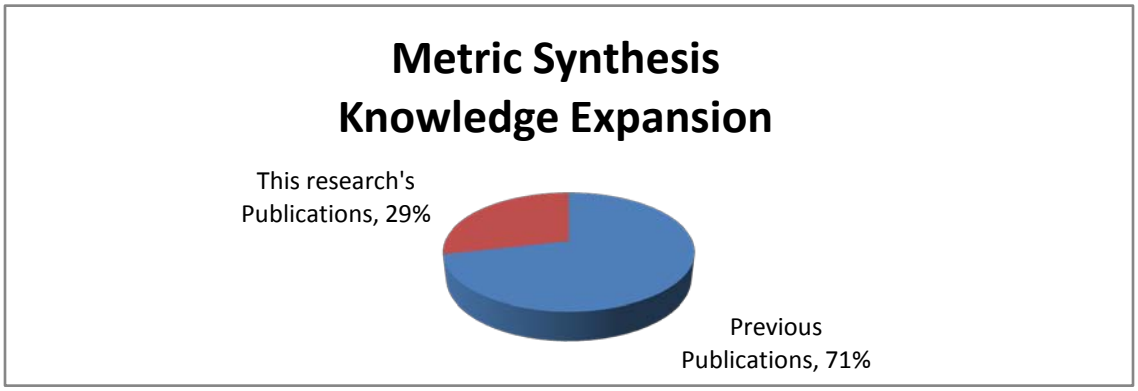
Figure 63: Research Publications by Topic and Paper Type

To gain a better appreciation of the knowledge expansion within each area, the simple metric of the percentage of newly published papers versus previously published papers is useful. Figure 64 displays a comparison of this research's contributions (in terms of

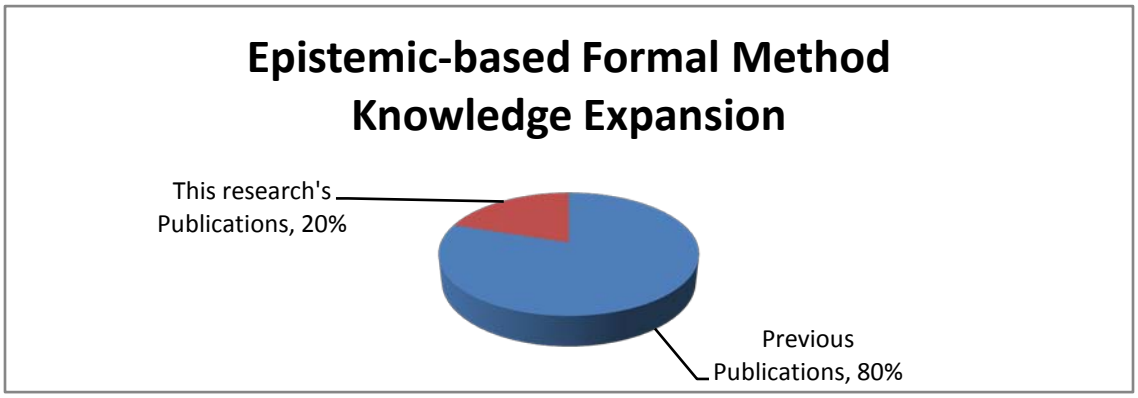
publications) versus the total number of publications that exist for the particular research area.



(a)



(b)



(c)

Figure 64: Knowledge Expansion by Subtopic

7.3 Recommendation for Future Research

The two prime areas for future research are in the anonymous network taxonomy and formal method topics. The area of anonymity metrics is active and continues to receive significant attention by other researchers in the field. Thus, an expansion of the conceptual taxonomy and formal models is in order.

For taxonomy, future work should more closely examine the last component – adversary capability – more completely to better articulate the overt and hidden adversary assumptions and implications for each anonymous protocol. This would make it easier to identify comparable anonymous system for further empirical or theoretical investigation as well as identifying gaps in anonymous protocol design.

For formal methods, immediate future work should relax the underlying PALM model assumption of logical omniscience and be applied toward a practical anonymous network such as Crowds or Tor. Another productive step would incorporate temporal and dynamic logic to provide a more expressive and quantitative means to (semi)-automatically verify anonymous protocols and properties. This would likely require the use of an appropriate theorem-prover and/or model checker. More interestingly, taking a modular or functional approach to analyzing a particular anonymous system, specific anonymity properties, and assumed adversary might prove most valuable. This combined approach would not only specify the anonymity properties in a modal logic as was done with the research herein but would also specify the anonymous system in process calculi and/or functions views. This process is represented in Figure 65.

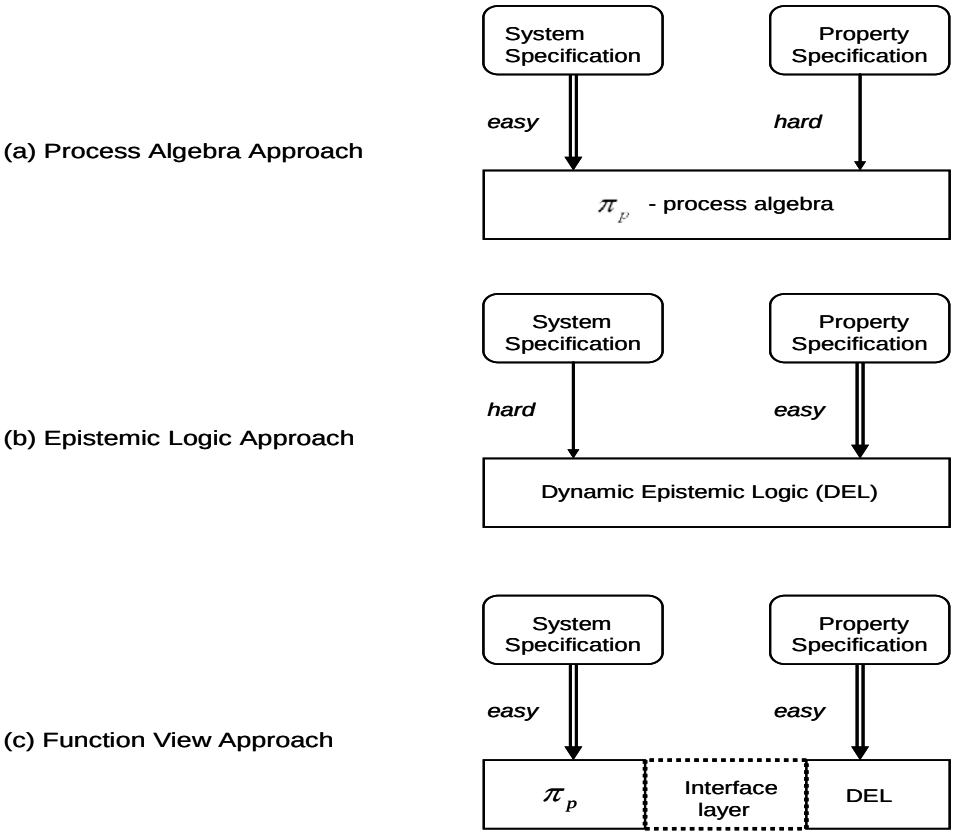


Figure 65: Modular Approach Example [HuS04]

In the process algebra approach in Figure 65(a), π_p-calculus represents the anonymous network behavior and is appropriate for modeling mobile networks. In the epistemic approach in Figure 65(b), a dynamic epistemic logic (DEL) can represent the desired anonymity properties and may include temporal logic and action models. In the function view approach in Figure 65(c), the interface layer has to be defined between the π_p-calculus system specification and DEL property specification. The primary contribution of this research would be to fill in the corresponding interface layer gap, an assuredly NP-hard problem, to allow formal reasoning about an adversary and how anonymity is preserved or degraded in an anonymous network.

BIBLIOGRAPHY

[AbF01] Abelson, H., M. Fischer, D. Weitzner, J. Pato, and J. Straggas. "Protecting privacy through anonymity tools." *Ethics and Law on the Electronic Frontier*. http://swiss.csail.mit.edu/6095/admin/admin-spring-2001/topics/anonymity.html.

