Insider Threats for File Sharing: Characterising File Sharing and its Protection

Rakan Alsowail and Ian Mackie
School of Engineering and Informatics,
University of Sussex, Falmer, Brighton, BN1 9RH, United Kingdom
ra216@sussex.ac.uk, i.mackie@sussex.ac.uk

ABSTRACT

This paper is about file sharing applied to the insider threat problem. We propose a new approach for classifying the insider threat problem and focus on one category that is related to file sharing. We characterise the protection required by the shared files against different types of insiders. We also characterise file sharing based on two factors: how and with whom files might be shared.

KEYWORDS

Insider threat, security, information security, file sharing, data confidentiality, file dissemination.

1 INTRODUCTION

File sharing has been a topic of interest in computer science right from the beginning—ever since files were created. Most of the research on file sharing is focused on specific domains and applications while little research studied file sharing more broadly [1]. The term *file sharing* is rarely defined in the literature, and if defined, it is tailored to a specific method of sharing. One exception is the study by Whalen et al. [1] who defined file sharing as "the activity of making specified file(s) available to an individual or group, with the option of granting specific right (e.g., ability to view, edit, delete) over those files". However, a general characterisation of how files might be shared is currently missing.

Shared files might be confidential content that needs to be protected against unauthorised disclosure, modification, or withholding. Generally speaking, such protections stem from three distinct fields of security: Communication security which is concerned with preventing different types of attacks on data transmitted over a network; Perimeter security which is concerned with preventing attacks on data stored inside a trusted internal network; and insider security which is concerned with preventing attacks on data by those who have authorised access.

According to the 2011 CyberSecurity Watch Survey, conducted by the U.S. Secret Service, the CERT Insider Threat Center, CSO Magazine, and Deloitte [2], 58% of the attacks are caused by outsiders (those unauthorised to access network systems or data) while 21% of the attacks are caused by insiders (those authorised to access network systems and data), and 21% from unknown sources. Even though the percentage of insider attacks is less than the external attacks, the consequences of insider attacks can be more severe. The survey indicated that 33% of the respondents consider insider attacks to be more costly and damaging. Consequently, insider attacks merit the same attention as external attacks.

Although there exists a large body of work in the literature to address the insider threat problem, little progress has been made due to the absence of clear answers to fundamental questions such as "What is an insider threat" [3]. Many definitions of insider and insider threat exist in the literature that complicated the research in insider threats as one solution to the insider problem might not be applicable to another insider problem [4]. We believe that the insider prob-

lem should be classified into several categories which can be defined, studied and solved independently, and later combined to solve the problem as a whole.

From the perspective of insider security, for a confidential file to be protected, it should only be shared with trusted individuals. Protecting the shared files from untrusted individuals who might strive to circumvent the protection is a dilemma for two reasons. First, each system has its own vulnerabilities and there is no system without vulnerabilities. Research efforts have proven that there is no system 100% secure against all deliberate attacks or misuses [5]. A brief look at the approach taken to protect commercial content, justifies this principle. Commercial content is protected by the use of Digital Rights Management systems that dictate how the content must be used by each individual. Although these systems are in place to protect commercial content, the content can still be obtained illegally in unprotected form. Second, the easiest way to circumvent any protection system used to protect confidential files is by exploiting the analog hole. All digital content must eventually be converted to human-perceptible form, known as the analog form, to be consumed by users. Once the digital content is converted to analog form, it will be in an unprotected form, and thus, it will be susceptible to unauthorised uses [6].

On the other hand, even if individuals are trusted to not violate the content policy deliberately, there is a chance of accidental violation. According to a survey conducted by Infosecurity Europe and PwC on 1,402 UK companies, 36% of the worst security breaches in the year were caused by inadvertent human error [7]. Also, AngloSec conducted a survey on 197 network, security, and compliance professionals, and found that the greatest security concern is employees accidentally jeopardising security through data leaks or similar errors [8]. Therefore, there should be appropriate level of protection that prevents accidental misuses on the shared files by trusted individuals who do not misuse the files intentionally.

The main purpose of this paper is to characterise the protection requirements and the activities of file sharing from the perspective of the insider threat problem. In addition, this paper proposes an approach to classify the insider threat problem and explicit definitions of insiders and insider threat problem to distinguish them from outsiders and external attacks. File sharing is treated in this paper as one category of the insider threat problem.

The rest of this paper is structured as follows. In Section 2 we review the literature and related work on insider threat and file sharing, respectively. In Section3, we propose a new approach for classifying the insider threat problem and focus on one category that is related to file sharing. In Section 4, we give our first contribution to characterising the protection required by the shared files against different types of insiders. In Section 5, we give our second contribution to characterising file sharing based on two factors: how and with whom files might be shared. In section 6, we discuss our current work. Finally, in Section 7, we conclude the paper with our future work.

2 RELATED WORK

2.1 Insider Security

Several definitions of insider and insider threat exist in the literature. Some authors have focused on the trust relationship when defining the term insider. For instance, RAND report [9] defined the insider as “an already trusted person with access to sensitive information and information systems”. Bishop [10] defined the insider as “a trusted entity that is given the power to violate one or more rules in a given security policy”. Other authors have focused on the abuse of given access privileges. For instance, Chinchani et al. [11] defined the insider as “legitimate users who abuse their privileges”. CERT report [12] defined the insider as “individuals who were, or previously had been, authorised to use the information systems they eventually employed to perpetrate harm”. Others defined the insider very broadly. For instance, Predd et al. [13] defined the insider as “someone with legitimate access to an organisation’s computers and networks”. RAND report [9] defined the insider again as “anyone with access, privilege, or knowledge of information sys-

tem and services". The former definition might include masqueraders who stole the credential of a legitimate user to get access to the computer or the network. The latter definition eliminates the need of trust and includes those who have knowledge of the system or the service even if they do not have access privileges. In 2008, a cross-disciplinary workshop on "Countering Insider Threats" [14] concluded that

> "an insider is a person that has been legitimately empowered with the right to access, represent, or decide about one or more assets of the organisation's structure"