[AbG97] Abadi, M. and A.D. Gordon, "Reasoning about Cryptographic Protocols in the Spi Calculus," *Concurrency Theory Lecture Notes in Computer Science*, vol. 1243, pp. 59-73, 1997.

[AdD03] Adi, K., M. Debbabi, and M. Mejri, "A New Logic for Electronic Commerce Protocols," *Theoretical Computer Science*, vol. 291, no. 3, pp. 223-283, 2003.

[AgF05] Aggarwal, G. , T. Feder, K. Kenthapadi, R. Motwani, R. Panigrahy, D. Thomas, and A. Zhu, "Approximation Algorithms for K-Anonymity," *Journal of Privacy Technology*, 2005.

[Alb02] Alberucci, L., "The Modal μ-calculus and the Logic of Common Knowledge," University of Bern, 2002.

[Ale07] Alex. "Boston PD Launch Anonymous Crime SMS Tip Service." *SMS Text News*. http://www.smstextnews.com/2007/06/boston_pd_launch_anonymous_crime_sms_tip_service.html. 15 June 2007.

[And01] Ross, A., *Security Engineering: A Guide to Building Dependable Distributed Systems*: Wiley, 2001.

[Ano02] Many-Worlds Intrepretation of Quantum Mechanics, Stanford, http://plato.stanford.edu/entries/qm-manyworlds/, Accessed.

[Ano04] Anonymous, "Anonymous Crime: Abuse of Anonymity Threatens the Internet," Church of Scientology International, 2004.

[Ano06] Anonymous. "Overview of Privacy." *Privacy International*. http://www.privacyinternational.org/article.shtml?cmd%5B347%5D=x-347-543673.

[Ano07g] Anonymous, "Anonymous Crime Reports on the Cards," www.dutchnews.nl, 2007.

[AtH99] Ateniese, G., A. Herzberg, H. Krawczyk, and G. Tsudik, "Untraceable Mobility or How to Travel Incognito," *Computer Networks*, vol. 31, no. 8, pp. 871-884, 1999.

[BaL07] Bao, L., "A New Approach to Anonymous Multicast Routing in Ad Hoc Networks," *Proceedings of the CHINACOM*, 2007.

[Bar95] Barrett, G., "Model Checking in Practice: The T9000 Virtual Channel Processor," *IEEE Transactions on Software Engineering*, vol. 21, no. 2, pp. 69-78, 1995.

[BeK85] Bergastra, J.A. and J.W. Klop, "Algebra for Communicating Processes with Abstraction (ACP)," *Journal of Theoretical Computer Science*, vol. 37, no. 1, pp. 77-121, 1985.

[BeP01] Berthold, O. , A. Pfitzmann, and Standtke. R, "The Disadvantages of Free MIX Routes and How to Overcome Them," *Proceedings of the International Workshop on Designing Privacy Enhancing Technologies (PET)*, Berkeley CA, USA, 2001.

[BeR03] Bettstetter, C., G. Resta, and P. Santi, "The Node Distribution of the Random Waypoint Mobility Model for Wireless Ad Hoc Networks," *IEEE Transactions on Mobile Computing*, vol. 2, no. 3, pp. 257-269, 2003.

[BeS03] Beresford, A. R. and F. Stajano, "Location Privacy in Pervasive Computing," *IEEE Pervasive Computing*, vol. 2, no. 1, pp. 46-55, 2003.

[BhP05] Bhargava, Mohit and Catuscia Palamidessi, "Probabilistic Anonymity," in *CONCUR 2005 - Concurrency Theory*, vol. 2653, *Lecture Notes in Computer Science*: Springer Berlin / Heidelberg, pp. 171-185, 2005.

[Blo70] Bloom, B.H., "Space/time Trade-offs in Hash Coding with Allowable Erros," *Communications of the ACM*, vol. 13, no. 7, pp. 422-426, 1970.

[BlV06] Blackburn, P., J. van Benthem, and Frank Wolter, *Handbook of Modal Logic*. North Holland, 2006.

[BoE04] Boukerche, A., K. El-Khatib, L. Xu, and L. Korba, "SDAR: a secure distributed anonymous routing protocol for wireless and mobile ad hoc networks," *Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks*, 2004.

[Bou89] Boute, R. T., "Representational and Denotational Semantics of Digital Systems," *Transactions on Computers*, vol. 38, no. 7, pp. 986-999, 1989.

[BoW05] Borisov, N. and J. Waddle, "Anonymity in Structured Peer-to-Peer Networks," Computer Science Division (EECS), University of California, Berkeley CA, 94720 USA *UCB/CSD-05-1390*, May 14, 2005.

[Boy97] Boyan, J., "The Anonymizer," in *Computer-Mediated Communication (CMC) Magazine*, 1997.

[BrA06] Bringsjord, S., K. Arkoudas, and Y. Yang, "New Architectures, Algorithms and Designs that Lead to Implemented Machine Reasoning over Knowledge in Epistemic and Deontic Formats, in the Service of Advanced Wargaming,"

Rensselaer Polytechnic Institute, Rome, New York *AFRL-IF-RS-TR-2006-264*, August, 2006.

[BrH84] Brookes, S.D., C.A.R. Hoare, and A.W. Roscoe, "A Theory of Communicating Sequential Processes," *Journal of the ACM*, vol. 31, no. 3, pp. 560-599, 1984.

[BuG03] *The National Strategy to Secure Cyberspace*, 2003.

[Cha81] Chaum, D. L., "Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms," *Communications of the ACM*, vol. 24, no. No. 2, pp. 84-88, 1981.

[Cha88] Chaum, D. L., "The Dining Cryptographers Problem: Unconditional Sender and Recipient Untraceability," *Journal of Cryptology*, vol. 1, no. 1, pp. 65-75, 1988.

[ChH04] Chen, T., T. Han, and J. Lu, "Towards a Modal Logic for pi-Calculus," *Proceedings of the 28th Annual International Computer Software and Applications Conference (COMPSAC)*, 2004.

[ChK03] Cheng, C., H. T. Kung, K.S. Tan, and S. Bradner, "ANON: an IP-layer Anonymizing Infrastructure," *Proceedings of the DARPA Information Survivability Conference and Exposition*, 2003.

[ChP05] Chatzikokolakis, K. and C. Palamidessi, "A Framework for Analyzing Probabilistic Procotols and its Application to Partial Secrets Exchange," Ecole Polytechnique, pp. 16, 2005.

[ChP07] Chatzikokolakis, K. , C. Palamidessi, and P. Panangaden, "Anonymity Protocols as Noisy Channels," *Symposium on Trustworthy Global Computing*, 2007.

[ChW06] Chen, S., X. Wang, and S. Jajodia, "On the Anonymity and Traceability of Peer-to-Peer VoIP Calls," *IEEE Network*, vol. 20, no. 5, pp. 32-37, 2006.

[CoB95] Cooper, D.A. and K.P. Birman, "Preserving Privacy in a Network of Mobile Computers," *Proceedings of the IEEE Symposium on Security and Privacy* 1995.

[Com05] Rights, Commissioner for Human, "Universal Declaration of Human Rights," United Nations Department of Public Information, 2005.

[Cot01] Anonymizer Inc., Mixmaster Protocol Version 3, http://www.eskimo.com/~rowdenw/crypt/Mix/draft-moeller-v3-01.txt, Accessed.

[CoW08] Coull, G. F., C. V. Wright, A. D. Keromytis, F. Monrose, and M. Reiter, "Taming the Devil: Techniques for Evaluating Anonymized Network Data," *Proceedings of the Internet Society (ISOC) Symposium on Network and Distributed Systems Security (SNDSS)*, San Diego CA, USA, 2008.

[Cra76] Craven, J.B., "Personhood: The Right to Be Let Alone " *Duke Law Journal*, vol. 1976, no. 4, pp. 699-720, 1976.

[Cre01] Creese, S., "Data Independent Induction: CSP Model Checking of Arbitrary Sized Networks," Oxford University, 2001.

[Dai98] Dai, W., "PipeNet 1.1," 1998.