With regard to insider threat, Predd et al. [13] defined insider threat as "an insider's action that puts an organisation or its resources at risk". RAND report [9] defined it as "malevolent (or possibly inadvertent) actions by an already trusted person with access to sensitive information and information systems". Hunker and Probst [15] defined it as "an insider threat is [posed by] an individual with privileges who misuses them or whose access results in misuse". The CERT Insider Threat Center's current definition of insider threats as follows:

> "A malicious insider threat to an organisation is a current or former employee, contractor, or other business partner who has or had authorised access to an organisation's network, system, or data and intentionally exceeded or misused that access in a manner that negatively affected the confidentiality, integrity, or availability of the organisation's information or information systems". [16]

Due to the differences and contradictory definition of insider and insider threats that complicated the problem to be solved, many authors are urging the community to establish a framework or taxonomy for distinguishing among different types of insider threats [14, 15]. They mentioned that each determining factor for an insider can be used for a taxonomy, for example based on distinctions between: Malicious and accidental threats; Doing something intentionally (for malice, or good reasons which nonetheless may result in damage) versus events that occur accidentally; Obvious and stealthy acts; Acts by masqueraders (e.g, an individual with a stolen password), traitors (malicious legitimate users) and naive or accidental use that results in harm; A combination of factors such as access types; aim or intentionality or reason for misuse; level of technical and the system consequences of insiders threats.

Bellovin [17] identified three different types of insider attack which are misuse of access, defence bypass, and access control failure. He stated that access control failure attacks can be prevented by purely technical means, while the other two attacks require combination of technical and non-technical means. Hunker and Probst [15] identified three different approaches, which current works in the field revolve around, to solve the insider threat problem. These approaches are technical approach, socio-technical approach, and sociological approach. The authors noted that technical approaches are focused on policy languages, access control and monitoring, while socio-technical approaches are focused on policy, monitoring and profiling, prediction, forensics and response work. Sociological approaches are focused on motivation, organisational culture, human factors and privacy and legal aspects. Silowash et al. [18] analysed cases of insider threat from the CERT insider threat database, which contains more than 700 cases of insider threat, and observed that malicious insider activities can be classified into four classes as follows.

- IT sabotage: an insider's use of IT to direct specific harm at an organisation or an individual. Example of this are destroying critical data, planting logical bomb to delete data at critical times, etc.
- Theft of IP: an insider's use of IT to steal IP from the organisation. This category includes industrial espionage involving outsiders. Examples of usually stolen IP assets are proprietary software, business plans, product details, and customer information.
- Fraud: an insider's use of IT for the unauthorised modification, addition, or deletion

of an organisation's data (not programs or systems) for personal gain, or theft of information that leads to an identity crime (e.g., identity theft or credit card fraud).
- Miscellaneous: cases in which the insider's activity was not for IP theft, fraud, or IT sabotage.

2.2 File Sharing

A wide variety of file sharing methods exist, and they differ from one another in the way that they allow users to control the what, how, and with whom to share [19]. Various studies have been conducted to investigate these properties.

Olson et al. [20] conducted a pilot study and a more formal survey to explore preferences for general information sharing by investigating what information people are willing to share, and with whom. The authors believe that their findings can provide guidance to the design of access control and interfaces. Voida et al. [19] conducted a survey and follow-up interviews at medium-sized research organisation to explore users' current practices and needs around file sharing. Based on their findings, they identified a number of critical characteristics of file sharing methods including universality, addressing, visibility, notification, and the differentiation between push- and pull-oriented sharing. They developed a prototype of a set of user interface features called a sharing palette, providing a platform for exploration and experimentation with new modalities of sharing.

Whalen et al. [21] conducted an online survey and follow up interviews at a medium-sized industrial research laboratory to address the issue of users' experience of file sharing and access control by gathering information on how and why people share files; the types of information shared; and how, when and why people limit access to those files. Based on the results of the study, the authors suggest guidelines to improve methods for appropriate content protection.

In another study, Whalen et al. [1] conducted a web-based survey at a medium-size university to investigate the fundamental issues regarding how files are shared and the difficulties encountered when managing files in collaborative environments. Their results show that there are a number of positive and negative factors that have an impact on peoples choice of file sharing methods. The positive factors are: the convenience and the ease of use of the method, the widespread availability of the method in order to reach all recipients, and the suitability of the method to the organisation or task at hand. The negative factors are: the limit on file space or file size, lack of access control or security features and the inability to reach all recipients.

Unlike the study of Voida et al. [19] and Whalen et al. [21, 1] which focused on subjects within a single organisation, all of whom had access to similar, established file sharing methods, Dalal et al. [22] conducted in-depth interviews with respondents across various domains in their homes, home offices, or in cafes where people worked to examine how file sharing and access controls are used, not used or circumvented in order to get work done. Analysing their results, they derived a set of design criteria for more effective file sharing system [22].

In contrast to previous studies which have focused on asking users themselves to report on how they share and protect files, Smetters and Good [23] conducted an automated survey of access control in a medium-sized corporation to collect behavioural data over time by analysing digital record of actual user behaviour as they believe that users' self-descriptions of their own behaviour can be incomplete or inaccurate. They used automated data mining to examine how users in a medium-sized corporation utilise two common access control features: the definition of access control groups, and the permissions settings, or ACLs, that users set on folders and documents. They found that access control polices which are applied by users to their content are quite complex. Based on the results of their study, they derived a number of suggestions for the design of both access control systems themselves, and the interfaces used to manage them [23].

Mazurek et al. [24] conducted semi-structured interviews with 33 non-technical computer users in 15 households to examine the current access control attitudes, needs, and practices of home

users when they share files inside and outside their homes. They found that people have complex policies that ever-changing over time which are inadequately addressed in current file sharing and access control methods; a finding supported by Olson et al. [20], Whalen et al. [21, 1], and Voida et al. [19]. Based on the results of their study, they have generated several guidelines for developers of access control systems aimed at home users.

Hart et al. [25] surveyed 23 blogging and social networking sites such as Blogger, Facebook, Flickr, YouTube, and MySpace to determine what access control and privacy features are currently available. They found that a lot of content-sharing sites provide primitive access control mechanisms which make a file entirely private or public while others allow more flexible control by offering private/friends/public access control model. The authors asserted that these models failed to support people's needs, and thus, proposed a method of access control for content-sharing sites that specify access control polices in terms of the content being mediated. Whalen et al. [26] pointed out that a potential solution for file sharing problems, such as exposing sensitive files accidentally, is to provide the user with clear information about file sharing settings and activities. Therefore, they explored existing research on awareness in collaborative environments, and used it to develop a framework for file sharing awareness. The authors used this awareness framework to develop a prototype for a file manager that facilitates file sharing by making sharing activity and settings more visible to the user.

Table 1 summarises the results of the above studies of file sharing with respect to answering the following fundamental questions: with whom the file is shared, what type of file is shared, how the file is shared and protected. The previous studies investigated these questions in details and provided valuable answers which could lead to better design of file sharing methods and access control models. However, the question of how the file is shared has been answered improperly. They merely answered the question of how people share their files by enumerating the methods of sharing files that people utilised. Such answers are applicable to the question of what methods people utilise to share their files rather than how the files are shared as we believe that the files can be shared in different ways using the same method.

3 OUR APPROACH

By surveying the previous work of insider security, we argue that the insider problem is significant and that no single definition can encompass the problem as a whole, which most researchers attempt to do. In the literature, insiders have always been defined and differentiated from outsiders by either being inside the network perimeter, trusted, authorised, or knowledgeable of the information system. Definitions based on these factors are either ambiguous or insufficient. To make progress and find a solution to the insider problem, we suggest that the problem should be classified into several categories which can be defined, studied and solved independently and which later can be combined to solve the problem as a whole. There are three factors which play an important role in classifying the insider problem which are: the type of activity that deals with an asset in an organisation; the type of asset that needs to be protected; and the type of attack that targeted the asset.

The activity. The activities are identified by the organisation for its partners, contractors, and employees to perform a particular job and might be different from one organisation to another. The activity will differentiate insiders from outsiders as an insider will be a person who is a legitimately given an activity by an organisation to perform a particular job. Therefore, the activity will lead to identifying who is the insider and what the insider is doing. The type of activity that insiders perform in an organisation are various and organisation-specific. Examples of activities that are given to insiders are file sharing, updating customer information, installing software to organisation's devices, setting up organisation's network, provisioning authorisation credentials to organisation's employees, etc.

Table 1: Summary of previous studies on file sharing

	With whom the file is shared	What type of file is shared	How the file is shared	How the file is protected
Olson et al. [20]	-The public, co-workers, managers and trusted co-workers, family and spouse.	-Email content, credit card number, transgression, work related documents, work email and desk phone number.	-	-
Voida et al. [19]	-Similar to Olson et. al.-With an average of 7 individuals or group	-34 different types of files e.g. business documents, paper drafts, music, ideas, schedules, and TV show	-Email (43%), shared network folders (16%) and posting content to a web site (11%)	-
Whalen et al. [1]	-Over 69% shared with two to four groups such as friends, family, research group, general public and colleagues. -25% shared with five to twenty groups.	-Only focused on sensitive files, such as email, personal financial or medical information, professional data or documents of an organisation, professional data or documents governed by law.	-Email (42%), shared network folders (14.7%), peer-to-peer program (10.3%) and file copy protocol (10.3%)	Various methods to control access to their sensitive files, some are technical (passwords, permissions) and others are socially-controlled such as hiding files.
Whalen et al. [21]	-	-	-Email (98%), shared network folder (55%), commercial content management systems (25%) and portable devices (25%)	Passwords; permissions/ access control lists; physical controls (e.g., safeguard in office or on person); encryption; obscurity (e.g., given files innocuous names, hidden directories); and deleting/relocating sensitive files.
Dalal et al. [22]	-With employees in professional sharing -With friends and family in personal sharing.	-In professional sharing: revolve around project work such as technical specifications, meeting minutes, and action items, proposals, reports.-In personal sharing: revolve around multimedia relational in nature such photograph and video.	Email (100%), - 80% used a wide variety of social software, such as wikis, blogs, social networking sites (including MySpace and Facebook) hosted services (such as Yahoo! Briefcase) public websites for sharing image and multimedia files (including Flickr and YouTube) and online forums and games.	-
Mazurek et al. [24]	-Family, friends, co-workers and strangers.	-Music, photo, video, private documents, school work, work files, and other personal documents.	-	-User accounts, password, encryption, limiting physical access to devices, and hide and delete sensitive files.

The asset. The assets that need to be protected are identified by an organisation based on a clear description of activities in the organisation, such that each activity will involve one or more assets to deal with. For example, if an activity in an organisation is employees sharing files with each other, the asset will be the file being shared which contains sensitive information. Another examples of activities and assets are an IT administrator who provisioning authorisation credentials to an organisation's employees where the asset here is the authorisation credential, a software developer who writes software scripts to an organisation computer where the asset can be the software itself or the computers that run the scripts, a network administrator who sets up the organisation's network and maintains it where the asset is the network.

Generally, the assets can be of three types which are the network which connects devices together, the devices which contains the data, or the data itself.

The attack. The attacks that targeted the asset can be generally of three types which are availability attacks, confidentiality attacks and integrity attacks, each of which can be performed in different ways which might require either physi-

cal security or IT security. Choosing which type of attacks to prevent is determined by the type of protection required for the chosen asset. For instance, if the asset is the network which needs to be available all the time, availability attacks should be prevented. On the other hand, if the asset is data that needs to be secret, confidentiality attacks should be prevented and so on. Therefore, the asset will determine which type of attacks should be prevented.

Based on these three factors, we can define the insider precisely as a person who is legitimately given an activity by an organisation to deal with the organisation's assets, and define the insider problem as particular types of attacks that performed by insiders on particular types of assets of an organisation during particular types of activities. Therefore, we can classify the insider problem into several categories based on these three factors such that each examination of different ways of performing a particular type of attacks targeted a particular type of asset during a particular type of activity will result in a unique class of the insider problem which can be defined, studied and solved independently. For example, one class of insider problem is examining the different ways of performing confidentiality attacks on sensitive files by employees when they share them with each other. Another class might be examining the different ways of performing availability attacks on an organisation's network by IT administrators when they maintain it, or examining the different ways of performing integrity attacks on customers information by employees when they update them etc.