[Dal86] Dalenius, T., "Finding a needle in a haystack - or identifying anonymous census record.," *Journal of Official Statistics*, vol. 2, no. 3, pp. 329-336, 1986.

[DaO05] Dawar, A. and M. Otto, "Modal Characterisation Theorems over Special Classes of Frames," *Proceedings of the 20th Annual IEEE Symposium on Logic in Computer Science*, 2005.

[DaR03] Danezis, G., R. Dingleldine, and N. Mathewson, "Mixminion: Design of a Type III Anonymous Remailer Protocol," *Proceedings of the IEEE Symposium on Security and Privacy (SP)*, 2003.

[DcS02] Diaz, C., S. Seys, J. Claessens, and B. Preneel, "Towards Measuring Anonymity," *Proceedings of the Privacy Enhancing Technologies workshop (PET)*, 2002.

[DeH04] Deng, J., R. Han, and S. Mishra, "Intrusion Tolerance and Anti-Traffic Analysis Strategies for Wireless Sensor Networks," *Proceedings of the IEEE International Conference on Dependable Systems and Networks (DSN)*, 2004.

[Del05] Dellarocas, C., "Reputation Mechanisms," University of Maryland, R.H. Smith School of Business, College Park, MD, June, 2005.

[DeP06] Deng, Y., J. Pang, and W. Peng, "Measuring Anonymity with Relative Entropy," *Proceedings of the 4th International Workshop on Formal Aspects in Security and Trust*, 2006.

[Dia05c] Diaz, C., "Anonymity and Privacy in Electronic Services," Katholieke Universiteit Leuven, 2005.

[DiC02] Diaz, C., J. Claessens, S. Seys, and B. Preneel, "Information Theory and Anonymity," *Proceedings of the 23rd Symposium on Information Theory in the Benelux*, Louvain la Neuve, Belgium, 2002.

[Dij06] Dijiang, Huang, "On Measuring Anonymity For Wireless Mobile Ad-hoc Networks," *Proceedings of the 31st IEEE Local Computer Networks Conference*, 2006.

[DiM04] Dingledine, R., N. Mathewson, and P. Syverson, "Tor: The Second Generation Onion Router," *Proceedings of the 13th USENIX Security Symposium*, 2004.

[DiP04] Diaz, C. and B. Preneel, "Anonymous Communication: WHOLES - A Multiple View of Individual Privacy in a Networked World," *Proceedings of the SICS Workshop*, Sweden, 2004.

[DiP04] Diaz, C. and B. Preneel, "Taxonomy of Mixes and Dummy Traffic," *Proceedings of the 3rd Working Conference on Privacy and Anonymity in Networked and Distributed Systems*, Toulouse, France, 2004.

[DoY83] Dolev, D. and A.C. Yao, "On the Security of Public Key Protocols," *IEEE Transactions on Information Theory*, vol. 29, no. 2, pp. 198-208, 1983.

[EdS07] Edman, Matthew, Fikret Sivrikaya, and Bulent Yener, "A Combinatorial Approach to Measuring Anonymity," *Proceedings of the IEEE Intelligence and Security Informatics*, 2007.

[Egg05] Eggendorfer, T., "Anonymous Surfing with Java Anonymous Proxy," in *Linxu Magazine*, 2005, pp. 44-46.

[Egr81] Egorychev, G.P., "The Solution of Van Der Waerden's Problem for Permanents," *Advances in Mathematics*, vol. 42, pp. 299-305, 1981.

[EiO07] Eijck, Jan van and Simona Orzan, "Epistemic Verification of Anonymity," *Electronic Notes in Theoretical Computer Science*, vol. 168, pp. 159-174, 2007.

[EiR85] Eisberg, R. and R. Resnick, *Quantum Physics of Atoms, Molecules, Solids, Nuclei, and Particles*, 2nd ed. Santa Barbara: Hamilton Printing Company, 1985.

[EiV89] van Eijk, P.H.J., C.A. Vissers, and M Diaz, "The Formal Description Technique LOTOS," *Computer Networks and ISDN Systems*, vol. 14, no. 1, pp. 25-29, 1989.

[ElK03] Il-Khatib, K., L. Korba, R. Song, and G. Yee, "Secure Dynamic Distributed Routing Algorithm for Ad hoc Wireless Networks," *Proceedings of the International Conference on Parallel Processing Workshops (ICPPW)*, 2003.

[Fal05] Falikman, D.I., "Proof of the Van Der Waerden's Conjecture on the Permanent of a Doubly Stochastic Matrix," *Mathematical Notes*, vol. 29, no. 6, pp. 475-479, 2005.

[Fal81] Falikman, D.I., "Proof of the Van Der Waerden's Conjecture on the Permanent of a Doubly Stochastic Matrix," *Mat. Zametki*, vol. 29, no. 6, pp. 931-938, 1981.

[FDR97] Ltd., Formal Systems (Europe), "Failures-Divergence Refinement: FDR2 User Manual," 1997.

[Fra06] Fraser, N.A., "Mitigiating Distributed Denial of Service Attacks in an Anonymous Routing Environment: Client Puzzles and Tor," Air Force Institute of Technology, 2006.

[Fre09] Anonymity Bibliography, http://www.freehaven.net/, Accessed.

[FrM02] Freedman, M.J. and R. Morris, "Tarzan: A Peer-to-Peer Anonymizing Network Layer " *Proceedings of the 9th ACM Conference on Computer and Communications Security (CCS)*, Washington DC, USA, 2002.

[GaH05] Garcia, F.D., I. Hasuo, W. Pieters, and P. van Rossum, "Provable Anonymity," *Proceedings of the ACM Workshop on Formal Methods in Security Engineering* Fairfax VA, USA, 2005.

[Gar03] Gardner, W. B., "Bridging CSP and C++ with selective formalism and executable specifications," *Proceedings of the First ACM and IEEE International Conference on Formal Methods and Models for Co-Design (MEMOCODE)* 2003.

[GeL04] Gedik, B. and L. Ling, "A Customizable k-Anonymity Model for Protecting Location Privacy," Georgia Institute of Technology, 2004.

[GeL05] Gedik, B. and L. Ling, "Location Privacy in Mobile Systems: A Personalized Anonymization Model," *25th IEEE International Conference on Distributed Computing Systems (ICDCS)*, pp. 620-629, 2005.

[GeL07] Gedik, B. and L. Liu, "Protecting Location Privacy with Personalized k-Anonymity: Architecture and Algorithms," *IEEE Transactions on Mobile Computing*, vol. 6, no. 9, 2007.

[GhK06] Ghinita, G., P. Kalnis, and S. Skiadopoulos, "PRIVE: Anonymous Location-Based Queries in Distributed Mobile Systems," Department of Computer Science, National University of Singapore, Singapore, 2006.

[Gol05] Goldblatt, R., *Mathematical Modal Logic: A View of Its Evolution*, vol. 6: Elsevier BV, 2005.

[GoR02] Goel, S., M. Robson, M. Polte, and E.G. Sirer, "Herbivore: A Scalable and Efficient Protocol for Anonymous Communication," Cornell, 2002.

[GoR03] Goel, S., M. Robson, M. Polte, and E. Sirer, "Herbivore: A Scalable and Efficient Protocol for Anonymous Communication," Cornell University, 2003.

[GoR96] Goldschlag, D. M., M. G. Reed, and P. Syverson, "Hiding Routing Information," *Proceedings of the Information Hiding Workshop*, Cambridge, UK, 1996.

[GoS99] Goldberg, I. and A. Shostack, "Freedom Network 1.0 Architecture and Protocols," 1999.

[GoT77] Goguen, J.A., J.W. Thatcher, E.G. Wagner, and J.B. Wright, "Initial Algebra Semantics and Continuous Algebras," *Journal of the ACM*, vol. 24, no. 1, pp. 68-95, 1977.

[GoW98] Goldberg, I. and D. Wagner, "TAZ Servers and the Rewebber Network:Enabling Anonymous Publishing on the World Wide Web," *First Monday*, vol. 4, no. 3, 1998.