Our concern in this paper is not to classify the insider problem thoroughly, rather we have provided an approach for such classification. However, we are interested in one class of the insider problem which is examining the different ways of performing confidentiality and integrity attacks on sensitive files by employees of an organisation when they share sensitive files with each other. Since file sharing is not only an activity that is performed by an organisation's employees but also it is an activity that can be performed among friends, family members, or colleagues, we will look at this class of the insider problem from broader perspective to include any individuals performing such activity. In other words, our focus will be on examining the different ways of performing confidentiality and integrity attacks on the shared files by the sharers when they share sensitive files with each other whether those sharers are employees, friends, family members, or colleagues. Although the asset is clearly identified in this class of the insider problem, the activity which is file sharing and the types of attacks on the shared file is still vague. Therefore, the following sections characterise file sharing and attacks that can be performed on shared files.

4 PROTECTING SHARED FILES

The shared files can be sensitive, which means that they need to be protected against unauthorised disclosure (confidentiality protection), unauthorised modification (integrity protection), or unauthorised withholding (availability protection). In this paper, we are only concerned with the confidentiality and integrity of the shared files. Protection of the shared files can be realised from two different angles: protecting the shared files while in transit, and protecting the shared files when they are received by the recipients. In this section we characterise attacks and misuses on the shared files initiated by different types of insiders.

4.1 Protecting the Shared Files in Transit

This type of protection prevents attacks on the file while it is transferred from the owner to the recipients. We divided these attacks into confidentiality attacks and integrity attacks as follows:

Confidentiality attacks. These attacks lead to the disclosure of the shared files to unauthorised users and can performed in two ways. First, someone eavesdrops or monitors the communication between the owner and the recipient to obtain knowledge about the files. We refer to such attacker *Interceptor*. Second, someone pretends to be the original recipient to deceive the owner and

obtain the files. We refer to such attacker *Masquerader*. Therefore, there should be two types of protections to prevent unauthorised disclosure of the shared files in transit as follows. Protecting the confidentiality of files from interceptor and protecting the confidentiality of files from Masquerader.

Integrity attacks. These attacks lead to unauthorised modification to the shared files by unauthorised users. The attacker in such attacks pretends to be the original owner to deceive the recipient by sending them files as if they were originated by the original owner. These files can either be an entirely new files or modified version of the original files. We refer to such attacker *Masquerader*. Therefore, there should be one type of protection to prevent unauthorised modification of the shared files in transit which is protecting the integrity of files from Masquerader.

4.2 Protecting the Shared Files at the Recipient

This type of protection prevents misuses on the file after it has been received by legitimate recipients. These misuses can affect the confidentiality and integrity of the files. Such misuses can be committed by three different entities which are *Malicious* recipients, *Naive* recipients or *Masqueraders*. *Malicious* recipients are untrusted legitimate recipients who deliberately misuse the shared files. *Naive* recipients are trusted recipients who accidentally misuse the shared files. *Masqueraders* are unauthorised users who accidentally acquire a device of a trusted legitimate user which contains the shared files and misuse these files. Therefore, misuses can be deliberate which are committed by *Malicious* or accidental which are committed by *Naive* or *Masqueraders*.

Protection against *Malicious* and *Naive* recipients are different from protection against *Masquerader*. *Malicious* and *Naive* recipients are already allowed to view the files, therefore, confidentiality of the files is achieved by not allowing them to redistribute the files to unauthorised users. Also, they may or may not be allowed to modify the files, therefore, integrity of the files is achieved by not allowing them to modify it in an unauthorised manner. On the other hand, *Masqueraders* are unauthorised users, therefore, confidentiality is achieved by not allowing them to view or redistribute the files, and integrity is achieved by not allowing them to modify the files.

Moreover, protection against *Naive* recipients is different from protection against *Malicious* recipients. The former is trusted to not redistribute or modify the files in an unauthorised manner, while the latter is untrusted and might strive to circumvent any protection to misuse the files. Therefore, we divided misuses which can be committed by the three entities broadly into confidentiality misuses and integrity misuses as follows:

Confidentiality misuses. Confidentiality misuses are those misuses which lead to the disclosure of the shared files to unauthorised users and which can be done in two ways. First, the shared file can be copied and sent to an unauthorised user through a file sharing method. Second, the device of a legitimate recipient which contains the shared file can be acquired by an unauthorised user, which we refer here to as a Masquerader, who discloses the shared files.

In the first case, the file can be redistributed in three ways. First, the file can be redistributed accidentally by a Naive legitimate recipient. Second, the file can be redistributed deliberately by a Malicious legitimate recipient. Third, the file can be redistributed accidentally by a Masquerader who found a device of a legitimate recipient unattended. In the second case, the file can be disclosed to Masqueraders in two ways. First, an unauthorised user steals the device of a Naive legitimate recipient. Second, a Malicious recipient lends his device to an authorised user.

Therefore, there should be five different types of protections to prevent unauthorised disclosure of the shared files at the recipients as follows. Protecting the confidentiality of files from accidental redistribution by Naive; protecting the confidentiality of files from accidental redistribution by Masquerader; protecting the confidentiality of files from deliberate redistributions by Malicious; protecting the confidentiality of files from accidental disclosure by Naive to Masqueraders; pro-

tecting the confidentiality of files from deliberate disclosure by Malicious to Masqueraders. Since the last two types of protection have similar impact which is disclosing the file to Masqueraders, we refer to them as protecting the confidentiality of files from accidental or deliberate disclosure to Masqueraders.

Integrity misuses. Integrity misuses are those misuses which lead to unauthorised modification to the shared files. Such unauthorised modification can be either modifying the shared files that do not allow any modification or modifying the shared file, that allowing partial modification, in an unauthorised manner. In both cases, the file can be modified in three ways. First, the file can be modified accidentally by a Naive legitimate recipient. Second, the file can be modified deliberately by a Malicious legitimate recipient. Third, the file can be modified accidentally by a Masquerader who found a device of a legitimate recipient unattended.

Therefore, there should be three different types of protections to prevent unauthorised modification of the shared files at the recipients as follows. Protecting the integrity of files from accidental modification by Naive; protecting the integrity of files from accidental modification by Masqueraders; protecting the integrity of files from deliberate modification by Malicious.

Below we classify the aforementioned protections into two types which are protection of files in transit and protection of the files at the recipients.

Protection of files in transit: this can be further divided into confidentiality protection and integrity protection.

- Confidentiality protection
  - Protecting the confidentiality of files in transit from *interceptor*
  - Protecting the confidentiality of files in transit from *Masquerader*
- Integrity protection
  - Protecting the integrity of files in transit from *Masquerader*

Protection of files at the recipients: this can be further divided into protection against accidental misuses when sharing with trusted recipient and protection against deliberate misuses when sharing with untrusted recipient.

Accidental misuse: this can be further divided into accidental misuse of confidentiality and accidental misuse of integrity.

- Accidental misuse of confidentiality:
  - Protecting the confidentiality of files at the recipients from accidental redistribution by *Naive*
  - Protecting the confidentiality of files at the recipients from accidental redistribution by *Masquerader*
  - Protecting the confidentiality of files at the recipients from accidental disclosure to *Masquerader*
- Accidental misuse of integrity:
  - Protecting the integrity of files at the recipients from accidental modification by *Naive*
  - Protecting the integrity of files at the recipients from accidental modification by *Masquerader*

Deliberate misuse: this can be further divided into deliberate misuse of confidentiality and deliberate misuse of integrity

- Deliberate misuse of confidentiality:
  - Protecting the confidentiality of files at the recipients from deliberate redistribution by *Malicious*
  - Protecting the confidentiality of files at the recipients from deliberate disclosure to *Masquerader*
- Deliberate misuse of integrity:
  - Protecting the integrity of files at the recipients from deliberate modification by *Malicious*

There are therefore eleven types of protections that might be required to protect files in transit or at the recipient, to protect files from trusted recipients or untrusted recipient etc. This classification of protection is depicted in Figure 1. Our focus in this paper is on protecting the files against accidental misuses when sharing the files with trusted individuals (i.e. against confidentiality and integrity misuses by Naive and Masquerader).

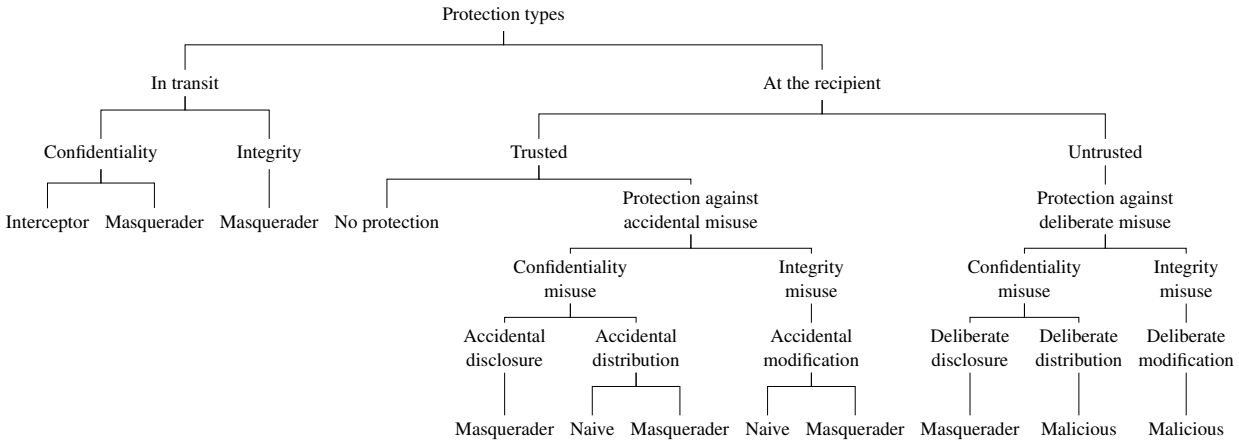

Figure 1: Types of protection of the shared files

5 CHARACTERISING FILE SHARING

Protecting a shared file against accidental misuses can be realised by controlling the operations that can be performed on the file (e.g. *Read, Write and Send*) in accordance to the policy of the file owner, such that an operation can be performed on a file only if it satisfies the policy of that file. For example, to protect the confidentiality of files at the recipients from accidental redistribution by *Naive*, *Send* operation can be aborted or allowed if the recipient is an authorised user. There are various restrictions that can be used to control file operations, such that each combination of them satisfies a file sharing scenario. Such restrictions can be deduced by observing how and with whom files can be shared, therefore, this section characterises the file sharing activity in general based on these two factors.

How the files can be shared. Shared files vary in terms of their sensitivity, some files are confidential while others are non-confidential. Differentiating confidential from non-confidential files is very helpful in characterising how files should be shared, because each of which is shared differently. People share confidential files with selected individuals while share non-confidential files with everyone. Available file sharing methods can either allow people to share files with selected individuals (suitable for confidential files) or allow people to share files with everyone (suitable for non-confidential files). A few file sharing methods provide both options. We will use the term *sharing* to refer to sharing files with selected individuals and the term *publishing* to share files with everyone.

Whether the files are published or shared, there are situations where the owner of the files might want to receive a new version of files with new changes made by the recipients, or the owner might want to update the files that the recipients have. We refer to such situations as sharing or publishing in *a dynamic mode*. It is adequate for a collaborative project where a group of members may work on a set of documents collectively. An example of a method that allows sharing in a dynamic mode is Dropbox where a file can be shared and updated by the owner or the recipients such that both can observe changes made to the shared files.

However, there are also situations where the owner of the files might not want to receive a new version of the published or shared files from the recipients or update the files that the recipients have. We refer to such situations as sharing or publishing in *a static mode*. An example of a method that allows sharing in a static mode is an email attachment where neither the owner nor the recipients can observe changes made on the shared files by others.

Furthermore, owners of files might want their shared files to be only viewed but not edited, edited but not viewed, viewed and edited, or not

viewed and not edited. We refer to such rights as *ReadOnly*, *WriteOnly*, *ReadWrite*, and *NoReadOrWrite*, respectively. The latter is useful when sharing the file with cloud storage providers (e.g. Dropbox).