[GrG03] Gruteser, Marco and Dirk Grunwald, "Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking," *Proceedings of the Proceedings of MobiSys 2003: The 1st International Conference on Mobile Systems, Applications, and Services*, San Francisco, CA, 2003.

[GrT96] Grassman, W.K. and J.P. Tremblay, *Logic and Discrete Mathematics: A Computer Science Perspective*: Prentice-Hall, Inc., 1996.

[GuF02] Guan, Y., X. Fu, R. Bettati, and W. Zhao, "An Optimal Strategy for Anonymous Communication Protocols," *Proceedings of the 22nd International Conference on Distributed Computing Systems*, College Station, TX, 2002.

[GuF04] Guan, Yong, Xinwen Fu, and Riccardo Bettati, "An Optimal Strategy for Anonymous Communication Protocols," *Proceedings of the 22nd International Conference on Distributed Computing Systems*, College Station, TX, 2002.

[GuF04] Guan, Y., X. Fu, R. Bettati, and W. Zhao, "A Quantitative Analysis of Anonymous Communications," *IEEE Transactions on Reliability*, vol. 53, no. 1, pp. 103-115, 2004.

[HaC02] Hall, A. and R. Chapman, "Correctness by construction: Developing a Commerical Secure System," in *IEEE Software*, vol. 19: Prentice Hall, pp. 18-25, 2002.

[HaJ01] Hartenstein, H. , K. Jonas, M. Liebsch, M. Stiemerling, R. Schmitz, and D. Westhoff, "Scalable Anonymous Connections in the Context of MIP and AAA," *Proceedings of the 10th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises* 2001.

[HaJ06] Haack, C. and A. Jeffrey, "Pattern-matching Spi-calculus," *Journal of Information and Computation*, vol. 204, no. 8, pp. 1195-1263, 2006.

[Hal05] *CSCE531: Discrete Math Lecture Notes*, 2006.

[HaL05] Han, J., Y. Liu, R. Xiao, and L. Ni, "A Mutual Anonymous Peer-to-peer Protocol Design," *Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, 2005.

[HaO02] Halpern, J. and K. O'Neill, "Secrecy in Multiagent Systems," *Proceedings of the 15th IEEE Computer Security Foundations Workshop*, 2002.

[HaO03] Halpern, J. Y. and K. R. O'Neill, "Anonymity and Information Hiding in Multiagent Systems," *Proceedings of the 16th IEEE Computer Security Foundations Workshop*, 2003.

[HeP00] Herescu, O.M. and C. Palamidessi, "Probabilistic Asynchronous Spi Calculus," *Lecture Notes in Computer Science*, vol. 1784, pp. 146--??, 2000.

[Hoa04] Hoare, C.A.R., *Communicating Sequential Processes*, 2004.

[Hoa69] Hoare, C.A.R., "An Axiomatic Basis for Computer Programming," *Communications of the ACM*, vol. 12, no. 10, pp. 576-580, 1969.

[Hoh82] Hohle, U., "Entropy with respect to Plausibility Measures," *Proceedings of the 1st IEEE International Symposium on Multiple-Valued Logic*, 1982.

[HqW04] He, Q., D. Wu, and P. Khosla, "Quest for Personal Control over Mobile Location Privacy," in *IEEE Communications Magazine*, vol. 45, 2004, pp. 130-136.

[HuD01] Hustadt, U., C. Dixon, R. A. Schmidt, M. Fisher, J. J. Meyer, and W. van der Hoek, "Reasoning about agents in the KARO framework," *Proceedings of the Eigth International Symposium on Temporal Representation and Reasoning (TIME)* Cividale del Friuli, Italy, 2001.

[Hui04] Huimin, Lin, "Formal Methods for Trustworthy Mobile Computing," *Proceedings of the 28th Annual International Computer Software and Applications Conference (COMPSAC'04)* 2004.

[Hum98] Watch, Human Rights, "Limits of Tolerance: Freedom of Expression and the Public Debate in Chile," Human Rights Watch *1-56432-192-4*, Nov 1998, 1998.

[HuR04] Huth, Michael and Mark Ryan, *Logic in Computer Science: Modelling and Reasoning about Systems*, 2nd ed. Cambridge, United Kingdom: The Press Syndicate of the University of Cambridge, 2004.

[HuS04] Hughes, D. and V. Shmatikov, "Information Hiding, Anonymity and Privacy: A Modular Approach," *Journal of Computer Security*, vol. 12, no. 1, pp. 3-36, 2004.

[IEE09] "IEEE Xplore," IEEE, 2009.

[IEE99] IEEE, "IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification," 1999.

[Iye02] Iyengar, V.S., "Transforming Data to Satisfy Privacy Constraints," *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery in Databases and Data Mining (KDD)*, Edmonton AB, Canada, 2002.

[JaJ01] Jakobsson, M. and A. Juels, "An Optimally Robust Hybrid Mix Network," *Proceedings of the 12th Annual ACM Symposium on Principles of Distributed Computing* Newport, Rhode Island, United States 2001.

[JiK05] Jiang, S. and R. Kumar, "Supervisory Control of Discrete Event Systems with CTL* Temporal Logic Specifications," *SIAM Journal on Control and Optimization*, 2005.

[Jon04] Jones, A., "Anonymous Communication on the Internet," Indiana University, pp. 13, 2004.

[Jor07] Jordan, M., "Anonymous Crime Report." Greenville, NC: East Carolina University Police Department, 2007.

[JuF05] Jurca, R. and B. Faltings, "Enforcing Truthful Strategies in Incentive Compatible Reputation Mechanisms," *Internet and Network Economics (WINE)*, vol. 3828, pp. 268-277, 2005.

[JuF06] Jurca, R. and B. Faltings, "Minimum Payments that Reward Honest Reputation Feedback," *Proceedings of the ACM Conference on Electronic Commerce (EC)*, Ann Arbor MI, USA, 2006.

[JuF07] Jurca, R. and B. Faltings, "Collusion Resistant, Incentive Compatible Feedback Payments," *Proceedings of the ACM Conference on E-Commerce (EC)*, San Diego CA, USA, 2007.

[KaG06] Kalnis, P., G. Ghinita, K. Mouratidis, and D. Papdias, "Perserving Anonymity in Location Based Services," Computer Science Department, National Universityof Singapore, Singapore, *Technical*, June, 2006.

[KaM06] Kawabe, Y., K. Mano, H. Sakurada, and Y. Tskuada, "Theorem-proving Anonymity on Infinite-State Systems," *Information Processing Letters*, vol. 101, no. 1, pp. p 46-51, 2007.

[KaM07] Kao, J.C. and R. Marculescu, "Real-Time Anonymous Routing for Mobile Ad Hoc Networks," *Proceedings of the IEEE International Conference on Local Computer Networks*, Tampa FL, USA, 2007.

[KeE98] Kesdogan, D., J. Egner, and R. Buschkes, "Stop-and-go Mixes Providing Probabilistic Security in an Open System," *Proceedings of the Second International Workshop on Information Hiding*, 1998.

[Ker07] Krebs, B., "Attacks Prompt Update for 'Tor' Anonymity Network," in *washingtonpost.com*, 2007, pp. 1.

[KeR08a] Kelly, D., R. Raines, R. Baldwin, B. Mullins, and M. Grimaila, "Analyzing Anonymity in Cyberspace," *Proceedings of the 3rd International Conference on i-Warfare and Security (ICIW)*, Omaha NE, USA, 2008.

[KeR08b] Kelly, D., R. Raines, R. Baldwin, B. Mullins, and M. Grimaila, "A Framework for Classifying Anonymous Networks in Cyberspace," *Proceedings of the 3rd International Conference on I-Warfare and Security (ICIW)*, Omaha NE, USA, 2008.

[KeR08c] Kelly, D., R. Raines, R. Baldwin, B. Mullins, and M. Grimaila, "A Survey of the State of the Art in Anonymity Metrics," *Proceedings of the 1st Workshop on Network Data Anonymization (NDA 2008)*, 2008.

[KeR08e] Kelly, D., R. Raines, R. Baldwin, B. Mullins, and M. Grimaila, "Towards Mathematically Modeling the Anonymity Reasoning Ability of an Adversary," *Proceedings of the IPCCC International Workshop on Information and Data Assurance*, Austin TX, USA, 2008.

[KeR09] Kelly, D., R. Raines, R. Baldwin, B. Mullins, and M. Grimaila, "Towards a Tree-based Taxonomy of Anonymous Networks," *Proceedings of the 6th Annual Consumer Communications & Networking Conference (CCNC)*, Las Vegas NV, USA, 2009.

[KeR09a] Kelly, D., R. Raines, R. Baldwin, B. Mullins, and M. Grimaila, "Towards a Taxonomy of Wired and Wireless Anonymous Networks," *Proceedings of the International Conference on Communications*, Dresden, Germany, 2009.

[Kes01] Kesdogan, D., "Evaluation of anonymity providing techniques using queuing theory," *Proceedings of the 26th Annual IEEE Conference on Local Computer Networks (LCN)* Tampa FL, USA, 2001.

[KiG06] Kifer, D. and J. Gehrke, "Injecting Utility into Anonymized Datasets," *Proceedings of the ACM SIGMOD International Conference on Management of Data*, Chicago IL, USA, 2006.

[KoH03] Kong, J. and X. Hong, "ANODR: ANonymous On Demand Routing with Untraceable Routes for Mobile Adhoc Networks," *Proceedings of the 4th ACM International Symposium on Mobile Ad-hoc Networking & Computing (MobiHoc)*, Annapolis MD, USA, 2003.

[KoH05] Kong, J., X. Hong, M. Y. Sanadid, and M. Gerla, "Mobility Changes Anonymity: Mobile Ad Hoc Networks Need Efficient Anonymous Routing," *Proceedings of the 10th IEEE Symposium on Computers and Communications (ISCC)*, 2005.

[KoH07] Kong, Jiejun, Xiaoyan Hong, and Mario Gerla, "An Identity-Free and On-Demand Routing Scheme against Anonymity Threats in Mobile Ad Hoc Networks," *IEEE Transactions on Mobile Computing*, vol. 6, no. 8, pp. 888-902, 2007.

[KoL07] Kong, J., J. Liu, X. Hong, D. Wu, and M. Gerla, *On Performance Cost of On-demand Anonymous Routing Protocols in Mobile Ad Hoc Networks* Springer US, 2007.

[Kon05] Kong, J., "Formal Notions of Anonymity for Peer-to-peer Networks," University of California, pp. 20, 2005.

[KoS04] Kouzmin, E. V., N. V. Shilov, and V. A. Sokolov, "Model Checking Mu-calculus in Well-Structured Transition Systems," *Proceedings of the 11th International Symposium on Temporal Representation and Reasoning (TIME'04)*, 2004.

[KoV98] Ko, Y. and N. Vaidva, "Location-Aided Routing (LAR) in Mobile Ad Hoc Networks," *Proceedings of the 4th International Conference on Mobile Computing and Networking*, Dallas, USA, 1998.

[Kri63] Kripke, S., "A Semantic Analysis of Modal Logic I: Normal Modal Propositional Calculi," *Zeitschrift fur Matematische Logik und Grundlagen der Mathematik*, vol. 9, pp. 67-96, 1963.

[LeD06] LeFevre, K., D. J. DeWitt, and R. Ramakrishnan, "Mondrian Multidimensional K-Anonymity," *Proceedings of the 22nd International Conference on Data Engineering (ICDE)*, 2006.

[Lem07] Lemos, R. "MySpace bars 29,000 sex offenders." *www.securityfocus.com*.

[LeS02] Levine, B. and C. Shields, "Hordes: A Multicast-based Protocol for Anonymity," *Journal of Computer Security*, vol. 10, no. 3, pp. 213-240, 2002.

[Lew18] Lewis, C.I., *A Survey of Symbolic Logic*. Berkeley: University of California Press, 1918.

[LhM04] Li, H., Y. Min, and Z. Li, "Evaluation of Dispersity Routing Strategies in Anonymous Communication," *Proceedings of the 10th Asia-Pacific Conference on Communications and 5th International Symposiwn on Multi-Dimensional Mobile Communications*, 2004.

[Li06] Li, L.I., "Reputation, Trust, and Rebates: How Online Auction Markets Can Improvve Their Feedback Mechanisms," *Proceedings of the Western Economic Association*, NYU, 2006.

[LiH06] Liu, J., X. Hong, J. Kong, Q. Zheng, N. Hu, and P. Bradford, "A Hierarchical Anonymous Routing Scheme for Mobile Ad-Hoc Networks," *Proceedings of the Military Communications Conference (MILCOM)*, 2006.

[LiK05] Liu, J., J. Kong, X. Hong, and M. Gerla, "Performance Evaluation of Anonymous Routing Protocols in MANETs," University of Alabama, pp. 6, 2005.

[LiL07] Li, Ninghui, Tiancheng Li, and Suresh Venkatasubramanian, "t-Closeness: Privacy Beyond k-Anonymity and l-Diversity," *Proceedings of the IEEE 23rd International Conference on Data Engineering*, 2007.

[Liu07] Liu, Ling, "From Data Privacy to Location Privacy: Models and Algorithms," *Proceedings of the VLDB*, Vienna, Austria, 2007.

[LiX06] Liu, X. and L. Xiao, "hiREP: Hierarchical Reputation Management for Peer-to-Peer Systems," *Proceedings of the International Conference on Parallel Processing (ICPP)* 2006.

[Luc06] Publius, the Pseudonym and Poetry, http://idtrail.org/index2.php?option=com_content&do_pdf=1&id=525, Accessed.

[LuF04] Lu, T., B. Fang, Y. Sun, and X. Cheng, "WonGoo: A Peer-to-Peer Protocol for Anonymous Communication," *Journal of Parallel and Distributed Processing Techniques and Applications*, vol. 3, pp. 1102-1106, 2004.

[LuF05] Lu, T., B. Fang, Y. Sun, and X. Cheng, "Performance Analysis of WonGoo System," *Proceedings of the Fifth International Conference on Computer and Information Technology (CIT)*, 2005.

[MaG06] Machanavajjhala, A., J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "l-Diversity: Privacy Beyond k-Anonymity," *Proceedings of the 22nd International Conference on Data Engineering*, Atlanta GA, USA, 2006.

[Mar01] Marx, G.T. "Identity and Anonymity: Some Conceptual Distinctions and Issues for Research " *Documenting Individual Identity*. http://web.mit.edu/gtmarx/www/identity.html.

[Mar99] Marx, G.T., "What's in a Name? Some Reflections on the Sociology of Anonymity " *International Journal on Information Society*, vol. 15, no. 2, pp. 99-112, 1999.

[MaW04] Meyerson, A. and R. Williams, "On the Complexity of Optimal K-Anonymity," *Proceedings of the 23rd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS '04)*, Paris, France, 2004.

[Mcc06] McCullagh, D. . "Create an e-annoyance, Go To Jail." http://news.com.com/Create+an+e-annoyance,+go+to+jail/2010-1028_3-6022491.html.

[Men05] Menzel, C., "Prior's Model Logic," Standford Encyclopedia of Philosophy, 2005.

[Mer06] Merkle, P.B., "Extended Defense Systems: I. Adversary-Defender Modeling Grammar for Vulnerability Analysis and Threat Assessment," Sandia Corporation, a Lockheed Martin Company *SAND2006-1484*, March, 2006.

[Mic61] Michael, J., "Justices of the Peace Act," England, 1361.

[Mil89] Milner, R., *Communication and Concurrency*: Prentice-Hall, 1989.

[MiR06] Miranda, H. and L. Rodriques, "A Framework to Provide Anonymity in Reputation Systems," *Proceedings of the 3rd Annual Mobile and Ubiquitous Systems: Networking & Services*, 2006.

[MiX06] Misra, S. and G. Xue, "SAS: A Simple Anonymity Scheme for Clustered Wireless Sensor Networks," *Proceedings of the IEEE International Conference on Communications (ICC)*, 2006.