Moreover, owners of files might want the rights which are ReadOnly, WriteOnly, and ReadWrite to be more restrictive. For instance, owners might require their shared files to be read *a limited number of times, a limited period of time (e.g. for three days starting from 2014/6/6), on a specific time (e.g. only Monday from 9am - 3pm), or at a specific location (e.g. only in London) which are abbreviated as Ln, Lp, St, and Sl, respectively*. Therefore, such restrictions might be used as parameters for the different rights mentioned above.

Based on mode of sharing and publishing, the rights, and the restrictions over these rights, we identify the different ways of how files can be shared as illustrated in Table 2. Each cell in the table marked with letter T indicates a way of how file is shared. Combination of multiple T in the same row illustrates how the file is shared in a complex scenario.

It should be noted that not all the rights are suitable for every sharing or publishing mode. For example, publishing or sharing in static mode with WriteOnly is not sensible as content that is written to the file will not be observed by anyone. Also, publishing in a static or dynamic mode and sharing in a dynamic mode with NoReadOrWrite is not sensible. This should be obvious as by not allowing anyone to read a file or write to it, there is no need for publishing the file in the first place, and there is no need to share a file in a dynamic mode, which allows changes made on files observed by both the owner or the recipients, since the file cannot be updated.

Also, not all the restrictions are suitable for every right such as restrictions over NoReadOrWrite. Note that NoReadOrWrite is only suitable for sharing in static mode (i.e. letter T highlighted with red) and restrictions over this right are not sensible. Also, restricting the rights on specific time and specific location for a published file is not sensible since such restrictions allow subset of the recipients who satisfied these restrictions to be able to exercise the different rights, while the file is intended to be shared with everyone. Therefore, such restrictions are only suitable over rights of sharing rather than publishing. Cells in the Table 2 marked with the letter F indicate a way of file sharing that is not useful in practice.

With whom the files can be shared. Files can be propagated in different ways, we use the following terminology to characterise the different ways of files propagation.

Terminology:

- $\bar{O}$: a particular owner of files who might or might not be known in advance.
- $I\bar{n}G$: a set of owners of files whom their numbers and identities are known in advance and share their files with each other.
- $\bar{G}$: a set of owners of files whom their numbers and identities are known in advance and do not share their files with each other.
- $\bar{M}$: a set of owners of files whom their number and identities are not known in advance and do not share their files with each other.
- O: a particular recipient who is known in advance.
- G: a set of recipients whom their number and identities are known in advance and whom receive the same copies of the shared files.
- M: a set of recipients whom their numbers and identities are not known and whom receive the same copies of the shared files.

In general, files can be shared either with O, G and M. However, the received files by the recipients who can be O, G and or M, might belong to $\bar{O}$, $I\bar{n}G$, $\bar{G}$ or $\bar{M}$. Therefore, including $I\bar{n}G$ as a category of sharing we have 11 different categories that can describe all the possible ways of with whom the file can be shared.

- $\bar{O} \rightarrow O$ (OneToOne): This describes a situation when a particular owner of files wants to share his files with a particular recipient who is known in advance. For example, Alice wants to share her file only with Bob but no one else.
- $\bar{O} \rightarrow G$ (OneToGroup): This describes a situation when a particular owner of files wants to share his files with a set of recipi-

Table 2: How files can be shared

How file can be shared		ReadOnly				WriteOnly				ReadWrite				NoreadOrWrite			
		Ln	Lp	St	Sl	Ln	Lp	St	Sl	Ln	Lp	St	Sl	Ln	Lp	St	Sl
Publish	Static	T	T	F	F	F	F	F	F	T	T	F	F	F	F	F	F
	Dynamic	T	T	F	F	T	T	F	F	T	T	F	F	F	F	F	F
Share	Static	T	T	T	T	F	F	F	F	T	T	T	T	T	T	T	T
	Dynamic	T	T	T	T	T	T	T	T	T	T	T	T	F	F	F	F

ents whom their numbers and identities are known in advance, and whom receive the same copies of the shared files. For example, Alice wants to share her file only with her colleagues Bob, Carol, and Dave but no one else.

- $\bar{O} \rightarrow M$ (OneToMany): This describes a situation when a particular owner of files wants to share his files with a set of recipients whom their numbers and identities are not known, and whom receive the same copies of the shared files. For example, Alice wants to share her file with everyone on the internet regardless of whom they are.
- $I\bar{n}G$ (InGroup): This describes a situation when owners of files whom their numbers and identities are known in advance want to share their files with each other. For example, Alice, Bob, and Carol want to share their files only with each other but no one else.
- $I\bar{n}G \rightarrow O$ (InGroupToOne): This describes a situation when a set of owners of files whom their numbers and identities are known in advance and share their files with each other, want to share their shared files with a particular recipient who is known in advance. For example, Alice, Bob and Carol who are sharing their files with each other want to share these shared files only with their colleague Dave but no one else.
- $I\bar{n}G \rightarrow G$ (InGroupToGroup): This describes a situation when a set of owners of files whom their numbers and identities are known in advance and share their files with each other, want to share their shared files with a set of recipients whom their numbers and identities are known in advance, and whom receive the same copies of the shared files. For example, Alice, Bob and Carol who are sharing their files with each other want to share these shared files only with their colleagues in the same department but no one else.
- $I\bar{n}G \rightarrow M$ (InGroupToMany): This describes a situation when a set of owners of files whom their numbers and identities are known in advance and share their files with each other, want to share their shared files with a set of recipients whom their numbers and identities are not known and whom receive the same copies of the shared files. For example, Alice, Bob and Carol who are sharing their files with each other want to share these shared files with everyone on the internet regardless of whom they are.
- $\bar{G} \rightarrow O$ (GroupToOne): This describe a situation when a set of owners of files whom their numbers and identities are known in advance and whom do not share their files with each other, want to share their files with a particular recipient who is known in advance. For example, Alice, Bob and Carol who work in the same company want to share their files only with Dave who is their employer but not with each other or anyone else.
- $\bar{G} \rightarrow G$ (GroupToGroup): This describe a situation when a set of owners of files whom their numbers and identity are known in advance and whom do not share their files with each other, want to share their files with a set of recipients whom their numbers and identities are known in advance, and whom receive the same copies of the shared files. For example, Alice, Bob and Carol who work in the same company want to share their files only with employees of the HR department but not with each other or anyone else.