[Mol03] Mouller, B., "Provably Secure Public-key Encryption for Length-preserving Chaumian Mixes," *Proceedings of the Topic in Cryptology CT-RSA*, San Francisco, CA, USA, 2003.

[MoS06] Morales, A. and G. Sciavicco, "Using Temporal Logic for Spatial Reasoning: Spatial Propositional Neighborhood Logic," *Proceedings of the 13th International Symposium on Temporal Representation and Reasoning* (*TIME)*, 2006.

[MuW08] Murdoc, S. and R. Watson, "Metrics for Security and Performance in Low-Latency Anonymity Networks," *Proceedings of the Eighth International Symposium on Privacy Enhancing Technologies*, Leuven, Belgium, 2008.

[MwX06] Min, Wu and Ye Xiaojun, "Towards the Diversity of Sensitive Attributes in k-Anonymity," *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and International Agent Technology Workshops (WI-IAT '06)*, 2006.

[NeC06] Nergiz, M. E. and C. Clifton, "Thoughts on k-Anonymization," *Proceedings of the 22nd International Conference on Data Engineering Workshops (ICDEW)*, 2006.

[NeM03] Newman, R., I. Moskowitz, P. Syverson, and A. Serjantov, "Metrics for Traffic Analysis Prevention " *Proceedings of the Privacy Enhancing Technologies Workshop*, Dresden, Germany, 2003.

[Nis97] Nissenbaum, H. "Toward an Approach to Privacy in Public: Challenges of Information Technology." *Ethics & Behavior*. http://www.leaonline.com/doi/abs/10.1207/s15327019eb0703_3?cookieSet=1&journalCode=eb.

[Nis98] Nissenbaum, H., "Protecting Privacy in an Information Age: The Problem of Privacy in Public," *Law and Philosophy*, vol. 17, pp. 5-6, 1998.

[Nis99] Nissenbaum, H., "The Meaning of Anonymity in an Information Age," *The Information Society*, vol. 15, pp. 141-144, 2001.

[OrL06] Orgun, M. A. and C. Liu, "Reasoning about Dynamics of Trust and Agent Beliefs," *Proceedings of the IEEE International Conference on Information Reuse and Integration*, 2006.

[Pal05] Palamidessi, Catuscia, "Anonymity in Probabilistic and Nondeterministic Systems," *Proceedings of the Workshop "Essays on Algebraic Process Calculi" (APC)*, 2005.

[PaM86] Pfitzmann, A. and M. Waidner, "Networks without User Observability - Design Options," *Proceedings of the Workshop on the Theory and Application of Cryptographic Techniques*, Linz, Austria, 1986.

[PaM86] Ptfizmann, Andreas and Michael Waidner, "Networks without User Observability - Design Options," *Proceedings of the EUROCRYPT'85, Lecture Notes in Computer Science*, 1986.

[PaO02] Parker, C.J., C. O'Brien, and J.J. Werner, "Roberson v. Rochester Folding Box Company," 1902.

[Pas00] Pascual, A.E., "Anonymous and Untraceable Communications," 21 June, 2000.

[PaV97] Parrow, J. and B. Victor, "The Update Calculus," *Proceedings of the Algebraic Methodology and Software Technology (AMAST)*, Sydney, 1997.

[PeB03] Perkins, C., E. Belding-Royer, and S. Das, "Ad hoc On-demand Distance Vector (AODV) Routing," RFC 2561, July 2003, 2003.

[Pel05] Peleska, J., *Applied Formal Methods - From CSP to Executable Hybrid Specifications*, vol. 3525: Springer Berlin / Heidelberg, 2005.

[PfK00] Pfitzmann, A. and M. Kohntopp, "Anonymity, Unobservability, and Pseudonymity – A Proposal for Terminology," *Proceedings of the Workshop on Design Issues in Anonymity and Unobservability*, Berkeley, CA, USA, 2000.

[PfP91] Pfitzmann, A., B. Pfitzmann, and M. Waidner, "ISDNMixes: Untraceable Communication with Very Small Bandwidth Overhead," *Proceedings of the GI/ITG: Communication in Distributed Systems*, Mannheim, Germany, 1991.

[Pos81] Posner, R.A., "The Economics of Privacy," *American Economic Review*, vol. 71, no. 2, pp. 405-409, 1981.

[RaM06] Rahman, M., M. Mambo, A. Inomata, and E. Okamoto, "An Anonymous On-Demand Position-Based Routing in Mobile Ad Hoc Networks," *Proceedings of the International Symposium on Applications and the Internet (SAINT)* 2006.

[RaS93] Rackoff, C. and D. R. Simon, "Cryptographic Defense Against Traffic Analysis," *Proceedings of the 25th annual ACM Symposium on the Theory of Computation (STOC)*, San Diego CA, 1993.

[ReP02] Rennhard, M. and B. Plattner, "Introducing MorphMix: Peer-to-Peer based Anonymous Internet Usage with Collusion Detection " *Proceedings of the WPES*, Washington, DC, 2002.

[ReP03] Rennhard, M. and B. Plattner, "Practical Anonymity for the Masses with Mix-Networks," *Proceedings of the Twelfth IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WET ICE)*, 2003.

[ReR06] Reiter, M. and A. Rubin, "Crowds: Anonymity for Web Transactions, Ed., 31 Jan, 2006.

[ReR88] Reed, M. G. and A.W. Roscoe, "A Timed Model for Communicating sequential Processes," *Journal of Theoretical Computer Science*, vol. 58, no. 1-3, pp. 249-261, 1988.

[ReR98] Reiter, M.K. and A.D. Rubin, "Crowds: Anonymity for Web Transactions," *ACM Transactions on Information and System Security*, vol. 1, no. 1, pp. 66-92, 1998.

[ReS98] Reed, M. G., P. F. Syverson, and D. M. Goldschlag, "Anonymous Connections and Onion Routing," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 4, 1998.

[Rig95] MIT, Anonymity on the Internet Must be Protected, http://www.swiss.ai.mit.edu/6805/student-papers/fall95-papers/rigby-anonymity.html, Accessed.

[Rig95] Rigby, K., "Anonymity on the Internet Must Be Protected," in *Ethics and Law on the Electronic Frontier*, 1995.

[RiS78] Rivest, R.L., A. Shamir, and L. Adleman, "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems " *MIT/LCS/TM-82*, 1978.

[RmR98] Reiter, M. and A. Rubin, "Crowds: Anonymity for web transactions," *ACM Transactions on Information and System Security*, pp. 66-92, 1998.

[Rob07] Roberston, C. "Be a Good Citizen: Report Crime with Anonymous Email." *www.majon.com*. http://www.majon.com/articles/Law_and_Politics/anonymous_email_418.html.

[RsZ04] Ray, S. and Z. Zhang, "An Efficient Anonymity Protocol for Grid Computing," *Proceedings of the Fifth IEEE/ACM International Workshop on Grid Computing*, 2004.

[RyS01] Ryan, P. Y. and S. Schneider, *Modelling and Analysis of Security Protocols*: Addison-Welsey, 2001.

[SaM95] Samfat, D., R. Molva, and N. Asokan, "Untraceability in Mobile Networks," *ACM MOBICOM*, pp. 26-36, 1995.

[SaP06] Sampigethaya, K. and R. Poovendran, "A Survey on Mix Networks and Their Secure Applications," *IEEE*, vol. 94, no. 12, pp. 2142-2181, 2006.

[SaS98] Samarati, P. and L. Sweeney, "Protecting Privacy when Disclosing Information: k-Anonymity and Its Enforcement through Generalization and Suppression," *Proceedings of the IEEE Symposium on Research in Security and Privacy*, 1998.

[Sch02] Schmidt, M., "Subscriptionless Mobile Networking: Anonymity and Privacy Aspects within Personal Area Networks," *Proceedings of the IEEE in Wireless Communications and Networking Conference (WCNC)*, 2002.

[Sch94] Schervish, P.G., "The Sound of One Hand Clapping: The Case For and Against Anonymous Giving " *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, vol. 5, no. 1, pp. 1-26, 1994.