- $\bar{M} \rightarrow O$ (ManyToOne): This describes a situation when a set of owners of files whom their numbers and identities are not known in advance and whom do not share their files with each other, want to share their files with a particular recipient who is known in advance. For example, applicants to a particular job want to share their documents files only with Alice who is the employer but no one else.
- $\bar{M} \rightarrow G$ (ManyToGroup): This describes a situation when a set of owners of files whom their numbers and identities are not known in advance and whom do not share their files with each other, want to share their files with a set of recipients whom their numbers and identities are known in advance, and whom receive the same copies of the shared files. For example, applicants to a particular job want to share their documents files only with Alice, Bob and Carol, who are the employees responsible for recruiting new staff, but no one else.

Figure 2 illustrates these categories and classifies them to either publish or share. Therefore, these categories can substitute the term share or publish in Table 2 accordingly to provide a comprehensive characterisation of file sharing that can describe all possible ways of how and with whom files can be shared. Note that we excluded situations that do not make sense such as $M' \rightarrow M$ and $G' \rightarrow M$, since any of the owners can be of the recipients and vise versa.

6 DISCUSSION

Our characterisation of the protections required by the shared files illustrates the different ways of how files can be misused by different types of insiders. This characterisation makes it clear which type of insider misuse needs to be prevented in a particular sharing scenario. For instance, misuses by Masqueraders need not to prevented if the machine containing the file resides in a locked room where unauthorised users cannot access. Also, deliberate misuses by Malicious insiders need not to be prevented if the file is shared with trusted recipients. A major advantage of this characterisation is the avoidance of the chaos exists in the literature with respect to distinguishing insider attacks from external attacks, and between insiders attacks themselves. We listed different protections requirements to prevent different insiders misuses so that one can select the desired protection requirements for a particular sharing scenario and develop a mechanism to enforce it.

Our characterisation of file sharing can be used to specify different policies to control operations on files to meet the protection requirements of the shared files discussed in Section 5.2, particularly, protecting the shared files against accidental misuses. For example, our characterisation of how the files can be shared is useful to specify policies that are concerned with Read and Write operations to protect the file against accidental modification by Naive or Masquerader and accidental disclosure to Masquerader. To protect the file from accidental modification by Naive, the file should be shared with ReadOnly or NoReadOrWrite right, so that Write operation cannot be performed on the file. To protect the file from accidental disclosure to and accidental modification by a Masquerader, rights of the file to perform Read and Write operations should be more restricted. For example, rights such as ReadOnly and ReadWrite can be restricted to be exercised only on a specific time (e.g. working hours) or location (e.g. office building), so that whether the device of a legitimate user is stolen from home or found unattended outside working hours, Read and Write operations cannot be performed.

On the other hand, our characterisation of with whom the files can be shared is useful to specify policies that are concerned with Send operation to protect the file against accidental redistribution by Naive or Masquerader. For example, a file that needs to be shared with a particular user should be categorised as OneToOne, so that Send operation can only be performed if the number of the recipients is one and the recipient is an authorised user.

Such policies can be specified and modelled, for example, as an Access Control Matrix which is a conceptual model that specifies the rights that each subject possesses for each object. Each row of the matrix corresponds to a subject and each

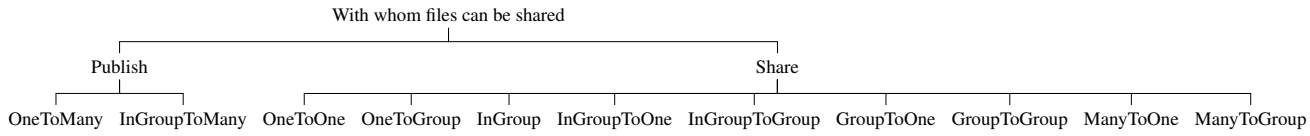

Figure 2: With whom files can be shared

column to an object. Each cell of the matrix specifies a set of rights for the subject in the row to the object in the column. For example as shown in Table 3, UserA is only allowed to perform read and send operations on File1, provided that read operation has not been performed more than three times and the user location while performing it should be London, and provided that send operation is performed to send the file to a particular person who is specified in the policy. UserB is only allowed to perform write and send operations on File1, provided that write operation is performed during the hours 9am-5pm and send operation is performed to send the file to a particular group whom are specified in the policy. With respect to File2, UserA is only allowed to perform read and write operation without any restriction. UserB is not allowed to perform any operation, however, he can keep the file, while UserC cannot access the file at all.

Table 3: Access Control Matrix for file sharing

	Objects	
Subjects	File1	File2
UserA	$R\bar{W}$(Ln:3, Sl:London); OneToOne	RW(-); -
UserB	$\bar{R}W$(St:9AM-5PM); OneToGroup	$\bar{R}\,\bar{W}$(-); -
UserC	RW(Lp:2014-2015); OneToMany	-

7 CONCLUSION and FUTURE WORK

This paper has studied one category of the insider threat problem that is concerned with file sharing. We have given a classification of the insider threat problem and defined the insider and insider threat problem precisely. We characterised the protection requirements of the shared files against different type of insiders. We also characterised file sharing activities, motivated from the development of extensive use-case scenarios.

We are currently working to develop a mechanism to enforce the different protections types against accidental misuses on the shared files (See Figure 1). Since software is the major cause of many breaches in security, a promising approach to create a secure software is to write it in a type-safe programming language. Therefore, we take a type-based approach to enforce security policies which is a language-based technique to provide security in programs. Generally, system security requirements can be divided into two concerns which are access control and information flow control. The former places restrictions on the release of the resources, while the latter on its propagation. Access control requirement can be specified by our characterisation of how files can be shared while information flow control can be specified by our characterisation of with whom files can be shared.