[ScS71] Scott, D.S. and C. Strachey, *Toward a Mathematical Semantics for Computer Languages*: Polytechnic Inst. of Brooklyn Press, 1971.

[ScS96] Schneider, S. and A. Sdiropoulos, "CSP and Anonymity," *Proceedings of the Fourth European Symposium on Research in Computer Security (ESORICS)* 1996.

[SeD02] Serjantov, A. and G. Danezis, "Towards an Information Theoretic Metric for Anonymity," *Proceedings of the Privacy Enhancing Technologies workshop (PET)*, 2002.

[SeM96] Seidel, K. and C. Morgan, "Hierarchical Reasoning in Probabilistic CSP," Oxford University Computing Laboratory *PRG-TR-3-96*, 1996.

[Sen02] Sentz, K., "Combination of Evidence in Dempster-Shafer Theory," Binghamton University, Binghamton, NY *SAND 2002-0835*, April, 2002.

[SeP06] Seys, S. and B. Preneel, "ARM: Anonymous Routing Protocol for Mobile Ad hoc Networks," *Proceedings of the 20th International Conference on Advanced Information Networking and Applications (AINA)*, Vienna, AU, 2006.

[Ser05] Serjantov, A., "On the Anonymity of Anonymity Systems, http://www-sop.inria.fr/everest/events/cassis05/Transp/serjantov.ppt," serjantov.ppt, Ed., 2005.

[Sha48] Shannon, C.E., "A Mathematical Theory of Communication," *Bell System Technical Journal*, vol. 27, pp. 379-423, pp. 623-656, 1948.

[Sha76] Shafer, G., *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton University Press, 1976.

[ShB02] Sherwood, R., B. Bhattacharjee, and A. Srinivasan, "P5: A Protocol for Scalable Anonymous Communication " *Proceedings of the Security and Privacy*, 2002.

[ShL00] Shields, C. and B. Levine, "A protocol for anonymous communication over the Internet," *Proceedings of the 7th ACM conference on Computer and Communications Security (CCS'00)*, Athens, Greece, 2000.

[Sik94] Sikka, V., "Logical Entailment," Stanford University, 1994.

[SoK05] Song, R., L. Korba, and G. Yee, "AnonDSR: Efficient Anonymous Dynamic Source Routing for Mobile Ad-hoc Networks," *Proceedings of the 3rd ACM Workshop on Security of Ad hoc and Sensor Networks* Alexandria VA, USA 2005.

[Ste03] Stepney, S., "CSP/FDR2 to Handel-C Translation," Department of Computer Science, University of York, 2003.

[SuK04] Suto, H., H. Kawakami, and O. Katai, "Researches on operators' mental model for structure of operations based on situation theory," *Proceedings of the SICE*, 2004.

[SuP03] Sunderam, V., J. Pascoe, and R. Loader, "Towards a Framework for Collaborative Peer Groups," *Proceedings of the 3rd IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGrid)*, 2003.

[Swe02] Sweeney, L., "k-Anonymity: A model for protecting privacy," *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, vol. 10, no. 5, pp. 557-570, 2002.

[SyC06] Sy, D., R. Chen, and L. Bao, "ODAR: On-Demand Anonymous Routing in Ad Hoc Networks," University of California, Irvine, pp. 1-10, 2006.

[SyG95] Syverson, P. F. and J. W. Gray, III, "The Epistemic Representation of Information Flow Security in Probabilistic Systems," *Proceedings of the Eighth IEEE in Computer Security Foundations Workshop* County Kerry, Ireland, 1995.

[SyS99] Syverson, P. F. and S. G. Stuart, "Group Principals and the Formalization of Anonymity," *Proceedings of the World Congress on Formal Methods in the Development of Computing Systems (FM)*, Berlin, 1999.

[TgH04] Toth, G. and Z. Hornak, "Measuring Anonymity in a Non-adaptive, Real-time System," *Privacy Enhancing Technologies workshop (PET)*, vol. 3424, pp. 226-241, 2004.

[TgH04a] Toth, G., Z. Hornak, and F. Vajda, "Measuring Anonymity Revisited," *Proceedings of the Ninth Nordic Workshop on Secure IT Systems*, Espoo, Finland, 2004.

[ThS05] Thati, P., K. Sen, and N. Marti-Oliet, "An Executable Specification of Asynchronous Pi-Calculus Semantics and May Testing in Maude 2.0," *Electronic Notes in Theoretical Computer Science*, vol. 71, pp. April, 2005.

[TiO05] Tillwick, H. and M. Olivier, "Towards a Framework for Connection Anonymity," *Proceedings of the SAICSIT*, 2005.

[Uni04] of, Commerce United States Department. "Overview of the Privacy Act of 1974, 2004 Edition." http://www.usdoj.gov/oip/1974indrigacc.htm.

[Uni05] Government, United States, "United States Code, Title 47," Cornell University Law Library, 2005.

[Uni97] United Nations Educational, Scientific and Cultural Organization, "Declaration on the Responsibilities of the Present Generations Towards Future Generations," 1997.

[Unk12] Unknown, "The Right of Privacy," *The Virginia Law Register*, vol. 18, no. 7, pp. 550-553, 1912.

[UnS01] Unypoth, A. and P. Sewell, "Nomadic Pict: Correct Communication Infrastructure for Mobile Computation," *Proceedings of the 28th Annual Symposium on Principles of Programming Languages (POPL)*, 2001.

[VaD92] Valacich, J. S., A. R. Dennis, L. M. Jessup, and J. F. Nunamaker, Jr., "A Conceptual Framework of Anonymity in Group Support Systems," *Proceedings of the 25th Hawaii International Conference on System Sciences* Kauai HI, USA, 1992.

[Ven97] Venne, M. "An Attempt to Define the Five Most Important Privacy Issues." *Electronic Frontier Canada*. http://www.efc.ca/pages/media/ledevoir.23sep97.html.

[Wai90] Waidner, M., "Unconditional Sender and Recipient Untraceability in Spite of Active Attacks," *Proceedings of the Workshop on the Theory and Application of Cryptographic Techniques on Advances in Cryptology*, Houthalen, Belgium 1990.

[Wai90] Waidner, M., "Unconditional Sender and Recipient Untraceability in Spite of Active Attacks," *Proceedings of the Workshop on the Theory and Application of Cryptographic Techniques on Advances in Cryptology*, Houthalen, Belgium 1990.

[Wal01] Walton, T. "Internet Privacy Law." http://www.netatty.com/privacy/privacy.html.

[WaN07] Wang, P., P. Ning, and D. Reeves, "A k-Anonymous Communication Protocol for Overlay Networks," *Proceedings of the ASIACCS*, Singapore, 2007.

[Wei99] Weiss, G., *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. Cambridge, MA, London, England: MIT Press, 1999.

[Wik07] Wikipedia, "Anycast, http://en.wikipedia.org/wiki/Anycast," 2007.

[Wik07a] Equivalence Relation, http://en.wikipedia.org/wiki/Equivalence_relation, http://en.wikipedia.org/wiki/Equivalence_relation, Accessed.

[WiK07c] Wikipedia, "Formal Semantics of Programmin Languages," Wikipedia, 2007.

[Woo06] Woo, J. "The Right Not to be Identified: Privacy and Anonymity in the Interactive Media Environment." *SAGE Journals Online*.

[WrS05] Wright, M., S. Stepney, J.A. Clark, and J. Jacob, "Designing Anonymity: A Formal Basis for Identity Hiding," York University, Heslington, York, 21 June, 2005.

[WsB90] Warren, S.D. and L.D. Brandeis, "The Right to Privacy," vol. 4 Harv. L. Rev, pp. 193-220, 1890.

[WuB05] Wu, X. and E. Bertino, "Achieving K-anonymity in Mobile Ad Hoc Networks," *Proceedings of the 1st IEEE ICNP Workshop on Secure Network Protocols (NPSec).* , 2005.

[WuB05a] Wu, X. and B. Bhargava, "AO2P: Ad hoc On-demand Position-based Private Routing Protocol," in *IEEE Transactions on Mobile Computing*, vol. 4, pp. 335-348, 2005.