Next, we formalise our approach using a type system to formally analyse access control and information flow whereby our characterisation of file sharing are represented as security type annotations and access control and information flow polices are enforced through type checking to prevent accidental misuses.

REFERENCES

[1] T. Whalen, E. Toms, and J. Blustein. File sharing and group information management. Workshop on Personal Information Management (PIM 2008), 2008.

[2] Software Engineering Institute. 2011 CyberSecurity Watch Survey. Software Engineering Institute, Carnegie Mellon University, 2011.

[3] J. Hunker. Taking Stock and Looking Forward - An Outsider's Perspective on the Insider Threat. In S. J. Stolfo, S. M. Bellovin, A. Keromytis, S. Hershkop, S. W. Smith, and S. Sinclair, editors, *Insider Attack and Cyber Security - Beyond the Hacker*, volume 39 of *Advances in Information Security*, pages 195–213. Springer, 2008.

[4] Matt Bishop and Carrie Gates. Defining the insider threat. In *In Proceedings of the 2008 Cyber Security and Information Infrastructure Research Workshop*, 2008.

[5] K. Scarfone and P. Mell. The common configuration scoring system (ccss): Metrics for software security configuration vulnerabilities. Technical Report 7502, National Institute of Standards and Technology, December 2010.

[6] S. Haber, B. Horne, J. Pato, T. Sander, and R. E. Tarjan. If piracy is the problem, is drm the answer? In E. Becker, W. Buhse, D. Gnnewig, and N. Rump, editors, *Digital Rights Management*, volume 2770 of *LNCS*, pages 224–233. Springer, 2003.

[7] Infosecurity Europe and PwC. 2013 information security breaches survey. Technical report, Department for Business, Innovation & Skills, April 2013.

[8] AlgoSec. The state of network security 2013: Attitudes and opinions, 2013.

[9] R. Anderson and R. Brackney. Understanding the insider threat. In *Proceedings of a March 2004 Workshop. Prepared for the Advanced Research and Development Activity (ARDA).* `http://www.rand.org/publications/CF/CF196`, 2004.

[10] M. Bishop. Position: "insider" is relative. In *Proceedings of the 2005 workshop on New security paradigms*, NSPW '05, pages 77–78. ACM, 2005.

[11] R. Chinchani, A. Iyer, H. Q. Ngo, and S. Upadhyaya. Towards a theory of insider threat assessment. In *Proceedings of the 2005 International Conference on Dependable Systems and Networks*, DSN '05, pages 108–117, Washington, DC, USA, 2005. IEEE Computer Society.

[12] K. Michelle and E. Kowalski. Insider Threat Study: Computer System Sabotage in Critical Infrastructure Sectors. May 2005.

[13] J. Predd, S. L. Pfleeger, J. Hunker, and C. Bulford. Insiders behaving badly. *IEEE Security & Privacy*, 6(4):66–70, 2008.

[14] C. W. Probst, J. Hunker, M. Bishop, and D. Gollmann. 08302 summary – countering insider threats. In Matt Bishop, Dieter Gollmann, Jeffrey Hunke, and Christian W. Probst, editors, *Countering Insider Threats*, number 08302 in Dagstuhl Seminar Proceedings, Dagstuhl, Germany, 2008.

[15] J. Hunker and C. W. Probst. Insiders and insider threats: An overview of definitions and mitigation techniques. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, 2(1):4–27, 2011.

[16] CERT. The CERT Insider Threat Center @ONLINE, April 2013.

[17] S. M. Bellovin. The Insider Attack Problem Nature and Scope. In S. J. Stolfo, S. M. Bellovin, A. Keromytis, S. Hershkop, S. W. Smith, and S. Sinclair, editors, *Insider Attack and Cyber Security - Beyond the Hacker*, volume 39 of *Advances in Information Security*, pages 1–4. Springer, 2008.

[18] G. Silowash, D. Cappelli, A. Moore, R. Trzeciak, T. J. Shimeall, and L. Flynn. Common Sense Guide to Mitigating Insider Threats 4th Edition, December 2012.

[19] Stephen Voida, W. Keith Edwards, Mark W. Newman, Rebecca E. Grinter, and Nicolas Ducheneaut. Share and share alike: exploring the user interface affordances of file sharing. In Rebecca E. Grinter, Tom Rodden, Paul M. Aoki, Edward Cutrell, Robin Jeffries, and Gary M. Olson, editors, *CHI*, pages 221–230. ACM, 2006.

[20] J. S. Olson, J. Grudin, and E. Horvitz. A study of preferences for sharing and privacy. In *Proceedings of CHI 05*, pages 1985–1988. ACM Press, 2005.

[21] T. Whalen, D. Smetters, and E. F. Churchill. User experiences with sharing and access control. In *In CHI 06: CHI 06 extended abstracts on Human factors in computing systems*, pages 1517–1522. ACM Press, 2006.

[22] B. Dalal, L. Nelson, D. Smetters, N. Good, and A Elliot. Ad-hoc guesting: when exceptions are the rule. In *Proceedings of the 1st Conference on Usability, Psychology, and Security*, UPSEC'08, pages 9:1–9:5, Berkeley, CA, USA, 2008. USENIX Association.

[23] D. K. Smetters and N. Good. How users use access control. SOUPS '09. ACM, 2009.

[24] M. L. Mazurek, J. P. Arsenault, J. Bresee, N. Gupta, I. Ion, C. Johns, D. Lee, Y. Liang, J. Olsen, B. Salmon, R. Shay, K. Vaniea, L. Bauer, L. F. Cranor, G. R. Ganger, and M. K. Reiter. Access control for home data sharing: Attitudes, needs and practices. In *CHI 2010: Conference on Human Factors in Computing Systems*, CHI '10, pages 645–654, New York, NY, USA, 2010. ACM.

[25] M. Hart, R. Johnson, and A. Stent. More content-less control: Access control in the web 2.0. *Control*, pages 1–3, 2006.

[26] T. Whalen, E. G. Toms, and J. Blustein. Information displays for managing shared files. In *Proceedings of the 2nd ACM Symposium on Computer Human Interaction for Management of Information Technology*, CHiMiT '08, pages 5:1–5:10, New York, NY, USA, 2008. ACM.