[Wux04] Wu, X., "DISPOSER: DIstributed Secure POSition SERvice in Mobile Ad Hoc Networks," Department of Computer Sciences *CSD TR # 04-027*, 2004.

[Wux05] Wu, X., "VDPS: Virtual Home Region based Distributed Position Service in Mobile Ad Hoc Networks," *Proceedings of the 25th IEEE International Conference on Distributed Computing Systems (ICDCS)*, 2005.

[WxB05] Wu, X. and B. Bhargava, "AO2P: Ad Hoc On-Demand Position-Based Private Routing Protocol," *IEEE Transactions on Mobile Computing*, vol. 4, no. 4, pp. 335-348, 2005.

[XiB05] Xiaoxin, Wu and E. Bertino, "Achieving K-anonymity in Mobile Ad Hoc Networks," *Proceedings of the 1st IEEE ICNP Workshop on Secure Network Protocols (NPSec'05)*, 2005.

[XiL06] Xiao, L., X. Liu, Gu. W., D. Xuan, and Y. Liu, "A Design of Overlay Anonymous Multicast Protocol," *Proceedings of the Parallel and Distributed Processing Symposium (IPDPS)*, 2006.

[XiR07] Lin, X., R. Lu, H. Zhu, P. Ho, X. Shen, and Z. Cao, "ASRPAKE: An Anonymous Secure Routing Protocol with Authenticated Key Exchange for Wireless Ad Hoc Networks," *Proceedings of the IEEE International Conference on Communications, 2007 (ICC '07)*, 2007.

[XiT07] Xiao, X. and Y. Tao, "m-Invariance: Towards Privacy Preserving Re-publication of Dynamic Datasets," *Proceedings of the SIGMOD*, Beijing, China, 2007.

[XiX03] Xiao, L., Z. Xu, and X. Zhang, "Mutual Anonymity Protocols for Hybrid Peer-to-Peer Systems," *Proceedings of the 23rd International Conference on Distributed Computing Systems*, 2003.

[XiX03a] Xiao, L., Z. Xu, and X. Zhang, "Low-Cost and Reliable Mutual Anonymity Protocols in Peer-to-Peer Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 14, no. 9, pp. 829-840, 2003.

[Yag83] Yager, R., "Entropy and Specificity in a Mathematical Theory of Evidence," *International Journal of General Systems*, vol. 9, no. 4, pp. 249-260, 1983.

[YaI04] Yamamoto, H., K. Ishida, and T. Ohta, "Trust Formation in a C2C Market: Effect of Reputation Management System," *Proceedings of the Seventh International Workshop on Trust in Agent Societies*, 2004.

[YaI05] Yamamoto, H., K. Ishida, K. Arai, and H. Deguchi, *Evolution of Cooperative Behavior in C2C market: Effect of Reputation Management System*, 2005.

[YoF04] Younis, O. and S. Fahmy, "Distributed Clustering in Ad Hoc Sensor Networks: A Hybrid, Energy-Efficient Approach," *Proceedings of the INFOCOM*, 2004.

[ZaM99] Zacharia, G., A. Moukas, and P. MAes, "Collaborative Reputation Mechanisms in Electronic Marketplaces," *Proceedings of the 32nd Hawaii International Conference on System Sciences*, 1999.

[ZaW05] Yanchao, Zhang, Liu Wei, and Lou Wenjing, "Anonymous Communications in Mobile Ad Hoc Networks," *Proceedings of the 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, 2005.

[ZhH04] Zhu, Y. and Y. Hu, "TAP: A Novel Tunneling Approach for Anonymity in Structured P2P Systems," *Proceedings of the International Conference on Parallel Processing (ICPP '04)*, 2004.

[ZhH06] Zhu, H., J. He, and J. Bowen, "From Algebraic Semantics to Denotational Semantics for Verilog," *Proceedings of the 11th IEEE International Conference on Engineering of Complex Computer Systems (ICECCS)*, 2006.

[ZhL06] Zhang, Z., W. Liu, and Y. Fang, "MASK: Anonymous On-Demand Routing in Mobile Ad Hoc Networks," *IEEE Transactions on Wireless Communications*, vol. 5, no. 9, pp. 2376-2385, 2006.

[ZhN05] Renyi, Z., V. L. Narasimhan, and S. Sastry, "Algebraic Semantics for Complete Interaction Sequence," *Proceedings of the IEEE Region 10 TENCON*, 2005.

[ZhW04] Zhu, B., Z. Wna, M. S. Kankanhalli, F. Bao, and R. H. Deng, "Anonymous Secure Routing in Mobile Ad-Hoc Networks," *29th Annual IEEE International Conference on Local Computer Networks*, pp. 102-108, 2004.

[ZhZ05] Zhuang, L., F. Zhou, B. Zhao, and A. Rowstron, "Cashmere: Resilient Anonymous Routing," *Proceedings of the 2nd Symposium on Networked Systems Design & Implementation (NSDI)*, Boston MA, 2005.

REPORT DOCUMENTATION PAGE		Form Approved OMB No. 074-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
26-03-2009	**Dissertation**	09/2005 – 03/2009

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER n/a
A Taxonomy For and Analysis of Anonymous Communications Networks	5b. GRANT NUMBER n/a
	5c. PROGRAM ELEMENT NUMBER n/a
6. AUTHOR(S)	5d. PROJECT NUMBER 09-295
Douglas J. Kelly	5e. TASK NUMBER n/a
	5f. WORK UNIT NUMBER n/a

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)	8. PERFORMING ORGANIZATION REPORT NUMBER
Air Force Institute of Technology (AFIT) Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way WPABF, OH 45433-7765	AFIT/DCS/ENG/09-08
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
National Security Agency (NSA) POC: Christine Nickell 9800 Savage Road, Suite 6722 Ft. George G. Meade, MD 20755-6000	NSA/CSS 11. SPONSOR/MONITOR'S REPORT NUMBER(S) n/a

12. DISTRIBUTION/AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Any entity operating in cyberspace is susceptible to debilitating attacks. With cyber attacks intended to gather intelligence and disrupt communications rapidly replacing the threat of conventional and nuclear attacks, a new age of warfare is at hand. In 2003, the United States acknowledged that the speed and anonymity of cyber attacks makes distinguishing among the actions of terrorists, criminals, and nation states difficult. Even President Obama's Cybersecurity Chief-elect feels challenged by the increasing sophistication of cyber attacks. Indeed, the rising quantity and ubiquity of new surveillance technologies in cyberspace enables instant, undetectable, and unsolicited information collection about entities. Hence, anonymity and privacy are becoming increasingly important issues. Anonymization enables entities to protect their data and systems from a diverse set of cyber attacks and preserve privacy.

This research provides a systematic analysis of anonymity degradation, preservation and elimination in cyberspace to enchance the security of information assets. This includes discovery/obfuscation of identities and actions of/from potential adversaries. First, novel taxonomies are developed for classifying and comparing the wide variety of well-established and state-of-the-art anonymous networking protocols. These expand the classical definition of anonymity and are the first known to capture the peer-to-peer and mobile ad hoc anonymous protocol family relationships. Second, a unique synthesis of state-of-the-art anonymity metrics is provided. This significantly aids an entities ability to reliably measure changing anonymity levels; thereby, increasing their ability to defend against cyber attacks. Finally, a novel epistemic-based model is created to characterize how an adversary reasons with knowledge to degrade anonymity. This offers multiple anonymity property representations and well-defined logical proofs to ensure the accuracy and correctness of current and future anonymous network protocol design.

15. SUBJECT TERMS

Privacy, anonymity, anonymous protocols, metrics, taxonomy, formal methods, probability, communication networks

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
REPORT	ABSTRACT	c. THIS PAGE			Dr. Richard A. Raines, AFIT (ENG)
U	U	U	UU	257	19b. TELEPHONE NUMBER *(Include area code)* (927) 255-6565, ext 4278; email: Richard.Raines@afit.edu

Standard Form 298 (Rev: 8-98)
Prescribed by ANSI Std. Z39-18